cancer la sân: predicția ariei tumorii și clasificare

PROIECT DE DIPLOMĂ

Autor: **Anca-Elena ANDREESCU**

Conducător științific: **Prof. dr. ing. Eva-Henrietta DULF**

|  |  |
| --- | --- |
| DECAN  **Prof. dr. ing. Liviu MICLEA** | Vizat,  DIRECTOR DEPARTAMENT AUTOMATICĂ  **Prof. dr. ing. Honoriu VĂLEAN** |

Autor: **Anca-Elena ANDREESCU**

Cancer la sân: predicția ariei tumorii și clasificare

1. **Enunțul temei:** *Realizarea unui model care poate să ofere o predicție cât mai corectă a ariei tumorii, pentru cancerul la sân. De asemenea, pentru același set de date, s-a dezvoltat un model, pentru a oferi un diagonistic, dacă tumoarea are potențial malign sau benign.*
2. **Conținutul proiectului:** *(enumerarea părților componente) Pagina de prezentare, Declarație privind autenticitatea proiectului, Sinteza proiectului, Cuprins, Titlul capitolului 1, Titlul capitolului 2,… Titlul capitolului n, Bibliografie, Anexe.*
3. **Locul documentării:** *Universitatea Tehnică din Cluj-Napoca*
4. **Data emiterii temei:**
5. **Data predării:**

Semnătura autorului

Semnătura conducătorului științific

**Declaraţie pe proprie răspundere privind**

**autenticitatea proiectului de diplomă**

Subsemnatul(a) **Anca-Elena ANDREESCU**  , legitimat(ă) cu CI/BI seria VX nr. 943559 , CNP 6020317385573 ,

autorul lucrării:

Cancer la sân: predicția ariei tumorii și clasificare

elaborată în vederea susținerii examenului de finalizare a studiilor de licență la **Facultatea de Automatică și Calculatoare**, specializarea **Automatică și Informatică Aplicată,** din cadrul Universității Tehnice din Cluj-Napoca, sesiunea Iulie 2023 a anului universitar 2022-2023, declar pe proprie răspundere, că această lucrare este rezultatul propriei activități intelectuale, pe baza cercetărilor mele și pe baza informațiilor obținute din surse care au fost citate, în textul lucrării, și în bibliografie.

Declar, că această lucrare nu conține porțiuni plagiate, iar sursele bibliografice au fost folosite cu respectarea legislației române și a convențiilor internaționale privind drepturile de autor.

Declar, de asemenea, că această lucrare nu a mai fost prezentată în fața unei alte comisii de examen de licență.

În cazul constatării ulterioare a unor declarații false, voi suporta sancțiunile administrative, respectiv, *anularea examenului de licență*.

Data Prenume NUME

(semnătura)

**SINTEZA**

proiectului de diplomă cu titlul:

Cancer la sân: predicția ariei tumorii si clasificare

Autor: **Anca-Elena ANDREESCU**

Conducător științific: **Titlu. ing. Prenume NUME**

1. Cerințele temei: Realizarea unui model care să poata să ofere o acuratețe mărită atât în procesul de predicție a ariei tumorii, cât și pentru partea de caracterizare a tumorii: malignă sau benignă.

2. Soluții alese: Utlizarea mediului de dezvoltare Pycharm. Cu ajutorul limbajului de programare Python, s-a antrenat o rețea neuronală artificială (ANN) care rezolvă problema de regresie bazata pe predictia ariei. O altă rețea artificială a fost construită pentru rezolvarea problemei bazate pe împărțirea tumorii in cele doua clase: malignă si benignă

3. Rezultate obținute: Pentru partea de regresie s-a obținut o eroare medie pătratică de de 0.0001 si un coeficient de determinare( R^2) de 0.99. De asemenea, pentru reteau neuronală care se ocupă de împărțirea tumorii în două clase, s-a obținut o acuratețe de 98%.

4. Testări și verificări: Din momentul în care retele neuronale au fost construite, s-au încercat mai multe combinații între parametrii, astfel s-au modificat valori pentru: numărul de straturi ascunse, numărul de neuroni de pe starturi, epoci. De asemenea, setul de date a jucat un rol important în antrenarea rețelei. Verificările s-au pus observa în valoarea erorii mediei pătratice, coeficientul de determinare și în modul în arătau graficele intre datele care se dorea sa fie prezise, și cele prezise.

