S V E U Č I L I Š T E U S P L I T U
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

Katarina Babić

DETEKCIJA LAŽIRANOG SIGNALA U GLOBALNOM NAVIGACIJSKOM SATELITSKOM SUSTAVU

DOKTORSKA DISERTACIJA

Split, svibanj 2025.

S V E U Č I L I Š T E U S P L I T U
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I BRODOGRADNJE

Katarina Babić

Detekcija lažiranog signala u globalnom navigacijskom satelitskom sustavu

DOKTORSKA DISERTACIJA

Split, svibanj 2025.

Doktorska disertacija je izrađena na Zavodu za elektroniku i računarstvo, Fakulteta elektrotehnike, strojarstva i brodogradnje

Mentor: prof. dr. sc. Dinko Begušić
Rad br.

PODACI ZA BIBLIOGRAFSKU KARTICU

Ključne riječi: globalni navigacijski satelitski sustav, napad lažiranjem, detekcija, strojno učenje, radio frekvencijski otisak
Znanstveno područje: elektrotehnika i informacijska tehnologija
Znanstveno polje: elektrotehnnika, računarstvo
Znanstvena grana: telekomunikacije i informatika, obrada informacija
Institucija na kojoj je rad izrađen: Sveučilište u Splitu, Fakultet elektrotehnike, strojarstva i brodogradnje
Mentor rada: prof. dr. sc. Dinko Begušić
Broj stranica: 156
Broj slika: 70
Broj tablica: 11
Broj korištenih bibliografskih jedinica: 138

Povjerenstvo za ocjenu doktorske disertacije:

1. Prof. dr. sc. Josip Lorincz, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
2. Prof. dr. sc. Toni Perković, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
3. Prof. dr. sc. Joško Radić, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
4. Prof. dr. sc. Maja Stella, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
5. Izv. prof. dr. sc. Miljenko Mikuc, Fakultet elektrotehnike i računarstva (FER), Zagreb

Povjerenstvo za obranu doktorske disertacije:

1. Prof. dr. sc. Josip Lorincz, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
2. Prof. dr. sc. Toni Perković, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
3. Prof. dr. sc. Joško Radić, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
4. Prof. dr. sc. Maja Stella, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split
5. Izv. prof. dr. sc. Miljenko Mikuc, Fakultet elektrotehnike i računarstva (FER), Zagreb

Disertacija obranjena dana: rujan 2025.

Detekcija lažiranog signala u globalnom navigacijskom satelitskom sustavu

Sažetak:

U današnjem modernom svijetu, usluge navigacije i pozicioniranja postale su svakodnevnica bez koje je gotovo nemoguće zamisliti današnji svijet. Jedna od najvažnijih infrastruktura u današnjem modernom svijetu je globalni navigacijski satelitski sustav GNSS (engl. *Global Navigation Satellite System*) koji omogućava usluge pozicioniranja, navigacije i vremenske reference. Povećano zanimanje za razvoj i integraciju usluga navigacije i pozicioniranja u širok raspon prijamnika čini ih osjetljivima na različite sigurnosne napade kao što su napadi ometanja (engl. *jamming*) i lažiranja (engl. *spoofing*). Napad lažiranjem podrazumijeva namjerno odašiljanje lažnih satelitskih signala s namjerom da prijamnik lažne signale pogrešno protumači kao autentične u svrhu preuzimanja navigacijskog sustava prijamnika i lažiranja lokacije prijamnika. Usluge za pozicioniranje i navigaciju dostupne su u skoro svim, čak i najmanjim uređajima i zato kako bi se ublažile smetnje u prijamnicima sustava GNSS i zajamčila pouzdana rješenja, otkrivanje interferencija i klasifikacija tipa signala postaje od iznimne važnosti. Stoga, u ovoj doktorskoj disertaciji predlaže se integrirani pristup za detekciju napada lažiranjem i klasifikaciju tipa signala korištenjem metoda radio frekvencijskog otiska RFF (engl. *Radio Frequency Fingerprinting*) i strojnog učenja u pre-korelacijskoj fazi. Motivacija za primjenu metoda radio frekvencijskog otiska proizlazi iz činjenice da navedene metode nisu obrađene i primjenjivane u području detekcije lažiranih signala u sustavu GNSS te ovo istraživanje daje značajan doprinos zajednici sustava GNSS. Rezultati detekcije lažiranih signala i klasifikacije tipa signala za dvije satelitske konstelacije GPS i Galileo na skupu podataka OAKBAT prikazani su za različite tipove ulaznih klasifikacijskih podataka koje su korištene u modelima strojnog učenja. Performanse modela su evaluirane korištenjem standardnih parametara. Visoka točnost klasifikacije tipa signala pokazuje da je predloženi pristup pouzdan za detekciju lažiranih signala.

Ključne riječi:
Sustav GNSS, napad lažiranjem, metoda radio frekvencijskog otiska, detekcija, klasifikacija, strojno učenje, obrada signala, spektrogram, diskretna valićna transformacija.

Spoofing signal detection in global navigation satellite system

Abstract:

In today's modern world, navigation and positioning services have become an everyday necessity, without which it is almost impossible to imagine today's world. One of the most important infrastructures in today's modern world is the Global Navigation Satellite System (GNSS), which provides positioning, navigation and time reference services. Increased interest in the development and integration of navigation and positioning services in a wide range of receivers makes them vulnerable to various security attacks such as jamming and spoofing attacks. A spoofing attack involves the deliberate transmission of fake satellite signals with the intention that the receiver misinterprets the fake signals as authentic ones in order to take over the receiver's navigation system and spoof the receiver's location. Positioning and navigation services are available in almost all, even the smallest devices, and therefore, in order to mitigate interference in GNSS receivers and guarantee reliable solutions, interference detection and signal type classification becomes of utmost importance. Therefore, in this doctoral dissertation, an integrated approach is proposed for the detection of spoofing attacks and the classification of signal type using radio frequency fingerprinting (RFF) methods and machine learning in the pre-correlation phase. The motivation for the application of radio frequency fingerprinting methods stems from the fact that the aforementioned methods have not been processed and applied in the field of detection of fake signals in the GNSS system, and this research makes a significant contribution to the GNSS community. The results of fake signal detection and signal type classification for two satellite constellations GPS and Galileo on the OAKBAT dataset are presented for different types of input classification data that were used in machine learning models. Model performance was evaluated using standard parameters. The high accuracy of signal type classification shows that the proposed approach is reliable for detecting fake signals.

Keywords:
GNSS system, spoofing attack, radio frequency fingerprinting method, detection, classification, machine learning, signal processing, spectrogram, discrete wavelet transform.

Zahvala

Ovu doktorsku disertaciju posvećujem svom ocu Ivanu koji je moj životni uzor u svemu i koji mi je bio velika potpora kroz cijeli doktorski studij sa svojim savjetima i konstruktivnim komentarima. Znam da je ovaj moj uspjeh i njegovo ostvarenje sna.

Sadržaj

Popis tablica

Popis slika

1. Uvod

Globalni navigacijski satelitski sustav GNSS (engl. *Global Navigation Satellite System*), koji čine četiri globalna navigacijska sustava: GPS (engl. *Global Positioning System*), GLONASS (*Navigazionnaya Sputnikovaya Sistema*), Galileo (engl. *European Global Navigation Satellite System*) i BeiDou (engl. *Chinese Global Navigation Satellite System*) [1], predstavlja jednu od najvažnijih infrastruktura u današnjem modernom svijetu koji omogućuje mnoge kritične usluge koje zahtijevaju pouzdanost primljenih signala. Neke od tih usluga su pozicioniranje, navigacija i sinkronizacija. Primjerice, stabilna i precizna sinkronizacija je od ključne važnosti u mobilnim mrežama za uspješno povezivanje baznih postaja i prijenos podataka u stvarnom vremenu te za usluge navigacije i pozicioniranja. Zbog svoje rasprostranjene upotrebe, sustav GNSS je veoma izložen različitim prijetnjama [1], [12] od kojih je najčešći napad lažiranjem (engl. *spoofing attack*) koji je i glavni predmet istraživanja ove doktorske disertacije. Osim napada lažiranjem, interferencije koje se javljaju u prijamnicima sustava GNSS su višestazno prostiranje GNSS signala (engl. *multipath*) i ometanje GNSS signala (engl. *jamming*).

Napad lažiranjem temelji se na odašiljanju lažnih signala u svrhu zavaravanja prijamnika i preuzimanja njegovog navigacijskog sustava. Pošto je snaga primljenog satelitskog signala jako slaba, zbog utjecaja radiofrekvencijskih smetnji, može doći do smanjene točnosti pozicioniranja i određivanja vremena ili čak do potpunog nedostatka rješenja za navigaciju. Upravo zbog toga, prijamnik uzima one signale koji imaju veću snagu i na temelju njih izračuna svoj položaj. Laka dostupnost softverski definiranih radija povećava održivost izvođenja takvih napada. Uređaji koji su najranjiviji na napade lažiranjem su mobilni telefoni koji se danas najviše koriste za usluge navigacije. Uspješnost izvođenja napada lažiranjem ovisi i o vrsti napada, primjerice, pojednostavljeni napad lažiranjem je najjednostavniji za izvesti korištenjem softverski definirang radija kao što je HackRF One. Veća pozornost ovim napadima u zajednici sustava GNSS posvećena je u otvorenoj literaturi 2008. tek nakon što su Humphreys i kolege razvili sustav za izvođenje napada lažiranjem signala sustava GPS te ga uspješno testirali na komercijalnom standardnom prijamniku [24]. Ovi napadi su opasni i iz sigurnosnog aspekta prilikom korištenja u vojnim svrhama za preusmjeravanje dronova, aviona, brodova, itd.

Korištenje prijamnika sustava GNSS koji su otporni na napad lažiranjem ključno je za sigurno pozicioniranje, navigaciju, vrijeme i sinkronizaciju. Stoga, učinkovita detekcija na-

pada lažiranjem i klasifikacija tipa signala postaje od iznimne važnosti. Iako su u tu svrhu predložene različite metode za detekciju, to je još uvijek važna tema istraživanja u ovom području. Posljednjih godina, metode strojnog učenja sve više imaju svoju primjenu u detekciji lažnih signala. Neke od metoda za detekciju ovih napada koje su još uvijek u svojim začecima u području sustava GNSS te nisu dovoljno istražene su metode radio frekvencijskog otiska RFF (engl. *Radio Frequency Fingerprinting*).

Glavni cilj ovog istraživanja je primjena metoda radio frekvencijskog otiska u kombinaciji s modelima strojnog učenja za učinkovitu detekciju lažnih signala i klasifikaciju tipa signala.

1.1. Pregled dosadašnjeg istraživanja i motivacija

Kao što je već navedeno, autori u radu [24] po prvi put spominju izvedbu napada lažiranjem i tek tada se više pozornosti počelo posvećivati ovim napadima.U [13] i [47], autori detaljno prikazuju vrste napada lažiranjem i obrambene tehnike koje se razmatraju ili razvijaju. Strategija za detekciju napada lažiranjem na kriptografski zaštićene GNSS signale je prikazana u [77]. U dosadašnjim istraživanjima mnogi istraživači su se bavili različitim metodama za detekciju napada lažiranjem. Najjednostavnija metoda za detekciju napada lažiranjem je metoda koja prati snagu signala. Ova metoda je osnova za mnoge druge metode detekcije napada lažiranjem [48]. Druge popularne metode detekcije su metode podatkovnih bitova: praćenje i analiza poruka Nacionalne udruge za pomorsku elektroniku NMEA (engl. *National Marine Electronics Association*) [32], [33], praćenje i usporedba vremena dolaska ToA (engl. *Time of Arrival*) [34] i smjera dolaska DoA (engl. *Direction of Arrival*) [35], [36], [37]. Metode obrade signala druga su kategorija kojoj pripadaju metode temeljene na snazi [78], [75], metode temeljene na anteni [76], [64] i metode temeljene na distorziji korelacije [41] i praćenju korelacijskog vrha [17], [18], [42], [43].

Detekcija napada lažiranjem na temelju praćenja vrijednosti omjera snage signala nosioca i šuma (C/N_0) prikupljenih na pametnom telefonu prikazana je u [104]. Rezultati eksperimenta su pokazali da su normalne vrijednosti C/N_0 od 0 dB-Hz do 35 dB-Hz dok su za vrijeme napada lažiranjem te vrijednosti 50 dB-Hz i više. U radu [68], detekcija lažnih signala se provodi na temelju C/N_0 vrijednosti tijekom promatranog vremenskog razdoblja i Pearsonovog koeficijenta korelacije. Rezultati mjerenja potvrđuju da su vrijednosti C/N_0 za vrijeme napada lažiranjem veće od 35 dB-Hz. U slučaju napada lažiranjem veća je korelacija između C/N_0 vrijednosti za dva različita satelita PRN1 i PRN3 te Pearsonov koeficijent korelacije iznosi 0.99. S druge strane, u uvjetima bez napada lažiranjem, niska je korelacija između C/N_0 vrijednosti za satelite PRN1 i PRN3 te je Pearsonov koeficijentom -0.76 zbog različitih trendova. Autori u [29] fokusiraju se na klasifikaciju nekoliko vrsta signala: autentični, lažni, višestazni i ometajući te pokazuju da je klasifikacija temeljena na snazi i distorziji korelacije (degradacija oblika funkcije korelacije između autentičnog i lažnog sig-

nala) najbolja kod detekcije namjernih smetnji kao što je napad lažiranjem. Lažni signali lako se razlikuju od autentičnih signala zbog njihove visoke prosječne snage i visokog stupnja distorzije korelacije.

Rad [17] fokusiran je na otkrivanje napada lažiranjem s malim kašnjenjem korištenjem K-najbližih susjeda metode strojnog učenja. Ključni korak za detekciju lažnog signala je detekcija vrhova signala korelacije. Detekcija se temelji na otkrivanju lažnog signala procjenom broja vrhova koji premašuju unaprijed postavljeni prag kada prijamnik uhvati signal. Ako postoji samo autentični signal u primljenom signalu, vrijednost samo jednog vrha signala korelacije premašit će unaprijed postavljeni prag. Ukoliko postoje lažni signali, tada postoje dva ili više vrhova signala korelacije koji su veći od postavljenog praga i ova metoda otkrivanja lažnih signala valjana je kada je fazna razlika između lažnog i autentičnog signala velika, tj. veća od dva čipa. Kada je fazna razlika između autentičnog i lažnog signala, primjerice, jedan čip, broj vrhova je i dalje jedan, pa je teško detektirati lažne signale. Eksperimentalni rezultati provedeni u ovom radu pokazali su da predloženi algoritam može otkriti lažne signale s kašnjenjem većim od 0.6 čipova u odnosu na autentične signale i da ima visoku točnost. Autori u [18] pokazuju da generativna kontradiktorna mreža GAN (engl. *Generative Adversarial Network*) može doseći više od 98% točnosti kada fazna razlika između lažnog i autentičnog signala prelazi 0.5 čipova i može se primijeniti na situacije u kojima je lažni signal visoko sinkroniziran s autentičnim signalom.

Budući da tradicionalne metode praćenja kvalitete signala SQM (engl. *Signal Quality Monitoring*) imaju nisku točnost detekcije, autori u [43] predstavljaju novu poboljšanu SQM metodu koja se temelji na primjeni statističkog Kolmogorov - Smirnovljevog testa. Ova metoda testirana je na scenarijima skupa podataka Texas Spoofing Test Battery (TEXBAT) [20] i rezultati pokazuju poboljšanje točnosti detekcije napada lažiranjem za različite razine snage signala.

Metode strojnog i dubokog učenja nedavno su postale najpopularniji pristupi za otkrivanje napada lažiranjem u sustavu GNSS. Razne studije analiziraju izvedbu različitih metoda strojnog učenja. Autori u [67] pokazuju da se modeli klasificiranja i regresijskog stabla odlučivanja ističu u odnosu na druge metode strojnog učenja za klasifikaciju GPS signala. Slično tome, rezultati u [25] pokazuju da metoda potpornih vektora SVM (engl. *Support Vector Machine*) daje najbolje rezultate za detekciju lažnih GPS signala. Međutim, u [26], usporedba nekoliko metoda pokazuje da metoda k-najbližih susjeda KNN (engl. *K-Nearest Neighbors*) nadmašuje SVM. Rezultati u [26], [45], [46], [40] pokazuju da je SVM metoda pouzdan pristup za detekciju lažnih signala. Semanjski i ostali u [46] nadopunjuju eksperimente i rezultate dobivene u [45]. Uz laboratorijski generirane skupove podataka lažnih signala korištene u [45], skupovi lažnih signala u realnom vremenu su dodani u [46] u fazi treniranja modela. Točnost SVM metode se poboljšala s 75.82% na 95.54% uz korištenje svih parametara. Metoda višeparametarske detekcije, primijenjena na skupovima podataka TEXBAT i Oak Ridge Spoofing and Interference Test Battery (OAKBAT) [22], koju su pred-

ložili Chen i ostali u [40] značajno poboljšava detekciju lažnih signala u usporedbi s tradicionalnim pristupima. Autori u [103] uspoređuju rezultate klasifikacije za dva skupa podataka: vlastiti i SatGrid [107] te pokazuju da SVM i neuralne mreže imaju najbolje rezultate klasifikacije tipa signala. Neuralne mreže NN (engl. *Neural Networks*) kao pristup dubinskom učenju također se često koriste kao što je prikazano u studijama [49], [50], [51]. Autori u [52] koriste arhitekturu konvolucijske neuralne mreže CNN (engl. *Convolutional Neural Network*) za detekciju različitih interferencija uključujući i napad lažiranjem te postižu visoku točnost od 99.69% u klasifikaciji interferencija.

Iako postoje različite metode za otkrivanje napada lažiranjem, značajni istraživački napori usmjereni su na primjenu i razvoj novih metoda. Jedna od metoda koja nije mnogo istražena u kontekstu sustava GNSS je metoda radio frekvencijskog otiska. Kod ove metode, provodi se analiza jedinstvenih karakteristika signala za otkrivanje nedosljednosti koja ukazuje na napade lažiranjem. U kontekstu Wi-Fi, Interneta stvari i mobilnih mreža [53], [54], [55], [56], [57], RFF metoda se često koristi. U kontekstu sustava GNSS, postoji nekoliko istraživačkih radova u kojima je primijenjena RFF metoda. Istraživanje različitih pristupa temeljenih na radio frekvencijskom otisku prikazano je u [58]. U radu [106], autori daju pregled najnovijih metoda za detekciju napada lažiranjem te navode metode radio frekvencijskog otiska kao obećavajući neistraženi pristup za detekciju napada lažiranjem.

Pregled metoda radio frekvencijskog otiska za detekciju napada lažiranjem i metoda za detekciju napada lažiranjem temeljena na radio frekvencijskom otisku u pre-korelacijskoj fazi dani su u [59]. Njihov pristup sastoji se od identificiranja relevantnih značajki, primjene izvlačenja značajki, prethodne obrade podataka i upotrebe klasifikatora temeljenih na strojnom i dubokom učenju (SVM, KNN, CNN). Njihovi rezultati pokazuju da kombiniranje različitih značajki u SVM modelu daje najbolje rezultate. Slično, autori u [60] koriste SVM i logističku regresiju za klasifikaciju radiofrekvencijskih otisaka (značajki) kako bi se utvrdilo je li signal autentičan ili lažan, te postižu točnost preko 90%. SVM se u [61] također primjenjuje na tri skupa podataka u pre-korelacijskoj i postkorelacijskoj fazi za klasifikaciju autentičnih i lažnih signala. Rezultati pokazuju da klasifikacija u pre-korelacijskoj domeni daje veću točnost (99.99%) u usporedbi s klasifikacijom u postkorelacijskoj domeni (87.72%). Razlog je dodatna obrada filtara u postkorelacijskoj domeni što kao rezultat daje veću složenost diskriminacije značajki. Detekcija napada lažiranjem u [62] temelji se na konvolucijskom autokoderu. Potvrda predloženog pristupa provedena je na skupu podataka TEXBAT, koji je jedan od najčešće korištenih skupova podataka za detekciju napada lažiranjem u sustavu GNSS [20]. Autori u [21] također koriste skup podataka TEXBAT za provjeru valjanosti metode praćenja kvalitete signala (SQM) koja se temelji na mjerenjima kvalitete korelacijske funkcije u realnim scenarijima napada lažiranjem. Metoda detekcije napada lažiranjem koja se temelji na simulaciji u idealnim uvjetima za RFF identifikaciju prikazana je u [63]. Ova metoda izdvaja RFF značajke iz primljenih signala pomoću dubokog učenja. Evaluiraju se dvije metode klasifikacije temeljene na dubokom učenju: jedna

se fokusira samo na učenje karakteristika fizičkog sloja signala, dok druga izdvaja RFF značajke u vremensko-frekvencijskoj domeni. Oba pristupa pokazuju učinkovitost u otkrivanju napada lažiranjem. U radu [52], autori koriste konvolucijske neuralne mreže za klasifikaciju različitih tipova ometajućih signala te u svom istraživanju isključuju napad lažiranjem. Učinkovitost metode prikazana je u dvije studije: praćenje i klasifikacija putem zemaljske postaje i sa satelita u niskoj orbiti Zemlje LEO (engl. *Low Earth Orbit*). Ulazi u konvolucijsku mrežu uključuju vremensko-frekvencijske vizualne prikaze sirovih uzoraka i izvedenih statističkih metrika. Rezultati pokazuju da predložena metoda postiže visoku točnost od 99.69% u klasificiranju smetnji, čak i uz malu snagu smetnji, te se može implementirati u stvarnom vremenu za praćenje ometača. Nadalje, kratkotrajna Fourierova transformacija STFT (engl. *Short-time Fourier Transform*) i Wigner-Villeova transformacija (WVT) dvije su korištene transformacije za analizu signala u vremensko-frekvencijskoj domeni. U radu se integriraju izdvojene značajke kako bi se poboljšale performanse klasifikacije.

Prema stanju literature i koliko je autoru poznato, primjena diskretne valićne transformacije DWT (engl. *Discrete Wavelet Transform*), konkretno različitih vrsta Daubechiesovih valića db4 i db8, i spektrograma u kombinaciji s modelima strojnog učenja i različitim vrstama klasifikacijskih podataka te različitim satelitskim konstelacijama, GPS i Galileo, za detekciju lažiranja još nisu predložene. Sličan pristup predstavljen je u [52] za detekciju ometanja i klasifikaciju različitih vrsta ometanja na temelju konvolucijskih neuralnih mreža i Wigner-Villeove transformacije. Nadalje, u ovom istraživanju predložen je pristup u kojem se računska složenost smanjuje korištenjem prethodno izdvojenih značajki za klasifikaciju. Svi rezultati za diskretnu valićnu transformaciju u ovoj doktorskoj disertaciji prikazani su za aproksimacijske koeficijente na razini dekompozicije 1. Model strojnog učenja koji se ističe svojim rezultatima kako u literaturi, tako i u ovom doktorskom istraživanju, i čiji rezultati za spektrograme i diskretnu valićnu transformaciju su prikazani u disertaciji je model SVM.

1.2. Hipoteze

Uzevši u obzir pregled dosadašnjeg istraživanja, zaključeno je da metode radio frekvencijskog otiska za detekciju napada lažiranjem u sustavu GNSS nisu istražene i obrađene te da su u svojim začecima. U ovoj doktorskoj disertaciji predložen je integrirani pristup za detekciju i klasifikaciju lažiranih signala koji se temelji na kombinaciji RFF metode i strojnog učenja te su stoga hipoteze doktorske disertacije:

1. Prva hipoteza je da se integracijom i primjenom RFF metoda (npr. spektrogram i DWT) i metoda strojnog učenja u pre-korelacijskoj fazi može poboljšati učinkovitost detekcije napada lažiranjem u sustavu GNSS. Motivacija za primjenu RFF metoda proizlazi iz činjenice da ove metode nisu obrađene i primjenjivane u području detekcije lažiranih signala u sustavu GNSS.

2. Druga hipoteza jest da se primjenom postupka dekompozicije signala i metoda strojnog učenja može razviti računski učinkovitiji algoritam za detekciju napada lažiranjem u GNSS sustavu. Mogući pristup je klasifikacija tipa signala na temelju unaprijed izdvojenih značajki generiranih iz slika i kofaznih i kvadraturnih I/Q (engl. *in-phase and quadrature*) komponenti signala.

1.3. Očekivani znanstveni doprinosi

Očekivani znanstveni doprinosi ove doktorske disertacije su sljedeći:

1. Novi pristup za detekciju lažiranih signala u sustavu GNSS temeljen na integraciji RFF spektrograma i diskretne valićne transformacije (DWT) sa strojnim učenjem u prekorelacijskoj fazi obrade signala. Istraživanje će se provesti za dvije vrste satelitskih konstelacija GPS i Galileo, a za evaluaciju predložene metode primijenit će se javno dostupni skup podataka OAKBAT.

2. Novi računski učinkovitiji algoritam za detekciju napada lažiranjem signala u sustavu GNSS primjenom postupka dekompozicije signala i strojnog učenja, pri čemu se klasifikacija tipa signala provodi na temelju unaprijed izdvojenih značajki generiranih iz slika i kofaznih i kvadraturnih I/Q komponenti signala.

1.4. Sadržaj disertacije

Ova doktorska disertacija podijeljena je na sedam poglavlja. U prvom poglavlju sažeto je prikazan uvod u temu istraživanja uključujući motivaciju i osnovne ciljeve istraživanja. Prikazana su najnovija znanstvena dostignuća u području istraživanja. Zaključno, definirane su hipoteze rada, očekivani znanstveni doprinosi te sažeti pregled sadržaja teksta doktorskog rada.

Drugo poglavlje opisuje osnovne značajke, način rada te strukturu signala sustava GNSS. Prikazana je i arhitektura sustava GNSS koja se sastoji od tri segmenta: svemirski, kontrolni i korisnički segment. Također, prikazani su i detalji za četiri sustava GNSS: GPS, Galileo, GLONASS i BeiDou. Opisan je i napad lažiranjem u prijamnicima sustava GNSS. Navedene su i osnovne vrste napada lažiranjem: pojednostavljeni napad, napad lažiranjem srednje razine složenosti te sofisticirani napad lažiranjem.

Pregled metoda strojnog učenja koje su korištene kroz različite eksperimente za detekciju lažiranih signala i klasifikaciju tipa signala u ovom istraživanju prikazan je u trećem poglavlju. To su sljedeće metode: metoda potpornih vektora SVM, K-najbližih susjeda KNN, stablo odlučivanja DT (engl. *Decision Tree*) i slučajne šume RF (engl. *Random Forest*). Kroz ovo poglavlje, prikazana je i primjena metoda strojnog učenja za detekciju napada lažiranjem u sustavu GNSS.

U prvom dijelu četvrtog poglavlja prikazani su eksperimenti izvođenja pojednostavljenog napada lažiranjem u različitim uvjetima: unutarnji i vanjski. Napad je u unutarnjim uvjetima izveden na hodniku Fakulteta elektrotehnike, strojarstva i brodogradnje (FESB) u Splitu dok je napad u vanjskim uvjetima izveden ispred zgrade istog Fakulteta. Prikazana je i korištena oprema potrebna za izvođenje napada lažiranjem koja uključuje softverski definirani radio, laptop, štapnu odašiljačku antenu, eksterni oscilator te pametne telefone (Android

i iPhone). Eksperimenti su izvedeni korištenjem simulatora otvorenog pristupa GPS-SDR-SIM za signale sustava GPS na frekvenciji L1 1575.42 MHz. Napadač u ovom eksperimentu je softverski definirani radio HackRF One kojim su odašiljani lažni signali sustava GNSS koji imaju veću snagu u odnosu na autentične signale. Eksperimentom u unutarnjim uvjetima analiziran je i domet napadača i vrijeme izračuna položaja pojedinog pametnog telefona ovisno o udaljenosti između napadača i prijamnika. Nadalje, prikazani su i rezultati detekcije napada lažiranjem i klasifikacije tipa signala korištenjem strojnog učenja na skupu podataka SatGrid u postkorelacijskoj fazi. Kroz ovo poglavlje, dan je i pregled različitih metoda za detekciju napada lažiranjem. Neke od prikazanih metoda i rezultata u literaturi su: klasične metode temeljene na promatranju omjera signal-šum, metode podatkovnih bitova koje uključuju praćenje i analizu poruka Nacionalne udruge za pomorsku elektroniku, smjera i vremena dolaska signala te metode obrade signala u koje spadaju metode temeljene na snazi, metode temeljene na anteni i metode temeljene na distorziji korelacije i praćenju vrha signala korelacije.

Predloženi integrirani pristup za detekciju lažiranih signala i klasifikaciju tipa signala korištenjem kombinacije metoda radio frekvencijskog otiska i strojnog učenja na skupu podataka OAKBAT za dvije satelitske konstelacije GPS i Galileo, prikazan je u petom poglavlju. Metode radio frekvencijskog otiska koje su korištene u ovom istraživanju su diskretna valična transformacija i spektrogram. Pristup temeljen na diskretnoj valičnoj transformaciji uključuje analizu i usporedbu dvije vrste valića Daubechies db4 i db8. Nadalje, sva analiza napravljena je za aproksimacijske koeficijente i jednu razinu dekompozicije. Za spektrograme, analiza uključuje usporedbu spektrograma koji imaju različite veličine prozora. U ovom istraživanju, po prvi put su metode radio frekvencijskog otiska primijenjene na skupove podataka OAKBAT sustava GPS i Galileo u statičkim uvjetima. Kroz peto poglavlje prikazana je i analiza računske složenosti za predložene metode i sve korištene ulazne klasifikacijske podatke. Dan je i prijedlog kako uz korištenje unaprijed definiranih odnosno izdvojenih značajki smanjiti računsku složenost.

U šestom poglavlju prikazani su model za evaluaciju predloženih metoda detekcije lažiranih signala u sustavu GNSS te vrste korištenih ulaznih klasifikacijskih podataka: slike, značajke slike i statističke i spektralne značajke. Svi modeli su evaluirani korištenjem standardnih parametara performansi modela (konfuzijska matrica, točnost, preciznost, odziv te mjera *F1*), koji su također detaljno prikazani kroz ovo poglavlje. Nadalje, prikazana je i primjena skupa podataka OAKBAT za evaluaciju metoda detekcije lažiranih signala sustava GNSS. Konačno, prikazani su rezultati evaluacije metoda detekcije lažiranih signala i klasifikacije tipa signala korištenjem standardnih parametara za evaluaciju i performanse modela. Prikazani su rezultati evaluacije metode temeljene na spektrogramu i metode temeljene na diskretnoj valičnoj transformaciji. Rezultati su prikazani za sve vrste korištenih značajki.

Dodatno, u šestom poglavlju prikazani su rezultati evaluacije metoda detekcije lažiranih signala i klasifikacije tipa signala korištenjem standardnih parametara za evaluaciju i perfor-

manse modela koji su navedeni u petom poglavlju. Prikazani su rezultati evaluacije metode temeljene na spektrogramu i metode temeljene na diskretnoj valićnoj transformaciji. Rezultati su prikazani za sve tri vrste korištenih značajki: slike, značajke slike i statističke i spektralne značajke signala. Kroz šesto poglavlje prikazana je i analiza računske složenosti za predložene metode i sve korištene ulazne klasifikacijske podatke. Dan je i prijedlog kako uz korištenje unaprijed definiranih odnosno izvučenih značajki smanjiti računsku složenost.

Konačno, u zadnjem sedmom poglavlju izvedeni su zaključci, potvrđen ostvareni znanstveni doprinos i hipoteza provedenog istraživanja.

2. Napad lažiranjem u globalnom navigacijskom satelitskom sustavu GNSS

Sustav GNSS je nevidljivi dio tehnologije na koji se ljudi svakodnevno oslanjaju npr. korištenje mobilnih navigacijskih aplikacija. Svrha navigacijskog satelitskog sustava je pružanje usluga pozicioniranja i navigacije u realnom vremenu bilo kada i bilo gdje. Osnovne značajke sustava GNSS su opisane u ovom poglavlju.

2.1. Osnovne značajke sustava GNSS

Pod pojmom GNSS sustav podrazumijeva se bilo koja konstelacija satelita koja pruža usluge pozicioniranja, navigacije i mjerenja vremena. GNSS se temelji na konstelaciji satelita koji odašilju signale iz svemira prema zemaljskoj površini. Signali prenose podatke o položaju i vremenu na GNSS prijamnik te prijamnik koristi te podatke za određivanje položaja odnosno pozicioniranje.

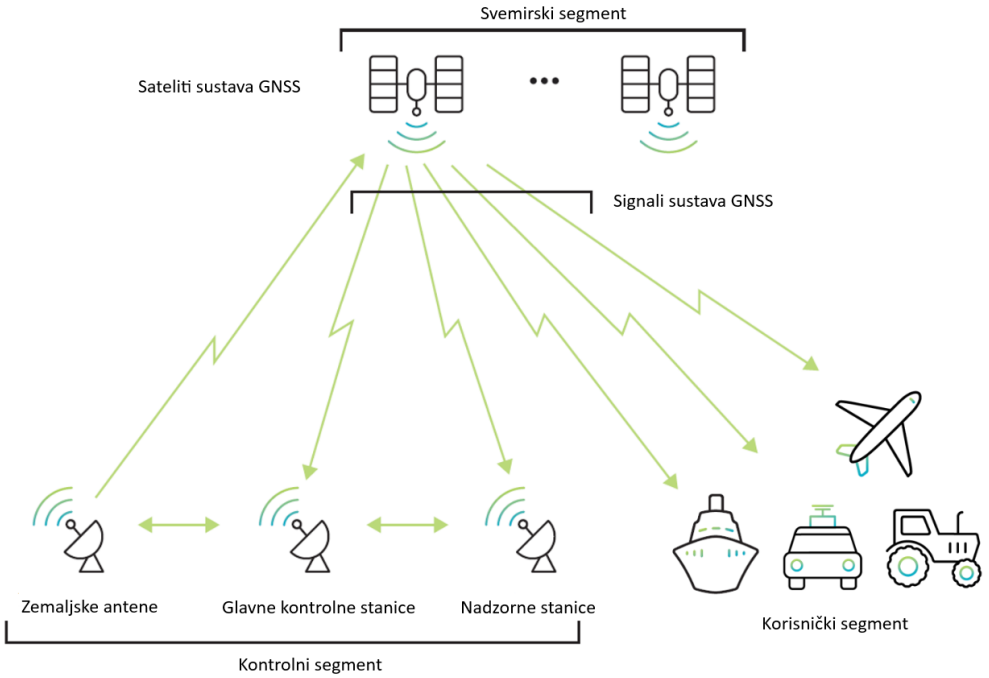
2.1.1. Satelitski sustav GPS

Ovaj radionavigacijski sustav je u najširoj civilnoj upotrebi danas. Poznat je i kao NAVSTAR (engl. *Navigation System with Time and Ranging*) te je prvotno razvijen u vojne svrhe od strane Ministarstva obrane SAD-a. Američki kongres je dozvolio i civilnu upotrebu. Prvi satelit lansiran je 1978. godine, a puna konstelacija je ostvarena 1995. godine. GPS sateliti konstantno odašilju dva signala nosioca u L pojasu (L1 i L2). Signali nosioci vrlo su važni jer na Zemlju donose informacije sa satelita koje prijamniku omogućuju da utvrdi točnu lokaciju [2].

Osnovni segmenti satelitskog navigacijskog sustava GPS su sljedeći (slika 2.1):

1. svemirski,
2. kontrolni,
3. korisnički.

Svemirski segment GPS sustava sastoji se od 31 satelita, ravnomjerno raspoređena u 6 orbitalnih ravnina, koji svakih 12 sati obiđu Zemlju na udaljenosti od približno 20 200



Slika 2.1: Arhitektura sustava GNSS [1].

kilometara. Osnovna zadaća ovog segmenta je odašiljanje radio signala pomoću kojih se mjere udaljenost te pružanje točnih informacija o položaju i vremenu korisnicima bilo gdje u svijetu. Orbitalne ravnine ne rotiraju u odnosu na udaljene zvijezde i centrirane su na Zemlji. Orbite su raspoređene tako da je najmanje šest satelita uvijek vidljivo sa svih strana Zemljine površine. Od 2019. godine 31 satelit se nalazi u GPS konstelaciji te je devet satelita vidljivo u bilo kojem trenutku s bilo kojeg mjesta na Zemlji. Dodatni sateliti poboljšavaju preciznost mjerenja.

Kontrolni segment odnosi se na zemaljske postaje smještene u cijelom svijetu u blizini ekvatora. Koriste se za praćenje, kontrolu i slanje informacija svakom GPS satelitu. Glavni zadatak kontrolnog ili zemaljskog segmenta je praćenje satelita u svrhu određivanja orbita i vremena, sinkronizacija vremena satelita te odašiljanje poruka satelitima. Kontrolni ili zemaljski segment sastoji se od: glavne kontrolne stanice, alternativne glavne kontrolne stanice, četiri dodijeljene zemaljske antene i šest dodijeljenih nadzornih stanica. Glavna kontrolna stanica nalazi se u bazi zračnih snaga u Colorado Springsu u SAD-u i odgovorna je za cjelokupno upravljanje lokacijama daljinskog nadzora i prijenosa. Osim toga, zadaće su joj i praćenje GPS satelita, nadziranje njihovih prijenosa, prikupljanje podataka nadzornih stanica, sinkronizacija vremena i prosljeđivanje podataka zemaljskim stanicama. Šest nadzornih stanica provjerava točnu visinu, položaj, brzinu i ukupno stanje satelita u orbiti. Kontrolni segment koristi mjerenja prikupljena od strane nadzornih stanica za predviđanje ponašanja orbite i sata svakog satelita. Podaci o predviđanjima prenose se korisnicima satelitima za prijenos. Kontrolni segment osigurava da orbite i satovi GPS satelita ostanu unutar prihvatljivih granica. Stanica može pratiti do 11 satelita istovremeno. Ova provjera obavlja

se dva puta dnevno za svaku stanicu, nakon što sateliti završe svoje putovanje oko Zemlje. U slučaju da se zapaze nekakvi problemi, proslijedi ih se glavnoj kontrolnoj stanici. Četiri zemaljske antene nadziru i prate satelite od horizonta do horizonta. Osim toga, satelitima prenose informacije o korekcijama.

Korisnički segment uključuje bilo koga tko koristi/ima GPS prijamnik. Ovaj se segment sastoji od stotina tisuća američkih i savezničkih vojnih korisnika usluge preciznog pozicioniranja i desetaka milijuna civilnih, komercijalnih i znanstvenih korisnika usluge standardnog pozicioniranja [2].

2.1.2. Satelitski sustav Galileo

Galileo je europski navigacijski satelitski sustav nastao kao zajednička inicijativa Europske svemirske agencije ESA (engl. *European Space Agency*) i Europske komisije, koji pruža vrlo točan, zajamčeni servis globalnog pozicioniranja u stvarnom vremenu s preciznošću od metra pod civilnom kontrolom. Prvi Galileo testni satelit GIOVE-A, lansiran je 2005. godine, a prvi satelit koji je kasnije postao dio operativnog sustava lansiran je 2011. Do srpnja 2018. godine, 26 od planiranih 30 satelita, uključujući i rezervne, bili su u orbiti. Potpuno raspoloživi Galileo sustav sastoji se od 24 operabilna satelita plus 6 rezervnih u orbiti, smještenih u 3 krušne orbite na 23 222 km visine iznad Zemlje.

Sustav Galileo sastoji se od svemirskog segmenta (sateliti u svemiru), zemaljskog segmenta na nekoliko lokacija te korisničkog segmenta (slika 2.1). Svemirski segment sustava Galileo definiran je kao 24/3/1 Walker konstelacija. To predstavlja 24 satelita nominalne srednje Zemljine orbite MEO (engl. *Medium Earth Orbit*) raspoređena u 3 orbitalne ravnine. Konstelaciju je moguće nadopuniti pomoćnim Galileo satelitima koji zauzimaju orbitalne utore koji nisu dio osnovne konstelacije. Zemaljski Galileo segment sastoji se od dva Galileo kontrolna centra smještena u Oberpfaffenhofeu u Njemačkoj i u Fucinu u Italiji. Svaki Galileo kontrolni centar upravlja kontrolnim funkcijama koje podržava Segment zemaljske kontrole i funkcijama misije koje podržava Segment zemaljske misije. Segment zemaljske kontrole nadzire i kontrolira satelite i bazira se na Galileo kontrolnom centru u Oberpfaffenhofenu, a povezan je s telemetrijskim, pratećim i telekomunikacijskim postajama u Kiruni (Švedska) i Kourou (Francuska Gvajana). Segment zemaljske misije nalazi se u drugom Galileo kontrolnom centru (Fucino) i osigurava najsuvremenije navigacijske performanse Galilea. Galileo korisnički segment sastoji se od svih kompatibilnih prijamnika i uređaja koji prikupljaju Galileo signale i izračunavaju svoju lokaciju. Postoje različite korisničke zajednice ovisno o primjeni te pokrivaju širok raspon, od prijevoza do aplikacija za mjerenje vremena. Galileo sateliti odašilju signale na nekoliko osnovnih frekvencija koje se nalaze u L pojasu (1.1 do 1.6 GHz): E1 (1575.42 MHz), E5 (1191.795 MHz) koji se sastoji od E5a (1176.45 MHz) and E5b (1207.14 MHz), te E6 (1278.75 MHz) [3]. Navedene frekvencije koriste se za civilne svrhe s tim da su E1, E5, E5a namijenjene pružanju bes-

platne usluge svim korisnicima bez potrebe za autorizacijom dok je E6 namijenjen pružanju komercijalne usluge, koja je dostupna isključivo ovlaštenim korisnicima.

2.1.3. Satelitski sustav GLONASS

GLONASS je ruski satelitski navigacijski sustav. Prvi GLONASS satelit odaslan je 1982. godine, a sustav je 1993. godine proglašen potpuno operativnim. Postojalo je razdoblje u kojem su performanse GLONASS-a opale te se Rusija obavezala dovesti sustav na potreban minimum od 18 aktivnih satelita. Trenutno GLONASS ima punu raspodjelu od 25 satelita u konstelaciji. Dizajn GLONASS sustava sličan je dizajnu GPS sustava i sastoji se od tri dijela: kontrolni segment, svemirski segment, korisnički segment, koja su definirana na vrlo sličan način kao i segmenti sustava GPS. Konstelacija GLONASS-a, ovisno o lokaciji, omogućava vidljivost različitog broja satelita. Potrebna su barem četiri satelita u vidokrugu kako bi GLONASS prijamnik mogao izračunati svoju poziciju u tri dimenzije te kako bi se sat prijamnika sinkronizirao sa satom sustava. Geometrija GLONASS konstelacije ponavlja se otprilike jednom svakih osam dana. Satelitski signal GLONASS identificira satelit i uključuje: podatke o pozicioniranju, brzini i ubrzanju za izračunavanje satelitskih lokacija, informacije o "zdravstvenom" stanju satelita te odmaku GLONASS vremena od UTC vremena. GLONASS sateliti odašilju signale na nekoliko frekvencija L1 (1602 MHz), L2 (1246 MHz), L3 (1201 MHz) i L5 (1176.45 MHz) od kojih se frekvencije L1, L3 i L5 koriste u civilne svrhe dok se L5 koristi u vojne svrhe [4].

2.1.4. Satelitski sustav BeiDou

Satelitski sustav BeiDou kineski je navigacijski satelitski sustav. Prvi BeiDou sustav, poznat i kao BeiDou-1, sastojao se od tri satelita koji su nudili ograničenu pokrivenost i navigacijske usluge, uglavnom za korisnike u Kini i susjednim regijama. Druga generacija sustava, BeiDou-2, imala je djelomičnu konstelaciju od 10 satelita u orbiti. Treća generacija BeiDou navigacijskog satelitskog sustava osigurava globalnu pokrivenost za mjerenje vremena i navigaciju te može poslužiti kao alternativa američkom GPS-u, ruskom GLONASS-u i europskom Galileu. BeiDou-1 bio je eksperimentalni regionalni navigacijski sustav koji se sastojao od tri radna i jednog rezervnog satelita. Sateliti su bili bazirani na kineskom geostacionarnom komunikacijskom satelitu DFH-3 i svaki je imao težinu lansiranja od 1 tone. Za razliku od prethodno opisanih satelitskih sustava, BeiDou-1 je koristio satelite u geostacionarnoj orbiti što znači da sustav ne zahtijeva veliku konstelaciju satelita, no ipak ograničava pokrivenost područja na Zemlji odakle su sateliti vidljivi. BeiDou-2 u potpunosti zamjenjuje sustav BeiDou-1 te isti nije njegovo proširenje. Prema posljednjim podacima iz 2023. aktivno je u orbitama 7-11 BeiDou-2 satelita. Sateliti su odašiljali signale na tri frekvencije B1, B2 i B3 [5]. Treća faza razvoja sustava BeiDou-3 uključuje ukupno 30 operativnih satelita. Time su uvedene nove frekvencije civilnih signala B1C/B1I/B1A (1575.42 MHz), otvore-

nih signala B2a/B2b (1191.795 MHz), signala B3I/B3Q/B3A (1268.52 MHz) i Bs signala (2492.028 MHz) za eksperimentalno emitiranje u S pojasu (2-4 GHz). Novi civilni signal B1C odašilje se na novoj frekvenciji 1575.420 MHz kao i signali na L1 u GPS-u. Otvoreni signali B2a (1176.450 MHz) i B2b (1207.140 MHz) odašilju se na frekvencijama koje su identične frekvencijama Galileo signala E5a i E5b, te postoji mogućnost njihove zajedničke obrade (1191.795 MHz). Novi BDS-3 signali omogućuju bolju kompatibilnost i interoperabilnost s drugim GNSS-ima. Sustav omogućuje dvije vrste usluga: otvoreni i autorizirani servis. Otvoreni servis za civilnu upotrebu je besplatan, a osigurava točnost apsolutnog pozicioniranja. Autorizirani servis omogućuje pouzdanije određivanje pozicije, brzine i vremena, te komunikacijski servis i viši stupanj integriteta. Besplatna civilna usluga ima točnost praćenja lokacije od 10 m, dok ograničena vojna usluga ima točnost od 10 cm.

2.2. Način rada sustava GNSS

Pod pojmom GNSS sustav podrazumijeva se bilo koja konstelacija satelita koja pruža usluge pozicioniranja, navigacije i mjerenja vremena. GNSS se temelji na konstelaciji satelita koji odašilju signale iz svemira prema zemaljskoj površini. Signali prenose podatke o položaju i vremenu na GNSS prijamnik te prijamnik koristi te podatke za određivanje položaja odnosno pozicioniranje. GNSS prijamnik se sastoji od antene i jedinice za obradu (prijamnika) [1]. Satelitski signali se prikupljaju pomoću antene, a jedinica za obradu pretvara prikupljene informacije u oblik razumljiv korisniku tj. zemljopisne koordinate. Sami položaj antene određuje stvarna mjerenja, primjerice ako se antena nalazi na nekom teško dostupnom položaju kao što je urbani kanjon, u samim mjerenjima će postojati mnogo reflektiranih signala nastalih višestaznim prostiranjem signala. Postoje različiti GNSS prijamnici koji ne mogu primati sve GNSS signale. Primjerice, GPS prijamnik može primati samo GPS signale dok GLONASS prijamnik može primati samo signale s GLONASS satelita. Također, postoje i složeniji prijamnici koji mogu primati signale s više satelita tzv. multi-konstelacijski GNSS prijamnici [70].

Za određivanje položaja moraju biti poznati sljedeći elementi:

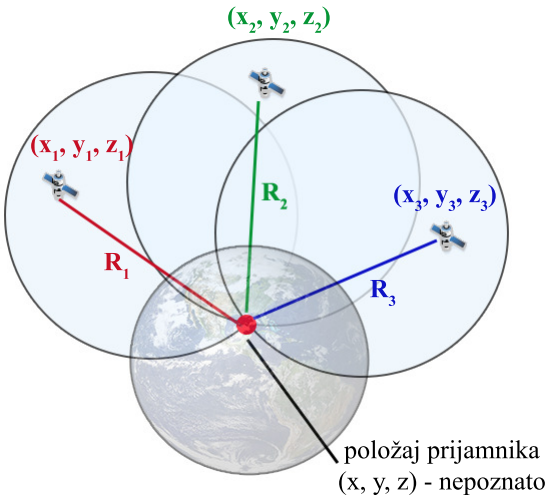
- položaj satelita,
- vrijeme odašiljanja signala,
- vrijeme prijama signala,
- brzina prostiranja signala.

Određivanje položaja temelji se na mjerenju vremena propagacije (širenja) satelitskog radijskog signala od satelitske odašiljačke antene do antene korisničkog prijamnika. GNSS pozicioniranje na visokoj razini jednostavno se temelji na konceptu trilateracije. Kako bismo

odredili nepoznati položaj (x,y,z) prijamnika kao što je prikazano na slici 2.2, pretpostavimo da su položaji triju GNSS satelita unaprijed poznati (sateliti šalju prijamniku informacije o položaju preko navigacijske poruke). Kada prijamnik dobije i prati dolazne GNSS signale od tri satelita, može odrediti vrijeme propagacije signala Δt (vrijeme odašiljanja minus vrijeme prijema). Pošto su GNSS signali elektromagnetski valovi koji se šire brzinom svjetlosti $c \approx 3 \times 10^8$ m/s, udaljenosti od prijamnika do tri satelita (R_1, R_2, R_3) se dobiju množenjem c s Δt i te udaljenosti se nazivaju pseudoudaljenosti. Pseudoudaljenost predstavlja pravu udaljenost na koju je dodana mala (pozitivna ili negativna) korekcija udaljenosti uzrokovana pogreškom sata prijamnika. Mjerenje pseudoudaljenosti zahtijeva precizno poznavanje vremena odašiljanja signala sa satelita i vremena prijama signala na prijamniku [14]. Konačno, skup trilateracijskih jednadžbi se može postaviti kao:

$$c(\Delta t^m) = \sqrt{(x-x_m)^2+(y-y_m)^2+(z-z_m)^2}, \quad m = 1,2,3, \tag{2.1}$$

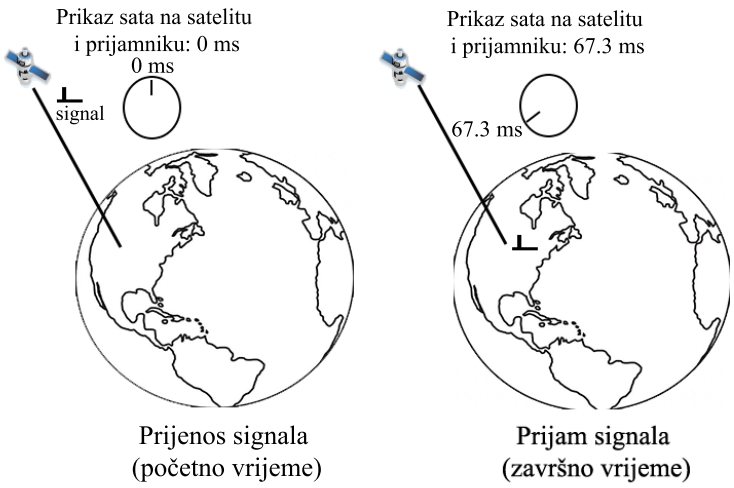
gdje su x_m, y_m i z_m poznate koordinate triju satelita [10]. Nepoznate koordinate prijamnika se određuju rješavanjem triju jednadžbi s tri nepoznanice. Iako su dovoljna samo tri satelita, točnost i preciznost povećat će se s većim brojem satelita, pa se za izračun položaja najčešće koriste četiri satelita.



Slika 2.2: GNSS trilateracija [10].

Preduvjet za određivanje položaja i vremena je sinkronizacija odnosno vremensko usklađivanje svih elemenata (satelit i prijamnik) sustava na zajedničko vrijeme GNSS sustava. Sinkronizacija elemenata sustava omogućuje mjerenje vremena propagacije satelitskog signala na način da satelit označava trenutak odašiljanja signala, a prijamnik trenutak prijama signala. Na slici 2.3 prikazana je sinkronizacija satova satelita i prijamnika za vrijeme propagacije signala [6].

Za početno vrijeme propagacije signala prikaz sata na satelitu i prijamniku je 0 ms. Svaki satelit prenosi svoj točni položaj i točno vrijeme do Zemlje s određenom frekvencijom ovisno



Slika 2.3: Sinkronizacija satova satelita i prijamnika [6].

o frekvencijskom pojasu i vrsti satelitskog sustava. Ovi signali putuju brzinom svjetlosti i prema tome treba približno 67.3 ms da dosegnu Zemljinu površinu neposredno ispod satelita što je prikazano kao završno vrijeme propagacije signala.

Satelitski navigacijski sustav koristi visoko postavljene satelite na način da se iz bilo koje točke na tlu može povući crta do četiri satelita. Svaki satelit ima do četiri atomska sata (najtočniji sat koji ima najveću grešku od 1 sekunde u 30 milijuna godina). Za još veću preciznost, atomski satovi rade korekciju ili sinkronizaciju iz kontrolne točke na Zemlji. Bez atomskog sata ne bi bio izvediv ni GPS, navigacija bi bila otežana, svemirski letovi se ne bi mogli tako precizno planirati, itd. Atomski sat na bazi cezija je sat koji koristi elektromagnetsko zračenje, koje nastaje kod prijelaza između dviju hiperfinih razina osnovnog stanja atoma cezija -133 na temperaturi od 0 K. I atomski i običan mehanički sat za mjerenje vremena koriste titranje ili osciliranje, ali kod atomskog sata je ono određeno masom jezgre atoma i silom gravitacije, te elektrostatičkom oprugom između pozitivnog naboja jezgre i elektronskog oblaka [7].

Na točnost položaja i vremena utječu dva faktora [66]:

1. Korisnička pogreška udaljenosti URE (engl. *User Range Error*) je razlika između navigacijskih podataka satelita (položaj i sat) i istinitih vrijednosti, projiciranih na vidokrug korisnika. URE je funkcija kvalitete emitiranog signala i podataka.

2. Geometrijsko smanjenje preciznosti GDOP (engl. *Geometric Dilution of Precision*) je mjera kvalitete geometrije (distribucija satelita na nebu) koju definiraju sateliti i prijamnik odnosno opisuje jakost trenutne satelitske konfiguracije ili geometrije na točnost podataka prikupljenih prijamnikom. GDOP je učinak geometrije satelita na pogrešku položaja i grubo se definira kao omjer pogreške položaja i pogreške dometa.

GDOP ovisi samo o položaju satelita (broj vidljivih satelita i koliko su visoko na nebu - geometrija). Kada su vidljivi sateliti blizu jedan drugome na nebu, geometrija je slaba, a GDOP vrijednost visoka. To potencijalno smanjuje kvalitetu pozicioniranja za nekoliko metara. S druge strane, kada su sateliti međusobno udaljeni, geometrija je jaka, a GDOP vrijednost niska što je prikazano na slici 2.4. Što je veći broj satelita, to je bolja vrijednost GDOP-a i obrnuto [14].

