



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Computer aided medical diagnosis systems – P13

Identify chance of thyroid cancer for patients

Autor: Marinescu Șerban Nicolae

Grupa: 30131

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Cuprins

1. Scopul Proiectului	3
a. Obiective	3
b. Specificații	3
2. Studiu bibliografic	4
3. Analiză, proiectare, implementare	6
4. Concluzii	17
a. Rezultate obținute	17
b. Direcții de dezvoltare	17

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

1. Scopul Proiectului

Scopul proiectului este de a crea o aplicatie care detecteaza sansa cancerului tiroidian ale diferitelor pacienti pe baza imaginilor cu ultrasunete ale glandelor tiroide. Aplicatiea este bazata pe retele neuronale si se urmareste folosirea ei pentru a detecta cat se poate de repede acest tip de cancer, astfel existand o sansa ridicata pentru a se putea trata.

a. Obiective

Obiectivele acestui proiect ar fi crearea unei retele neuronale si antrenarea acesteia cu ecografii de tiroida pentru a putea face o delimitare intre cele cu cancer si cele fara.

Odata ce reseaua a invatat sa faca clasificarea va trebui sa fie testata mai departe cu alte imagini cu ultrasunete in care va trebui sa identifice daca pacientul corespunzator ecografiei sufera de cancer tiroidian sau nu.

b. Specificații

Ca specificatii urmarim sa obtinem o retea neuronala cat se poate de precisa pentru a reusi sa identifice cancerul tiroidian cu o acuratete ridicata. Avem nevoie de un set de date pentru antrenare, acesta fiind compus din ecografii de tiroida atat cu cancer cat si fara pentru a putea ajuta reseaua sa faca diferente intre acestea. De asemenea vom folosi si un set de date pentru test dupa crearea retelei, sa vedem daca performantele au fost atinse.

Pentru performante ridicate este nevoie de utilizarea unei retele cu un numar ridicat de neuroni si folosirea unui set de date cu foarte multe imagini. Acest lucru constituie o problema insa, din cauza resurselor limitate ale fiecarui computer si a mediului de programare utilizat.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

2. Studiu bibliografic

Unul dintre articolele științifice studiate este intitulat: „Application of deep learning to the diagnosis of cervical lymph node metastasis from thyroid cancer with CT”, redactat de Jeong Hoon Lee, Eun Ju Ha, Ju Han Kim, afiliați la Division of Biomedical Informatics, Seoul National University Biomedical Informatics (SNUBI), Seoul National University College of Medicine, Seoul, 110799, Republic of Korea, Department of Radiology, Ajou University School of Medicine, Wonchon-Dong, Yeongtong-Gu, Suwon, 443-380, South Korea.

Scopul l-a reprezentat dezvoltarea unui sistem de diagnosticare asistată de computer (CAD) bazat pe învățarea profundă pentru utilizare în diagnosticul CT al metastazei ganglionilor limfatici cervicali (LNM) la pacienții cu cancer tiroidian.

Metode utilizate constau într- un total de 995 de imagini CT axiale care au inclus ganglioni limfatici benini (n = 647) și maligni (n = 348) au fost colectate de la 202 pacienți cu cancer tiroidian cărora li s-au efectuat CT pentru planificare chirurgicală între iulie 2017 și ianuarie 2018. Seturile de date au fost împărțite aleatoriu. în formare (79,0%), validare (10,5%) și testare (10,5%) seturi de date. Opt modele de rețea neuronală convoluțională profundă (CNN) au fost folosite pentru a clasifica imaginile în ganglioni limfatici metastatici sau benini. Pe ImageNet au fost utilizate rețele pregătite și a fost selectat algoritmul cu cele mai bune performanțe. Regiunile discriminatorii specifice clasei au fost vizualizate cu hărțile de atenție utilizând o metodă de punere în comun a mediei globale.

Rezultate

Aria de sub curba ROC (AUROC) pentru algoritmi testati a variat între 0,909 și 0,953. Sensibilitatea, specificitatea și acuratețea algoritmului cu cele mai bune performanțe au fost toate de 90,4%, respectiv. Atenție, harta termică a evidențiat subregiuni importante pentru o revizuire clinică ulterioară.

Concluzie

Un sistem CAD bazat pe învățarea profundă ar putea clasifica cu precizie LNM cervical la pacienții cu cancer tiroidian pe CT preoperator cu un AUROC de 0,953. Dacă această abordare are utilitate clinică va necesita o evaluare într-un cadru clinic.

Link catre articol: <https://pubmed.ncbi.nlm.nih.gov/30877461/>

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Un alt articol studiat este intitulat: „Multi-channel convolutional neural network architectures for thyroid cancer detection”, scris de Xinyu Zhang, Vincent C S Lee, Jia Rong, Feng Liu, Haoyu Kong, afiliați la Department of Data Science and AI/Faculty of IT, Monash University, Melbourne, Victoria, Australia, West China Hospital of Sichuan University, Chengdu City, Sichuan Province, China, Department of Human-Centred Computing/Faculty of IT, Monash University, Melbourne, Victoria, Australia.

