

Lokalizacija na prostem s pomočjo radiofrekvenčnih odtisov tehnologije BLE v testnem omrežjem LOG-a-TEC

Blaž Bertalanč¹, Grega Morano^{1,2}, Gregor Cerar¹

¹Department of Communication Systems, Jožef Stefan Institute, Slovenia

²Jožef Stefan International Postgraduate School, Slovenia

E-pošta: {blaz.bertalanic | grega.morano | gregor.cerar }@ijs.si

LOG-a-TEC Testbed outdoor fingerprinting localization using BLE beacons

While the Global Positioning System (GPS) provides high accuracy it places a significant strain on the device's battery. In search of alternative techniques for outdoor localization, several approaches have been explored and recently Bluetooth Low Energy (BLE) is becoming a viable alternative to GPS for outdoor localization. Despite its popularity, access to open-source datasets for outdoor localization is limited. In this paper, we present a new openly available BLE fingerprint-based localization dataset that has been collected on LOG-a-TEC testbed at the Jožef Stefan Institute, Ljubljana, Slovenia. The presented dataset was also used to develop a machine learning model that is capable of correctly classifying fingerprints with an average F1-score of 96.1%. We also provide insight into the importance of each node to the performance of the model.

1 Uvod

Lokalizacijske storitve (ang. localization-based services, LSB), so postale sestavni del našega vsakdana, tako za določanje lokacije v zaprtih prostorih kot na prostem. Najpogosteje uporabljena tehnologija za zunanjo lokalizacijo je globalni sistem pozicioniranja (ang. Global Positioning System, GPS), ki se uporablja v številnih aplikacijah. Kljub splošno znanim prednosti tehnologije GPS, je njena glavna pomanjkljivost ta, da zahteva neposredno vidljivost vsaj štirih satelitov, kar lahko v primeru nizkega razmerja med signalom in šumom (ang. signal to noise ratio, SNR) omeji delovanje na območjih s slabšo vidljivostjo satelitov, kot so nekatera mestna območja z visokimi nebotičniki ali področja z neugodnim reliefom. Druga pomanjkljivost zaradi nizkega SNR pri lokalizaciji GPS je velika poraba energije [1], ki lahko znatno vpliva na življenjsko dobo baterije mobilnih naprav.

S tehnološkim napredkom so sodobne pametne naprave postale opremljene s številnimi senzorji z majhno porabo energije, ki nam omogočajo raziskovanje v alternativne možnosti lokalizaciji GPS, ki so bolj energijsko učinkovite, zlasti v primer lokalizacije znotraj prostorov. Takšne alternative večinoma temeljijo na uporabi radijskih tehnologij, kot so celična [2], WiFi/Bluetooth [3, 4], ultraširokopasovna (UWB) [5] ali radiofrekvenčna identifikacija (RFID) [6]. Vendar lokalizacija z uporabo radij-

skih frekvenc ni omejena le na lokalizacijo znotraj prostorov, ampak lahko predstavlja tudi alternativo sistemu GPS v mestnih okoljih.

Lokalizacija z radiofrekvenčnim prstnim odtisom (ang. fingerprint) na podlagi prejete moči signala (ang. received signal strength, RSS) postaja priljubljena alternativa sistemu GPS [7, 8]. Pri tej metodi se sprva se opravi niz meritev RSS na vnaprej določenih točkah v mreži, ki se jih shrani v podatkovno zbirko prstnih odtisov. Meritve RSS v realnem času se nato primerjajo z meritvami v zbirki, s čimer se oceni najbolj podobne lokacije. V zadnjem času so metode strojnega učenja postale priljubljeno orodje za razvoj lokalizacijskih storitev, zlasti za določanje lokacij s prstnimi odtisi. Model strojnega učenja je naučen s pomočjo zajete podatkovne zbirke prstnih odtisov in se pozneje uporabi za ocenjevanje lokacije na podlagi meritev v realnem času. Študije kažejo, da pristopi z uporabo strojnega učenja bistveno izboljšajo delovanje lokalizacijskih storitev, ki temelji na prstnih odtisih, v primerjavi s preprosto primerjavo vrednosti RSS [9].