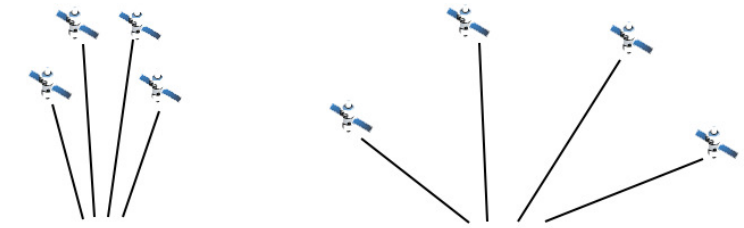
GDOP se može izraziti kao niz zasebnih komponenti [23]:

(a) Horizontalno smanjenje preciznosti HDOP (engl. *Horizontal Dilution of Precision*) je mjera točnosti u 2D položaju (zemljopisna širina i dužina). HDOP vrijednosti su tipično između 1 i 2.

(b) Položajno smanjenje preciznosti PDOP (engl. *Position Dilution of Precision*) označava mjeru preciznosti položaja (HDOP + VDOP). Sateliti raširenи nebom obično će imati dobru (nižu PDOP vrijednost) geometriju. Sateliti skupljeni čvrsto na određenom dijelu neba obično će imati lošu (veću PDOP vrijednost) geometriju. PDOP vrijednosti koje se smatraju dobrima za pozicioniranje su male, poput 3. Vrijednosti veće od 7 se smatraju lošima.

(c) Vertikalno smanjenje preciznosti VDOP (engl. *Vertical Dilution of Precision*) je mjera točnosti u 1-D položaju (visina).

(d) Vremensko smanjenje preciznosti TDOP (engl. *Time Dilution of Precision*) je mjera preciznosti vremena. Visoki TDOP uzrokuje pogreške sata prijamnika što rezultira do povećanja pogreški položaja.

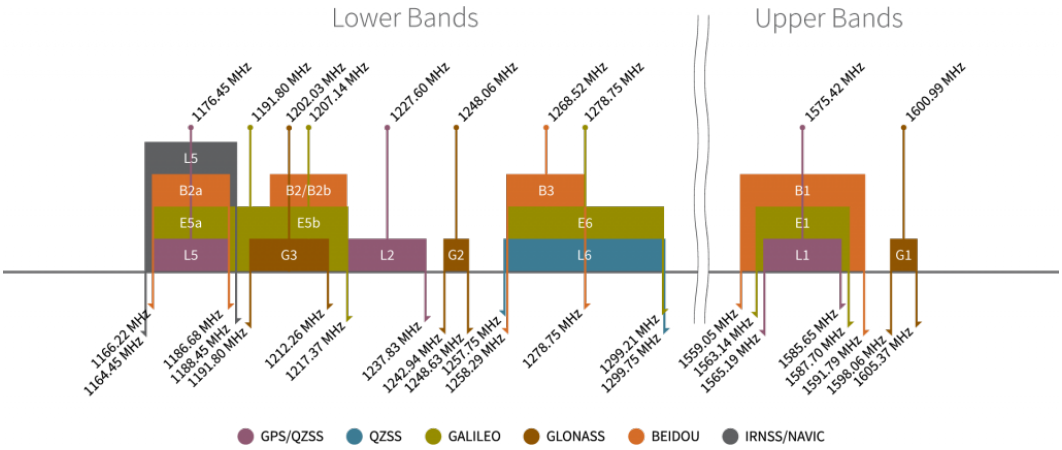


Slika 2.4: Ilustracija dobre i loše geometrije satelita.

2.3. Struktura signala sustava GNSS

GNSS sustavi rade isključivo u L pojasu (1.1 – 1.6 GHz). Ovi frekvencijski pojasevi koriste se za satelitske sustave iz razloga što omogućuju lak prolazak kroz atmosferu, manje gubitke s povećanjem udaljenosti te pouzdan prijenos signala i što valovi L pojasa prodiru kroz oblake, maglu, kišu, oluje i vegetaciju te GNSS jedinice mogu primati točne podatke

u svim vremenskim uvjetima, danju ili noću. Prikaz frekvencijskih pojaseva koji se nalaze unutar L pojasa u sustavu GNSS su prikazani na slici 2.5. Osobitost svih signala sustava GNSS je modulacija harmonijskog radio signala (signal nosioc) s karakterističnim pseudoslučajnim nizom PRN (engl. *Pseudorandom Noise Code*). PRN kod je binarni niz brojeva 0 i 1. Ovaj kod se neprekidno ponavlja u intervalima od nekoliko milisekundi do sekunde i olakšava mjerenje vremena propagacije signala. Svaki prijamnik po PRN nizu razlikuje svaki pojedinačni satelit koji emitira na istoj frekvenciji [14].



Slika 2.5: Prikaz frekvencijskih pojaseva unutar L pojasa u sustavu GNSS [9].

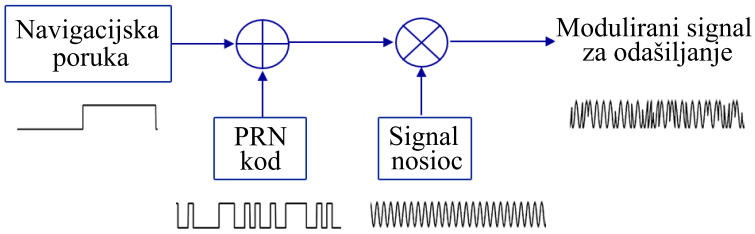
GNSS sateliti kontinuirano odašilju signale na dvije ili više frekvencija u L pojasu s tim da postoji testna frekvencija Bs (2492.02 MHz) kod BeiDou satelita koja je u eksperimentalnoj fazi. Ovi signali sadrže PRN kodove i navigacijske poruke pomoću kojih se računa vrijeme propagacije od satelita do prijamnika i koordinate satelita u bilo kojoj epohi.

Osnovne komponente signala sustava GNSS su [8]:

1. signal nosioc - radio frekvencijski sinusoidalni signal na određenoj frekvenciji,
2. pseudoslučajni kod PRN je binarni niz nula i jedinica koji je jedinstven za svaki satelit. PRN kodovi su matematički modelirani (imaju nisku međusobnu korelaciju) na način da omogućavaju svim satelitima emitiranje na istoj frekvenciji bez da ometaju jedan drugoga. Svaki satelit u GNSS konstelaciji ima jedinstveni PRN kod koji emitira kao dio navigacijske poruke te tako omogućava prijamniku da točno identificira satelit od kojeg prima signal. Vrijeme propagacije signala od satelita do prijamnika mjeri se pomoću PRN koda na način da prijamnik stvara lokalnu kopiju PRN koda, uspoređuje primljeni PRN kod s vlastitom kopijom te pomiče (korelira) lokalni kod sve dok ne postigne podudarnost s primljenim signalom. Taj pomak (kašnjenje) predstavlja vrijeme propagacije signala od satelita do prijamnika na temelju kojeg se računa udaljenost korištenjem brzine svjetlosti. Nadalje, PRN kod omogućuje sinkronizaciju odnosno točno usklađivanje vremena između satelita i prijamnika te da svi sateliti koriste istu frekvenciju, ali različite kodove (metoda višestrukog pristupa s kodnom podjelom CDMA (engl. *Code Division Multiple Access*)).

3. navigacijska poruka je binarno kodirana poruka koja pruža informacije o satelitskim efemeridama (položaj i brzina satelita), parametrima za usuglašavanje satova, almanahu (raspored satelitskih orbitalnih parametara kako bi prijamnik dobio informaciju o vidljivosti satelita u određenom trenutku), zdravstvenom statusu satelita (aktivni satelit jer prijamnik ne prati satelite koji nisu aktivni) i drugim komplementarnim informacijama. Navigacijske poruke se odašilju brzinom od najmanje 50 bit/s s trajanjem od 20 ms. Bitne satelitske efemeride i parametri sata ponavljaju se svakih 30 s.

Kao primjer, glavne komponente signala GPS L1 C/A su prikazane na slici 2.6.



Slika 2.6: Struktura GNSS signala.

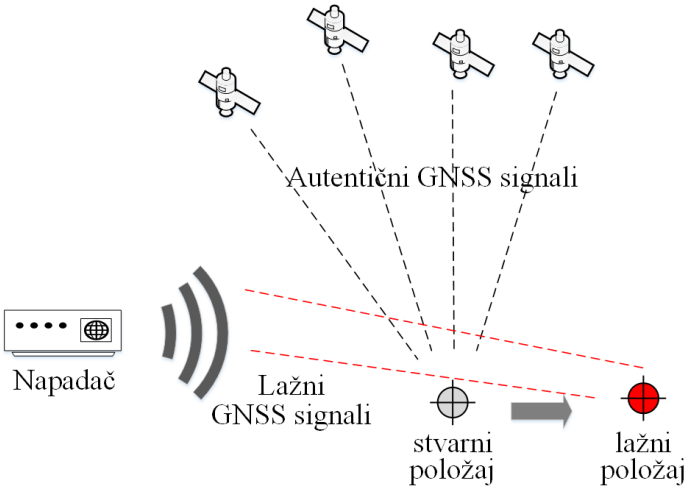
Svaki GPS satelit odašilje dva različita koda: civilni kod C/A (engl. *coarse acquisition*) i enkriptirani kod P(Y) (engl. *precision/secure*) koji je rezerviran za vojne i ovlaštene civilne korisnike. Svaki C/A kod je jedinstvena sekvenca od 1023 bita i ponavlja se svaku milisekundu. C/A kod se prenosi na jednoj frekvenciji L1 dok se P kod prenosi na dvije frekvencije (L1 i L2).

Za zbrajanje navigacijske poruke i PRN koda koristi se operacija *xor* tj. zbrajanje po modulu 2. Ako su oba bita 0 ili 1, rezultat je 0. Ako su bitovi različiti (jedan bit 0, drugi 1), rezultat je 1. Binarni signal modulira signal nosioc korištenjem digitalne modulacije s binarnim faznim pomakom ili BPSK (engl. *Binary Phase Shift Keying*) u kojoj se podaci prenose mijenjanjem ili moduliranjem dviju različitih faza signala nosioca. Ovim postupkom nastaje modulirai signal koji se odašilje. Bit 0 ostavlja signal nosioc nepromijenjen dok se za bit 1 množi signal nosioc s -1 (ekvivalent za fazni pomak sinusnog signala za 180°). Kada kod prelazi s 0 na 1 ili obrnuto, faza signala nosioca se mijenja za 180°.

2.4. Napad lažiranjem u sustavu GNSS

Mobilni telefoni su vrlo osjetljivi na napade lažiranjem lokacije. Ovakvi napadi se često koriste u vojnim svrhama npr. za preusmjeravanje aviona ili dronova na lažnu lokaciju. Napad lažiranjem predstavlja veliki sigurnosni problem i zato je potrebno razviti nove algoritme i metode za sprječavanje ovih napada te poboljšati postojeće metode.

Pod pojmom napad lažiranjem GNSS signala podrazumijeva se namjerno odašiljanje lažnih GNSS signala s namjerom da prijamnik lažne signale pogrešno protumači kao autentične te u svrhu lažiranja lokacije prijamnika. Osnovne zadaće GNSS prijamnika su primiti i razdvojiti signale sa satelita, izračunati pseudoudaljenosti za svaki satelit na temelju vremena prijama signala, demodulirati navigacijsku poruku kako bi se dobili efemeris (engl. *ephemeris*) podaci te procijeniti položaj, brzinu i vrijeme prijamnika tzv. PVT (engl. *Position, Velocity, Time*) rješenje [14].



Slika 2.7: Napad lažiranjem [14].

Slika 2.7 prikazuje jednostavan napad lažiranjem. Dakle, napadač (engl. *spoofer*) odašilje lažne signale, koji su veoma slični autentičnim GNSS signalima. Lažni signali imaju veću snagu u odnosu na autentične kako bi se prijamnik zavarao i uzeo te signale. Nakon primanja lažnih signala, prijamnik pokazuje lažnu lokaciju na kojoj se zapravo ne nalazi. U načelu, lažni signal mora imati određene značajke podataka koje odgovaraju onima stvarnog satelitskog signala.

Općenito, primljeni signali sustava GNSS se mogu matematički opisati kao kombinacija nekoliko signala [13]:

$$y(t) = Re\left\{\sum_{i=1}^{N} A_i D_i[t-\tau_i(t)]C_i[t-\tau_i(t)]e^{j[\omega_c t-\phi_i(t)]}\right\}, \tag{2.2}$$

gdje je N broj signala sustava GNSS, ω_c nominalna frekvencija signala nosioca, A_i amplituda signala, $D_i(t)$ tok podataka signala (navigacijska poruka), $C_i(t)$ PRN kod, $\tau_i(t)$ faza koda, $\phi_i(t)$ faza nosioca, za svaki signal i.

Napadač odašilje slične signale, u kojima pokušava reproducirati nosioc i PRN kod te se lažni signal može prikazati kao:

$$y_s(t) = Re\left\{\sum_{i=1}^{N_s} A_{si} D_i[t-\tau_{si}(t)]C_i[t-\tau_{si}(t)]e^{j[\omega_c t-\phi_{si}(t)]}\right\}, \tag{2.3}$$

gdje su $\tau_{si}(t)$, $\phi_{si}(t)$ i A_{si} faze kodova, faze nosioca i amplitude lažnih signala. Njihove vrijednosti ovise o vrsti napada i razlikuju se od vrijednosti stvarnih signala. Napadač pokušava što bolje procijeniti tokove podataka koji su označeni s $\hat{D_i}(t)$. Pseudoslučajni nizovi lažnih signala moraju odgovarati stvarnim pseudoslučajnim nizovima kako bi se omogućilo uspješno lažiranje.

Ukupan primljeni signal tijekom napada lažiranjem jednak je:

$$y_{tot(t)} = y(t) + y_s(t) + v(t), \tag{2.4}$$

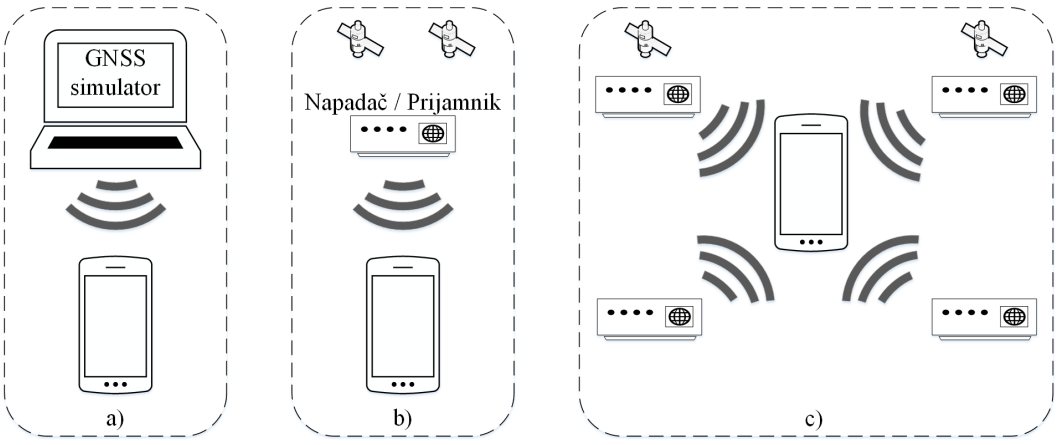
gdje je $v(t)$ primljeni šum. Primljeni šum ponekad može uključivati komponentu šuma koja se dodaje od strane napadača. Izvori bijelog šuma u GNSS prijamniku obično se opisuju temperaturnim šumom antene i prijamnika. Temperatura antene modelira šum koji ulazi u antenu s neba, dok temperatura prijamnika modelira toplinski šum zbog gibanja naboja unutar uređaja kao što je prednji dio prijamnika. Dodatni šum se javlja prilikom propagacije signala od antene do prijamnika kao šum aktivne (npr. pojačalo) ili pasivne (kabel) komponente.

2.4.1. Vrste napada lažiranjem

U novijim istraživanjima, različite vrste napada lažiranjem klasificirane su na temelju kompleksnosti napadača te na temelju poteškoća u detekciji napada lažiranjem sa strane prijamnika.

Iako postoji više vrsta napada, sve se svodi na dva temeljna načina izvođenja napada:

- lažni signali kreiraju se na način da nalikuju autentičnim signalima,
- emitiraju se signali snimljeni negdje drugdje u neko drugo vrijeme.



Slika 2.8: Vrste napada lažiranjem.

1. **Pojednostavljeni napad lažiranjem** (engl. *simplistic*) prikazan je na slici 2.8a. Ovaj napad temelji se na korištenju simulatora GNSS signala za kreiranje lažnog signala i njegovo odašiljanje kako bi se zavarao prijamnik. Ovu vrstu napada je veoma lako implementirati jer se koristi jeftina oprema. S druge strane, pojednostavljeni napad je lako detektirati s obzirom na to da je potrebna velika snaga lažnog signala kako bi prijamnik zanemario autentični satelitski signal i uzeo lažni, a uz to lažni signal nije sinkroniziran sa satelitskom konstelacijom. Obično se ovi napadi izvode na način da se prvo omete autentični GNSS signal kako bi se prijamnik prisilio na ponovno prikupljanje i zaključavanje na lažni signal. Rezultat pojednostavljenog napada su većinom skokovi u PVT izračunima [11].

   U [15], autori su pokazali da je lako lažirati lokacije pametnih telefona pomoću pojednostavljenog napada lažiranjem. Predloženi pristup je jednostavan i ekonomičan jer za izvođenje lažnog napada koristi jeftini SDR (HackRF One) [73] i simulator otvorenog pristupa GPS-SDR-SIM [71] koji je distribuiran pod MIT licencom [72]. Lažni signal je kreiran korištenjem simulatora GPS-SDR-SIM na temelju lokacije na koju se lažno želi locirati pametni telefon i navigacijske datoteke. Zatim se lažni kreirani signali prenose na SDR koji ih pretvara u RF signale. Eksperimentalni setup sastoji se od: HackRF One (predajnik koji odašilje GPS L1 signal), pametnog telefona (prijamnik) i ANT 500 antene. Parametri koji su promatrani u ovom eksperimentu su: broj vidljivih satelita, SNR satelita te lokacija pametnog telefona. U provedenim eksperimentima udaljenost između predajnika i prijamnika varira od 1 m do 7 m. Na udaljenostima do 5 m prijamnik dobiva signal dok na udaljenostima većim od 5 m prijamnik ne može primiti signal. Eksperimentom je zaključeno da je raspon prijenosa HackRF One 5 m. Pametni telefon je uspješno lažno lociran na željenu lokaciju (Mahatma Gandhi Institute of Technology (MGIT)) umjesto svoje stvarne lokacije (Chaitanya Bharathi Institute of Technology (CBIT)). Pokazano je da jeftini setup može lako preuzeti navigacijski sustav pametnog telefona.

   Autori u [16] istražuju učinke napada lažiranjem na jedinice za pozicioniranje i navigaciju na masovnom tržištu koje su integrirane u obične Android pametne telefone. Za izvođenje napada se također koriste HackRF One i GPS-SDR-SIM. Pokazano je da pametni telefoni imaju odličnu otpornost na pojednostavljene lažne napade (simplistic) ističući potencijalne slabosti koje treba zaštititi pomoću praktičnih obrambenih mehanizama i protumjera za lažne napade.

2. **Napad lažiranjem srednje razine složenosti** (engl. *intermediate*) ili napad temeljen na prijamniku prikazan je na slici 2.8b. Kod ove vrste napada, napadač ima ugrađen prijamnik koji prati i prikuplja parametre autentičnog satelitskog signala kako bi u skladu s tim signalom generirao lažirani signal te ga odašiljao ciljnom prijamniku. Ova vrsta napada je složena jer lažirani signali trebaju biti sinkronizirani s autentičnim

signalima. Izvedivost ovog napada je dokazana kao i mogućnost promjene položaja prijamnika bez podizanja upozorenja ili stvaranja diskontinuiteta u PVT rješenju [24].

3. **Sofisticirani napad lažiranjem** (engl. *sophisticated*) je najsloženija vrsta napada koja je prikazana na slici 2.8c. Ova vrsta napada koristi nekoliko napadača srednje razine koji generiraju i prenose lažne GNSS signale [11]. U ovom slučaju, napad se ne može jednostavno detektirati gledajući kut dolaska signala zbog toga što signali dolaze iz različitih kuteva i od različitih napadača. Međutim, ovi napadi imaju mnogo veću razinu složenosti zbog procesa sinkronizacije i komunikacije između svakog pojedinačnog odašiljača, što ga čini vrlo teškim za realizaciju i neprikladnim za scenarije u realnom vremenu. Također, sofisticirani napad lažiranjem nije isplativ ni što se tiče ekonomske strane jer zahtijeva dodatnu i skupu opremu (nekoliko napadača tj. predajnika i antena) [13].

3. Metode strojnog učenja za detekciju lažiranih signala

Strojno učenje dio je umjetne inteligencije koji se temelji na obučavanju računala i strojeva da uče i predviđaju kao ljudi odnosno gdje je sustavima omogućeno automatsko učenje i poboljšanje iz iskustva tj. treniranja. Za proces učenja, potrebno je imati neka opažanja ili uzorke kako bi se istražili temeljni obrasci skriveni u dostupnim podacima. Računalni sustavi uče te obrasce, koji su zapravo funkcije ili granice odlučivanja, bez ljudske intervencije. Uzorci koje sustav koristi za treniranje nazivaju se skupovi za treniranje. Nakon što se sustav istrenira na skupu za treniranje, potrebno ga je testirati na nekim novim ulaznim podacima i taj skup se naziva skup za testiranje.

Postoje četiri osnovne kategorije strojnog učenja s obzirom na to kako model uči iz danih podataka [80]:

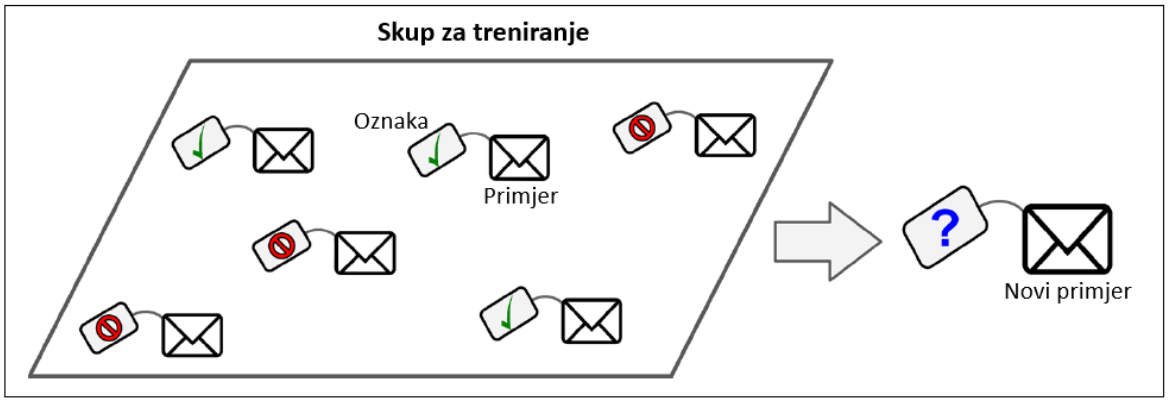
1. Nadzirano učenje zahtijeva označavanje podataka od strane čovjeka.

2. Nenadzirano učenje ne zahtijeva prvotno označavanje podataka nego model sam pokušava naučiti karakteristike podataka.

3. Polunadzirano učenje je kombinacija nadziranog i nenadziranog učenja. Model se prvo trenira na označenim podacima, što odgovara nadziranom učenju. Zatim koristi neoznačene podatke i pokušava pronaći uzorke u neoznačenim podacima na temelju onoga što je naučio u označenim podacima i to odgovara nenadziranom učenju.

4. Podržano učenje sastoji se od sustava učenja, koji se naziva agent koji promatra okolinu. Agent odabire i izvodi akcije, a zauzvrat dobiva pozitivne ili negativne nagrade. Na taj način pokušava naučiti najbolju strategiju kako bi dobio najveću nagradu.

Nadalje, postoje dvije temeljne vrste tehnika prediktivnog modeliranja u strojnom učenju: klasifikacija i regresija. Glavna razlika između klasifikacije i regresije je u vrsti rezultata koji se predviđa. Naime, predviđeni izlazi u klasifikaciji su diskretne/kategoričke vrijednosti dok regresija ima kontinuirane/numeričke izlazne vrijednosti. Klasifikacijski modeli imaju za cilj klasificirati ulazne podatke u jednu od unaprijed definiranih klasa. Tipičan nadzirani zadatak učenja je klasifikacija. Filtar za neželjenu poštu dobar je primjer za to: istreniran je s mnogo primjera e-pošte zajedno s njihovom klasom (ulazna ili neželjena pošta) i mora naučiti kako klasificirati nove e-poruke (slika 3.1).



Slika 3.1: Primjer nadziranog učenja - klasifikacija neželjene pošte [80].

U ovom istraživanju koriste se klasifikacijski modeli nadziranog učenja koji predviđaju tip satelitskog signala koji se prima na prijamniku. Korišteni modeli nadziranog učenja su opisani i uspoređeni u sljedećim poglavljima.

3.1. Primjena metoda strojnog učenja za detekciju lažiranih signala u sustavu GNSS

Kao što je već navedeno, metode strojnog učenja su pouzdan i učinkovit pristup za detekciju napada lažiranjem i klasifikaciju tipa signala. Model K-najbližih susjeda korišten je u radu [17] za otkrivanje napada lažiranjem s malim kašnjenjem. Detekcija se provodila promatranjem vrhova signala korelacije i postignuta je točnost detekcije oko 95% za kašnjenje oko 0.4 čipa. Za klasifikaciju signala sustava GPS, u [67] koriste stablo odlučivanja, linearni SVM te slučajne šume. SVM metoda pokazuje najbolje rezultate za detekciju lažnih signala sustava GPS u [25]. S druge strane, model KNN ostvaruje nabolje rezultate u [26]. U radovima [40], [44], [45], [46] za detekciju se koristi model SVM koji daje vrlo dobre rezultate. Modeli SVM, K-najbližih susjeda i stablo odlučivanja koriste se u radovima [29], [30] i [103]. Nadalje, radovi [19], [59], [60], [61] isto koriste model SVM kao model s najboljim performansama za detekciju lažnih signala. Optimizirani algoritam slučajne šume korišten je u radu [65] za detekciju napada lažiranjem.

3.2. Metoda potpornih vektora (SVM)

Metoda potpornih vektora je vrsta nadziranog algoritma strojnog učenja koji se koristi za klasifikaciju i regresiju. Posebno je koristan za rješavanje složenih klasifikacijskih problema gdje podaci nisu linearno odvojivi. Ova metoda se temelji na ideji maksimalne margine odnosno osnovni cilj je postaviti hiperravninu u N-dimenzionalnom prostoru tako da bude najviše udaljena od primjera iz dviju klasa koje su najčešće označene kao -1 i +1 te se stoga

postiže manja osjetljivost na šum ili odstupanje u podacima. Dakle, margina je udaljenost od hiperravnine do najbližeg primjera sa svake strane. Jednostavno rečeno, ideja je pronaći optimalnu hiperravninu koja će maksimizirati marginu između dvije klase.

Model potpornih vektora je najjednostavniji i opisan je izrazom [83]:

$$h(\mathbf{x};\mathbf{w},w_0) = \mathbf{w}^T\mathbf{x} + w_0, \tag{3.1}$$

gdje je h rezultat modela (predikcija), $\mathbf{x}$ vektor ulaznih podataka, $\mathbf{w}$ vektor težina, T transpozicija kako bi se omogućilo skalarno množenje 2 vektora i w_0 slobodan član (bias). Često se umjesto oznake w_0 koristi oznaka b. Bias pomjera hiperravninu lijevo ili desno. Ukoliko je $w_0 = 0$, onda hiperravnina uvijek prolazi kroz središte koordinatnog sustava, što može ograničiti model i smanjiti njegovu sposobnost pravilnog razdvajanja podataka. Granica između klasa je hiperravnina koja je određena točkama x za koje vrijedi $h(\mathbf{x};\mathbf{w},w_0) = 0$, odnosno:

$$\mathbf{w}^T\mathbf{x} + w_0 = 0. \tag{3.2}$$

Predikcija modela $h(x)$ ovisi o tome je li uzorak na jednoj ili na drugoj strani hiperravnine što se može detektirati na temelju predznaka pa je predikcija oznake (klase) jednaka:

$$y = sgn(h(x)) = \begin{cases} +1, & \text{ako } \mathbf{w}^T\mathbf{x} + w_0 > 0 \\ -1, & \text{ako } \mathbf{w}^T\mathbf{x} + w_0 < 0 \end{cases}. \tag{3.3}$$

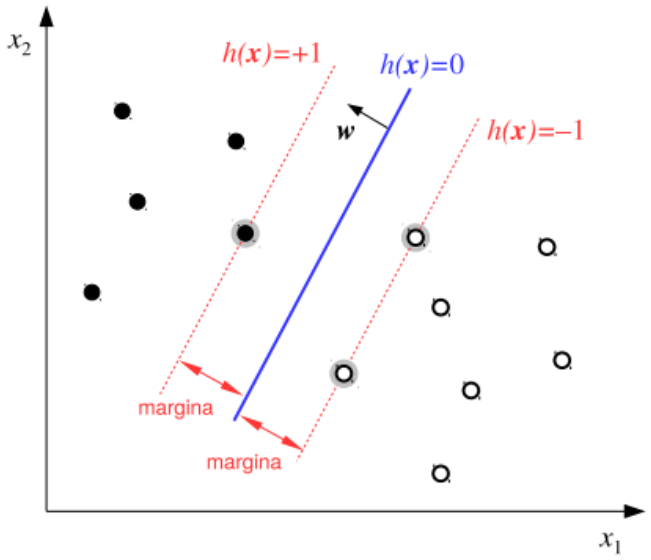
Točke koje su najbliže margini nazivaju se potporni vektori i zadovoljavaju sljedeće uvjete:

$$\mathbf{w}^T\mathbf{x}_i + w_0 = \pm 1. \tag{3.4}$$

Potporni vektori leže na marginama udaljenim $\frac{1}{\|\mathbf{w}\|}$ (ovaj izraz predstavlja širinu margine) od hiperravnine. Oni su važni zato što utječu na granicu odlučivanja.

Slika 3.2 prikazuje ilustraciju metode potpornih vektora u situaciji kada imamo uzorke dviju klasa $y = +1$ (crni kružići) i $y = -1$ (bijeli kružići). Pravac za koji vrijedi $h(x) = 0$ je granica. Uzorci koji se nalaze s lijeve strane pravca su oni koji pripadaju klasi +1 i za njih vrijedi $h(x) = +1$, a oni koji se nalaze s desne strane pravca pripadaju klasi -1 i za njih vrijedi $h(x) = -1$. Tri uzorka za koje vrijedi $|yh(x) = 1|$ nalaze se na margini i zovemo ih potporni vektori.

Margina predstavlja udaljenost hiperravine do najbližeg uzorka. Budući da je cilj ove metode maksimalna margina, potrebno je pronaći takvu hiperravninu koja će maksimizirati tu udaljenost:



Slika 3.2: Ilustracija metode potpornih vektora [81].

$$\arg\max_{\mathbf{w},w_0}\left\{\frac{1}{\|\mathbf{w}\|}\min_i\left\{y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)}+w_0)\right\}\right\}, \tag{3.5}$$

gdje se za svaku hiperravninu (funkcija *argmax* radi iteracije po svim hiperravninama) računa minimalna udaljenost do najbližeg uzorka te se uzima ona hiperravnina za koju je ta udaljenost najveća. Za hiperravninu koja maksimizira marginu, uzorci s lijeve i desne strane jednako su udaljeni od nje.

Konačan izraz za optimizacijski problem maksimalne margine odnosno tzv. problem kvadratnog programiranja je:

$$\arg\min\ \frac{1}{2}\|\mathbf{w}\|^2, \tag{3.6}$$

uz uvjet:

$$y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)}+w_0)\geqslant 1, \quad i=1,\ldots,N, \tag{3.7}$$

gdje je i redni broj uzorka i N ukupan broj uzoraka.
Ukoliko podaci nisu linearno odvojivi, SVM koristi jezgrenu funkciju (engl. *kernel function*) za mapiranje podataka u višedimenzionalni prostor gdje postaju linearno odvojivi. Najčešće korištene jezgrene funkcije su linearna, polinomna i radijalna bazna funkcija (engl. *Radial Basis Function - RBF*). Nakon što se pronađe optimalna hiperravnina, SVM se može koristiti za klasificiranje novih podatkovnih točaka na način da se provjeri na kojoj strani hiperravnine se nalaze. Ukratko, metoda potpornih vektora je moćan algoritam za rješavanje složenih problema klasifikacije pronalaženjem optimalne hiperravnine koja maksimizira marginu između klasa.

3.3. Metoda K-najbližih susjeda (KNN)

Metoda K-najbližih susjeda je neparametarska metoda nadziranog učenja. Iako se KNN algoritam može koristiti i za probleme regresije i klasifikacije, obično se koristi kao algoritam klasifikacije, polazeći od pretpostavke da se slične točke mogu pronaći jedna blizu druge. To je vrsta učenja temeljenog na instancama ili lijenog učenja, gdje model ne uči funkciju iz podataka za treniranje, već umjesto toga pohranjuje podatkovne točke za treniranje u memoriju kako bi napravio predviđanja na temelju sličnosti s novom podatkovnom točkom. Stoga se svi izračuni izvode u trenutku kada se traži predviđanje.

Metoda KNN koristi udaljenost za klasifikaciju ili predviđanje grupiranja jedne podatkovne točke. Polazi se od pretpostavke da se slične podatkovne točke nalaze u blizini jedna drugoj te se oznaka klase dodjeljuje na temelju većine glasova tj. promatranoj podatkovnoj točki dodjeljuje se ona klasa koja je najčešće zastupljena oko nje [84], [85].

Algoritam KNN sastoji se od nekoliko koraka [87]:

1. Odabir K koji definira koliko će susjeda biti provjereno da bi se odredila klasa određene podatkovne točke. Primjerice, ako je $K = 1$, instanca će biti dodijeljena istoj klasi kao i njezin najbliži susjed.

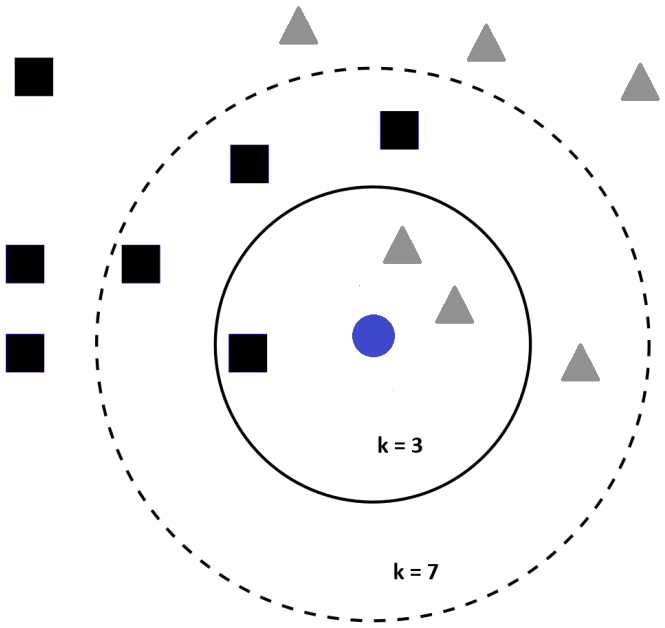
2. Odrediti metriku udaljenosti. Kako bi se odredilo koje su podatkovne točke najbliže točki za koju treba odrediti klasu, treba se izračunati udaljenost između promatrane točke i ostalih podatkovnih točaka. Ove metrike udaljenosti pomažu u formiranju granica odlučivanja, koje dijele upitne točke u različite regije. S obzirom na novu podatkovnu točku, algoritam izračunava udaljenost između nove točke i svih točaka u skupu podataka za treniranje kako bi pronašao najbliže susjede nove podatkovne točke.

   Slika 3.3 prikazuje primjer algoritma K-najbližih susjeda s dvije klase: crni kvadratići i sivi trokutići. Nova podatkovna točka označena je plavim krugom i treba je dodijeliti jednoj od klasa. Ukoliko se uzme $K = 3$, nova podatkona točka će biti klasificirana kao sivi trokutić. S druge strane, ukoliko se uzme $K = 7$, ona će biti klasficirana kao crni kvadratić.

   Postoje različite metrike udaljenosti i neke od njih za promatrane točke x i y su [86], [87]:

   - Euklidska udaljenost, poznata i kao $L2$ norma, mjeri najkraću udaljenost između dviju točaka u ravnoj liniji i definira se kao:

$$d_{L2}(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{\frac{1}{2}}, \tag{3.8}$$



Slika 3.3: Primjer algoritma K-najbližih susjeda [88].

gdje $|x_i - y_i|$ predstavlja apsolutnu razliku između koordinata x i y u svakoj dimenziji, i predstavlja indeks koji prolazi kroz sve elementa vektora i n predstavlja dimenziju vektora.

- Manhattan udaljenost, poznata i kao $L1$ norma, predstavlja udaljenost između dviju točaka izračunavanjem zbroja apsolutnih razlika njihovih koordinata:

$$d_{L1}(x,y) = \sum_{i=1}^{n} |x_i - y_i|. \tag{3.9}$$

- Minkowski udaljenost, poznata i kao L_p norma, predstavlja generalizaciju Euklidske i Manhattan udaljenosti i može se definirati kao:

$$d_{L_p}(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}, \tag{3.10}$$

gdje je p parametar koji određuje vrstu udaljenosti. Kad je $p = 1$, Minkowski udaljenost postaje Euklidska udaljenost, a za $p = 2$, Minkowski udaljenost postaje Manhattan udaljenost.

3. Odabir K-najbližih susjeda na temelju izračunate udaljenosti. Susjedi su K točaka za treniranje koji se nalaze najbliže novoj podatkovnoj točki. U slučaju binarne klasifikacije, granica odlučivanja je pravac koji razdvaja dvije klase. S druge strane, u slučaju klasifikacije s više klasa, granica je hiperravnina koja razdvaja različite klase.

4. Dodjela klase novoj podatkovnoj točki. Nakon pronalaska K-najbližih susjeda, algoritam dodjeljuje novu podatkovnu točku klasi koja je najčešća među njegovim K-

najbližim susjedima. Drugim riječima, algoritam koristi većinsku klasu K-najbližih susjeda kao predviđenu klasu za novu podatkovnu točku.

5. Algoritam ponavlja korake 2-4 za svaku novu podatkovnu točku.

Prilikom korištenja algoritma KNN, treba uzeti u obzir neke bitne stvari:

- Vrijednost K može utjecati na performanse algoritma. Mala vrijednost može dovesti do pretreniranja modela (engl. *overfitting*) dok velika vrijednost može uzrokovati pristranost (engl. *underfitting*). Pretreniranje znači da model pored stvarnih podataka može učiti ili memorirati neke greške ili šum. S druge strane, pristranost predstavlja nemogućnost modela da nauči stvarnu strukturu podataka bez obzira na njihovu količinu [89].

- Odabir metrike udaljenosti također može utjecati na performanse algoritma. Primjerice, Manhattan udaljenost može biti korisna u slučajevima kada značajke imaju različite jedinice ili skale i ne mogu se usporediti na istoj skali. Izračun Minkowski udaljenosti je računalno vrlo zahtjevan zato što uključuje uzimanje p-te potencije razlika između koordinata. Stoga, odabir velike vrijednosti p može dovesti do sporijeg rada algoritma.

- Algoritam KNN može biti računski vrlo zahtjevan posebno ukoliko se radi o velikim skupovima podataka pa i o tome treba voditi računa prilikom odabira ovog algoritma.

3.4. Metoda stabla odlučivanja (DT)

Metoda stabla odlučivanja ima algoritam za klasifikaciju i regresiju koji razdvaja podatke u hijerarhijsku strukturu s granama (odlukama), čvorovima i listovima (krajnjim klasama ili vrijednostima). Svaki čvor u stablu predstavlja odluku temeljenu na nekom kriteriju (npr. prag vrijednosti značajke), a listovi predstavljaju konačne odluke. Svaki čvor klasifikacijskog stabla odlučivanja sadrži test koji se primjenjuje na značajki uzorka (npr. uzorak automobil sa značajkom snaga motora), svaka grana predstavlja ishod testa, a svaki list predstavlja oznaku klase. Dubina stabla predstavlja ukupan broj čvorova od korijena do najdaljeg lista [90].

Algoritam se sastoji od nekoliko koraka [91]:

1. Početak na vršnom čvoru (engl. *root node*) koji predstavlja cijeli skup podataka. Tu se skup podataka počinje dijeliti na temelju različitih značajki ili uvjeta.

2. Traženje najvažnije značajke ili pitanja koje dijeli podatke u najrazličitije skupine.

3. Grananje - na temelju odgovora na to pitanje podaci se dijele na manje podskupove, stvarajući nove grane. Svaka grana predstavlja mogući put kroz stablo. Čvorovi koji nastaju razdvajanjem vršnog čvora poznati su kao čvorovi odlučivanja. Ovi čvorovi predstavljaju srednje odluke ili uvjete unutar stabla.

4. Algoritam nastavlja postavljati pitanja i dijeliti podatke u svakoj grani sve dok ne dođe do konačnih čvorova lista (engl. *leaf nodes*), koji predstavljaju predviđene ishode ili klasifikacije.

Iako postoji više načina za odabir najbolje značajke na svakom čvoru, dvije metode, Gini nečistoća (engl. *Gini impurity*) i srednji uzajamni sadržaj informacije (engl. *information gain*), predstavljaju popularne kriterije razdvajanja za modele stabla odlučivanja. Oni pomažu procijeniti kvalitetu svakog uvjeta ispitivanja i koliko će dobro moći klasificirati uzorke [92].

- Gini nečistoća ili Gini indeks predstavlja vjerojatnost netočnog klasificiranja nasumične podatkovne točke u skupu podataka ako je ista označena na temelju distribucije klase skupa podataka. Računa se tako da se zbroje vjerojatnosti odabira svakog uzorka pomnožene s vjerojatnošću krive klasifikacije tog uzorka. Ima vrijednost između 0 i 1. Gini indeks 0 znači da su uzorci savršeno homogeni i da su svi elementi slični (spadaju u jednu klasu), dok Gini indeks 1 znači maksimalnu nejednakost među elementima [93]. Definira se kao:

$$\text{Gini} = 1 - \sum_{i=1}^{n} p_i^2, \tag{3.11}$$

  gdje je n broj klasa, p_i vjerojatnost klasifikacije značajke u čvoru.

- Srednji uzajamni sadržaj informacije koristi koncept entropije. Entropija je definirana u fizici, a mjera koja mjeri količinu informacije $H(S)$ koju nosi neka poruka ekvivalentna je entropiji i definira se kao:

$$H(S) = -\sum_{i=1}^{n} p_i \log_2(p_i), \tag{3.12}$$

  gdje je S skup podataka.

  Ukoliko je izvršena normalizacija, vrijednosti entropije mogu biti između 0 i 1. Ako svi uzorci u skupu podataka, pripadaju jednoj klasi, tada je entropija jednaka 0. Ukoliko je polovica uzoraka klasificirana kao jedna klasa, a druga polovica kao druga klasa, entropija je maksimalna i iznosi 1. Kako bi se odabrala najbolja značajka za podjelu i pronašlo optimalno stablo odlučivanja, trebao bi se koristiti atribut s najmanjom količinom entropije. Drugim riječima, što je entropija manja, čvor je bolje razdvojen.

Srednji uzajamni sadržaj informacije predstavlja razliku u entropiji prije i nakon podjele na danoj značajki. Značajka s najvećim srednjim uzajamnim sadržajem informacije daje najbolju podjelu jer najbolje klasificira podatke u skupu za treniranje. Definiran je kao:

$$IG = H(S) - \sum_{j=1}^{k} \frac{|S_j|}{|S|} H(S_j), \tag{3.13}$$

gdje je S_j podskup podataka nakon podjele, j je redni broj vrijednosti značajke i k je broj mogućih vrijednosti značajke.

Stabla odlučivanja su dobra za lako razumijevanje podataka i brzu klasifikaciju (brzo treniranje skupa podataka i brza predikcija) te su pogodna za velike skupove podataka. Ne zahtijevaju normalizaciju podataka niti skaliranje te je stoga jednostavna priprema podataka za korištenje ovog modela. S druge strane, postoje i neki nedostaci kao što su: pretreniranje kada model pored stvarnih podataka uči i greške ili šum; mala promjena u podacima može dovesti do značajne promjene u strukturi stabla; često postiže nižu točnost u usporedbi s nekim drugim metodama primjerice slučajnim šumama. Stoga se za bolju točnost i robusnost često koriste slučajne šume koje su opisane ispod.

3.5. Metoda slučajne šume (RF)

Metoda slučajne šume su metoda nadziranog strojnog učenja koja se koristi za klasifikaciju i regresiju. Temelji se na kombiniranju nekoliko stabala odlučivanja zbog poboljšanja točnosti i smanjivanja pretreniranja [94]. Umjesto da koristi jedno stablo, model kreira više stabala i trenira svako stablo na slučajnom podskupu podataka (engl. *bagging* metoda). Za zadatke klasifikacije, konačna odluka je klasa koju je odabrala većina stabala [95], [96].

Algoritam slučajne šume sastoji se od nekoliko koraka što je prikazano i na slici 3.4:

1. Iz originalnog skupa podataka nasumično se uzima podskup istih podataka veličine n, ali s ponavljanjem što znači da se neki uzorci mogu pojaviti više puta dok drugi mogu biti potpuno izostavljeni. Na taj način se formira tzv. *bootstrap* uzorak koji se koristi za treniranje jednog stabla odlučivanja te se tako smanjuje pretreniranje (svako stablo vidi samo dio podataka) i povećava raznolikost modela (svako stablo uči na drugačijem uzorku). Ovaj postupak ponavlja se za svako stablo u modelu.

2. Slučajni podskup od m značajki odabire se iz ukupnog skupa od p značajki. To pomaže pri smanjenju korelacije između stabala.

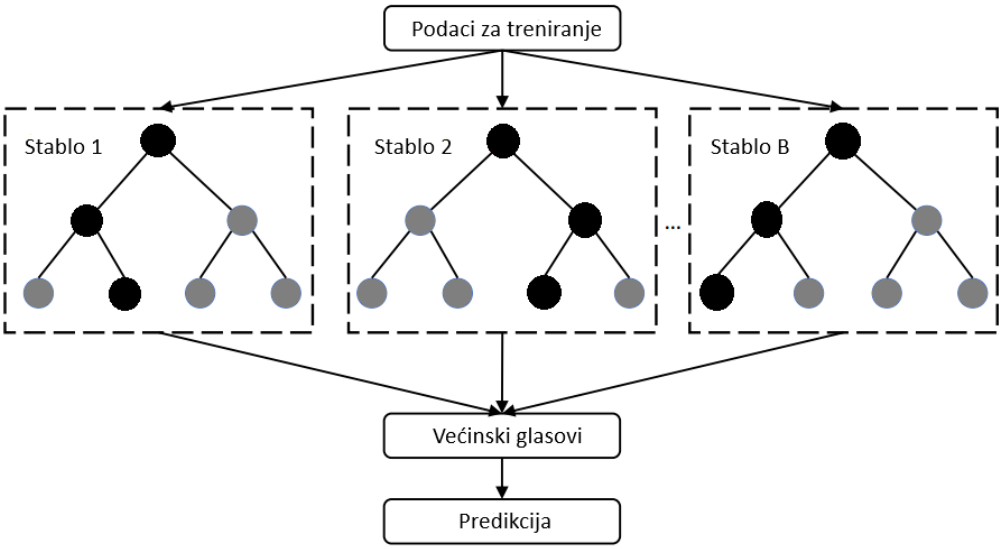
3. Stablo odlučivanja gradi se za svaki *bootstrap* uzorak i odabrane značajke koristeći određeni kriterij razdvajanja (prethodno spomenuti Gini nečistoća i srednji uzajamni

sadržaj informacija). Proces se ponavlja dok se ne ispuni unaprijed određeni kriterij zaustavljanja (npr. maksimalna dubina stabla, minimalni broj uzoraka po listu).

4. Konačna predikcija (klasa) modela dobije se zbrajanjem predviđanja svih stabala odlučivanja i definira se kao:

$$\hat{y} = \arg\max_{c} \sum_{b=1}^{B} 1(h_b(x) = c), \tag{3.14}$$

gdje je $\hat{y}$ konačna klasa, c klasa za koju se računaju glasovi, B broj stabala u šumi, $h_b(x)$ predikcija pojedinog stabla za ulazni podatak x i $1(h_b(x) = c)$ indikatorska funkcija koja daje 1 ukoliko je $h_b(x) = c$, a inače je 0 [96].



Slika 3.4: Primjer modela slučajnih šuma [97].

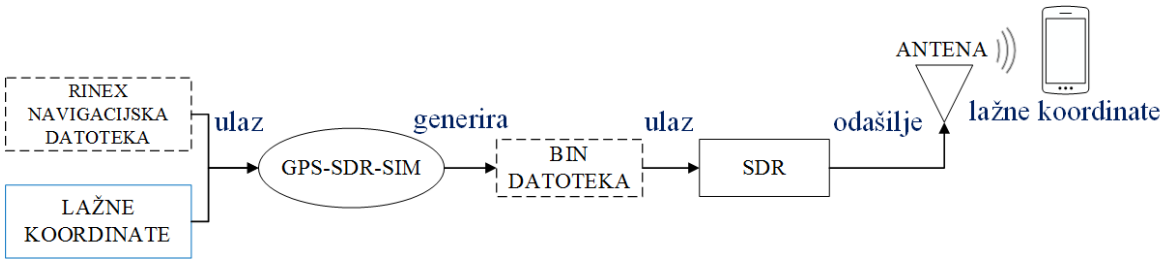
Algoritam slučajne šume je otporniji na pretreniranje jer koristi više različitih stabala i zbraja njihove odluke. Nadalje, dobar je za velike skupove podataka, može raditi s mješovitim tipovima podataka (numerički i kategorički) i postiže vrlo visoku točnost. *Bootstrapping* omogućava da RF izgradi više različitih stabala odlučivanja, što poboljšava robusnost i preciznost modela. Nedostatak ovog algoritma je visoka računska složenost u odnosu na pojedinačno stablo (više stabala = veća računska složenost).

4. Detekcija napada lažiranjem u sustavu GNSS

4.1. Izvođenje napada lažiranjem pomoću softverski definiranog radija

Glavni i najčešće korišteni dio opreme za izvođenje napada lažiranjem je softverski definirani radio. Jedan jeftini softverski definirani radio može vrlo lako preuzeti navigacijski sustav pametnih telefona i lažirati njihove lokacije što može biti vrlo opasno.

SDR sustavi sastoje se od analogne korisničke aplikacije (engl. *front-end*) i digitalne poslužiteljske aplikacije (engl. *back-end*). Analogni dio upravlja funkcijama za odašiljanje i primanje.



Slika 4.1: Blok dijagram.

Slika 4.1 prikazuje blok dijagram izvođenja pojednostavljenog napada lažiranjem. Za izvođenje napada potrebno je prikupiti vlastitu RINEX (engl. *Receiver Independent Exchange Format*) navigacijsku datoteku ili istu preuzeti s NASA-ine stranice [108] te definirati lažne koordinate na koje želimo staviti prijamnik ili pametni telefon. RINEX datoteka služi za zapisivanje neobrađenih podataka primljenih sa satelita. Korištenjem navigacijske datoteke i lažnih koordinata kao ulaz za program npr. GPS-SDR-SIM, kreira se bin datoteka koja se odašilje na softverski definirani radio i s njega dalje na prijamnik koji bi ovisno o udaljenosti i snazi predajnika trebao kroz nekoliko sekundi/minuta pokazivati lažnu lokaciju. Predajnik prenosi I/Q modulirane signale sustava GPS na frekvenciji L1 1575.42 MHz.

Na slici 4.2 prikazana je oprema potrebna za izvođenje napada lažiranjem korištenjem softverski definiranog radija: 1 - laptop, 2 - softverski definirani radio HackRF One, 3 - antena ANT500, 4 - eksterni oscilator i 5 - pametni telefoni. Osim hardverske opreme, simulator otvorenog pristupa GPS-SDR-SIM [71] instaliran je na laptopu i korišten za odašiljanje lažiranih signala sustava GPS.

Detaljni podaci o korištenoj opremi su [103], [104]:

1. Dell Vostro 15 3510, 11th Gen Intel(R) Core(TM) i7-1165G7 @2.80 GHz procesor i 16 GB RAM memorije,
2. SDR HackRF One [73] odašilje / prima bilo koji radio signal od 1 MHz do 6 GHz, to je poludupleksan primopredajnik s ograničenom snagom prijenosa,
3. ANT500 štapna antena,
4. TCXO eksterni oscilator za praćenje signala sustava GNSS,
5. pametni telefoni (Android, iPhone).

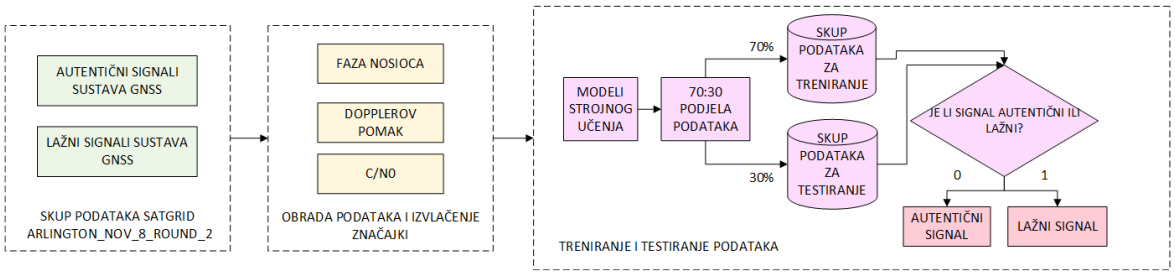
Slika 4.2: Oprema za izvođenje napada lažiranjem.

U našem eksperimentu, pojednostavljeni napad lažiranjem korištenjem softverski definiranog radija izveden je u unutarnjim uvjetima na hodniku i vanjskim uvjetima ispred Fakulteta elektrotehnike, strojarstva i brodogradnje, Sveučilište u Splitu.

4.1.1. Detekcija napada lažiranjem i klasifikacija tipa signala

Prvi dio istraživanja odnosi se na korištenje metoda strojnog učenja za detekciju napada lažiranjem i klasifikaciju tipa signala.

Na slici 4.3 prikazan je dijagram toka modela strojnog učenja za klasifikaciju tipa signala, koji se sastoji od četiri koraka:



Slika 4.3: Dijagram toka modela strojnog učenja za klasifikaciju signala [103].

1. odabir skupa podataka,
2. obrada podataka i izvlačenje značajki,
3. treniranje i testiranje podataka primjenom strojnog učenja,
4. klasifikacija tipa signala i detekcija lažiranih signala.

Prvi korak je prikupljanje skupa podataka (autentični i lažni signali). U drugom koraku se izvlače parametri po kojima će se vršiti klasifikacija signala. Zadnji korak je primjena metoda strojnog učenja tj. treniranje i testiranje modela na prikupljenim podacima. Kao rezultat model klasificira signale na autentične i lažne na temelju parametara korištenih za treniranje i testiranje.