5. Contribuții personale: Documentarea asupra modului optim de construire a rețelelor neuronale.

6. Surse de documentare:

Semnătura autorului

Semnătura conducătorului științific

Cuprins

[1 Introducere 2](#_Toc477457095)

[1.1 Context general 2](#_Toc477457096)

[1.2 Obiective 2](#_Toc477457097)

[1.3 Specificații 2](#_Toc477457098)

[2 Studiu bibliografic 3](#_Toc477457099)

[3 Analiză, proiectare, implementare 4](#_Toc477457100)

[4 Concluzii 5](#_Toc477457101)

[4.1 Rezultate obținute 5](#_Toc477457102)

[4.2 Direcții de dezvoltare 5](#_Toc477457103)

[5 Reguli de formatare 6](#_Toc477457104)

[5.1 Formatarea paginii 6](#_Toc477457105)

[5.2 Titluri și stiluri 6](#_Toc477457106)

[5.3 Figuri, tabele și ecuații 7](#_Toc477457107)

[5.3.1 Figuri 7](#_Toc477457108)

[5.4 Tabele 7](#_Toc477457109)

[5.5 Ecuații 7](#_Toc477457110)

[5.6 Referințe bibliografice 8](#_Toc477457111)

[6 Bibliografie 9](#_Toc477457112)

# Introducere

## Context general

Cancerul de sân a reprezentat dintotdeauna unul dintre principalele motive de deces în răndul femeilor, aproximativ 15% din numărul total. Din acest motiv, importanța găsirii unui diagnostic corect v-a putea duce la un tratament eficient aplicat pacientelor. [1] Șansele de viață a femeilor pot să fie influențate major de momentul în care este confirmat rezultatul medical, astfel se evidențiază importanța cunoașterii timpurie a diagonisticului.

De-a lungul anilor, aria medicinei s-a dezvoltat foarte mult , astfel reusind să ofere tratamente potrivite si mult mai eficiente pentru pacienti. Totusi, învățarea automată poate să aducă un beneficiu considerabil în clasificarea tumorii, dar si pentru modul în care aria tumorii o să evolueze pe parcursul timpului. [1] O astfel de predicție eficientă asupra particularității tumorii poate să îi ajute si pe medicii oncologi să poată oferii o îngrijire medicală adecvată. Pe parcursul anilor s-a încercat introducerea învățării automate în ajutorul medicinei, iar acest lucru a dus la lucruri inovatoare.

Prin această lucrare, s-au atins două puncte vitale care caraterizează o tumoare: modulul în care o să se dezvolte aria tumorii, bazat pe anumite intrări, respectiv dacă aceasta are o caracteristică malignă sau benignă. Tumoarea reprezintă o acumularea excesivă de celule , principalul motiv pentru care se formează este deoarece această organizare de celule se divide anormal de mult sau din simplu fapt că acestea nu mor într-un timp așteptat. Tumoarea benigna în mod normal nu cauzeaza probleme pacienților, dar chiar dacă au un ritm lent de creștere, pot să ajungă în punctul în care să influenteze funcționarea normala a celorlalte organe din corp. Pe de altă parte, tumoarea malignă se caracterizează printr-un ritm rapid si necontrolat. Ele reprezintă un real pericol pentru corpul uman, reușind să se răspandească local, dar și în zone îndepărtate prin intermediul sângelui.[2] Aceste aspecte arată importanța cunoașterii tipului tumorii, din acest motiv în această lucrare s-a studiat și încercat găsirea unui algoritm cât mai potrivit pentru aceasta sarcină. De sigur, cunoscând caracteristica tumorii maligne, creștere sporadică, se justifică și dorința medicilor de a cunoaște maniera de evoluție a mărimii tumorii. Această lucrare reflectă dorința de a ajuta și de a ușura modul de gestionare a tratamentului oncologic aplicat oamenilor.

Lucrarea conține capitole în care sunt explicate concis partea de implementare, concluzii, dar și testele realizate. Setul de date pe care s-au realizat aceste rețele neuronale este obținut de pe UI machine learning, sustrase dintr-o imagine digitalizată a unei mase mamare. Baza de date se numeste Breast Cancer Wisconsin (Diagnostic). Acesta include 569 de instanțe care descriu caracteristicile celulelor.

