



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

Computer aided medical diagnosis system

Autor: Goron Andreea-Lidia

Grupa: **30131**

AN UNIVERSITAR: 2021-2022

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

Cuprins

1.	Scopul Proiectului	3
a.	Obiective	3
b.	Specificații	3
2.	Studiu bibliografic	5
3.	Analiză, proiectare, implementare	7
4.	Concluzii	16
a.	Rezultate obținute	16
b.	Direcții de dezvoltare	17

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

1. Scopul Proiectului

Scopul proiectului este de a dezvolta o aplicație care detectează cancerul tiroidian, aceasta fiind bazată pe utilizarea Rețelelor Neuronale. Deoarece modul în care se realizează detecția de cancer tiroidian este costisitoare din punct de vedere material și al timpului alocat, dezvoltarea unei tehnologii bazată pe Deep Learning, orientată în direcția respectivă, este o modalitate excelentă de a ajuta sistemul medical din aceste puncte de vedere.

a. Obiective

- Prin dezvoltarea proiectului se dorește obținerea șanse de a exista cancer tiroidian într-o imagine cu ultrasunete preluată de la un pacient.
- Obiectivul principal constă în folosirea aplicației în domeniul medical, aceasta reprezentând un real ajutor pentru medicii oncologi.
- Dezvoltarea aplicației se bazează pe metode din domeniul Deep Learning, anume utilizarea Rețelelor Neuronale pentru detectarea cancerului tiroidian.

b. Specificații

Metodele tradiționale de diagnosticare a bolilor în general, se bazează în mare parte pe cunoștințele de specialitate ale medicilor cu privire la orice afecțiune dată. Cu toate acestea, această metodă de diagnosticare are o mare limitare, adică performanța sa depinde mult mai mult de experiențele și cunoștințele personale ale medicilor. Ca urmare, performanța de diagnosticare variază și este limitată. Odată cu dezvoltarea tehnologiei digitale, tehnicile de diagnosticare bazate pe imagini au fost utilizate pe scară largă pentru a ajuta medicii să investigheze problemele cu organele care se află sub piele și/sau adânc în interiorul corpului uman.

Prin dezvoltarea aplicației "Computer aided medical diagnosis system- Detection of Thyroid Cancer" s-a urmărit diagnosticarea cancerului tiroidian, utilizând rețeaua neuronală Neural Network Start, Pattern recognition and classification app. Rețeaua este capabilă să recunoască cancerul tiroidian, fiind antrenată cu imagini cu ultrasunete ce ilustrează glandele tiroide a mai multor pacienți, glande atât afectate cât și neafectate de cancer. Bazându-se în continuare pe un target binar, clasificarea se realizează în funcție de acesta: "1" denotă faptul că pacientul suferă de cancer, în timp ce "0" reflectă sănătatea tiroidei examinate. Aplicația reușește să returneze șansa aproximativă de a exista cancer tiroidian, aceasta fiind exprimată într-o valoare zecimală cuprinsă între limitele maxime, respectiv "1" și "0".

Datorită limitării de memorie a aplicației Matlab (maxim 2GB), în care s-a realizat partea de antrenare cu un set de imagini a rețelei neuronale, acuratețea aplicației este scăzută. În general, o rețea neuronală aferent antrenată și eficientă se bazează pe cel puțin 40-50 de neuroni. Aplicația descrisă în momentul de față suportă folosirea a maxim 20 de neuroni și un set de imagini de

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

antrenare în număr de 258. Prin urmare, eficiența aplicației este scăzută din motivele descrise mai sus.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

2. Studiu bibliografic

Prima lucrare de cercetare studiată pentru realizarea proiectului a fost “Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study “. Scopul principal al acestui studiu a fost de a utiliza modele de rețea neuronală convoluțională profundă (DCNN) pentru a îmbunătăți acuratețea diagnosticului cancerului tiroidian prin analiza datelor imagistice ecografice din ecografii clinice. Metoda abordată în această lucrare este dezvoltarea și antrenarea modelului DCNN pe setul de antrenament, „131 731 imagini cu ultrasunete de la 17 627 de pacienți cu cancer tiroidian și 180 668 imagini de la 25 325 controale din baza de date de imagistică tiroidiană de la Spitalul de Cancer Tianjin”. Imaginile cu glanda tiroidă care au fost considerate ca neavând cancer au fost excluse din setul de antrenament și numai persoanele cu suspectare de cancer tiroidian au fost supuse unui examen patologic pentru a confirma diagnosticul. S-a constatat că Modelul DCNN a arătat o sensibilitate similară și o „specificitate îmbunătățită în identificarea pacienților cu cancer tiroidian în comparație cu un grup de radiologi calificați”. Așadar, prima concluzie fixată din studiul de cercetare respectiv a fost importanța dimensiunii setului de antrenament necesar pentru o acuratețe ridicată a rezultatului obținut și faptul că tehnologia bazată pe rețeaua neuronală DCNN a reușit să investigheze cu succes cancerul tiroidian, fiind antrenată doar cu imagini ecografice ce conțin organul afectat de cancer.

