

PROIECT INGINE	ERIA REGLARII AUTOMATE II				
NUME student	Scoartă Tudor-Claudiu	GRUPA:	30135	Nota	

Detecția cancerului de sân folosind rețele neuronale și procesarea imaginilor

Autor: Scoarță Tudor-Claudiu

Grupa: 30135

AN UNIVERSITAR: 2022-2023

PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

Cuprins

1.		Scopul Proiectului	3
a	١.	Obiective	3
b).	Specificații	3
2.		Studiu bibliografic	3
3.		Analiză, proiectare, implementare	5
1.		Îmbunătățirea calității imaginii	5
		Pași algoritm:	5
		Afișare rezultate ale îmbunătățirii imaginii:	7
1	I.	Segmentare	8
		Thresholding (prăguire) cu algoritmul lui Otsu:	8
		Algoritmul pentru decuparea imaginii:	9
		Imagine îmbunătățită și segmentată:	10
I	II.	. Algoritmul pentru obținerea datelor imaginii (Algoritmul GLCM (Gray Level Co-Occurrence	
N	N	atrix))	11
ľ	V	. Clasificare bazată pe rețele neuronale	12
4.C	0	ncluzii	15
a	١.	Rezultate obținute	15
k).	Direcții de dezvoltare	16
5.B	ib	oliografie	17

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

1. Scopul Proiectului

Scopul acestui proiect este de a dezvolta o metodă cât mai precisă pentru detecția cancerului de sân, utilizând tehnici avansate de procesare a imaginilor și clasificarea bazată pe rețele neuronale. Prin analiza detaliată a imaginilor medicale, dorim să identificăm semnele precoce ale cancerului de sân, întro etapă cât mai timpurie, când tratamentul are cele mai mari șanse de succes. Prin utilizarea algoritmilor de învățare automată, vom antrena o rețea neuronală să recunoască anomalii și să ofere diagnosticuri precise. Scopul final este de a îmbunătăți rata de supraviețuire a pacienților și calitatea vieții acestora prin diagnostic precoce și tratament eficient.

a. Objective

- -antrenarea unei rețele neuronale pe un set de mamografii digitalizate
- -prelucrarea mamografiilor astfel încât să se elimine informația inutilă
- -pe baza unei mamografii pe care utilizatorul o încarcă, programul returnează diagnosticul pentru țesutul analizat

b. Specificații

- după cum aminteam și mai sus, funcționalitatea principală a aplicației o reprezintă categorisirea unei imagini
- -aplicația va returna un diagnostic într-o interfață grafică unde utilizatorul poate să încarce altă mamografie digitalizată, să vadă rezultatul și procesul prin care trece imaginea pentru a putea fi utilizată de către modelul antrenat
- din punctul de vedere al performanței, ne dorim o performanță cât mai bună, însă această nu este posibilă întodeauna. Cauzele pot și multiple, printre care amintin: date de antrenament defectuoase sau un număr mult mai mare de date aparținând unei categorii, raportat la celelalte.

2. Studiu bibliografic

Cancerul de sân reprezintă cel mai întâlnit tip de cancer la femei și este a doua cauză principală a deceselor din cauza tumorilor maligne, după cancerul pulmonar, în rândul populației feminine. Acest tip de boală se inițiază atunci când celulele glandei mamare suferă mutații genetice și încep să se înmulțească necontrolat, formând o tumoră.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

Cel mai des întâlnită manifestare a cancerului de sân constă în apariția unui nodul, ceea ce înseamnă o masă solidă, nedureroasă, cu margini neregulate. În general, această formațiune poate indica prezența unei afecțiuni canceroase. Cu toate acestea, este important să știți că și formațiunile moi, rotunjite și nedureroase pot fi de natură canceroasă. Numai investigațiile recomandate de medicul oncolog pot oferi informații precise despre tipul de leziune prezentă.