U eksperimentu su korištena dva skupa podataka [103]:

- SatGrid *Arlington_Nov_8_Round_2* skup podataka, koji sadrži autentične i lažirane signale sustava GPS prikupljene na različitim zemljopisnim lokacijama (Arlington Virginia, Blacksburg Virginia, Missouri Texas), vremenima i uvjetima okoline [107]. Autentični signali prikupljeni su na krovu istraživačkog centra Virginia Tech, Arlington, Virginia u trajanju od 50 minuta. Originalni podaci (*SatGrid:G25*) su zatim reproducirani na različitim razinama snage pomoću softverski definiranog radija USRP B210 u laboratorijskom okruženju za generiranje skupa lažiranih signala (*SatGrid:S10*). U našem eksperimentu korišteno je 10 000 uzoraka iz ovog skupa podataka.

- Vlastiti prilagođeni skup podataka je modificirana datoteka efemerida s podacima sa stvarnih satelita za određeni dan i sat koja je preuzeta sa stranice NASA [108]. Skup podataka modificiran je s lažnim datumom, vremenom i lokacijom te odašiljan sa softverski definiranog radija HackRF One na pametni telefon. Aplikacija GNSSLogger [109] je korištena za prikupljanje neobrađenih podataka na pametnom telefonu. RINEX 3.03 datoteka s neobrađenim podacima, koji su kasnije obrađeni u programu Matlab, generirana je u aplikaciji GNSSLogger. Prikupljeni skup podataka sastoji se od autentičnih i lažiranih podataka. S obzirom da je ovaj skup podataka odašiljan u kontroliranim unutarnjim uvjetima i da je napadač relativno blizu žrtvi, očekuje se da

je razlika između autentičnih i lažiranih podataka značajna te se može očekivati da metode strojnog učenja klasificiraju podatke s visokom točnošću. Nadalje, kako bismo testirali i potvrdili predložene modele i vlastiti skup podataka, namjerno su uvedene greške tj. nekim uzorcima dodijeljena je pogrešna klasa (autentičan ili lažiran). Ovaj prilagođeni skup podataka koji ima 5665 uzoraka omogućuje nam veću kontrolu nad eksperimentalnim uvjetima, a time i bolju prilagodbu modela stvarnim izazovima s kojima se možemo suočiti u stvarnosti.

Tablica 4.1: Značajke korištene za klasifikaciju tipa signala [103]

Značajka	*Tip*
Faza nosioca [rad]	ulaz
Dopplerov pomak [Hz]	ulaz
Omjer snage signala nosioca i šuma (C/N_0) [dB-Hz]	ulaz
klasa	izlaz

U sljedećem koraku izdvojene su tri najznačajnije značajke koje utječu na satelitski signal: faza nosioca, Dopplerov pomak i omjer signala nosioca i šuma C/N_0 što je prikazano u Tablici 4.1. Navedene značajke korištene su kao ulazni podaci za treniranje i testiranje modela. Izlazni podatak označava klase tj. tip signala: autentični (klasa 0) i lažirani (klasa 1). C/N_0 je najznačajnija značajka jer je prema vrijednosti ove značajke lako zaključiti je li signal lažiran ili ne - veća vrijednost ukazuje na napad lažiranjem. Napad lažiranjem teže je detektirati po ostalim značajkama, a pogotovo ako je razlika u fazi nosioca između autentičnog i lažiranog signala malena.

Rezultati eksperimenata

Slika 4.4 prikazuje tekstualnu RINEX datoteku s prikupljenim podacima na pametnom telefonu. Iz RINEX datoteke mogu se vidjeti podaci o prijamniku, vidljivim satelitima, parametrima koje prijamnik može prikupiti kao i brojčanim vrijednostima tih parametara, itd. Primjerice, posljednji stupac prikazuje vrijednosti C/N_0 i prema vrlo viskom vrijednostima može se pretpostaviti da su primljeni signali lažni.

Na slici 4.5 prikazane su snimke zaslona iz aplikacije GPS Test u trenutku kada pametni telefon još uvijek nije postigao "position fix", tj. nije izačunao svoj položaj koristeći podatke s minimalno četiri satelita što je prikazano na slici 4.5a. Slika 4.5b prikazuje trenutak kada prijamnik ima postignut "3D position fix" odnosno ima izračunat svoj 3D položaj na temelju lažnih signala odašiljanih od strane napadača što se vidi po vrlo visokim vrijednostima C/N_0. Dakle, može se zaključiti da je napad lažiranjem uspješno izveden.

Nakon izvlačenja značajki, 70% uzoraka svakog skupa podataka je korišteno za treniranje, a preostalih 30% za testiranje četiri različita modela strojnog učenja. Uzorci skupova

```
     3.03           OBSERVATION DATA    M                   RINEX VERSION / TYPE
GnssLogger          samsung 14          20240630 100136 UTC PGM / RUN BY / DATE
Google GnssLogger                                          MARKER NAME
Unknown                                                    MARKER NUMBER
Unknown             Unknown                                OBSERVER / AGENCY
Unknown             GnssLogger          v3.0.6.4           REC # / TYPE / VERS
Unknown             Unknown                                ANT # / TYPE
        0.0000        0.0000        0.0000                 ANTENNA: DELTA H/E/N
G    4 C1C L1C D1C S1C                                     SYS / # / OBS TYPES
J    4 C1C L1C D1C S1C                                     SYS / # / OBS TYPES
C    4 C2C L2C D2C S2C                                     SYS / # / OBS TYPES
  2024    06    30    10    01   36.0000000     GPS        TIME OF FIRST OBS
                                                           GLONASS SLOT / FRQ #
G L1C                                                      SYS / PHASE SHIFT
J L1C                                                      SYS / PHASE SHIFT
C L2C  0.00000                                             SYS / PHASE SHIFT
 C1C    0.000 C1P    0.000 C2C    0.000 C2P    0.000      GLONASS COD/PHS/BIS
                                                           END OF HEADER
> 2024 06 30 10 01 36.1000592  0 4
G28  22156268.03509  -6725560.30309      -656.12309        55.50009
G16  24083268.90008  -6581556.71308      -611.46108        52.80008
G26  24019421.50108  -6543454.44608       -76.62508        52.20008
G10  24930485.98408  -6575507.30108      -519.68608        50.60008
> 2024 06 30 10 01 37.0000058  0 4
G28  22156242.85209  -6724968.98409      -659.46609        55.50009
G16  24083234.72408  -6581004.87708      -613.51608        52.80008
G26  24019294.98808  -6543384.94108       -79.04808        52.20008
G10  24930435.31908  -6575038.02808      -522.33808        50.50008
> 2024 06 30 10 01 38.0000000  0 4
G28  22156450.30909  -6724309.23909      -660.87109        55.40009
G16  24083434.38508  -6580390.80808      -614.83308        52.80008
G26  24019392.12108  -6543305.69108       -80.02008        52.20008
G10  24930616.69308  -6574515.65408      -523.16108        50.50008
> 2024 06 30 10 01 39.0000001  0 4
G28  22156593.60909  -6723650.84209      -657.72109        55.40009
G16  24083566.89408  -6579778.87108      -610.86808        52.70008
G26  24019423.29908  -6543228.50908       -76.14808        52.10008
G10  24930733.61208  -6573995.33808      -519.28508        50.40008
```

Slika 4.4: Primjer prikaza podataka primljenih na pametnom telefonu u obliku RINEX datoteke.

za treniranje i testiranje prikupljeni su na istim lokacijama i to može dati vrlo visoku točnost klasifikacije. Predloženi model predviđa klase (autentičan ili lažan signal) korištenjem novih ulaznih podataka na kojima nije treniran. Budući da sve značajke imaju vrijednosti izmjerene na različitim skalama, potrebno je prilagoditi sve vrijednosti na zajedničku skalu. Min-max skaliranje značajki koristi se za dovođenje svih vrijednosti u raspon $[0,1]$. Ovo se također naziva normalizacija temeljena na jedinstvu (engl. *unity-based normalization*), a definira se kao:

$$X_n = \frac{X - X_{min}}{X_{max} - X_{min}}, \tag{4.1}$$

gdje je X_n normalizirana vrijednost značajke, X je originalna vrijednost značajke, X_{min} minimalna vrijednost značajke i X_{max} je maksimalna vrijednost.

Pet evaluacijskih metrika je korišteno za usporedbu odabranih modela strojnog učenja. Tablica 4.2 prikazuje predviđene točnosti klasifikacije i varijance za nove ulazne podatke za odabrane modele strojnog učenja. Može se zaključiti da svi modeli imaju viskou točnost klasifikacije te da su uzorci u korištenim skupovima podataka jako dobri (uzorci se mogu

(a) Nije postignut "position fix". (b) Postignut je "position fix".

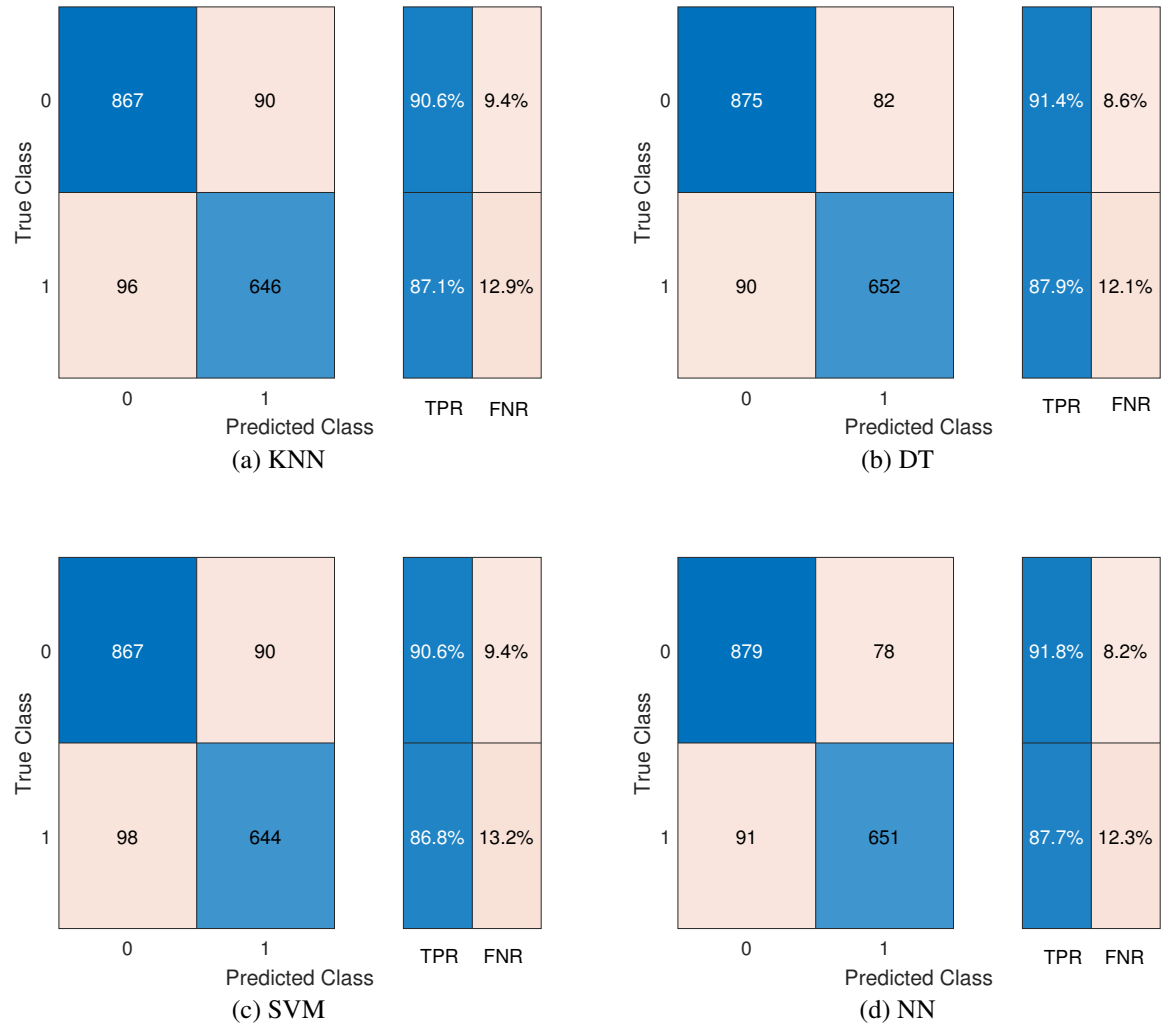
Slika 4.5: Snimke zaslona iz aplikacije GPS test za dva slučaja.

dobro klasificirati budući da značajke imaju specifične vrijednosti prema kojima im se lako dodjeljuje klasa). Skup podataka SatGrid ima nešto bolju točnost za sve modele zbog toga što su u naš skup podataka namjerno unesene pogreške zbog testiranja. To je i razlog zašto su varijance u našem skupu podataka veće. Točnost klasifikacije računa se prema izrazu (6.20).

Tablica 4.2: Točnost klasifikacije za odabrane modele strojnog učenja

Model	*Točnost [%]*	*Varijanca [%]*
	Skup podataka SatGrid	
SVM	**99.7**	0.3
KNN	99.67	0.33
Stablo odlučivanja	99.2	0.8
Neuralne mreže	98.03	1.97
	Vlastiti skup podataka	
Neuralne mreže	**90.05**	9.95
Stablo odlučivanja	89.88	10.12
KNN	89.05	10.95
SVM	88.93	11.07

Konfuzijska matrica za naš vlastiti skup podataka je prikazana na slici 4.6. Rezultati klasifikacije pokazuju manju točnost za sve metode u usporedbi sa skupom podataka SatGrid. Razlog tome je modifikacija klasa za korištene parametre kako bi se namjerno unijele pogreške radi testiranja ponašanja modela. Neuralne mreže 4.6d imaju najbolji rezultat klasifikacije s točnošću od 90.05%. 879 uzoraka je ispravno klasificirano kao autentični signali i 651 kao lažni signali. 78 uzoraka je krivo klasificirano kao lažni signali dok je 91 uzorak



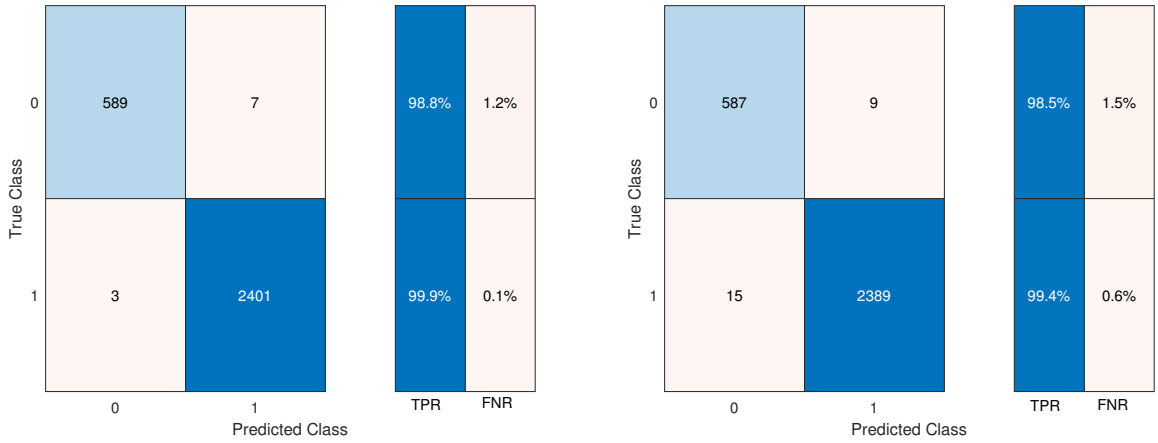
(a) KNN (b) DT

(c) SVM (d) NN

Slika 4.6: Konfuzijska matrica za pojedinu metodu strojnog učenja u našem skupu podataka [103].

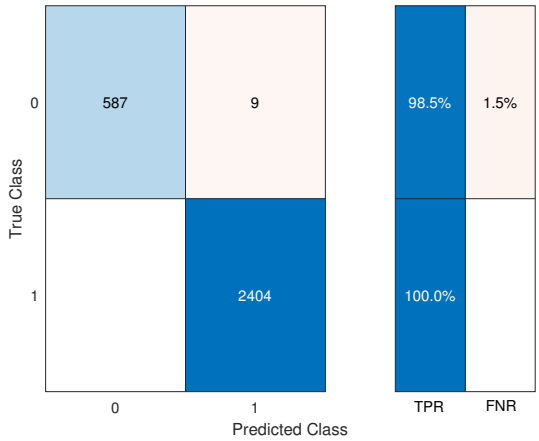
klasificirano kao autentični signali iako su zapravo lažni signali pa TPR iznosi 91.8%, TNR je 87.7%, FNR je 8.2% i FPR je 12.3%. Parametar C/N_0 ima najveći utjecaj na točnost klasifikacije u usporedbi s parametrima faza nosioca i Dopplerov pomak zbog vrijednosti koji se značajno razlikuju za autentične i lažne signale npr. C/N_0 za lažne signale iznosi 45-60 dB-Hz dok za autentične iznosi do 42 dB-Hz.

Slika 4.7 prikazuje konfuzijsku matricu za odabrane modele strojnog učenja u skupu podataka SatGrid. Iako je točnost za sve modele približno jednaka, SVM je za nijansu bolji od ostalih modela pa su rezultati za SVM detaljno analizirani. Može se vidjeti na slici 4.7c da je 587 uzoraka ispravno klasificirano kao autentični signali. 2404 uzoraka su ispravno klasificirani kao lažni signali dok je 9 uzoraka krivo klasificirano kao lažni signali klase 1 iako pripadaju autentičnim signalima klase 0. Svi lažni signali su ispravno klasificirani te stoga nema signala u kategoriji koja krivo predviđa lažne signale kao autentične (bijelo polje na slici). Prema ovim rezultatima klasifikacije, TPR za SVM is 98.5%, TNR je 100%, FNR je 1.5% i FPR je 0%. Za KNN, koji ima malo lošiju točnost, TPR je 98.8%, TNR je 99.9%,

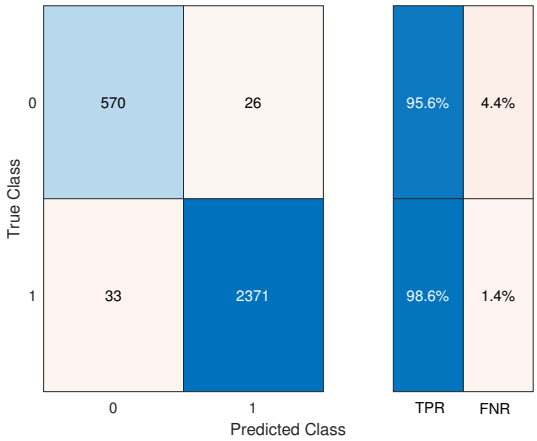


(a) KNN

(b) DT



(c) SVM



(d) NN

Slika 4.7: Konfuzijska matrica za pojedinu metodu strojnog učenja u skupu podataka SatGrid [103].

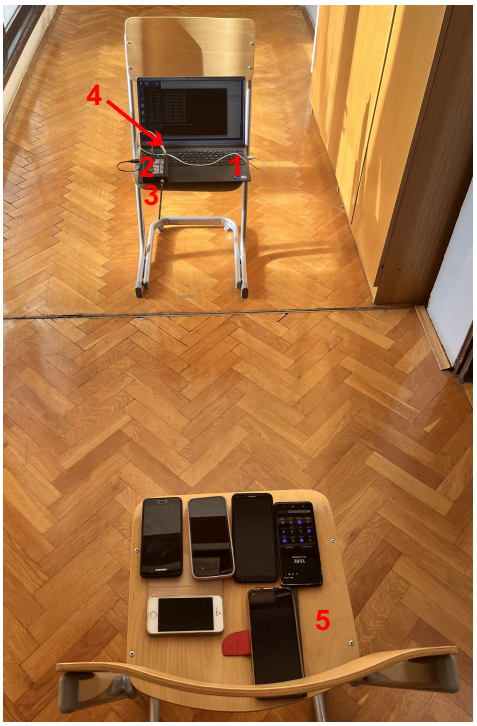
FNR je 1.2% i *FPR* je 0.1%. Sve vrijednosti za parametre u konfuzijskoj matrici odgovaraju vrijednostima izračunatima prema izrazima (6.20), (6.21) (6.22), (6.23), (6.24).

4.1.2. Ispitivanje dometa napadača tijekom napada lažiranjem

U ovom istraživanju izveden je i eksperiment u kojemu je ispitan domet napadača (HackRF One) ovisno o različitim razinama odašiljačke snage te udaljenostima od pametnih telefona. Uz to, prikazano je i vrijeme potrebno za izračun položaja za svaki pametni telefon korišten u eksperimentu [104]. Okolina izvođenja pojednostavljenog napada lažiranjem prikazana je na slici 4.8.

Tablica 4.3 prikazuje pametne telefone korištene u eksperimentu. Korišteni pametni telefoni imaju različite operacijske sustave i verzije. Osim hardverske opreme, simulator otvorenog pristupa GPS-SDR-SIM korišten je za generiranje i odašiljanje lažiranih signala sustava GPS.

Kako bi se kreirali lažirani signali, potrebno je imati podatke o dostupnim i vidljivim



(a) Unutarnji uvjeti

(b) Vanjski uvjeti

Slika 4.8: Izvođenje pojednostavljenog napada lažiranjem u unutarnjim uvjetima i vanjskim uvjetima [104], [105].

satelitima pa se u ovom eksperimentu koristila navigacijska datoteka koja je preuzeta sa stranice NASA's Archive of Space Geodesy Data [108] za određeni dan i sat. Navigacijska datoteka zajedno s lažnim koordinatama, datumom i vremenom (3.215912° N, 73.256822°E, 1.5.2024., 08:01 am) su učitane u simulator GPS-SDR-SIM koji generira bin datoteku lažiranih signala. Kreirana bin datoteka se odašilje s napadača HackRF One prema pametnim telefonima (odašilju se samo signali sustava GPS). Blok dijagram izvedenog eksperimenta prikazan je na slici 4.1.

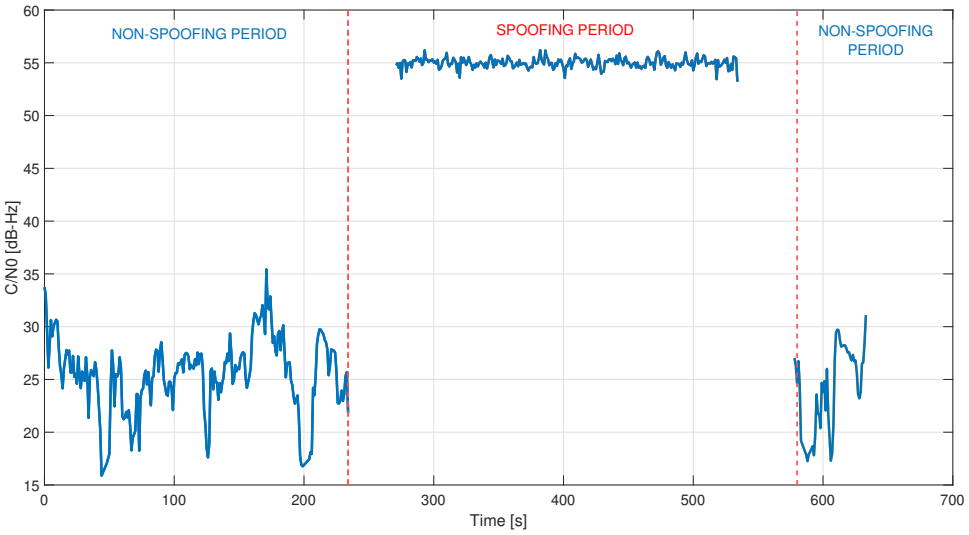
Rezultati eksperimenta u unutarnjim i vanjskim uvjetima

Neobrađena mjerenja prikupljena su kao RINEX 3.03 datoteka na pametnom telefonu Samsung Galaxy 10 korištenjem aplikacije GnssLogger [109] te su obrađena u programu Matlab. Razlog odabira pametnog telefona Samsung Galaxy 10 je napredna verzija njegovog operativnog sustava (Android 12.0) i mogućnost korištenja Google usluga (aplikacija GnssLogger jedna je od Googleovih usluga).

Slika 4.9 prikazuje vrijednosti C/N_0 tijekom napada lažiranjem za zajednički satelit G11. Eksperiment počinje prikupljanjem autentičnih signala sa satelita tijekom 270 sekundi. Nakon toga, napadač se aktivira i odašilje lažirane signale od 271 sekunde do 534 sekunde. Zadnjih 98 sekundi, prikupljaju se opet autentični signali. Može se primijetiti da postoje dva prijelaza - skok između prvog perioda bez lažiranja i perioda lažiranja te pad između peri-

Tablica 4.3: Pametni telefoni i njihovi operacijski sustavi korišteni u eksperimentu.

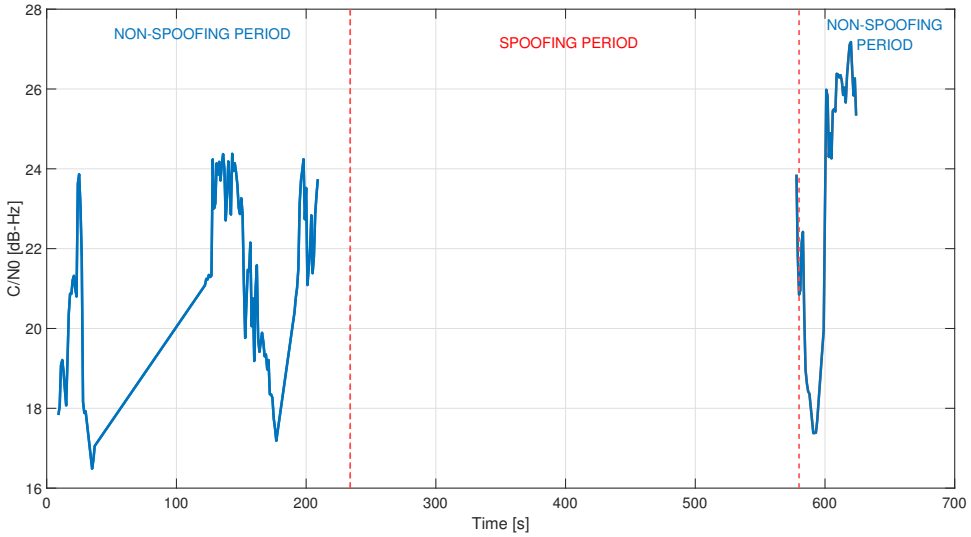
ID	*Pametni telefon*	*Operacijski sustav*
SP1	Samsung Galaxy A5	Android 7.0
SP2	Huawei Mate 10 lite	Android 8.0.0, EMUI 8.0.0
SP3	Huawei P40 lite	EMUI 12.0.0
SP4	iPhone 5S	iOS 12.5.5
SP5	iPhone 14	iOS 16.4.1
SP6	Samsung Galaxy S10	Android 12.0



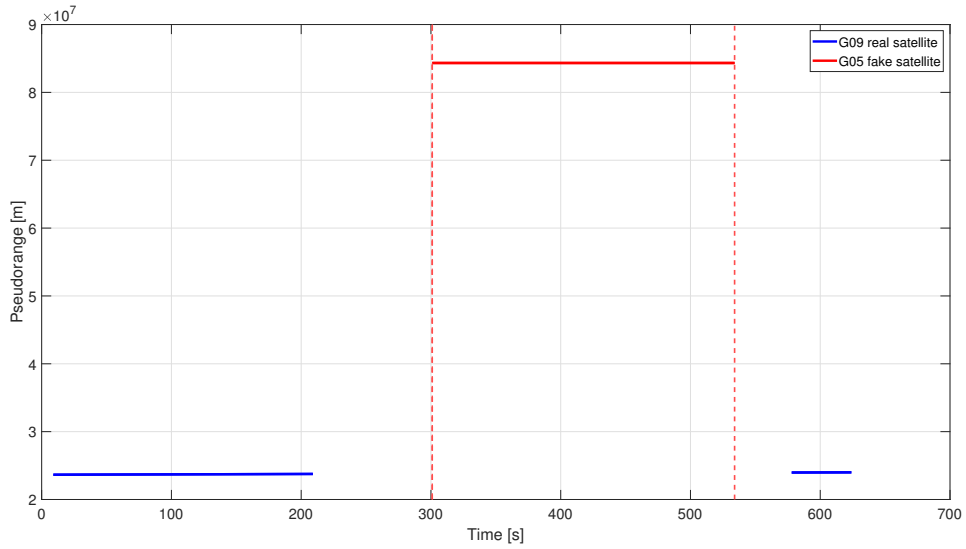
Slika 4.9: Vrijednosti C/N_0 u scenariju lažiranja za zajednički satelit G11 [104].

oda lažiranja i drugog perioda bez lažiranja. Ovi prijelazi se događaju onda kada prijamnik izgubi vezu s autentičnim satelitima i počinje pratiti lažirane signale i obrnuto. Normalna vrijednost parametra C/N_0 tijekom perioda bez lažiranja je 30-35 dB-Hz. Nakon što je napadač aktiviran, vrijednost C/N_0 poraste na 55 dB-Hz i prijamnik počinje pratiti lažirane signale. Sve ove vrijednosti C/N_0 promatrane su i preko aplikacije GPS Test.

Vrijednosti C/N_0 tijekom napada lažiranjem za autentični satelit G09 prikazane su na slici 4.10. Može se vidjeti da su vrijednosti tijekom perioda bez napada lažiranjem do 28 dB-Hz te da tijekom napada lažiranjem ovaj satelit nije vidljiv. S druge strane, u slučaju lažnog satelita G05, vrijednosti C/N_0 su od 50-60 dB-Hz samo tijekom napada lažiranjem zato što se satelit G05 pojavljuje samo u ovom periodu. Sve navedeno dovodi do zaključka da je C/N_0 najznačajniji parametar po kojem se mogu razlikovati autentični i lažni signali. Vrijednosti parametra pseudoudaljenost za autentični satelit G09 i lažni satelit G05 prikazani su na slici 4.11. Vrijednosti pseudoudaljenosti se mijenjaju u trenutku kada napadač krene odašiljati lažne signale u $t = 301$ s. Tijekom perioda lažiranja, signali satelita G09 se više ne prate jer se prijamnik zaključava na satelit G05 koji ima veću snagu signala.



Slika 4.10: Vrijednosti C/N_0 u scenariju lažiranja za autentični satelit G09 [104].



Slika 4.11: Pseudoudaljenosti za satelite G09 i G05 [104].

Domet napadača i vrijeme izračuna položaja prijamnika ("position fix time") za svaki pametni telefon ovisno o različitim razinama snage i udaljenostima od pametnih telefona su analizirani. Osim položaja pametnog telefona, u eksperimentu je uspješno lažirano i vrijeme i datum. Rezultati su promatrani kroz aplikaciju GPS Test i prikazani su na slici 4.12. Sa slike je vidljivo da je pametni telefon uspješno napadnut i da mu je lažirana lokacija koja pokazuje lokaciju Maldive s geografskom širinom 3.215912° N i geografskom dužinom 73.256822° E umjesto stvarne lokacije u Splitu s geografskom širinom 43.508133° dužinom 16.440193°. Snimka zaslona aplikacije prikazuje lažni datum i vrijeme na uspješno napadanutom pametnom telefonu. Datum na pametnom telefonu je 1.5. umjesto 5.5. kada je eksperiment izveden. Stvarno vrijeme na pametnom telefonu 14:40 dok aplikacija pokazuje

08:01.

Slika 4.12: Rezultati napada lažiranjem u aplikaciji GPS Test.

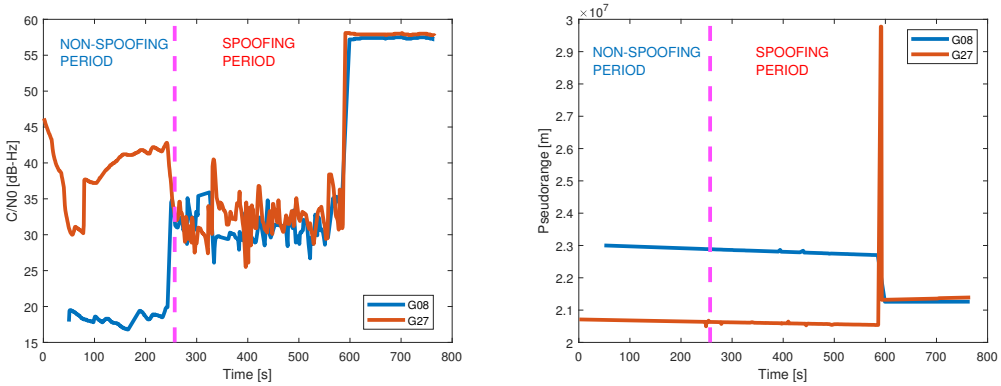
Parametri koji su promatrani tijekom eksperimenta u unutarnjim uvjetima za različite pametne telefone prikazani su u tablici 4.4.

Za pojačanje odašiljača (napadač HackRF One) od 20 dB i udaljenost od 2 m između napadača i pametnog telefona, uspješan napad lažiranjem izveden je samo za dva pametna telefona: Huawei P40 lite i Samsung Galaxy S10. Drugim riječima, postignut je izračun položaja za navedene pametne telefone. Kada se pojačanje odašiljača poveća na 40 dB i udaljenost na 6 m, na pametnim telefonima iPhone 14 i Samsung Galaxy S10, napad lažiranjem nije uspješno izveden, tj. pametni telefoni nisu uspjeli izračunati svoj položaj. Domet napadača u našem eksperimentu je 25 m za slučaj kada je napad lažiranjem uspješno izveden na dva pametna telefona. Kada je udaljenost između napadača i pametnih telefona veća od 25 m, napad lažiranjem nije uspješno izveden na nijednom pametnom telefonu. Iz dobivenih rezultata, može se zaključiti da ne postoji neko pravilo u brzini postizanja izračuna položaja tj. preuzimanja navigacijskih sustava pametnih telefona. Primjerice, vrijeme izračuna položaja za SP3 uz pojačanje od 40 dB i udaljenost od 6 m, veće je nego u slučaju istog pojačanja i udaljenosti od 25 m (Tablica 4.4). Pretpostavka je da su ovakvi rezultati povezani s predmemorijom pametnih telefona koju bi vjerojatno trebalo obrisati (isključiti pametni telefon) prije ponavljanja eksperimenta na nekoj drugoj udaljenosti. U slučaju operacijskog sustava iOS, najnovije verzije sustava su toliko zaštićene da je veoma teško preuzeti njihov

Tablica 4.4: Parametri promatrani tijekom eksperimentu u untarnjim uvjetima za različite pametne telefone [104].

ID	*Pojačanje napadača [dB]*	*Udaljenost [m]*	*Vrijeme izračuna položaja [s]*
SP1	20	2	no fix
SP2	20	2	no fix
SP3	20	2	185
SP4	20	2	no fix
SP5	20	2	no fix
SP6	20	2	75
SP1	40	6	55
SP2	40	6	53
SP3	40	6	51
SP4	40	6	60
SP5	40	6	no fix
SP6	40	6	no fix
SP1	40	25	30
SP2	40	25	no fix
SP3	40	25	37
SP4	40	25	no fix
SP5	40	25	no fix
SP6	40	25	no fix

navigacijski sustav. Također, za novije verzije sustava Android (12 i više) teže je preuzeti navigacijski sustav izvođenjem pojednostavljenog napada lažiranjem. U ovakvim slučajevima, neka druga vrsta napada lažiranjem bi mogla biti uspješnija.



(a) C/N_0 (b) Pseudoudaljenost

Slika 4.13: Vrijednosti parametara C/N_0 i pseudoudaljenosti tijekom napada lažiranjem u vanjskim uvjetima [105].

Analiza rezultata za izvođenje napada lažiranjem u vanjskim uvjetima prikazana je na slici 4.13 za dva zajednička satelita sustava GPS G08 i G27 za parametre C/N_0 i pseudoudaljenost. Vrijednosti pojačanja antene se postepeno povećavaju od 15 do 40 dB. Slika 4.13a

prikazuje vrijednosti C/N_0. Sa slike se vidi da napad lažiranjem počinje u trenutku $t = 257$ s uz pojačanje od 15 dB i tada pametni telefon ne postiže izračun položaja. Stoga, pojačanje antene se postepeno povećava na 25, 35 i konačno 40 dB kada pametni telefon uspijeva izračunati položaj odnosno napadač uspješno preuzima navigacijski sustav pametnog telefona. U trenutku $t = 600$ s vrijednost C/N_0 poraste na 57 dB-Hz i do kraja napada ima konstantnu vrijednost.
Što se tiče pseudoudaljenosti (slika 4.13b), ona je konstantna s nekim sitnim oscilacijama za oba satelita sve do trenutka kada prijamnik postiže izračun položaja u trenutku $t = 600$ s. U tom trenutku pseudoudaljenost ima nagli skok nakon čega opet ima konstantnu vrijednost. Na temelju dobivenih rezultata može se zaključiti da je vrijednost C/N_0 tijekom napada lažiranjem u unutarnjim uvjetima konstantna dok u vanjskim uvjetima ona oscilira s dosta manjim vrijednostima (do 40 dB-Hz) i tek postiže konstantnu vrijednost u trenutku kada prijamnik izračuna svoj položaj. Razlog tomu je što su u vanjskim uvjetima vrlo dobro vidljivi svi autentični sateliti pa prijamnik uzima i te signale, a ne samo lažne signale kao u kontroliranim uvjetima. Nadalje, utjecaj na sami prijam signala imaju i neke druge interferencije kao što je višestazno prostiranje [105].

4.2. Metode za detekciju ometanja i lažnih signala

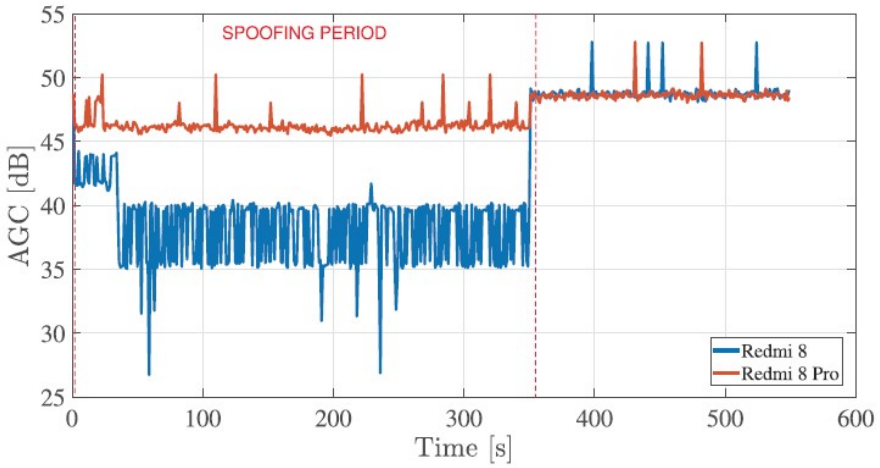
Metode detekcije lažnih GNSS signala imaju za primarni cilj otkrivanje napada lažiranjem kako bi upozorile prijamnik da podaci o njegovoj lokaciji i vremenu nisu točni. Potrebno je razumjeti svojstva različitih napada kako bi se razvila dobra obrana od samog napada [14]. Postoje različite metode detekcije lažnih signala: klasične metode temeljene na promatranju C/N_0, pseudoudaljenosti i različitih parametara, klasične metode koje se temelje na promatranju korelacijskih funkcija signala, metode temeljene na simulatorskom hardveru (npr. simulator poput Spirenta) koje nisu ekonomične [31], metode koje počivaju na korištenju niza antena, metode koje koriste NMEA poruke [32] te metode strojnog učenja.

Korisnički uređaj koji prima lažne signale i vjeruje da je autentičan može potaknuti opasno ponašanje zbog pogrešnog položaja ili ispravki vremena. Primjer je spomenut u [12], gdje je lažiranje GPS signala korišteno za krivo usmjeravanje drona u neplanirano ronjenje i za skretanje jahte s kursa. Stoga je obrana od prijevare usmjerena na otkrivanje napada kako bi se napadnuti prijamnik upozorio da su njegov izračunati položaj i pomak sata nepouzdani. Dan je prikaz različitih metoda napada lažiranjem.

4.2.1. Strojno učenje u kombinaciji s promatranjem klasičnih parametara i korištenjem softverski definiranog radija

U [68], autori prikazuju eksperimentalne rezultate osjetljivosti pametnih telefona na pojednostavljeni napad lažiranjem. Učinci osjetljivosti pametnih telefona se očituju kroz neo-

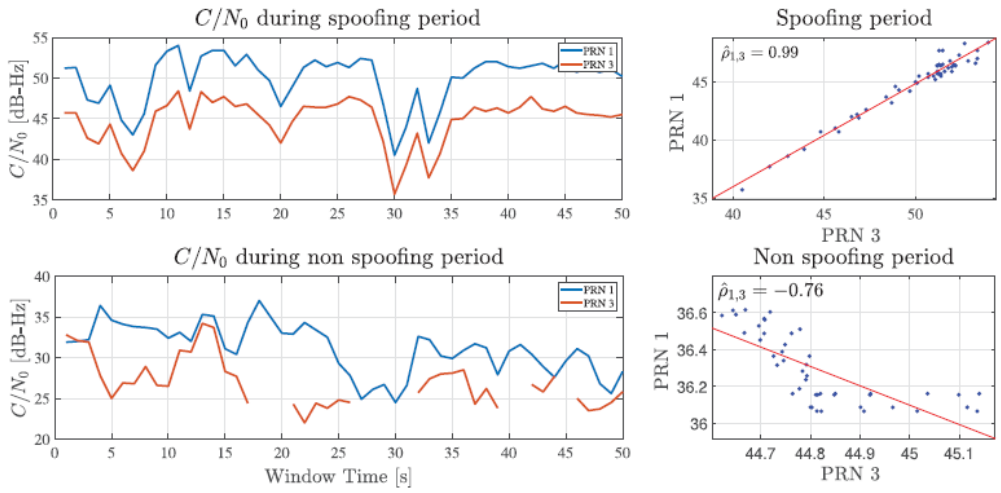
brađena mjerenja parametara npr. C/N_0, automatsko upravljanje pojačanjem AGC (engl. *Automatic Gain Control*), pseudoudaljenosti i procjene pozicije. Autori reproduciraju dva scenarija pojednostavljenog napada lažiranjem.



Slika 4.14: Usporedba parametra AGC između dva Android uređaja prilikom napada lažiranjem [68].

Slika 4.14 prikazuje AGC vrijednosti za dva Android uređaja (Redmi 8 i Redmi 8 Pro) prilikom napada lažiranjem. Napad lažiranjem traje od $0-350$ s. U trenutku $t = 350$ s kada napad lažiranjem završava, AGC vrijednost se povećava na svoju početnu razinu kao što se može vidjeti sa slike. Skok u AGC vrijednosti za Redmi8 uređaj može biti posljedica gubitka kačenja na autentične signale i ponovnog praćenja te kačenja na lažne signale. Velika snaga i postojanost lažnog signala mogu biti čimbenik u određivanju praznina u mjerenjima. Primjerice, ukoliko je lažni signal dovoljno snažan i postojan, GNSS prijamnik može izgubiti povezanost odnosno "kačenje" na signale na duži period što rezultira prazninom u GNSS mjerenjima. S druge strane, ako je lažni signal slab i manje postojan, prijamnik može zadržati kačenje na autentične signale i proizvesti kontinuirani izlaz, unatoč prisutnosti lažnih signala. Različiti prijamnici imaju različite osjetljivosti i druge značajke koje utječu na otpornost na napade lažiranjem.

Praćenje snage signala je najjednostavniji način detekcije napada lažiranjem jer je snaga lažnog signala puno veća u odnosu na autentični signal. Osim po većoj snazi signala, lažni signal se može detektirati po konstantnom Dopplerovom pomaku jer se napadač nalazi na istoj lokaciji i odašilje signale iz istog izvora. Kod satelitskih signala koji se odašilju sa satelita, Dopplerov pomak je dinamičan jer se kontinuirano mijenja s vremenom i različit je za svaki satelit zato što ovisi o brzini i smjeru gibanja satelita, položaju i brzini prijamnika te geometriji promatranja u datom trenutku. Dodatni parametri po kojima se mogu prepoznati lažni signali su konstantna pseudoudaljenost i konstantni kut elevacije jer napadač odašilje s fiksne lokacije. U slučaju dinamičkog napada, trebalo bi postojati nekoliko lokacija s kojih napadač odašilje i tada bi bilo teže detektirati lažne signale.



Slika 4.15: Usporedba C/N_0 za različite satelite tijekom i bez napada lažiranjem [68].

Klasična detekcija lažnog signala temeljena na C/N_0 je predložena u [75], gdje je izmjereni C/N_0 primljenih GNSS signala uspoređen s poznatom ili očekivanom vrijednosti. U [16] autori uz C/N_0 za detekciju napada lažiranjem prate i pseudoudaljenosti. S druge strane, autori u [29] skupa s pseudoudaljenostima i snagom signala razmatraju i distorziju korelacije. U [42] and [18], lažni GNSS signali su detektirani na temelju vrhova signala korelacije i faznih razlika između lažnih i autentičnih signala. Eksperimentalni rezultati osjetljivosti pametnih telefona na pojednostavljeni napad lažiranjem su prikazani u [68]. Slika 4.15 prikazuje usporedbu C/N_0 vrijednosti za GPS satelite PRN 1 i PRN 3 tijekom (gore) i bez (dolje) napada lažiranjem za Xiaomi Redmi 8. Tijekom napada lažiranjem, C/N_0 za oba satelita je u rangu 35-55 dB-Hz dok je u uvjetima bez napada vidljiva osjetna razlika u kojima C/N_0 ima vrijednosti od 20-40 dB-Hz s laganim trendom opadanja i diskontinuiteta pri nižim vrijednostima. Korelacija između vrijednosti C/N_0 za oba slučaja je potvrđena linearnom regresijom i Pearsonovim koeficijentom korelacije. U slučaju napada lažiranjem, postoji veća korelacija između C/N_0 vrijednosti za dva različita satelita PRN1 i PRN3 te Pearsonov koeficijent iznosi 0.99 dok je bez napada lažiranjem niska korelacija između vrijednosti C/N_0 za navedene satelite i Pearsonov koeficijent iznosi -0.76 zbog diskontinuiteta podataka i različitih trendova. Razlog tomu je što u normalnim uvjetima bez napada lažiranjem, signali različitih satelita dolaze iz različitih smjerova i njihove vrijednosti C/N_0 se ponašaju različito (vrijednost kod jednog satelita raste dok kod drugog opada) i nekorelirano te je Pearsonov koeficijent nizak i negativan. S druge strane, u uvjetima napada lažiranjem, svi lažni signali dolaze iz istog smjera jer se koristi jedan izvor odašiljanja i vrijednosti C/N_0 se mijenjaju slično u vremenu te je korelacija vrlo visoka. Učinci osjetljivosti pametnih telefona se ogledaju kroz njihova neobrađena mjerenja npr. C/N_0, pseudoudaljenosti i procjene položaja. Utjecaj napada lažiranjem na pametne telefone je analiziran u [74]. Autori predlažu tehnike za povećanje sigurnosti kao što je upotreba jeftinih senzora ubrzanja.

Pojednostavljeni napad lažiranjem je izveden u [69] pomoću softverski definiranog radija. GPS signali su snimljeni i ponovno odašiljani na pametne telefone. GPS Test aplikacija je korištena za praćenje rezultata napada tj. parametara: dostupni sateliti i njihov C/N_0. U slučajevima u kojima C/N_0 diskriminacija ima ograničenu učinkovitost, prijamnik može mjeriti apsolutnu snagu vrhova signala korelacije (maksimalna vrijednost korelacije između dva signala odnosno maksimalna amplituda), i ova metoda je učinkovita za detekciju i diskriminaciju izvora napada. Autori u [78] pokazuju da praćenje apsolutne snage signala značajno smanjuje područje osjetljivosti prijamnika u usporedbi s praćenjem C/N_0. U [79], autori predlažu metodu za detekciju napada lažiranjem i ometanja signala temeljenu na automatskoj kontroli pojačanja i C/N_0 opservacijama. Napad lažiranjem će vjerojatno biti detektiran kada se AGC vrijednost smanji, i C/N_0 je relativno konstantan ili čak povećan. Međutim, AGC nije dovoljan za detekciju prisustva lažnog signala, nego samo za podizanje upozorenja. Stoga bi se AGC trebao koristiti u kombinaciji s C/N_0.

U [67], autori uspoređuju performanse nekoliko nadziranih modela s nenadziranima u smislu točnosti, vjerojatnosti otkrivanja, vjerojatnosti pogrešnog otkrivanja, vjerojatnosti lažnog alarma, vremena obrade, vremena treniranja, vremena predviđanja i veličine memorije. Rezultati pokazuju da klasifikacijski i regresijski modeli stabla odlučivanja nadmašuju ostale nadzirane i nenadzirane modele u otkrivanju i klasificiranju GPS napada lažiranjem.

U [25] i [26] autori uspoređuju izvedbu nekoliko ML (engl. *Machine Learning*) algoritama u otkrivanju napada lažiranjem GPS signala. Autori u [25] provode k-fold analizu kako bi odabrali najbolji algoritam strojnog učenja između nekoliko algoritama. Na temelju njihovih rezultata, metoda potpornih vektora s polinomskom jezgrom nadmašuje ostale metode. S druge strane, rezultati i analiza ML algoritama u [26] pokazuje da algoritmi temeljeni na stablima odlučivanja daju bolje rezultate u odnosu na SVM (linearni i radijalni), K najbližih susjeda i ostale analizirane algoritme.

U [19], autori predlažu detekciju lažnih GNSS signala korištenjem SVM metode strojnog učenja uz kombinaciju stvarnih i simuliranih skupova podataka za provjeru i validaciju algoritama strojnog učenja. Rezultati pokazuju da je SVM metoda obećavajući pristup za detekciju lažnih signala. Međutim, ovo istraživanje ne analizira razloge za odabir određenih parametara te kombinaciju i sklonost prema određenim značajkama. Većina postojećih algoritama za otkrivanje napada lažiranjem koristi postojeći skup podataka TEXBAT koji je objavilo Sveučilište Texas [20], s relativno fiksnim scenarijima. Albright i ostali iz Nacionalnog laboratorija Oak Ridge, SAD, objavili su još jedan gotovi skup podataka OAKBAT [22] koji sadrži lažne signale GPS i Galileo, pružajući više testnih scenarija za istraživanje otkrivanja napada lažiranjem.

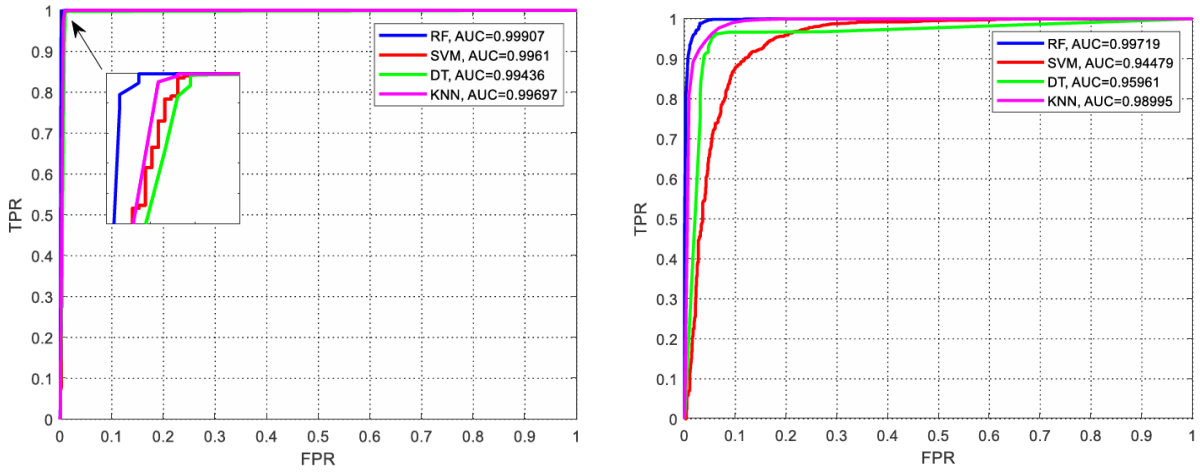
Autori u [40] predlažu GNSS više-parametarsku metodu zajedničke detekcije koja se također temelji na SVM metodi obradom i usporedbom skupova podataka TEXBAT i OAKBAT. Dobiveni rezultati pokazuju značajno poboljšanje u performansama otkrivanja lažnih signala u usporedbi s tradicionalnim jedno-parametarskim metodama. Nadalje, autori u radu

Algorithm	Accuracy	Precision	Recall	F1	AUC
RF	**99.67**	**99.57**	**99.82**	**99.70**	**99.91**
SVM	99.34	98.91	99.87	99.39	99.61
DT	99.31	99.08	99.63	99.36	99.43
KNN	99.50	99.27	99.81	99.54	99.69

Slika 4.16: Točnosti klasifikacije nekoliko modela strojnog učenja za skup podataka TEXBAT ds2 [65].

[65] koriste optimizirani algoritam slučajne šume za detekciju lažnih signala u skupu podataka TEXBAT. Njihov pristup temelji se na korištenju nekoliko klasifikacijskih parametara (C/N_0, pseudoudaljenost, Dopplerov pomak, faza nosioca, razlika vremena na prijamniku). Njihov optimizirani algoritam slučajne šume ostvaruje najbolje performanse i točnost klasifikacije od 99.7%. Uz ovaj algoritam, za detekciju su korišteni i SVM, KNN i DT koji su isto ostvarili točnost preko 99% kao što je prikazano na slici 4.16.

U istom radu autori koriste skupove podataka TEXBAT ds2 i ds3 koji se razlikuju po prednosti razine snage lažnih u odnosu na autentične signale (ds2 - prednost od 10 dB i ds3 - prednost od 1.3 dB). Krivulje operativnih karakteristika za skup podataka ds2 prikazane su na slici 4.17a dok su za skup ds3 prikazane na slici 4.17b. Iz slika se može vidjeti da za skup podataka ds2, u kojem je razina snage za lažne signale za 10 dB veća u odnosu na autentične signale, svi modeli strojnog učenja imaju podjednako dobre performanse. S druge strane za skup podataka ds3, kod kojeg je razina snage za lažne signale za 1.3 dB veća u odnosu na autentične signale, najbolje performanse ima algoritam slučajne šume.

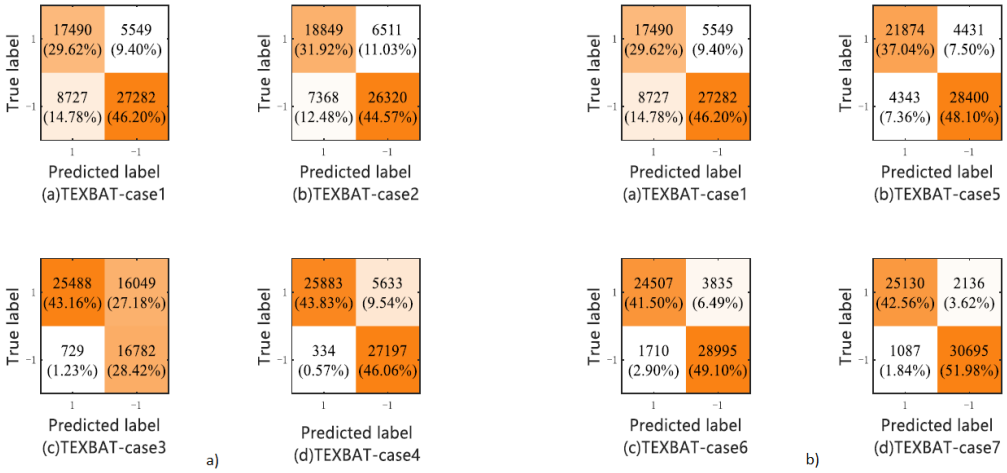


(a) ds2 (b) ds3

Slika 4.17: Krivulje operativnih karakteristika za nekoliko modela strojnog učenja za skupove podataka TEXBAT ds2 i ds3 [65].

