

UNIVERSITATEA BABEȘ-BOLYAI CLUJ-NAPOCA  
FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ  
SPECIALIZAREA MATEMATICĂ INFORMATICĂ

## LUCRARE DE LICENȚĂ

### Studiul evoluției rinosinuzitei cronice prin folosirea tehnicilor de învățare automată

Conducător științific  
Prof. DIOȘAN Laura

*Absolvent*  
*Ursu Teodora Maria*

2023



---

## ABSTRACT

---

Abstract: Chronic rhinosinusitis with nasal polyps is an ongoing inflammatory condition that affects the nasal sinuses and upper airways. It is characterized by persistent inflammation and the growth of nasal polyps in the nasal cavity and sinuses.

The treatment options for this condition include medication therapy or surgical removal of the polyps. However, the exact cause of polyp formation is still unknown.

The aim of these experiments was to estimate the POSE values after polyp removal surgery, at the 6-month and 12-month postoperative intervals. To achieve this, we utilized Decision Tree and Random Forest algorithms for classification and regression analysis.

The first chapter introduces the study's objectives and some analysis of the existing literature, and explains the underlying motivation.

In the second chapter, we provide a comprehensive overview of the subject, emphasizing the importance of the data used and the predicted outcomes from a medical perspective and a ML algorithms perspective.

The third chapter delves into the theory, introducing machine learning and its applications in medicine, and provides a detailed explanation of the Decision Tree and Random Forest models.

Chapter four examines the features used in the experiments, including the pre-processing steps taken to ensure their suitability. We also document the conducted experiments, each with its own unique characteristics. Additionally, we discuss the development of the PosePredict system, a user-friendly website that allows health-care professionals to utilize the generated models.

The paper presents all the conducted experiments, involving both classification and regression analyses. These experiments were performed simultaneously using both Decision Tree and Random Forest models. We also introduce PosePredict, a website that takes input data and uses the Random Forest model to predict the POSE value.

The results obtained showed an accuracy of at least 80% in predicting these values. Random Forest consistently outperformed Decision Tree in terms of performance. The experiments using classification also yielded better results.

# Cuprins

<b>1</b>	<b>Introducere</b>	<b>1</b>
1.1	Motivație și obiective . . . . .	1
1.2	Contribuții originale . . . . .	1
1.2.1	Analiza literaturii de specialitate . . . . .	2
1.2.2	Modelarea problemei ca problemă de regresie și problemă de clasificare . . . . .	3
1.2.3	Studiu experimental . . . . .	4
1.2.4	Dezvoltarea sistemului Pose Predict . . . . .	4
1.3	Structura lucrării pe capitole . . . . .	5
<b>2</b>	<b>Problematică</b>	<b>6</b>
2.1	Din punct de vedere medical . . . . .	6
2.2	Din punct de vedere al datelor . . . . .	7
<b>3</b>	<b>Elemente teoretice</b>	<b>9</b>
3.1	Machine Learning . . . . .	9
3.1.1	Scurtă introducere în Machine Learning . . . . .	9
3.1.2	Machine Learning în medicină . . . . .	10
3.1.3	Măsurarea performanței algoritmilor de ML . . . . .	10
3.2	Decision Tree . . . . .	12
3.2.1	Parametri . . . . .	12
3.2.2	Avantaje . . . . .	13
3.2.3	Dezavantaje . . . . .	13
3.2.4	Domenii de aplicabilitate . . . . .	14
3.3	Random Forest . . . . .	15
3.3.1	Parametri . . . . .	15
3.4	Avantaje . . . . .	16
3.4.1	Dezavantaje . . . . .	16
3.4.2	Domenii de aplicabilitate . . . . .	16