Detectarea precoce a nodulilor tiroidieni maligni care conduc la tratamente specifice pacientului poate reduce ratele de morbiditate și mortalitate. În prezent, specialiștii tiroidieni folosesc imagini medicale pentru a diagnostica, apoi urmează protocoalele de tratament, care au limitări din cauza ratelor de diagnostic fals pozitive umane nesigure.

Odată cu apariția învățării profunde, progresele în tehnicile de diagnosticare asistată de computer au produs o acuratețe promițătoare de detectare și predicție mai devreme; cu toate acestea, adoptarea clinicienilor lipsește cu mult.

Prezentul studiu adoptă rețeaua neuronală Xception ca structură de bază și proiectează un cadru practic, care cuprinde trei arhitecturi multicanal adaptabile care au fost evaluate pozitiv folosind seturi de date din lumea reală. Arhitecturile propuse depășesc tehnicile statistice și de învățare automată existente și au atins o rată de acuratețe a diagnosticului de 0,989 cu imagini cu ultrasunete și 0,975 cu scanări de tomografie computerizată prin arhitectura cu intrare unică cu două canale.

Mai mult, designul specific pacientului a fost implementat pentru detectarea cancerului tiroidian și a obținut o precizie de 0,95 pentru arhitectura cu două canale de intrare dublă și 0,94 pentru arhitectura cu patru canale. Evaluarea noastră sugerează că imaginile cu ultrasunete și tomografia computerizată (CT) oferă rezultate de diagnostic comparabile prin aplicații de diagnosticare asistate de computer.

Cu imaginile cu ultrasunete obținute rezultate puțin mai mari, CT, pe de altă parte, poate realiza designul diagnostic specific pacientului. În plus, cu cadrul propus, clinicienii pot selecta cea mai potrivită arhitectură atunci când iau decizii cu privire la un diagnostic de cancer tiroidian.

Cadrul propus încorporează, de asemenea, rezultate interpretabile ca dovezi, ceea ce poate îmbunătăți încrederea clinicienilor și, prin urmare, adoptarea acestora a tehnicilor de diagnostic asistate de computer propuse cu eficiență și acuratețe sporite.

Link catre articol: <https://pubmed.ncbi.nlm.nih.gov/35061759/>

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

3. Analiză, proiectare, implementare

Ca si problema generala ar fi antrenarea rețelei neuronale cu ecografiile de tiroida, atat cu cancere cat si fara pentru "a invata" rețeaua sa faca deosebirea intre acestea, dar si sa identifice in viitoare imagini aspecte asemanatoare cu cele de la antrenare astfel putand detecta existenta cancerului sau sansa ca un pacient sa aiba sau nu.

Ca si implemenate am utilizat mediul MATLAB, si cu ajutorul unui toolbox numit „Deep Learning Toolbox”, am reusit sa creez o rețea neuronală care ne va ajuta in identificarea sansei de cancer tiroidian.

Primul pas ar fi crearea unui input pentru rețea. In cazul nostru avem nevoie de mai multe imagini pentru antrenarea rețelei, si cel mai eficient mod ar fi crearea unui vector de imagini.

```

1 - clc
2 - close all
3 - clear
4 - %Specificam folderele in care se afla pozele pentru antrenarea rețelei
5 - imds1 = imageDatastore('C:\Users\Y510\Desktop\Facultate\AN 3\Semestrul 2\IRA2\Proiect\ImaginiT','IncludeSubfolders',1,'LabelSource','foldernames');
6 - %Citim toate imaginile din foldere si cream un cell array numit images
7 - %in care le vom stoca
8 - images=readall(imds1);
9 - %Parcurgem fiecare poza si o formatam
10 - for i=1:length(images)
11 -     image=rgb2gray(images{i}); %2d
12 -     image=image(:); %vector
13 -     Array(i)=image;
14 - end

```

Acest vector a fost creat in codul de mai sus. Cu ajutorul functiei „imageDatastore” am specificat calea catre folderele cu imagini. Numele acestor foldere am constituit etichete pe care le vom folosi mai tarziu, si anume „cancere” si „non cancere”. Insa pentru a utiliza acest vector de poze in rețea, mai intai trebuie sa reformatam imaginile. Imaginile in MATLAB sunt matrici de numere. La inceput am utilizat functia „rgb2gray” pentru a schimba formatul uint8 cu trei dimensiuni intr-o matrice normala 2D. Mai departe am transformat matricile in vectori normali si am salvat acesti vectori in variabila „Array” care reprezinta practic un vector de vectori, fiind de tip cell array.

```

15
16 %Variabila "Array" creata anterior reprezinta un cell array in care este
17 %salvata fiecare poza ca vector
18 %Mai departe trebuie reformatat acest cell array intr-un vector normal
19 %pentru a putea fi citit de rețea
20 - s = cellfun(@numel,Array);
21 - m = max(s);
22 - n = numel(Array);
23 - input = zeros(m,n);
24
25 %In final in variabila input avem salvate toate pozele (fiecare line din matrice
26 %reprezinta o poza)
27 - for i = 1:n
28 -     input(1:s(i),i) = Array{i};
29 - end

