

PRIMENA METODA MAŠINSKOG UČENJA U ANALIZI HODA PACIJENATA OBOLELIH OD NEURODEGENERATIVNIH BOLESTI**APPLICATION OF MACHINE LEARNING METHODS IN GAIT ANALYSIS OF PATIENTS SUFFERING FROM NEURODEGENERATE DISEASES**Svetlana Sekulić, *Fakultet tehničkih nauka, Novi Sad***Oblast – BIOMEDICINSKO INŽINJERSTVO**

Kratak sadržaj – U okviru rada opisana je primena mašinskog učenja u analizi hoda kod neurodegenerativnih oboljenja. Cilj je sagledavanje mogućnosti da se pomoću metoda mašinskog učenja formira klasifikator koji može da prepozna neurodegenerativna oboljenja kod pacijenata na osnovu snimanja karakteristike hoda.

Ključne reči: Faze hoda, odabir obeležja, klasifikacija.

Abstract – The paper describes the application of machine learning in the analysis of gait in neurodegenerative diseases. The goal is to consider the possibility of using machine learning methods to create a classifier that can recognize neurodegenerative diseases in patient based on the recording of gait characteristics.

Keywords: Phases of gait, feature selection, classification.

1. UVOD

Nervni sistem je najsloženiji sistem organizma koji reguliše ogroman broj složenih funkcija. U osnovi se sastoji od specijalizovanih ćelija čija je uloga da prenese utiske iz spoljašnje i unutrašnje sredine do centralnih struktura koji ih obrađuju i vraćaju do efektivnih organa obezbeđujući njihovo harmonično delovanje.

Osnovna jedinica, građe i funkcija nervnog sistema je nervna ćelija ili neuron. Neuron se sastoji od tela, kratkih nastavaka ili dendrida i dugih nastavaka ili aksona. Nastavci neurona imaju ulogu da preko dendrida prime informaciju, a preko aksona proslede informaciju. Proteini imaju veliku ulogu prilikom deobe nervne ćelije, i oni se smatraju genetskim realizatorima. Kada njihov nastanak nije adekvatan, dolazi do mutacije ćelija. Protein pod imenom „prion“, utiče na deobu nervnih ćelija, pa samim time utiče na nemogućnost njihovog funkcionisanja.

Ovakav protein je jedan od uzroka nastanka Parkinsonovog oboljenja, Hantingtonove bolesti i amiotrofične lateralne skleroze. Ova oboljenja se svrstavaju u neurodegenerativna oboljenja. Glavni simptom ovih oboljenja jeste tremor. Tremor izaziva nevoljno podrhtavanje dela tela ili čak čitavog tela. Prvo se pojavljuju na završecima donjih ekstremiteta i prilikom progresije nastavlja da se penje sve do gornjih ekstremiteta.

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Darko Stanišić, vanred. prof.

2. BAZA PODATAKA

Hod se najjednostavnije može definisati kao niz ravnomernih i naizmeničnih koordinisanih pokreta nogu i trupa čiji je cilj premeštanje tela sa jednog na drugo mesto u prostoru. Hod se deli na dve faze:

1. Faza oslonca: razlikuje se 5 podfaza, udar petom, oslonac na puno stopalo, međuoslonac, odvajanje pete i odvajanje prstiju.
2. Faza njihanja: razlikuje se 3 podfaze, inicijalno klaćenje, međuklaćenje i deceleracija.

Baza podataka sadrži vremenske intervale faza hoda. U eksperimentu je učestvovalo 64 učesnika. Učesnici su podeljeni u četiri klase:

1. Klasa 0: učesnici kod kojih nije detektovano nijedno oboljenje i učestvovalo je 16 osoba.
2. Klasa 1: učesnici kod kojih je detektovana amiotrofična lateralna skleroza i učestvovalo je 13 osoba.
3. Klasa 2: učesnici kod kojih je detektovana Hantingtonova bolest i učestvovalo je 20 učesnika.
4. Klasa 3: učesnici kod kojih je detektovana Parkinsonova bolest i učestvovalo je 15 osoba.