Un studiu de caz al unui set de date din Arabia Saudită a explicat dezvoltarea instrumentelor bazate pe Deep Learning care pot servi ca sisteme ce reușesc să detecteze din timp apariția cancerului tiroidian. În plus, în acest studiu de caz s-a propus obținerea de astfel de sisteme cu cea mai mare precizie posibilă, utilizându-se mai puține funcții. Tehnicile utilizate în această lucrare includ „pădure aleatoare (RF), rețea neuronală artificială (ANN), mașină vector de suport (SVM) și Bayes naiv (NB), fiecare dintre acestea fiind selectată pentru capacitățile lor unice. Cea mai mare rată de precizie obținută a fost de 90,91% cu tehnica RF, în timp ce SVM, ANN și NB au obținut o acuratețe de 84,09%, 88,64% și, respectiv, 81,82%. Aceste niveluri au fost obținute utilizând doar șapte caracteristici din cele 15 disponibile.” În urma rezultatelor obținute, se poate afirma că tehnica RF este mai bună și, prin urmare, ”recomandată pentru această problemă specifică”.

Așadar, se pot remarca următoarele idei principale pentru dezvoltarea unei aplicații ce reușește să detecteze cancerul tiroidian, extrase din această lucrare de cercetare:

1. S-au utilizat patru modele de inteligență computațională dezvoltate pentru a preciza existența bolii cancerului tiroidian în mod preventiv.

2. Modelele de inteligență computațională au fost implementate pe un set de date din Arabia Saudită, cuprinzând peste 100 000 de resurse imagistice utilizate la antrenarea rețelelor neuronale.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

3. Modelele au obținut rezultate promițătoare cu tehnica Random Forest, obținând cel mai bun rezultat.

4. ”Cele mai bune rezultate au fost obținute utilizând setarea parametrilor optimizați ai tehnicilor cu abordarea de eliminare recursivă a caracteristicilor.”

Link-uri către articolele studiate:

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0262128>

<https://www.sciencedirect.com/topics/computer-science/machine-learning>

[https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045\(18\)30762-9/fulltext](https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(18)30762-9/fulltext)

<https://pubmed.ncbi.nlm.nih.gov/33647831/>

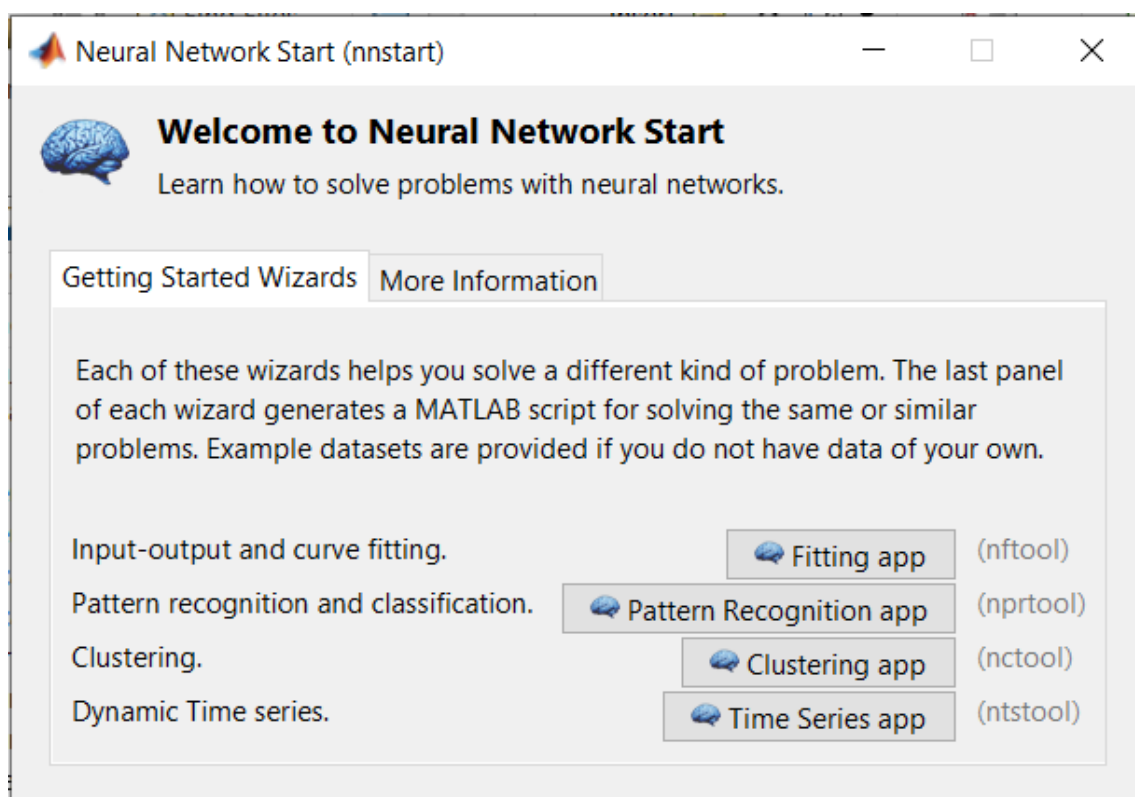
Pentru studiul rețelelor neuronale: <https://code-it.ro/introductere-in-retele-neuronale-teorie-si-aplicatii/>

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

3. Analiză, proiectare, implementare

Dezvoltarea aplicației se bazează pe ToolBox-urile oferite de Matlab, anume **Deep Learning Toolbox** pentru partea de input, target, testare și antrenare a rețelei neuronale și Design App pentru design-ul aplicației. În ceea ce privește rețeaua neuronală, s-a utilizat Neural Network Start, Pattern recognition and classification.