```

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Vectorul de tip cell este foarte usor de folosit pentru a stoca orice tip de variabile, insa nu este un format acceptat de toolboxul nostru pentru retea. Asadar trebuie transformat acest cell array intr-un vector normal. Am realizat acest lucru cu ajutorul functiei „cellfun” si „numel”. Aceste functii retuneaza numarul de elemente si de cell-uri din vector, si folosind un for, in final am creat o matrice „input” in care fiecare linie reprezinta vectorul de numere al unei imagini.

Acum ca am creat un input, vom avea nevoie de un target pentru ca reseaua sa asocieze imaginile incarcate la antrenare cu ce specificam noi. Am optat pentru target sa utilizez un vector binar de 0 si 1, punand in acest vector „unu” pentru imaginile cu cancer tiroidian si „zero” pentru imaginile fara. Acest lucru a putut fi implementat foarte usor cu acele etichete create de functia „ImageDatastore”. Am creat doua foldere pentru ecografi, cu numele „cancere” respectiv „non cancere”, iar functia a realizat etichetele fix cu aceste nume. Si in final cu ajutorul unui for am terminat si vectorul de target pentru retea.

2x2 [table](#)

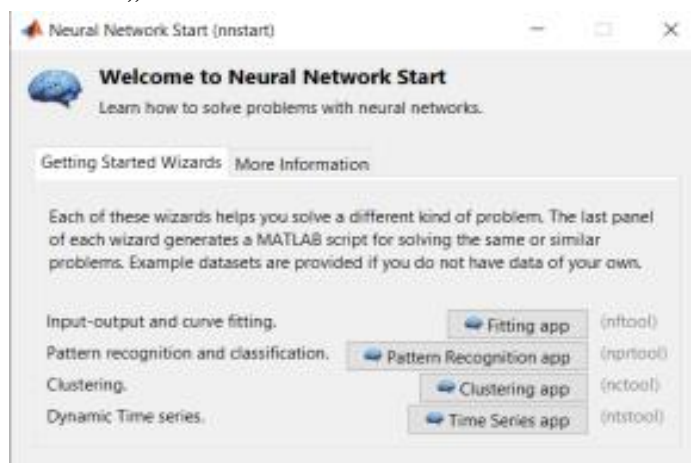
Label	Count
cancere	61
non cancere	60

```

31 %Partea de target
32 imds = imageDatastore('C:\Users\Y510\Desktop\Facultate\AN 3\Semestrul 2\IRA2\Proiect\ImaginiT', 'IncludeSubfolders',1, 'LabelSource', 'foldernames');
33 var=countEachLabel(imds)
34 labels=imds.Labels;
35 for i=1:length(labels)
36     if (labels(i)=='cancere')
37         target(i)=1;
38     else
39         target(i)=0;
40     end
41 end

```

In momentul de fata avem cei doi vectori, input si target cu care trebuie sa antrenam reseaua. Pasul urmator ar fi utilizarea toolboxului „Deep Learning” pentru a incepe crearea retelei. Se utilizeaza functia „nnstart”.



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Din acest meniu se alege opțiunea „Pattern Recognition app”, după care vom introduce variabilele noastre „input” și „target” în câmpurile corespunzătoare.

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: input ...
Target data defining desired network output.
Targets: target ...
Samples are: ☒ Matrix columns ☐ Matrix rows

Want to try out this tool with an example data set?
Load Example Data Set

Summary
Inputs 'input' is a 201600x258 matrix, representing static data: 258 samples of 201600 elements.
Targets 'target' is a 1x258 matrix, representing static data: 258 samples of 1 element.

To continue, click [Next].
Neural Network Start Welcome Back Next Cancel

Mai departe am ales opțiunile pentru Training, Validation și Test, optând pentru Test cu cele mai puține imagini pentru a lăsa mai multe pe Training și pe Validation. Rețeaua o vom testa noi manual în cele ce urmează.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Neural Pattern Recognition (nprtool)

Validation and Test Data

Set aside some samples for validation and testing.

Select Percentages

Randomly divide up the 258 samples:

Training:	80%	206 samples
Validation:	15% ▾	39 samples
Testing:	5% ▾	13 samples

Restore Defaults

Explanation

Three Kinds of Samples:

- Training:** These are presented to the network during training, and the network is adjusted according to its error.
- Validation:** These are used to measure network generalization, and to halt training when generalization stops improving.
- Testing:** These have no effect on training and so provide an independent measure of network performance during and after training.

Change percentages if desired, then click [Next] to continue.

Neural Network Start Welcome Back Next Cancel

Urmeaza selectia numarului de neuroni pentru retea. Eu am ales 20, insa pentru o acuratete mai buna si si performante dorite ar trebui alesi minim 50. Datorita resurselor limitate ale programului MATLAB nu am avut posibilitatea cresterii numarului de neuroni mai mult decat 20 ceea ce va duce din pacate la rezultate mai slabe.