<b>4</b>	<b>Studiu de caz</b>	<b>18</b>
4.1	Baza de date, preprocesare și normalizare . . . . .	18
4.2	Validare experimentală . . . . .	24
4.2.1	Experiment I . . . . .	24
4.2.2	Experiment II . . . . .	27
4.2.3	Experiment III . . . . .	35
4.2.4	Experiment IV . . . . .	41
4.2.5	Experiment V . . . . .	54
4.3	Dezvoltare sistem PosePredict . . . . .	66
4.3.1	Funcționalități: . . . . .	66
4.3.2	Proiectare: . . . . .	67
4.3.3	Implementare . . . . .	68
4.3.4	Implementare efectivă . . . . .	69
4.3.5	Modelele de predicție bazate pe Machine Learning folosite . .	70
4.3.6	Validare și testare . . . . .	77
<b>5</b>	<b>Concluzii</b>	<b>79</b>
	<b>Bibliografie</b>	<b>80</b>

# Capitolul 1

## Introducere

### 1.1 Motivație și obiective

Rinosinuzita cronică cu polipi nazali este o afecțiune inflamatorie cronică a cavităților nazale și a sinusurilor, caracterizată prin inflamație persistentă și creșterea polipilor în mucoasa nazală. Este considerată una dintre cele mai comune afecțiuni cronice ale cavității nazale și sinusurilor, afectând în special adulții, dar poate apărea și la copii.

Este important ca pacienții să fie supravegheați pe termen lung și să primească tratamente adecvate pentru a controla simptomele și a preveni recidiva.

Experimentele realizate vizează prezicerea evoluției recurenței bolii, la șase luni, doisprezece luni și un an și jumătate după operație. Recurența bolii se masoară folosindu-se sistemul de scor POSE(Perioperative Sinus Endoscopy), metrică care încearcă să fie precisă de către modele.

Pentru realizarea perzicerilor s-au folosit arbori de decizie. În paralel, în cadrul aceluiași proiect, au fost realizate experimente folosindu-se Random Forest, de către doctoranda ADEL BAJCSI. Rezultatele experimentelor folosind Random Forest au fost considerabil mai bune, putând afirma că modelul RF este mai potrivit pentru rezolvarea problemei.

### 1.2 Contribuții originale

În vederea realizării aplicației, am analizat literatura de specialitate în domeniu. Există diverse studii care analizează aplicații ale inteligenței artificiale în diagnosticarea sau prezicerea evoluției bolii în problema rinosinuzitei cu (sau fără) polipi nazali.

### 1.2.1 Analiza literaturii de specialitate

#### “Expert-level diagnosis of nasal polyps using deep learning on whole-slide imaging”

În acest articol se studiază rata de apariție a eozinofilelor postoperatorii. “S-a observat că prezența acestora în polipii nazali se află în corelație directă cu scoruri mari ale durerii, o calitate a vieții puternic afectată și o rată de recurență sporită[7]”.

În acest studiu sunt folosite 3 modele de rețele neuronale (Resnet50, Xception și Inception V3) pentru clasificarea și Inception V3 pentru regresie. Imaginile sunt obținute cu ajutorul unui scanner, care digitalizează și transformă mostrele în imagini whole-slide. După 100 de iterații prin setul de date, nu se mai observă îmbunătățiri ale erorii medii absolute, așa că sunt salvate rezultatele obținute.

În urma tuturor analizelor, s-a concluzionat că Inception V3 este cea mai fiabilă opțiune, iar modelul obținut a fost introdus în AICEP 1.0.

Studiul a fost continuat în vederea realizării platformei AICEP 2.0. De această dată, se folosește o rețea neuronală convoluțională, EfficientNet B5. EfficientNets reprezintă familia de rețele neuronale bazată pe modelul Neural Architecture Search.

Neural Architecture Search este o tehnică de automatizare a proiectării rețelelor neuronale artificiale. Tipul de rețea neuronală artificială care poate fi proiectată depinde de spațiul de căutare. Rezultatele celui de-al doilea studiu au fost comparativ mai bune față de primul.