Zastupljenost oba pola u bazi je jednaka, starosna granica je između 20 i 80 godina. Svaki učesnik je morao da prošetna hodnikom dužine 77m i napravi oko 100 koraka. Vremenski intervale svakog učesnika su predstavljani u .ts fajl. Baza sadrži 64 .ts fajlova u kojima su prikazane vrednosti u sekundama i procentima svih faza za više od 100 merenja svake klase. Svaki fajl sadrži trinaest kolona koje su prikazane na slici 1.

KOLONE	OBJAŠNENJE
1. ELAPSED TIME(SEC)	Protoklo vreme, razlika između početka koraka i završetka koraka.
2. LEFT STRIDE INTERVAL(SEC)	Interval koraka izmeren između jednog udarca petom leve noge do sledećeg udarca o pod.
3. RIGHT STRIDE INTERVAL(SEC)	Interval koraka izmeren između jednog udarca petom desne noge do sledećeg udarca o pod.
4. LEFT SWING INTERVAL (SEC)	Vremenski period kada leva noga nije u kontaktu sa podom.
5. RIGHT SWING INTERVAL (SEC)	Vremenski period kada desna noga nije u kontaktu sa podom.
6. LEFT SWING INTERVAL (%)	Procentat vremenskog perioda kada leva noga nije u kontaktu sa podom.
7. RIGHT SWING INTERVAL (%)	Procentat vremenskog perioda kada desna noga nije u kontaktu sa podom.
8. LEFT STANCE INTERVAL (SEC)	Vremenski interval leve noge kada je bila na th.
9. RIGHT STANCE INTERVAL (SEC)	Vremenski interval desnog stopala na th.
10. LEFT STANCE INTERVAL (%)	Vremenski interval levog stopala na th izražen u procentima.
11. RIGHT STANCE INTERVAL (%)	Vremenski interval desnog stopala na th izražen u procentima.
12. DOUBLE SUPPORT INTERVAL (SEC)	Vremenski interval kada su oba stopala na podu.
13. DOUBLE SUPPORT INTERVAL (%)	Vremenski interval kada su oba stopala na podu izražen u procentima.

Slika 1. Prikaz kolona

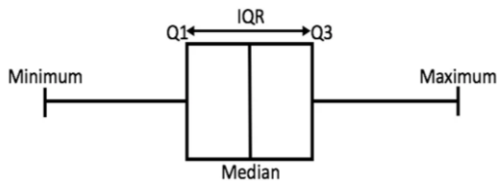
Imena kolona predstavljaju „obeležja“ ili „atribute“. Obeležje je bilo koji merljivi pokazatelj.

2.1. ODABIR OBELEŽJA

Odabir obeležja zahteva analizu podataka. Analiza podataka je izvršena u dve etape:

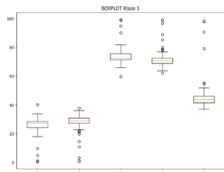
1. Iscrtavanje boxplot-ova
2. Računanje srednje vrednosti, medijana, interkvartalnog i dinamičkog opsega i standardne devijacije.

Boxplot ili „box and whisker plot“ je način da se prikaže raspodela vrednosti na osnovu 5 ključnih tačaka, minimum, prvi kvartal, medijan, treći kvartal i maksimum. Primer boxplot-a prikazan je na slici 2.

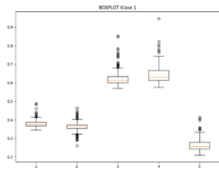


Slika 2. Primer boxplot-a

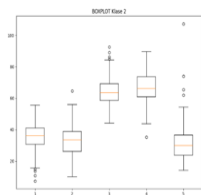
Izabrano je 4 učesnika iz svake klase. Iz klase 1 odabran je učesnik pod rednim brojem 4, klase 2 pod rednim brojem 13, iz klase 3 pod redim brojem 2 i iz klase 0 pod rednim broj 15. Slike boxplot-ova pomenutih učesnika su prikazane na slikama od 2 do 9.



