

Klasifikacija mamograma primenom nove metode za selekciju najznačajnijih osobina slike

Marina Milošević, Dragan Janković, Đorđe Damjanović i Aleksandar Peulić

Apstrakt—Ovaj rad predstavlja sistem za klasifikaciju mamograma u jednu od dve kategorije, mamogrami sa tumorom i mamogrami bez tumora. Predložen sistem sastoji se od četiri modula: predobrade, koja podrazumeva uklanjanje šuma sa mamograma i izdvajanje područja od interesa, izdvajanja, selekcije osobina teksture slike i klasifikacije mamograma. Nakon predobrade originalnih snimaka, za svaki mamogram je izdvojeno ukupno 20 osobina slike. Da bi se povećala tačnost klasifikacije, primenom nove selekcionarne metode izvršen je odabir nekoliko najrelevantnijih osobina. Analizirani su rezultati predviđanja tri klasifikatora: SVM (Support Vector Machine) klasifikatora, Bayes-ovog i k-NN klasifikatora. U cilju poređenja eksperimentalnih rezultata, sistem za detektovanje tumora dojke testiran je na mamogramima iz MIAS (Mammographic Image Analysis Society) baze slika i mamogramima iz Kliničkog centra u Kragujevcu.

Ključne reči—Detektovanje tumora dojke; GLCM osobine; Selekcija osobina; Klasifikacija.

I. UVOD

Tumor dojke je najčešći oblik tumora kod žena širom sveta. Mamografija je veoma važna dijagnostička metoda, jer često može da otkrije tumor pre nego što žena oseti bilo kakve promene [1]. Digitalna mamografija donela je veliki napredak u računarskoj obradi slike.

Tumor dojke najčešće je zvezdastog oblika ali zvezdasta senka istovremeno je i veoma nesiguran pokazatelj tumora. Centralna masa tumora nije lako uočljiva jer nema jasno definisane ivice. Njihove veličine variraju od nekoliko milimetara do nekoliko centimetara [2]. Iz navedenih razloga tumor nije moguće detektovati na osnovu oblika ili veličine.

Kod mnogih sistema za obradu slike, osnovni cilj obrade je izvlačenje informacija iz slike pomoću kojih se može izvršiti računarska interpretacija i analiza slike. Takve informacije nazivaju se osobine (engl. *features*). Ulazni mamogram se mora prvo obraditi nekim od postupaka za popravku kvaliteta ili restauracije. Posle pretprocesiranja, iz slike se izdvajaju karakteristične osobine. Izdvajanje osobina ima za cilj da bez značajnog gubitka informacija smanji količinu podataka koji će se obrađivati u nastavku procesa analize mamograma. Izdvajanje osobina se može izvršiti izdvajanjem ivica, određivanjem oblika, opisom teksture, itd.

Marina Milošević – Fakultet tehničkih nauka, Univerzitet u Kragujevcu, Svetog Save 65, 32000 Čačak, Srbija (e-mail: marina.milosevic@ftn.kg.ac.rs).

Dragan Janković – Elektronski fakultet, Univerzitet u Nišu, Aleksandra Medvedeva 14, 18000 Niš, Srbija (e-mail: Dragan.Jankovic@elfak.ni.ac.rs).

Đorđe Damjanović – Fakultet tehničkih nauka, Univerzitet u Kragujevcu, Svetog Save 65, 32000 Čačak, Srbija (e-mail: djordje.damjanovic@ftn.kg.ac.rs).

Aleksandar Peulić – Fakultet tehničkih nauka, Univerzitet u Kragujevcu, Svetog Save 65, 32000 Čačak, Srbija (e-mail: aleksandar.peulic@ftn.kg.ac.rs).

Algoritmi za klasifikaciju vrše kategorizaciju uzoraka u odgovarajuće klase prema klasifikacionoj šemi. Uzorak je sačinjen od jedne ili više osobina. Tačnost klasifikacije uzoraka (mamograma) zavisi od izbora odgovarajućih osobina koje se dovode na ulaz klasifikatora. Šema klasifikacije obično je bazirana na skupu uzoraka koji je već ranije klasifikovan. Ovaj skup uzoraka naziva se trening skup (*training set*) ili skup za obučavanje, a sam proces se naziva obučavanje ili učenje.