Așa cum s-a menționat mai sus, **Deep Learning Toolbox** pune la dispoziție utilizarea rețelei neuronale necesară pentru diagnosticarea cancerului.



După alegerea opțiunii necesare, anume Pattern recognition and classification, se vor selecta input-ul și target-ul necesar pentru antrenarea rețelei. De remarcat este faptul că rețeaua suportă ca și input doar un format de tip Array, așadar imaginile necesare antrenării vor fi incluse într-un Array de imagini pentru ca rețeaua să poată funcționa. După cum s-a specificat anterior, target-ul rețelei este întocmit pe bază de valori binare, “1” și “0”, acestea ilustrând șansa maximă și minimă ca organul analizat să fie afectat de cancer.

PROIECT INGINERIA REGLARII AUTOMATE II				
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota

Neural Pattern Recognition (nprtool)

Select Data

What inputs and targets define your pattern recognition problem?

Get Data from Workspace

Input data to present to the network.

Inputs: (none) ...

Target data defining desired network output.

Targets: (none) ...

Samples are: ☒ Matrix of x rows

Summary

No inputs selected.

No targets selected.

Want to try out this tool with an example data set?

[Load Example Data Set](#)

Select inputs and targets, then click [Next].

[Neural Network Start](#) [Welcome](#) [Back](#) [Next](#) [Cancel](#)

Neural Pattern Recognition (nprtool)

Select Data

What inputs and targets define your pattern recognition problem?

Get Data from Workspace

Input data to present to the network.

Inputs: (none) ...

Target data defining desired network output.

Targets: (none) ...

Samples are: ☒ Matrix of x rows

Summary

No inputs selected.

No targets selected.

Want to try out this tool with an example data set?

[Load Example Data Set](#)

Select inputs and targets, then click [Next].

[Neural Network Start](#) [Welcome](#) [Back](#) [Next](#) [Cancel](#)

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

Codul prin care s-a creat vectorul de input și target-ul Rețelei Neuronale este prezentat mai jos:

```

1 - clc
2 - close all
3 - clear
4 - DirectorFile = imageDatastore('C:\Users\Lidia\Desktop\Proiect IRA\ImaginiT','IncludeSubfolders',1,'LabelSource','foldernames')
5 - Images = imageDatastore('C:\Users\Lidia\Desktop\Proiect IRA\ImaginiT','IncludeSubfolders',1,'LabelSource','foldernames')

```

-Partea de input:

```

6 - %% Inputul Rețelei Neuronale
7 - images=readall(DirectorFile)
8 - for i=1:length(images) %Citim toate imaginile din folder si le vom stoca intr-un vector de imagini
9 -     image=rgb2gray(images{i}); %2d
10 -    image=image(:); %vector
11 -    Array{i}=image;
12 - end
13 -
14 - %Parte necesara pentru ca vectorul de input sa poata fi recunoscut de retea
15 - %Transformam vectorul de vectori (Array{i}) intr-un vector normal
16 - functie = cellfun(@numel,Array); %Numarul de array-uri din Array, cu cate elemente contine fiecare
17 - maxim = max(functie); %Maximul de elemente din fiecare vector
18 - nume = numel(Array); %Numarul de array-uri incluse in Array
19 - input = zeros(maxim,nume); %Initializare input cu dimensiunile respective
20 -
21 - %Creare input:
22 - for ii = 1:nume
23 -     input(1:functie(ii),ii) = Array{ii};
24 - end
25 -

```

Se poate observa că primul pas este citirea imaginilor din folder, acestea fiind salvate în variabila DirectorFile. Inițial, imaginile sunt în format 3D, fiind nevoie ca ulterior acestea să fie transformate în format 2D prin apelul funcției rgb2gray, ca în final să se poată obține un vector de imagini. Deoarece inițial Array{i} reprezintă un “vector de vectori”, pentru a putea fi transformat într-un vector de imagini standard, este nevoie de porțiunea de cod cuprinsă între liniile 16 și respectiv 23. În acest mod s-a realizat partea de input pentru Rețeaua Neuronală.

-Partea de target:

```

25 - %% Partea de Target
26 - %Se realizeaza partea de target in functie de numele folder-ului unde
27 - %sunt salvate imaginile -> cancere / noncancere
28 - Labels=Images.Labels
29 - for i=1:length(Labels)
30 -     if (Labels(i)=='cancere')
31 -         target(i)=1;
32 -     else
33 -         target(i)=0;
34 -     end
35 - end

```

Pentru partea de target este luat în considerare numele folderelor unde sunt salvate imaginile de antrenare a rețelei. Acestea sunt împărțite în 2 foldere principale, “cancere” și “noncancere” și clasificate după existența cancerului în imagini. Așadar, imaginilor din folderul “cancere” le va fi asociat un target de valoarea 1, în timp ce celorlalte imagini li se vor asocia un target de valoare 0.

PROIECT INGINERIA REGLARII AUTOMATE II				
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota

În cele ce urmează, după finalizarea părții anterioare și alegerea numărului de neuroni (în cazul nostru se utilizează 20 de neuroni), se poate antrena rețeaua neuronală.