S druge strane, autori u [45] koriste tri sintetički generirana (simulirana) skupa podataka lažnih signala sa Spirent simulatorom za treniranje i verifikaciju i dva skupa podataka za provjeru valjanosti modela stvorena korištenjem softverski definiranih radija LimeSDR i HackRF. Autori koriste C-SVM metodu nadziranog strojnog učenja za otkrivanje lažnih signala. U [46], autori nadopunjuju eksperimente i rezultate dobivene u [45]. Uz laboratorijski generirane skupove podataka lažnih signala koji su u [45] korišteni za treniranje modela, dodani su skupovi podataka lažnih signala u stvarnom vremenu u fazi treniranja C-SVM metode.

Slika 4.18 prikazuje konfuzijsku matricu za detekciju napada lažiranjem uz korištenje različitih parametara - a) i kombinacije različitih parametara b). Sa slika je vidljivo da se točnost SVM metode poboljšala u slučaju sedam, u kojem je korišteno svih devet parametara, sa 75.82% na 95.54%.



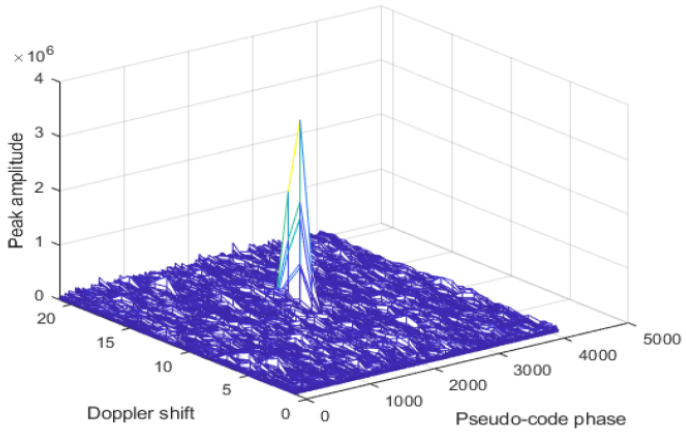
Slika 4.18: Konfuzijska matrica za detekciju napada lažiranjem u skupu podataka TEXBAT [40].

U preglednom radu [27], dane su preporuke za istraživače te je zaključeno da su ML metode obećavajući pristup za primjenu u sustavu GNSS.

Budući da su bespilotne zračne letjelice (UAS) vrlo osjetljive na ovu vrstu napada, autori u [28] provode usporedbu nekoliko modela nadziranog strojnog učenja koji se temelje na stablu kako bi otkrili lažne napade i prikupili stvarne GPS signale pomoću SDR-a. U [44], autori vrednuju pet modela strojnog učenja temeljenih na instancama za otkrivanje lažnih GPS signala. Također, autori koriste SDR jedinicu za prikupljanje i izdvajanje značajki satelitskih signala te simuliraju tri vrste napada lažiranjem (pojednostavljeni napad, napad srednje razine složenosti i sofisticirani napad). Rezultati pokazuju da Nu-SVM ima najbolje performanse.

Autori u [29] predlažu navigaciju u okruženju u kojem se događa napad lažiranjem GNSS signala uzimajući u obzir primljenu snagu, funkciju izobličenja korelacije i pseudoudaljenosti. U skupu podataka se koriste i stvarna i lažna mjerenja. Strojno učenje prikazuje

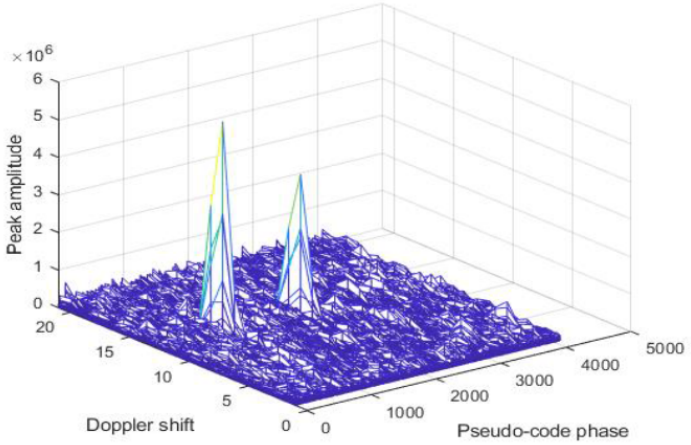
autentična mjerenja iz dostupnog skupa pomoću parametara kao što su primljena snaga i izobličenje korelacijske funkcije. U radu je korišteno nekoliko metoda strojnog učenja za klasifikaciju i detekciju lažnih signala. Kao najbolje metode, pokazale su se neuralne mreže i linearni SVM s točnošću od 98.20%.



Slika 4.19: Stvarni satelitski signal u fazi snimanja [17].

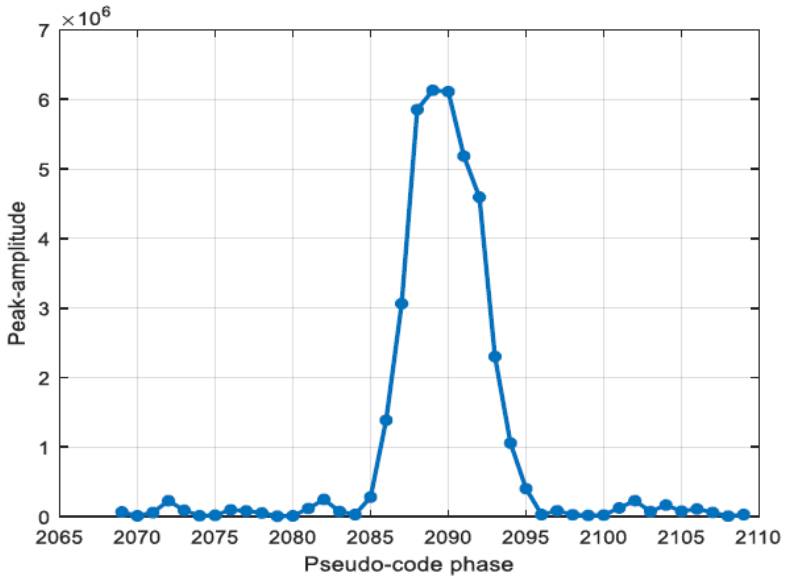
4.2.2. Tradicionalna metoda promatranja korelacijske funkcije - SQM

Metoda detekcije lažnih GNSS signala na temelju vrhova signala korelacije SQM i fazne razlike između lažnog i autentičnog signala je korištena u [17], [18] i [42]. Rad [17] je fokusiran na detekciju lažnih signala s malim kašnjenjem korištenjem KNN metode strojnog učenja. Detekcija broja vrhova signala je ključan korak za detekciju lažnog signala. Detekcija se temelji na otkrivanju lažnog signala na način da se procijeni broj vrhova koji prelaze unaprijed postavljeni prag kada prijamnik uhvati signal. Ako u primljenom signalu postoji samo pravi GNSS signal, vrijednost samo jednog vrha signala korelacije će premašiti unaprijed postavljeni prag što je prikazano na slici 4.19.



Slika 4.20: Lažni signal postoji u fazi snimanja uz kašnjenje od 100 čipova [17].

Kada postoje lažni signali, onda postoje dva ili više vrha signala korelacije koji su veći od postavljenog praga (4.20) i ovakav način detekcije lažnih signala vrijedi kada je fazna razlika između lažnog signala i stvarnog signala velika tj. veća od 2 čipa. Kada je fazna razlika između stvarnog i lažnog signala manja npr. 1 čip što je slučaj na slici 4.21, broj vrhova je i dalje 1 pa je teško detektirati lažne signale. Eksperimentalni rezultati provedeni u ovom radu su pokazali da predloženi algoritam može detektirati lažne signale s kašnjenjem većim od 0.6 čipova te da ima visoku točnost. Autori u [18] pokazuju da generativna suparnička mreža GAN može doseći više od 98% točnosti kada fazna razlika između lažnog i autentičnog signala prelazi 0.5 čipova i može se primijeniti na situacije u kojima je lažni signal dobro sinkroniziran s autentičnim signalom.



Slika 4.21: Lažni signal postoji u fazi snimanja uz kašnjenje od 1 čip [17].

Autori se u [30] fokusiraju na klasifikaciju GNSS signala i svrstavaju ih u klase: autentični, višestazni, lažni ili ometani. Značajke koje koriste za klasifikaciju signala su prosječna snaga i izobličenje korelacije. Različite metode strojnog učenja su testirane korištenjem testa točnosti i konfuzijske matrice. Lažni i ometani signali lako se razlikuju od autentičnih signala zbog njihove visoke prosječne snage i visokog stupnja izobličenja korelacije. Stoga je u slučaju namjernih ometanja (interferencija) ovakva metoda klasificiranja moćan alat za navigacijske aplikacije koje koriste GNSS prijamnik.

4.2.3. Detekcija pomoću NMEA poruka

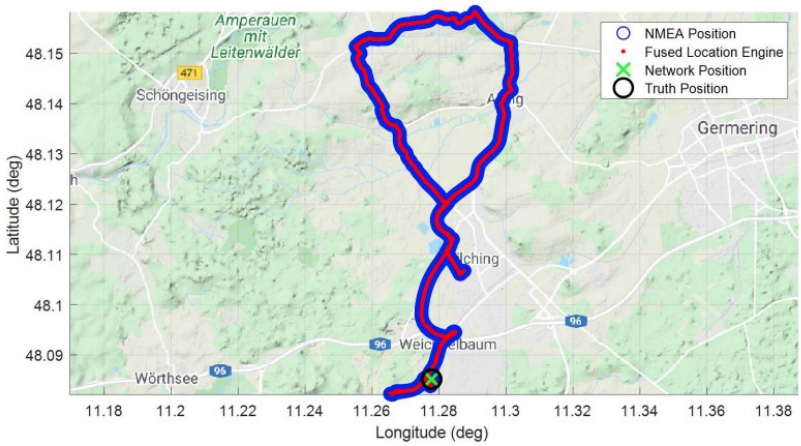
Autori u [32] predlažu pristup temeljen na korištenju NMEA poruka od GNSS prijamnika (pametni telefoni i komercijalni ublox prijamnik) za detekciju i identifikaciju sumnjivih potencijalno lažnih signala. NMEA 0183 poruke sadrže informacije o vidljivim satelitima, položaju prijamnika, brzini i vremenu te za njihovu obradu nije potrebno značajno proce-

siranje. Korištenjem NMEA poruka zaobiđena su velika proračunska opterećenja koja su potrebna za dobivanje i obradu neobrađenih mjerenja. Slika 4.22 prikazuje vrste NMEA poruka i njihove opise.

NMEA Message Type	Description
GSV	GNSS satellites in view PRN, Elevation, Azimuth, C/No
GSA	GNSS DOP and active satellites
GGA	GNSS fix data Time, Position, DOP
RMC	Recommended minimum specific data Time, Position, Velocity
VTG	Track made good and ground speed Velocity, Heading
GRS (not available for smartphones)	Range residuals for active satellites

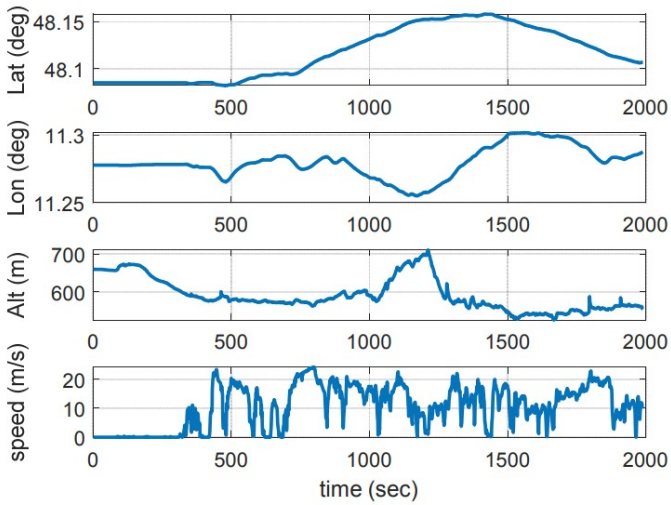
Slika 4.22: Definicija NMEA poruka prikupljenih od strane GNSS prijamnika [32].

Promatrana su tri različita scenarija: u prvom scenariju napadač je emulirao vožnju koja počinje od zgrade i radi petlju oko obližnjeg područja, u drugom scenariju napadač se udaljava od zgrade i vraća na početak i treći sceanrij je isti kao drugi samo što napadač ima dodatno prigušenje. U prvom scenariju, lokacije svih pametnih telefona su uspješno lažirane. Iako su pametni telefoni bili u stacionarnom položaju na stolu unutar zgrade, NMEA poruke su zabilježile da su uređaji u pokretu u okolnom području (slika 4.23).



Slika 4.23: Putanja kretanja uređaja tijekom uspješnog napada lažiranjem [32].

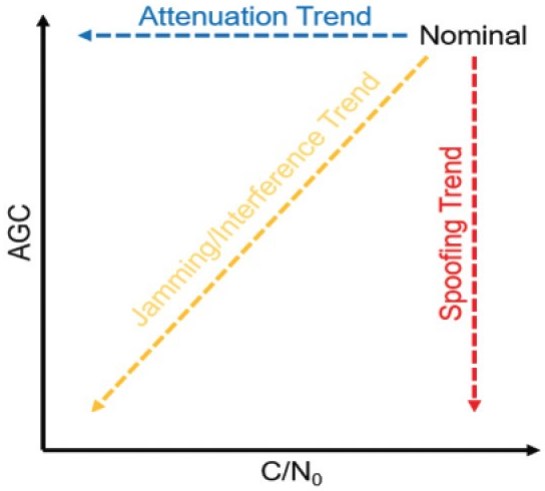
Za drugi scenarij, napad lažiranjem utjecao je na točnost pozicioniranja, ali potpuno očekivana lažna putanja nije uočena dok je za treći scenarij napad lažiranjem bio uspješan i uočena je očekivana lažna putanja. Slika 4.24 prikazuje položaje i brzine zabilježene od strane pametnih telefona (NMEA poruke za položaj, brzinu i vrijeme). Iako su uređaji u stacionarnom stanju, logovi su zabilježili da su u dinamičnom stanju pod napadom lažiranjem.



Slika 4.24: Pozicije i brzine prijamnika pametnih telefona.

4.2.4. Metoda detekcije lažiranja na temelju parametara pametnih telefona

Ometajući signal se može detektirati kao i lažni signal promatranjem parametara C/N_0 i AGC. Autori u radu [38] opisuju kako se preko prethodno navedenih parametara mogu razlikovati lažni i ometajući signal. Ukoliko se i AGC i C/N_0 smanje, vjerojatniji je ometajući signal, a ako se AGC smanji i C/N_0 ostane konstantan, vjerojatniji je lažni signal. Ako je AGC konstantan, onda je malo vjerojatan bilo koji oblik smetnje, a slabi signal se može pripisati prigušenju.



Slika 4.25: Očekivani trend za AGC i C/N_0 [39].

U radu [39], autori predlažu rješenje za detekciju ometanja i napada lažiranjem korištenjem izvornih parametara (između ostalih AGC i C/N_0) lokacije unutar Androida. Ovo rješenje povećava robusnost proračuna pozicije i vremena u Android sustavima i implementirano je u GNSSAlarm Android aplikaciji koja sadrži indikatore za AGC i C/N_0 . Ako AGC

padne ispod postavljenog praga i C/N_0 padne na jednak iznos ili više, smetnje su vjerojatne i odgovarajući indikatori postati žuti što je vidljivo na slici. Ako se dogodi isti scenarij, a C/N_0 ne padne proporcionalno, indikatori postaju crveni i upozoravaju na napad lažiranjem što je prikazano na slici 4.25.

5. Integrirani pristup za detekciju lažiranih signala korištenjem metoda radio frekvencijskog otiska i strojnog učenja

U ovom dijelu prikazan je predloženi integrirani pristup za detekciju napada lažiranjem korištenjem metoda radio frekvencijskog otiska (spektrogram i diskretna valićna transformacija) i modela strojnog učenja. Cilj istraživanja u ovoj doktorskoj disertaciji je primjena neistraženih metoda radio frekvencijskog otiska za detekciju napada lažiranjem i klasifikaciju tipa signala (autentični ili lažni) u području GNSS.

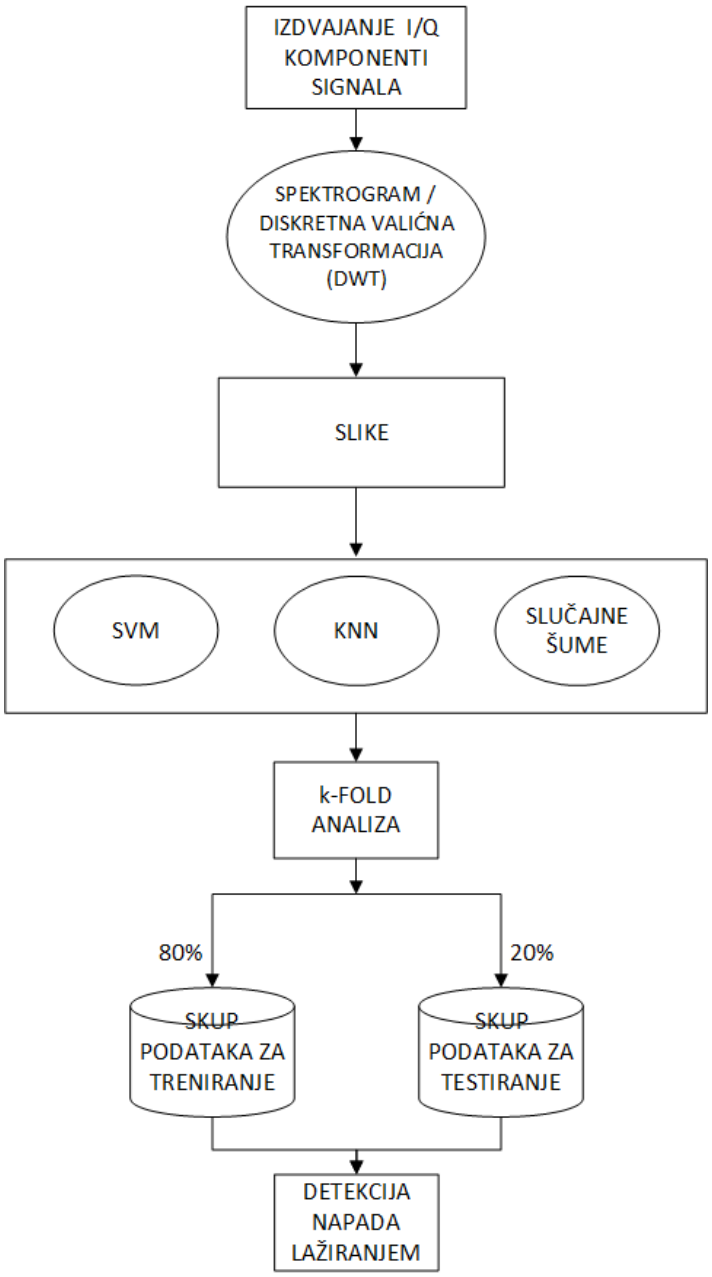
5.1. Predloženi pristup i metode radio frekvencijskog otiska

Predloženi integrirani pristup za detekciju napada lažiranjem i klasifikaciju tipa signala sastoji se od tri koraka:

1. Primjena metoda radio frekvencijskog otiska na odabrane skupove podataka.
2. Primjena odabranih modela strojnog učenja na generirane skupove podataka.
3. Evaluacija modela korištenjem standardnih parametara performansi modela.

Verifikacija predloženog pristupa napravljena je na postojećim skupovima podataka OAKBAT u statičkim uvjetima za dvije različite konstelacije signala GPS i Galileo. OAKBAT skupovi podataka su simulirani skupovi u kojima su podaci zapisani u binarnom obliku. Skupovi podataka se razlikuju u razinama snage između autentičnih i lažnih signala što je prikazano u tablici 6.1. Blok dijagram predloženog integriranog pristupa za detekciju napada lažiranjem uz primjenu slika kao klasifikacijskog podatka prikazan je na slici 5.1. Signali su izdvojeni kao I/Q komponente nakon čega su na njih primijenjene metode radio frekvencijskog otiska, spektrogram i diskretna valićna transformacija. Rezultat prvog koraka su generirani skupovi slika spektrograma i diskretne valićne transformacije. Na generirane skupove slika, primijenjen je algoritam klasifikacije A1 za modele strojnog učenja: metoda potpornih vektora, K-najbližih susjeda i slučajne šume. Unakrsna provjera valjanosti k podskupova primijenjena je na sve navedene modele strojnog učenja. Uz odabrani $k = 5$, 80%

podataka je korišteno za treniranja, a 20% za testiranje. Svi modeli su evaluirani korištenjem standardnih parametara performansi modela.



Slika 5.1: Blok dijagram predloženog integriranog pristupa za detekciju lažiranih signala uz primjenu slika kao klasifikacijskog podatka.

5.1.1. Metoda temeljena na primjeni diskretne valićne transformacije (DWT)

Diskretna valićna transformacija preslikava diskretni signal u niz valića na različitim skalama. DWT dijeli signal na vremensku i frekvencijsku komponentu. U usporedbi s Fourierovom transformacijom, koja daje samo podatke o frekvenciji, diskretna valićna transfor-

macija dodatno daje vremensku lokalizaciju što je čini prikladnom za analizu nestacionarnih ili prijelaznih signala.

U kontekstu detekcije napada lažiranjem u sustavu GNSS, diskretna valićna transformacija može se koristiti za detekciju anomalija koje ukazuju na prevaru. Primjene ove transformacije za detekciju napada lažiranjem su sljedeće:

- Dekompozicija signala na različite frekvencijske komponente koje omogućuju izvlačenje značajki.
- Klasifikacija tipa signala (autentični ili lažni) korištenjem metoda strojnog učenja.
- Računalna učinkovitost čini ovu transformaciju prikladnom za obradu u stvarnom vremenu kada je potrebna neposredna detekcija, npr. u GNSS prijamnicima.

Na temelju generiranih skupova spektrograma i diskretne valićne transformacije izvršena je detekcija napada lažiranjem te klasifikacija tipa signala korištenjem metoda strojnog učenja u programu MATLAB.

Matematički model

Valić je kratkotrajna oscilacija lokalizirana u vremenu koja zadovoljava sljedeće uvjete:

1. Uvjet prihvatljivosti, što znači da valna funkcija ima srednju vrijednost 0:

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0. \tag{5.1}$$

2. Uvjet normalizacije što znači da valna funkcija ima konačnu energiju:

$$\int_{-\infty}^{\infty} |\psi(t)|^2\, dt = 1. \tag{5.2}$$

Uvjet (1) označava da valićna funkcija ima jednaka pozitivna i negativna područja tako da se mogu uhvatiti i niske i visoke frekvencije.

Diskretna valićna transformacija definirana je kao:

$$W_{j,k} = \sum_{n=0}^{N-1} x[n] \cdot \psi_{j,k}[n], \tag{5.3}$$

gdje je $\psi_{j,k}[n]$ valićna osnovna funkcija na određenoj skali j i pomak k za svaki diskretni uzorak n. Valićna osnovna funkcija nastaje iz matičnog (majčinog) valića $\psi(t)$:

$$\psi_{j,k}[n] = \frac{1}{\sqrt{2^j}} \psi\left(\frac{n - k \cdot 2^j}{2^j}\right), \tag{5.4}$$

gdje je skala $\frac{1}{\sqrt{2^j}}$ faktor normalizacije, 2^j je skala koja komprimira ili širi valić, $k \cdot 2^j$ pomiče valić u vremenu, indeksi j i k određuju frekvenciju i položaj valića u vremenu [100], [101], [102].

Dekompozicija diskretnog signala $x[n]$ provodi se korištenjem niskopropusnih i visokopropusnih filtara za hvatanje niskofrekventnih i visokofrekventnih komponenti. Rezultati dekompozicije su koeficijenti aproksimacije i detalja na svakoj razini.

Koeficijenti aproksimacije A_j su definirani na sljedeći način:

$$A_j[n] = \sum_k x[k] \cdot g[n-2k], \tag{5.5}$$

gdje je $g[n]$ izlaz niskopropusnog filtra.

Koeficijenti detalja D_j u koje spadaju vertikalni, horizontalni i dijagonalni koeficijenti su definirani kao;

$$D_j[n] = \sum_k x[k] \cdot h[n-2k], \tag{5.6}$$

gdje je $h[n]$ izlaz visokopropusnog filtra.

Izvorni signal može se rekonstruirati korištenjem koeficijenata aproksimacije i detalja kako slijedi:

$$x[n] = \sum_k A_j[k] \cdot g[n-2k] + \sum_k D_j[k] \cdot h[n-2k]. \tag{5.7}$$

Diskretna valićna transformacija dijeli signal na vremensku i frekvencijsku komponentu. U usporedbi s Fourierovom transformacijom, koja daje samo podatke o frekvenciji, DWT dodatno daje vremensku lokalizaciju što ga čini prikladnim za analizu nestacionarnih ili prijelaznih signala.

Koraci za izvođenje analize diskretne valićne transformacije su sljedeći:

1. odabir skupa podataka,
2. učitavanje I/Q podataka iz .bin datoteke,
3. podjela podataka u segmente zbog lakšeg procesiranja,
4. izvođenje 2D valićne jednorazinske dekompozicije na I/Q podacima,
5. izvlačenje 4 vrste koeficijenata: aproksimacijski, horizontalni, vertikalni i dijagonalni,
6. kreiranje i spremanje koeficijenata u oblike slike u .tiff formatu.

S obzirom na velike procesorske zahtjeve i računsku složenost metoda radio frekvencijskog otiska te nemogućnost obrade podataka odjednom, uvedena je obrada nekoliko jednakih dijelova (segmenata) signala paralelno. Umjesto obrade jednog po jednog dijela signala, više

dijelova signala obrađeno je u svakoj iteraciji što omogućuje kompromis između upotrebe memorije i brzine računanja. Optimalno trajanje dijela signala za diskretnu valićnu transformaciju je 4 ms. Broj segmenata signala po iteraciji je 100 što znači da je 400 ms podataka obrađeno u jednoj iteraciji. Nakon obrade ovih 100 segmenata signala, slijedi nova iteracija u kojoj se novi skup segmenata obrađuje do kraja datoteke u kojoj su pohranjeni podaci.

5.1.2. Metoda temeljena na primjeni spektrograma

Kao što je već spomenuto, RFF je metoda koja se koristi za prepoznavanje i razlikovanje pojedinačnih radio uređaja na temelju njihovih jedinstvenih karakteristika prilikom slanja radio signala. Jedna od mogućih metoda za stvaranje jedinstvenih radiofrekvencijskih otisaka radio signala je korištenje spektrograma. Spektrogram je grafički prikaz snage signala u vremenskoj i frekvencijskoj domeni. Detekcija napada lažiranjem korištenjem spektrograma radi na način da se određene značajke signala izdvajaju iz spektrograma. Te značajke su specifične za određeni uređaj (predajnik) i teško ih je replicirati. Temeljem izdvojenih značajki, detektira se prisutnost lažiranih signala. Lažirani signali obično imaju varijacije u spektru zbog razlika u hardveru uređaja za lažiranje te bi ih stoga trebalo lako detektirati.

Kod detekcije napada lažiranjem, spektrogrami se kreiraju iz podataka signala. Detekcija napada lažiranjem izvršena je na način da se određene značajke signala izdvajaju iz spektrograma te se na temelju odabranih značajki detektira prisutnost lažnog signala. Lažni signali obično imaju varijacije u spektru zbog razlika u hardveru uređaja za lažiranje. Analiza spektrograma u ovom istraživanju odnosi se na pre-korelacijsku fazu te je izvedena u nekoliko koraka:

1. odabir skupa podataka,
2. učitavanje I/Q podataka iz .bin datoteke,
3. podjela podataka u segmente zbog lakšeg procesiranja,
4. kreiranje spektrograma na temelju I/Q komponenti signala,
5. spremanje spektrograma u obliku slike u .tiff formatu.

Parametri korišteni za izradu spektrograma su: veličina prozora (engl. *window size*), preklapanje (engl. *overlap*) i (engl. *nfft*). Veličina prozora definira duljinu segmenta svakog signala koji se koristi za izračun kratkotrajne Fourierove transformacije STFT. Bolja frekvencijska rezolucija postiže se s većom veličinom prozora, ali s lošijom vremenskom rezolucijom. Za parametar preklapanja obično se koristi 50% (polovica podataka iz jednog prozora dijeli se sa sljedećim) jer omogućuje glatke prijelaze između segmenata te je zato u ovom radu korišten parametar preklapanja 50%. Parametar preklapanja definira broj točaka koje se koriste za Fourierovu transformaciju. Veće vrijednosti parametra preklapanja

daju bolju frekvencijsku rezoluciju i više podatkovnih točaka u frekvencijskoj domeni, ali s većim računalnim zahtjevima. Optimalno trajanje dijela signala za spektrogram je 2 ms. Broj segmenata signala po iteraciji je 100 što znači da je 200 ms podataka obrađeno u jednoj iteraciji nakon čega slijede ostale iteracije do kraja datoteke s podacima.

Matematički model

Glavne matematičke metode i modeli koji se koriste za opisivanje i analizu spektrograma temelje se na obradi signala (Fourierova transformacija). Glavni matematički koraci za kreiranje spektrograma su sljedeći:

- Fourierova transformacija - koristi se za pretvorbu signala iz vremenske u frekvencijsku domenu. Ova transformacija omogućuje dobivanje spektra signala. Diskretna Fourierova transformacija definirana je sljedećim izrazom:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, 2, \ldots, N-1, \tag{5.8}$$

  gdje je k diskretni frekvencijski indeks, N je broj točaka u signalu [98].

- Kratkotrajna Fourierova transformacija STFT - koristi se za generiranje spektrograma signala. STFT omogućuje analizu promjena frekvencija tijekom vremena, što je ključno u otkrivanju lažiranih signala. STFT daje dvodimenzionalni prikaz signala u vremenskoj i frekvencijskoj domeni, što je osnova za spektrogram. STFT se definira kao:

$$X[m,k] = \sum_{n=0}^{N-1} x[n] w[n-m] e^{-j\frac{2\pi}{N}kn}, \tag{5.9}$$

  gdje je $x[n]$ diskretni vremenski signal, $w[n]$ prozorska funkcija sa središtem u $n = m$, N je broj diskretnih frekvencijskih točaka, m je vremenski indeks koji definira položaj prozora $w[n-m]$, i k je frekvencijski indeks [99].

- Spektrogram - pokazuje intenzitet signala za svaku frekvenciju tijekom vremena što omogućuje analizu jedinstvenih karakteristika signala. Spektrogram se dobije kvadriranjem apsolutne vrijednosti kratkotrajne Fourierove transformacije:

$$S[m,k] = |X[m,k]|^2, \tag{5.10}$$

  gdje $X[m,k]$ predstavlja kratkotrajnu Fourierovu transformaciju diskretnog signala $x[n]$, $S[m,k]$ je spektrogram koji predstavlja kvadrat magnitude kratkotrajne Fourierove transformacije.

5.1.3. Klasifikacijski algoritam koji koristi slike kao ulazne poodatke

Na temelju definiranog pristupa za detekciju napada lažiranjem prikazanog blok dijagramom na slici 5.1 može se definirati odgovarajući algoritam za klasifikaciju A1 koji koristi slike kao ulazni klasifikacijski podatak. U okviru algoritma mogu se koristiti različiti modeli strojnog učenja (primjerice SVM, KNN, RF...). U tablici 5.1 prikazani su koraci algoritma A1 uz primjenu modela SVM i njihova računska složenost. Prikaz pseudokoda algoritma A1 dan je u nastavku teksta.

Algoritam A1 Klasifikacijski algoritam A1 koji koristi slike kao ulazni klasifikacijski podatak

```
1:  Učitavanje slika:
2:     slike = učitaj('putanja/do/slika');
3:  Obrada slika i generiranje torbe značajki BoF:
4:     obrada = generiraj_torbu_značajki(slike);
5:  Podjela podataka u k podskupova i unakrsna provjera valjanosti k podskupova:
6:     kf = kfold_podjela(podaci, 'kfold', K);
7:  for i = 1 to R do
8:      for j = 1 to k do
9:          Podjela podataka na skupove za treniranje i testiranje:
10:            Trening_podaci;
11:            Test_podaci;
12:         Treniranje modela:
13:            model = treniraj_model(Trening_podaci, Test_podaci);
14:         Predikcija podataka:
15:            predikcija = predvidi_podatke();
16:         Evaluacija modela i izračun parametara performansi modela:
17:                         [točnost, preciznost, odziv, mjera F1, konfuzijska matrica] =
    evaluiraj(predikcija);
18:     end for
19: end for
20: Generiranje krivulja ROC i PR:
21:    kreiraj_ROC();
22:    kreiraj_PR();
23: Spremanje svih rezultata i grafova.
```

U nastavku je dan prikaz značenja pojedinih linija u pseudokodu algoritma A1. Linije 1 i 2 algoritma A1 prikazuju učitavanje skupa podataka slika na način da se koristi funkcija za učitavanje slika kojoj je potrebno definirati samu putanju do foldera. U linijama 3 i 4 koristi se BoF pristup za obradu slika i generiranje torbe značajki. Podjela podataka u k podskupova je definirana linijom 6. Linije 7 i 8 definiraju *for* petlje za R iteracija algoritma i k podskupova. Podjela podataka na skupove za treniranje i testiranje prikazana je linijama 9, 10 i 11. Treniranje modela korištenjem funkcije *treniraj_model* prikazana je linijom 13. Linija 15 predstavlja predikciju podataka na način da se uzima novi uzorak iz skupa za testiranje i za njega se radi klasifikacija i odluka da li je uzorak autentičan ili lažan. Evaluacija modela i izračun parametara performansi modela kao što su: točnost, preciznost, odziv, mjera *F1* i konfuzijska matrica prikazani su linijom 16. Linije 18 i 19 predstavljaju završetak obje *for* petlje. Linije 21 i 22 prikazuju generiranje krivulja ROC i PR i konačno linija 23 predstavlja

spremanje svih rezultata i grafova u programu Matlab (slike se spremaju u formatu .fig).

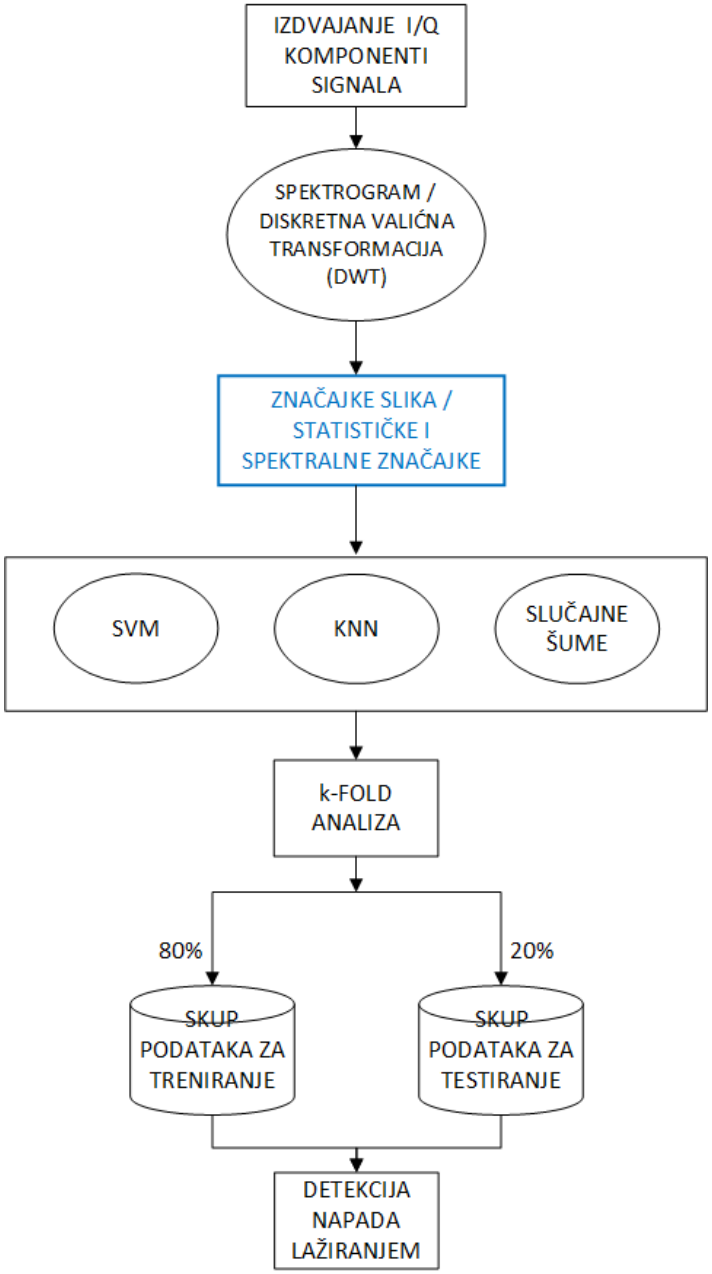
5.2. Analiza računske složenosti i prijedlog računski učinkovitog algoritma

U ovom poglavlju prikazan je prijedlog računski učinkovitijeg modificiranog algoritma A1M u kojemu se umjesto korištenja slika koriste prethodno izvučene značajke. Definirane su dvije inačice algoritma i to: A1M1 u kojemu se koriste izvučene značajke slika i A1M2 u kojemu se koriste statističke i spektralne značajke. Nadalje, prikazana je i analiza računske složenosti za predložene metode radio frekvencijskog otiska i metode strojnog učenja korištene za klasifikaciju tipa signala korištenjem osnovnog klasifikacijskog algoritma A1 i računski učinkovitijeg modificiranog algoritma A1M.

5.2.1. Prijedlog modificiranog računski učinkovitijeg algoritma

Na slici 5.2 prikazan je blok dijagram pristupa za detekciju lažiranih signala uz primjenu računski učinkovitijeg algoritma A1M koji koristi prethodno izvučene značajke kao klasifikacijski podatak. Prvotno predloženi pristup koristi slike kao ulazne podatke za klasifikacijske modele i ima veću računsku složenost zbog učitavanja i obrade svake pojedine slike koristeći BoF pristup. Vrijeme izvršavanja algoritma za predložene pristupe prikazano je u tablicama 6.2, 6.3, 6.4 i 6.5. Kako bi se smanjilo vrijeme izvođenja pojedinog algoritma i računska složenost, predložen je računski učinkovitiji pristup koji umjesto slika koristi značajke slike i statističke i spektralne značajke kao ulazne klasifikacijske podatke. Iz tablice 6.2 može se vidjeti da je vrijeme izvođenja algoritma koji koristi značajke slike i statističke i spektralne značajke smanjeno značajno u odnosu na algoritme koji koriste slike kao ulazne klasifikacijske podatke. Temeljem pristupa prikazanog blok dijagramom na slici 5.2, u nastavku je prikazan pseudokod za modificirani računski učinkovitiji algoritam A1M s dvije inačice.

U nastavku je dan prikaz značenja pojedinih linija u pseudokodu modificiranog algoritma A1M. Linije 1 i 2 prikazuju učitavanje značajki iz .csv datoteke. Prebacivanje učitanih značajki iz .csv datoteke u tablicu te dodjela klasa podacima prikazano je linijama 3-7. Uzorcima su dodijeljene klase: 1 za autentični signal i 2 za lažni signal. Da bi se ubrzala obrada podataka iz tablice, tablica je pretvorena u niz što je prikazano u linijama 6 i 7. Podjela podataka u k podskupova je definirana linijom 9. Linije 10 i 11 definiraju *for* petlje za R iteracija algoritma i k podskupova. Podjela podataka na skupove za treniranje i testiranje prikazana je linijama 13 i 14. Treniranje modela korištenjem funkcije *treniraj_model* prikazana je linijom 16. Linija 18 predstavlja predikciju podataka na način da se uzima novi uzorak iz skupa za testiranje i za njega se radi klasifikacija i odluka da li je uzorak autentičan ili lažan. Evaluacija modela i izračun parametara performansi modela kao što su: točnost, preciznost, odziv, mjera *F1* i konfuzijska matrica prikazani su linijom 20. Linije 21 i 22 predstavljaju



Slika 5.2: Blok dijagram predloženog pristupa za detekciju lažiranih signala uz primjenu računski učinkovitijeg algoritma koji koristi prethodno izvučene značajke kao klasifikacijski podatak.

završetak obje *for* petlje. Linije 24 i 25 prikazuju generiranje krivulja ROC i PR i konačno linija 26 predstavlja spremanje svih rezultata i grafova u programu Matlab (slike se spremaju u formatu .fig). Modificirani algoritam A1M razlikuje se od osnovnog algoritma A1 u vrsti ulaznih klasifikacijskih podataka i načinu njihove obrade.

Algoritam A1M Modificirani računski učinkovitiji algoritam klasifikacije A1M s dvije inačice: A1M1 koristi značajke slike kao ulazne podatke i A1M2 koristi statističke i spektralne značajke kao ulazne podatke

Učitavanje značajki iz .csv datoteke:
    `značajke = učitaj('putanja/do/datoteke');`
Prebacivanje učitanih značajki u tablicu.
Dodjela klasa podacima:
    `autentični = klasa 1;  lažni = klasa 2;`
Pretvaranje tablice u niz:
    `tablica_u_niz();`
Podjela podataka u k podskupova i unakrsna provjera valjanosti k podskupova:
    `kf = kfold_podjela(podaci, 'kfold', k);`
for $i = 1$ to R **do**
    for $j = 1$ to k **do**
        Podjela podataka na skupove za treniranje i testiranje:
            `Trening_podaci;`
            `Test_podaci;`
        Treniranje modela:
            `model = treniraj_model(Trening_podaci, Test_podaci);`
        Predikcija podataka:
            `predikcija = predvidi_podatke();`
        Evaluacija modela i izračun parametara performansi modela:
            `[točnost, preciznost, odziv, mjera F1, konfuzijska matrica] = evaluiraj(predikcija);`
    end for
end for
Generiranje krivulja ROC i PR:
    `kreiraj_ROC();`
    `kreiraj_PR();`
Spremanje svih rezultata i grafova.

5.2.2. Analiza računske složenosti za algoritam klasifikacije SVM

Tablica 5.1 prikazuje korake za algoritam klasifikacije A1, A1M1 i A1M2 za model SVM i računsku složenost za svaki pojedini dio algoritma gdje je:

- N - ukupan broj uzoraka / slika u skupu podataka,
- D - broj detektiranih značajki po slici. Značajke se ne računaju za svaki piksel slike nego se koristi samo podskup piksela s fokusom na informativne dijelove slike kao što su rubovi i kutovi. Slika se podijeli u mrežu (npr. svakih 8x8 piksela) i značajke se računaju samo na tim točkama. Nakon što se izračunaju značajke na "grid" točkama, rangiraju se po jakosti ili važnosti odnosno uzima se 70% najistaknutijih značajki odnosno one koje najviše doprinose razlikovanju slike.
- K - broj vizualnih riječi (klastera) u BoF modelu. Iz svih slika izvuku se lokalne značajke odnosno vizualne riječi i te se značajke grupiraju u klastere koji predstavljaju vizualne riječi (broj klastera = broj vizualnih riječi u riječniku = dimenzija značajki svake slike u BoF modelu). Zatim se svaka slika prikazuje kao histogram vizualnih riječi koji ide u klasifikator,

- I - broj iteracija klasterizacije,
- R - broj ponavljanja unakrsne provjere valjanosti k-podskupova,
- k - broj podskupova u unakrsnoj provjeri valjanosti,
- M - broj uzoraka za treniranje po iteraciji, $M = N \cdot \frac{k-1}{k}$
- d - dimenzija BoF vektora ovisi o broju vizualnih riječi. Ovaj vektor predstavlja histogram koji broji koliko puta je neka značajka dodijeljena kojoj riječi (klasteru).
- T - broj uzoraka za testiranje, $T = \frac{N}{k}$ (npr. ako je $N = 1000$ i $k = 5$, prema izrazu dobije se da je $T = 200$, a $M = N - T = N \cdot \frac{k-1}{k} = 800$ što odgovara omjeru 80%:20%),
- S - broj potpornih vektora,
- n_{trees} - broj stabala u modelu slučajnih šuma.

Iz tablice se može vidjeti da se algoritam A1 koji koristi slike kao ulazne klasifikacijske podatke razlikuje od druga dva algoritma A1M1 i A1M2 u dva koraka, a to su generiranje torbi značajki (engl. *Bag of Features*) u kojem se koriste modeli za obradu slike te korak u kojem se izvlače značajke iz svake generirane slike korištenjem metode SURF (engl. *Speeded-Up Robust Feature*). U algoritmima A1M1 i A1M2, ulazni podaci su brojčane vrijednosti iz .csv datoteka i nema dodatne obrade slike te je zbog toga računska složenost za ove algoritme manja u odnosu na algoritam A1. Što se tiče ostalih modela strojnog učenja, KNN i RF, kod njih se u odnosu na SVM razlikuje korak treniranja i testiranja modela. Sve slike koje su se koristile kao ulazni podaci su u 2D formatu.

Tablica 5.1: Koraci za algoritam klasifikacije A1, A1M1 i A1M2 za SVM model i odgovarajuća računska složenost [122], [123], [124].

Korak	**Opis koraka**	**Algoritam / Vrsta podataka**			**Računska složenost**
		A1 / slike	A1M1 / značajke slike	A1M2 / statističke i spektralne značajke	
A	Učitavanje podataka	✓	✓	✓	$O(N)$
B	Generiranje torbe značajki (*Bag of Features*)	✓	–	–	$O(N \cdot D \cdot K \cdot I)$
C	Izvlačenje značajki	✓	–	–	$O(N \cdot D \cdot K)$
D	Dodjela klasa podacima	–	✓	✓	$O(N)$
E	Pretvorba tablice u niz	–	✓	✓	$O(N \cdot d)$
F	Podjela podataka u k-podskupove	✓	✓	✓	$O(N)$
G	Unakrsna provjera valjanosti k-podskupova (petlja, R ponavljanja)	✓	✓	✓	$O(R)$
H	Treniranje i testiranje modela u svakoj iteraciji	✓	✓	✓	$O(k)$
I	Treniranje modela s Gaussovom jezgrom	✓	✓	✓	$O(M^2 \cdot d)$
J	Predikcija podataka	✓	✓	✓	$O(T \cdot S \cdot d)$
K	Izračun parametara za performanse modela	✓	✓	✓	$O(1) - O(N)$
L	Generiranje krivulja ROC i PR	✓	✓	✓	$O(1) - O(N)$
M	Grafički prikaz i spremanje podataka	✓	✓	✓	$O(N)$

Prema tablici 5.1 prikazan je izračun ukupne računske složenosti za algoritam klasifikacije SVM A1 i SVM A1M.

Ukupna računska složenost za algoritam klasifikacije A1 SVM dobije se zbrajanjem računskih složenosti za svaki pojedini dio koda i dana je sljedećim izrazom:

$$\begin{aligned} O_{SVM_{A1}}(ukupno) = A+B+C+F+G+H+I+J+K+L+M = \\ O(N)+\underline{O(N \cdot D \cdot K \cdot I)}+O(N \cdot D \cdot K)+O(N)+\underline{O(R)}+\underline{O(k)}+\underline{O(M^2 \cdot d)} \\ +\underline{O(T \cdot S \cdot d)}+O(N)+O(N)+O(N) = \\ O(N \cdot D \cdot K \cdot I)+O(R)+O(k)+O(M^2 \cdot d)+O(T \cdot S \cdot d) = \\ O(N \cdot D \cdot K \cdot I)+O(R \cdot k(M^2 \cdot d+T \cdot S \cdot d)). \end{aligned} \tag{5.11}$$

Članovi A, C, F, K, L i M sadržani su u članu B i imaju linearnu složenost te se stoga mogu zanemariti u krajnjem izrazu za ukupnu računsku složenost. Član $O(T \cdot S \cdot d)$ koji se odnosi na testiranje modela isto ima linearnu složenost i zanemariv je u odnosu na kvadratnu složenost treniranja. Stoga se može definirati dominantna računska složenost kao:

$$O_{SVM_{A1}}(dominantna) = O(N \cdot D \cdot K \cdot I)+O(M^2 \cdot d). \tag{5.12}$$

Dominantni dio računske složenosti predstavlja generiranje torbe značajki BoF $O(N \cdot D \cdot K \cdot I)$ te treniranje modela s Gaussovom jezgrom $O(M^2 \cdot d)$ koje ima kvadratnu složenost po broju uzoraka za treniranje. Razlozi za navedeno su zato što se u BoF modelu vrši obrada slike na način da se iz svih slika izvuku lokalne značajke i svaka slika se zatim prikazuje kao histogram vizualnih riječi koji se koristi za klasifikaciju. Treniranje modela s Gaussovom jezgrom je također dominantan član zbog toga što ima kvadratnu složenost jer se u procesu treniranja nelinearnog modela računa matrica jezgrene funkcije između svakog para uzoraka za treniranje što je u ovom slučaju $MxM = M^2$ izračuna jezgrene funkcije [128]. Dakle, $M^2 \cdot d$ predstavlja donju granicu treniranja jezgrene funkcije: kvadratno zbog matrice jezgrene funkcije dimenzija MxM i linearno po dimenziji značajke d [129]. Tsang [127] navodi $= O(m^3)$ i $O(m^2)$ kao računske složenosti za standardni model SVM koji ima m uzoraka za treniranje. Računska složenost ide čak i do kubne složenosti u slučaju kada se koriste veliki skupovi podataka. Računska složenost od n^3 gdje je n broj uzoraka za treniranje za standardni SVM navedena je i u radu [130].

Proces treniranja i testiranja se ponavlja k puta po svakom od R ponavljanja u unakrsnoj provjeri valjanosti k podskupova što je u kodu prikazano kao:

for $i = 1$ to R **do**
  for $j = 1$ to k **do**
    $Mdl = fitcsvm(...)$;
    $predict(Mdl, ...)$;
  end for
end for

Ukupna računska složenost za klasifikacijski model SVM A1M1 prikazana je izrazom:

$$\begin{aligned} O_{SVM_{A1M1}}(ukupno) = A+D+F+G+H+I+J+K+L+M = \\ O(N)+O(N)+O(N \cdot d)+O(N)+O(R)+O(k)+O(M^2 \cdot d) \\ +O(T \cdot S \cdot d)+O(N)+O(N)+O(N) = \\ O(N \cdot d)+O(R)+O(k)+O(M^2 \cdot d)+O(T \cdot S \cdot d) = \\ O(N \cdot d)+O(R \cdot k \cdot (M^2 \cdot d)). \end{aligned} \tag{5.13}$$

Najdominantniji član računske složenosti je treniranje SVM klasifikatora s Gaussovom jezgrenom funkcijom $M^2 \cdot d$ koji ima kvadratnu složenost po broju uzoraka za treniranje dok su ostali dijelovi zanemarivi. Isto vrijedi i za algoritam klasifikacije A1M2 SVM koji kao ulazne klasifikacijske podatke koristi statističke i spektralne značajke.

Može se zaključiti da računska složenost za model SVM ovisi o korištenoj jezgrenoj funkciji te o hardverskoj izvedbi korištene opreme koja utječe na brzinu izvršavanja algoritma. Linearna jezgrena funkcija je najbrža i ima nanjnižu složenost dok je polinomska

najzahtjevnija i ima najveću računsku složenost. Gaussova jezgrena funkcija ima najbolju računsku složenost od nelinearnih jezgrenih funkcija [131]. Nadalje, možemo zaključiti da pristupi koji koriste značajke slike i statističke i spektralne značajke imaju manju računsku složenost i manje vrijeme izvršavanja (Tablice 6.2 i 6.3) u odnosu na pristup koji koristi slike kao ulazni klasifikacijski podatak zbog obrade svake pojedine slike.

5.2.3. Analiza računske složenosti za algoritam klasifikacije KNN

Računska složenost za algoritam klasifikacije KNN razlikuje se u odnosu na SVM samo u fazi treniranja i testiranja modela. Stoga je ukupna računska složenost za algoritam klasifikacije A1 za model KNN jednaka:

$$\begin{aligned} O_{KNN_{A1}}(ukupno) = A+B+C+F+G+H+Trening+Test+K+L+M = \\ O(N)+O(N\cdot D\cdot K\cdot I)+O(N\cdot D\cdot K) \\ +O(N)+O(R)+O(k)+O(1)+O(T\cdot M\cdot d)+O(N)+O(N)+O(N) = \\ O(N\cdot D\cdot K\cdot I)+O(R)+O(k)+O(1)+O(T\cdot M\cdot d) = \\ O(N\cdot D\cdot K\cdot I)+O(R\cdot\frac{N^2\cdot d}{k}) \end{aligned} \tag{5.14}$$

Drugi član dobiven je iz izraza $T\cdot M\cdot d$ nakon uvrštavanja $T=\frac{N}{k}$ i $M=\frac{N\cdot(k-1)}{k}$. Klasifikator KNN ne trenira model nego samo pohranjuje uzorke za treniranje i stoga je računska složenost za treniranje jednaka $O(1)$ te je zanemariva. Dominantna računska složenost za KNN odnosi se na generiranje BoF značajki i obradu slike te testiranje modela. Faza testiranja modela uključuje usporedbu svakog testnog uzorka T sa svakim uzorkom za treniranje M odnosno za svaki testni uzorak mora se izmjeriti udaljenost do svakog uzorka za treniranje i ukupno se napravi N^2 mjerenja udaljenosti. Dominantna računska složenost iznosi:

$$O_{KNN_{A1}}(dominantna) = O(N\cdot D\cdot K\cdot I)+O(\frac{N^2\cdot d}{k}) \tag{5.15}$$

Ukupna računska složenost za algoritam klasifikacije A1M1 za model KNN dobivena je zbrajanjem računskih složenosti za svaki pojedini dio algoritma i prikazana je izrazom:

$$\begin{aligned} O_{KNN_{A1M1}}(ukupno) = A+D+E+F+G+H+Trening+Test+K+L+M = \\ O(N)+O(N)+O(N\cdot d)+O(N)+O(R)+O(k) \\ +O(1)+O(T\cdot M\cdot d)+O(N)+O(N)+O(N) = \\ O(N\cdot d)+O(R)+O(k)+O(1)+O(T\cdot M\cdot d) = \\ O(N\cdot d)+O(R\cdot \frac{N^2\cdot d}{k}). \end{aligned} \tag{5.16}$$

Dominantna računska složenost iznosi:

$$O_{KNN_{A1M1}}(dominantna) = O(\frac{N^2\cdot d}{k}). \tag{5.17}$$

Ova dominantna složenost je kvadratna u odnosu na sve ostale članove i dolazi iz KNN klasifikacije zbog toga što se računa udaljenost između svih parova uzoraka pri klasifikaciji svakog testa tj. uspoređuje se svaki novi testni uzorak s pohranjenim uzorkom za treniranje. Kvadratnu računsku složenost samo za KNN klasifikaciju prikazuju i autori u radu [132] i ona iznosi $O(n_{TD}^2)$, gdje je n_{TD}^2 veličina skupa za treniranje. Računska složenost za KNN ovisi o veličini skupa podataka što izravno utječe na vrijeme potrebno za usporedbu svakog novog testnog uzorka sa svakim uzorkom za treniranje. KNN nije dobar izbor za velike skupove podataka zbog toga što zahtijeva spremanje svih uzoraka za treniranje što je memorijski veoma zahtjevno i povećava složenost. Za algoritam A1M2 za model KNN koji koristi statističke i spektralne značajke, računska složenost je ista kao i za algoritam A1M1 koji koristi značajke slike jer se koriste podaci iz .csv datoteka. Može se zaključiti da algoritmi A1M1 i A1M2 za model KNN imaju manju računsku složenost u odnosu na algoritam A1 KNN.