Nizko energijski Bluetooth (ang. Bluetooth Low Energy, BLE) se je zaradi svoje energetske učinkovitosti pogosto uporabljal za lokalizacijo na podlagi prstnih odtisov v zaprtih prostorih, vendar se je izkazalo, da ga je mogoče uporabiti tudi za zunanjo lokalizacijo, ki lahko vključuje sledenje zasedenosti parkirišča, navigacijo na prostem in drugo. Zaradi pomanjkanje prosto dostopnih zbirk podatkov iz realnih okolij je razvoj modelov strojnega učenja za zaznavanje BLE radiofrekvenčnih prstnih odtisov na prostem oteženo, zlasti za primere mest in zelenih površin.

V tem članku predstavljamo nov javno dostopen nabor podatkov o meritvah BLE RSS za lokalizacijo na podlagi radiofrekvenčnih odtisov, pridobljenih s testnim omrežjem LOG-a-TEC [10]. Pridobljeni nabor podatkov je uporabljen za razvoj modela strojnega učenja za lokalizacijo s prstnimi odtisi BLE. Naši glavni prispevki so naslednji:

- Nov nabor podatkov BLE brezžičnih prstnih odtisov, ki so bili zajeti s testno napravo LOG-a-TEC v urbanem in vegetacijskem območju. Podatki so prosto dostopni na: <http://log-a-tec.eu/datasets-ble.html>.
- Model strojnega učenja za lokalizacijo na podlagi BLE radiofrekvenčnih odtisov naučen na zajeti zbirki podatkov.

Članek je organiziran na naslednji način. V razdelku 2 obravnavamo sorodno delo. V razdelku 3 predstavimo metodologijo pridobivanja podatkovnih nizov, v razdelku 4 korake predhodne obdelave podatkovnih nizov, razpravo o izbiri modela ML in metodološki vidiki razvoja lokalizacijskih modelov, v razdelku 5 pa podamo splošno oceno. Z razdelkom 6 zaključimo članek.

2 Pregled področja

Lokalizacija z uporabo tehnologije Bluetooth je postala popularna za razvoj LSB, predvsem za lokalizacijo v notranjih prostorih. Faragher *et al.* [11] je med prvimi predlagal uporabo BLE za lokalizacijo v notranjih prostorih, pri čemer je dosegel natančnost napovedi lokacije pod 2,6 m. Njegov sistem je bil sestavljen iz prenosnih nizkocenovnih BLE naprav z baterijskim napajanjem, ki jih je bilo mogoče enostavno namestiti na poljubno območje.

Podoben pristop kot mi, uporabo meritev BLE RSS za učenje modela strojnega učenja, so za lokalizacijo v zaprtih prostorih uporabili v [12], kjer so ustvarili mrežo velikosti $2,5\text{ m} \times 4,5\text{ m}$, pri čemer so bili izmerjeni položaji med seboj oddaljeni 0,5 m. Na vsaki točki so zbrali 5000 vzorcev oddanih referenčnih signalov, ki so jih uporabili za učenje algoritmov SVM (ang. support vector machine) in logistične regresije (ang. logistic regression) s povprečno napako lokalizacije 50 cm. Drug podoben pristop uporabe BLE za lokalizacijo znotraj prostorov so predlagali v [13], kjer so uporabili mrežo $3\text{ m} \times 3\text{ m}$, pri čemer so bile lokacijske točke oddaljene 0,3 m (1 čevlj). Uporabili so algoritem strojnega učenja imenovan naključna drevesa (ang. random forest) in dosegli natančnost lokalizacije do 96%. V obeh predstavljenih delih je bila lokalizacija opredeljena kot klasifikacijski problem, ki določa, kje v mreži se nahaja naprava. Nekateri raziskovalci so lokalizacijo BLE v notranjih prostorih opredelili kot regresijski problem, pri čemer so za določitev natančne lokacije naprave uporabili različne algoritme strojnega učenja. Eno takih del so predstavili v [14], kjer so uporabili regresijo s pomočjo Gaussovega procesa, s katero so poskušali določiti natančno trajektorijo aktivne naprave v prostoru velikosti $10\text{ m} \times 25\text{ m}$.

V zadnjih letih je BLE pritegnil pozornost raziskav za lokalizacijo na prostem. V [15] so predlagali uporabo radiofrekvenčnih odtisov BLE za zaznavanje zasedenosti parkirišč, kjer so v mreži 150 parkirnih mest lahko zaznali zasedenost z 90,7-odstotno natančnostjo. V [16] so predstavili rešitev za lokalizacijo ljudi na prehodih za pešce s pomočjo tehnologije BLE, ter dosegli 99 oziroma 97,7-odstotno natančnost lokalizacije.