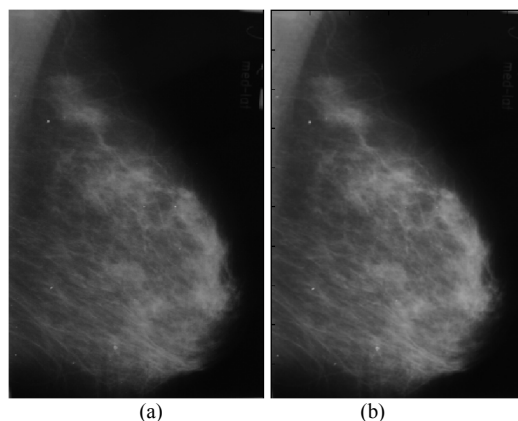
II. PRETPROCESIRANJE SLIKE

Mamogrami su medicinske slike koje je veoma teško analizirati. Pre same analize mamograma potrebno je omogućiti što bolje uslove za izdvajanje željenih objekata. To se postiže u fazi predobrade, prvenstveno eliminisanjem šuma i poboljšanjem kvaliteta slike. Za određivanje odgovarajućih karakteristika mamograma nije potrebno obrađivati celu sliku. Nakon uklanjanja šuma sa mamograma, prvi korak je određivanje područja od interesa (engl. Region of Interest-ROI), koji podrazumeva izdvajanje dela slike koji predstavlja dojku iz pozadine.

A. Uklanjanje šuma sa mamograma

Uklanjanje šuma sa slike je važan deo mnogih aplikacija za obradu slike, a postojeće metodologije uklanjanja šuma su veoma različite. Izbor metoda adekvatnog datoj primeni zavisi kako od konteksta tako i od cilja obrade.

Uklanjanje šuma sa mamograma izvršeno je primenom jedno-nivovske Diskretne Wavelet Transformacije (DWT), pomoću Haar-ovog talasića i vrednosti praga 50. Prag se obično bira na osnovu statističkih pretpostavki o šumu i o samom signalu koji se estimira. Na Sl. 1. prikazani su originalni i estimirani mamogram.



Sl. 1. Rezultat uklanjanja šuma sa mamograma iz kliničkog centra: (a) Originalni mamogram, (b) Estimirani mamogram.

B. Izdvajanje područja od interesa

Određivanje područja od interesa podrazumeva izdvajanje dojke iz pozadine. Ako se pritom radi o skeniranim analognim slikama, proces same segmentacije je zahtevniji, jer skenirane slike često sadrže nepoželjne objekte koje je potrebno ukloniti. Skenirani mamogrami takođe često imaju problem nejednako osvetljene pozadine, koji nastaje kao posledica nesavršenosti kako samog filma, tako i procesa naknadne digitalizacije. Procedura izdvajanja područja od interesa sastoji se od nekoliko osnovnih koraka, opisanih u narednim poglavljima.

1) Povećanje kontrasta

Kontrast slike povećan je primenom jednostavne logaritamske operacije. Logaritamska operacija definisana jednačinom:

$$g(x, y) = \log[k + f(x, y)] \quad (1)$$

primenjena je na originalnu sliku $f(x, y)$, gde je $g(x, y)$ slika sa povećanim kontrastom. Vrednost ulaznog parametra logaritamske funkcije k se može menjati zavisno od toga koliko je potrebno povećati kontrast. U ovom algoritmu kontrast mamograma je povećan postavljanjem vrednosti parametra $k=7$. Ova operacija, iako je primenjena na celu sliku, značajno povećava kontrast u područjima koja se nalaze u blizini ivice dojke, koje karakterišu male gustine tkiva i slabo definisani detalji slike. Cilj primene ove procedure na mammogram je određivanje aproksimacije ivice dojke što je moguće bliže stvarnoj ivici dojke [3]. Rezultat primene procedure za povećanje kontrasta može se videti poređenjem originalnog mamograma i mamograma sa povećanim kontrastom na Sl. 2 (a) i (b).

2) Kvantizacija

Kvantizacija je proces predstavljanja vrednosti odmeraka diskretnog signala ili slike, pomoću konačnog skupa dozvoljenih vrednosti. Kod digitalne reprezentacije, upotrebom n bita po odmerku i upotrebom samo pozitivnih celih brojeva, postoji tačno 2^n kvantizacionih nivoa koji pripadaju opsegu $[0, 2^n - 1]$. Za dobijanje binarne slike, korišćeno je $n=1$ bit za predstavljanje svakog piksela slike. Svaki piksel slike predstavljen je jednom od moguće dve vrednosti, "0" što odgovara crnoj boji ili "1" što odgovara beloj boji.