5.2.4. Analiza računske složenosti za algoritam klasifikacije RF

Kao i za prethodna dva algoritma, u izračunu računske složenosti koriste se koraci iz Tablice 5.2. Ukupna računska složenost za algoritam klasifikacije A1 za model RF prikazana je izrazom:

$$\begin{aligned} O_{RF_{A1}}(ukupno) = A+B+C+F+G+H+Trening+Test+K+L+M = \\ O(N)+O(N\cdot D\cdot K\cdot I)+O(N\cdot D\cdot K)+O(N)+O(R)+O(k)+ \\ O(n_{trees}\cdot M\cdot logM\cdot d)+O(n_{trees}\cdot T\cdot logM)+O(N)+O(N)+O(N) = \\ O(N\cdot D\cdot K\cdot I)+O(R\cdot k\cdot n_{trees}\cdot M\cdot logM\cdot d). \end{aligned} \tag{5.18}$$

Ukupna računska složenost dobivena je nakon sređivanja i uvrštavanja $M \approx N$ jer je faza testiranja zanemariva u odnosu na treniranje.
Dominantna računska složenost iznosi:

$$O_{RF_{A1}}(dominantna) = O(N \cdot D \cdot K \cdot I) + O(N \cdot logN \cdot d). \tag{5.19}$$

Treniranje modela $N \cdot logN \cdot d$ predstavlja najdominantniji član računske složenosti zbog toga što predstavlja složenost izgradnje jednog stabla tj. stablo se gradi rekurzivnim dijeljenjem čvorova i svaki čvor prima jednu podskupinu uzoraka te se zatim sortiraju vrijednosti značajki u svakom stablu. $N \cdot logN$ dolazi iz rekurzivne podjele i sortiranja u stablu dok d označava broj značajki koje se analiziraju u svakom čvoru. Ostatak izraza $R \cdot k \cdot n_{trees}$ su konstante i ne utječu na složenost nego samo povećavaju broj operacija multiplikativno. Tradicionalna vremenska računska složenost za algoritam RF koji kao ulazne podatke koristi slike prikazana je u radu [133] i iznosi $n \cdot logn$ dok potpuna parametarska računska složenost kako je nazivaju autori iznosi $F(s \cdot (m+1) \cdot n \cdot v \cdot logn)$, gdje je n broj uzoraka u skupu za treniranje, s broj stabala u modelu, m je broj klasa i v je broj značajki. Drugi najdominantniji član se odnosi na generiranje torbe značajki BoF zbog postupka učenja vizualnih riječi gdje se sve značajke slike grupiraju u K klastera. Nadalje, velik broj lokalnih deskriptora iz svake slike treba biti obrađen i pri tome se za svaku točku mjeri udaljenost do svih centara klastera tijekom I iteracija. Ovaj dio je dominantan pogotovo ako se koristi veliki broj slika i slike visoke rezolucije.

Računska složenost za algoritme klasifikacije A1M1 i A1M2 za model RF ima isti izraz za ukupnu računsku složenost:

$$\begin{aligned} O_{RF_{A1M1}}(ukupno) = A + D + E + F + G + H + Trening + Test + K + L + M = \\ O(N) + O(N) + O(N \cdot d) + O(N) + O(R) + O(k) + O(n_{trees} \cdot M \cdot logM \cdot d) \\ + O(n_{trees} \cdot T \cdot logM) + O(N) + O(N) + O(N) \\ = O(N \cdot d) + O(R \cdot k \cdot trees \cdot N \cdot logn \cdot d). \end{aligned} \tag{5.20}$$

Najdominantniji član računske složenosti je $N \cdot logN \cdot d$ iz istog razloga koji je naveden u slučaju kad su ulazni podaci slike. Računska složenost algoritma koji koristi slike kao ulazne podatke je veća u odnosu na algoritam koji koristi unaprijed definirane značajke odnosno brojčane vrijednosti.

5.2.5. Usporedba računske složenosti za algoritam klasifikacije A1, A1M1 i A1M2 za SVM, KNN i RF

Tablica 5.2 prikazuje računsku složenost za algoritam klasifikacije A1, A1M1 i A1M2 za korištene modele strojnog učenja SVM, KNN i RF.

Tablica 5.2: Računska složenost za algoritam klasifikacije A1, A1M1 i A1M2 za modele strojnog učenja SVM, KNN i RF.

Algoritam / Klasifikacijski podaci	SVM	KNN	RF - 100 / 200
A1 / slike	$O(N \cdot D \cdot K \cdot I + R \cdot k \cdot (M^2 \cdot d + T \cdot S \cdot d))$	$O(N \cdot D \cdot K \cdot I + R \cdot \frac{N^2 \cdot d}{k})$	$O(N \cdot D \cdot K \cdot I + R \cdot k \cdot n_{trees} \cdot N \cdot logN \cdot d)$
A1M1 / značajke slika	$O(N \cdot d + R \cdot k \cdot M^2 \cdot d)$	$O(N \cdot d + R \cdot \frac{N^2 \cdot d}{k})$	$O(N \cdot d + R \cdot k \cdot n_{trees} \cdot N \cdot logN \cdot d)$
A1M2 / statističke i spektralne značajke	$O(N \cdot d + R \cdot k \cdot M^2 \cdot d)$	$O(N \cdot d + R \cdot \frac{N^2 \cdot d}{k})$	$O(N \cdot d + R \cdot k \cdot n_{trees} \cdot N \cdot logN \cdot d)$

Računska složenost algoritma klasifikacije A1 za model SVM sastoji se od dva dijela kao što je prikazano u tablici 5.2. Prvi dio odnosi se na generiranje torbe značajki dok se drugi dio odnosi na unakrsnu provjeru valjanosti te treniranje i testiranje modela. Dominatni dijelovi su BoF pristup koji uključuje obradu slike i treniranje modela kako prikazuje izraz (5.12). Najdominantniji član je M^2 koji raste kvadratno za Gaussovu jezgru (može ići i do M^3). Računska složenost ovog modela može se smanjiti smanjenjem rezolucije slike, prebacivanjem slike u sivu skalu, korištenjem linearne umjesto Gaussove jezgre, smanjivanjem broja ponavljanja validacije R te smanjivanjem broja podskupova k (iz toga razloga je u ovom istraživanju broj ponavljanja jednak 3, a broj podskupova 5). Ako se usporedi algoritam klasifikacije A1 koji kao ulazne podatke koristi i obrađuje slike tj. generira torbu značajki s algoritmima klasifikacije A1M1 i A1M2 u kojima se koriste značajke slike i statističke i spektralne značajke, može se zaključiti da je algoritam koji koristi BoF pristup složeniji jer uključuje obradu slike, izvlačenje ključnih značajki, klasteriranje u vizualne riječi i kodiranje svake slike u histogram riječi. S druge strane, algoritmi koji koriste unaprijed izračunate značajke su mnogo lakši za procesiranje (manja složenost, brža obrada i manji memorijski zahtjevi) jer nema potrebe za izvlačenjem značajki niti klasteriranjem.

Kod modela KNN ne vrši se eksplicitni trening podataka tj. model ne uči parametre ni funkcije nego pamti sve podatke iz skupa za treniranje te je računska složenost za treniranje zanemariva i iznosi $O(1)$ [125]. Algoritam klasifikacije A1 za model KNN koji ima slike kao ulazne podatke ima BoF pristup i testiranje modela kao najdominantnije korake ((5.15)). Kvadratna složenost po broju podataka (tablica 5.2) proizlazi iz mjerenja udaljenosti između svakog testnog uzorka i uzorka za treniranje tj. mora računati udaljenost do svih točaka što može biti kompleksno za velike skupove podataka. Klasifikacija modela KNN je najjednostavnija za implementaciju jer nema potrebe za treniranjem, ali pohranjuje sve podatke što je memorijski zahtjevno. Što se tiče algoritma klasifikacije A1M1 i A1M2 za model KNN, njihova računska složenost je manja jer ne koriste BoF pristup i obradu slika. Najdominantniji

član je treniranje modela koje isto ima kvadratnu složenost.

Za algoritam klasifikacije A1 za model RF, generiranje torbe značajki i treniranje modela su najdominantniji članovi računske složenosti [126] što je dano izrazom (5.19). Za algoritme klasifikacije A1M1 i A1M2 za model RF koji koriste unaprijed izračunate značajke kao ulazne podatke za klasifikaciju, najdominantniji član je treniranje modela jer predstavlja cijeli proces izgradnje jednog stabla. Algoritmi A1M1 i A1M2 imaju manju složenost jer ne koriste BoF pristup i obradu slike. Model RF pogodan je za velike skupove podataka, a složenost ovisi o broju stabala, broju značajki i veličini skupa podataka.

Najveću računsku složenost ima model KNN zato što ima kvadratnu složenost po broju uzoraka N^2 za sve algoritme klasifikacije A1, A1M1 i A1M2, a u kombinaciji s generiranjem torbe značajki i obradom slike predstavlja najkompleksniji pristup. S druge strane, najmanju složenost ima model SVM koji koristi unaprijed izvučene značajke jer nema kvadratnu složenost po broju uzoraka N^2 niti procesiranje slika, i ako M nije prevelik, ovo je najjednostavniji pristup po računskoj složenosti.

5.2.6. Analiza računske složenosti za diskretnu valićnu transformaciju

Algoritam diskretne valićne transformacije sastoji se od sljedećih koraka i pripadajućih računskih složenosti:

1. Učitavanje I/Q podataka iz .bin datoteke i računska složenost

   $O(S)$, gdje je $S = D \cdot P$ ukupan broj uzoraka po iteraciji, D je broj uzoraka u promatranom djeliću signala, a P broj djelića signala po iteraciji.

2. Pretvorba 1D kompleksnog signala u 2D kvadratnu matricu (kombinacija $I + j \cdot Q$, j se odnosi na imaginarnu jedinicu) dimenzija *nxn* - $O(n^2)$

3. Primjena dvodimenzionalne jednorazinske diskretne valićne transformacije na skup podataka - $O(n^2)$

4. Izdvajanje statističkih značajki za prvu razinu aproksimacijskih koeficijenata - $O(n^2)$

5. Izvlačenje spektralnih značajki i primjena brze Fourierove transformacije - $O(n^2 \cdot logn)$

6. Spremanje značajki u .csv datoteku - $O(1)$

7. Generiranje i spremanje jednorazinskih aproksimacijskih koeficijenata (*A1img*) u obliku slike u *.tiff* formatu - $O(n^2)$

8. Izvlačenje sljedećih značajki iz slike u *.tiff* formatu i spremanje u .csv datoteku:

   - histogram po RGB kanalu - $O(n^2 \cdot C)$, gdje je C broj kanala slike (3 kanala za RGB)

- gradijentna magnituda - $O(n^2 \cdot C)$
- teksturne značajke - $O(n^2 \cdot C + G \cdot C)$, gdje je G broj GLCM (engl. *Gray-Level Co-occurrence Matrix*) metrika (kontrast, korelacija, energija i homogenost). Ukupan broj ovih GLCM značajki je 12 tj. 3 kanala puta 4 značajke.

Ukupna računska složenost algoritma diskretne valićne transformacije za cijelu obradu .bin datoteke dobije se zbrajanjem računskih složenosti po koracima algoritma i prikazana je izrazom:

$$\begin{aligned} O_{DWT}(ukupno) = 1+2+3+4+5+6+7+8 = \\ O(S)+O(n^2)+O(n^2)+O(n^2)+ \\ O(n^2 \cdot logn)+O(1)+O(n^2)+O(n^2 \cdot C)+O(G \cdot C) = \\ O(S)+O(n^2 \cdot logn)+O(n^2 \cdot C). \end{aligned} \tag{5.21}$$

Konstanta G može se zanemariti budući da je $n^2 >> G$.

S obzirom da se signal pretvara u 2D sliku, vrijedi da je jedna dimenzija kvadratne slike n jednaka kvadratnom korijenu broja uzoraka signala S, u izraz (5.21) umjesto n^2 uvrstimo S odnosno $n = \sqrt{S}$ te se izraz za ukupnu računsku složenost može zapisati kao:

$$\begin{aligned} O_{DWT}(ukupno) = O(S)+O(S)+O(S)+O(S)+O(S \cdot logS) \\ +O(1)+O(S)+O(S \cdot C)+O(G \cdot C) = O(S)+O(S \cdot logS). \end{aligned} \tag{5.22}$$

Dominantna računska složenost odnosi se na primjenu dvodimenzionalne diskretne valićne transformacije, primjenu brze Fourierove transformacije te izvlačenje značajki iz slika i iznosi:

$$O_{DWT}(dominantna) = O(S \cdot logS). \tag{5.23}$$

Ako pretpostavimo da je R ukupan broj iteracija, dobivamo modificirani izraz za dominantnu računsku složenost:

$$\begin{aligned} O_{DWT}(ukupno) = O(S)+O(S)+O(S)+O(S)+O(S \cdot logS) \\ +O(1)+O(S)+O(S \cdot C)+O(G \cdot C) = O(R \cdot S \cdot logS). \end{aligned} \tag{5.24}$$

Ako P i D rastu, onda složenost ide preko $O(n^2)$ što je memorijski vrlo zahtjevno. Račun-

ska složenost se povećava s korištenjem više dekompozicijskih razina. Kako bi se smanjila računska složenost, u ovom istraživanju korištena je samo jedna dekompozicijska razina što omogućava bržu i jednostavniju obradu .bin datoteke.

Primjer izračuna računske složenosti

Parametri koji su korišteni za izračun diskretne valićne transformacije su sljedeći:

- frekvencija uzorkovanja $f_s = 5MHz$,
- trajanje segmenta $T_s = 4ms$,
- broj segmenata po iteraciji $P = 100$,
- broj uzoraka po segmentu $D = f_s \cdot T_s = 5MHz \cdot 4ms = 20000$,
- broj uzoraka po iteraciji $S = D \cdot P = 20000 \cdot 100 = 2000000$

Kada se u izraz (5.24) uvrste korišteni parametri, dobije se $O(10 \cdot 2 \cdot 10^6 \cdot log2 \cdot 10^6) = 42 \cdot 10^6$ operacija po iteraciji odnosno $378 \cdot 10^8$ za svih 900 iteracija.

5.2.7. Analiza računske složenosti za spektrogram

Koraci algoritma za kreiranje spektrograma i odgovarajuća računska složenost su:

1. Učitavanje I/Q uzoraka iz .bin datoteke (funkcija *fread*) i složenost $O(S)$. Svaki spektrogram koristi $S = D \cdot P = f_s \cdot T_{piece}$ uzoraka, f_s je frekvencija uzorkovanja signala, S je ukupan broj uzoraka u spektrogramu, D je broj uzoraka po djeliću signala, T_{piece} je trajanje jednog djelića signala, P je broj djelića po spektrogramu.
2. Generiranje spektrograma (STFT) pomoću funkcije *spectrogram* u programu Matlab [136]. Za svaki prozor se radi brza Fourierova transformacija nad *nfft* točaka pa je složenost $O(T \cdot \text{nfft} \cdot \log \text{nfft})$ [137] gdje T predstavlja broj vremenskih prozora i jednak je $1 + \left\lfloor \frac{S-W}{H} \right\rfloor = \frac{2 \cdot S}{W}$ [135], H predstavlja preklapanje i jednak je $\frac{W}{2}$, W predstavlja širinu prozora za spektrogram npr. 512, 1024.
3. Pretvorba snage u logaritamsku skalu i prikaz spektrograma u obliku slike - $O(T \cdot \text{nfft})$.
4. Spremanje slike u format *.tiff* - $O(T \cdot \text{nfft})$.

Ukupna računska složenost za jedan spektrogram jednaka je zbroju računskih složenosti prema koracima korištenog algoritma i prikazana je izrazom:

$$O_{spektrogram}(ukupno) = 1+2+3+4 = \\ O(S)+O(T\cdot nfft\cdot lognfft)+O(T\cdot nfft)+O(T\cdot nfft). \tag{5.25}$$

Nakon što se uvrsti $T=\frac{2\cdot S}{W}$, izraz postaje:

$$O_{spektrogram}(ukupno) = O(S)+O(\frac{2\cdot S}{W}\cdot nfft\cdot lognfft)+O(\frac{2\cdot S}{W}\cdot nfft)+O(\frac{2\cdot S}{W}\cdot nfft). \tag{5.26}$$

Iz izraza se može izostaviti konstanta 2 jer ne utječe na složenost. S obzirom na to da su 3. i 4. član izraza (5.25) sadržani u 2. članu, izraz za ukupnu računsku složenost može se zapisati kao:

$$O_{spektrogram}(ukupno) = O(S)+\underbrace{O(\frac{2\cdot S}{W}\cdot nfft\cdot lognfft)}_{dominatno}. \tag{5.27}$$

Najdominantniji član računske složenosti je generiranje spektrograma odnosno 2. član u izrazu (5.27). Obrada cijele .bin datoteke tj. ukupan broj R generiranih spektrograma ima računsku složenost $O\left(R\cdot\left(S+\frac{S}{W}\cdot \mathrm{nfft}\cdot\log \mathrm{nfft}\right)\right)$.

Primjer izračuna računske složenosti

Parametri koji su korišteni za izračun spektrograma su sljedeći:

- frekvencija uzorkovanja $f_s = 5MHz$,
- trajanje segmenta $T_s = 2ms$,
- broj uzoraka po segmentu $D = f_s\cdot T_s = 5MHz\cdot 2ms = 10000$,
- broj segmenata po spektrogramu $P = 1000$,
- širina prozora $W = 1024$,
- preklapanje $H=\frac{W}{2}$,
- broj točaka po prozoru $nfft = 2\cdot W$,
- ukupan broj uzoraka po spektrogramu $S = D\cdot P = 10^6$.

Kada se u izraz (5.27) uvrste korišteni parametri, dobije se $=(10^6+\frac{210^6}{1024}\cdot 2048\cdot log2048) = 45\cdot 10^6$ operacija po spektrogramu odnosno $81\cdot 10^9$ za svih 1800 spektrograma.

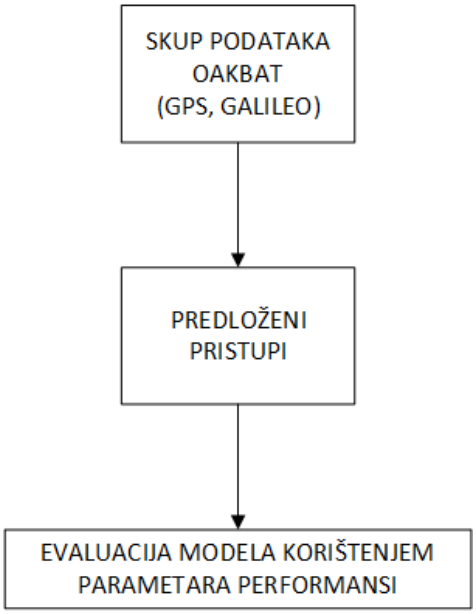
Ako se usporede korišteni algoritmi radio frekvencijskog otiska, može se zaključiti da je spektrogram računski složeniji od diskretne valićne transformacije s jednom razinom dekompozicije zbog višestrukih izračuna brze Fourierove transformacije. Spektrogram može biti brži od višerazinske diskretne valićne transformacije ako se koristi mali broj *nfft* točaka i manja veličina prozora.

6. Evaluacija metoda detekcije lažiranih signala i klasifikacije tipa signala primjenom integriranog pristupa

U ovom poglavlju prikazani su rezultati detekcije lažiranih signala i klasifikacije tipa signala na temelju predloženog integriranog pristupa koji uključuje dvije metode radio frekvencijskog otiska i nekoliko modela strojnog učenja. Prikazani su rezultati klasifikacije za različite vrste ulaznih podataka.

6.1. Model za evaluaciju predloženih metoda detekcije lažiranih signala u sustavu GNSS

Model za evaluaciju predloženih metoda za detekciju lažnih signala u sustavu GNSS prikazan je na slici 6.1.



Slika 6.1: Blok dijagram korištenog evaluacijskog modela.

Blok dijagram evaluacijskog modela za predložene pristupe prikazan je na slici 6.1. Na korištene skupove podataka OAKBAT primijenjena su dva pristupa: prvotno predloženi pris-

tup koji za klasifikaciju koristi slike kao ulazne podatke i računski učinkovitiji pristup koji koristi značajke slike i statističke i spektralne značajke. Predloženi pristupi evaluirani su korištenjem standardnih parametara za evaluaciju i performanse modela. Navedeni parametri performansi modela prikazani su u poglavlju 6.4.

6.2. Vrste korištenih klasifikacijskih podataka

U istraživanju su korištene tri vrste ulaznih klasifikacijskih podataka: slike, značajke slike te statističke i spektralne značajke. Navedene značajke izdvojene su za sve koeficijente diskretne valićne transformacije (aproksimacijski A, horizontalni H, dijagonalni D, i vertikalni V) za sve korištene skupove podataka.

Slike koje su korištene kao klasifikacijski podaci su u formatu .tiff. Veličina slike ovisi o frekvenciji uzorkovanja, trajanju jednog segmenta i broju segmenata po iteraciji. Primjerice, ako je frekvencija uzorkovanja 5 MHz, trajanje jednog segmenta 4 ms i broj segmenata po iteraciji 100, broj uzoraka po segmentu je 20000, a ukupan broj uzoraka po iteraciji je 100 x 20000 = 2 000 000 uzoraka. Prema tome, rezolucija slike je 1414 x 1414 piksela.

Statističke značajke su sljedeće [112], [113]:

1. Srednja vrijednost (μ) predstavlja aritmetičku sredinu svih vrijednosti u skupu podataka (u ovom istraživanju to su aproksimacijski koeficijenti). Računa se kao:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{6.1}$$

   gdje je N broj aproksimacijskih koeficijenata, x_i je svaki pojedinačni koeficijent.

2. Varijanca (σ^2) prikazuje prosječnu kvadratnu udaljenost vrijednosti od srednje vrijednosti. Daje uvid u raspon i raznolikost podataka te je ključna za procjenu šuma ili varijabilnosti u signalu. Varijanca se računa prema izrazu:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2. \tag{6.2}$$

3. Standardna devijacija (σ) mjeri raspršenost podataka oko srednje vrijednosti. Što je standardna devijacija veća, to su podaci udaljeniji od srednje vrijednosti. Izračunava se kao kvadratni korijen varijance prema izrazu:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2} \tag{6.3}$$

4. Maksimalna vrijednost predstavlja najveći koeficijent u skupu podataka i koristi se za detekciju vrhova u signalu koji mogu ukazivati na nagle promjene. Definirana je izrazom:

$$x_{\max} = \max(x_1, x_2, \ldots, x_N). \tag{6.4}$$

5. Minimalna vrijednost predstavlja najmanji koeficijent u skupu podataka. Daje informaciju o donjoj granici signala i može ukazivati na anomalije ili specifične obrasce. Definirana je izrazom:

$$x_{\min} = \min(x_1, x_2, \ldots, x_N). \tag{6.5}$$

6. Koeficijent spljoštenosti (engl. *kurtosis*) predstavlja stupanj zaobljenosti određene distribucije. Visoka vrijednost znači da je većina vrijednosti koncentrirana oko sredine, s prisutnošću odstupanja, dok niska vrijednost ukazuje na ravnomjerniju raspodjelu. Koeficijent spljoštenosti izračunava se prema izrazu:

$$\text{kurtosis} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^4. \tag{6.6}$$

7. Asimetrija (engl. *skewness*) je mjera asimetrije ili distorzije simetrične distribucije. Ukoliko je asimetrija pozitivna, rep raspodjele je dulji s desne strane (asimetrija s desne strane), dok negativna asimetrija ukazuje na rep s lijeve strane (asimetrija s lijeve strane). Računa se kao:

$$\text{skewness} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \mu}{\sigma}\right)^3. \tag{6.7}$$

Sve ove značajke računaju se za koeficijente aproksimacije jer sadrže većinu energije signala i predstavljaju njegovu najvažniju strukturnu informaciju na nižoj frekvenciji, što ih čini korisnima za klasifikaciju i analizu.

Izvučene spektralne značajke su [114], [115]:

1. Centroid spektra *SC* (engl. *Spectral Centroid*) označava dominantnu frekvenciju, odnosno mjesto gdje se nalazi centar spektra. Može se smatrati medijanom spektra, a izračunava se prema izrazu:

$$SC = \frac{\sum_{k=0}^{N-1} f_k \cdot |X(k)|}{\sum_{k=0}^{N-1} |X(k)|}, \tag{6.8}$$

gdje je f_k frekvencija $k-tog$ segmenta, $X(k)$ je amplituda (ili magnituda) Fourierova spektra na frekvenciji f_k i N broj frekvencijskih komponenti u spektru.

2. Raspršenost spektra *SS* (engl. *Spectral Spread*) je drugi centralni moment spektra. Predstavlja standardnu devijaciju oko spektralnog centroida *SC* koji je definiran izrazom (6.8) i računa se prema izrazu:

$$SS = \sqrt{\frac{\sum_{k=0}^{N-1}(f_k - SC)^2 \cdot |X(k)|}{\sum_{k=0}^{N-1}|X(k)|.}} \tag{6.9}$$

3. Entropija spektra *SE* (engl. *Spectral Entropy*) je mjera neuređenosti ili nasumičnosti u raspodjeli spektralne snage signala. Izračunava se pomoću Shannonove entropije:

$$SE = -\sum_{k=0}^{N-1} P(k) \cdot \log_2(P(k)), \tag{6.10}$$

gdje je $P(k)$ normalizirana magnituda spektra u frekvencijskom segmentu k.

Značajke izvučene iz slika su značajke obrade slike i analize teksture. U ovom istraživanju, ove značajke su izdvojene za tri kanala (crveni, zeleni i plavi) jer se za analizu koriste RGB slike.

Značajke obrade slike i analize teksture su [116], [118], [117], [119], [120], [121], [138]:

- Srednja vrijednost histograma (μ_H) je srednja vrijednost svih intenziteta piksela unutar određenog kanala boje (npr. crveni, zeleni, plavi). Naziva se srednja vrijednost histograma jer se srednja vrijednost računa na temelju histograma kao distribucije. Ona pruža informaciju o ukupnoj svjetlini slike - svjetlije slike imaju veće prosječne vrijednosti, dok tamnije slike imaju manje prosječne vrijednosti. Računa se prema izrazu:

$$\mu_H = \frac{1}{N}\sum_{i=0}^{N-1} I(i), \tag{6.11}$$

gdje je $I(i)$ intenzitet za $i-ti$ piksel, a N je ukupan broj piksela.

- Standardna devijacija histograma (σ_H) mjeri disperziju vrijednosti intenziteta piksela u odnosu na srednju vrijednost. Velika standardna devijacija ukazuje na slike s visokim kontrastom ili velikim razlikama u svjetlini između piksela, dok mala standardna devijacija označava ravnomjerniju raspodjelu svjetline. Računa se prema izrazu:

$$\sigma_H = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(I(i) - \mu_H)^2}. \tag{6.12}$$

μ_H predstavlja srednju vrijednost histograma koja je definirana izrazom (6.11).

- Srednja veličina gradijenta (G_μ) predstavlja prosječnu snagu rubova slike, tj. koliko brzo se intenzitet mijenja između susjednih piksela. Veće vrijednosti označavaju slike s izraženim rubovima i detaljima. Izraz za srednju veličinu gradijenta je:

$$G_\mu = \frac{1}{N}\sum_{i=0}^{N-1}\sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}, \tag{6.13}$$

gdje su $\frac{\partial I}{\partial x}$ i $\frac{\partial I}{\partial y}$ su gradijenti intenziteta piksela duž x i y osi. Gradijent duž osi x daje informaciju o tome koliko se intenzitet boje mijenja u vodoravnim linijama dok gradijent duž osi y daje informaciju o promjeni intenziteta boje u vertikalnim linijama. Računaju se pomoću metode Sobel operatora (metoda za detekciju rubova na slici koja aproksimira parcijalne derivacije slike po vodoravnom i okomitom smjeru kako bi se izračunala promjena intenziteta piksela).

- Standardna devijacija gradijenta (G_σ) pokazuje varijacije u čvrstoći rubova. Ako postoji velika varijacija, to znači da slika ima različite teksture, od glatkih područja do područja s izraženim detaljima. Računa se kao:

$$G_\sigma = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}\big(G(i) - G_\mu\big)^2}, \tag{6.14}$$

gdje je $G(i)$ veličina gradijenta za $i-ti$ piksel i G_μ srednja veličina gradijenta definirana izrazom (6.13).

Sljedeće značajke su teksturne značajke matrice supojavljivanja razina sive boje GLCM. GLCM razmatra odnose između susjednih piksela, pružajući bogate informacije o teksturi. Dakle, iz svake RGB komponente slike dobivene diskretnom valićnom transformacijom izvlače se teksturne značajke korištenjem metode GLCM.

- Kontrast (C) mjeri razliku između intenziteta susjednih piksela. Visoki kontrast označava grublje teksture, dok nizak kontrast označava glatke ili ujednačenije teksture. Računa se prema izrazu:

$$C = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-j)^2 \cdot P(i,j), \tag{6.15}$$

gdje je $P(i,j)$ element normalizirane matrice GLCM tj. vjerojatnost da piksel sive razine i ima susjedni piksel sive razine j i N je broj sivih razina na slici.

- Korelacija (ρ) označava stupanj povezanosti intenziteta između piksela na određenoj udaljenosti. Visoka korelacija znači da na slici postoji jasno definiran uzorak ili struktura. Izraz prema kojem se računa korelacija je:

$$\rho = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{(i-\mu_i)\cdot(j-\mu_j)\cdot P(i,j)}{\sigma_i \cdot \sigma_j}, \tag{6.16}$$

gdje su μ_i, μ_j i σ_i, σ_j srednje vrijednosti i standardne devijacije za retke i stupce i i j.

- Energija E ili kutni drugi moment *ACM* (engl. *Angular Second Moment*) predstavlja ujednačenost teksture. Više vrijednosti energije pokazuju da slika ima pravilne ili ponavljajuće uzorke, dok niže vrijednosti ukazuju na slučajniji raspored piksela. Energija se dobije prema izrazu:

$$E = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)^2. \tag{6.17}$$

- Homogenost (H) mjeri sličnost intenziteta između susjednih piksela. Visoka homogenost ukazuje na glatke teksture bez velikih razlika u intenzitetu. Računa se kao:

$$H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{P(i,j)}{1+|i-j|}. \tag{6.18}$$

Sve ove značajke su nakon izvlačenja sa slika spremljene u .csv datoteke.

6.3. Primjena skupa podataka OAKBAT za evaluaciju metoda detekcije lažiranih signala sustava GNSS

Skupovi digitaliziranih radio signala OAKBAT razvijeni su za istraživanje napada lažiranjem u području GNSS te su komplementarni skupovima podataka TEXBAT. Za razliku od skupova podataka TEXBAT koji se temelje samo na signalima sustava GPS, OAKBAT dodatno sadrži i Galileo signale što je i razlog korištenja ovoga skupa podataka u ovom istraživanju. OAKBAT se sastoji od 8 skupova koji sadrže samo GPS L1 C/A signale i 8 skupova podataka koji sadrže samo Galileo E1 signale. Kreiran je pomoću komercijalnog GNSS simulatora Orolia Vulnerability Test System (VTS), koji se sastoji od dva Orolia GSG-6 serije višefrekventnih i višekonstelacijskih GNSS simulatora. Parametri korišteni za kreiranje OAKBAT skupa podataka su sljedeći:

- okolina 'otvoreno nebo',
- datum i vrijeme početka svih scenarija - March 19, 2020, 09:59:42 UTC,
- snage za GPS i Galileo satelite: 123,5 dBm and -125 dBm.
- frekvencija uzorkovanja 5 MHz.

Skupovi podataka korišteni u ovom istraživanju prikazani su u tablici 6.1:

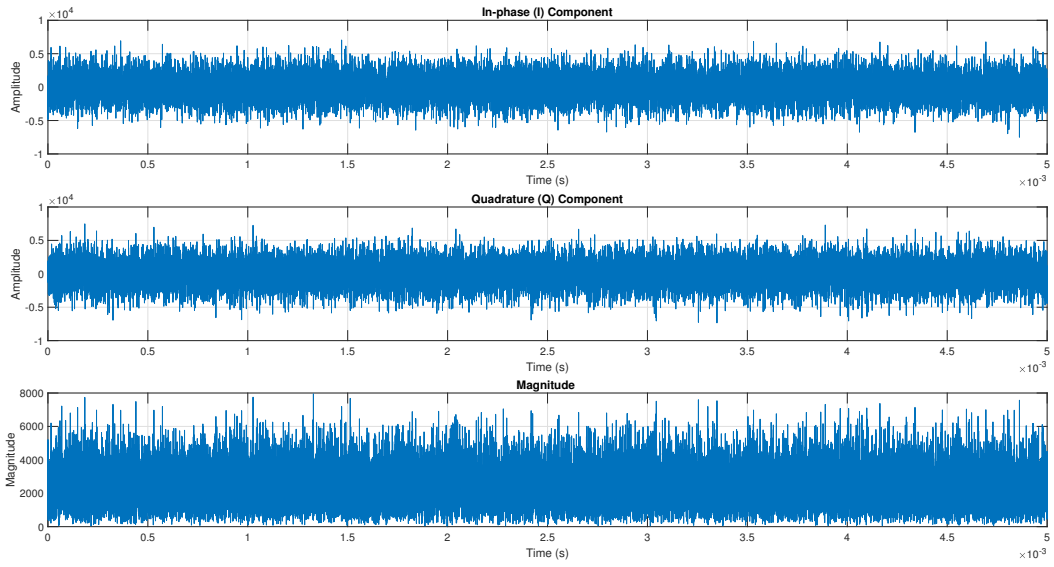
Skupovi podataka odabrani za analizu u ovom istraživanju su statički autentični i lažni scenariji u sustavima GPS i Galileo. U statičkim scenarijima prijamnik se nalazi na poznatoj, fiksnoj lokaciji odnosno u stanju mirovanja. Lažni scenariji sustava GPS su os2, os3 i os4

Tablica 6.1: Korišteni skupovi podataka OAKBAT [22]

Skup podataka	Tip	Prednost razine snage [dB]	Konstelacija	Scenarij
cleanStatic	statički	-	GPS	autentični
cleanStatic	statički	-	Galileo	autentični
os2	statički	+ 10	GPS	lažni
os3	statički	+ 1.3	GPS	lažni
os4	statički	+ 0.4	GPS	lažni
os10	statički	+ 10	Galileo	lažni
os11	statički	+ 1.3	Galileo	lažni
os12	statički	+ 0.4	Galileo	lažni

i razlikuju se po prednosti razine snage lažnih signala od +10, +1.3, +0.4 dB u odnosu na autentične signale. Lažni scenariji sustava Galileo imaju sve iste karakteristike kao scenariji sustava GPS.

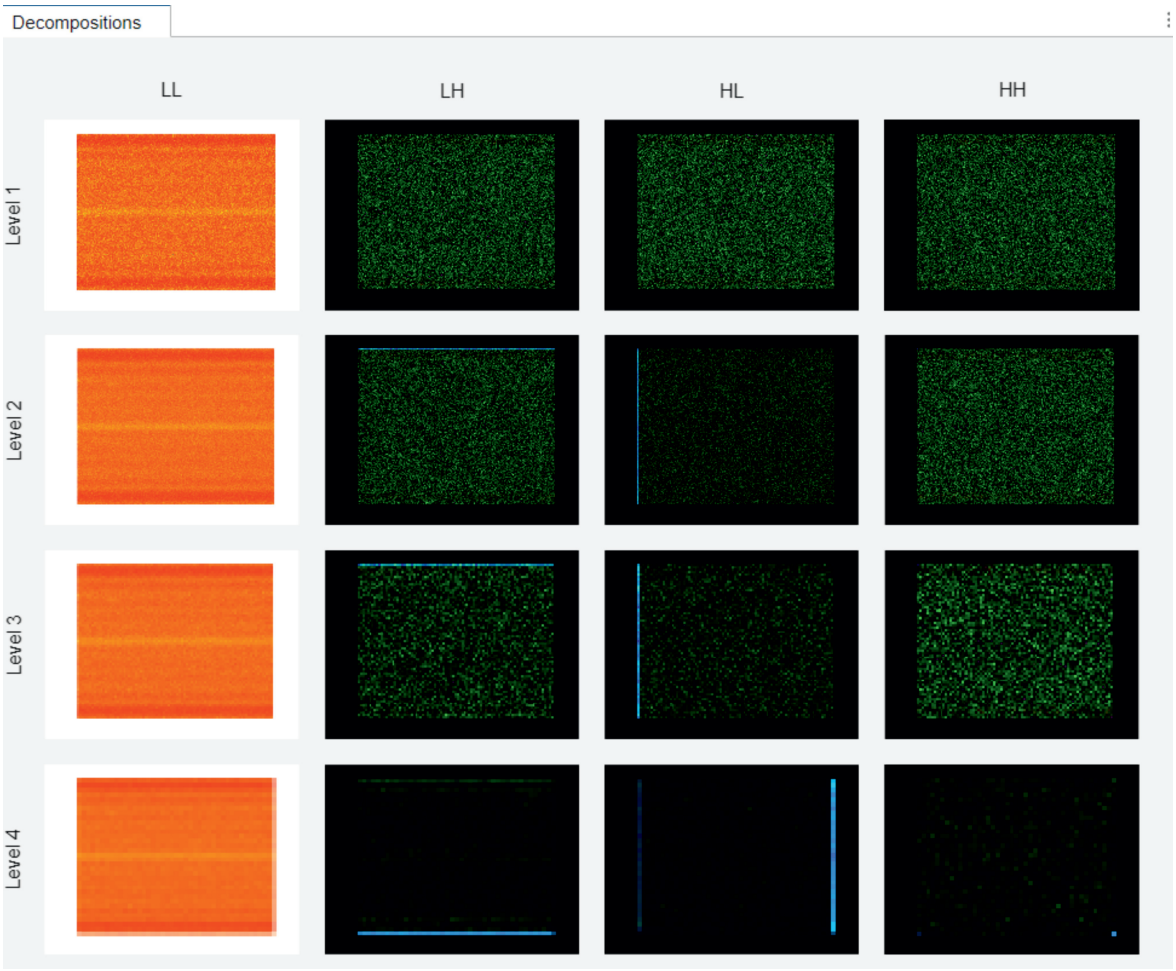
S obzirom na velike procesorske zahtjeve i računsku složenost RFF metoda te nemogućnost obrade podataka odjednom, uvedena je obrada nekoliko jednakih dijelova (segmenata) signala paralelno. Umjesto obrade jednog po jednog dijela signala, više dijelova signala obrađeno je u svakoj iteraciji što omogućuje kompromis između upotrebe memorije i brzine računanja. Optimalno trajanje dijela signala za spektrogram je 2 ms. Broj segmenata signala po iteraciji je 100 što znači da je 200 ms podataka obrađeno u jednoj iteraciji. Nakon obrade ovih 100 segmenata signala, slijedi nova iteracija u kojoj se novi skup segmenata obrađuje do kraja datoteke u kojoj su pohranjeni podaci. U slučaju diskretne valične transformacije, trajanje segmenta signala postavljeno je na 4 ms zbog ograničenih računalnih resursa (manji broj obrađenih uzoraka po segmentu signala).



Slika 6.2: Prikaz I/Q komponenti signala sustava GPS u skupu podataka os2 tijekom prvih 5 ms.

Slika 6.2 prikazuje I/Q komponente u skupu podataka os2 u prvih 5 ms. Gornji graf

prikazuje kofazne I komponente, a kvadraturne Q komponente su prikazane na srednjem grafu. Q komponente su ortogonalne prema I komponentama (90 stupnjeva) i pružaju dodatne informacije za obradu signala. Donji graf na slici predstavlja ukupnu snagu izračunatu iz I/Q komponenti. Može se vidjeti da magnituda varira s nekim vrhovima koji ukazuju na varijacije u snazi signala.



Slika 6.3: Primjer dekompozicije slike korištenjem diskretne valićne transformacije.

Primjer dekompozicije slike korištenjem diskretne valićne transformacije prikazan je na slici 6.3. Tip valića korišten u analizi je Daubechies db4 s razinom dekompozicije 1. Lijevi stupac na slici predstavlja koeficijente aproksimacije, sljedeći stupac prikazuje horizontalne koeficijente, zatim vertikalne koeficijente, a posljednji stupac predstavlja dijagonalne koeficijente. Izvorna slika može se rekonstruirati iz razdvojenih razina. Može se vidjeti da je razina 1 najdetaljnija i najreprezentativnija za dekompoziciju i rekonstrukciju slike. S druge strane, razina 4 nema puno detalja i stoga nema veliki utjecaj na rekonstrukciju kao niti na analizu podataka. To je razlog zašto je samo razina 1 uključena u analizu u ovom istraživanju.

6.4. Parametri performansi modela

Klasifikacijski modeli imaju diskretan izlaz te su potrebni parametri koji uspoređuju diskretne klase u nekom obliku. Stoga, za evaluaciju modela, koriste se parametri koji su opisani ispod.

1. Konfuzijska matrica ili matrica zabune (engl. *confusion matrix*) poznata i kao matrica pogreške, posebna je vrsta tablice koja omogućuje vizualizaciju performansi algoritma. Svaki redak u matrici predstavlja uzorke u stvarnoj klasi (engl. *true*), a svaki stupac predstavlja uzorke u predviđenoj (engl. *predicted*) klasi. Konfuzijska matrica sastoji se od četiri polja [82]:

   - stvarno pozitivni (engl. *true positive* - T_p) - predikcija je 1 i stvarna oznaka je 1,
   - stvarno negativni (engl. *true negative* - T_n) - predikcija je 0 i stvarna oznaka je 0,
   - lažno pozitivni (engl. *false positive* - F_p) - predikcija je 1 a stvarna oznaka je 0,
   - lažno negativni (engl. *false negative* - F_n) - predikcija je 0 a stvarna oznaka je 1.

   Primjer 2x2 konfuzijske matrice za binarnu klasifikaciju je [82]:

   $$\begin{array}{c|cc} & y_{predicted}=1 & y_{predicted}=0 \\ \hline y_{true}=1 & T_p & F_p \\ y_{true}=0 & F_n & T_n \end{array}. \tag{6.19}$$

2. Točnost (engl. *accuracy*) se definira kao broj točno predviđenih uzoraka podijeljen s ukupnim brojem predviđenih uzoraka i računa se kao:

   $$ACC = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \times 100, \tag{6.20}$$

   gdje T_p označava broj stvarno pozitivnih predviđenih uzoraka, F_n broj lažno negativnih predviđenih uzoraka, F_p definira broj lažno pozitivnih predviđenih uzoraka i T_n definira broj stvarno negativnih predviđenih uzoraka.

3. Preciznost (engl. *precision*) ili pozitivna prediktivna vrijednost predstavlja broj stvarno pozitivnih predviđenih uzoraka podijeljen s ukupnim brojem pozitivnih predviđenih uzoraka ($T_p + F_p$). Vrijednost ovog parametra je između 0 i 1 i računa se kao:

   $$P = \frac{T_p}{T_p + F_p}, \quad 0 < P < 1. \tag{6.21}$$

   Preciznost odgovara na pitanje: "Od svih slučajeva koji su predviđeni kao pozitivni, koliko ih je bilo pozitivnih?" Što je precizost veća, manji je broj lažno pozitivnih uzoraka. Niska preciznost ($P < 0.5$) znači da klasifikator ima veliki broj lažno pozitivnih

uzoraka koji mogu biti rezultat neuravnotežene klase ili nepodešenih hiperparametara modela. Što je vrijednost preciznosti bliža 1, znači da model nije propustio nijedan pravi pozitivan rezultat i da je u stanju dobro klasificirati ispravno i netočno označavanje uzoraka.

4. Odziv (engl. *recall*) ili osjetljivost ili stopa stvarnih pozitiva daje postotak pozitivnih rezultata dobro predviđenih određenim modelom. Drugim riječima, daje udio stvarnih pozitivnih uzoraka koje je model ispravno klasficirao. Ova se mjera naziva odziv jer nam govori koliko se pozitivnih uzoraka odazvalo klasifikatoru. Odgovara na pitanje: "Od svih stvarnih pozitivnih slučajeva, koliko ih je model ispravno predvidio?" Odziv se računa prema izrazu:

$$R = TPR = \frac{T_p}{T_p + F_n}, \quad 0 < R < 1. \tag{6.22}$$

   Nizak odziv ($R < 0.5$) znači da klasifikator ima veliki broj lažno negativnih uzoraka koji mogu biti rezultat neuravnotežene klase ili nepodešenih hiperparametara modela. U idealnom slučaju, $R = 1$, tj. sve pozitivne uzorke klasifikator označava kao takve. Idealno je kada su svi pozitivno klasificirani uzorci stvarno pozitivni ($P = 1$) i, obrnuto, svi pozitivni uzorci su također klasificirani kao pozitivni ($R = 1$).

5. Ispadanje (engl. *fall-out*) ili stopa lažnih pozitiva FPR je udio lažno pozitivnih uzoraka F_p u skupu svih negativnih uzoraka $F_P + T_n$. U idealnom slučaju, $FPR = 0$, tj. klasifikator niti jedan negativan uzorak neće lažno proglasiti pozitivnim. Definiran je izrazom:

$$FPR = \frac{F_p}{F_p + T_n}. \tag{6.23}$$

6. Specifičnost (engl. *specificity*) ili stopa stvarno negativnih predstavlja udio stvarno negativnih uzoraka T_n (TN) u skupu svih negativnih uzoraka $T_n + F_p$ i prikazuje se kao:

$$TNR = \frac{T_n}{T_n + F_p}. \tag{6.24}$$

7. Stopa lažno negativnih uzoraka (engl. *False Negative Rate*) predstavlja udio pozitivnih uzoraka F_n koje je model pogrešno prepoznao kao negativne i definira se izrazom:

$$FNR = \frac{F_n}{T_p + F_n}. \tag{6.25}$$

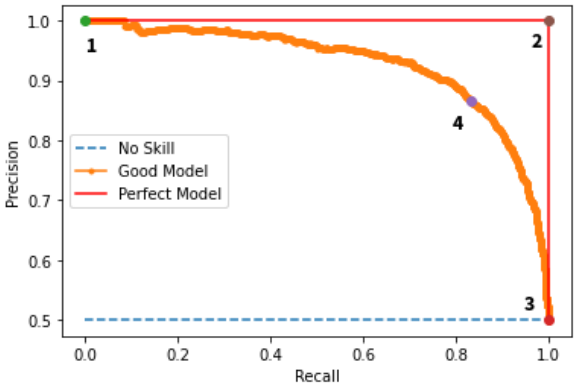
8. Mjera *F1* (engl. *F1-score*) ključni je parametar u evaluaciji performansi modela. *F1* mjeri performanse modela balansiranjem preciznosti i odziva koje su dvije izravno

suprotstavljene mjere (visok odziv znači nisku preciznost i obrnuto), pružajući jedinstvenu ocjenu koja odražava i lažno pozitivne i lažno negativne rezultate. Računa se kao harmonijska sredina preciznosti i odziva. Harmonijska sredina (recipročna vrijednost aritmetičke sredine recipročnih vrijednosti) se koristi kako bi se odziv i preciznost mogli kombinirati na istoj skali. Mjera *F1* definirana je izrazom:

$$F1 = \frac{2}{\frac{1}{P}+\frac{1}{R}} = \frac{2T_p}{2T_p+F_p+F_n}, \quad 0 < F1 < 1. \tag{6.26}$$

$F1$ postiže najbolju vrijednost na 1, a najlošiju na 0. Primjerice, ako je $P = 0.1$ i $R = 0.8$ mjera $F1$ iznosi 0.178, dok bi aritmetička sredina iznosila 0.4 i zbog toga treba uzimati stroži kriterij kod vrednovanja klasifikatora.

9. Krivulja preciznosti i odziva (engl. *PR curve*) je grafički prikaz preciznosti i odziva klasifikatora na različitim pragovima klasifikacije. Ona pokazuje koliko je dobro predviđena manjinska klasa odnosno koliko točno su napravljena pozitivna predviđanja i detektirani stvarni pozitivni rezultati. Ova krivulja je važan alat za procjenu performansi modela u neuravnoteženim skupovima podataka. Pomaže u odabiru optimalnog praga koji učinkovito uravnotežuje preciznost i odziv. Os x na PR krivulji predstavlja odziv dok os y predstavlja preciznost kao što je prikazano na slici 6.4.

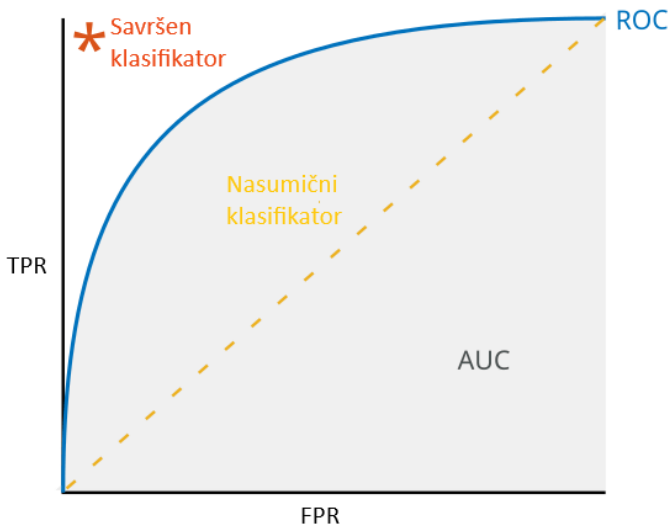


Slika 6.4: Primjer PR krivulje [110].

Što se ova krivulja više približava gornjem desnom kutu grafa, to je model sposobniji u postizanju visoke preciznosti i odziva istovremeno, što ukazuje na robusnu izvedbu u razlikovanju između klasa, bez obzira na njihovu učestalost u skupu podataka. Drugim riječima, visoko područje ispod krivulje predstavlja i visok odziv i visoku preciznost. Visoka preciznost postiže se malim brojem lažno pozitivnih uzoraka, a visok odziv postiže se malim brojem lažno negativnih uzoraka. Model koji ima visok odziv i nisku preciznost vraća većinu relevantnih uzoraka, ali udio netočno klasificiranih uzoraka je visok. S druge strane, model koji ima visoku preciznost i nizak odziv vraća vrlo malo relevantnih uzoraka, ali većina njegovih predviđenih uzoraka točna je u usporedbi sa

stvarnim klasama. Idealan model koji ima visoku i preciznost i odziv vratit će većinu relevantnih uzoraka, s većinom ispravno označenim klasama.

10. Krivulja operativnih karakteristika ROC (engl. *Receiver Operating Characteristic*) predstavlja vizualni prikaz performansi modela preko svih pragova. Os x na krivulji prikazuje lažnu pozitivnu stopu FPR dok je na osi y prikazana stvarna pozitivna stopa TPR. Klasifikacijski model je bolji što je krivulja ROC viša kao što je prikazano na slici 6.5. Stoga, što je veća površina ispod krivulje, to je klasifikator bolji. Površina ispod krivulje predstavlja vrijednost AUC (engl. *Area Under Curve*). Vrijednost ROC AUC iznosi 0-1.



Slika 6.5: Primjer ROC krivulje [111].

ROC AUC vrijednost kreće se od 0 do 1. Vrijednost 0.5 označava nasumično pogađanje (engl. *random guessing*) tj. nasumični klasifikator što je označeno žutom bojom na slici. Vrijednost 1 označava savršenu klasifikaciju. Savršen klasifikator ima TPR jednak 1 i FPR jednak 0 za sve vrijednosti praga. Nasumični klasifikator ima iste vrijednosti za FPR i TPR za sve vrijednosti praga. Većina klasifikatora upada u vrijednosti od 0.5 do 1 te su rijetki izuzeci za koje je $AUC < 0.5$.

Za sve korištene modele strojnog učenja, primijenjena je unakrsna provjera valjanosti k-podskupova kako bi se pouzdanije procijenila izvedba modela. U unakrsnoj provjeri valjanosti k-podskupova, ulazni podaci dijele se na k podskupova. Model se trenira na $k-1$ podskupova dok se preostali podskup koristi za testiranje. Ovaj proces ponavlja se k puta, osiguravajući da svaki podskup bude korišten samo jednom kao podskup za testiranje. Dodatno, interval pouzdanosti (engl. *Confidence Interval CI*) od 95% za sve modele izračunava se korištenjem prikupljenih parametara. Koraci za izračun intervala pouzdanosti su:

1. Izračun srednje vrijednosti dobivenih točnosti definiran je izrazom:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} ACC_i, \tag{6.27}$$

gdje je ACC_i točnost za iteraciju i i N je broj iteracija.

2. Izračun standardne devijacije:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(ACC_i - \mu)^2}. \tag{6.28}$$

3. Procjena intervala pouzdanosti

$$\mathrm{CI} = \mu \pm Z \cdot \frac{\sigma}{\sqrt{N}}, \tag{6.29}$$

gdje je Z rezultat za odabranu razinu pouzdanosti (npr. $Z = 1.96$ za 95%).

6.5. Rezultati evaluacije metode detekcije temeljene na primjeni diskretne valićne transformacije

Za otkrivanje lažnih signala i klasifikaciju tipa signala, valići Daubechies db4 i db8 integrirani su s metodama strojnog učenja. Ovi tipovi valića često se koriste za obradu signala u sutavu GNSS zbog svojih matematičkih svojstava. Ovi valići omogućuju učinkovitu analizu promjena signala u vremenu i frekvenciji, što je ključno za otkrivanje anomalija koje se javljaju tijekom napada lažiranjem. Nadalje, valići db4 i db8 imaju sposobnost otkrivanja suptilnih, ali značajnih razlika između autentičnih i lažnih signala te nude dobar kompromis između vremenske i frekvencijske rezolucije.

Što se tiče parametara koji se koriste u modelima strojnog učenja, Gaussova jezgra se koristi u modelu SVM jer daje najbolje rezultate. U modelu KNN parametar broj susjeda postavljen je na 5, a korištena udaljenost je euklidska, jer te vrijednosti pokazuju najbolju izvedbu modela. Parametar broj stabala postavljen je na 100 i 200 kako bi se ispitale performanse modela kada se taj broj poveća. Također, ugrađena metoda provjere valjanosti, predviđanje (engl. *Out-Of-Bag - OOB*), koristi se u nasumičnim šumama za procjenu modela. Svako se stablo uvježbava na nasumičnim uzorcima, a OOB uzorci izostavljeni su i zatim korišteni za procjenu modela, dajući internu procjenu izvedbe bez potrebe za zasebnim skupom podataka za testiranje.