Neural Pattern Recognition (nprtool)

Train Network

Train the network to classify the inputs according to the targets.

Train Network
Train using scaled conjugate gradient backpropagation. (trainscg)

Results

	Samples	CE	%E
Training:	206	5.71033e-1	26.21359e-0
Validation:	39	9.02699e-1	35.89743e-0
Testing:	13	8.95657e-1	30.76923e-0

Training automatically stops when generalization stops improving, as indicated by an increase in the cross-entropy error of the validation samples.

Notes

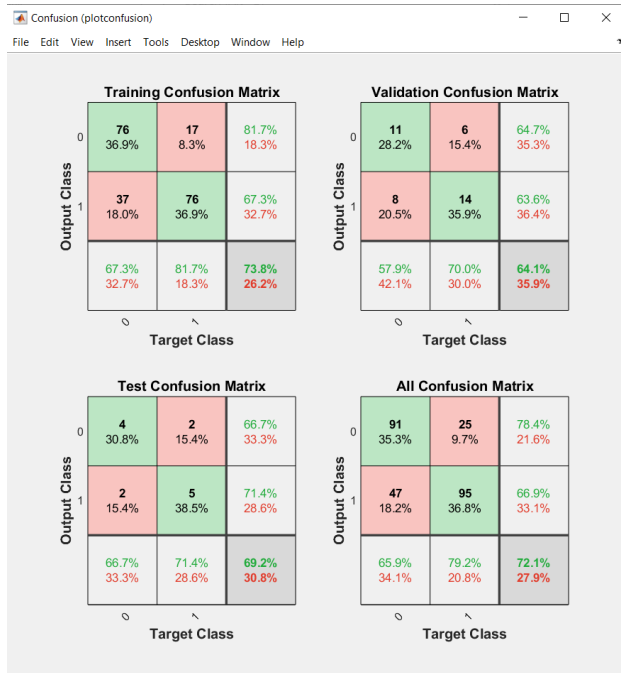
- Training multiple times will generate different results due to different initial conditions and sampling.
- Minimizing Cross-Entropy results in good classification. Lower values are better. Zero means no error.
- Percent Error indicates the fraction of samples which are misclassified. A value of 0 means no misclassifications, 100 indicates maximum misclassifications.

Open a plot, retrain, or click [Next] to continue.

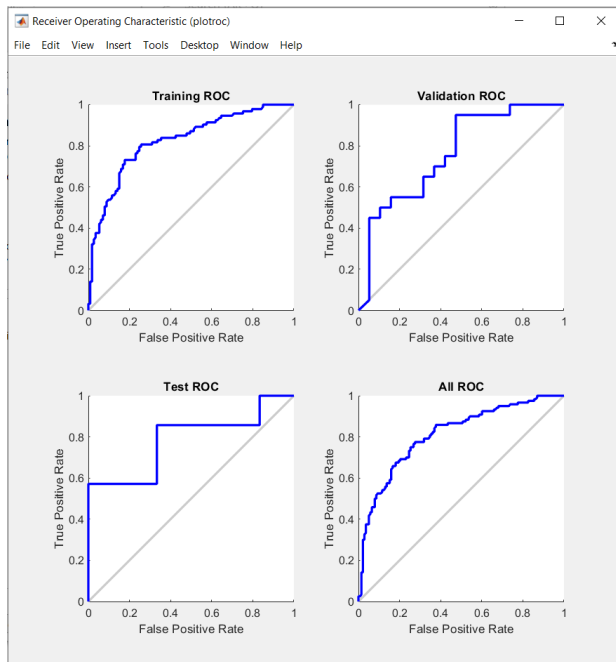
Neural Network Start Welcome Back Next Cancel

PROIECT INGINERIA REGLARII AUTOMATE II				
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota

-Se pot ilustra grafic Matricile de Confuzie, acestea reprezentând rezumatul rezultatelor predicțiilor unei probleme de clasificare.



-De asemenea, prin graficele ROC se pot observa grafic performanțele unui model de clasificare. Datorită resurselor limitate specificate anterior, s-au obținut următoarele grafice ROC pentru un set de antrenare de 258 de imagini cu ultrasunete și 20 de neuroni utilizați:



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

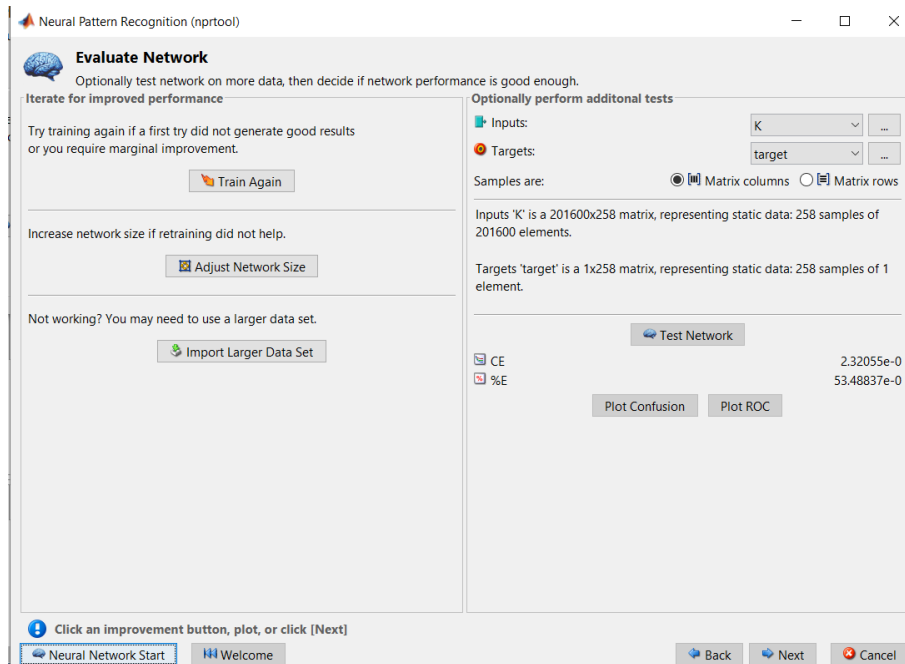
O parte importantă după antrenarea rețelei neuronale este aceea de testare. Prin intermediul testării, am ajuns la concluziile referitoare la acuratețea aplicației create. Așadar, după ce rețeaua a fost configurată funcțional, am creat un nou folder cu imagini care nu sunt incluse în cele pentru antrenarea rețelei și am realizat testarea cu acestea.

```

39 %% Testare
40 - img=imageDatastore('C:\Users\Lidia\Desktop\Proiect IRA\test');
41 - images1=readall(img)
42 -
43 - for i=1:length(images1)
44 -     image=rgb2gray(images1{i}); %2d
45 -     image=image(:); %vector
46 -     array{i}=image;
47 -
48 - % % %figure,imshow(image);
49 - end
50 - s = cellfun(@numel,array); %
51 - m = max(s);
52 - n = numel(array);
53 - tests = zeros(m,n);
54 -
55 - for ii = 1:n
56 -     tests(1:s(ii),ii) = array{ii};
57 - end
58 - K = imresize(tests,[201600 258]);
59 - K=im2double(K);
60 - nnstart
61 -

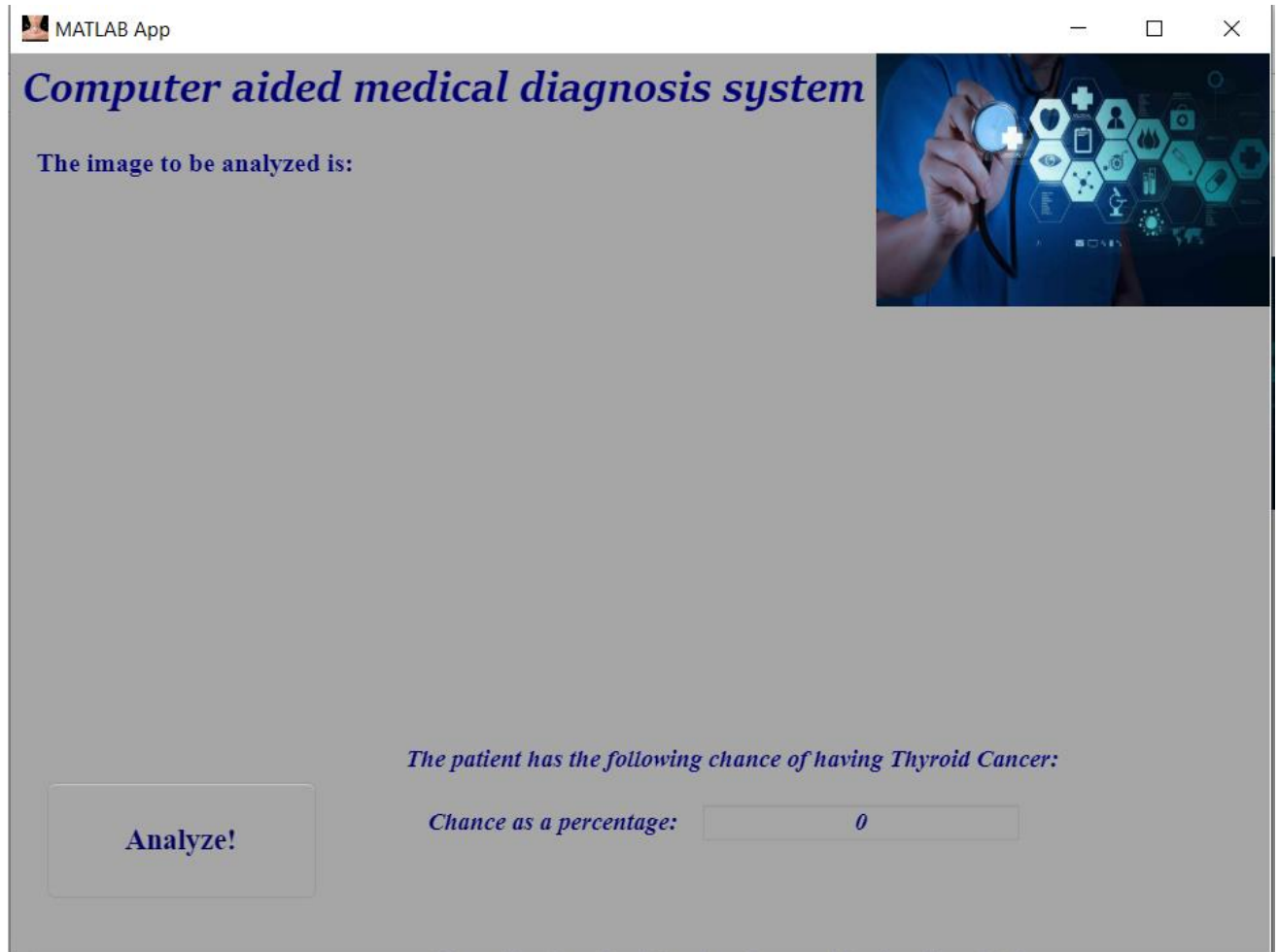
```