Neophodno je, na odgovarajući način, ograničiti opseg datog ulaznog signala, tako da odgovara ulaznom dinamičkom opsegu kvantizera. Donja granica dinamičkog opsega kvantizera postavljena je na vrednost 128, a gornja na vrednost 255. To znači da je svakom pikselu čija vrednost pripada opsegu $[128, 255]$, dodeljena vrednost "1", a svakom pikselu čija je vrednost van pomenutog opsega, dodeljena je vrednost "0". Na Sl. 2(c) prikazana je binarna verzija mamograma sa povećanim kontrastom.

3) Poboljšavanje binarne slike

Da bi male bele površine koje ne pripadaju dojci bile uklonjene, dojka je uobličena primenom jednog od elemenata za struktuiranje objekata slike. Na Sl. 2(d) prikazana je poboljšana verzija binarne slike.

4) Uklanjanje pozadine

Procedura za uklanjanje pozadine dojke biće opisana za LMLO (Left MLO) tip mamograma. Procedura uklanjanja pozadine kod RMLO (Right MLO) tipa mamograma je veoma slična, tako da nije potrebno opisivati obe.

Da bi pozadina dojke bila uklonjena, neophodno je kreirati binarnu masku. Binarna maska (Sl. 2(e)) kreirana je primenom algoritma opisanog u daljem tekstu.

Algoritam za formiranje binarne maske

Korak 1: Inicijalna vrsta je prva vrsta.

Korak 2: Skeniraj vrstu s desne na levu stranu.

Korak 3: Ako je piksel crn pređi na sledeći piksel i ponovi

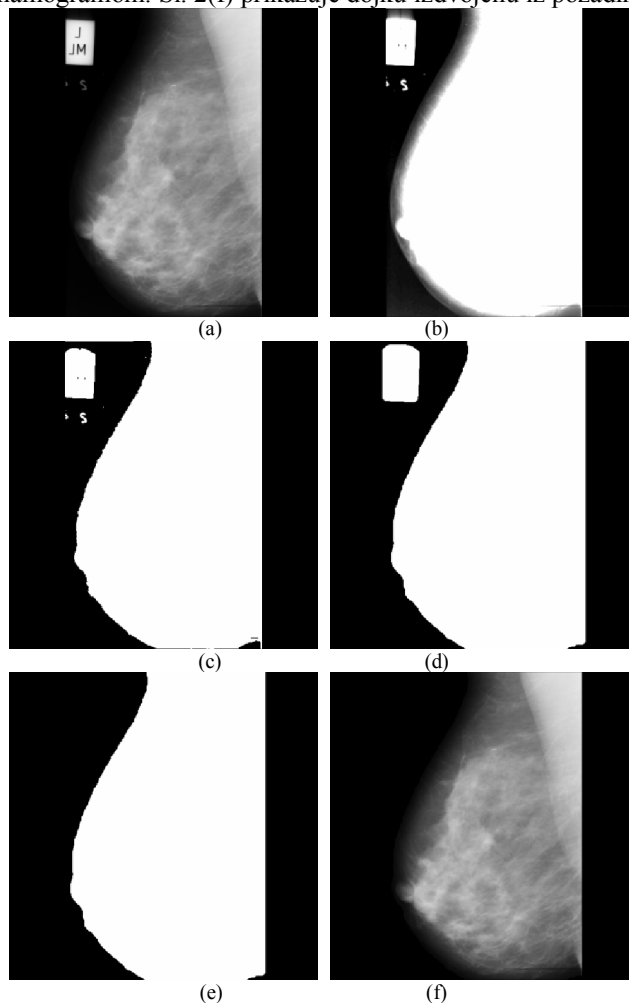
Korak 3. U suprotnom, ako je piksel beo pređi na Korak 4.

Korak 4: Prelazi na sledeći piksel sve dok je piksel beo. Kada piksel postane crn idi na Korak 5.

Korak 5: Tekući piksel zameni crnim pikselom i pređi na sledeći. Ako je tekuća kolona poslednja kolona, pređi na Korak 6, u suprotnom, ponovi Korak 5.

Korak 6: Ponovi korake 2-5 za sledeću vrstu.

Područje od interesa izdvojeno je množenjem binarne maske, kreirane na opisan način, sa originalnim mamogramom. Sl. 2(f) prikazuje dojku izdvojenu iz pozadine.