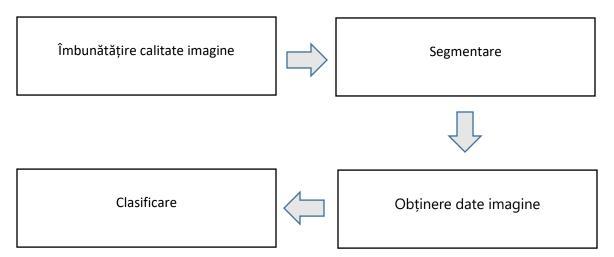
În situații de suspiciune avansată, se poate efectua o biopsie pentru a obține o confirmare diagnostică. Prin utilizarea unui sistem de asistență medicală, putem reduce numărul de biopsii inutile. Acest lucru ajută și medicul să ia decizii mai informate cu privire la necesitatea unui procedeu invaziv sau, în unele cazuri, să monitorizeze evoluția nodulului în mod regulat. Astfel, se poate asigura o abordare mai precisă și eficientă în evaluarea și tratamentul pacientului.

În trecut, s-au utilizat rețele neuronale pentru a clasifica nodulii detectați ca fiind benigni sau maligni, fără a lua în considerare în mod obligatoriu caracteristicile adiționale ale acestuia, care afectează pielea mamelonului și areola. Clasificarea se bazează pe imagini de antrenament cu rezultate cunoscute.

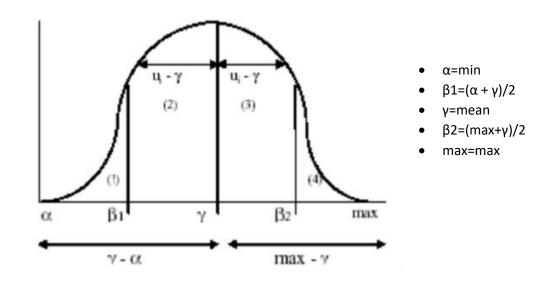
Această abordare este abordată și în acest studiu, unde clasificarea se realizează pe baza unor imagini ale căror rezultate sunt cunoscute, având scopul de a evidenția eficiența rețelei pe imaginile din setul de testare. Ulterior, pot fi efectuate teste utilizând imagini noi și evaluând rezultatul obținut.

PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

3. Analiză, proiectare, implementare



I. Îmbunătățirea calității imaginii



Pași algoritm:

- 1. Inițializare parametrii:
 - -se inițializează matricea imaginii
 - -se caută valoarea minimă, maximă și cea medie pentru griul din imagine

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

```
data2 = im2double(imread('image.jpg'));
data = rgb2gray(data2);
sizes = size(data);
height = sizes (1);
width = sizes (2);
newgraylevel = zeros(height,width);
fuzzydata = zeros(height,width);
enhanceddata = zeros(height,width);
min = 256;
maks = 1;
total = 0;
```

- β1 se calculează astfel: (valoare minimă + valoare medie)/2
- β2 se calculează astfel: (valoare maximă + valoare medie)/2

```
if data(i,j) < min
min = data(i,j);
end
if data(i,j) > maks
maks = data(i,j);
end

means = mean2(data);
b1 = (min+means)/2;
b2 = (maks+means)/2;
```

2. Fuzzificare:

-se aplică regulile de calcul pentru modificarea intensității pixelilor astfel, unde ui este intensitatea unui pixel:

1. Dacă $\alpha \le ui \le \beta 1$, atunci $P = 2((ui - \alpha)/(\gamma - \alpha))2$

```
if (data(i,j)>= min) && (data(i,j) < b1)
newgraylevel(i,j) = 2*(((data(i,j)-min)/(means-min))^2);
end</pre>
```

2. Dacă $\beta 1 \le ui \le \gamma$, atunci $P = 1 - 2((ui - \gamma)/(\gamma - \alpha))2$

```
if (data(i,j)>= b1) && (data(i,j) < means)
newgraylevel(i,j) = 1-(2*(((data(i,j)-means)/(means-
min))^2));
end</pre>
```