	MAE (%)	RMSE (%)	$R^2$
<b>AICEP 1.0</b>	4.30	7.16	0.92
<b>AICEP 2.0</b>	1.64	2.79	0.98

Tabela 1.1: analiza AICEP 1.0 vs AICEP 2.0

#### “Using Machine Learning to Recognize Chronic Rhinosinusitis”

În acest articol se pune problema dacă persoanele cu rinosinuzită cronică vorbesc diferit suficient cât să fie recunoscuți de un algoritm de învățare automată (machine learning). Pentru a afla răspunsul, subiecții au înregistrat un test la o frecvență de eșantionare de 44100 kHz. Fișierele înregistrate au fost prelucrate, iar semnalul a fost ulterior procesat folosind transformata Fourier rapidă (FFT), transformata wavelet continuă (CWT) și spectrogramele.

Datele obținute din spectrograme au fost transformate în vectori, iar apoi un model a fost antrenat pe acei vectori. Pentru a urmări eficacitatea algoritmului, s-a urmărit funcția de precizie - raportul dintre valorile pozitive care sunt cu adevărat

pozitive (dacă algoritmul spune că 5 subiecți au rinosinuzită cronică, dar doar 3 dintre ei o au, precizia va fi de 0.6) și funcția recall - raportul dintre valorile pozitive care au fost găsite (dacă au fost identificați 8 pacienți din 10 care aveau afecțiunea, recall-ul va fi de 0.8).

	precizie	recall
<b>nas neobturat</b>	0.64	0.46
<b>nas obturat</b>	0.58	0.74

Tabela 1.2: tabel comparativ pentru raportul de clasificare [8]

### **"Automated classification of nasal polyps in endoscopy video-frames using hand-crafted and CNN features"**

În acest articol se studiază eficacitatea utilizării învățării automate (Machine Learning) în ceea ce privește rinosinuzita cronică. Pentru aceasta, s-a folosit un set de date preluat de la 80 de pacienți, iar algoritmul de învățare automată "vanilla" (simplu) și clasificatorii de învățare adâncă (deep learning) au fost urmăriti. În vederea realizării studiului, s-au utilizat imagini endoscopice WSI (Whole Slide Imaging), care au fost corelate cu analizele medicale.

Concluziile studiului indică faptul că algoritmi de deep learning sunt mai eficienți, prezentând o rată de performanță de **98,3%**, în comparație cu performanța de **96,3%** a algoritmului "vanilla" de învățare automată.[9] Totuși, ambele tipuri de experimente au evidențiat rezultate foarte bune în detectarea rinosinuzitei cronice.

## **1.2.2 Modelarea problemei ca problemă de regresie și problemă de clasificare**

Regresia este o tehnică utilizată în machine learning pentru a crea modele care pot prezice o valoare numerică continuă, pe baza unui set de variabile de intrare. Acest proces implică înțelegerea relației dintre variabilele de intrare și țintă, astfel încât să se poată dezvolta o funcție sau un model matematic capabil să estimeze acea valoare pentru date noi.

Clasificarea este o tehnică utilizată pentru a dezvolta modele capabile să atribuie obiectelor o clasă predefinită, prin împărțirea lor în categorii distincte. Scopul principal al clasificării este de a crea un model care poate face predicții corecte pentru noi date, prin identificarea caracteristicilor relevante care diferențiază clasele.

Astfel, regresia și clasificarea reprezintă două abordări diferite în machine learning. Regresia se concentrează pe estimarea unei valori continue, în timp ce clasifi-



carea se concentrează pe atribuirea unei clase discrete.

În primele experimente realizate s-a folosit regresia, în încercarea prezicerii unei valori POSE concrete. Ulterior, s-a considerat împărțirea valorilor POSE în 3 categorii, după cum urmează:

- pentru o valoare POSE între 0 și 8: 0
- pentru o valoare POSE între 8 și 16: 1
- pentru o valoare POSE între 16 și 32: 2

Valorile prezise în cazul clasificării reprezintă clasele din care fac parte valorile POSE.

### 1.2.3 Studiu experimental

În vederea găsirii modelelor optime care pot prezice valorile POSE cât mai corect, s-au încercat 5 experimente, fiind modificate caracteristicile datelor (features) folosite, parametri arborilor și s-a încercat, comparativ, atât regresia cât și clasificarea. Totodată, experimentele au fost realizate în paralel, cu decision tree și cu random forest.