Sl. 2. Rezultati dobijeni u svakoj pojedinačnoj fazi procedure za izdvajanje područja od interesa: (a) Originalni mamogram, (b) Mamogram sa povećanim kontrastom, (c) Binarizovan mamogram, (d) Poboljšana verzija binarnog mamograma, (e) Binarna maska, (f) Izdvojen ROI.

III. KORIŠĆENE METODE

Procedura klasifikacije medicinskih slika, opisana u radu, sastoji se od tri osnovna koraka: (1) Izdvajanje osobina slike, (2) Selekcija osobina i (3) Klasifikacija slike.

A. Izdvajanje osobina slike

Osnovni cilj obrade slike je izvlačenje osobina iz slike pomoću kojih se može izvršiti računarska interpretacija i analiza slike. Tekstura slike opisuje prostorno uređenje nivoa sivog piksela u jednom regionu. Osobine tekture slike su veoma važne za klasifikaciju medicinskih slika i najčešće su korišćeno sredstvo za detektovanje tumora dojke na mamogramu [4]. Postoji veliki broj metoda za izdvajanje različitih osobina slike, a izbor se vrši zavisno od toga koje su osobine potrebne za dalju analizu.

Za izdvajanje osobina tekture slike korišćena je kookurensna matrica. Svaki element kookurensne matrice predstavlja procenu verovatnoće pojavljivanja para piksela određenih inteziteta u određenom regionu, na nekom rastojanju d . Zbog toga je ova statistička metoda dobila naziv *statistika združenog pojavljivanja nivoa sivog* (engl. Gray level cooccurrence - GLC). GLC matrice (GLCM) se mogu izračunavati za 4 pravca: 0° , 45° , 90° i 135° . Broj nivoa sivog na slici određuje dimenzije GLC matrice. Iz kookurensnih matrica se može izračunati veliki broj osobina tekture.

U ovom radu, GLC matrica je izračunata na dva različita načina i rezultati klasifikacije slike za tako izdvojene osobine su upoređeni. Kod prvog načina izračunavanja, koji je Haralick [5] predložio u svom radu, GLC matrica je izračunata za 4 pravca ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) i rastojanje između piksela $d=1$. Nakon izdvajanja osobina tekture iz svake GLC matrice, za svaku pojedinačnu osobinu tekture slike izračunate su dve vrednosti, prosečna vrednost i rang. U drugom slučaju, GLC matrica je izračunata samo za pravac 0° i rastojanje $d=1$. Eksperimentalni rezultati su pokazali da su primenom druge metode za izdvajanje osobina tekture slike dobijeni mnogo bolji rezultati klasifikacije. Prema tome, u daljem tekstu biće analizirane samo GLCM osobine tekture izračunate za pravac 0° i rastojanje $d=1$.

Analizirano je ukupno 20 osobina slike. Osobine tekture, $f1-f13$, je predložio Haralick, Soh [6] je definisao osobine $f14-f18$ a, $f19$ i $f20$ su osobine koje je opisao Clausi [7].

Za definisanje GLCM osobina slike korišćene su sledeće oznake:

$P(i,j)$ – Pikel u i -toj vrsti i j -toj koloni GLC matrice.

N_g – broj nivoa sivog.

$$\sum_i X_i = \sum_{i=0}^{N_g} X_i, \quad \sum_j X_j = \sum_{j=0}^{N_g} X_j$$

$$P_x(i) = \sum_{j=1}^{N_g} P(i,j), \quad P_y(i) = \sum_{i=1}^{N_g} P(i,j)$$

$$P_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j), i+j=k, k=2,3,\dots,2N_g$$

$$P_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j),$$

$$|i-j|=k, k=0,1,\dots,N_g-1$$

$$\mu_x = \sum_i \sum_j iP(i,j), \quad \mu_y = \sum_i \sum_j jP(i,j)$$

$$\mu = \sum_{i,j} iP(i,j)$$

$$\sigma_x = \sum_i \sum_j (i - \mu_x)^2 P(i,j),$$

$$\sigma_y = \sum_i \sum_j (j - \mu_y)^2 P(i,j)$$

Obeležja tekture slike:

1. Angular Second Moment (Energy)

$$f1 = \sum_i \sum_j \{P(i,j)\}^2$$

2. Contrast

$$f2 = \sum_{n=0}^{N_g-1} n^2 \{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \}, |i-j|=n$$

3. Correlation

$$f3 = \frac{\sum_i \sum_j (ij)P(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