3 Meritve

V tem poglavju najprej predstavimo testno omrežje LOG-a-TEC, s katerim smo izvedli meritve in nato podrobneje opišemo metodo in nastavitve eksperimenta za zajem podatkov.

3.1 Testno omrežje LOG-a-TEC 3.0

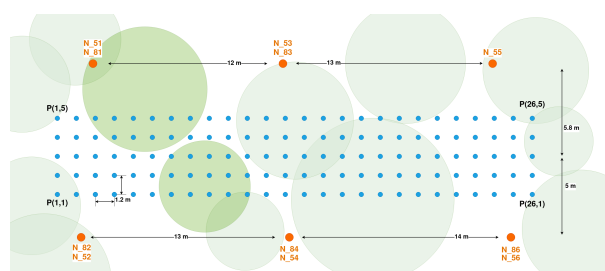
Testno omrežje LOG-a-TEC 3.0, ki se nahaja v parku Inštituta "Jožef Stefan" v Ljubljani, je namenjeno preučevanju

različnih brezžičnih tehnologij z realnimi napravami v realnem okolju, ter tako omogoča zanesljivo testiranje in vrednotenje aplikacij interneta stvari, komunikacije med napravami in mobilnih omrežij naslednjih generacij. Po meri zasnovana vozlišča lahko gostijo več različnih radijskih vmesnikov, ki omogočajo eksperimentiranje s tehnologijami kot so nizko-energijski Bluetooth, UWB, LoRaWAN, WiFi, Zigbee in 6LoWPAN. 77 vozlišč, nameščenih v notranjem in zunanjem testnem okolju, tako tvori heterogeno brezžično testno omrežje, katerega je mogoče upravljati in krmiliti preko centralnega strežnika za upravljanje. Dostop do vozlišč je urejen prek spletnega vmesnika, ki omogoča hkraten dostop večim uporabnikom in ponuja sistem za rezervacijo naprav ter popolnoma avtomatiziran mehanizem izvajanja eksperimentov in zbiranja rezultatov [10]. Za dodatno podporo in nadzor pri izvajanju eksperimentov testno omrežje ponuja orodje za spremljanje poskusov v realnem času [17], ki nadalje omogoča zbiranje meritev iz vseh vozlišč in vizualizacijo podatkov.

3.2 Zajem podatkov



Slika 1: a) Postavitev uporabljenih vozlišč v testnem omrežju, b) Zelenje v parku in drog ulične svetilke s pritrjenim vozliščem 55.



Slika 2: Podrobna mapa področja lokalizacije.

Podatke smo pridobili z uporabo zunanjega dela testnega omrežja LOG-a-TEC, kot prikazuje slika 1. Pri izvajanju meritev smo uporabili 25 vozlišč, opremljenih z Bluetooth modulom WL1837MOD, ki so pritrjena na droge uličnih svetilk in okoliške stavbe na višinah med 2 in 3,9 m nad tlemi. Testno omrežje je umeščeno v park inštituta, ki je gosto posejan z drevesi in visokim grmičevjem ter obdan s stavbami pisarn, kar je razvidno na sliki 1. Slika 2 prikazuje podrobnejši načrt testnega okolja, na kateri so visoka drevesa označena s svetlo zelenimi krogi, gostejše in visoko grmovje pa s temno zelenimi krogi. Ista slika prikazuje tudi področje lokalizacije,

ki je sestavljeno iz 5×26 točk (prikazanih z modrimi pikami), ki so med seboj oddaljene 1,2 m in pokrivajo 150 kvadratnih metrov. Na vsaki točki je bil na 1,6 m visokem stojalu fiksno nameščen namenski oddajnik BLE, ki je v 100 ms intervalih pošiljal oglaševalne pakete BLE z močjo -2 dBm. Okoliška vozlišča so shranjevala zaznane pakete in njihove karakteristike za vsako točko približno eno minuto, nato pa je operater prestavil oddajnik na naslednjo točko. Meritev je bila izvedena pozno popoldne, ko v okolici parka ni bilo ljudi. Ker je testno omrežje postavljeno v realno okolje, so vozliščne naprave zaznale pakete BLE tudi od drugih naprav (npr. brezžične tipkovnice in slušalke v okoliških pisarnah), katere smo pa po koncu meritve izločili iz podatkov.