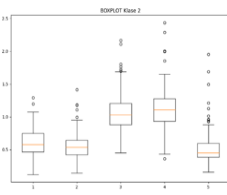
Slika 2. Boxplot klase 1, procenat faze hoda



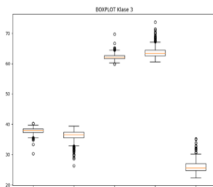
Slika 3. Boxplot klase 1, vremenski period



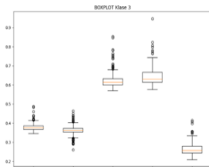
Slika 4. Boxplot klase 2, procenat faze hoda



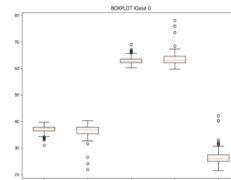
Slika 5. Boxplot klase 2, vremenski period faza hoda



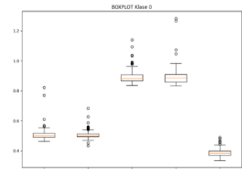
Slika 6. Boxplot klase 3, procenat faza hoda



Slika 7. Boxplot klase 3, vremenski period



Slika 8. Boxplot klase 0, procenat faza hoda



Slika 9. Boxplot klase 0, vremenski period faza hoda.

Kružići prikazani na slikama se nazivaju „outlier“. „Outlier“ je vrednost koja se nalazi izvan opsega ukupnog uzorka tj. vrednost koje se ne nalaze u opsegu dinamičkog intervala. Slike prikazuju da dosta uzorka ispada iz opsega i da postoje odstupanja.

Boxplot-ovi kod klase 1,3 i 0 su dosta uži, dok kod klase 2 raspon boxplot-a je širi. Kod svih obeležja 3 i 4, koja predstavljaju faze oslonca u toku levog i desnog koraka, vrednosti između 60% i 80%, što je u skladu sa očekivanih oko 60%, kod obeležja vezanih za fazu njihanja imamo takođe očekivane vrednosti 20% i 40%, dok trajanje faze duplog oslonca, obeležje 5, ima još manju vrednost.

Poređenje obeležja koja čine procenat faza hoda i obeležja koja predstavljaju vremenski period, može se primetiti da više vrednosti odstupaju kod vremenskog perioda. Zbog toga za dalju obradu koriste se obeležja:

1. Left swing interval % (LSI)
2. Right swing interval % (RSI)
3. Left stance interval % (LSI)
4. Right stance interval % (RSI)
5. Double Support interval % (DSI)

2.2. ANALIZA PODATAKA

Svaki učesnik sadrži više od 100 merenja svake faze hoda, tako da su izračunati sledeći parametri:

1. Srednja vrednost :

$$srv = (x_1 + x_2 + \dots + x_n)/n \quad (1)$$

2. Medijan:

$$med = (n + 1)/2 \quad (2)$$

3. Dinamički opseg:

$$dim = \max(x) - \min(x) \quad (3)$$

4. Interkvartalni opseg:

$$IQR = Q_3 - Q_1 \quad (4)$$

5. Standardna devijacija:

$$std = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (5)$$

Obeležje	Klasa	SR	Medijan	DIM	IQR	STD
LSI	1	25,98	27,22	39,51	4,48	5,3
	2	35,686	36,4	48,3	10,38	8,369
	3	37,812	37,95	9,95	1,15	1,069
	0	36,169	36,15	11,55	1,33	1,24
RSI	1	28,66	29,08	52,56	3,67	5,916
	2	33,298	33,85	54,49	12,7	8,831
	3	36,069	36,55	13,13	1,84	2,037
	0	36,143	36,17	10	1	1,06
LSI	1	74,016	72,78	39,51	4,48	5,39
	2	63,314	63,6	48,3	10,36	8,369
	3	62,188	62,05	9,95	1,15	1,069
	0	63,831	63,85	11,54	1,33	1,24
RSI	1	71,34	70,92	52,55	3,67	5,916
	2	66,702	66,15	54,5	12,72	8,831
	3	63,931	63,45	13,13	1,84	2,073
	0	63,857	63,83	10	1	1,06
DSI	1	45,49	43,92	61,08	4,71	8,976
	2	31,44	29,88	92,99	12,85	11,22
	3	26,177	25,58	12,87	2,35	2,27
	0	27,676	27,58	7,8	1,612	1,198