3. Dacă $\gamma \le ui \le \beta 2$, atunci $P = 1 - 2((ui - \gamma)/(max - \gamma))2$

PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

```
if (data(i,j)>= means) && (data(i,j) < b2)
newgraylevel(i,j) = 1-(2*(((data(i,j)-means)/(maks-
means))^2));
end</pre>
```

4. Dacă $\beta 2 \le ui \le max$, atunci $P = 2((ui - \gamma)/(max - \gamma))2$

```
if (data(i,j)>= b2) && (data(i,j) < maks)
newgraylevel(i,j) = 2*(((data(i,j)-means)/(maks-
means))^2);
end</pre>
```

3. Modificare:

Pentru fiecare pixel se recalculeză valoarea care este obținută din noua valoare de gri ridicată la puterea a doua.

```
fuzzydata(i,j) = newgraylevel(i,j)^2;
```

4.Defuzzificare:

Pentru toți pixelii, se calculează noile valori de contrast, care sunt obținute din valorile fuzzy calculate și înmulțite cu matricea inițială a imaginii.

```
enhanceddata(i,j) = fuzzydata(i,j)*data(i,j);
```

Afișare rezultate ale îmbunătățirii imaginii:

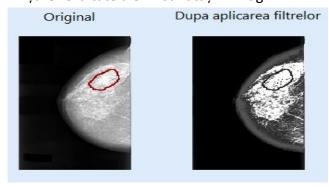


figure
subplot(1,2,1);
imshow(data)
subplot(1,2,2);
imshow(enhanceddata)

PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

II. Segmentare

Thresholding (prăguire) cu algoritmul lui Otsu:

Scopul algoritmului lui Otsu este de a segmenta imaginile prin împărțirea lor în două clase, respectiv fundal (valoarea setată la 0) și obiect (valoarea setată la 1), utilizând un anumit nivel ca barieră. Pașii efectuați pentru realizarea pragului Otsu sunt următorii:

1. Ni este numărul pixelilor la nivelul I, iar N este numărul total de pixeli

$$p_i = n_i/N$$
, $p_i \ge 0$, $\sum_{i=1}^{L} p_i = 1$.

p = counts / sum(counts);

2.Se împart in cele 2 clase: CO reprezentând fundalul, iar C1 obiectul (zona de interes).

$$\omega_0 = \operatorname{Pr} (C_0) = \sum_{i=1}^{k} p_i = \omega(k)$$

$$\omega_1 = \operatorname{Pr} (C_1) = \sum_{i=k+1}^{L} p_i = 1 - \omega(k)$$

$$\mu(k) = \sum_{i=1}^{k} i p_i$$

$$C_{\circ}$$
 -> omega = cumsum(p);
 C_{1} -> 1 - omega;

mu = cumsum(p .* (1:num_bins)');

$$\mu_T = \mu(L) = \sum_{i=1}^{L} i p_i$$

mu_t = mu(end);

3. Variația de clasă se obține prin următoarea formulă:

$$\sigma_W^2 = \omega_0 \sigma_0^2 + \omega_1 \sigma_1^2$$

$$\sigma_B^2 = \omega_0 (\mu_0 - \mu_T)^2 + \omega_1 (\mu_1 - \mu_T)^2$$

$$= \omega_0 \omega_1 (\mu_1 - \mu_0)^2$$

due to (9)) and

$$\sigma_T^2 = \sum_{i=1}^L (i - \mu_T)^2 p_i$$

 $sigma_b_squared = (mu_t * omega - mu).^2 ./ (omega .* (1 - omega));$

4. Se calculează valoarea de threshold:

$$\sigma_B^2(k^*) = \max_{1 \le k \le L} \sigma_B^2(k).$$

$$S^* = \{k; \omega_0 \omega_1 = \omega(k)[1 - \omega(k)] > 0, \quad \text{or } 0 < \omega(k) < 1\}.$$