4. Sum of Squares: Variance

$$f4 = \sum_i \sum_j (i - \mu)^2 P(i,j)$$

5. Inverse Difference Moment (Homogeneity)

$$f5 = \sum_i \sum_j \frac{P(i,j)}{1 + (i-j)^2}$$

6. Sum Average

$$f6 = \sum_{i=2}^{2N_g} iP_{x+y}(i)$$

7. Sum Variance

$$f7 = \sum_{i=2}^{2N_g} (i - f8)^2 P_{x+y}(i)$$

8. Sum Entropy

$$f8 = -\sum_{i=2}^{2N_g} P_{x+y}(i) \log \{P_{x+y}(i)\}$$

9. Entropy

$$f9 = -\sum_i \sum_j P(i, j) \log(P(i, j))$$

10. Difference Variance

$$f10 = \sum_{i=0}^{N_g-1} i^2 P_{x-y}(i)$$

11. Difference Entropy

$$f11 = -\sum_{i=0}^{N_g-1} P_{x-y}(i) \log \{P_{x-y}(i)\}$$

12. Information Measure of Correlation 1

$$f12 = \frac{HXY - HXY1}{\max \{HX, HY\}}$$

13. Information Measure of Correlation 2

$$f13 = (1 - \exp[-2.0(HXY2 - HXY)])^{\frac{1}{2}}$$

Gde su: $HXY = -\sum_i \sum_j P(i, j) \log(P(i, j))$

$$HXY1 = -\sum_i \sum_j P(i, j) \log \{P_x(i) P_y(j)\}$$

$$HXY2 = -\sum_i \sum_j P_x(i) P_y(j) \log \{P_x(i) P_y(j)\}$$

14. Autocorrelation

$$f14 = \sum_i \sum_j (ij) P(i, j)$$

15. Dissimilarity

$$f15 = \sum_i \sum_j |i-j| P(i, j)$$

16. Cluster Shade

$$f16 = \sum_i \sum_j (i+j - \mu_x - \mu_y)^3 P(i, j)$$

17. Cluster Prominence

$$f17 = \sum_i \sum_j (i+j - \mu_x - \mu_y)^4 P(i, j)$$

18. Maximum Probability

$$f18 = \max_{i,j} P(i, j)$$

19. Inverse Difference Normalized

$$f19 = \sum_i \sum_j \frac{P(i, j)}{1 + |i-j| / N_g^2}$$

20. Inverse Difference Moment Normalized

$$f20 = \sum_i \sum_j \frac{P(i, j)}{1 + (i-j)^2 / N_g^2}$$

B. Selekcija osobina slike

Selekcijom izdvojenih osobina slike smanjuje se broj osobina na osnovu kojih se vršiti klasifikacija sa ciljem da se poveća tačnost klasifikacije a, smanji vreme izračunavanja i kompleksnost izračunavanja [8]. U cilju poboljšanja efikasnosti faze selekcije osobina slike, ovaj rad predlaže jednostavan algoritam za selekciju osobina, kao i proceduru koja kombinuje rezultate selekcije predloženog algoritma i rezultate statističkog t-testa.

Algoritam za selekciju osobina slike:

Korak 1: Izdvoj osobinu iz N normalnih mamograma. Neka to bude osobina A_i ($i=1, 2, \dots, N$).

Korak 2: Izdvoj osobinu iz N mamograma sa tumorom. Neka to bude osobina B_i ($i=1, 2, \dots, N$).

Korak 3: Izračunaj sumu S_A za N normalnih mamograma.

$$S_A = \sum_{i=1}^N A_i$$

Korak 4: Izračunaj sumu S_B za N mamograma sa tumorom.

$$S_B = \sum_{i=1}^N B_i$$

Korak 5: Izračunaj količnik K:

Ako je $S_A > S_B$

$$K = S_A / S_B$$

U suprotnom

$$K = S_B / S_A.$$

Korak 6: Ponovi korake 1 - 5 za svih 20 osobina.

Korak 7: Na osnovu vrednosti količnika K sortiraj osobine slike u opadajućem redosledu.

Korak 8: Označi sedam najznačajnijih osobina slike.

C. Klasifikacija mamograma

Za klasifikaciju mamograma korišćena su tri različita klasifikatora: klasifikator zasnovan na potpornim vektorima (*engl. Support Vector Machine - SVM*), Bayes-ov klasifikator i k-NN (*engl. K-Nearest Neighbor*) klasifikator.