Neural Pattern Recognition (nprtool)

Network Architecture

Set the number of neurons in the pattern recognition network's hidden layer.

Hidden Layer

Define a pattern recognition neural network. (patternnet)

Number of Hidden Neurons: 20

Restore Defaults

Recommendation

Return to this panel and change the number of neurons if the network does not perform well after training.

Neural Network

Change settings if desired, then click [Next] to continue.

Neural Network Start Welcome Back Next Cancel


PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Neural Pattern Recognition (nprtool)

Train Network




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

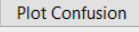
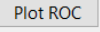
Train Network
Train using scaled conjugate gradient backpropagation. (trainscg)

 Retrain

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

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

 Plot Confusion  Plot ROC

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.

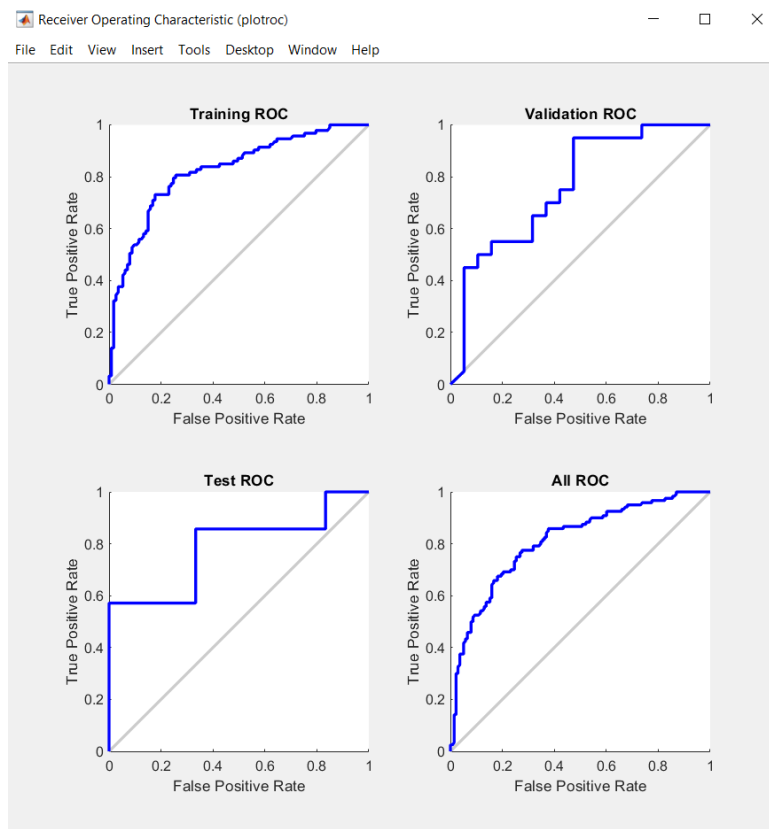
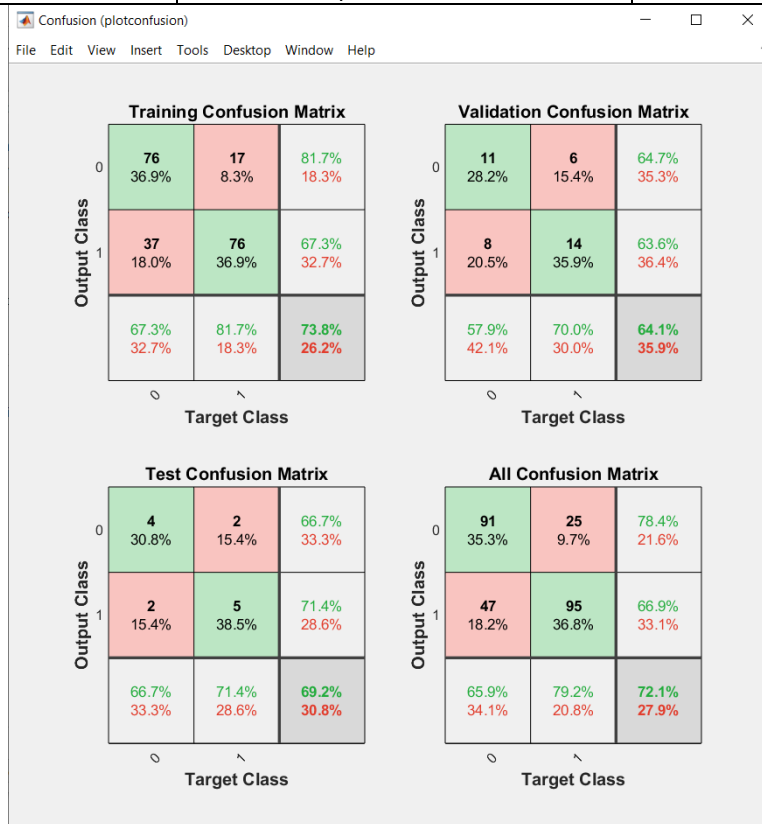
Open a plot, retrain, or click [Next] to continue.