Parametri performansi modela za algoritme klasifikacije A1, A1M1 i A1M2 za različite modele strojnog učenja za valiće db4 i db8 sustava GPS prikazani su u tablici 6.2. Prikazane su dobivene klasifikacijske vrijednosti nekoliko parametara za metode SVM, KNN, RF-100

i RF-200 za aproksimacijske koeficijente u skupovima podataka os2 i os4. Parametri performansi koji su analizirani su: srednja vrijednost točnosti modela, interval pouzdanosti od 95%, mjera $F1$, odziv i preciznost. Nadalje, dodan je još i parametar koji pokazuje vrijeme potrebno za izvođenje pojedinog modela. Iz tablice možemo vidjeti da su primjerice srednje vrijednosti točnosti podjednake za oba korištena valića db4 i db8 jer oba valića pripadaju istoj obitelji valića Daubechies i imaju slične karakteristike s razlikom u stupnju složenosti. Što se tiče klasifikacije sa slikama kao ulaznim podacima, najveću točnost za obje vrste valića db4 i db8 ima model SVM u skupu podataka os2 i ona iznosi 98.94% odnosno 99.43%. Rezultati za značajke slika i spektralne i statističke su svi jednako dobri, a ističu se modeli RF-100 i RF-200 u statističkim s točnošću od 100% i vrijednostima $F1$, R i P koje iznose 1. Jedina razlika između ova dva modela je u vremenu izvođenja jer RF-200 ima više stabala u odnosu na RF-100 te je potrebno i više vremena za njegovo izvođenje. Vrijednosti R i P su sve veće od 0.7 što ukazuje na dobre klasifikacijske modele. Vrijeme izvođenja modela koji kao ulazni podatak imaju slike je dosta veće u odnosu na modele koji kao ulazni podatak uzimaju brojčane vrijednosti. Najveće vrijeme izvođenja je 1202 sekunde za model RF-200 u skupu podataka os2 jer je ovaj model najkompleksniji zbog broja stabala.

Slika 6.6 prikazuje krivulje operativnih karakteristika za algoritam klasifikacije A1 za različite modele strojnog učenja korištenjem valića db4 skupovima podataka GPS i Galileo. Os x predstavlja stopu lažno pozitivnih uzoraka dok je stopa stvarno pozitivnih uzoraka prikazana na osi y. Na slikama 6.6a and 6.6c, krivulje ROC imaju slično ponašanje za sve modele strojnog učenja (SVM, KNN, RF-100 i RF-200) u skupovima podataka os2 i os10 kod kojih je snaga lažnih signala za 10 dB veća u odnosu na autentične signale, što je slično ponašanju modela za skup podataka TEXBAT ds2 koji ima istu prednost razine snage od 10 db za lažne signale i koji je prikazan na slici 4.17a. Svi modeli imaju visoku točnost klasifikacije, a SVM se ističe kao najbolji u oba slučaja. Rezultati klasifikacije tipa signala za skup podataka GPS os2 su bolji u usporedbi sa skupom podataka Galileo os10. Slični rezultati dobiveni su i za skupove podataka os4 i os12 kao što je prikazano na slikama 6.6b and 6.6d. Iako se SVM ističe kao najbolji model s najvećom točnošću, krivulje ROC imaju značajna izobličenja zbog niže točnosti klasifikacije u tim skupovima podataka. Razlog tomu je mala razlika u snazi između autentičnih i lažnih signala zbog koje je teško razlikovati autentični od lažnog signala u skupovima podataka os4 i os12 za razliku od skupova podataka os2 i os10 koji imaju veću razliku u snazi između autentičnih i lažnih signala.

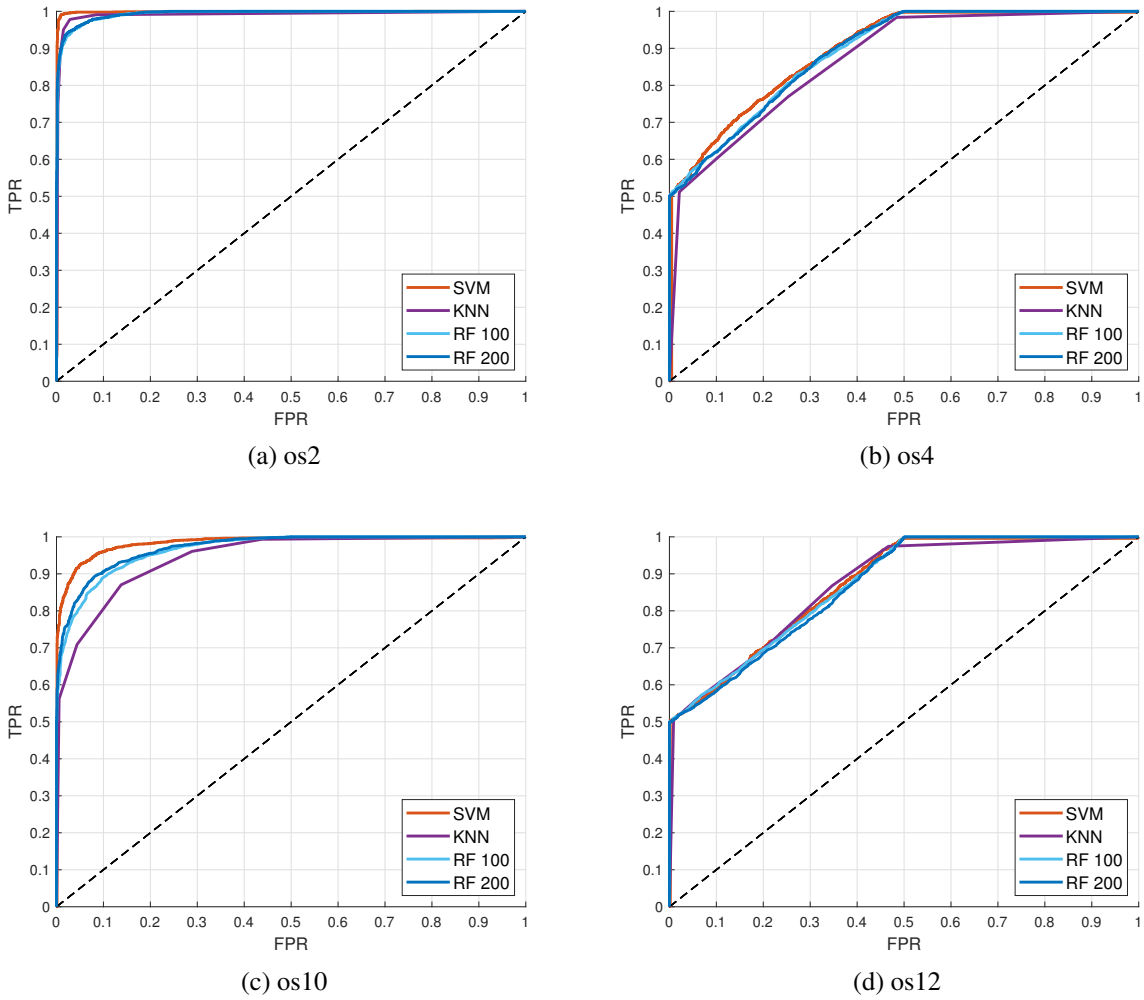
Dobiveni rezultati klasifikacije za ove skupove podataka očekivani su s obzirom na prednost razine snage od 10 dB i 0.4 dB za lažne signale u odnosu na autentične signale u skupovima podataka os2 i os10 odnosno os4 i os12.

Na slici 6.7, prikazana je distribucija mjere *F1* za različite skupove podataka sustava GPS i Galileo. Skupovi podataka razlikuju se po prednosti razine snage lažnih nad autentičnim signalima. Svaka boja predstavlja pojedini skup podataka kao što je prikazano u legendi. Isti valić db4 primijenjen je na sve skupove podataka i prikazani su rezultati za algoritam

Tablica 6.2: Parametri performansi modela za valiće db4 i db8 sustava GPS za algoritme klasifikacije: a) A1 (plavo), b) A1M1 (crveno), c) A1M2 (zeleno).

Skup podataka i vrsta koef.	ML model	Srednja točnost [%]	95% interval	Mjera *F1*	Odziv	Preciznost	Vrijeme izvođenja [s]
db4							
(clean + os2) approx.	SVM	98.94	[98.6885, 99.2004]	0.9895	0.9882	0.9907	606
	KNN	96.77	[96.5297, 97.0259]	0.9683	0.952	0.9852	589
	RF - 100	95.43	[95.047, 95.8049]	0.9537	0.9658	0.9419	594
	RF - 200	95.31	[94.7946, 95.835]	0.9524	0.9675	0.9378	1202
(clean + os4) approx.	SVM	78.20	[76.9726, 79.3978]	0.7829	0.7792	0.7867	764
	KNN	75.82	[75.0259, 76.6037]	0.7553	0.7642	0.7467	585
	RF - 100	77.33	[76.0106, 78.6561]	0.7742	0.7713	0.777	617
	RF - 200	77	[76.0039, 78.1073]	0.7711	0.7693	0.773	741
(clean + os2) approx.	SVM	99.77	[99.6352, 99.9203]	0.9978	0.9989	0.9967	9.4
	KNN	99.72	[99.5032, 99.9413]	0.9972	0.9956	0.9989	2.47
	RF - 100	99.75	[99.6306, 99.88]	0.9976	0.9985	0.9967	6.3
	RF - 200	99.78	[99.6689, 99.8867]	0.9978	0.9989	0.9967	10.28
(clean + os4) approx.	SVM	96	[95.2177, 96.7823]	0.9605	0.9493	0.9719	2.16
	KNN	91.94	[91.0319, 92.857]	0.9209	0.9043	0.9381	3.13
	RF - 100	93.93	[93.4139, 94.4739]	0.9398	0.931	0.9489	6.64
	RF - 200	94.41	[93.9267, 94.8881]	0.9445	0.9373	0.9519	12.4
(clean + os2) approx.	SVM	99.68	[99.5362, 99.8342]	0.9969	0.9967	0.997	6.8
	KNN	99.93	[99.8616, 99.9903]	0.9993	0.9993	0.9993	3.46
	RF - 100	100	[100, 100]	1	1	1	5.2
	RF - 200	100	[100, 100]	1	1	1	8
(clean + os4) approx.	SVM	99.72	[99.616, 99.8285]	0.9972	0.9956	0.9989	5.1
	KNN	99.93	[99.8425, 100]	0.9993	0.9985	1	2.9
	RF - 100	99.98	[99.9452, 100]	0.9998	1	0.9996	7.1
	RF - 200	100	[99.9135, 100]	0.9996	1	0.9993	7.9
db8							
(clean + os2) approx.	SVM	99.43	[99.2103, 99.6415]	0.9943	0.9944	0.9941	568
	KNN	96.02	[95.7493, 96.2877]	0.9614	0.934	0.9904	528
	RF - 100	95.66	[95.09, 96.2433]	0.9564	0.9621	0.9507	522
	RF - 200	95.88	[95.4406, 96.3372]	0.9586	0.9651	0.9522	531
(clean + os4) approx.	SVM	75.69	[74.6357, 76.7166]	0.7597	0.7508	0.7689	628
	KNN	75.18	[74.051, 76.3194]	0.7506	0.7545	0.7467	637
	RF - 100	75.53	[74.8393, 76.2348]	0.758	0.7499	0.7663	632
	RF - 200	75.6	[74.6291, 76.5561]	0.7615	0.7445	0.7793	654
(clean + os2) approx.	SVM	99.91	[99.8388, 99.976]	0.9991	0.9989	0.9993	3.3
	KNN	99.85	[99.7474, 99.9563]	0.9985	0.9981	0.9989	2.9
	RF - 100	99.87	[99.7804, 99.9603]	0.9987	0.9974	1	6.3
	RF - 200	99.83	[99.7169, 99.9497]	0.9983	0.9967	1	10
(clean + os4) approx.	SVM	96.02	[95.4397, 96.5973]	0.9606	0.9507	0.9707	3.5
	KNN	94.33	[93.6389, 95.0277]	0.9442	0.9306	0.9581	4
	RF - 100	93.76	[93.223, 94.2955]	0.9378	0.9342	0.9145	8.8
	RF - 200	94.13	[93.4557, 94.8035]	0.9418	0.9344	0.9493	15
(clean + os2) approx.	SVM	99.78	[99.699, 99.8566]	0.9978	0.9985	0.997	4.9
	KNN	99.98	[99.9452, 100]	0.9998	0.9996	1	2.8
	RF - 100	100	[100,100]	1	1	1	4.87
	RF - 200	100	[100,100]	1	1	1	7.8
(clean + os4) approx.	SVM	99.70	[99.5585, 99.8489]	0.997	0.9981	0.9959	5.3
	KNN	99.85	[99.7346, 99.9691]	0.9985	0.9996	0.9974	3.7
	RF - 100	99.98	[99.9452, 100]	0.9998	1	0.9996	5.87
	RF - 200	100	[100,100]	1	1	1	7.85

klasifikacije A1 za model SVM. Os x predstavlja vrijednosti mjere *F1*, koje su metričke vrijednosti za evaluaciju performansi modela (balans između odziva i preciznosti). Os y

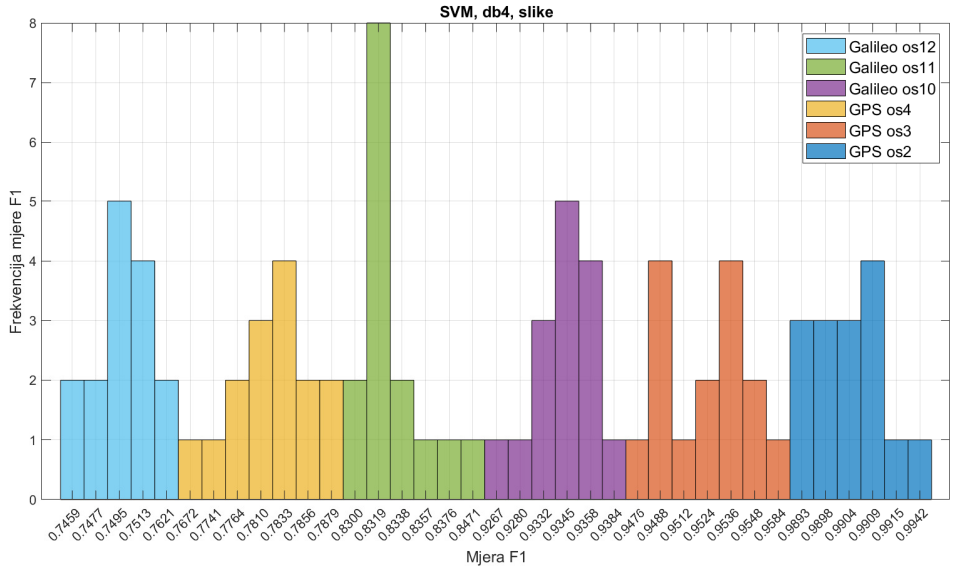


(a) os2 (b) os4

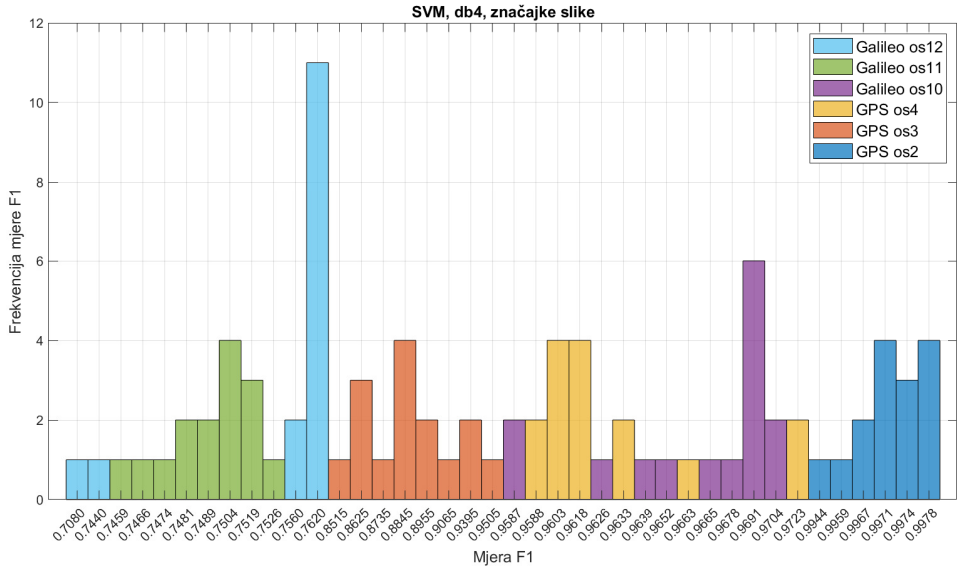
(c) os10 (d) os12

Slika 6.6: Krivulje operativnih karakteristika za algoritam klasifikacije A1 za različite modele strojnog učenja korištenjem valića db4 u skupovima podataka GPS (a) os2 i (b) os4 te Galileo (c) os10 i (d) os12.

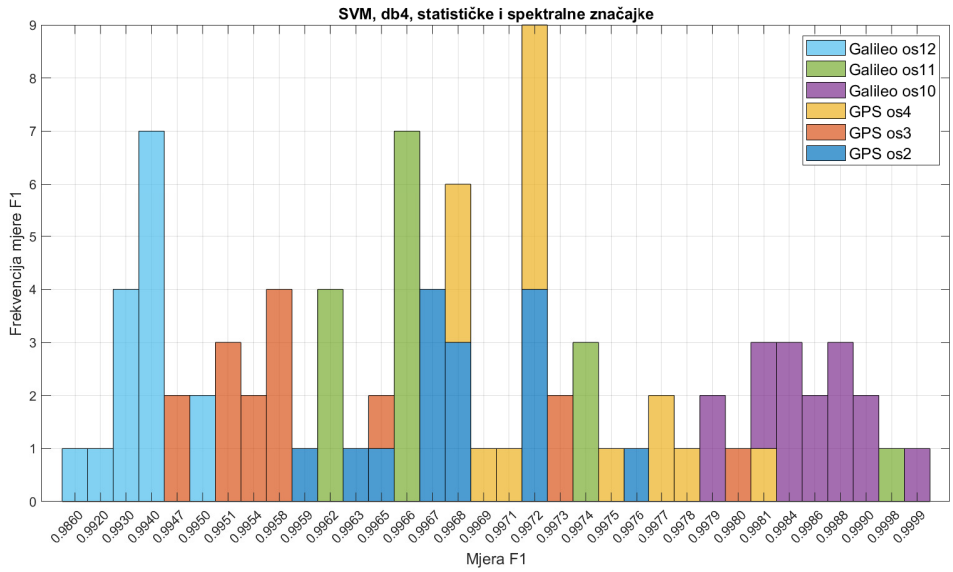
predstavlja broj pojavljivanja određene vrijednosti mjere *F1* unutar svakog skupa podataka. Kao što se može vidjeti na slici, vrijednost *F1* koja iznosi 0.9942 za skup podataka os2, pojavljuje se jednom u skupu podataka os2 te ovaj skup podataka postiže najveće vrijednosti mjere *F1*. Signali sustava Galileo isto pokazuju dobre vrijednosti mjere *F1* posebno za skup podataka os10 koji je ekvivalentan skupu os2. Skupovi podataka os4 i os12 imaju najmanje vrijednosti mjere *F1*. Ovi rezultati su očekivani s obzirom da lažni signali u skupovima os2 i os10 imaju veću razinu snage od autentičnih u odnosu na skupove podataka os3 i os4 te os11 i os12. Generalno, može se zaključiti da signali sustava GPS imaju veće vrijednosti mjere *F1* u odnosu na Galileo što ukazuje na veću preciznost klasifikacije kod signala sustava GPS.



Slika 6.7: Distribucija mjere F1 za skupove podataka sustava GPS os2, os3, os4 i Galileo os10, os11, os12 za algoritam klasifikacije A1 za model SVM.



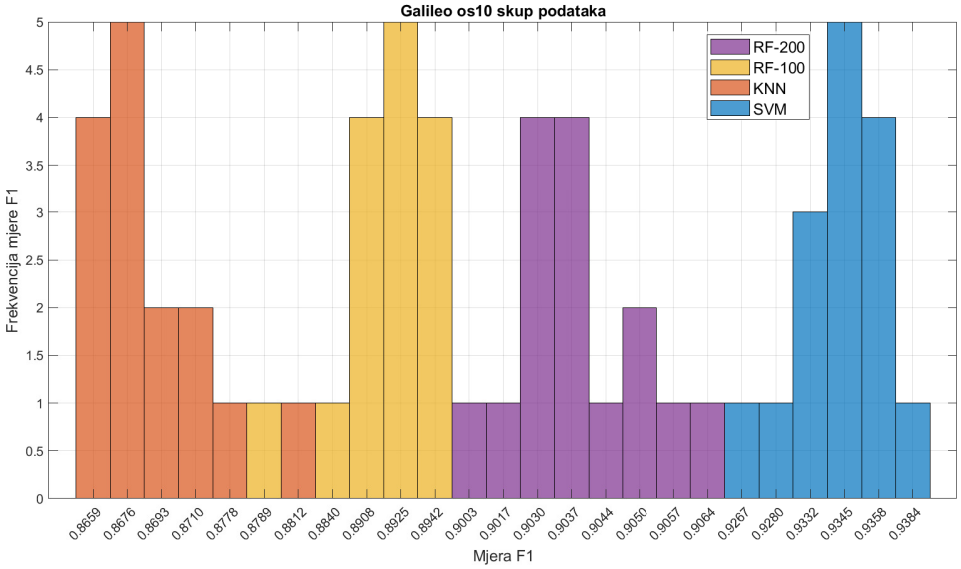
Slika 6.8: Distribucija mjere F1 za različite skupove podataka sustava GPS os2, os3, os4 i Galileo os10, os11, os12 za algoritam klasifikacije A1M1 za model SVM.



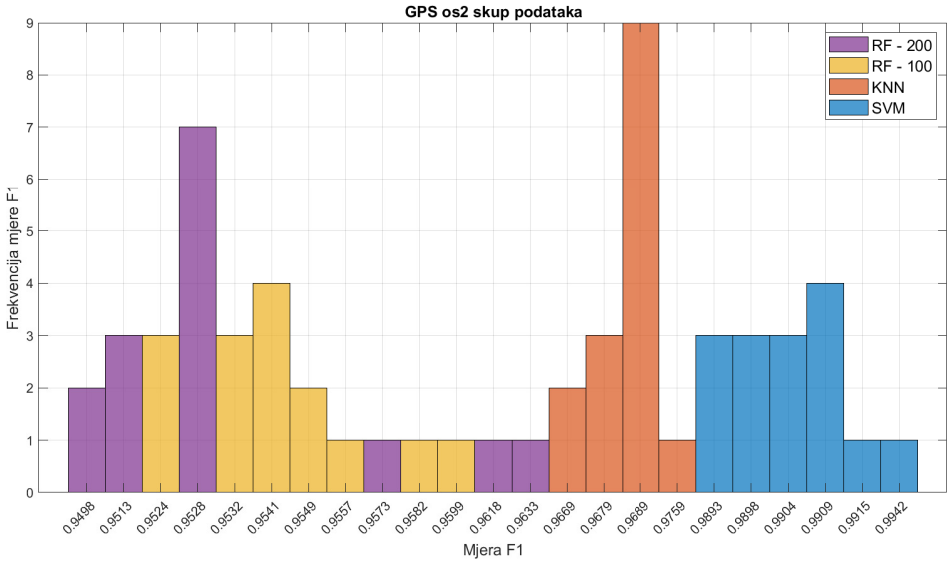
Slika 6.9: Distribucija mjere F1 za različite skupove podataka sustava GPS os2, os3, os4 i Galileo os10, os11, os12 za algoritam klasifikacije A1M2 za model SVM.

Slično, slika 6.8 prikazuje histogram distribucije *F1* mjere za skupove podataka sustava GPS i Galileo za algoritam klasifikacije A1M1. Iz slike je vidljivo da skup os2 sustava GPS i u ovom slučaju ima najbolje vrijednosti mjere *F1* (0.9978). Svi skupovi podataka sustava GPS imaju visoke vrijednosti mjere *F1*. S druge strane, skupovi os11 i os12 sustava Galileo ima najmanje vrijednosti mjere *F1* (0.708).

Na slici 6.9 prikazan je histogram distribucije mjere *F1* za algoritam klasifikacije A1M2 za različite skupove podataka sustava GPS i Galileo. Svi skupovi podataka imaju vrlo visoke vrijednosti mjere *F1*, a kao najbolji se ističe Galileo os10 koji pokazuje najveći broj uzoraka u najvišem području *F1* (0.998 - 0.999).



Slika 6.10: Distribucija mjere F1 za različite modele strojnog učenja primijenjene za algoritam klasifikacije A1 u skupu podataka os10 sustava Galileo.

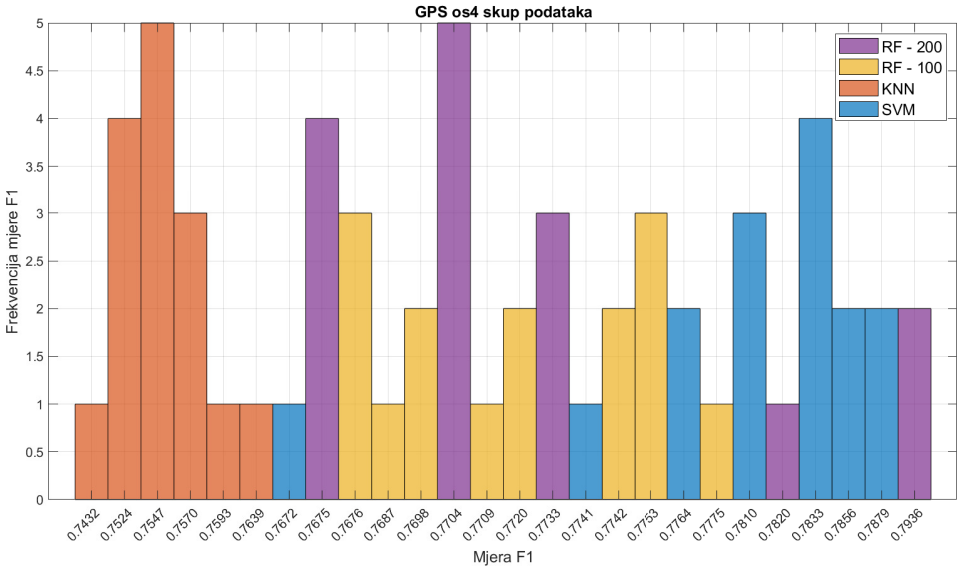


Slika 6.11: Distribucija mjere F1 za različite modele strojnog učenja primijenjene za algoritam klasifikacije A1 u skupu podataka os2 sustava GPS sa slikama kao ulaznim podacima.

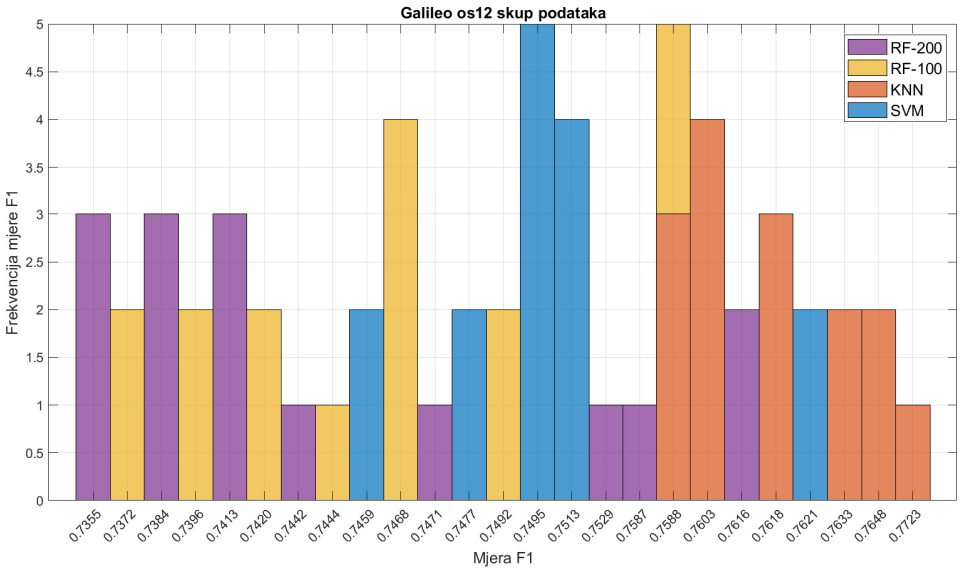
Slika 6.10 prikazuje distribuciju mjere *F1* za četiri različita modela strojnog učenja (KNN, RF-100, RF-200 i SVM) koji su primijenjeni za algoritam klasifikacije A1 u skupu podataka os10 sustava Galileo. Sa slike je vidljivo da model SVM postiže najviše vrijednosti mjere *F1*, s izraženom koncentracijom u rasponu 0.9003 - 0.9384, što ga čini najuspješnijim modelom za klasifikaciju signala u ovom skupu. Modeli RF - 100 i RF - 200 imaju podjednake vrijednosti u rasponu 0.88 - 0.9 dok model KNN ima najniže vrijednosti u rasponu 0.8659 - 0.8812.

Model SVM pokazuje najviše vrijednosti mjere *F1* (0.9893 - 0.9942) i za algoritam klasifikacije A1 u skupu podataka os2 sustava GPS kao što je vidljivo na slici 6.11.

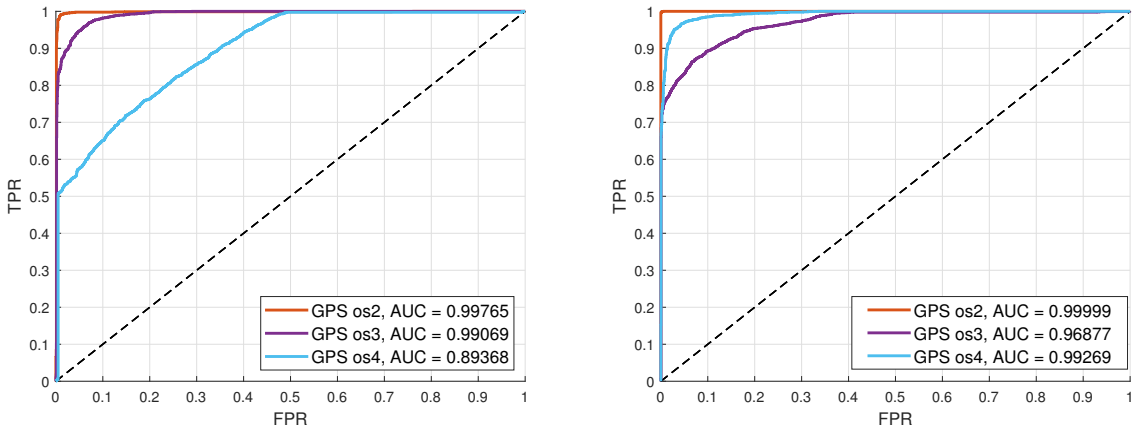
Na slikama 6.12 i 6.13 prikazane su distribucije mjere *F1* za algoritam klasifikacije A1 za različite modele strojnog učenja primijenjene u skupu podataka os4 sustava GPS i os12 sustava Galileo. Sa slika je vidljivo da model SVM pokazuje najviše vrijednosti mjere *F1* za oba promatrana skupa, npr. u slučaju skupa podataka Galileo os12 (0.7588 - 0,7723). Kao što je i očekivano zbog manje razlike u snazi između autentičnih i lažnih signala, točnost klasifikacije je niža.



Slika 6.12: Distribucija mjere F1 za različite modele strojnog učenja primijenjene za algoritam klasifikacije A1 u skupu podataka os4 sustava GPS.

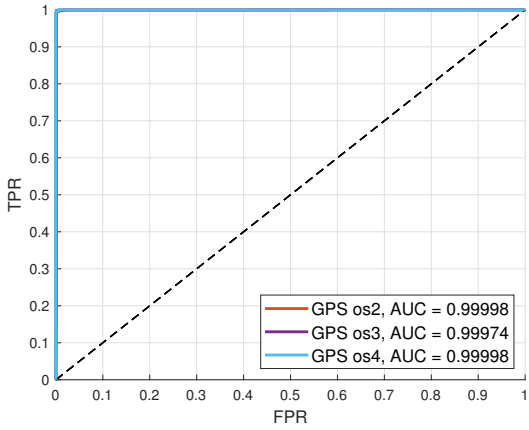


Slika 6.13: Distribucija mjere F1 za algoritam klasifikacije A1 za različite modele strojnog učenja u skupu podataka os12 sustava Galileo.

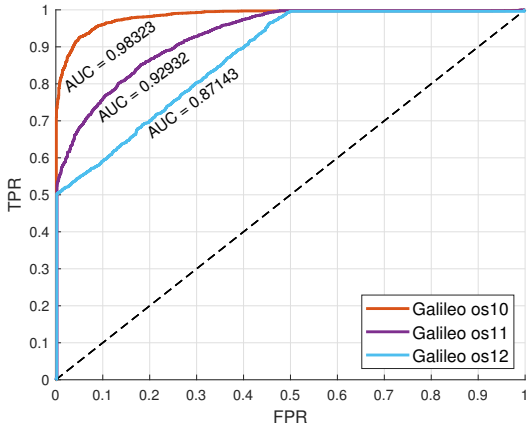


(a)

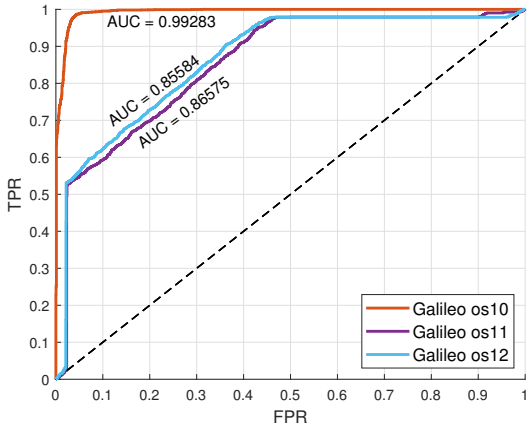
(b)



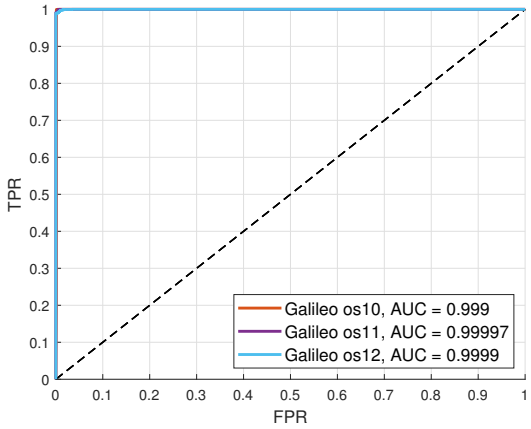
(c)



(d)



(e)



(f)

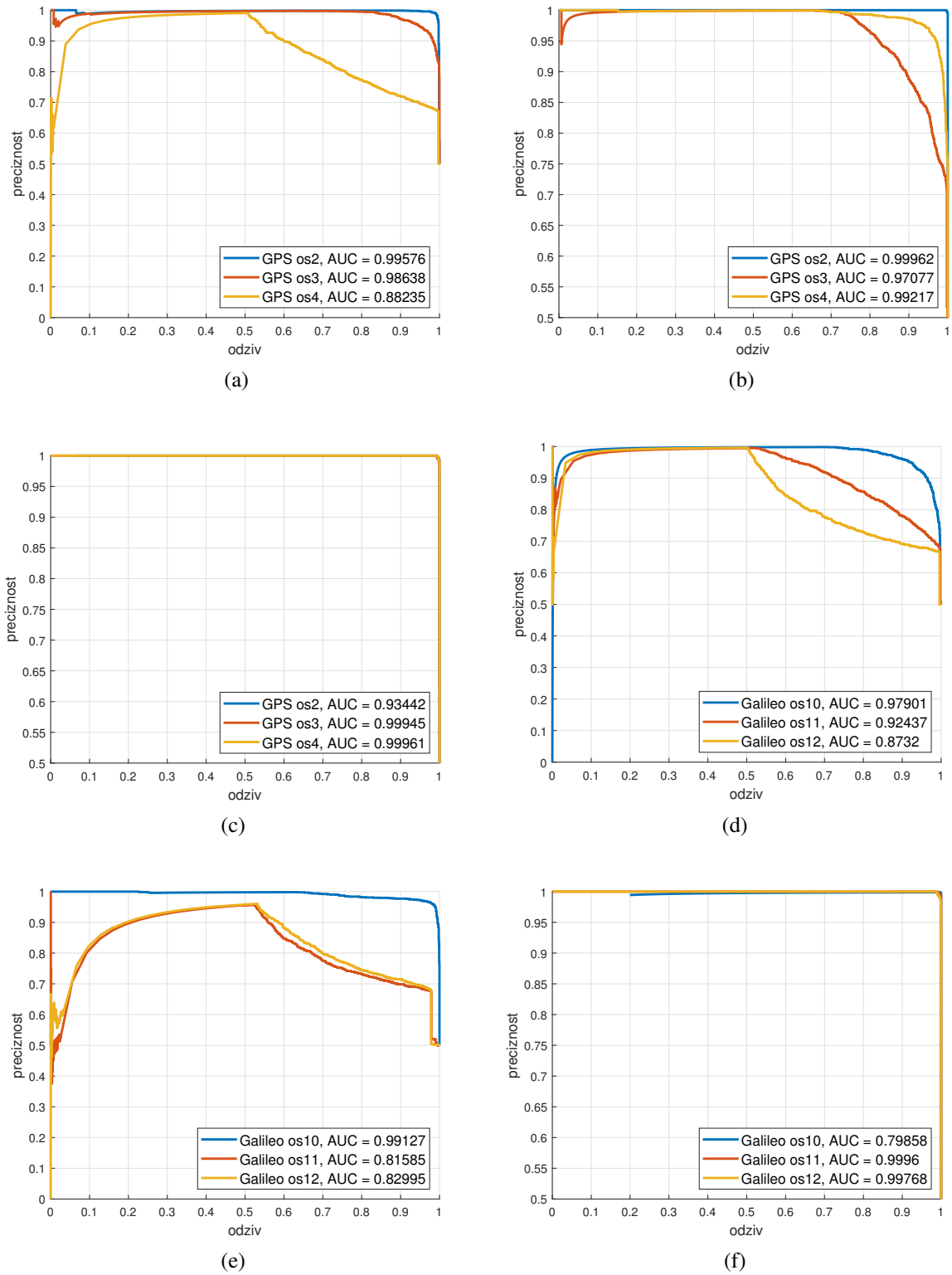
Slika 6.14: Krivulje operativnih karakteristika za valić db4 i algoritam klasifikacije za model SVM u skupovima podataka GPS os2, os3, os4 (a) A1, (b) A1M1, (c) A1M2; i Galileo os10, os11, o12 (d) A1, (e) A1M1; (f) A1M2.

Krivulje operativnih karakteristika za algoritam klasifikacije A1, A1M1 i A1M2 za model SVM u skupovima podataka GPS i Galileo prikazane su na slici 6.14. Korišten je valić

db4. Iz priloženih slika vidljivo je da sve krivulje imaju dosta dobre oblike, ali ističu se krivulje na slikama 6.14c i 6.14f (ulazni podaci su statističke i spektralne značajke signala) koje imaju gotovo savršen oblik i vrijednost AUC gotovo jednaku 1 što ukazuje na vrlo visoku točnost klasifikacije. Najmanja točnost klasifikacije postignuta je u slučaju kad se kao ulazni podaci za klasifikaciju uzmu slike i za skupove podataka GPS i Galileo. U slučaju kad su ulazni podaci slike i značajke slike, rezultati za Galileo su nešto lošiji u odnosu na GPS. Razlozi tomu mogu biti naprednija modulacijska tehnika koju koristi Galileo, geografska pokrivenost i satelitska geometrija jer GPS ima širu i stabilniju pokrivenost dok je Galileo još uvijek u razvoju i može imati manji broj vidljivih satelita u svakom trenutku kao i više grešaka u signalnim karakteristikama što rezultira lošijom klasifikacijom.

Krivulje preciznosti i odziva za i valić db4 i algoritme klasifikacije A1, A1M1 i A1M2 za model SVM u skupovima podataka GPS i Galileo prikazane su na slici 6.15. Iz slika je vidljivo da statističke i spektralne značajke kao ulazni podaci za klasifikaciju modela SVM imaju najveću točnost i vrijednosti odziva R i preciznosti P su gotovo jednake 1 što je blizu idealnog slučaja. Primjerice, za slučaj kada je ulazni klasifikacijski podatak slika, točnost za skup podataka os4 sustava GPS je 78.2%, odziv je 0.7792, a preciznost 0.7867 i krivulja PR je dosta izobličena. U slučaju kada su ulazni podaci značajke slika, točnost za skup podataka os12 sustava Galileo je 76.43%, odziv je 0.7712, a preciznost 0.7515 te je krivulja PR za os12 Galileo dosta izobličena u odnosu na os4 GPS . Sve vrijednosti preciznosti i odziva veće su od 0.7 koji je prag za dobar klasifikacijski model. Preciznost manja od 0.5 ($P < 0.5$) znači da klasifikator ima veliki broj lažno pozitivnih uzoraka koji mogu biti rezultat neuravnotežene klase ili nepodešenih hiperparametara modela. S druge strane, odziv manji od 0.5 ($R < 0.5$) označava veliki broj lažno negativnih uzoraka u klasifikacijskom modelu. Što je vrijednost preciznosti i odziva bliža 1, znači da model nije propustio nijedan pravi pozitivan rezultat odnosno da sve pozitivne uzorke klasifikator označava kao takve. Idealno je kada su svi pozitivno klasificirani uzorci stvarno pozitivni ($P = 1$) i, obrnuto, svi pozitivni uzorci su klasificirani kao pozitivni ($R = 1$).

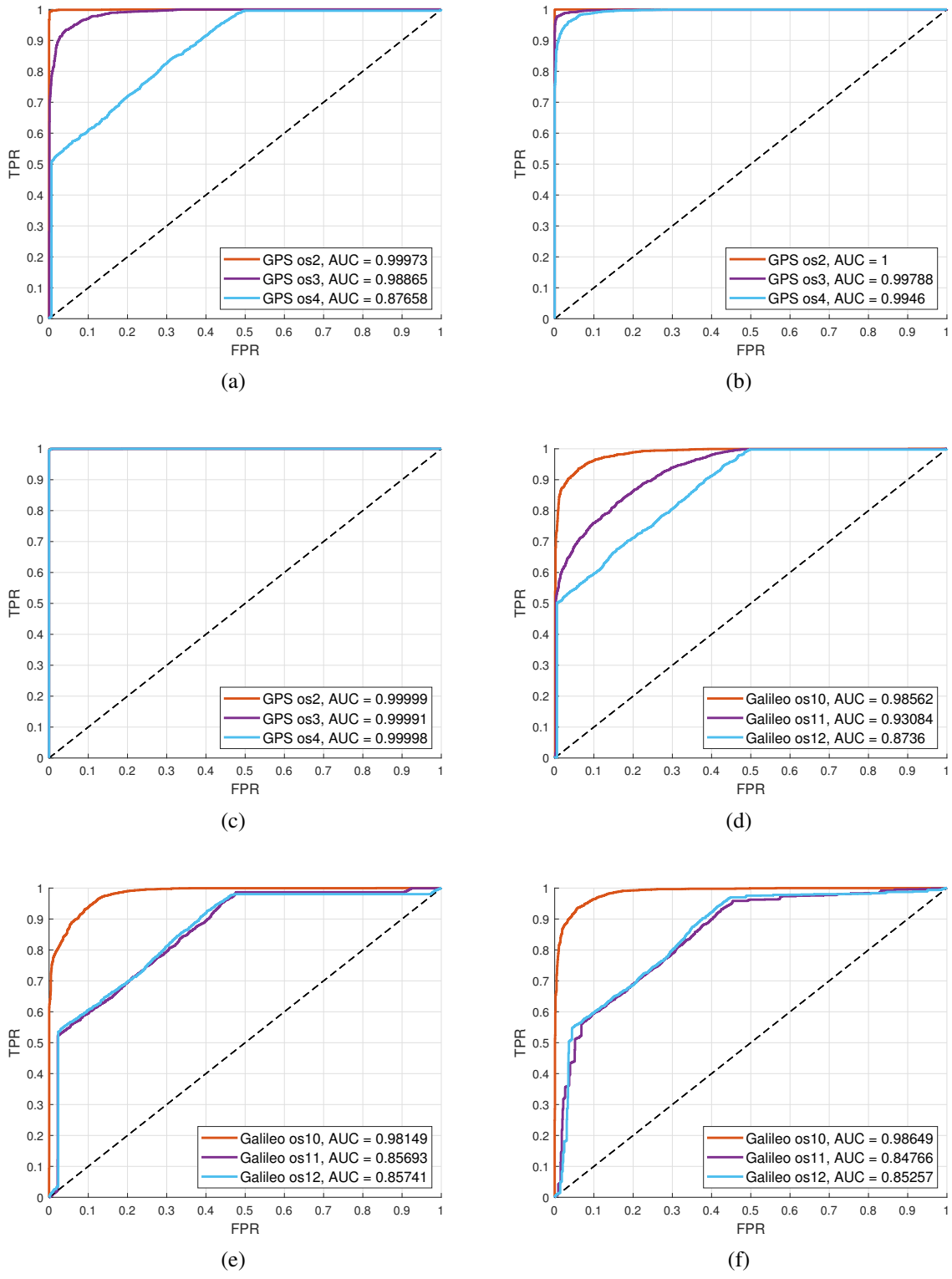
Vrijednosti parametara performansi modela dobivene za skupove podataka os10 i os12 sustava Galileo prikazane su u tablici 6.3. Kod sustava Galileo, dobiveni su nešto lošiji rezultati za sve parametre u odnosu na sustav GPS. Ti dobiveni rezultati pokazuju značajnu razliku u točnostima između skupova podataka os10 i os12 što je i očekivano s obzirom na razliku u snazi između autentičnih i lažnih signala.



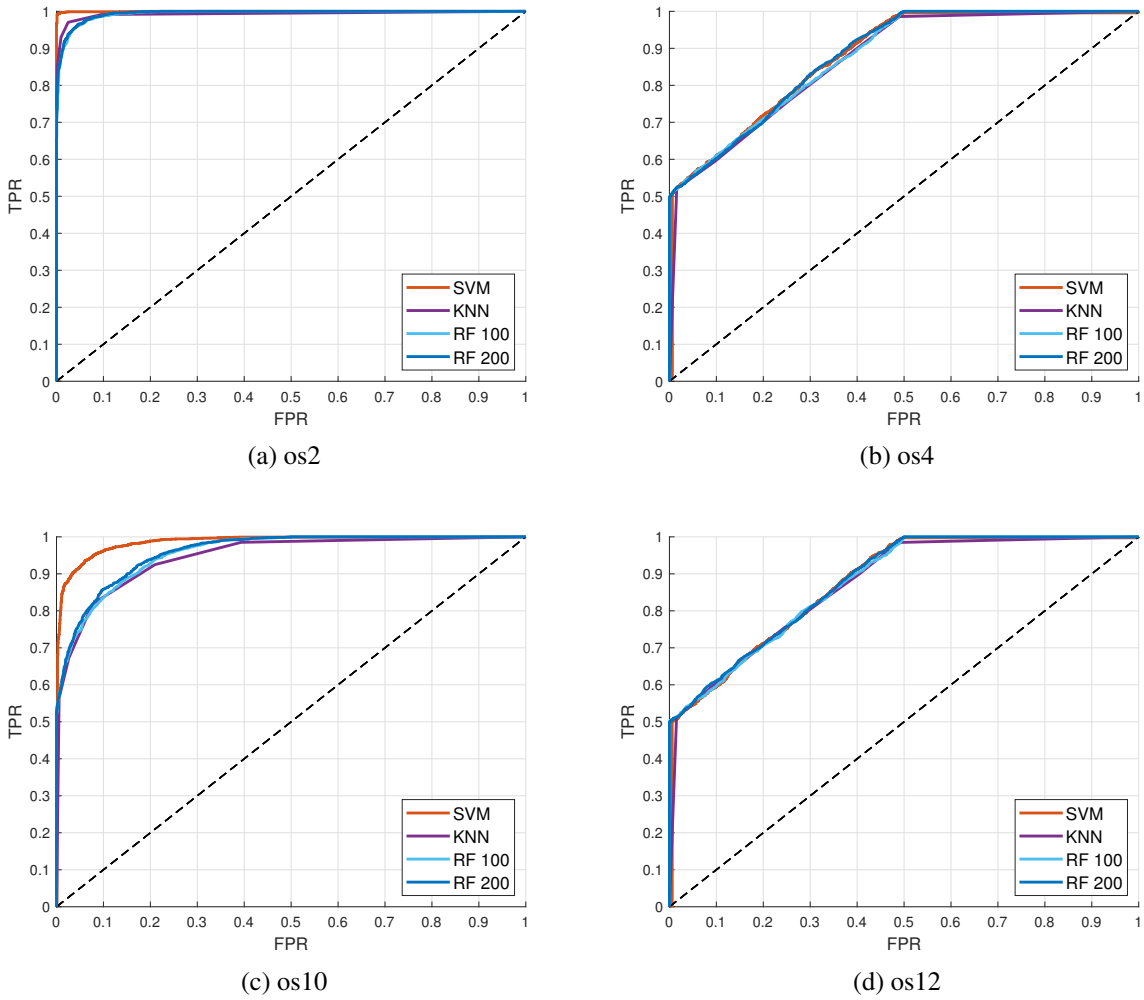
Slika 6.15: Krivulje preciznosti i odziva za valić db4 i algoritam klasifikacije za model SVM u skupovima podataka GPS os2, os3, os4 (a) A1, (b) A1M1, (c) A1M2; i Galileo os10, os11, o12 (d) A1, (e) A1M1; (f) A1M2.

Tablica 6.3: Parametri performansi modela za algoritam klasifikacije za različite modele strojnog učenja za valiće db4 i db8 sustava Galileo: a) A1 (plavo), b) A1M1 (crveno), c) A1M2 (zeleno).

Skup podataka i vrsta koef.	ML model	Srednja točnost [%]	95% interval	Mjera *F1*	Odziv	Preciznost	Vrijeme izvođenja [s]
db4							
	SVM	93.94	[92.8414, 94.0474]	0.9346	0.9322	0.937	684
(clean + os10) approx.	KNN	86.63	[85.927, 87.3322]	0.8657	0.8693	0.8622	722
	RF - 100	89.46	[88.5827, 90.3432]	0.895	0.8922	0.8978	569
	RF - 200	90.35	[89.7567, 90.947]	0.9039	0.9001	0.9078	604
	SVM	75	[73.9226, 76.1145]	0.7512	0.7483	0.7541	522
(clean + os12) approx.	KNN	74.72	[73.6974, 75.747]	0.762	0.7199	0.8093	510
	RF - 100	74.70	[73.3187, 76.0887]	0.7481	0.745	0.7511	548
	RF - 200	73.94	[72.5136, 75.3753]	0.738	0.742	0.7341	533
	SVM	97	[96.5481, 97.5259]	0.9701	0.9785	0.9619	8.7
(clean + os10) approx.	KNN	92.74	[92.2071, 93.2744]	0.9272	0.9303	0.9241	2.3
	RF - 100	95.63	[94.9238, 96.3354	0.9563	0.9563	0.9563	6.5
	RF - 200	95.55	[95.035, 96.0761]	0.9556	0.9552	0.9559	11.6
	SVM	76.43	[75.4082, 77.4436]	0.7612	0.7712	0.7515	3.6
(clean + os12) approx.	KNN	74.35	[72.9636, 75.7401]	0.7438	0.7431	0.7444	2
	RF - 100	77.29	[76.15, 78.4426]	0.7718	0.7758	0.7678	7.45
	RF - 200	77.77	[76.7125, 78.8431]	0.777	0.7799	0.7741	13.14
	SVM	99.87	[99.7804, 99.9603]	0.9987	0.9989	0.9985	4.5
(clean + os10) approx.	KNN	99.88	[99.8, 99.9788]	0.9989	0.9978	1	1.9
	RF - 100	99.94	[99.8862, 100]	0.9994	0.9989	1	5.04
	RF - 200	99.94	[99.8862, 100]	0.9994	0.9989	1	8.72
	SVM	99.33	[98.9738, 99.6929]	0.9933	0.9941	0.9926	4.4
(clean + os12) approx.	KNN	98.83	[98.5562, 99.1104]	0.9885	0.9786	0.9985	2
	RF - 100	99.88	[99.7444, 99.9222]	0.9989	0.9993	0.9985	5.2
	RF - 200	99.83	[99.8176, 99.9602]	0.9983	0.9989	0.9978	8.9
db8							
	SVM	93.59	[93.0548, 94.1304]	0.9362	0.9321	0.9404	599.51
(clean + os10) approx.	KNN	87.09	[86.4575, 87.7277]	0.8686	0.8394	0.9	562.73
	RF - 100	86.74	[86.1848, 87.2966]	0.8671	0.8693	0.8648	568.87
	RF - 200	87.54	[86.9246, 88.1495]	0.8753	0.8761	0.8744	586.87
	SVM	75.30	[74.6467, 75.9459]	0.7508	0.7573	0.7444	564.41
(clean + os12) approx.	KNN	75.48	[74.1129, 76.8501]	0.7554	0.7537	0.757	597.62
	RF - 100	75.11	[74.0661, 76.1561]	0.7521	0.7491	0.7552	272.4
	RF - 200	75.19	[74.2103, 76.1601]	0.7483	0.7591	0.7378	583.12
	SVM	91.81	[88.7672,94.8623]	0.9175	0.9244	0.9107	5.52
(clean + os10) approx.	KNN	95.20	[94.6956, 95.7119]	0.9521	0.9509	0.9533	2.78
	RF - 100	96.26	[95.6088, 96.9097]	0.9626	0.9616	0.9637	7
	RF - 200	95.83	[95.2137, 96.453]	0.9584	0.9571	0.9596	10.49
	SVM	75.22	[73.531, 76.9135]	0.7466	0.764	0.73	2.62
(clean + os12) approx.	KNN	76.5	[75.1772, 77.8228]	0.7620	0.7717	0.7526	1.94
	RF - 100	76.59	[75.5782, 77.6069]	0.7636	0.7711	0.7563	7.69
	RF - 200	75.85	[74.4598, 77.2439]	0.7591	0.7572	0.7611	12.46
	SVM	93.46	[90.7088, 96.2171]	0.934	0.9433	0.9248	10.85
(clean + os10) approx.	KNN	95.02	[94.3518, 95.6853]	0.9504	0.9467	0.9541	3.07
	RF - 100	96.11	[95.4691, 96.7531]	0.9614	0.9554	0.9674	17.76
	RF - 200	96.28	[95.7771, 96.7785]	0.9629	0.9589	0.9967	11.94
	SVM	74.54	[73.0168, 76.0573]	0.7431	0.7497	0.7367	5.12
(clean + os12) approx.	KNN	76.85	[75.8626, 77.8411]	0.7666	0.773	0.7604	8.09
	RF - 100	76.67	[75.3565,77.97678]	0.7671	0.7657	0.7685	15.25
	RF - 200	77.26	[75.9364,78.5821]	0.7722	0.7736	0.7707	12.51



Slika 6.16: Krivulje operativnih karakteristika za valić db8 i algoritam klasifikacije za model SVM u skupovima podataka GPS os2, os3, os4 (a) A1, (b) A1M1, (c) A1M2; i Galileo os10, os11, o12 (d) A1, (e) A1M1; (f) A1M2.