Proces klasifikacije sastoji se od dve osnovne faze: (1) faza formiranja modela tj. obučavanje, (2) faza testiranja. U fazi obučavanja, klasifikatoru se predaju poznati podaci. U fazi testiranja, klasifikator vrši klasifikaciju primeraka test skupa u unapred utvrđene klase.

IV. MERE ZA ANALIZU TAČNOSTI

Za analizu tačnosti klasifikacije korišćene su tri mere: Preciznost (Accuracy - AC), Senzitivnost (Sensitivity - SE) i Specifičnost (Specificity - SP) [9]. Ukupan broj ispravno klasifikovanih mamograma definiše preciznost klasifikacije. Senzitivnost meri tačnost u pozitivnim primercima, a specifičnost predstavlja tačnost u negativnim primercima. Pomenute mere za analizu tačnosti klasifikacije definisane su sledećim izrazima:

$$AC = (TP+TN) / (TP+FP+TN+FN) \quad (2)$$

$$SE = TP / (TP+FN) \quad (3)$$

$$SP = TN / (TN+FP) \quad (4)$$

Gde su TP – stvarno pozitivni (engl. *True Positive Rate*) i TN – stvarno negativni (engl. *True Negative Rate*) rezultati predviđanja koji predstavljaju ispravnu klasifikaciju a, FP – lažno pozitivni (engl. *False Positive Rate*) i FN – lažno negativni (engl. *False Negative Rate*) rezultati predviđanja koji predstavljaju dva moguća tipa greške.

V. EKSPERIMENTALNI REZULTATI

Klasifikovane su dve grupe mamograma: mamogrami iz MIAS baze slika i mamogrami iz Kliničkog centra u Kragujevcu. Testirano je ukupno 100 mamograma, 25 normalnih i 25 mamograma sa tumorom iz MIAS baze i 25 normalnih i 25 mamograma sa tumorom iz kliničkog centra. Posle preprocesiranja ulazne slike, iz slike je izdvojeno 20 osobina i označeno je 5-6 najznačajnijih osobina. Najrelevantnije osobine postavljene su na ulaz klasifikatora koji svrstava sliku u jednu od dve kategorije. Izbor klasifikatora za generisanje preciznih modela predviđanja predstavlja složen zadatak. Iz tog razloga implementirana su tri različita klasifikatora, SVM, Bayes-ov i k-NN klasifikator. Eksperiment je izveden za obe grupe mamograma odvojeno i napravljeno je poređenje rezultata oba eksperimenta.

Za svaki mamogram izdvojeno je ukupno 20 teksturalnih osobina slike. Pet – šest najznačajnijih osobina slike odabrano je primenom procedure koja kombinuje rezultate predloženog algoritma za selekciju osobina i statističkog testa – t-testa.

Predložena procedura za selekciju najznačajnijih osobina slike:

Korak 1: Izdvoj 20 osobina iz slike.

Korak 2: Primenom predloženog algoritma za odabir najrelevantnijih osobina slike kreiraj skup $n1$ koji čini 7 najznačajnijih osobina.

Korak3: Primenom t-testa kreiraj skup $n2$ koji čini 7 najznačajnijih osobina.

Korak 4: Nađi presek skupova $n1$ i $n2$, $n1 \cap n2 = n$.

Korak 5: Nađi skup p koji čine osobine koje se nalaze na prvom, drugom ili trećem mestu u skupovima $n1$ i $n2$ a, ne pripadaju skupu n .

Korak 6: Unija skupova n i p predstavlja skup osobina slike koje se predaju klasifikatoru.

U Tabeli I prikazane su osobine mamograma iz MIAS baze označene kao najznačajnije primenom predloženog algoritma za selekciju (skup $n1$) i primenom t-testa (skup $n2$). Vrednosti u koloni *Redni broj* označavaju značaj osobine za klasifikaciju slike.