4 Metodologija

V tem poglavju opišemo postopke predobdelave podatkov, predlagamo model strojnega učenja in predstavimo metodologijo eksperimenta lokalizacije, kjer podrobneje opišemo učni model in ovrednotimo rezultate.

4.1 Predpriprava podatkov

Neobdelane meritve, katerih zajem smo opisali v razdelku 3, vsebujejo časovne vrste različnih dolžin. Vsaka časovna vrsta vsebuje zaporedne zapise vrednosti sprejete moči signala (RSS), katerih statična porazdelitev je bila pravilna za vsa vozlišča. Zaradi tehničnih omejitev, časovni žig prejetih paketov med vozlišči ni usklajen (sinhroniziran), sami paketi pa tudi ne vsebujejo informacij o sosledju. Hkrati je zaradi izgubne narave brezžične komunikacije malo verjetno, da vsa vozlišča uspešno sprejmejo vsak poslan oglaševalni paket. Vozlišča #63, #64 in #65 na primer niso vedno v vidnem polju z oddajnikom in so začela prejemati pakete šele, ko je bil oddajnik postavljen blizu prehoda. To je povzročilo neenako število vzorcev na položaj (med 910 in 12 317) oz. neenako število vzorcev na vozlišče (med 0 in 889). Neobdelane meritve skupno vsebujejo 723 147 vzorcev na napravo.

Podatke iz več časovnih vrst različnih dolžin smo morali preoblikovati v vektorje značilk R dolžine N z ustrezno oznako položaja P , kjer se je nahajal oddajnik D . Da bi odpravili omejitev nabora meritev zaradi neusklajenih podatkov, smo uporabili posebno metodo povečanja, s katero smo pridobili več vzorcev. Metoda upošteva porazdelitev vzorcev RSS, namesto da obravnava meritve RSS v vsakem vozlišču kot časovno vrsto. Z metodo smo tako naključno vzorčili zbrane podatkovne točke iz vsakega vozlišča, z enako verjetnostjo za vsak merilni položaj. To je učinkovito povečalo število vzorcev, zaradi nepristranskega vzorčenja pa se je ohranila prvotna statična porazdelitev vzorcev RSS v vsakem vozlišču. Če vzorcev ni bilo na voljo, smo predpostavili, da je bil sprejem -140 dBm, ki je vrednost nekoliko pod območjem občutljivosti sprejemnika. Po razširitvi je končni nabor podatkov sestavljen iz 500 000 vzorcev vektorjev značilk R dolžine N , kjer N ustreza številu uporabljenih vozlišč, to je 25.

4.2 Predlagan model

Cilj našega pristopa je natančna ocena položaja brezžičnega oddajnika v parku s pomočjo radiofrekvenčnega odtisa. Nalogo smo opredelili kot klasifikacijski problem, pri katerem smo poskušali oceniti položaj oddajnika do njegove najbližje točke v mreži.

Za rešitev navedenega problema je smo izbrali klasifikacijski algoritem naključni gozd (ang. Random Forest), saj se je izkazal za učinkovitega pri reševanju nalog lokalizacije [18, 19]. Uporabili smo knjižnico XGBoost [20] (implementacijo naključnega gozda) zaradi njegove učinkovitosti na velikih naborih podatkov, pri čemer smo uporabili že privzete parametre določene v implementaciji knjižnice.

4.3 Učenje modela in vrednotenje

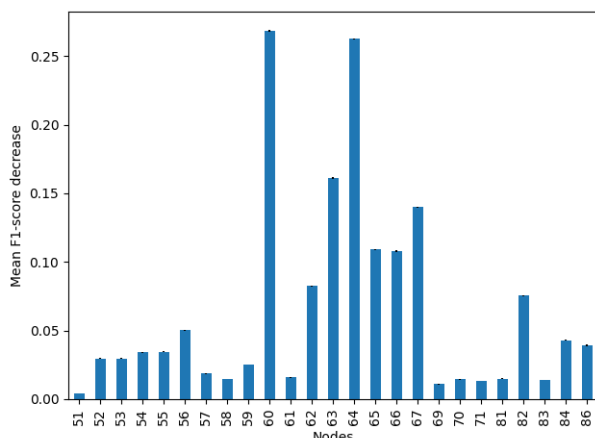
Za učenje modela smo uporabili 10-kratno tehniko stratificiranega naključnega deljenja. Pri tem smo podatke premešali in jih razdelili na učno in testno množico v razmerju 9 : 1. Postopek smo ponovili desetkrat, da smo zagotovili verodostojnost rezultatov.