Se va observa în cele ce urmează că se obține o eroare mare, anume aproximativ 53.4%. Aceasta este datorată insuficienței de poze pentru antrenare și de absența numărului de neuroni necesari pentru o acuratețe ridicată.



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

În ceea ce privește design-ul aplicației, prin intermediul toolbox-ului Design App, s-a realizat următoarea interfață grafică:



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

Cu ajutorul butonului “Analyze!”, se va încărca o anumită imagine conținută în folderul “test”, urmând mai apoi să se returneze șansa exprimată în procente în partea dreaptă jos, prin apelul funcției corespunzătoare rețelei neuronale antrenate. Funcția respectivă a fost generată automat după realizarea pașilor necesari antrenării rețelei neuronale. Codul pentru implementarea butonului, cu comentariile ce explică funcționalitatea, este prezentat mai jos:

```

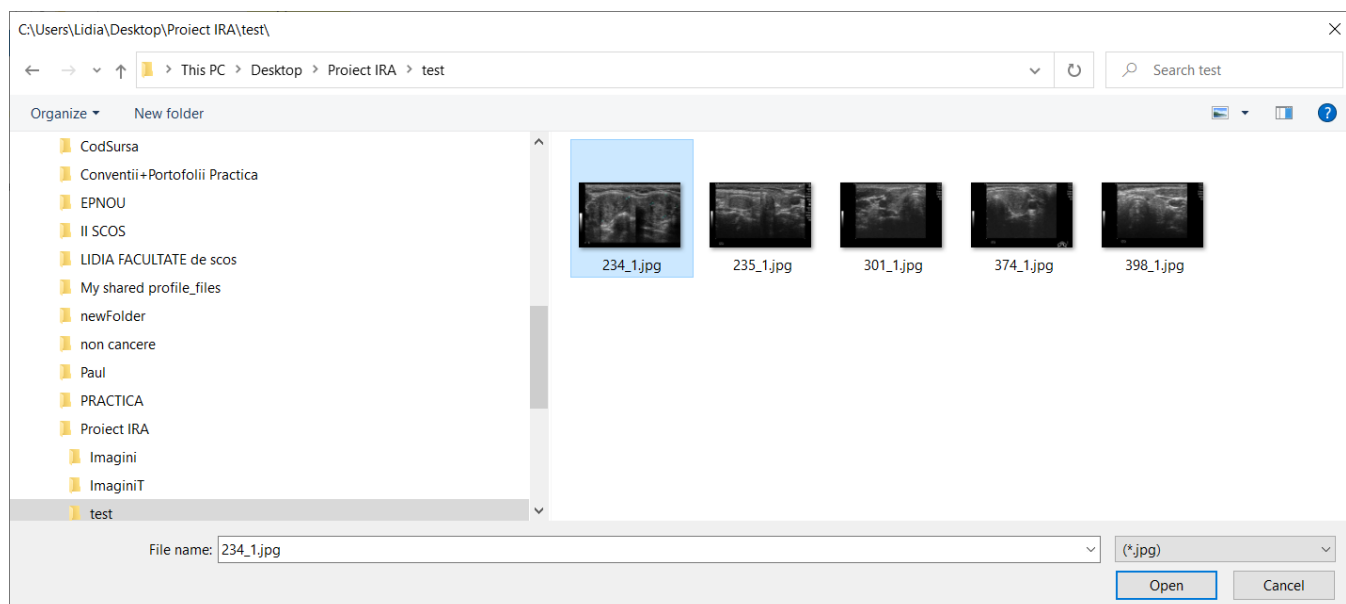
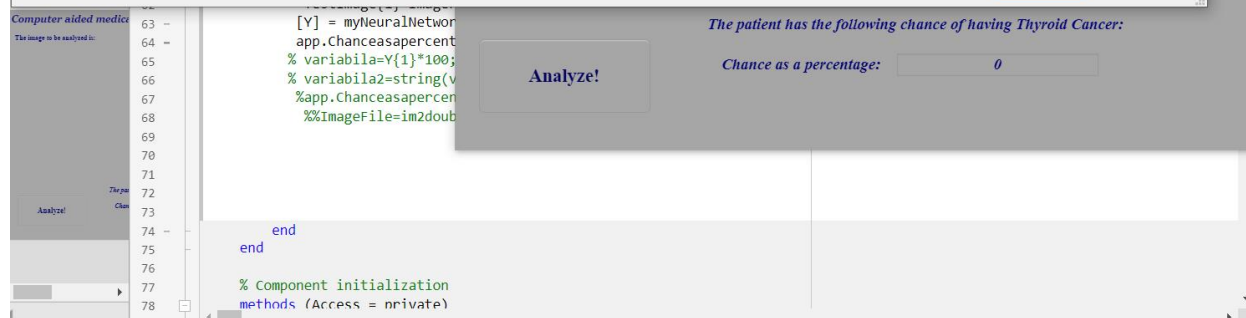
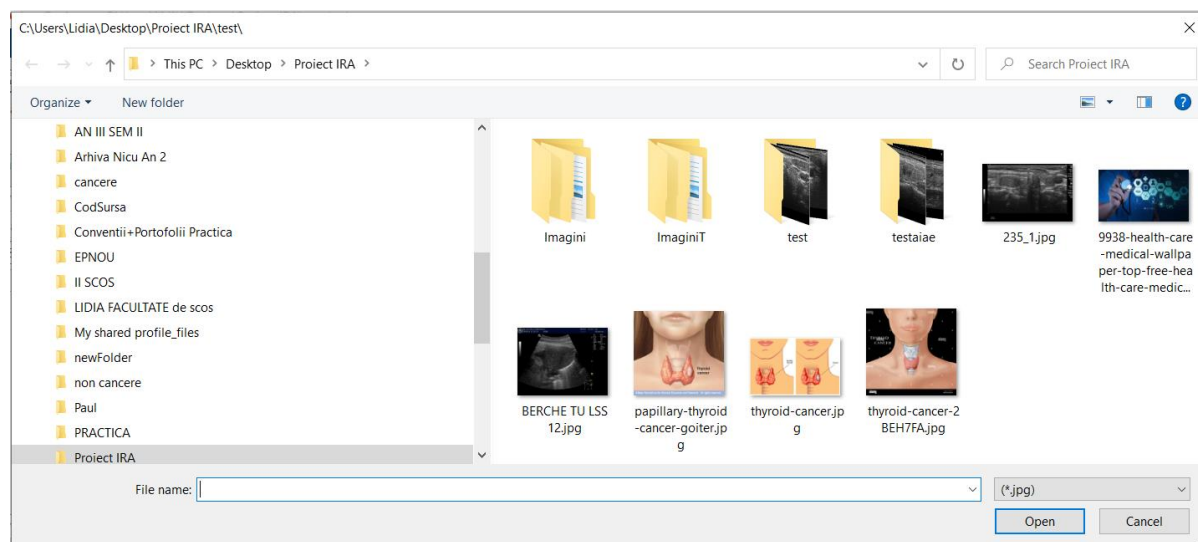
50 % Button pushed function: AnalyzeButton
51 function AnalyzeButtonPushed(app, event)