Tabela 1. Analiza podataka procenta faza hoda

Tabela 1 pokazuje dobru raspodelu podataka, faza hoda čije vrednosti predstavljaju procenat faza hoda. Boxplot-ovi su pokazali prisustvo outlier-a. Odabir obeležja, sa manje outlier-a je smanjio njihov uticaj, ali ne i uklonio. Uklanjanje outlier-a je izvršeno računanjem srednje vrednosti, standardne devijacije, maksimalne vrednosti i minimalne vrednosti za svakog učesnika.

3. KLASIFIKACIJA FAZA HODA

Klasifikacioni problemi mašinskog učenja spadaju u nadgledano učenje. Nadgledano učenje za cilj ima da nauči mašinu samostalno da donosi odluke već poznatih podataka tj. potrebno je naučiti mašinu određenom preslikavanju. Za klasifikaciju izabrana su tri modela:

1. Metod k najbližeg suseda (KNN):

Metod K najbližeg suseda predstavlja veoma intuitivan metod klasifikacije novih uzoraka na osnovu sličnosti sa postojećim uzorcima iz skupa za obuku. Potrebno je izabrati vrednost parametra K, pa za nepoznati uzorak pronaći k najbližih trening – uzoraka, i na kraju svrstati nepoznati uzorak u klasu koja se u najvećoj meri pojavljuje u tom podskupu od k uzoraka

2. Metod vektora nosača (SVM):

Vektori nosači predstavljaju jedan od modela mašinskog učenja sa supervizijom, koji se koristi za klasifikacionu i regresionu analizu. Uloga SVM-a je da na osnovu ulaznog skupa uzoraka odredi kojoj klasi nepoznati uzorak pripada.

3. Neuralne mreže (MLP):

Neuralne mreže predstavljaju novu generaciju informacijske analize, a karakteriše ih mogućnost učenja, pamćenja i generalizacije na osnovu prethodnih informacija. One se opisuju kao matematički model kojima se oponaša rad neurona ljudskog mozga.

3.1. EVALUACIJA

Matrica konfuzije je tabelarni prikaz brojeva ispravno i pogrešno klasifikovanih objekata na osnovu kojih se mog vršiti ocene klasifikacije. Kreirani su parovi: „srednja vrednost i standardna devijacija“, „srednja i minimalna vrednost“, „srednja i maksimalna vrednost“ i „srednja vrednost i medijan“. Svaki par je obučen korišćenjem

pomenutih modela. Rezultati matrice konfuzije svih parova i modela su prikazani u tabeli 2.

	SREDNJA VREDNOST											
	KNN			SVM			MLP					
STD	14	0	1	1	13	1	1	1	12	1	2	1
	2	3	2	6	4	1	2	6	3	5	3	2
	4	3	12	1	2	2	11	5	1	2	14	3
	2	1	5	7	2	4	2	7	2	7	1	5
MIN	12	0	3	1	11	0	4	1	11	0	4	1
	0	1	4	8	1	3	5	4	2	3	4	4
	8	4	6	2	8	0	11	1	9	2	6	3
	6	4	1	4	5	2	6	2	5	4	4	2
MAX	13	0	2	1	11	2	2	1	11	2	2	1
	3	1	1	8	5	2	1	5	1	5	2	5
	4	2	12	2	4	3	10	3	4	3	13	0
	2	4	6	3	3	4	5	3	4	4	3	4
MED	12	0	2	2	14	0	3	2	10	0	4	2
	1	2	4	6	3	3	1	6	2	2	2	7
	8	4	6	2	7	0	9	4	8	1	7	4
	5	3	1	6	5	4	4	2	3	5	3	4