În introducere familiarizați cititorul cu motivația lucrării, plasați lucrarea într-un context care să permită cititorului să înțeleagă obiectivele.

Descrieți importanța lucrării, de ce merita să o faceți, plasați ideile într-un context larg.

Susțineți studiul: de ce exact aceasta aplicație/implementare. Comentați asupra aspectelor teoretice sau practice care v-au făcut să o alegeți.

În funcție de natura lucrării, ar putea fi necesar să prezentați informații de fond asupra domeniului în care se încadrează aplicația, mai ales dacă aveți o lucrare într-un domeniu multidisciplinar. În acest caz, puteți introduce terminologia pe care o utilizați în continuare.

Descrieți pe scurt lucrarea: ce conține fiecare capitol.

## Obiective

Lucrarea de față are două obiective principale, care doresc să ofere o întelegere mai profundă a modului in care o sa se dezvolte aria tumorii prezente în zona mamară, în plus reușește să ofere o clasificare a tipul tumorii cu o acuratețe de aproximativ 98%.

Predicția dimensiunii tumorii si clasificarea tipului acesteia, au fost realizate cu ajutorul unui set de date de tip numeric. Acestea conțin informații importante legate de diametru, textură, rază ( distanța medie de la punctele de pe perimetru), compactitate, concavitate (severitatea porțiunilor concave ale conturului), puncte concave (numărul de porțiuni concave ale conturului), simetrie, netezimea ( variația locală a lungimilor razei), diagnostic ( B = benignă, M= malignă), aria tumorii.

Pentru reteaua neuronala care reaizeaza clasificarea tumroii s-a folosit ca si output coloana de „ diagnostic”, iar pentru cea care se ocupă cu pedicția de arie s-a folosit coloana cu numele „arie”.

Toate aceste informații provenite de la mase mamare au contribuit la realizarea rețelelor neuronale artificale, avand ca și scop final perfecționarea diagnosticului și tratamentului cancerului la sân, furnizând resurse medicale asistate de calculator care pot sprijini medicii în luarea deciziilor clinice. Prin aceste modele, medicii ar trebui să beneficieze de o prividere de ansamblu asupra modului în care urmează să gestioneze situația pacinetului. Tot odata, aceste modele reușesc să elimine mai mult eroarea umană, riscul unei diagnosticări eronate.

## Specificații

În specificațiile lucrării detaliați cerințele. Descrieți ce intenționați să obțineți. Vă puteți referi la funcțiile aplicației, interfață, nivele de performanță, structuri de date, elemente, securitate, fiabilitate, calitate, limitări, etc.

# Studiu bibliografic

Conține o analiză a ceea ce s-a realizat/studiat anterior. Arătați că ați studiat materiale bibliografice și că ați înțeles ceea ce ați citit.

Puteți include diferite puncte de vedere asupra problemei pe care o rezolvați în lucrare.

Nu uitați să citați corespunzător autorii oricărei idei extrase dintr-o sursă bibliografică.

# Analiză, proiectare, implementare

Aceasta parte a lucrării este flexibilă și depinde foarte mult de natura lucrării, poate fi organizată în mai multe capitole și conține contribuțiile personale ale autorului.

Includeți:

* + Detalii referitoare la analiză și proiectare:
    - descrierea metodelor pe care le-ați aplicat pentru rezolvarea problemei,
    - descrierea materialelor, procedurilor
    - calcule, tehnici, descrierea echipamentelor
    - metodologia de proiectare
    - informațiile necesare pentru ca cineva să poată reface lucrarea
  + Implementare :
    - Descrieți detaliile tehnice ale implementării aplicației: mediul de implementare, modul de prezentare, modul de utilizare al aplicației, etc.
  + Testare si validare :
    - Descrieți metodologia de testare a aplicației și rezultatele
    - Includeți experimentele pe care le-ați realizat, analiza rezultatelor pe care le-ați obținut.

# Concluzii

## Rezultate obținute

Evidențiați toate rezultatele pe care le-ați obținut și trageți concluzii din ele. Puteți prezenta o analiză critică a ceea ce ați realizat comparativ cu alte lucrări/studii anterioare.

Includeți o listă a contribuțiilor pe care le-ați avut în domeniul temei abordate.

## Direcții de dezvoltare

Descrieți direcțiile posibile de dezvoltare.