TABELA I

NAJZNAČAJNIJE OSOBINE MAMOGRAMA IZ MIAS BAZE

Redni Broj	Skup osobina n1	Skup osobina n2
1	Cluster Shade (f16)	Correlation (f3)
2	Cluster Prominence(f17)	Cluster Shade (f16)

3	Contrast (f2)	Cluster Prominence(f17)
4	Difference Variance (f10)	Contrast (f2)
5	Dissimilarity (f15)	Difference Variance (f10)
6	Difference Entropy (f11)	Inverse Difference Moment Normalized (f20)
7	Maximum Probability (f18)	Informaiton Measure of Correlation 2 (f13)

Skup n je definisan kao presek skupova $n1$ i $n2$. Prema tome, osobine f2, f10, f16 i f17 čine skup n . Prve tri pozicije u skupu $n1$ zauzimaju osobine slike koje pripadaju skupu n , tako da se ne mogu naći u skupu p . Osobina f3, koja se nalazi na prvom mestu u skupu $n2$, pridružuje se skupu p . Osobine koje zauzimaju drugo i treće mesto u skupu $n2$ takođe pripadaju skupu n pa se ne mogu pridružiti skupu p . Konačno, skup p sadrži samo jednu osobinu slike – f3.

Unijom skupova n i p određene su najrelevantnije osobine za klasifikaciju mamograma iz MIAS baze slika a, to su osobine f2, f3, f10, f16 i f17.

Zatim, predložena procedura za selekciju najznačajnijih osobina mamograma primenjena je na mamograme iz kliničkog centra. U Tabeli II prikazane su osobine mamograma iz kliničkog centra označene kao najznačajnije primenom predloženog algoritma za selekciju (skup $n1$) i primenom t-testa (skup $n2$). Vrednosti u koloni *Redni broj* označavaju značaj osobine za klasifikaciju slike.

TABELA II

NAJZNAČAJNIJE OSOBINE MAMOGRAMA IZ KLINIČKOG CENTRA

Redni Broj	Skup osobina n1	Skup osobina n1
1	Variance (f4)	Information Measure of Correlation 1 (f12)
2	Autocorrelation (f14)	Sum Variance (f7)
3	Sum Variance (f7)	Correlation (f3)
4	Cluster Prominence (f17)	Autocorrelation (f14)
5	Sum Average (f6)	Variance (f4)
6	Sum Entropy (f8)	Sum Average (f6)
7	Entropy (f9)	Maximum Probability (f18)

Skup n kod mamograma iz kliničkog centra čine osobine f4, f6, f7 i f14. Prve tri pozicije u skupu $n1$ zauzimaju osobine slike koje pripadaju skupu n , tako da se ne mogu naći u skupu p . Skup p sadrži dve osobine iz skupa $n2$ – f3 i f12.

Unijom skupova n i p određene su najrelevantnije osobine za klasifikaciju mamograma iz kliničkog centra a, to su osobine f3, f4, f6, f7, f12 i f14.

Odabranih pet (šest) osobina slike korišćeno je za klasifikaciju mamograma pomoću SVM klasifikatora, Bayes-ovog klasifikatora i k-NN klasifikatora. U Tabeli III prikazani su rezultati klasifikacije 10 slučajno izabranih mamograma.

TABELA III
REZULTATI KLASIFIKACIJE 10 SLUČAJNO IZABRANIH UZORAKA

	f2, f3, f10, f16, f17		f3, f4, f6, f7, f12, f14	
	MIAS	Klinički centar	MIAS	Klinički centar
SVM	6 ispravnih 4 pogrešna	7 ispravnih 3 pogrešna	5 ispravnih 5 pogrešnih	9 ispravnih 1 pogrešan
Bayes	7 ispravnih 3 pogrešna	7 ispravnih 3 pogrešna	4 ispravna 6 pogrešnih	9 ispravnih 1 pogrešan
k-NN	5 ispravnih 5 pogrešnih	6 ispravnih 4 pogrešna	3 ispravna 7 pogrešnih	7 ispravnih 3 pogrešna

SVM klasifikator i Bayes-ov klasifikator daju slične rezultate klasifikacije dok rezultati dobijeni korišćenjem k-NN klasifikatora su značajno lošiji kod oba tipa mamograma. Najbolji rezultati klasifikacije u Tabeli III su rezultati dobijeni klasifikacijom mamograma iz kliničkog centra, korišćenjem SVM i Bayes-ovog klasifikatora i osobina slike odabranih u eksperimentu sa slikama iz kliničkog centra, f3, f4, f6, f7, f12 i f14.

Rezultati klasifikacije mamograma iz kliničkog centra su bolji od rezultata klasifikacije mamograma iz MIAS baze slika čak i u slučaju klasifikacije bazirane na 5 najznačajnijih osobina dobijenih u eksperimentu sa MIAS slikama.

Vrednosti pokazatelja tačnosti predloženog sistema za detektovanje tumora dojke prikazane su u Tabeli IV.