Neural Network Start Welcome Back Next Cancel

Dupa antrenarea retelei putem vedea ploturile corespunzatoare pentru Confusion si ROC. Matricea de Confuzie reprezinta rezumatul rezultatelor predictiilor intr-o problema de clasificare. Curba ROC reprezinta un grafic care ne arata performantele unui model de clasificare la toate pragurile de clasificare

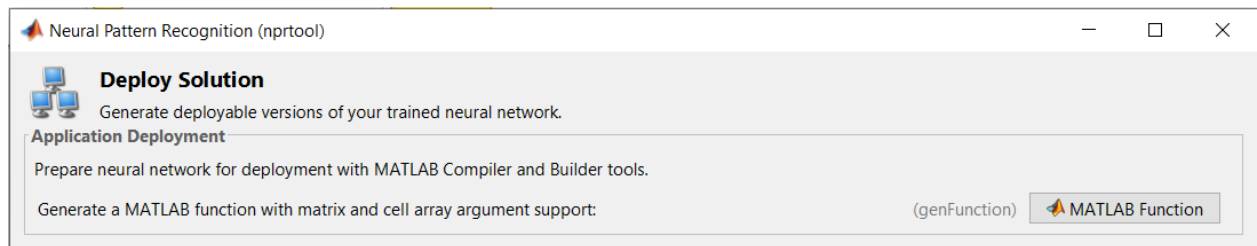
In partea de Confusion putem vedea procentele pentru cazurile cand retea credea ca o imagine are cancer si chiar a avut, sau cand credea ca are si nu a avut, etc. Aceasta matrice ne arata toate aceste combinatii de cazuri, atat pe Training cat si pe Validation sau test.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Graficul ROC ne arata performantele modelului de clasificare. Observam ca in toate cazurile curbele urmaresc oarecum referinta cu gri. In cazul testului, datorita faptului ca am folosit putine imagini apar mai multe drepte in loc de o curba.



In final utilizand aceasta functie a toolboxului, MATLAB ne va genera o functie a Retelei Neuronale cu care vom putea testa ecografiile viitoare pentru sansa de cancer tiroidian.

Iar pentru aceasta parte de test am optat pentru un alt toolbox oferit de MATLAB, si anume „App Designer” pentru a putea crea o interfata grafica in care vom testa imaginile de la ecograf. Aceasta interfata va ajuta pe orice doreste sa foloseasca aceasta aplicatie, eliminand problemele de utilizare a functiei create anterior deoarece pune la dispozitie butoane si indicatoare usor de urmat.

Ca si parte de cod, avem doar un singur buton care va face totul pentru noi. Cu ajutorul functiei „uigetfile” ni se va deschide un explorer pentru a selecta imaginea dorita pentru test. Aceasta poza urmeaza sa fie afisata in interfata, dupa care va fi modificata pentru a putea fi folosita de functie, precum am aranjat si imaginile la partea de input.

```
% Deschidere explorer pt a selecta imaginea
[filename,filepath]=uigetfile('*.jpg');
fullname=[filepath,filename];
Image=imread(fullname);

% Afisare imagine
imagesc(app.UIAxes,Image);

% Testare
% Formatare imagine pentru a putea utiliza retea
Image=rgb2gray(Image);
Image=Image(:);
Image=im2double(Image);
Image=Image/255;
X{1}=Image;

% Utilizare functie si afisare procentaj
[Y,~,~] = myNeuralNetworkFunction(X);
[Y,~,~] = NeuralNetworkcu20Neuroni(X);
```