Tabela 2. Matrice konfuzije parova

Tabela 2 pokazuje da je KNN kod prvog para tačno klasifikovao 36 uzoraka, SVM 32 uzorka i MLP 36 uzoraka. Kod trećeg para KNN je klasifikovao 23 uzorka, SVM 27 uzoraka i MLP 22 uzorka. KNN kod trećeg para je klasifikovao 29 uzorka, SVM 26 uzorka i MLP 33 uzorka. Kod četvrtog para KNN je tačno klasifikovao 26 uzorka, SVM 25 uzorka i MLP 23 uzorka. Klasifikatori su se najbolje pokazali kod prvog para. U svrhu poboljšavanja klasifikacije kreirani su tripleti:

1. Srednja vrednost, standardna devijacija i minimalna vrednost
2. Srednja vrednost, standardna devijacija i maksimalna vrednost
3. Srednja vrednost, standardna devijacija i medijan

SREDNJA VREDNOST I STANDARDNA DEVIJACIJA

	SREDNJA VREDNOST I STANDARDNA DEVIJACIJA											
	KNN			SVM			MLP					
MIN	13	0	2	1	13	1	1	1	13	0	2	1
	2	2	3	6	4	2	2	5	2	3	3	5
	6	4	9	1	2	2	12	4	2	4	10	4
	3	1	4	7	2	4	2	7	2	7	1	5
MAX	14	0	0	1	10	1	1	4	12	2	2	0
	4	2	1	6	4	2	1	6	4	5	2	2
	5	2	11	2	3	1	10	6	1	1	13	5
	1	4	7	3	3	6	1	5	2	5	4	4
MED	14	0	1	1	14	0	1	1	12	1	1	2
	2	1	3	7	4	2	2	5	2	4	2	5
	4	4	10	2	2	2	11	5	1	2	13	4
	3	1	4	7	2	5	2	6	2	6	2	5

Tabela 3. Matrice konfuzije tripleta

Tabela 3 prikazuje matrice konfuzije tripleta. Kod prvog tripleta KNN je tačno klasifikovao 31 uzorak, SVM 33 uzorka i MLP 31 uzorak. Kod drugog tripleta KNN je tačno klasifikovao 30 uzorka, SVM 27 uzorka i MLP 34 uzorka. Kod trećeg tripleta KNN je tačno klasifikovao 32 uzorka, SVM 33 uzorka i MLP 34 uzorka.

3.2. TAČNOST KLASIFIKACIJE

Tačnost klasifikacije predstavlja udeo tačno klasifikovanih uzoraka u ukupnom broju uzoraka tj. zbir dijagonale matrice konfuzije kroz ukupan broj uzoraka. Računanje tačnosti se vrši preko formule:

$$\text{Tačnost} = \frac{\sum_{i=0}^n TP}{\text{Ukupan broj uzoraka}} \quad (6)$$

U tabelama 4 i 5 prikazane su vrednosti tačnosti klasifikatora parova i tripleta.

Parovi/Model	TAČNOST MODELA		
	KNN	SVM	MLP
„ <i>srv</i> i <i>std</i> “	0,56	0,5	0,56
„ <i>srv</i> i <i>min</i> “	0,35	0,42	0,34
„ <i>srv</i> i <i>max</i> “	0,45	0,40	0,51
„ <i>srvt</i> i <i>med</i> “	0,40	0,39	0,36

Tabela 4. Tačnost modela korišćenjem parova

U tabeli 4 klasifikatori su tačnije klasifikovali uzorke kod prvog para, gde su KNN i MLP klasifikovali sa tačnošću od 56 % i SVM sa tačnošću od 50%.

Tripleti/Model	TAČNOST MODELA		
	KNN	SVM	MLP
„ <i>srv</i> , <i>std</i> i <i>min</i> “	0,48	0,53	0,48
„ <i>srv</i> , <i>std</i> i <i>max</i> “	0,46	0,42	0,55
„ <i>srv</i> , <i>std</i> i <i>med</i> “	0,5	0,52	0,53

Tabela 5. Tačnost modela korišćenjem stripleta

Kreiranje tripleta je nešto malo poboljšalo klasifikaciju kod SVM klasifikatora, gde je kod KNN-a i MLP-a, nešto malo smanjena tačnost klasifikacije.