# Reguli de formatare

## Formatarea paginii

* + Dimensiunea paginii: A4
  + Margini: 2.5 cm (sus, jos, stânga, dreapta)
  + Antet și subsol: 1.27 cm de la marginea paginii
  + În antetul paginii (header): titlul capitolului, centrat, stil: Header\_style
  + În subsolul paginii: numărul paginii, centrat

## Titluri și stiluri

Titlurile capitolelor și subcapitolelor se marchează cu stilurile Heading 1 – 4, conform documentului model anexat în format Word. Descrierea stilurilor utilizate în document este prezentată în Tabelul 5.1.

Tabelul 5.1. Stiluri utilizate în acest document

| Nr. | Stil | Utilizat pentru | Format |
| --- | --- | --- | --- |
| 1 | Normal | Text normal | Font: (Default) Cambria, 12 pt, Justified, Line spacing: Multiple 1.1 li, Space After: 6 pt |
| 2 | Titlu | Titlul proiectului, prima pagină | Font: 24 pt, Small caps, Centered Line spacing: single, Space Before: 126pt, After: 0 pt, |
| 3 | Titlu2 | Titlul proiectului, pagina de prezentare | Font:14pt, Bold, Centered |
| 4 | Heading 1 | Titlurile capitolelor (nivel 1) | Font: 24 pt, Indent: Left: 0 cm Hanging: 0.76 cm, Space Before: 24pt, After: 12pt |
| 5 | Heading 2 | Titlurile subcapitolelor (nivel 2) | Font: 14 pt, Bold, Indent: Left: 0 cm  Hanging: 1.02 cm, Space Before: 18pt, After: 12pt |
| 6 | Heading 3 | Titlurile secțiunilor (nivel 3) | Font: Bold, Indent: Left: 0 cm Hanging: 1.27 cm, Space Before: 6 pt, After: 6pt |
| 7 | Heading 4 | Titlurile secțiunilor (nivel 4) | Font: Italic, Indent: Left: 0 cm Hanging: 1.52 cm, Space Before: 2 pt, After: 0 pt |
| 8 | Caption | Legenda figurilor și tabelelor | Font: Italic, Font color: Text 1, Line spacing: single, Space After: 10 pt, |
| 9 | Header\_style | Antetul paginii | Font: 10 pt, Italic, Centered, Border: Bottom: (Single solid line, Background 1, 0.5 pt Line width) |

## Figuri, tabele și ecuații

### Figuri

Figurile se inserează în text centrate, cu etichetă de numerotare și legendă (Caption) în partea de jos a figurii. Numărul figurii include și numărul capitolului, după exemplul prezentat în Figura 5.1.



Figura 5.1. Figură exemplu, stil: Caption

## Tabele

Tabelele se inserează în text centrate, cu etichetă și legendă (Caption) în partea de sus a tabelului, aliniată la stânga. Numărul tabelului include și numărul capitolului, după cum este prezentat, de exemplu, în Tabelul 5.1.

## Ecuații

Ecuațiile se inserează în text centrate, cu numerotare în partea dreaptă. Numărul ecuației include și numărul capitolului, conform exemplului din relația (5.1).

|  |  |
| --- | --- |
|  | (5.1) |

## Referințe bibliografice

Se recomandă ca citarea referințelor bibliografice să fie făcută în formatul IEEE.

În secțiunea Bibliografie sunt prezentate exemple pentru: o citare a unui capitol dintr-o carte [1], un articol publicat într-o revistă [2] și un articol publicat la o conferință [3].

Detalii cu privire la formatul citării diverselor tipuri de referințe pot fi găsite în [4] sau [5].

Referințele bibliografice se pot insera în text utilizând facilitățile Word de a adăuga surse și bibliografie unui document (References -> Citations & Bibliography). Dacă formatul IEEE pentru bibliografie nu este instalat implicit în Word, se poate descărca gratuit de la:

<https://bibword.codeplex.com/wikipage?title=Styles&referringTitle=Home>

Instrucțiunile de instalare pentru diferite versiuni de Word se pot obține de la aceeași adresă.

6.Bibliografie

[1] Yue, Wenbin, Zidong Wang, Hongwei Chen, Annette Payne, and Xiaohui Liu. 2018. "Machine Learning with Applications in Breast Cancer Diagnosis and Prognosis" Designs 2, no. 2: 13.

[2] Patel A. Benign vs Malignant Tumors. JAMA Oncol. 2020;6(9):1488.