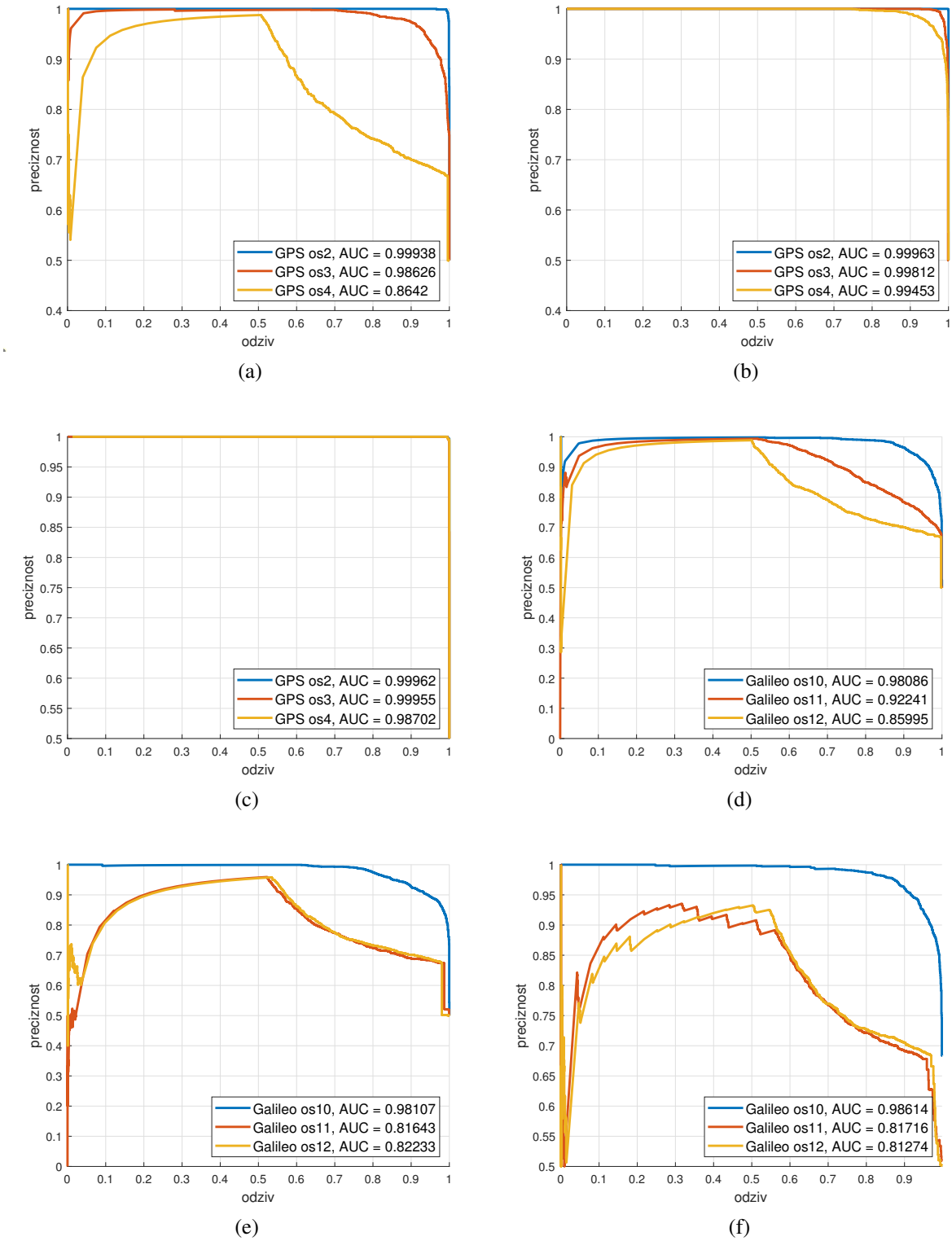
(a) os2 (b) os4

(c) os10 (d) os12

Slika 6.17: Krivulje operativnih karakteristika za različite modele strojnog učenja za algoritam klasifikacije A1 korištenjem valića db8 u skupovima podataka GPS (a) os2 i (b) os4 te Galileo (c) os10 i (d) os12.

Na slici 6.16 prikazane su krivulje operativnih karakteristika i pripadajuće vrijednosti AUC za algoritam klasifikacije A1 za model SVM i valić db8 u sustavima GPS i Galileo. Vidljivo je da kod klasifikacije na temelju slika, najbolju vrijednost AUC ima skup podataka os2 sustava GPS i ona iznosi 0.99973 (slika 6.16a) dok os4 ima znatno nižu vrijednost AUC = 0.87658. Kod sustava Galileo (slika 6.16d), slično se ponašaju skupovi podataka os10 (AUC = 0.98562) i os12 (AUC = 0.8736) koji su ekvivalentni skupovima podataka os2 i os4 sustava GPS. Za klasifikaciju na temelju značajki slika svi modeli imaju jako dobre rezultate za sve skupove podataka, a ističe se skup podataka os2 koji ima AUC = 1 (slika 6.16b). Rezultati za klasifikaciju koja koristi spektralne i statističke značajke pokazuju najbolje rezultate za skupove podataka sustava GPS (slika 6.16c) dok su kod sustava Galilea rezultati nešto lošiji npr. za os11 vrijednost AUC je 0.84766 (slika 6.16f). Skupovi podataka sustava GPS postižu bolje klasifikacijske performanse od skupova sustava Galileo u svim scenarijima. Može se zaključiti da upotreba obje vrste valića db4 i db8 daje vrlo dobre klasifikacijske rezultate za

sve vrste korištenih ulaznih podataka.



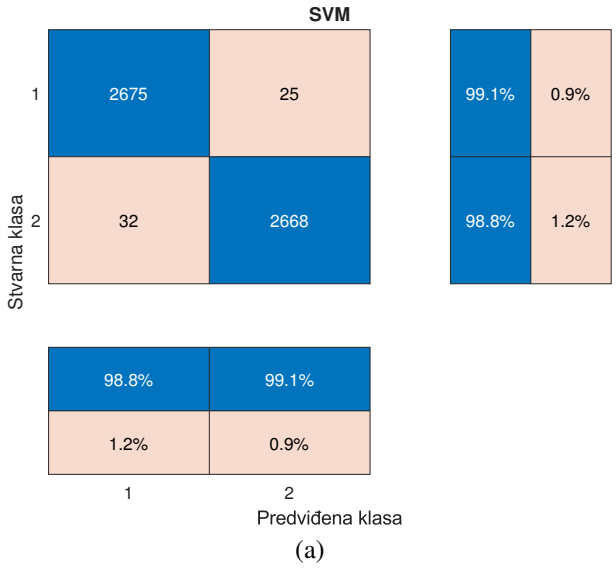
Slika 6.18: Krivulje preciznosti i odziva za valić db8 i algoritam klasifikacije za model SVM u skupovima podataka GPS os2, os3, os4 (a) A1, (b) A1M1, (c) A1M2; i Galileo os10, os11, o12 (d) A1, (e) A1M1; (f) A1M2.

Krivulje operativnih karakteristika za četiri korištena modela strojnog učenja za klasifi-

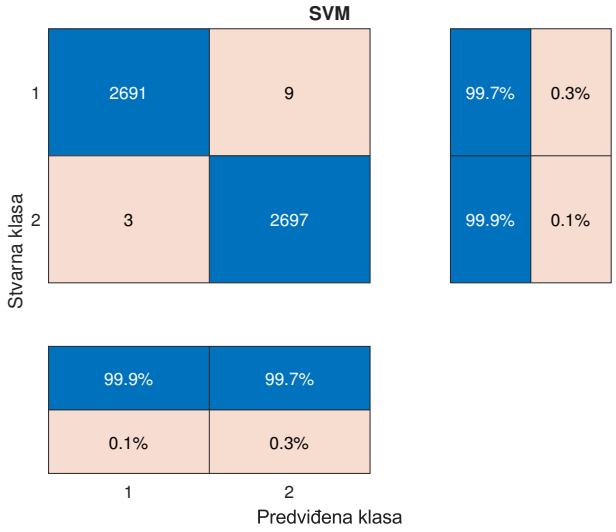
kaciju tipa signala temeljenu na slikama korištenjem valića db8 u skupovima podataka GPS i Galileo prikazane su na slici 6.17. Vidljivo je da je krivulja za model SVM u skupu podataka os2 približno jednaka idealnom slučaju tj. AUC = 0.99973 (slika 6.17a) odnosno stopa lažno pozitivnih FPR je jednaka 0. Kod skupa os10, model SVM isto ima najbolje rezultate s vrijednosti AUC = 0.98562 (slika 6.17c). Svi ostali modeli imaju podjednako dobre rezultate i nešto lošije nego model SVM. Kao i u slučaju valića db4, krivulje operativnih karakteristika su izobličene za skupove podataka os4 (slika 6.17b) i os12 (slika 6.17d) za sve modele strojnog učenja.

Slika 6.18 prikazuje krivulje preciznosti i odziva za klasifikaciju korištenjem modela SVM i valića db8 u skupovima podataka GPS i Galileo temeljenu na različitim ulaznim podacima. Svi skupovi podataka sustava GPS ostvaruju jako dobre vrijednosti AUC, ali os2 se ističe s najvišom preciznošću i odzivom. S druge strane, skup podataka os10 sustava Galileo jasno nadmašuje os11 i os12 za sve vrste ulaznih podataka. Generalno, bolje performanse pokazuju skupovi podataka sustava GPS u odnosu na Galileo.

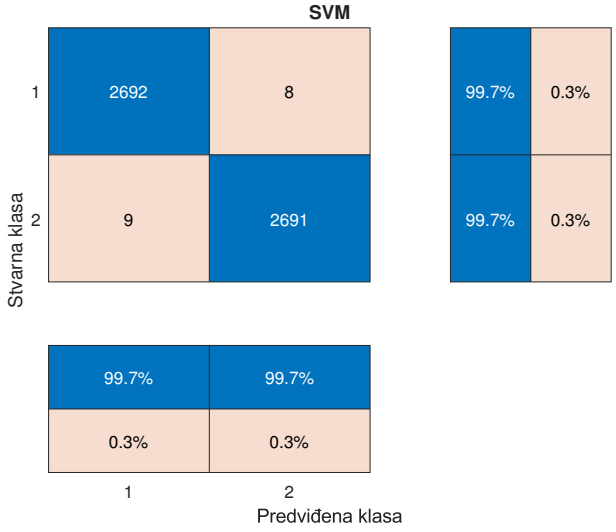
Ukupna sumirana konfuzijska matrica kroz sve iteracije za model SVM u skupu podataka os2 za aproksimacijske koeficijente i valić db4 prikazana je za tri različita tipa ulaznih podataka na slici 6.19. Klasa 1 predstavlja autentične signale dok klasa 2 predstavlja lažne signale. Sumirana matrica gradi se na osnovi svih 5400 testnih uzoraka kroz 15 modela (5 podskupova po 3 ponavljanja). Od toga je 5 x 180 (20%) autentičnih uzoraka i 5 x 180 lažnih uzoraka što je ukupno 1800. Primjerice, u slučaju kad su klasifikacijski podaci slike (slika 6.19a), model je ispravno klasificirao 2675 autentičnih uzoraka kao autentične (klasa 1), 25 autentičnih uzoraka je pogrešno klasificirano kao lažni. Nadalje, 2668 lažnih uzoraka je ispravno klasificirano kao lažni (klasa 2) i 32 lažna uzorka su pogrešno klasificirana kao autentični. Vrijednost 98.8% predstavlja preciznost, a 99.1% odziv modela. Za ostale tipove ulaznih podataka (slike 6.19b i 6.19c), model je napravio manje pogrešaka prilikom klasifikacije.



(a)



(b)

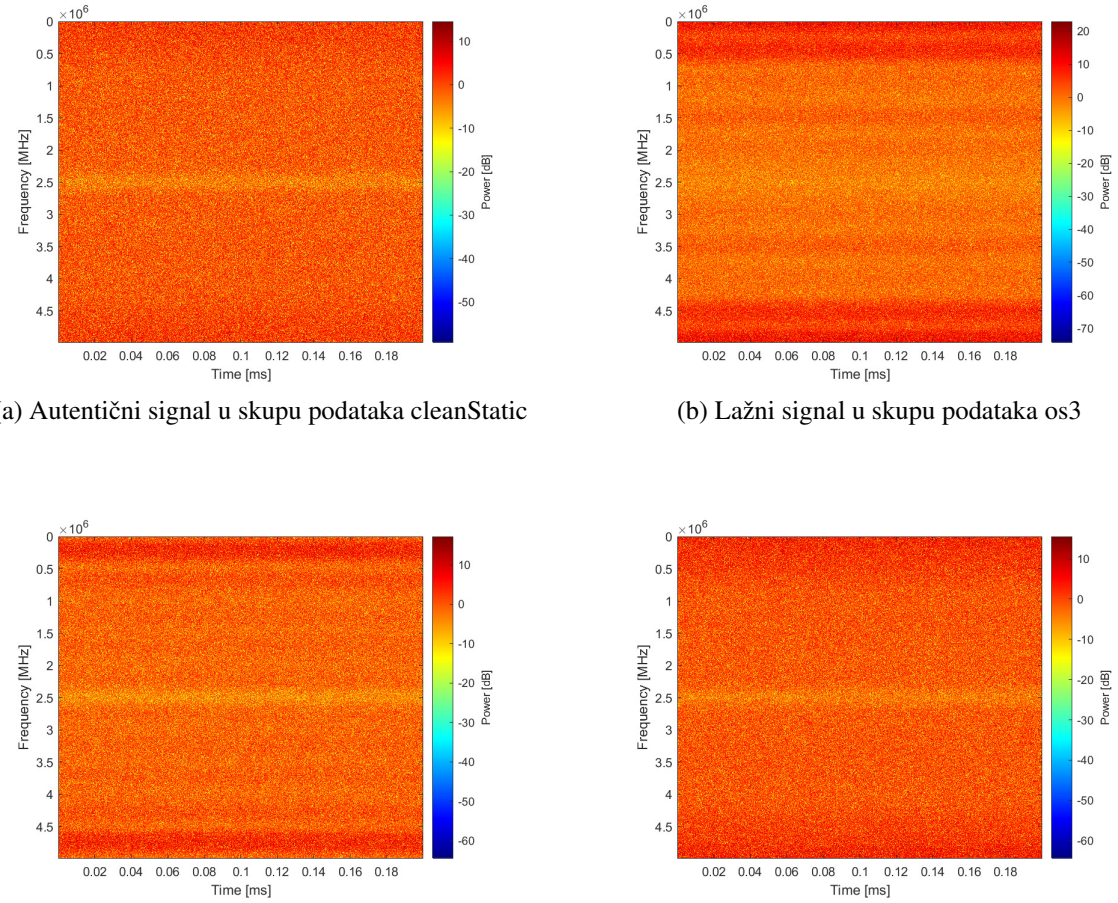


(c)

Slika 6.19: Ukupna sumirana konfuzijska matrica kroz sve iteracije za model SVM u skupu podataka GPS os2 za aproksimacijske koeficijente i valić db4 temeljeno na vrsti ulaznih podataka: a) slika, b) značajke izvučene iz slike, c) statističke i spektralne značajke.

6.6. Rezultati evaluacije metode detekcije temeljene na primjeni spektrograma

U ovom istraživanju analizirani su rezultati klasifikacije tipa signala na temelju generiranih skupova spektrograma za različite vrijednosti veličine prozora - 512 i 1024. Sukladno veličini prozora, uzete su vrijednosti parametara preklapanja (50% od veličine prozora) i broja točaka brze Fourierove transformacije *nfft* (dva puta veći od veličine prozora).



(a) Autentični signal u skupu podataka cleanStatic

(b) Lažni signal u skupu podataka os3

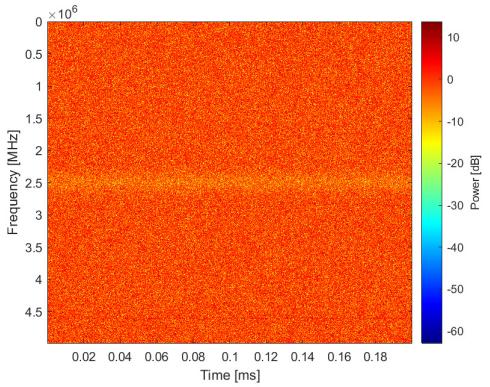
(c) Lažni signal u skupu podataka os2

(d) Lažni signal u skupu podataka os4

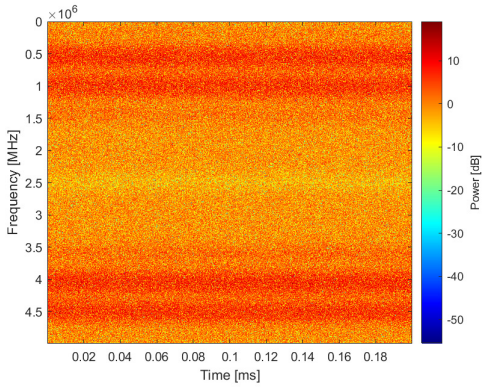
Slika 6.20: Spektrogrami za autentične i lažne signale sustava GPS uz širinu prozora 512.

Na slici 6.20 prikazani su spektrogrami koji su dobiveni za četiri skupa podataka koji uključuju samo signale sustava GPS i širinu prozora 512. Slika 6.20a prikazuje spektrogram za autentični signal u skupu podataka *cleanStatic* dok su lažni signali u skupovima os2, os3 i os4 prikazani na slikama 6.20b, 6.20c i 6.20d. Može se primijetiti da postoji više crvenih linija u spektrogramima za skupove os2 i os3 u kojima lažni signali imaju veću snagu od autentičnih signala za +10 dB i +1,3 dB što ukazuje na napad lažiranjem. Spektrogrami za skupove podataka *cleanStatic* i os4 gotovo su isti jer autentični i lažni signali imaju gotovo istu snagu (lažni signali imaju samo +0,4 dB veću snagu što je gotovo nezamjetno).

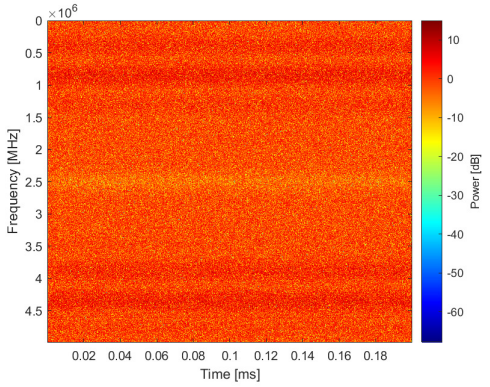
Slika 6.21 prikazuje spektrograme generirane za četiri skupa podataka za signale sustava



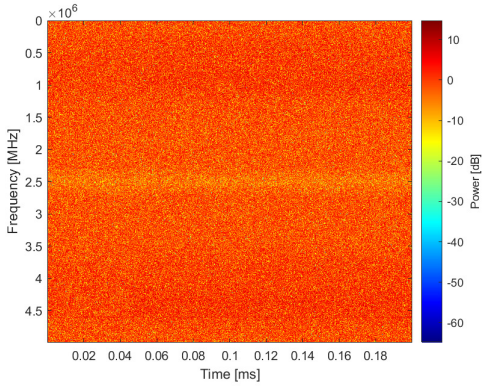
(a) Autentični signal u skupu podataka cleanStatic



(b) Lažni signal u skupu podataka os10



(c) Lažni signal u skupu podataka os11



(d) Lažni signal u skupu podataka os12

Slika 6.21: Spektrogrami za autentične i lažne signale sustava Galileo i širinu prozora 512.

Galileo i širinu prozora 512. Spektrogram za autentični signal prikazan je na slici 6.21a dok su spektrogrami za lažne signale u skupovima podataka os10, os11 i os12 sustava Galileo prikazani na slikama 6.21b, 6.21c i 6.21d. Spektrogrami za skupove podataka os10 i os11 vrlo su slični spektrogramima za skupove podataka os2 i os3 jer imaju iste karakteristike što se tiče prednosti razine snage lažnih u odnosu na autentične signale. Nadalje, spektrogrami za autentične signale i skup podataka os12 su slični kao kod istih skupova u sustavu GPS. Kada se veličina prozora poveća na 1024, dobije se bolja frekvencijska, a lošija vremenska rezolucija i spektrogram pokazuje jasnije razdvojene frekvencije, ali je zamućen u vremenu.

Na temelju generiranih skupova spektrograma (slike u *.tiff* formatu) za signale sustava GPS (7200 slika) i Galileo (7200 slika) izvršena je klasifikacija tipa signala korištenjem nekoliko modela strojnog učenja: SVM, KNN, RF-100 i RF-200. Zbog problema pretreniranja modela, u svim modelima napravljena je unakrsna provjera valjanosti k-podskupova. Broj podskupova postavljen je na 5 i prema tome postotak (ovisi o vrijednosti k) podataka za treniranje je 80%, a za testiranje 20%. Svi podaci se koriste i za treniranje i za testiranje samo u različitim iteracijama.

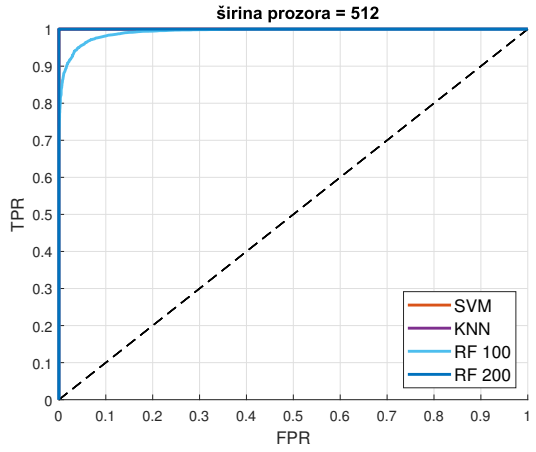
Tablica 6.4 prikazuje parametre performansi modela za klasifikaciju tipa signala za spek-

trograme širine prozora 512 i 1024 u sustavu GPS za različite ulazne podatke. Iz tablice se može zaključiti da svi korišteni modeli imaju dobre performanse. Performanse modela su nešto lošije za skup podataka os4 u kojem se nalaze autentični i lažni signali kod kojih je razlika u snazi 0.4 dB. Najlošije performanse modela ima model KNN za širinu prozora 1024 i skup podataka os4 za značajke slike. Mogući razlog za to je da vizualne razlike nisu izražene kao kod skupa podataka os2, pa značajke nisu dovoljno razlučive, a model KNN se oslanja na udaljenosti značajki - ako su uzorci slični, ne može ih pouzdano razlikovati. Vrijeme izvođenja pojedinog modela bitno se razlikuje za različite tipove klasifikacijskih podataka. Primjerice, vrijeme izvršavanja za model SVM sa slikama i širinom prozora 512 je 1100 sekundi dok za značajke slike kao ulazne podatke iznosi 7.7 sekundi. Ako se usporedi vrijeme izvođenja modela za različite širine prozora, 512 i 1024, može se vidjeti da je u većini slučajeva to vrijeme veće za širinu prozora 1024 što je i očekivano s obzirom na veći broj uzoraka po spektrogramu. Vrijeme izvršavanja pojedinog modela osim o broju i veličini uzoraka, dimenzionalnosti značajki, tipu modela i njegovoj složenosti, ovisi značajno i o hardverskim resursima računala kao što su: procesor, radna memorija te predmemorija (*cache*).

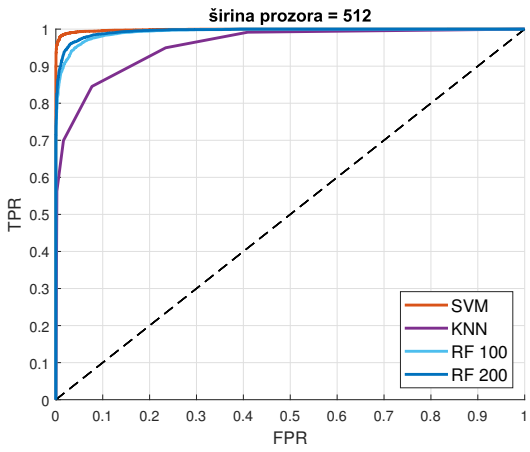
Parametri performansi modela u sustavu Galileo prikazani su u tablici 6.5. Iz tablice je vidljivo da i signali sustava Galileo pokazuju vrlo dobre performanse, iako nešto lošije u odnosu na sustav GPS što se posebno odnosi na značajke slike za skup podataka os12 za obje širine prozora. Primjerice, model SVM za skup podataka os12 i značajke slike ima točnost od 81.6% i preciznost od 0.757 dok navedeni parametri za isti skup podataka os4 sustava GPS ima vrijednosti 92.18% i 0.9106.

Na slici 6.22 prikazane su krivulje operativnih karakteristika za nekoliko modela strojnog učenja za klasifikaciju tipa signala u sustavima GPS i Galileo za spektrograme i širine prozora 512. Ulazni podaci u sve modele strojnog učenja su slike. Slike 6.22a i 6.22b prikazuju krivulje za skupove podataka os2 i os4 u sustavu GPS dok su na slikama 6.22c i 6.22d prikazani rezultati za skupove podataka os10 i os12 u sustavu Galileo. Sa slika je vidljivo da model SVM ima najbolje rezultate za sve skupove podataka. U skupovima podataka os4 i os12, najlošije rezultate je ostvario model KNN. Ako se usporede krivulje operativnih karakteristika za iste modele strojnog učenja i spektrograme s većom širinom prozora 1024 koje su prikazane na slici 6.23, vidljivo je da su bolji rezultati ostvareni s povećanjem širine prozora za sve korištene modele. I u slučaju širine prozora 1024, model SVM pokazuje najbolje performanse za sve skupove podataka.

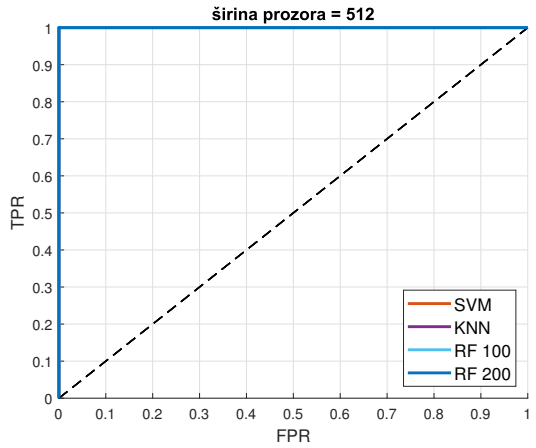
Na slikama 6.24 i 6.25 prikazane su krivulje operativnih karakteristika za iste modele strojnog učenja za signale sustava GPS i Galileo i spektrograme širine prozora 512 i 1024 za značajke slika kao ulazne podatke. Krivulje za sve modele i spektrograme širine prozora 512 imaju vrlo dobre performanse te da skoro sve krivulje idu uz gornji lijevi rub grafa ($TPR = 1$, $FPR = 0$). U slučaju širine prozora 1024, krivulje imaju dosta izobličeniji oblik, ali i dalje vrijednost AUC veću od 0.5 (crna dijagonalna isprekidana crta) za skupove podataka os4 i



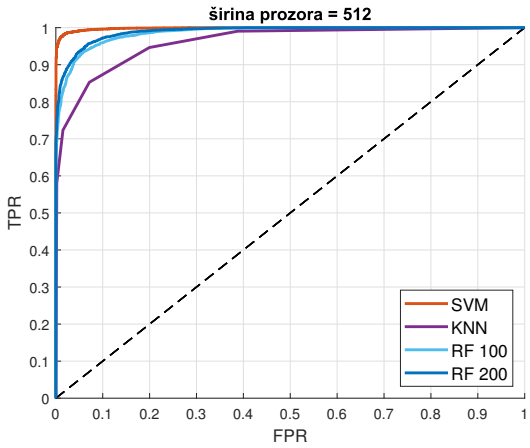
(a) os2



(b) os4



(c) os10

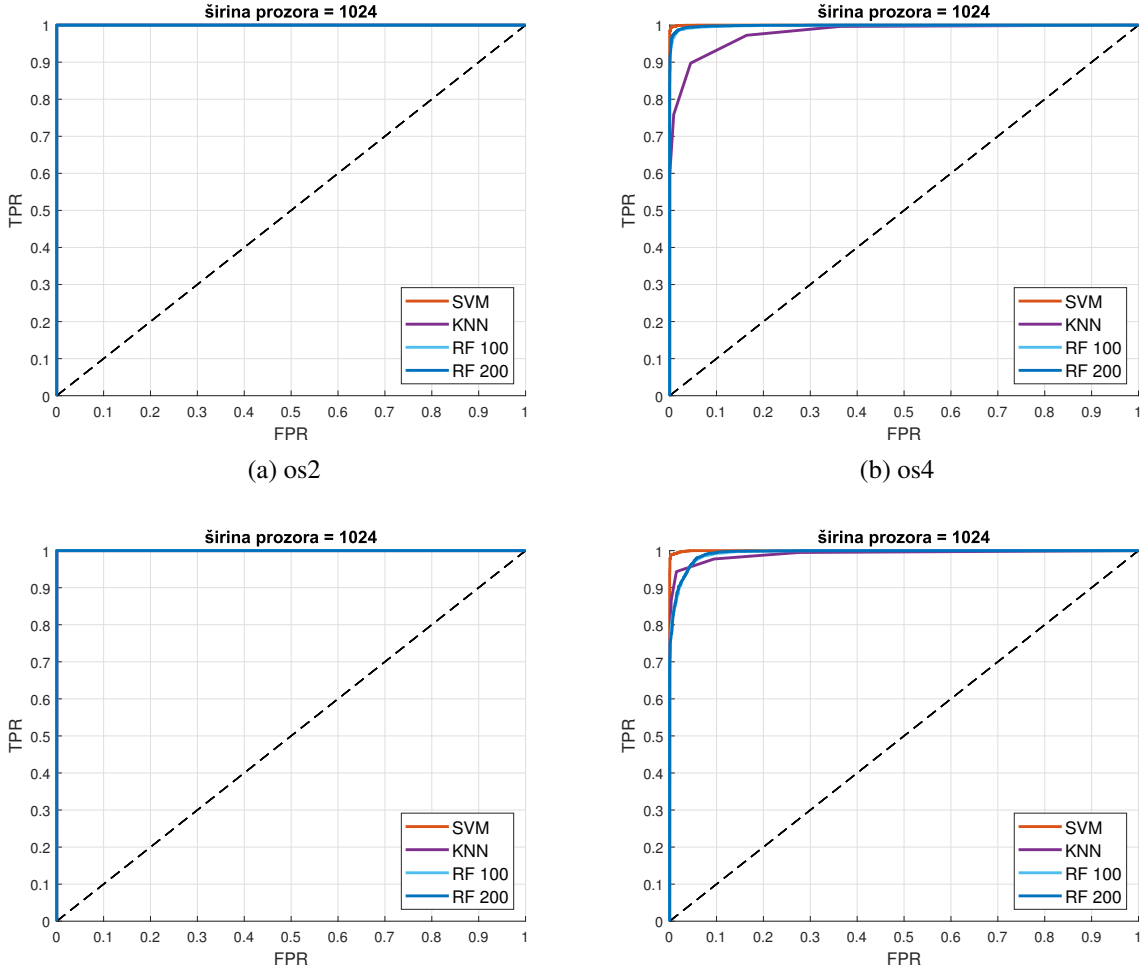


(d) os12

Slika 6.22: Krivulje operativnih karakteristika za različite modele strojnog učenja za klasifikaciju tipa signala za skupove podataka sustava GPS (a) os2 i (b) os4 te Galileo (a) os10 i (b) os12 i širinu prozora 512 za slike kao ulazne podatke.

os12 u kojima je teže razlikovati autentične i lažne signale zbog male razlike u snazi. Ako se usporede krivulje operativnih karakteristika za različite ulazne klasifikacijske podatke (slike i značajke slike) te različite širine prozora, može se zaključiti da za slike kao klasifikacijski podatak, veći prozor (1024) daje bolje rezultate jer čuva kompletnu informaciju, bolju teksturu i frekvencijsku dubinu. S druge strane, značajke iz slike ne rastu proporcionalno s veličinom prozora (ne rastu s kompleksnošću slike) i ne mogu iskoristiti dodatnu složenost tj. ne skaliraju dovoljno da uhvate sve obrasce. Značajke iz slike gube dio svoje diskriminativne snage (sposobnost razlikovanja klasa) zbog sačimanja informacija, dok korištenje cijele slike omogućuje bolju razdvojenost klasa.

Na slici 6.26 prikazana je distribucija mjere F1 za SVM klasifikator s ulaznim podacima u obliku slika. Vidljivo je da skupovi os2 i os10 uz širinu prozora 1024 rezultiraju gotovo savršenom klasifikacijom ($F1 \approx 1$). S druge strane, niže vrijednosti mjere $F1$ dobivene su za os4 i os10 pri širini prozora 512, što ukazuje na to da slike manje razlučivosti ne sadrže



(a) os2 (b) os4

(c) os10 (d) os12

Slika 6.23: Krivulje operativnih karakteristika za različite modele strojnog učenja za klasifikaciju tipa signala za skupove podataka sustava GPS (a) os2 i (b) os4 te Galileo (a) os10 i (b) os12 i širinu prozora 1024 za slike kao ulazne podatke.

dovoljno informacija za preciznu klasifikaciju. Povećanje širine prozora na 1024 značajno poboljšava diskriminativnost vizualnih značajki, osobito kod skupova podataka os4 i os12 u kojima je teže razlikovati autentične i lažne signale, što potvrđuje važnost pravilnog odabira veličine prozora pri generiranju klasifikacijskih slika. Primjerice, SVM klasifikator je 5 puta postigao vrijednost *F1* jednaku 0.9806 za Galileo os12 - 512 u intervalu [0.9806, 0.9811].

Distribucija mjere *F1* za SVM klasifikator kada se koriste značajke iz slike, prikazana je na slici 6.27. U usporedbi s prethodnim pristupom klasifikacije gdje su korištene cijele slike (slika 6.26), ovdje je vidljiv širi raspon vrijednosti mjere *F1* (od 0.7 do 0.99), što ukazuje na nižu stabilnost i točnost klasifikatora. Skup podataka os2 pokazuje najbolje rezultate za obje širine prozora. Najmanje vrijednosti mjere *F1* ima skup podataka os12 za obje širine prozora (u slučaju 512 prosječna vrijednost mjere *F1* iznosi 0.8 dok za 1024 iznosi 0.76). U usporedbi sa slikama kao klasifikacijskim podacima, značajke iz slika rezultiraju manjom diskriminativnom snagom i većim gubitkom informacija, osobito kod skupova podataka u

Tablica 6.4: Parametri performansi modela za algoritam klasifikacije za različite modele strojnog učenja za spektrograme i širine prozora 512 i 1024 u sustavu GPS: a) A1 (plavo), b) A1M1 (crveno).

Skup podataka i vrsta koef.	ML model	Srednja točnost [%]	95% interval	Mjera *F1*	Odziv	Preciznost	Vrijeme izvođenja [s]
širina prozora = 512							
(clean + os2) approx.	SVM	99.93	[99.8725, 99.9955]	0.9994	0.9994	0.9994	1100
	KNN	100	[100, 100]	1	1	1	1446
	RF - 100	99.97	[99.9431, 100]	0.9997	0.9994	1	1327
	RF - 200	99.97	[99.9431, 100]	0.9997	0.9994	1	1671
(clean + os4) approx.	SVM	98.46	[98.1601, 98.7715]	0.985	0.9786	0.9915	1334
	KNN	88.42	[87.7521, 89.0812]	0.8885	0.8565	0.923	1316
	RF - 100	95.34	[94.9933, 95.6919]	0.9538	0.9543	0.9626	1359
	RF - 200	96.37	[96.1185, 96.6222]	0.964	0.9568	0.9713	2000
(clean + os2) approx.	SVM	99.91	[99.8292, 99.9856]	0.9991	0.9994	0.9987	7.7
	KNN	99.92	[99.8737, 99.9782]	0.9993	0.9994	0.9991	4.6
	RF - 100	99.95	[99.9194, 99.988]	0.9995	0.9991	1	5.9
	RF - 200	99.92	[99.8585, 99.9749]	0.9992	0.9989	0.9994	10.1
(clean + os4) approx.	SVM	92.18	[91.6993, 92.671]	0.921	0.9316	0.9106	4.1
	KNN	89.87	[89.2081, 90.5511]	0.9025	0.8706	0.9369	2.8
	RF - 100	88.56	[87.7666, 89.363]	0.8806	0.9215	0.8431	5.5
	RF - 200	88.81	[88.0408, 89.5888]	0.8832	0.9241	0.8457	10.3
širina prozora = 1024							
(clean + os2) approx.	SVM	99.99	[99.9726, 100]	0.9999	0.9998	1	1312
	KNN	100	[100, 100]	1	1	1	1259
	RF - 100	99.97	[99.9431, 100]	0.9997	0.9994	1	1264
	RF - 200	99.97	[99.9431, 100]	0.9997	0.9994	1	1398
(clean + os4) approx.	SVM	99.37	[99.2581, 99.4827]	0.9937	0.9892	0.9983	1552
	KNN	92.93	[92.0689, 93.1903]	0.9284	0.903	0.9552	1831
	RF - 100	98.13	[97.8682, 98.4096]	0.9815	0.976	0.987	1458
	RF - 200	98.37	[98.1638, 98.5769]	0.9838	0.9784	0.9893	1664
(clean + os2) approx.	SVM	99.83	[99.762, 99.9046]	0.9983	0.9994	0.9972	7.2
	KNN	99.88	[99.8176, 99.9602]	0.9989	0.9989	0.9989	3.6
	RF - 100	99.91	[99.8649, 99.9685]	0.9992	0.9994	0.9989	5.7
	RF - 200	99.91	[99.8585, 99.9749]	0.9992	0.9994	0.9989	10.2
(clean + os4) approx.	SVM	90.83	[90.3717, 91.295]	0.9079	0.912	0.9039	5.1
	KNN	64.6	[63.9161, 65.2691]	0.7291	0.5904	0.9531	2.8
	RF - 100	90.23	[89.8589, 90.604]	0.8998	0.9234	0.8774	7.9
	RF - 200	90.37	[89.7765, 90.9828]	0.9014	0.9243	0.8796	12.3

kojima je razlika u snazi autentičnih i lažnih signala malena pa je teško razlikovati autentične od lažnih signala. Ovaj rezultat dodatno potvrđuje važnost očuvanja kompletne strukture slike kod izgradnje klasifikacijskih modela.

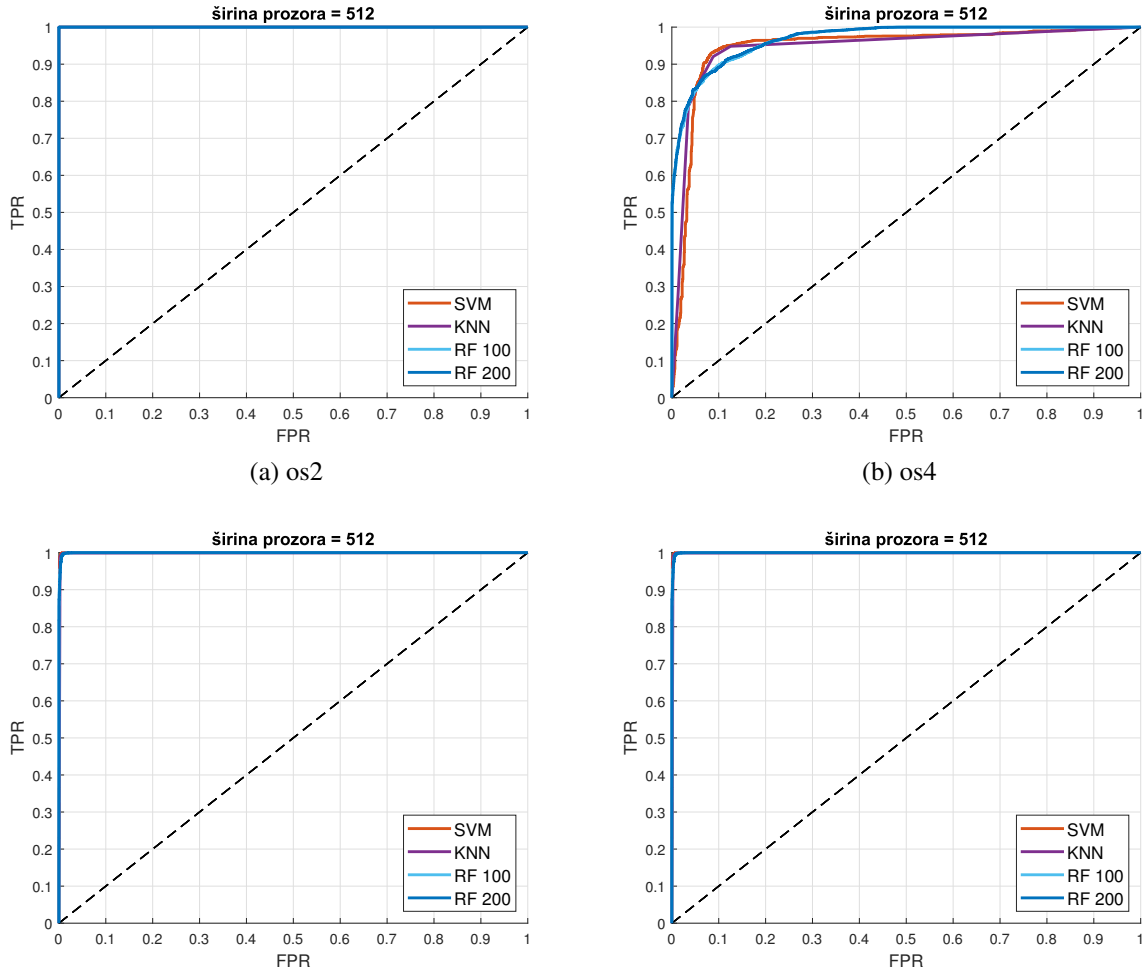
Na slici 6.28 prikazana je distribucija mjere *F1* za model KNN u slučaju kada su ulazni podaci slike. Iako i KNN pokazuje dobre rezultate za sve skupove podataka, njegovi rezultati su nešto lošiji u skupovima podataka os4 i os12 bez obzira na širinu prozora. Ako se usporedi distribucija mjere *F1* za model KNN u slučaju kada se kao ulazni podaci uzmu značajke slike kao što je prikazano na slici 6.29, performanse modela opadaju za skup podataka os4. Primjerice, mjera *F1* za model SVM u skupu podataka os4 iznosi 0.9079 dok za KNN iznosi 0.7291. Razlog ovomu može biti i to što KNN sporo uči (engl. *lazy learner*) tj. ne trenira model nego samo memorira. Iz navedenih rezultata može se zaključiti da je SVM dosljedno

Tablica 6.5: Parametri performansi modela za algoritam klasifikacije za različite modele strojnog učenja za spektrograme uz širine prozora 512 i 1024 u sustavu Galileo: a) A1 (plavo), b) A1M1 (crveno).

Skup podataka i vrsta koef.	ML model	Srednja točnost [%]	95% interval	Mjera *F1*	Odziv	Preciznost	Vrijeme izvođenja [s]
širina prozora = 512							
	SVM	99.97	[99.9431, 100]	0.9997	0.9994	1	1467
(clean + os10) approx.	KNN	100		1	1	1	1661
	RF - 100	99.97	[99.9431, 100]	0.9997	0.9994	1	1407
	RF - 200	99.97	[99.9431, 100]	0.9997	0.9994	1	1443
	SVM	98.02	[97.8472, 98.2083]	0.9805	0.9692	0.992	1950
(clean + os12) approx.	KNN	89	[88.6542, 89.4569]	0.8946	0.863	0.9285	1764
	RF - 100	93.75	[93.1077, 94.4108]	0.938	0.9317	0.9444	2113
	RF - 200	94.27	[93.7154, 94.8402]	0.9432	0.9363	0.9502	1917
	SVM	99.62	[99.5093, 99.7314]	0.9962	0.998	0.9944	4.5
(clean + os10) approx.	KNN	99.47	[99.362, 99.5824]	0.9947	0.9968	0.9926	3
	RF - 100	99.26	[99.1555, 99.4001]	0.9928	0.9941	0.9915	8.8
	RF - 200	99.37	[99.2461, 99.4946]	0.9937	0.9952	0.9922	23
	SVM	81.6	[80.8974, 82.3248]	0.8046	0.8585	0.757	3.6
(clean + os12) approx.	KNN	78.41	[77.7179, 79.1154]	0.8134	0.7163	0.9411	2.4
	RF - 100	81.62	[81.0608, 82.1799]	0.8088	0.8428	0.7774	7.6
	RF - 200	81.71	[81.0458, 82.3801]	0.8082	0.8496	0.7707	10
širina prozora = 1024							
	SVM	99.99	[99.9726, 100]	0.9999	1	0.9998	2060
(clean + os10) approx.	KNN	99.97	[99.9431, 100]	0.9997	1	0.9994	1910
	RF - 100	99.97	[99.9431, 100]	0.9997	0.9994	1	1821
	RF - 200	99.97	[99.9431, 100]	0.9997	0.9994	1	1869
	SVM	99.15	[98.9596, 99.3553]	0.9916	0.9852	0.9981	1925
(clean + os12) approx.	KNN	96.44	[96.2071, 96.6818]	0.9652	0.9456	0.9856	1499
	RF - 100	95.6	[95.1108, 96.0929]	0.9562	0.9532	0.9591	1407
	RF - 200	95.68	[95.2467, 96.1237]	0.9569	0.956	0.9578	1483
	SVM	98.98	[98.7489, 99.2141]	0.9898	0.9937	0.9859	12.7
(clean + os10) approx.	KNN	98.84	[98.6592, 99.026]	0.9884	0.9914	0.9854	2.5
	RF - 100	98.38	[98.1546, 98.6232]	0.9837	0.9934	0.9743	5.2
	RF - 200	98.48	[98.3023, 98.6606]	0.9847	0.9941	0.9754	11
	SVM	78.15	[77.6551, 78.6597]	0.7777	0.7918	0.7641	4.6
(clean + os12) approx.	KNN	75.92	[74.9903, 76.843]	0.7933	0.6948	0.9244	3.2
	RF - 100	75.62	[74.6156, 76.6436]	0.7627	0.7431	0.7833	5.5
	RF - 200	76.13	[75.5783, 76.6994]	0.765	0.7536	0.7767	8.6

precizan i stabilan, što potvrđuje njegovu superiornost kao klasifikatora za detekciju napada lažiranjem.

Slika 6.30 prikazuje ukupnu sumiranu konfuzijsku matricu kroz sve iteracije za klasifikaciju slika spektrograma širine prozora 1024 korištenjem modela SVM za skupove podataka sustava GPS os2 i os4 te Galileo os10 i os12. Sumirana matrica dobiva se na osnovi svih 10 800 testnih uzoraka kroz 15 modela od čega 5400 uzoraka pripada klasi 1 (autentični) i 5400 klasi 2 (lažni). Za skupove podataka os2 i os10 (slike 6.30a, 6.30c), model pogrešno klasificira samo jedan uzorak. Najviše pogrešno klasificiranih uzoraka (lažno pozitivne predikcije) ima skup podataka os12 (slika 6.30d). Unatoč tome što ima visoku točnost klasifikacije 99.15%, 10 autentičnih uzoraka pogrešno su klasificirani kao lažni i 81 lažni kao autentični. Za skup os4 (slika 6.30b) manji je broj pogrešno klasficiranih uzoraka u obje klase.

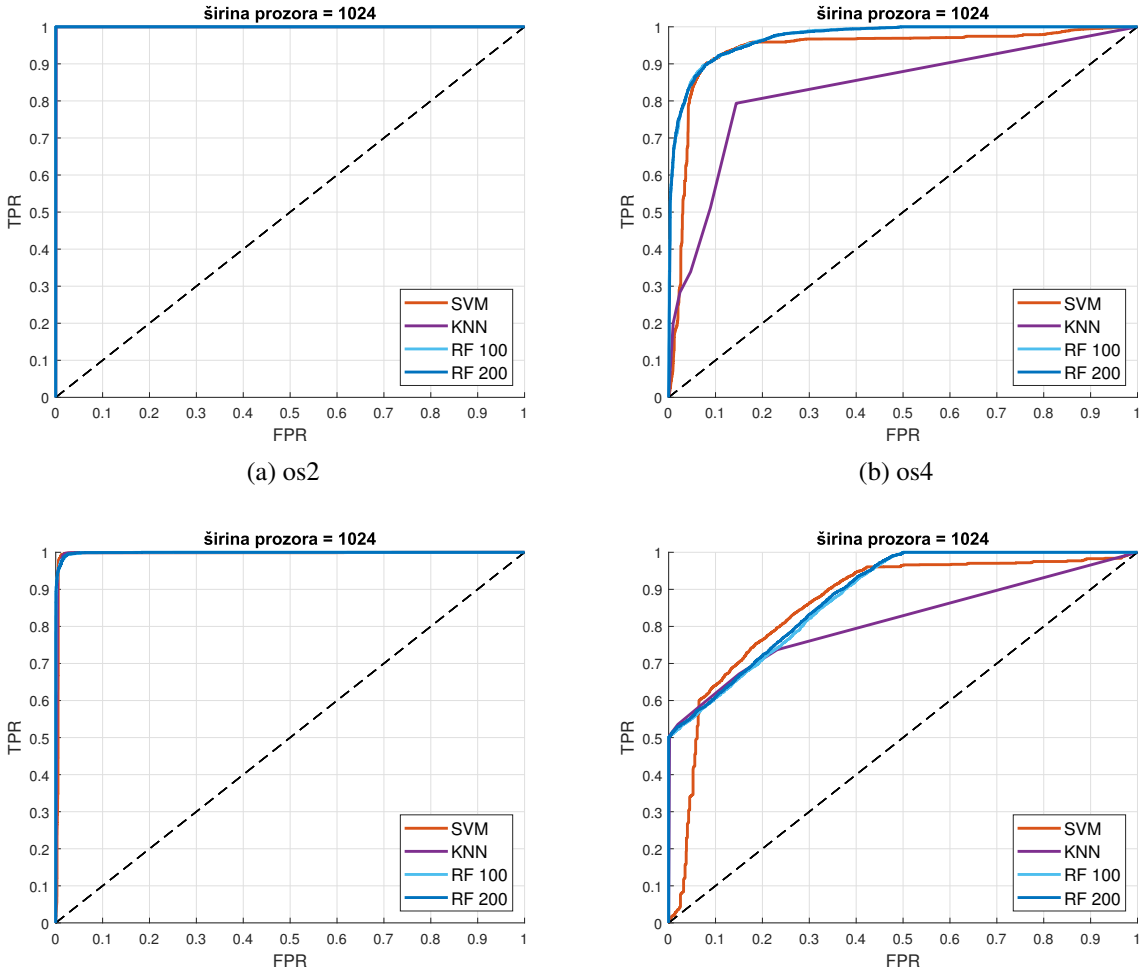


(a) os2 (b) os4

(c) os10 (d) os12

Slika 6.24: Krivulje operativnih karakteristika za različite modele strojnog učenja za klasifikaciju tipa signala za za skupove podataka sustava GPS (a) os2 i (b) os4 te Galileo (a) os10 i (b) os12 i širinu prozora 512 za značajke slike kao ulazne podatke.

Na slici 6.31 prikazane su krivulje preciznosti i odziva za skupove podataka GPS os2 i os4 te skupove podataka Galileo os10 i os12 za klasifikaciju slika spektrograma širine 512 (slike 6.31a i 6.31c) i 1024 (slike 6.31b i 6.31d) za modele strojnog učenja SVM i KNN. Iz slika je vidljivo da širina prozora od 1024 donosi znatno poboljšanje u performansama preciznosti i odziva, a posebno za skupove podataka os4 i os12. Model SVM je dosljedno najbolji za sve skupove podataka dok model KNN ima značajan pad preciznosti uz rast odziva za sve skupove podataka i širine prozora osim u slučaju kada je širina prozora 1024 za skupove podataka Galileo os10 i os12 kada ima nešto bolje performanse preciznosti. Iako je klasifikator KNN postigao izuzetno visoke performanse, rezultati klasifikatora bili su isključivo binarni (0 ili 1) što ukazuje na jednoglasno glasanje susjeda, ali i ograničava mogućnost evaluacije modela putem analize krivulje preciznosti i odziva. U ovakvim slučajevima, preporučuje se korištenje modela koji vraćaju kontinuirane vjerojatnosti za potrebe evaluacije krivulje preciznosti i odziva kao što je primjerice model SVM.