52     % Incarcare imagine si afisare
53     [filename,filepath]=uigetfile('*.jpg','C:\Users\Lidia\Desktop\Proiect IRA\test\'); %Deschidere folder test
54     fullname=[filepath,filename];
55     ImageFile=imread(fullname); %Citire imagine
56     imagesc(app.UIAxes,ImageFile); %Afisare imagine in aplicatie
57     %Testare retea cu imaginea selectata
58
59     ImageFile=rgb2gray(ImageFile); %Transformare in matrice 2d
60     ImageFile=ImageFile(:); %Transformare in vector
61     ImageFile=im2double(ImageFile); %Reconfigurare tip (in double)
62     ImageFile=ImageFile/255; %Deoarece avem numere cuprinse intre 0 si 255 in array, vom imparti cu 255 toate aceste elemente pentru a lucra cu numere mai mici
63     TestImage{1}=ImageFile; %Punem imaginea intr-un cellarray pentru a putea fi data ca parametru functiei
64     [Y] = myNeuralNetworkFunction(TestImage{1}); %Apelare functie
65     app.ChanceasapercentageEditField.Value=Y; %Afisare rezultat
66     % variabila=Y{1}*100;
67     % variabila2=string(variabila);
68     %app.ChanceasapercentageEditField.Value=strcat(variabila2,'%');
69     %%ImageFile=im2double(ImageFile);
70
71
72
73
74
75 end
76 end
77