Uspešnost nastalih modelov smo ocenili s standardno metriko F1. Rezultat F1 je izražen kot harmonično povprečje med Natančnost $= \frac{TP}{TP+FP}$ in Priklic $= \frac{TP}{TP+FN}$, kjer TP, FP in FN pomenijo resnično pozitivne, lažno pozitivne in lažno negativne rezultate. Ker je bila za vrednotenje modela uporabljena 10-kratna stratificirana porazdelitev, je končni rezultat povprečna ocena F1 za vseh 10 scenarijev učenja in testiranja.

Za bolj poglobljeno analizo izdelanega modela smo uporabili postopek permutacijske pomembnosti za vrednotenje značilk [21]. V tem postopku se najprej eno od značilk izključi (v našem primeru eno vozlišče), nato se model ponovno nauči in opazuje razliko v uspešnosti. Proces se ponovi za vsako izmed značilk. Celoten postopek tako pokaže, prispevek posameznih značilk (t.j., vozlišč) k uspešnosti modela.

5 Rezultati

V prejšnjih poglavjih opisan učni model v povprečju doseže F1 rezultat 96,1 %. Najnižji F1 rezultat je bil 80,9 % pri referenčnih točkah $P(8, 2)$ in $P(10, 4)$. Preverili smo tudi, koliko je vsako vozlišče v testnem okolju LOG-ATEC prispevalo h končnemu rezultatu F1. Kot prikazuje slika 3, analiza pomembnosti značilk pove, da je šest vozlišč, ki so največ prispevala k uspešnosti učnega modela: #60, #64, #63, #67, #65 in #66 v padajočem vrstnem redu. To nakazuje, da so pri oceni pozicije najbolj prispevala vozlišča, ki niso vedno v vidnem polju oddajnika, kot so vozlišča za stavbo oz. nad prehodom. Sklep, ki izhaja iz rezultatov je skladen z našo predpostavko, da model lahko predvidi, da je oddajnik na določenem delu območja, ko zgoraj omenjena vozlišča uspešno prejemajo pakete. Na primer, ko vozlišče #64 začne prejemati oglaševalne pakete, se oddajnik zelo verjetno nahaja v desnem delu parka, saj je takrat v vidnem polju z vozliščem.



Slika 3: Pomembnost brezžičnih vozlišč pri oceni pozicije, dobljena s permutacijo na celotnem učnem modelu.

6 Zaključek

V tem članku predstavimo nov, javno dostopen set podatkov z meritvami sprejete moči signalov BLE paketov, namenjen lokalizaciji na podlagi radiofrekvenčnega odtisa, pridobljen s testnim omrežjem LOG-a-TEC. Na podlagi zajetih podatkov smo razvili učni model, ki temelji na algoritmu naključnega gozda in lahko oceni lokacijo naprave BLE s povprečnim rezultatom F1 96,1 %. Poleg tega smo opisali tudi delovanje modela in njegovih napovedi z izračunom pomembnosti, ki ga ima posamezno vozlišče na končno oceno lokacije. Kot je razvidno iz rezultatov, k oceni položaja najbolj prispevajo vozlišča, ki niso vedno v vidnem polju oddajnika.

Pri prihodnjem delu nameravamo izboljšati metodo pridobivanja paketov z dodajanjem časovnih žigov. Ta nadgradnja bi pomagala izboljšati modeliranje povezav z odkrivanjem manjkajočih in ne-zaznanih paketov. Načrtujemo tudi razširitev našega nabora zbirk podatkov z bolj raznolikimi scenariji. Na primer scenariji s posebnimi človeškimi dejavnostmi na območju in njihovim ponavljanjem v različnih delih leta (npr. poletje, zima) in vremenskih razmerah (npr. nevihta, vroč dan).