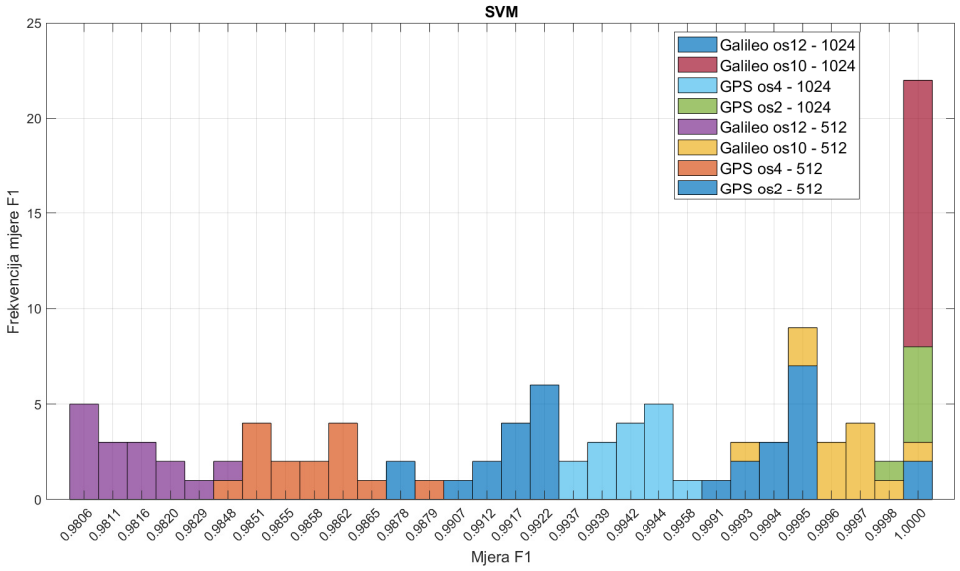
(a) os2

(b) os4

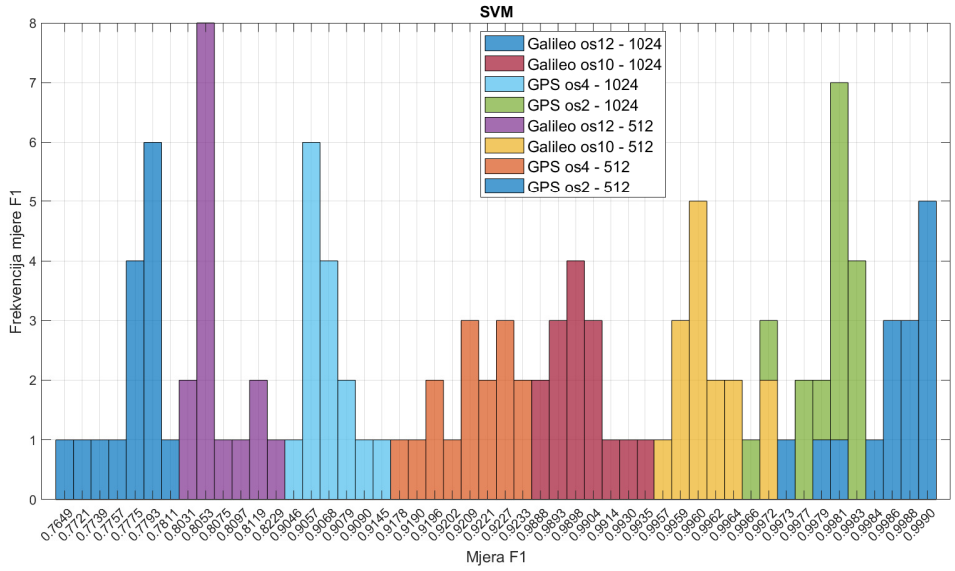
(c) os10

(d) os12

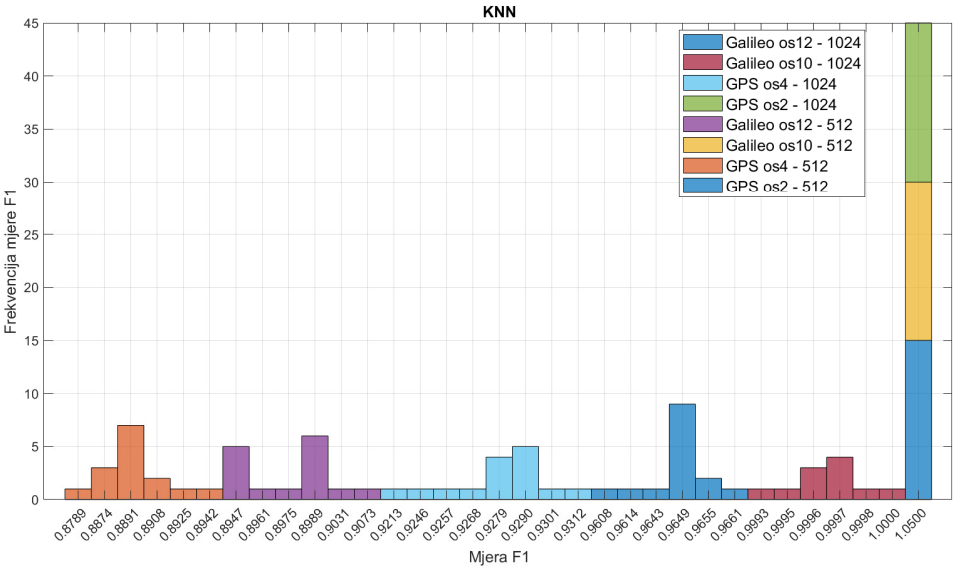
Slika 6.25: Krivulje operativnih karakteristika za različite modele strojnog učenja za klasifikaciju tipa signala za za skupove podataka sustava GPS (a) os2 i (b) os4 te Galileo (a) os10 i (b) os12 i širinu prozora 1024 za značajke slike kao ulazne podatke.



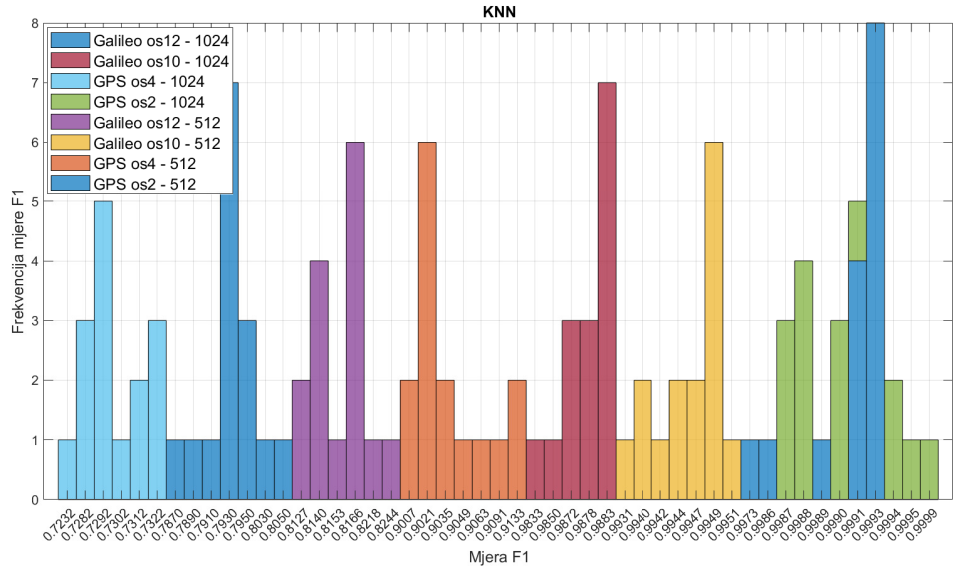
Slika 6.26: Distribucija mjere F1 za model SVM u skupovima podataka GPS os2 i os4 te Galileo os10 i os12 i širinu prozora 512 i 1024 za slike kao ulazne podatke.



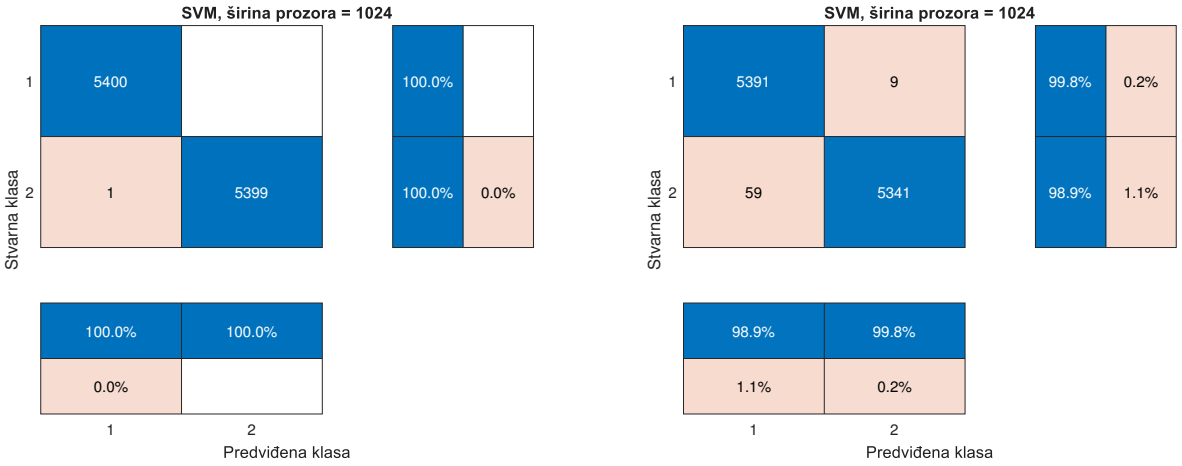
Slika 6.27: Distribucija mjere F1 za model SVM u skupovima podataka GPS os2 i os4 te Galileo os10 i os12 i širinu prozora 512 i 1024 za značajke slika kao ulazne podatke.



Slika 6.28: Distribucija mjere F1 za model KNN u skupovima podataka GPS os2 i os4 te Galileo os10 i os12 i širinu prozora 512 i 1024 za slike kao ulazne podatke.

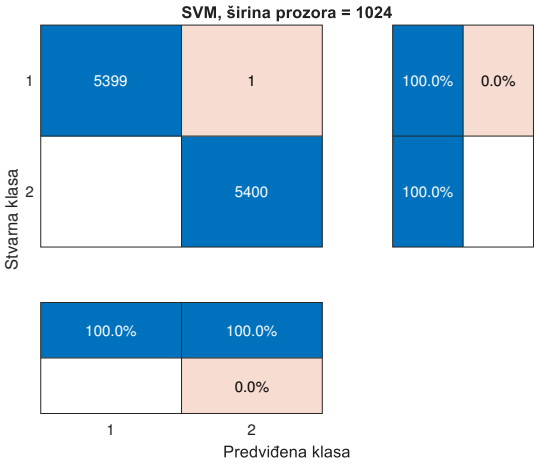


Slika 6.29: Distribucija mjere F1 za model KNN u skupovima podataka GPS os2 i os4 te Galileo os10 i os12 i širinu prozora 512 i 1024 za značajke slika kao ulazne podatke.

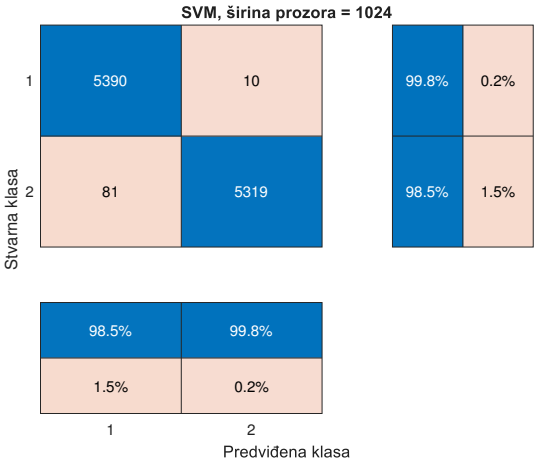


(a) os2

(b) os4

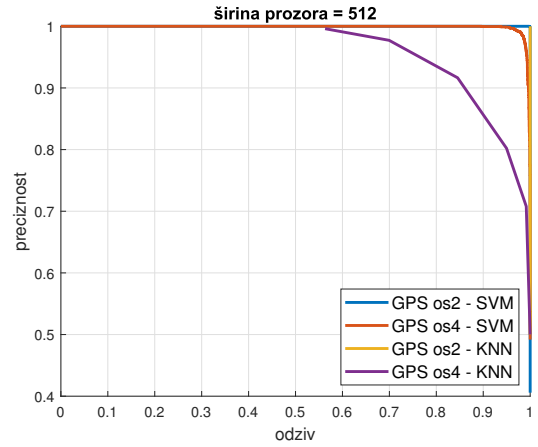


(c) os10

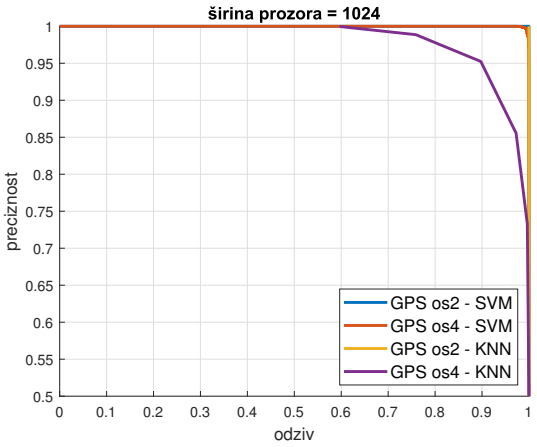


(d) os12

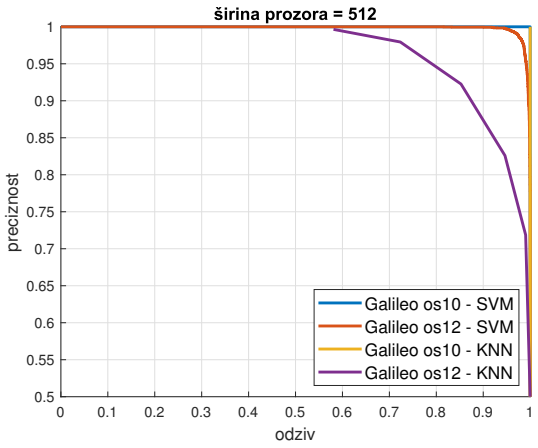
Slika 6.30: Ukupna sumirana konfuzijska matrica kroz sve iteracije za klasifikaciju slika spektrograma širine prozora 1024 korištenjem modela SVM za skupove podataka sustava GPS (a) os2i (b) os4, te Galileo (c) os10 i (d) os12.



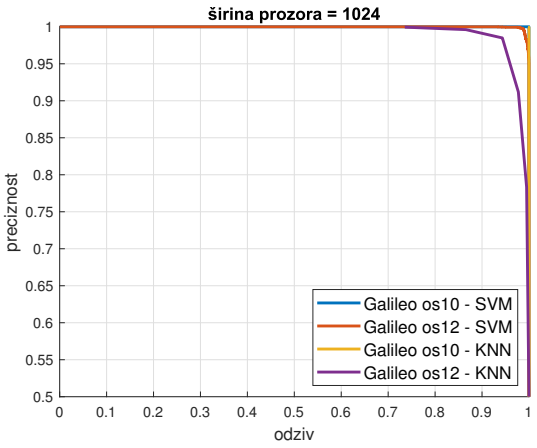
(a) GPS os2 i os4



(b) GPS os2 i os4



(c) Galileo os10 i os12



(d) Galileo os10 i os12

Slika 6.31: Krivulje preciznosti i odziva za modele SVM i KNN za skupove podataka sustava GPS i Galileo i širine prozora 512 i 1024 za slike kao ulazne podatke.

Zaključak

Globalni navigacijski satelitski sustav predstavlja jednu od najvažnijih infrastruktura u današnjem modernom svijetu jer se koriste za pozicioniranje, navigaciju i sinkronizaciju te imaju vrlo veliku primjenu u svim aspektima života. Takva rasprostranjena upotreba čini ove sustave izloženima različitim prijetnjama uključujući i zlonamjerne prijetnje kao što je napad lažiranjem. Dostupnost jeftinih uređaja kao što je softverski definirani radio povećava održivost izvođenja takvih napada. Najosjetljiviji na napade lažiranjem su pametni telefoni koji se najčešće koriste za usluge pozicioniranja i navigacije. Utjecaj napada lažiranjem na neki GNSS prijamnik se ogleda u preuzimanju navigacijskog sustava i lažiranju lokacije prijamnika što je jako opasno u slučaju preusmjeravanja aviona, brodova, dronova itd.

Učinkovito otkrivanje napada lažiranjem od ključne je važnosti za ublažavanje takvih napada. Iako su u tu svrhu predložene različite metode, to je još uvijek važna tema istraživanja. Kao jedna od metoda koja još nije dovoljno obrađena i istražena te je u svojim začecima u ovom području je metoda radio frekvencijskog otiska. Stoga je motivacija ovog istraživanja u primjeni navedene metode za detekciju napada lažiranjem. U ovom istraživanju, predložen je integrirani pristup za detekciju lažiranih signala i klasifikaciju tipa signala korištenjem kombinacije metoda radio frekvencijskog otiska i strojnog učenja u pre-korelacijskoj fazi. Korištene metode radio frekvencijskog otiska su diskretna valićna transformacija i spektrogram. Integrirani pristup je po prvi put primijenjen na statičke scenarije skupova podataka OAKBAT za dvije različite konstelacije signala, a to su GPS i Galileo. Nadalje, predložena su dva pristupa: prvi pristup koji koristi slike kao klasifikacijski podatak i drugi pristup koji koristi unaprijed izvučene značajke kao klasifikacijski podatakte su na temelju toga definirani algoritmi klasifikacije A1 i A1M s dvije inačice: A1M1 i A1M2. Definirani algoritmi su primijenjeni na različite modele strojnog učenja. Što se tiče diskretne valićne transformacije, napravljena je usporedba i evaluacija modela za dva različita tipa valića, Daubechies db4 i db8 te su u procesu generiranja skupova slika diskretne valićne transformacije korištene samo prva dekompozicijska razina i aproksimacijski koeficijenti zato što daju naviše informacija. Za spektrograme, napravljena je analiza i usporedba rezultata kada se koriste različite širine prozora. Konkretno su u istraživanju korištene dvije širine prozora 512 i 1024.

Evaluacija modela korištenjem standardnih metrika za performanse pokazuje da svi modeli imaju visoku točnost klasifikacije za obje korištene metode radio frekvencijskog otiska. Model koji se ističe sa svojim performansama za sve tipove ulaznih klasifikacijskih podataka

i skupova podataka je SVM. Pokazuje vrlo visoku točnost klasifikacije za obje vrste primijenjenih valića db4 i db8 te za spektrograme. Primjerice, za skup podataka os2 sustava GPS u slučaju korištenja valića db4 i db8 i slika kao ulaznih podataka, SVM postiže točnost od 98.94% odnosno 99.43%. Isto ponašanje modela SVM postiže se i u slučaju skupa podataka os10 sustava Galileo gdje je postignuta točnost od 93.94% za valić db4 odnosno 93.59% za db8. Kod spektrograma je postignuta točnost od 99.93% za skup podataka os2 sustava GPS te 99.97% za skup podataka os10 sustava Galileo. Očekivano, nešto niže točnosti postignute su za skupove podataka os4 i os12 u kojima je prednost razine snage lažnih u odnosu na autentične snaga manja (0.4 dB) nego za skupove os2 i os10 (10 dB). Kako bi se neki model pokazao pouzdanim, potrebno je pratiti sve parametre performansi modela.

Drugi doprinos ovog istraživanja je prijedlog modificiranog računski učinkovitijeg algoritma klasifikacije A1M s dvije inačice A1M1 i A1M2. Predloženi algoritam klasifikacije A1M koristi unaprijed izvučene značajke kao ulazne podatke za klasifikaciju u usporedbi s prvotno predloženim algoritmom A1 koji koristi slike kao ulazne podatke. Rezultati primjene algoritma A1M za različite modele strojnog učenja (SVM, KNN, RF) pokazuju smanjenje vremena izvođenja i računske složenosti za svaki pojedini model strojnog učenja. Primjerice, vrijeme izvođenja algoritma A1 za model SVM za valić db4 u skupu os2 GPS sustava iznosi približno 600 sekundi dok za algoritam A1M1 to vrijeme iznosi oko 9.5 sekundi čime je pokazano značajno smanjenje vremena izvođenja algoritma A1M u odnosu na algoritam A1. Najveću računsku složenost ima algoritam klasifikacije A1 za model KNN koji kao ulazne klasifikacijske podatke koristi slike i kod kojega je dominatna računska složenost po broju uzoraka kvadratna N^2 i u kombinaciji s kreiranjem torbe značajki predstavlja najkompleksniji pristup. S druge strane, algoritam klasifikacije A1M za model SVM koji koristi unaprijed izvučene značajke ima najmanju složenost. Računska složenost algoritma A1 koji kao ulazne klasifikacijske podatke koristi slike, može se smanjiti smanjenjem rezolucije slike, broja uzoraka, broja ponavljanja validacije te broja podskupova kao i prebacivanjem slike u sivu skalu. Kod modela SVM, korištenje linearne umjesto nelinearne jezgrene funkcije smanjuje složenost dok ju kod modela KNN smanjuje manji broj uzoraka jer se smanjuje potreba za izračunom udlajenosti između svakog testnog uzorka i uzorka za treniranje. Ako se usporede računske složenosti za dvije metode radio frekvencijskog otiska, diskretnu valićnu transformaciju i spektrogram, može se zaključiti da je računska složenost spektrograma $O(81 \times 10^9)$ veća u odnosu na diskretnu valićnu transformaciju s jednom razinom dekompozicije $O(378 \times 10^8)$ zbog višestrukih izračuna brze Fourierove transformacije. Složenost kod spektrograma se može smanjiti korištenjem manjeg broja *nfft* točaka i manje veličine prozora. Kod diskretne valićne transformacije, složenost ovisi o broju razina dekompozicije te o korištenim koeficijentima. Općenito za obje metode vrijedi da se računska složenost može smanjiti korištenjem slika manje rezolucije i jednog RGB kanala za obradu što opet ne daje dovoljno informacija za pouzdanu klasifikaciju.

Zaključno, predloženi integrirani pristup temeljen na kombinaciji RFF metoda i modela

strojnog učenja daje vrlo dobre rezultate za detekciju napada lažiranjem. Nadalje, predloženi algoritmi klasifikacije pokazuju vrlo visoke performanse u klasifikaciji lažiranih signala što ukazuje na njihovu učinkovitost i pouzdanost za detekciju napada lažiranjem te da ovo istraživanje daje značajan doprinos zajednici sustava GNSS. Nadalje, modificirani računski učinkovitiji algoritam klasfikacije A1M pokazuje da korištenje unaprijed definiranih značajki kao ulaznih podataka smanjuje računsku složenost i vrijeme izvođenja pojedinog modela strojnog učenja.

Literatura

[1] Novatel, ”What are Global Navigation Satellite Systems?”, https://novatel.com/tech-talk/an-introduction-to-gnss/what-are-global-navigation-satellite-systems-gnss, s Interneta, 15.11. 2022.

[2] ”Global Positioning System”, https://en.wikipedia.org/wiki/Global_Positioning_System#Principles, s Interneta, 5.6.2023.

[3] ”Galileo System”, https://www.gsc-europa.eu/galileo/system, s Interneta, 5.6.2023.

[4] ”GLONASS”,https://novatel.com/an-introduction-to-gnss/chapter-3-\satellitesystems/glonass-global-navigation-satellite-system-russia, s Interneta, 5.6.2023.

[5] ”BeiDou”, https://en.wikipedia.org/wiki/BeiDou, s Interneta, 5.6.2023.

[6] P. J. G. Teunissen, O. Montebruck, “Handbook of Global Navigation Satellite Systems“, Springer International Publishing, August 2017., https://link.springer.com/content/pdf/bfm:978-3-319-42928-1/1.pdf

[7] “Atomic clock“, Wikipedia, https://en.wikipedia.org/wiki/Atomic_clock, s Interneta, 15.12.2022.

[8] J. Sanz Subirana, JM. Juan Zornoza, M. Hernandez-Pajares, “GNSS signal“, Navipedia, 2011., https://gssc.esa.int/navipedia/index.php/GNSS_signal, s Interneta, 1.12.2022.

[9] “GNSS Constellations, Radio Frequencies and Signals“, https://www.tallysman.com/gnss-constellations-radio-frequencies-and-signals/, s Interneta, 5.6.2023.

[10] Y. Peng, W. A. Scales, “Ionospheric Remote Sensing with GNSS,” Encyclopedia, vol. 1, no. 4, pp. 1246–1256, Nov. 2021, MDPI, doi: 10.3390/encyclopedia1040094.

[11] E. Garbin Manfredini, “Signal processing techniques for GNSS anti-spoofing algorithms“, PhD thesis, 2017, doi: 10.6092/polito/porto/2672749.

[12] M. L. Psiaki, T. E. Humphreys and B. Stauffer, “Attackers can spoof navigation signals without our knowledge. Here’s how to fight back GPS lies,“ in IEEE Spectrum, vol. 53, no. 8, pp. 26-53, August 2016, doi: 10.1109/MSPEC.2016.7524168.

[13] M. L. Psiaki and T. E. Humphreys, “GNSS Spoofing and Detection,“ in Proceedings of the IEEE, vol. 104, no. 6, pp. 1258-1270, June 2016, doi: 10.1109/JPROC.2016.2526658.

[14] K. Radoš, "Signal processing for the purpose of eliminating interference in the receivers of the GNSS system", University of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, September 2023.

[15] K. K. Songala, S. R. Ammana, H. C. Ramachandruni and D. S. Achanta, "Simplistic Spoofing of GPS Enabled Smartphone," 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE), Bhubaneswar, India, 2020, pp. 460-463, doi: 10.1109/WIECON-ECE52138.2020.9397980.

[16] A. Rustamov, N. Gogoi, A. Minetto and F. Dovis, "Assessment of the Vulnerability to Spoofing Attacks of GNSS Receivers Integrated in Consumer Devices," 2020 International Conference on Localization and GNSS (ICL-GNSS), Tampere, Finland, 2020, pp. 1-6, doi: 10.1109/ICL-GNSS49876.2020.9115489.

[17] J. Li, W. Li, S. He, Z. Dai and Q. Fu, "Research on Detection of Spoofing Signal with Small Delay Based on KNN," 2020 IEEE 3rd International Conference on Electronics Technology (ICET), Chengdu, China, 2020, pp. 625-629, doi: 10.1109/ICET49382.2020.9119515.

[18] J. Li, X. Zhu, M. Ouyang, W. Li, Z. Chen and Q. Fu, "GNSS Spoofing Jamming Detection Based on Generative Adversarial Network," in IEEE Sensors Journal, vol. 21, no. 20, pp. 22823-22832, 15 Oct.15, 2021, doi: 10.1109/JSEN.2021.3105404.

[19] S. Semanjski, A. Muls, I. Semanjski and W. De Wilde, "Use and Validation of Supervised Machine Learning Approach for Detection of GNSS Signal Spoofing," 2019 International Conference on Localization and GNSS (ICL-GNSS), Nuremberg, Germany, 2019, pp. 1-6, doi: 10.1109/ICL-GNSS.2019.8752775.

[20] T. E. Humphreys, J.A. Bhatti, D.P. Shepard, and K.D. Wesson, “The Texas Spoofing Test Battery: Toward a Standard for Evaluating GPS Signal Authentication Techniques,” Proc. ION GNSS, Nashville, TN, 2012.

[21] E. G. Manfredini, F. Dovis, and B. Motella, “Validation of a signal quality monitoring technique over a set of spoofed scenarios,” in *Proc. 7th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC)*, Noordwijk, Netherlands, 2014. doi: 10.1109/NAVITEC.2014.7045136.

[22] A. Albright, S. Powers, J. Bonior, and F. Combs, “Oak Ridge Spoofing and Interference Test Battery (OAKBAT) - GPS“, Oak Ridge National Lab. (ORNL), Oak Ridge, United States: N. p., 2020. doi:10.13139/ORNLNCCS/1664429.

[23] “Global Positioning System Standard Positioning Service Performance Standard“, $5^{(}th)$ edition, Department of Defense, United States of America, 2020, available at: https://www.gps.gov/technical/ps/.

[24] T. E. Humphreys, B. M. Ledvina, M. L. Psiaki, B. W. O’Hanlon, P. M. Kintner: “Assessing the Spoofing Threat: Development of a Portable GPS Civilian Spoofer”, 2008 ION GNSS Conference, 2008.

[25] A. Shafique, A. Mehmood and M. Elhadef, “Detecting Signal Spoofing Attack in UAVs Using Machine Learning Models,“ IEEE Access, vol. 9, pp. 93803-93815, 2021, doi: 10.1109/ACCESS.2021.3089847.

[26] F. Gallardo and A. P. Yuste, "SCER Spoofing Attacks on the Galileo Open Service and Machine Learning Techniques for End-User Protection," in IEEE Access, vol. 8, pp. 85515-85532, 2020, doi: 10.1109/ACCESS.2020.2992119.

[27] A. Siemuri, K. Selvan, H. Kuusniemi, P. Valisuo and M. S. Elmusrati, "A Systematic Review of Machine Learning Techniques for GNSS Use Cases," in IEEE Transactions on Aerospace and Electronic Systems, vol. 58, no. 6, pp. 5043-5077, Dec. 2022, doi: 10.1109/TAES.2022.3219366.

[28] G. Aissou, H. O. Slimane, S. Benouadah and N. Kaabouch, “Tree-based Supervised Machine Learning Models For Detecting GPS Spoofing Attacks on UAS,“ 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2021, pp. 0649-0653, doi: 10.1109/UEMCON53757.2021.9666744.

[29] B. Pardhasaradhi, R. R. Yakkati and L. R. Cenkeramaddi, “Machine Learning-Based Screening and Measurement to Measurement Association for Navigation in GNSS Spoofing Environment,“ in IEEE Sensors Journal, vol. 22, no. 23, pp. 23423-23435, 1 Dec.1, 2022, doi: 10.1109/JSEN.2022.3214349.

[30] R. R. Yakkati, B. Pardhasaradhi, J. Zhou and L. R. Cenkeramaddi, “A Machine Learning based GNSS Signal Classification,“ 2022 IEEE International Symposium on Smart Electronic Systems (iSES), Warangal, India, 2022, pp. 532-535, doi: 10.1109/iSES54909.2022.00116.

[31] A. Broumandan, S. Kennedy, J. Schleppe: “Demonstration of a Multi-Layer Spoofing Detection Implemented in a High Precision GNSS Receiver”, 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS), 2020.

[32] D. -K. Lee et al., “Detection of GNSS Spoofing using NMEA Messages,“ 2020 European Navigation Conference (ENC), Dresden, Germany, 2020, pp. 1-10, doi: 10.23919/ENC48637.2020.9317470.

[33] J. Spravil, C. Hemminghaus, M. von Rechenberg, E. Padilla, and J. Bauer, “Detecting maritime GPS spoofing attacks based on NMEA sentence integrity monitoring,” *Journal of Marine Science and Engineering*, vol. 11, no. 5, p. 928, 2023. doi: 10.3390/jmse11050928.

[34] V. Truong, A. Vervisch-Picois, J. Rubio Hernan, and N. Samama, “Characterization of the ability of low-cost GNSS receiver to detect spoofing using clock bias,” *Sensors*, vol. 23, pp. 1–18, 2024. doi: 10.3390/s23052735.

[35] Q. Yang and Y. Chen, “A GPS spoofing detection method based on compressed sensing,” presented at the *IEEE Int. Conf. on Signal Processing, Communications and Computing (ICSPCC)*, Xi’an, China, Oct. 25–27, 2022, pp. 1–5. doi: 10.1109/ICSPCC55723.2022.9984624.

[36] Y. Zhao, F. Shen, D. Xu, and Z. Meng, “A coprime array-based technique for spoofing detection and DoA estimation in GNSS,” *IEEE Sensors Journal*, vol. 22, pp. 22828–22835, 2022.

[37] S. Chen, S. Ni, T. Lei, L. Cheng, and X. Song, “GNSS spoofing detection via the intersection angle between two directions of arrival in a single rotating antenna,” *Sensors*, vol. 24, no. 4, p. 1116, 2024.

[38] E. G. Manfredini, D. M. Akos, Y.-H. Chen, S. Lo, T. Walter, and P. Enge, "Effective GPS Spoofing Detection Utilizing Metrics from Commercial Receivers," Proceedings of the 2018 International Technical Meeting of The Institute of Navigation, Reston, Virginia, January 2018, pp. 672-689.

[39] N. Spens, D.-K. Lee, F. Nedelkov, and D. Akos, "Detecting GNSS Jamming and Spoofing on Android Devices", NAVIGATION: Journal of the Institute of Navigation September 2022, 69 (3) navi.537; doi: 10.33012/navi.537.

[40] Z. Chen, J. Li, J. Li, X. Zhu and C. Li, “GNSS Multiparameter Spoofing Detection Method Based on Support Vector Machine,“ in IEEE Sensors Journal, vol. 22, no. 18, pp. 17864-17874, 15 Sept.15, 2022, doi: 10.1109/JSEN.2022.3193388.

[41] M. Turner, S. Wimbush, C. Enneking and A. Konovaltsev, “Spoofing Detection by Distortion of the Correlation Function,“ 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS), Portland, OR, USA, 2020, pp. 566-574, doi: 10.1109/PLANS46316.2020.9110173.

[42] B. Yang, M. Tian, Y. Ji, J. Cheng, Z. Xie and S. Shao, “Research on GNSS Spoofing Mitigation Technology Based on Spoofing Correlation Peak Cancellation,“ in IEEE Communications Letters, vol. 26, no. 12, pp. 3024-3028, Dec. 2022, doi: 10.1109/LCOMM.2022.3204944.

[43] W. Zhou, Z. Lv, G. Li, B. Jiao, and W. Wu, “Detection of spoofing attacks on global navigation satellite systems using Kolmogorov–Smirnov test-based signal quality monitoring method,” *IEEE Sensors Journal*, vol. 24, pp. 10474–10490, 2024.

[44] G. Aissou, S. Benouadah, H. El Alami and N. Kaabouch, “Instance-based Supervised Machine Learning Models for Detecting GPS Spoofing Attacks on UAS,“ 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2022, pp. 0208-0214, doi: 10.1109/CCWC54503.2022.9720888.

[45] S. Semanjski, I. Semanjski, W. De Wilde, and A. Muls, “Use of Supervised Machine Learning for GNSS Signal Spoofing Detection with Validation on Real-World Meaconing and Spoofing Data—Part I,” Sensors, vol. 20, no. 4, p. 1171, Feb. 2020, doi: 10.3390/s20041171.

[46] S. Semanjski, I. Semanjski, W. De Wilde, and S. Gautama, “Use of Supervised Machine Learning for GNSS Signal Spoofing Detection with Validation on Real-World Meaconing and Spoofing Data—Part II,” Sensors, vol. 20, no. 7, p. 1806, Mar. 2020, doi: 10.3390/s20071806.

[47] L. Meng, L. Yang, W. Yang, and L. Zhang, “A Survey of GNSS Spoofing and Anti-Spoofing Technology,” Remote Sensing, vol. 14, no. 19, p. 4826, Sep. 2022, doi: 10.3390/rs14194826.

[48] L. Zhang, L. Wang, R. Wu, and X. Zhuang, “A new approach for GNSS spoofing detection using power and signal quality monitoring,” *Measurement Science and Technology*, vol. 35, no. 12, pp. 1–18, 2024. doi: 10.1088/1361-6501/ad7629.

[49] A. Elango, S. Ujan and L. Ruotsalainen, “Disruptive GNSS Signal detection and classification at different Power levels Using Advanced Deep-Learning Approach,“ 2022 International Conference on Localization and GNSS (ICL-GNSS), Tampere, Finland, 2022, pp. 1-7, doi: 10.1109/ICL-GNSS54081.2022.9797026.

[50] P. Borhani-Darian, H. Li, P. Wu, and C. Pau, “Detecting GNSS spoofing using deep learning,” *EURASIP Journal on Advances in Signal Processing*, vol. 2024, no. 14, 2024. doi: 10.1186/s13634-023-01103-1.

[51] M. Marchand, A. Toumi, G. Seco-Granados, and J. A. Lopez-Salcedo, “Machine learning assessment of anti-spoofing techniques for GNSS receivers,” presented at the *WIPHAL 2023: Work-in-Progress in Hardware and Software for Location Computation*, CEUR Workshop Proc., Castellon, Spain, June 6–8, 2023.

[52] I. E. Mehr and F. Dovis, “A deep neural network approach for classification of GNSS interference and jammer,” *IEEE Transactions on Aerospace and Electronic Systems*, 2024. doi: 10.1109/TAES.2024.3462662.

[53] K. S. Kuciapinski, M. A. Temple, and R. W. Klein, “ANOVA-based RF DNA analysis: Identifying significant parameters for device classification,” presented at the *Int. Conf. on Wireless Information Networks and Systems (WINSYS)*, Athens, Greece, July 26–28, 2010, pp. 1–6.

[54] B. Danev, D. Zanetti, and S. Capkun, “On physical-layer identification of wireless devices,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 1, pp. 1–29, 2012. doi: 10.1145/2379776.2379782.

[55] G. Baldini, R. Giuliani, G. Steri, and R. Neisse, “Physical layer authentication of Internet of Things wireless devices through permutation and dispersion entropy,” presented at the *Global Internet of Things Summit (GIoTS)*, Geneva, Switzerland, June 6–9, 2017, pp. 1–6.

[56] G. Baldini, C. Gentile, R. Giuliani, and G. Steri, “Comparison of techniques for radiometric identification based on deep convolutional neural networks,” *Electronics Letters*, vol. 55, pp. 90–92, 2019. doi: 10.1049/el.2018.6229.

[57] M. K. M. Fadul, D. R. Reising, and M. Sartipi, “Identification of OFDM-based radios under Rayleigh fading using RF-DNA and deep learning,” *IEEE Access*, vol. 9, pp. 17100–17113, 2021. doi: 10.1109/ACCESS.2021.3053491.

[58] S. Gahlawat, *Investigation of RF Fingerprinting Approaches in GNSS*, Ph.D. dissertation, Tampere Univ., Tampere, Finland, 2020.

[59] W. Wang, I. Aguilar Sanchez, G. Caparra, A. McKeown, T. Whitworth, and E. S. Lohan, “A survey of spoofer detection techniques via radio frequency fingerprinting with focus on the GNSS pre-correlation sampled data,” *Sensors*, vol. 21, no. 9, 2021. doi: 10.3390/s21093012.

[60] R. Morales-Ferre, W. Wang, A. Sanz-Abia, and E.-S. Lohan, “Identifying GNSS Signals Based on Their Radio Frequency (RF) Features—A Dataset with GNSS Raw Signals Based on Roof Antennas and Spectracom Generator,” Data, vol. 5, no. 1, p. 18, Feb. 2020, doi: 10.3390/data5010018.

[61] W. Wang, E. S. Lohan, I. A. Sanchez, and G. Caparra, “Pre-correlation and post-correlation RF fingerprinting methods for GNSS spoofer identification with real-field measurement data,” presented at the *10th Workshop on Satellite Navigation Technology (NAVITEC)*, Noordwijk, Netherlands, Apr. 4–8, 2022, pp. 1–10. doi: 10.1109/NAVITEC53682.2022.9847540.

[62] X. Zhang, Y. Huang, Y. Tian, M. Lin, and J. An, “Noise-like features assisted GNSS spoofing detection based on convolutional autoencoder,” *IEEE Sensors Journal*, vol. 23, no. 20, pp. 25473–25486, 2023. doi: 10.1109/JSEN.2023.3311799.

[63] C. Guo and Z. Yu, “Robust RF fingerprint extraction scheme for GNSS spoofing detection,” presented at the *36th Int. Tech. Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+)*, Denver, CO, USA, Sept. 11–15, 2023, pp. 199–205. doi: 10.33012/2023.19302.

[64] J. Magiera, “A Multi-Antenna Scheme for Early Detection and Mitigation of Intermediate GNSS Spoofing,” Sensors, vol. 19, no. 10, p. 2411, May 2019, doi: 10.3390/s19102411.

[65] L. Junzhi *et al.*, “Performance testing and analysis of a new GNSS spoofing detection method in different spoofing scenarios,” *IEEE Access*, vol. 13, pp. 54779–54793, 2025. doi: 10.1109/ACCESS.2025.3553475.

[66] T. Bašić, “Globalni navigacijski satelitski sustavi, Systems - GNSS, Praktični primjeri naknadne obrade multi-GNSS mjerenja otvorenim i komercijalnim programima”, Stručno usavršavanje HKOIG 2020 – GF (11).

[67] T. T. Khoei, A. Gasimova, M. A. Ahajjam, K. A. Shamaileh, V. Devabhaktuni and N. Kaabouch, “A Comparative Analysis of Supervised and Unsupervised Models for Detecting GPS Spoofing Attack on UAVs,“ 2022 IEEE International Conference on Electro Information Technology (eIT), Mankato, MN, USA, 2022, pp. 279-284, doi: 10.1109/eIT53891.2022.9813826.

[68] A. Rustamov, A. Minetto and F. Dovis, "Improving GNSS Spoofing Awareness in Smartphones via Statistical Processing of Raw Measurements," in IEEE Open Journal of the Communications Society, vol. 4, pp. 873-891, 2023, doi: 10.1109/OJCOMS.2023.3260905.

[69] L. Huang and Q. Yang, “Low-cost GPS simulator GPS spoofing by SDR,” in Proceedings of DEFCON, 2015.

[70] M. Brkić, “Detekcija lažnog signala u sustavu GNSS“, Diplomski rad, Fakultet elektrotehnike, strojarstva i brodogradnje (FESB), Split, 2022.

[71] “Software-Defined GPS Signal Simulator,” Accessed: March 20, 2023. [Online]. Available: https://github.com/osqzss/gps-sdr-sim.

[72] “MIT Licence.” Accessed: March 20, 2023. [Online]. Available: https://opensource.org/licenses/mit-license.php.

[73] Great Scott Gadgets, “HackRF One”, accessed: April 15, 2023. [Online]. Available: https://greatscottgadgets.com/hackrf/.

[74] S. Ceccato, F. Formaggio, G. Caparra, N. Laurenti and S. Tomasin, "Exploiting side-information for resilient GNSS positioning in mobile phones," 2018 IEEE/ION Position, Location and Navigation Symposium (PLANS), Monterey, CA, USA, 2018, pp. 1515-1524, doi: 10.1109/PLANS.2018.8373546.

[75] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen, and G. Lachapelle, “GPS vulnerability to spoofing threats and a review of antispoofing techniques,” Int. J. Navig. Observ., vol. 2012, Jul. 2012, Art. no. 127072, doi: 10.1155/2012/127072.

[76] Y.-S. Lee, J. S. Yeom, and B. C. Jung, “A novel array antenna-based GNSS spoofing detection and mitigation technique,” presented at the *IEEE 20th Consumer Communications & Networking Conf. (CCNC)*, Las Vegas, NV, USA, Jan. 8–11, 2023, pp. 489–492. doi: 10.1109/CCNC51644.2023.10060423.

[77] T. E. Humphreys, "Detection Strategy for Cryptographic GNSS Anti-Spoofing," in IEEE Transactions on Aerospace and Electronic Systems, vol. 49, no. 2, pp. 1073-1090, APRIL 2013, doi: 10.1109/TAES.2013.6494400.

[78] A. Jafarnia-Jahromi, A. Broumandan, J. Nielsen, and G. Lachapelle, “GPS spoofer countermeasure effectiveness based on using signal strength noise powerand C/No observables,” Int. J. Satellite Commun. Netw., vol. 30, pp. 181–191, Jul. 2012, doi: 10.1002/sat.1012.

[79] N. Spens, D.-K. Lee, and D. Akos, “An application for detecting GNSS jamming and spoofing,” in Proc. 33rd Int. Tech. Meeting Satellite Div. Inst. Navig. (ION GNSS+), Sep. 2021, pp. 1981–1988, doi: 10.33012/2021.18027.

Method Using Dual Antennas", The 13th Asian Control Conference (ASCC 2022), Jeju Island, Korea, May 4-7, 2022.

[80] A. Géron, "Hands-On Machine Learning with Scikit-Learn and TensorFlow", O’Reilly, Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, 1st edn., 2017.

[81] J. Šnajder, "Strojno učenje: 8. Stroj potpornih vektora", UNIZG FER, ak. god. 2020./2021., v3.0.

[82] J. Šnajder, "Strojno učenje 1, 21. Vrednovanje modela", UNIZG FER, ak. god. 2022./2023.

[83] C. Cortes, V. Vapnik, "Support-vector networks", Machine Learning 20, 1995, pp. 273–297, doi: 10.1007/BF00994018.

[84] E. Fix and J. L. Hodges, Discriminatory analysis. nonparametric discrimination: Consistency properties, International Statistical Review / Revue Internationale de Statistique, 57, 3, 238, Dec. 1989.

[85] T. Cover and P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory, 13, 1, 21–27, 1967.

[86] V. Chugani, "Minkowski Distance: A Comprehensive Guide", Tutorial, DataCamp, October 9, 2024.

[87] J. Čulić Gambiroža, *Machine learning methods for efficient data reduction and reconstruction in the concept of Internet of Things*, Ph.D. dissertation, Univ. of Split, Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, 2023.

[88] R. Shah, "Introduction to k-Nearest Neighbors (kNN) Algorithm", Published in Artificial Intelligence in Plain English, March 3, 2021.

[89] T. Šmuc, "Strojno učenje 2", Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet, Zagreb, 2011.

[90] MLMath, "Math behind Decision Tree Algorithm", Medium, 2019.

[91] Anshul,"What is Decision Tree?", Analytics Vidhya, 2025.

[92] J.R. Quinlan, "Induction of decision trees", Machine Learning 1, 81–106, 1986., doi: 10.1007/BF00116251.

[93] I. Fabijanić, "Vrednovanje postupka semantičke segmentacije temeljenog na slučajnim šumama, Završni rad, Sveučilište u Zagrebu, Fakultet elektrtehnike i računarstva, 2015.

[94] T. K. Ho, The random subspace method for constructing decision forests, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, 8, 832–844, 1998.

[95] L. Breiman, J. H. Friedman, R. A. Olshen and C. J. Stone, Classification And Regression Trees, Routledge, Oct. 2017.

[96] L. Breiman, Machine Learning, 45, 1, 5–32, 2001.

[97] C. Wang, Y. Sun, W. Wang, H. Liu and B. Wang, Hybrid intrusion detection system based on combination of random forest and autoencoder, Symmetry, 15, 3, 2023.

[98] P. Cuff, “Lecture 7 ELE 301: Signals and Systems,” Online lecture, 2011. [Online]. Available: https://www.princeton.edu/ cuff/ele301/ (Accessed: Oct. 28, 2024).

[99] “Short-time Fourier transform (STFT),” Online. [Online]. Available: <URLhere> (Accessed: Oct. 28, 2024).

[100] S. Mallat, *A Wavelet Tour of Signal Processing*. United States of America: Academic Press, 2008.

[101] I. Daubechies, *Ten Lectures on Wavelets*. United States of America: SIAM, 1992.

[102] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. United States of America: Wellesley-Cambridge Press, 1996.

[103] K. Radoš, M. Brkić and D. Begušić, "GNSS Signal Classification based on Machine Learning Methods," 2024 47th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, 2024, pp. 806-811, doi: 10.1109/MIPRO60963.2024.10569174.

[104] K. Radoš, M. Brkić and D. Begušić, "Vulnerability of Smartphones on GNSS Simplistic Spoofing Attack," 2024 47th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia, 2024, pp. 812-817, doi: 10.1109/MIPRO60963.2024.10569351.

[105] M. Balić, K. Radoš and Z. Blažević, "GNSS Spoofing Attack in Real-time Static and Dynamic Scenarios," 2024 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2024, pp. 1-6, doi: 10.23919/SoftCOM62040.2024.10721889.

[106] K. Radoš, M. Brkić and D. Begušić, "Recent Advances on Jamming and Spoofing Detection in GNSS," Sensors 2024, 24, 4210. https://doi.org/10.3390/s24134210

[107] M. Foruhandeh, A. Z. Mohammed, G. Kildow, R. Gerdes and R. Berges, "SatGrid Dataset, realtime genuine and spoofing traces of GPS signals collected at different geographical locations, times and environmental conditions", University Libraries, Virginia Tech. Dataset, 2020, doi: 10.7294/SE62-7X13.

[108] NASA's Archive of Space Geodesy Data, https://cddis.nasa.gov/Data and Derived Products/GNSS/broadcast ephemeris data.html.

[109] "GnssLogger App" (Version 3.0.6.1) [Mobile App]." 2023. [Online].https://play.google.com/store/apps/details?id=com.google. android.apps.location.gps.gnsslogger&hl=en_US

[110] A. Dharwal, "Complete Guide to Understanding Precision and Recall Curves", Deep Tech, 2021.

[111] "Compare Deep Learning Models Using ROC Curves", Mathworks, Matlab, 2024.

[112] D. N. Joanes and C. A. Gill, "Comparing measures of sample skewness and kurtosis," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 47, no. 1, pp. 183–189, 1998. doi: 10.1111/1467-9884.00122.

[113] T. Giannakopoulos and A. Pikrakis, "Audio Features," in *Introduction to Audio Analysis: A MATLAB Approach*, Elsevier, 2014, pp. 59–103. doi: 10.1016/B978-0-08-099388-1.00004-2.

[114] M. Pilanci, "EE269 Signal Processing for Machine Learning – Lecture 3 Part II: Spectral Features," Stanford University, 2021. [Online]. Available: https://web.stanford.edu/class/ee269/Lecture3_spectral_features.pdf [Accessed: Feb. 22, 2025].

[115] N. Kulkarni, "Use of complexity based features in diagnosis of mild Alzheimer disease using EEG signals," *International Journal of Information Technology*, vol. 10, pp. 59–64, 2018. doi: 10.1007/s41870-017-0057-0.

[116] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2008. [Online]. Available: https://www.amazon.com/Digital-Image-Processing-Rafael-Gonzalez/dp/013168728X

[117] W. K. Pratt, *Digital Image Processing: PIKS Inside*, 3rd ed. New York: John Wiley & Sons, 2001.

[118] W. K. Pratt, *Digital Image Processing: PIKS Scientific Inside*, 4th ed. Hoboken, NJ: Wiley-Interscience, 2007.

[119] A. C. Bovik, Ed., *Handbook of Image and Video Processing*. San Diego, CA: Academic Press, 2000.

[120] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973. doi: 10.1109/TSMC.1973.4309314.

[121] M. Hall-Beyer, "GLCM Texture: A Tutorial v. 3.0," Univ. of Calgary, Mar. 2017. [Online]. Available: https://prism.ucalgary.ca/handle/1880/51900 [Accessed: Mar. 15, 2025].

[122] S. Shalev-Shwartz and S. Ben-David, "Support vector machines," in *Understanding Machine Learning: From Theory to Algorithms*, ch. 15, pp. 315–340, Cambridge Univ. Press, 2014. [Online]. Available: https

[123] L. Bottou and C.-J. Lin, "Support vector machine solvers," in *Large Scale Kernel Machines*, L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, Eds. Cambridge, MA: MIT Press, 2007, pp. 301–320. [Online]. Available: https://leon.bottou.org/papers/lskm-2007.

[124] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002. [Online]. Available: https://mitpress.mit.edu/9780262536578/learning-with-kernels/

[125] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2001.

[126] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324. [Online]. Available: https://link.springer.com/article/10.1023/A:1010933404324

[127] I. W. Tsang, J. T. Kwok, and P. M. Cheung, "Core vector machines: Fast SVM training on very large data sets," *Journal of Machine Learning Research*, vol. 6, pp. 363–392, 2005.

[128] N. Cesa-Bianchi, C. Gentile, and L. Zaniboni, "Analysis of learning algorithms for structured prediction," *Machine Learning*, vol. 101, no. 1–3, pp. 239–281, 2015.

[129] Y. Bengio, "Why is kernelized SVM much slower than linear SVM?", Quora, 2011. [Online]. Available: https://www.quora.com/Why-is-kernelized-SVM-much-slower-than-linear-SVM/answer/Yoshua-Bengio?ch=10&share=69b63c74&srid=uuoZN

[130] W. Li, Z. Huang, R. Lang, H. Qin, K. Zhou, and Y. Cao, "A real-time interference monitoring technique for GNSS based on a twin support vector machine method," Sensors, vol. 16, no. 3, p. 329, Mar. 2016, doi: 10.3390/s16030329.

[131] C. Savas, and F. Dovis, "The Impact of Different Kernel Functions on the Performance of Scintillation Detection Based on Support Vector Machines," Sensors, vol. 19, no. 23, p. 5219, 28. Oct. 2019, doi: 10.3390/s19235219.

[132] J.-S. Chen, and C.-M. Kuo, "An Efficient GNSS Coordinate Classification Strategy with an Adaptive KNN Algorithm for Epidemic Management," Mathematics, vol. 12, no. 4, p. 536, Feb. 2024, doi:?10.3390/math12040536.

[133] X. Zheng et al., "Full Parameter Time Complexity (FPTC): A Method to Evaluate the Running Time of Machine Learning Classifiers for Land Use/Land Cover Classification," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 2222-2235, 2021, doi: 10.1109/JSTARS.2021.3050166.

[134] S.-H. Ieng, E. Lehtonen, and R. Benosman, “Complexity analysis of iterative basis transformations applied to event-based signals,” *Front. Neurosci.*, vol. 12, p. 373, Jun. 2018, doi:10.3389/fnins.2018.00373.

[135] B. McFee, "9.1. Framing", Digital Signals Theory, 2022.

[136] "Spectrogram", [Online]. Available: https://www.mathworks.com/help/signal/ref/spectrogram.html

[137] M. Hasegawa-Johnson, "Lecture 5: Short-Time Fourier Transform and Filterbanks", ECE 417: Multimedia Signal Processing, Fall 2020, [Online]. Available: https://courses.grainger.illinois.edu/ece417/fa2020/slides/lec05.pdf

[138] M. O’Byrne, B. Ghosh, V. Pakrashi, and F. Schoefs, "Texture Analysis based Detection and Classification of Surface Features on Ageing Infrastructure Elements", BCRI2012 Bridge & Concrete Research in Ireland, 2012, Cork, Ireland.

Životopis

Katarina Babić

Katarina Babić rođena je 2. rujna 1992. godine u Splitu. Nakon završene Opće gimnazije u Tomislavgradu, 2011. upisuje studij Elektrotehnike i informacijske tehnologije na Fakultetu elektrotehnike, strojarstva i brodogradnje u Splitu. Diplomirala je s izvrsnim uspjehom u srpnju 2016. godine pod mentorstvom profesora Dinka Begušića.

Od veljače 2017. do studenog 2018. bila je zaposlena kao stručni suradnik I. vrste na Fakultetu elektrotehnike, strojarstva i brodogradnje u Splitu. Od studenog 2018. pa do danas zaposlena je na mjestu mlađeg istraživača na Fakultetu elektrotehnike, strojarstva i brodogradnje u Splitu. U studenom 2018. upisuje i poslijediplomski studij Elektrotehnike i informacijske tehnologije na Fakultetu elektrotehnike, strojarstva i brodogradnje u Splitu. Od 2017. godine radi na razvoju i održavanju informacijskih sustava za organizaciju znanstvenih skupova i znanstvene publikacije u području informacijske i komunikacijske tehnologije. Također, sudjeluje u organizacijskim i administracijskim poslovima za međunardonu znanstvenu konferenciju SoftCOM i međunarodni znanstveni časopis JCOMSS.

U okviru nastavnih djelatnosti na FESB-u, u ulozi asistenta sudjelovala je u izvođenju laboratorijskih i auditornih vježbi iz kolegija Inženjerska grafika i prezentacija, Komunikacijski sustavi i protokoli, Optički komunikacijski sustavi, Komunikacijski sustavi.

Njezini znanstveni interesi uključuju satelitske i navigacijske sustave s naglaskom na globalni navigacijski satelitski sustav i otkrivanje napada lažiranjem, obradu signala, strojno učenje. Ukupno je objavila 19 radova. Od toga je 7 radova objavljenih tijekom poslijediplomskog studija (2 znanstvena rada u časopisima A kategorije, 4 znanstvena rada i 1 stručni rad na međunarodnim znanstvenim skupovima). Osim toga, objavila je još i 4 znanstvena rada u znanstvenim časopisima, 2 stručna rada te 6 popularizacijskih radova.

Curriculum Vitae

Katarina Babić

Katarina Babić was born on September 2, 1992 in Split. After graduating from high school in Tomislavgrad in 2011, she enrolled the study of electrical engineering and information technology at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture at the University of Split. She received MSc. degree with the excellent marks in 2016, with thesis leaded by mentor prof. dr. sc. Dinko Begušić.

From February 2017 to November 2018, she has been employed as a first-class professional associate at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture in Split. From November 2018 to the present, she has been employed as a junior researcher at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture in Split. In November 2018, she enrolled in the postgraduate study of Electrical Engineering and Information Technology at the Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture in Split. Since 2017, she has been working on the development and maintenance of the information systems for the organization of scientific conferences and scientific publications in the field of information and communication technology. She also participates in organizational and administrative tasks for the international scientific conference SoftCOM and the international scientific journal JCOMSS.

As part of her teaching activities at FESB, she participated as a teaching assistant in conducting laboratory and auditory exercises in the courses Engineering Graphics and Presentation, Communication Systems and Protocols, Optical Communication Systems, Communication Systems.

Her scientific interests include Satellite and Navigation Systems with the emphasize on Global Navigation Satellite System and Spoofing Attack Detection, Signal Processing, Machine Learning. She published 19 papers in total. Of these, 7 papers were published during PhD study (2 scientific papers in a journals of A category, 4 scientific papers and 1 professional paper at the international scientific conferences). In addition, she also published 4 scientific papers in scientific journals, 2 professional papers and 6 popularization papers.