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

```
maxval = max(sigma_b_squared);
if isfinite(maxval)
idx = mean(find(sigma_b_squared == maxval));
level = (idx - 1) / (num_bins - 1);
else
level = 0.0;
```

Algoritmul pentru decuparea imaginii:

Pentru a putea decupa o imagine, trebuie să definim marginile acesteia. Pentru a determina dacă un pixel este marginal sau nu, se face o buclă în care determinăm dacă acel pixel are intensitatea 1. Dacă acesta este de intensitate 1, acel pixel este marginal. După definirea acestor margini, se decupează zona de interes.

```
%initialize the initial
limit for cropping the
image
batas atas = 1;
batas bawah = 1;
batas_kanan = 1;
batas_kiri = 1;
% find the lower bound of
the image
while (true)
for i=1:row
for j=1:col
if citra(i,j) == 1
batas bawah = i;
break;
end
end
end
break;
end
% find the right border of
the image while (true)
for j=1:col
for i=1:row
```

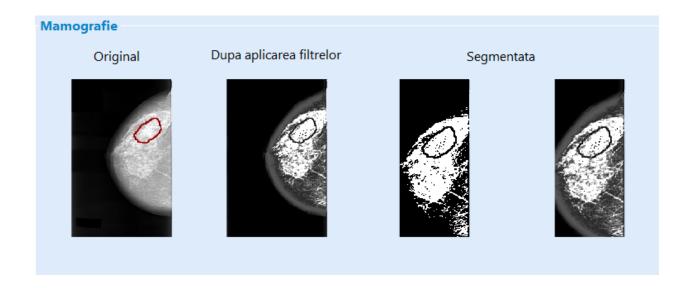
```
if citra(i,j) == 1
batas kanan = j;
break;
end
end
end
break;
% find the upper bound of the
image while (true)
for i=row:-1:1
for j=col:-1:1
if citra(i,j) == 1
batas atas = i;
break;
end
end
end
break;
end
% find the left border of the
while (true)
for j=col:-1:1
for i=row:-1:1
```

PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

```
for i=row:-1:1
if citra(i,j) == 1
batas_kiri = j;
break;
end
end
end
break;
end
data cropped double =
zeros (batas bawah-
batas atas+\overline{1}, batas kanan-
batas kiri+1);
data cropped int =
zeros (batas_bawah-
batas atas+1, batas kanan-
batas kiri+1);
%set crop
for i=batas atas:batas bawah
for j=batas kiri:batas kanan
%cuts the 'image' with the
calculated limit
data cropped double ((i-
batas_atas)+1, (j-
batas kiri) +1) = citra
```

```
(i,j);
%cuts the 'enhanceddata'
based on the calculated
limit
data_cropped_int ((i-
batas_atas)+1, (j-
batas_kiri)+1) =
enhanceddata(i,j);
end
```

Imagine îmbunătățită și segmentată:



PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

III. Algoritmul pentru obținerea datelor imaginii (Algoritmul GLCM (Gray Level Co-Occurrence Matrix))

Exemplu de matrice:

1	2	2	3
3	3	4	5
1	3	1	1
2	4	1	5

Contrast
$$\sum_{i,j=0}^{G-1} (i-j)^2 P(i,j)$$

Varianța
$$\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (p(i,j) - \mu)^2$$

Deviație standard
$$\sigma = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (p(i,j) - \mu)^2}$$

Kurtosis
$$K = \left\{ \frac{1}{MN} \sum_{i=1}^{M} \sum_{i=1}^{N} \left\lceil \frac{p(i,j) - \mu}{\sigma} \right\rceil^{4} \right\} - 3$$