```

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

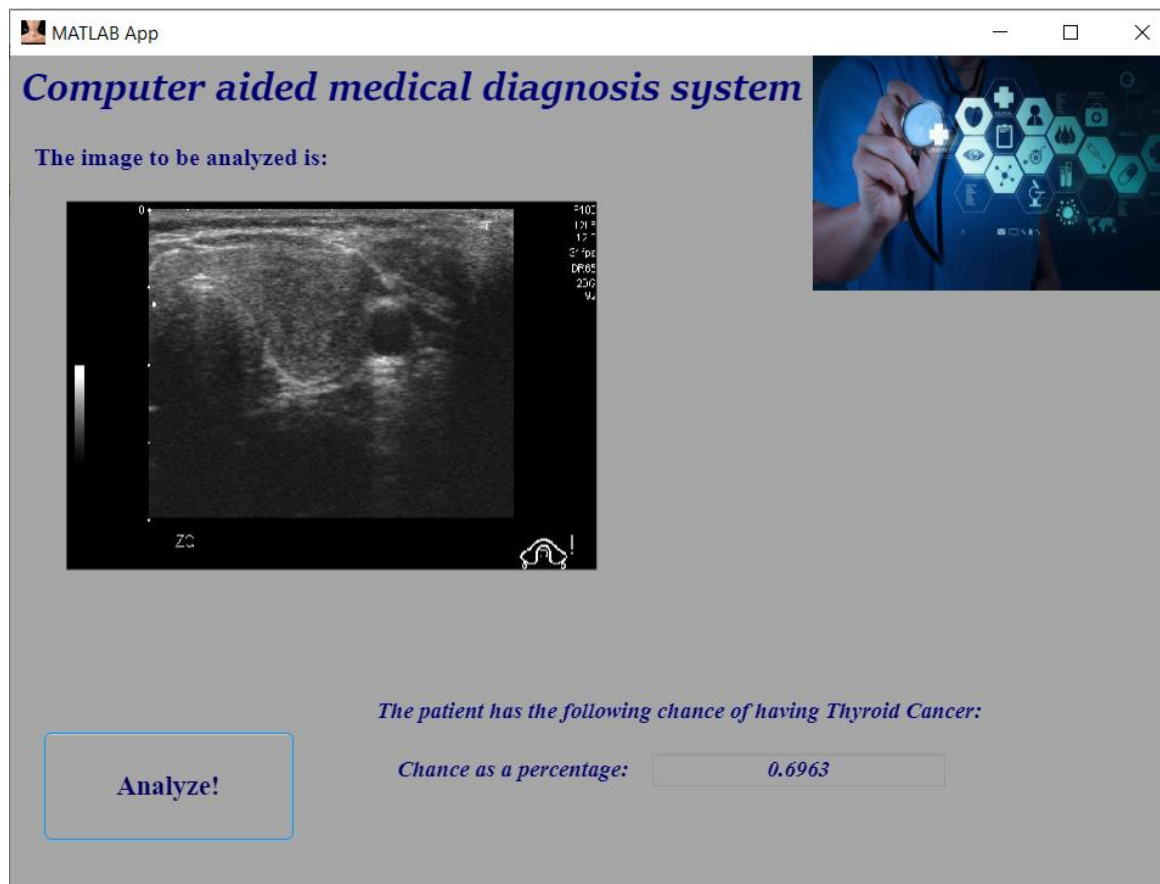
În final, este prezentată funcționalitatea aplicației:

-Alegerea imaginii cu ultrasunete din folderul “test” ce urmează să fie analizată:



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

-Rezultatul obținut în urma analizei:



4. Concluzii

a. Rezultate obținute

În concluzie, prin realizarea proiectului s-a reușit dezvoltarea unei aplicații a cărei funcționalitate se fundamentează pe folosirea rețelei neuronale Neural Network Start, Pattern recognition and classification. Aplicația reușește să genereze o șansă aproximativă exprimată în procente, cuprinsă între 0 și 1, aceasta reflectând cât la sută organul din imaginea cu ultrasunete analizată este afectat de cancer.

Datorită resurselor limitate de memorie, aplicația descrisă în acest proiect nu reușește să genereze cu exactitate șansa respectivă. Așadar, nefiind antrenată cu suficiente poze și neputând atribui rețelei un număr de minim 40 de neuroni, șansele tuturor imaginilor analizate vor oscila în jurul valorii corecte, însă nu se reușește

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Goron Andreea-Lidia	GRUPA:	30131	Nota	

atingerea acesteia. Se poate afirma așadar că împărțirea sau clasificarea imaginilor în “cancer”, respectiv “noncancer” este una lipsită de o acuratețe ridicată. În studiile de cercetare pe aceeași temă medicală, dezvoltate în trecut, s-a reușit, pe baza tehnologiei, dezvoltarea de aplicații care să reușească detectarea cancerului cu o precizie ridicată. S-au folosit pentru acestea alte tipuri de rețele neuronale, de exemplu, modelul DCNN, ResNet50, rețele care s-au antrenat cu peste 100 000 de imagini cu ultrasunete și s-a obținut o acuratețe ridicată, în jurul valorii de 90,4%.

Prin urmare, se poate exprima importanța extraordinară a setului de imagini de antrenare și se poate concluziona faptul că odată cu creșterea numărului de imagini se va obține garantat o acuratețe mai ridicată a rețelei neuronale antrenate corespunzător.

Cu toate acestea, prin dezvoltarea proiectului, s-au atins următoarele contribuții:

- Antrenarea Rețelei Neuronale prin intermediul unui toolbox. Prin această metodă, se reduce semnificativ timpul alocat creării rețelei neuronale de la zero, prin cod.

- S-a remarcat sensibilitatea funcționalității corecte a unei rețele neuronale, observându-se prin antrenare performanțele acesteia. S-a ajuns la concluzia că atât prin creșterea numărului de imagini cu ultrasunete pentru antrenare, cât și prin creșterea numărului de neuroni, rețeaua reușește să obțină o acuratețe mult mai ridicată.

- Aplicația dezvoltată poate fi utilizată pentru recunoașterea oricărui tip de cancer, nu doar cel tiroidian. Așadar, dezvoltată pe un sistem cu resurse mult mai eficiente din punctul de vedere al memoriei, se poate persista în direcția dezvoltării aplicației pentru o acuratețe mult mai ridicată a rezultatelor obținute.

b. Direcții de dezvoltare

Aplicația rezultată poate fi utilizată în domeniul medical, iar cu ajustările necesare pentru o acuratețe mult mai ridicată, aceasta reprezintă un real ajutor pentru sistemul medical, reducând considerabil timpul alocat detectării de cancer prin mijloacele clasice, anume experiența medicilor.