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

Outputul funcției este variabila „Y”, care se creează după folosirea funcției Rețelei Neuronale. Această variabilă este transmisă mai departe unui Edit Field pentru a putea fi văzută de utilizator. Variabila „Y” corespunde unui număr care reprezintă șansa ca pacientul a cărei ecografie a fost testată, să aibă cancer tiroidian.

```

a=Y{1}*100;
b='%';
c=string(a);
app.PercentageforpatienttohavethyroidcancerEditField.Value=strcat(c,b);

app.UltrasoundImageNameEditField.Value=filename;

for i=1:length(app.UltrasoundImageNameEditField.Value)
    A{i}=app.UltrasoundImageNameEditField.Value(i);
    if app.UltrasoundImageNameEditField.Value(i+1)=='_'
        break;
    end
end

B=length(A);
for i=1:B
    valoare(i)=A{i};

end
app.PatientNumberinDataBaseEditField.Value=valoare;

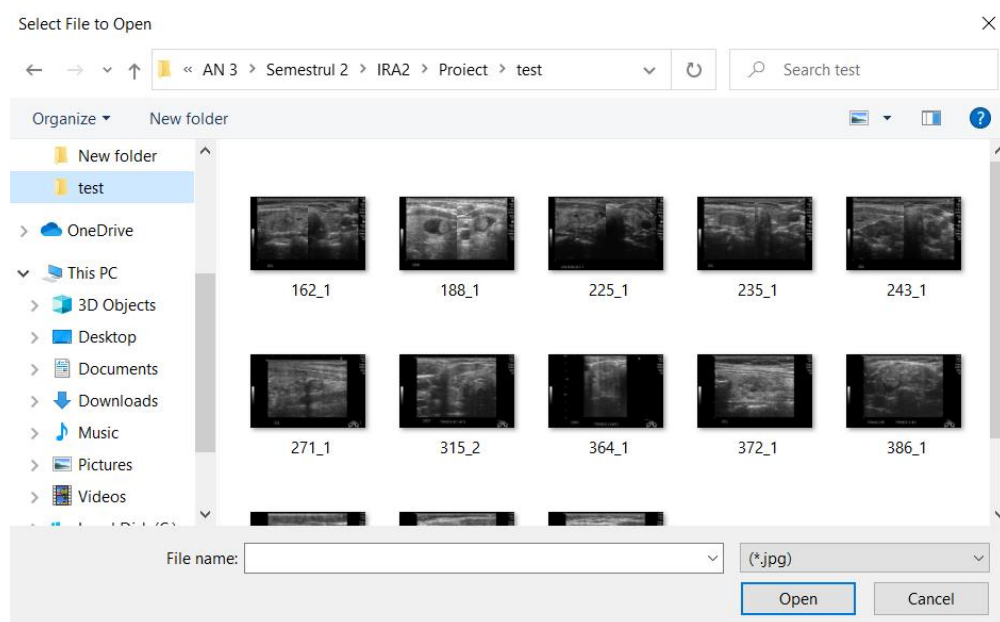
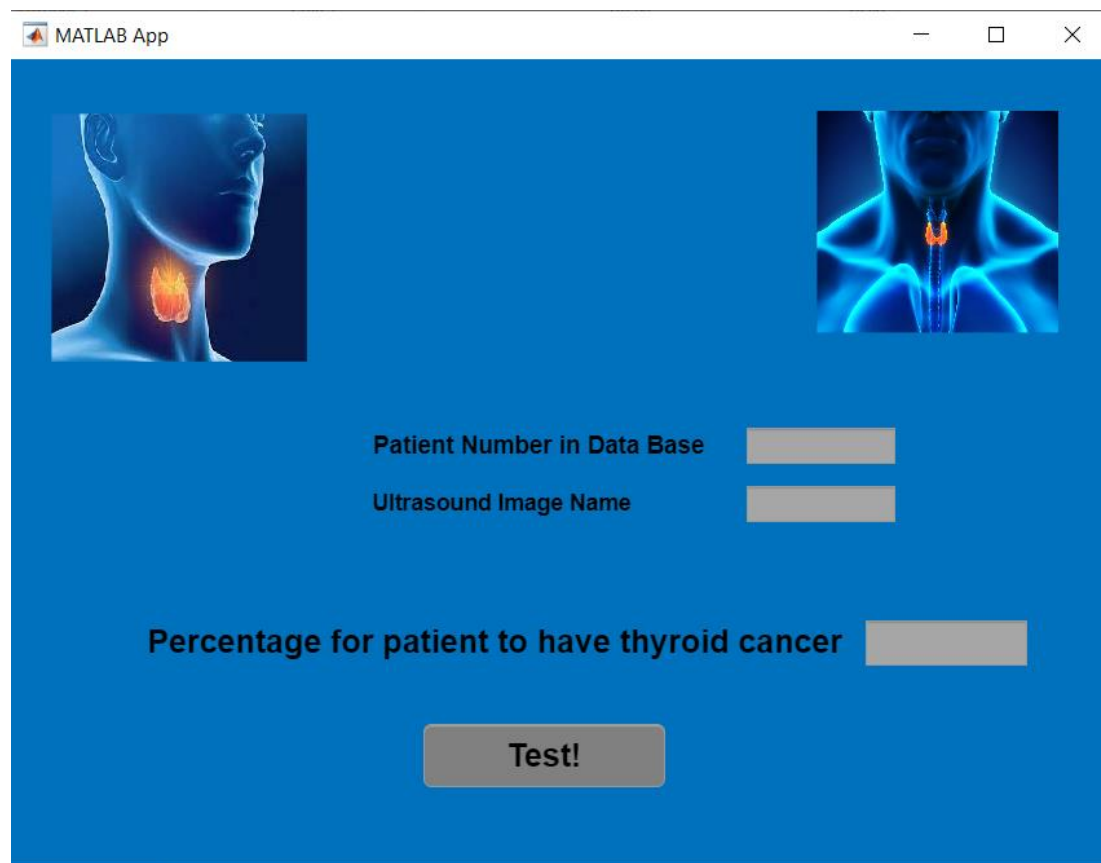
```

În codul de mai sus am aranjat să apară și indicatorul de procentaj „%” pentru a înțelege mai bine. De asemenea am optat și pentru câteva îmbunătățiri vizuale, pentru a ajuta medicul care ar folosi această aplicație, de exemplu să vadă numărul pacientului din baza de date, precum și numele ecografiei pentru a putea urmări mai bine și a pierde mai puțin timp să caute manual.