Media
$$\mu = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} p(i,j)$$

Zgomot redus
$$R = 1 - \frac{1}{1 + \sigma^2}$$

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

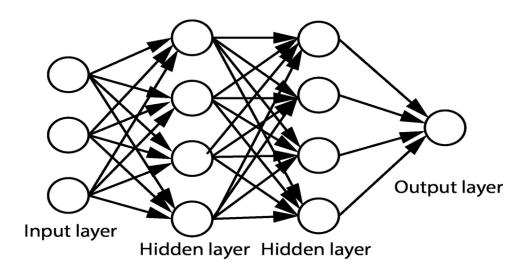
```
function [cont, vars, stds, kurt, men,
smo]=feature extraction(data cropped int)
glcm = graycomatrix(data cropped int);
%search for contrast, variance, standard deviation, curtosis, mean, and
smoothness values of segmented images
stats = graycoprops(glcm, {'contrast'});
cont = struct2array(stats);
vars = var(data cropped int(:));
stds = std(data cropped int(:));
kurt = kurtosis(data cropped int(:));
men = mean(data cropped int(:));
smo = 1 - (1/(1 + vars));
%mencetak nilai-nilai dari ekstraksi fitur
fprintf('\t\t %f\n', cont);
fprintf('\t\t %f\n', vars);
fprintf('\t\t %f\n', stds);
fprintf('\t\t %f\n', kurt);
fprintf('\t\t %f\n', men);
fprintf('\t\t %f\n', smo);
end
```

Fișier	Contrast	Deviație standard	Kurtosis	Media	Varianță	Zgomot redus
cancer001	1.027975	0.628802	9.45129	0.539653	0.395393	0.283356
cancer002	1.39875	0.685388	9.736556	0.57835	0.469757	0.319615
cancer004	0.737317	0.605021	7.130871	0.527126	0.36605	0.267962
cancer005	0.649256	0.952167	3.081397	0.95672	0.906621	0.475512
cancer008	0.791496	0.59201	14.38957	0.418011	0.350476	0.25952
normal001	0.1923	0.148253	34.16894	0.188447	0.021979	0.021506
normal002	0.369617	0.22563	27.85677	0.22431	0.050909	0.048443
normal003	0.313964	0.17338	6.763178	0.195019	0.03006	0.029183
normal005	0.750176	0.444575	15.89192	0.41034	0.197647	0.16503

IV. Clasificare bazată pe rețele neuronale

Rețeaua neuronală este un model creat pentru a imita funcția de învățare a creierului uman. Într-o rețea neurală, neuroni sunt grupați în straturi numite straturi neuronale. De obicei, fiecare neuron dintr-un strat este conectat la toți neuronii din stratul anterior și următor (cu excepția stratului de intrare și de ieșire). Informația este propagată într-o rețea neurală, strat cu strat, începând de la intrare până la ieșire, fie prin intermediul unului sau mai multor straturi ascunse. În funcție de algoritmul utilizat, informația poate fi propagată și în sens invers (backpropagation). Imaginea de mai jos prezintă o rețea neurală cu trei straturi neuronale.

PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	



Un tip de rețea neuronală este Rețeaua Neuronală Pe Mai Multe Straturi (MLP). MLP are un strat de intrare, mai multe straturi ascunse și un strat de ieșire. În figura de mai jos este reprezentat un nod dintr-o rețea de tip MLP:

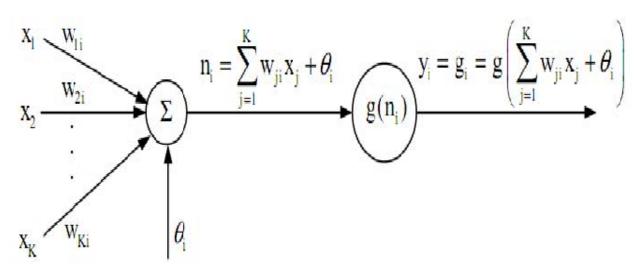
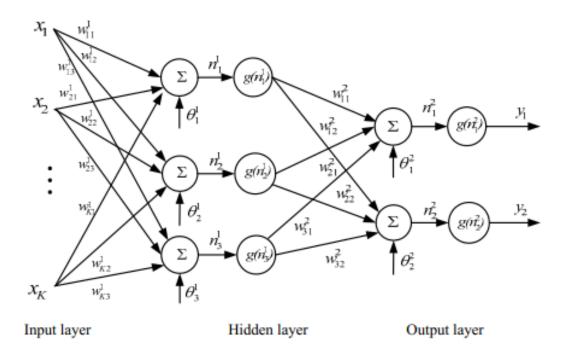


Figure 2: Single node in a MLP network

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

Dacă conectăm mai multe astfel de noduri în paralel, obținem o rețea neuronală de tip MLP astfel:



Pentru antrenarea modelului pentru mamograme vom folosi 5 straturi ascunse:

```
inputs = hasil';
targets = target';
hiddenLayerSize = 5;
net =
patternnet(hiddenLayerSize);
net.divideParam.trainRatio =
100/100;
net.trainParam.showWindow =
false;
net.trainParam.showCommandLi
ne = false;
[net,tr] =
train(net,inputs,targets);
```

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

Clasificare:

```
function
[result,c]=classification(ne
t, new_inputs, new_targets)
% Test Network
outputs =
sim(net,new_inputs);
[c,cm] =
confusion(new_targets,output
s);
if cm(1,1)==1 || cm(2,1)==1
result='Normal';
elseif cm(1,2)==1 ||
cm(2,2)==1
result='Cancer';
end
```

4.Concluzii

a. Rezultate obținute

În urma implementării s-a obținut o rețea cu o garanție satisfăcătoare pentru a detecta noduli de tip malign prezenți la nivelul mamografiilor digitalizate.

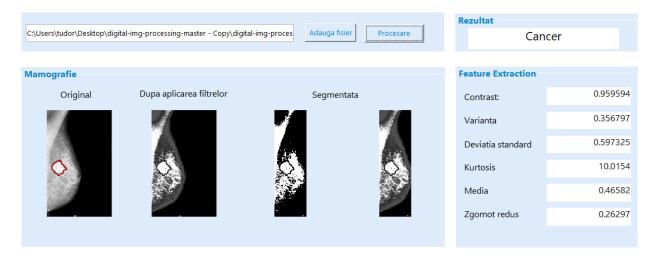
În articolul "Un studiu comparativ al algoritmilor de îmbunătățire a mamografiei digitale bazat pe teoria fuzzy" se explică despre îmbunătățirea calității imaginii datelor de mamografie pentru detectarea cancerului de sân folosind 5 (cinci) algoritmi bazați pe abordarea fuzzy. Unul dintre algoritmi utilizat în îndeplinirea acestei sarcini este algoritmul de distribuție a posibilității. În general, acest algoritm reduce nivelul de gri al pixelilor care au o valoare de gri între $\beta1$ și $\beta2$. Astfel, rezultatele obținute sunt imagini care evidențiază părțile foarte luminoase și foarte întunecate ale setului de date, care reprezintă țesutul pielii și pot fi prelucrate ulterior pentru analiza cancerului de sân.

La etapa de extragere a caracteristicilor, după încercarea mai multor caracteristici, am identificat 6 caracteristici specifice mamogramelor. Ne bazăm pe matricea de co-ocurență a nivelurilor de gri (GLCM), caracteristicile de intensitate și caracteristicile histogramei de intensitate. Din caracteristica GLCM luăm valoarea de contrast, din caracteristica de intensitate luăm media,

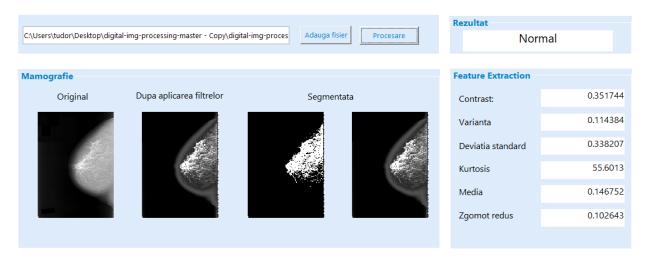
PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

varianța și deviația standard, iar din caracteristica histogramei de intensitate luăm kurtosisul și uniformitatea.