4. ZAKLJUČAK

U ovom radu predstavljena je mogućnost dijagnostikovanja neurodegenerativnih oboljenja primenom mašinskog učenja. Metode su testirane na bazi podataka koja sadrže rezultate dobijene snimanjem karakteristika hoda grupe subjekata. U eksperimentu je učestvovalo 64 učesnika, među kojima su bili pacijenti oboleli od amiotrofične lateralne skleroze, Hantingtonovog oboljenja, Parkinsonove bolesti, kao i učesnici kod kojih nije detektovano nijedno od ovih oboljenja. Detaljno su analizirani svi podaci sadržani u bazi i na osnovu njih su formirana karakteristična obeležja. Za dalji rad su izabrana obeležja koja predstavljaju procenat trajanja faza hoda. Prvo su formirani modeli koji su kao ulaze imali grupu obeležja dobijenih kreiranjem različitih parova. Parovi su „srednja vrednost i standardna devijacija“, „srednja i minimalna vrednost“, „srednja i maksimalna vrednost“ i „srednja vrednost i medijan“. Za klasifikaciju korišćena su tri modela KNN, SVM i MLP. Za svaki model prikazana je matrica konfuzije i izračunata tačnost klasifikacije. Matrice konfuzije su pokazale da najviše pogođenih uzoraka kod svih klasifikatora ima kod para „srednja vrednost i standardna devijacija“.

Za svrhu poboljšanje klasifikacije kreirani su tripleti: „srednja vrednost, standardna devijacija i minimalna vrednost“, „srednja vrednost, standardna devijacija i maksimalna vrednost“ i „srednja vrednost, standardna devijacija i medijan“.

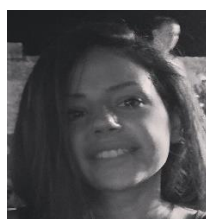
Kreiranje matrice konfuzije tripleta je pokazala da se broj tačno pogođenih uzoraka nije poboljšao. Računanje tačnosti klasifikacije je pokazalo da je tačnost kod KNN i MLP 56% kod prvog para, ali da je uvođenjem tripleta klasifikacija opala za oko 8%. Matrice su pokazale da najviše klasifikovanih uzoraka pripada učesnicima koji služe kao kontrola, a klasifikatori su najbolje prepoznali Hantingtonovo oboljenje sa oko 60% uspešnosti.

Parkinsonova bolest je uspešno prepoznata sa oko 50% slučajeva, dok je kod svih klasifikatora amiotrofičnolateralne skleroza prepoznavana u veoma niskom procentu. Ovaj rad pokazuje da obeležja dobijena iz trajanja faza hoda dobra osnova za formiranje klasifikatora za prepoznavanje neurodegenerativnog oboljenja i da se najbolji rezultati dobijaju kada se kao obeležja koriste srednja vrednost pojedinih faza i njihova standardna devijacija. Dalje istraživanje bazirano na rezultatima ovog rada bi moglo da dovede do poboljšanja performansi klasifikatora dodavanjem dodatnih obeležja koja nisu vezana za trajanje faza hoda.

5. LITERATURA

- [1] Vladimir Crnojević, “Prepoznavanje oblika”, Fakultet tehničkih nauka, Novi Sad, 2014.
- [2] Grazia Circirelli, Donato Impedovo, Vincenzo Dentamaro, Roberto Marani, “Human Gait Analysis in Neurodegenerative Diseases”, Biomed health, 2022.
- [3] <http://automatika.ftn.uns.ac.rs/images/predmeti/VirtueIni%20Senzori/Predavanja/support%20vector%20machin e.pdf>
- [4] <https://www.bbc.com/serbian/lat/srbija-55063175>

Kratka biografija:



Svetlana Sekulić rođena je u Vrbasu 1995. god. Mater rad na Fakultetu tehničkih nauka iz oblasti Biomedicinskog inženjstva odbranila 2023. god.
kontakt:
svetlanasekulic95@gmail.com