TABELA IV
PERFORMANSE PREDLOŽENOG SISTEMA ZA OBA TIP
MAMOGRAMA

	Mamogrami iz MIAS baze slika			Mamogrami iz kliničkog centra		
	AC	SE	SP	AC	SE	SP
SVM	64%	67%	50%	92%	85,70%	100%
Bayes	75%	71.50%	67%	92%	85,70%	100%
k-NN	56%	57%	33%	73%	80%	60%

VI. ZAKLJUČAK

Razvoj i unapređenje CAD (Computer Aided Diagnosis) sistema odvija se prilično brzo, ali njihova tačnost još uvek nije zadovoljavajuća. Zbog toga nije moguće osloniti se na dijagnozu koju oni nude, već ih je potrebno koristiti isključivo nakon donošenja dijagnoze, kako bi se ona potvrdila ili uočio mogući previd.

Mere izračunate iz GLC matrice su se pokazale kao veoma pouzdane, zbog čega je GLCM metoda jedna od najčešće korišćenih metoda za kvantitativnu karakterizaciju teksture, bilo u praktičnoj primeni, bilo u teorijskoj analizi, kao polazna tačka za ocenu novih algoritama. Pokazano je da su rezultati klasifikacije mamograma iz kliničkog centra mnogo bolji od rezultata dobijenih u slučaju klasifikacije mamograma iz

MIAS baze slika, za sve grupe izdvojenih osobina i sva tri klasifikatora.

Dalji rad biće usmeren ka razvijanju novih metoda za selekciju najrelevantnijih osobina mamograma, metoda za segmentiranje mamograma te klasifikaciju segmentiranih uzoraka, kao i povećanju broja klasnih grupa.

LITERATURA

- [1] R. A. Papadopoulos, D. I. Fotiadis, A. Likas, "An automatic microcalcification detection system based on a hybrid neural network classifier," *Artificial Intelligence in Medicine*, vol. 25, pp.149–167, 2002.
- [2] S. Liu, C. F. Babbs, E.J. Delp, "Multiresolution detection of speculated lesions in digital mammograms," *IEEE Transactions Image Processing*, vol. 10, pp. 874-884, Jun 2001.
- [3] R. M. Rangayyan, *Biomedical Image Analysis*, USA: CRC Press LLC, 2005.
- [4] N. Pradeep, H. Girisha, K. Karibasappa, "Segmentation and Feature Extraction of Tumors from Digital Mammograms," *Computer Engineering and Intelligent Systems*, vol. 3, no. 4, pp. 37- 46, 2012.
- [5] R. M. Haralick, K. Shanmugam, I. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man and cybernetics*, vol. 3, no. 6, pp. 610-621, 1973.
- [6] L. K. Soh, C. Tsatsoulis, "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices," *IEEE Transactions on geoscience and remote sensing*, vol. 37, no. 2, pp. 780-795, 1999.
- [7] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Canadian Journal of Remote Sensing*, vol. 28, no. 1, pp. 45–62, 2002.
- [8] M. A. Alolfe, W. A. Mohamed, A. M. Youssef, Y. M. Kadah, A. S. Mohamed, "Feature selection in computer aided diagnostic system for microcalcification detection in digital mammograms," 26th National Radio Science Conference, New Cairo, Egypt, 2009.
- [9] B. Verma, P. McLeod, A. Klevansky, "Classification of benign and malign patterns in digital mammograms for the diagnosis of breast cancer," *Expert System with Applications*, vol. 37, no. 4, pp. 3344-3351, 2010.

ABSTRACT

This paper presents a system for mammograms classification into one of two categories, mammograms with breast cancer and mammograms without breast cancer. The proposed system consists of four modules: preprocessing, which includes noise reduction and region of interest extraction, textural features extraction and selection and mammogram classification. After original images preprocessing, a total of 20 GLCM features are extracted for each mammogram. In order to improve the accuracy of classification stage, the most relevant features are selected by selection algorithm proposed in this paper. Prediction results of three different classifiers were analyzed: SVM classifier, Bayes and k-NN classifier. In order to compare experimental results, system for breast tumors detection was tested for mammographic images from Mammographic Image Analysis Society (MIAS) digital mammogram database and mammographic images from the Clinical Center in Kragujevac.

Mammogram classification applying a new method for the most important image features selection

Marina Milošević, Dragan Janković, Đorđe Damjanović and Aleksandar Peulić