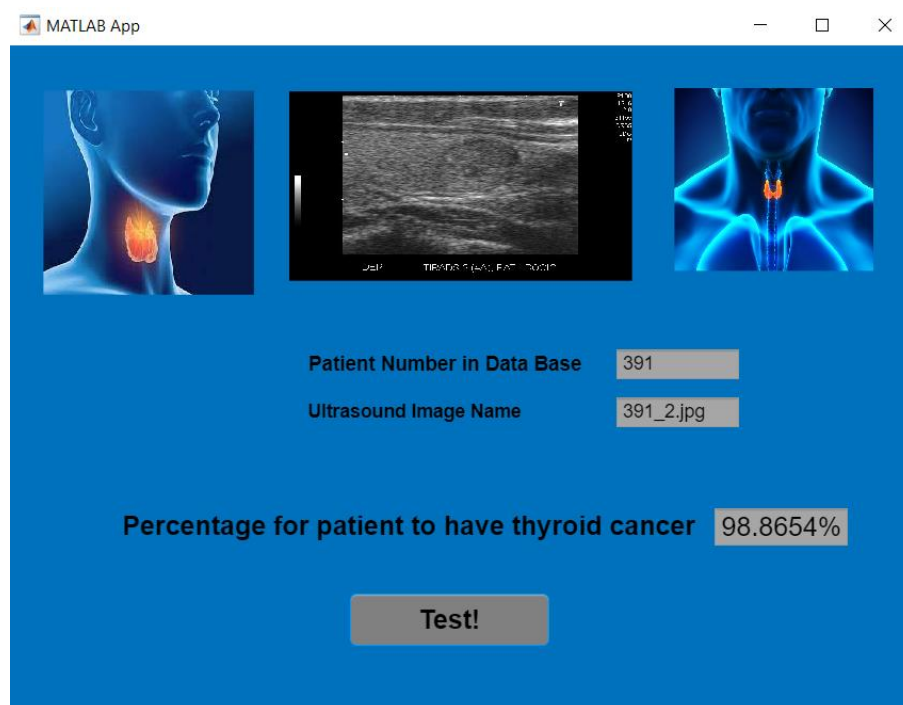
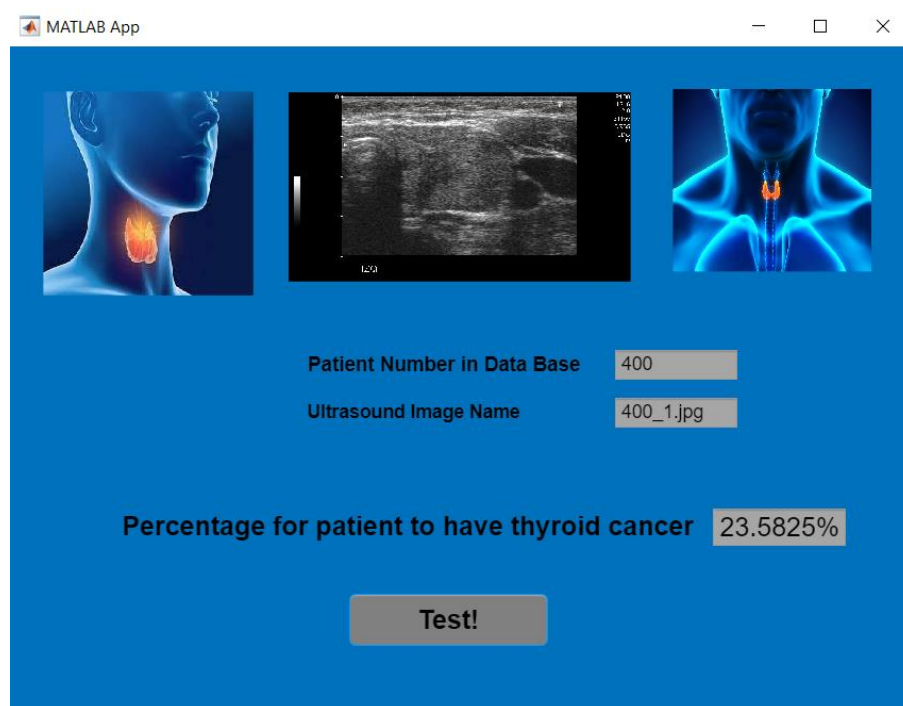
În cele ce urmează voi prezenta câteva poze sugerând folosirea aplicației și funcționalitatea acesteia.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