Zahvala

Delo je bilo podprto s strani Slovenske raziskovalne agencije (P2-0016 in J2-2507), in evropskega projekta Fed4FIRE+ (Grant 732 638).

Literatura

- [1] L. A. Tawalbeh, A. Basalamah, R. Mehmood, and H. Tawalbeh, "Greener and smarter phones for future cities: Characterizing the impact of gps signal strength on power consumption," *IEEE Access*, vol. 4, pp. 858–868, 2016.
- [2] M. Ibrahim and M. Youssef, "Cellsense: An accurate energy-efficient gsm positioning system," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 1, pp. 286–296, 2011.
- [3] Y. Wang, Q. Ye, J. Cheng, and L. Wang, "Rssi-based bluetooth indoor localization," in *2015 11th international conference on mobile ad-hoc and sensor networks (MSN)*. IEEE, 2015, pp. 165–171.
- [4] M. A. Youssef, A. Agrawala, A. U. Shankar, and S. H. Noh, "A probabilistic clustering-based indoor location determination system," Tech. Rep., 2002.

- [5] L. Zwirrello, T. Schipper, M. Harter, and T. Zwick, "Uwb localization system for indoor applications: Concept, realization and analysis," *Journal of Electrical and Computer Engineering*, vol. 2012, 2012.
- [6] C. Li, L. Mo, and D. Zhang, "Review on uhf rfid localization methods," *IEEE Journal of Radio Frequency Identification*, vol. 3, no. 4, pp. 205–215, 2019.
- [7] K. Whitehouse, C. Karlof, and D. Culler, "A practical evaluation of radio signal strength for ranging-based localization," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, no. 1, p. 41–52, jan 2007. [Online]. Available: <https://doi.org/10.1145/1234822.1234829>
- [8] H. Zou, Z. Chen, H. Jiang, L. Xie, and C. Spanos, "Accurate indoor localization and tracking using mobile phone inertial sensors, wifi and ibeacon," in *2017 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL)*, 2017, pp. 1–4.
- [9] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern wlan fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1974–2002, 2017.
- [10] M. Vucnik, T. Solc, U. Gregorc, A. Hrovat, K. Bregar, M. Smolnikar, M. Mohorcic, and C. Fortuna, "Continuous Integration in Wireless Technology Development," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 74–81, 2018.
- [11] R. Faragher and R. Harle, "Location fingerprinting with bluetooth low energy beacons," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 11, pp. 2418–2428, 2015.
- [12] P. Sthapit, H.-S. Gang, and J.-Y. Pyun, "Bluetooth based indoor positioning using machine learning algorithms," in *2018 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, 2018, pp. 206–212.
- [13] C. Jain, G. V. S. Sashank, V. N, and S. Markkandan, "Low-cost ble based indoor localization using rssi fingerprinting and machine learning," in *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2021, pp. 363–367.
- [14] D. D. Nguyen and M. Thuy Le, "Enhanced indoor localization based ble using gaussian process regression and improved weighted knn," *IEEE Access*, vol. 9, pp. 143 795–143 806, 2021.
- [15] P. Seymer, D. Wijesekera, and C.-D. Kan, "Secure outdoor smart parking using dual mode bluetooth mesh networks," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1–7.
- [16] K. Shin, R. McConville, O. Metatla, M. Chang, C. Han, J. Lee, and A. Roudaut, "Outdoor localization using ble rssi and accessible pedestrian signals for the visually impaired at intersections," *Sensors*, vol. 22, no. 1, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/1/371>
- [17] G. Morano, A. Hrovat, M. Vučnik, J. Puhon, G. Gardašević, D. Vasiljević, and T. Javornik, "Experiment Control and Monitoring System for LOG-a-TEC Testbed," *Sensors*, vol. 21, no. 19, p. 6422, 2021. [Online]. Available: <https://www.mdpi.com/1424-8220/21/19/6422>
- [18] L. Calderoni, M. Ferrara, A. Franco, and D. Maio, "Indoor localization in a hospital environment using random forest classifiers," *Expert Systems with Applications*, vol. 42, no. 1, pp. 125–134, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417414004497>
- [19] Y. Wang, C. Xiu, X. Zhang, and D. Yang, "Wifi indoor localization with csi fingerprinting-based random forest," *Sensors*, vol. 18, no. 9, p. 2869, 2018.
- [20] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [21] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.