Identificare cancer de san



Identificare cancer de san



b. Direcții de dezvoltare

Direcțiile de dezvoltare pot fi multiple, această aplicație având un potențial foarte mare. Datele pe care este antrenat modelul pot avea dimensiuni mai mari, astfel crescând precizia modelului. De

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

asemenea, prin prelucrarea imaginilor se poate trasa și un contur în jurul zonei de interes pentru medicul oncolog, astfel mutând zona de interes pe o suprafață redusă.

5.Bibliografie

- -https://rasbt.github.io/mlxtend/user_guide/classifier/MultiLayerPerceptron/
- -"Breast Cancer Classification using Support Vector Machine and Neural Network", Ebrahim Edriss Ebrahim Ali, Wu Zhi Feng, School of Information Technology and Engineering, Tianjin University of Technology and Education, Dagu Nanlu Road Tianjin, China
 - -https://www.mathworks.com/
- -https://www.mathworks.com/matlabcentral/fileexchange/26532-image-segmentation-using-otsu-thresholding
 - -https://users.utcluj.ro/~simona/apim/apim8.pdf
- -"ANALIZĂ ȘI RECUNOAȘTERE DE FORME PENTRU APLICAȚII CU IMAGINI DIGITALE", RUSU MARIANA, 122.03 – MODELARE, METODE MATEMATICE, PRODUSE PROGRAM
 - -https://users.utcluj.ro/~tmarita/IPL/IPCurs/
 - -https://www.researchgate.net/
 - -https://www.semanticscholar.org/
 - -https://chat.openai.com/
 - -https://users.utcluj.ro/~rdanescu/teaching_pi.html
- -"Classification of Breast Cancer Detection Using Artificial Neural Networks", Sulochana Wadhwani, A.K.Wadhwani, Monika Saraswat, Current Research in Engineering, Science and Technology (CREST) Journals
- -"AN EFFICIENT AUTOMATIC MASS CLASSIFICATION METHOD IN DIGITIZED MAMMOGRAMS USING ARTIFICIAL NEURAL NETWORK", Mohammed J. Islam, Majid Ahmadi, Maher A. Sid-Ahmed, International Journal of Artificial Intelligence & Applications (IJAIA), Vol.1, No.3, July 2010
- -https://github.com/st186/Detection-of-Breast-Cancer-using-Neural-Networks/blob/master/README.md
 - -https://github.com/vasudev-sharma/Breast-Cancer-Detection-using-Artificial-Neural-Networks
 - -https://github.com/Simonlyamu/Breast-Cancer-Prediction

PROIECT INGINE	RIA REGLARII AUTOMATE II				
NUME student	Scoarță Tudor-Claudiu	GRUPA:	30135	Nota	

- -https://github.com/ajitkoduri/Breast-Cancer-Tumor-Prediction
- -https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data
- -http://www.eng.usf.edu/cvprg/
- -R. Nithya and B. Santhi, "Classification of Normal and Abnormal Patterns in Digital Mammograms for Diagnosis of Breast Cancer," International Journal of Computer Applications (0975 8887), Vols. 28-No.6, pp. 21-25, 2011.
- -N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," Transactions on Systems, Man, and Cybernetics, Vols. SMC-9 No.1, pp. 62-66, 1979.
- -T. C. J. Astley S M, "Combining cues for mammographic abnormalities Proc," in *1st British Machine Vision Conference Oxford*, UK, 1990.