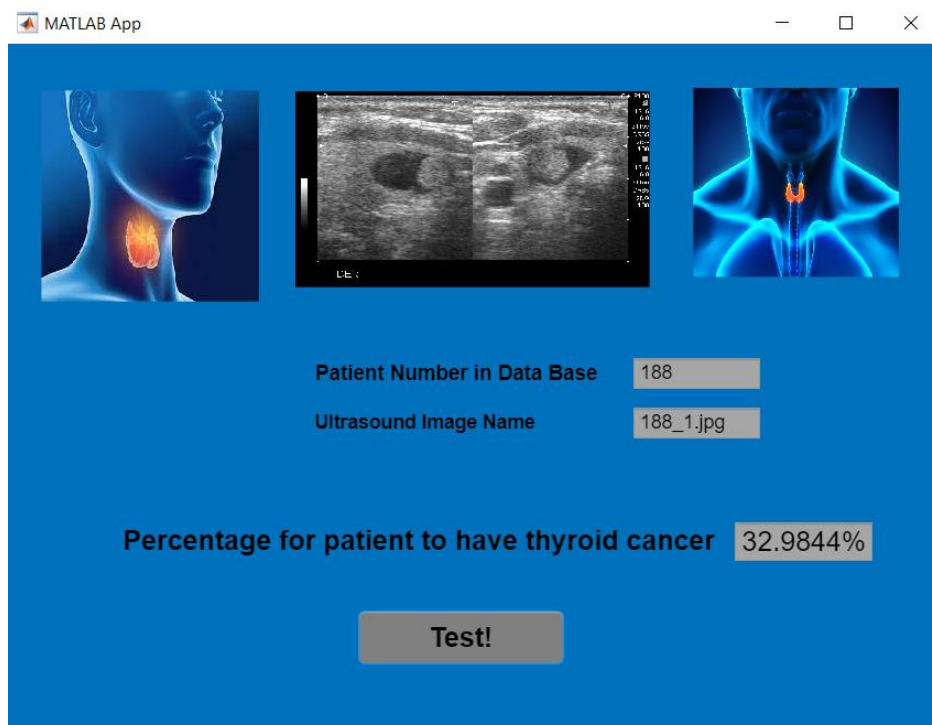
Designul Aplicației este prezentat mai jos.



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	



PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	



In cele de ma sus am prezentat functionarea aplicatiei pentru diferite imagini de test. Din pacate rezultatele afisate nu sunt de o acuratete dorita, in unele cazuri, dar nu in toate, ne arata ca ar avea cancer insa imaginea este fara, sau invers. Aceste performante slabute sunt datorita faptului ca mediul MATLAB nu are suficienta memorie la dispozitie pentru a crea o retea precisa, ceea ce a dus la numar de neuroni foarte mic, si imagini putine folosite la antrenare.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	

4. Concluzii

a. Rezultate obținute

Dupa cum spuneam, rezultatele obtinute sunt din pacate de o acuratete mica, datorita limitarilor hardware-ului si a software-ului, ceea ce nu am putut rezolva. Ca si incercare personala de imbunatatire, am optat pentru a imparti valorile vectorului de imagini la 255, deoarece numerele intr-o matrice de imagini sunt valori intre 0 si 255. Acest range este foarte mare si duce la dificultati pentru calculator cand lucreaza cu numere mari si cu multe imagini.

Aceasta modificare nu schimba functionarea rețelei, deoarece noi luam datele pozelor si lucram la fel cu toate, si daca lucram cu numere mai mici nu mai solicitam atat de mult calculatorul si putem avea rezultate mai bune.

Dupa aceasta schimbare am putut utiliza 20 de neuroni, inainte puteam doar 10, ceea ce a dus la performante mai bune, dar nu perfecte. Reteaua a fost antrenata la inceput cu 100 de poze cu cancer si fara, dar dupa modificarea facuta de mine am reusit sa dublez numarul de poze la antrenare, lucru care de asemenea a ajutat la imbunatatirea performantelor.

b. Direcții de dezvoltare

Acest proiect constituie o aplicatie foarte utila in primul rand pentru medici. Acestia putand foarte usor sa o utilizeze cu orice ecografie de tiroida, si sa afle procentajul sansei de cancer.

Daca aplicatie ar dezvoltata in viitor pe un program mai bun sau pe un calculator mai performant, ar duce la rezultate cu o acuratete mult mai mare, ceea ce ar insemna ca aplicatia ar putea fi folosita in spitale sa ajute medicii sa identifice cancerul tiroidian mult mai repede. Acest lucru ar reprezenta un avantaj enorm atat pentru medici cat si pentru pacienti, deoarece identificarea cancerului cat mai repede poate duce la tratarea acestuia .

O alta directie de dezvoltare ar fi utilizarea acestei aplicatii si pentru ecografii de alt tip. Acest lucru constituie si o contributie personala la crearea codului. Reteaua poate fi reantrenata cu imagini cu cancer de orice alt tip, ceea ce ar duce la un proiect universal de identificare al cancerului la pacienti.

PROIECT INGINERIA REGLARII AUTOMATE II					
NUME student	Marinescu Șerban Nicolae	GRUPA:	30131	Nota	