În primul experiment s-au folosit datele în forma lor inițială, iar analizele miR 125 și miR 203 nu au fost incluse. Datele de intrare și rezultatele au fost exprimate pentru fiecare fosă nazală. În cel de-al doilea experiment, caracteristicile pentru cele două fose nazale au fost însumate. Totodată, s-a adăugat markerul de vârstă, care indică dacă pacientul are sau nu peste 64 de ani. În cel de-al treilea experiment s-a studiat importanța prezenței atât a valorii HPQ-9, cât și a valorii SNOT ca și caracteristici ale datelor, fiind date puternic corelate. În urma acestor experimente s-a decis eliminarea valorii HPQ-9. În cel de-al patrulea experiment, s-a făcut trecerea de la regresie la clasificare. În experimentul numărul cinci s-a studiat importanța analizelor miR 125 și miR 203 în predicția evoluției bolii.

### 1.2.4 Dezvoltarea sistemului Pose Predict

Sistemul PosePredict reprezintă un website care facilitează folosirea modelelor generate. Printr-o interfață accesibilă și intuitivă, aplicația are ca scop crearea punții de legătură dintre doctori și lumea informatică. Analizele medicale necesare acestor modele sunt prelevate în cadrul unuiu sau a mai multor consulturi de specialitate. După realizarea consultului, medicul are posibilitatea, prin intermediul site-ului creat, să stocheze aceste informații și să afle imediat predicția evoluției bolii pacientului în 6, respectiv 12 luni.

## 1.3 Structura lucrării pe capitole

În capitolul doi al acestei lucrări este prezentată o introducere în tematica și problematica abordată. Pe de-o parte, se pune accentul pe importanța, din punct de vedere medical, a datelor pe care le folosim și a datelor prezise. Pe de altă parte, se prezintă, din punct de vedere al informațiilor pe care le aduc aceste date, importanța lor.

În capitolul trei sunt prezentate aspectele teoretice, precum o introducere în ceea ce înseamnă machine learning, machine learning în medicină, dar și o prezentare pe larg a celor două modele folosite: Decision Tree și Random Forest.

În capitolul patru sunt prezentate toate caracteristicile folosite, precum și procesele de prelucrare prin care au trecut în vederea folosirii lor în experimente. Tot în acest capitol sunt documentate experimentele efectuate, fiecare având propriile caracteristici. În ultima parte a capitolului este prezentată dezvoltarea sistemului PosePredict, un website ce vine în întâmpinarea medicilor, oferind un cadru accesibil pentru a se folosi de modelele obținute.

# Capitolul 2

## Problematică

### 2.1 Din punct de vedere medical

Rinosinuzita cronică cu polipi nazali este o afecțiune inflamatorie cronică a cavităților nazale și a sinusurilor, caracterizată prin inflamație persistentă și creșterea polipilor în mucoasa nazală. Este considerată una dintre cele mai comune afecțiuni cronice ale cavității nazale și sinusurilor, afectând în special adulții, dar poate apărea și la copii.

Boala este rezultatul unei reacții inflamatorii la diferiți stimuli, cum ar fi alergenii, infecțiile repetate ale căilor respiratorii superioare, factorii genetici și imunologici. Această inflamație cronică determină modificări structurale ale mucoasei nazale, iar pe măsură ce boala progresează, polipii se formează și cresc în cavitățile nazale și sinusuri.

Oamenii de știință nu au determinat pe deplin cauzele apariției polipilor nazali, nu este clar de ce inflamația dezvoltată determină formarea polipilor la unii pacienți și la alții nu.

”Polipii nazali se pot forma la orice vârstă, dar cel mai adesea se întâlnesc la adulții tineri și la cei de vârstă mijlocie. Se pot forma oriunde în sinusuri și în căile nazale dar apar de regulă în zona unde sinusurile de lângă ochi și din zona obrazului drenează în fosa nazală printr-un canal (complex osteomeatal)[1].”

Simptomele caracteristice ale rinosinuzitei cronice cu polipi nazali includ congestia nazală persistentă, secreția nazală abundentă și descărcare posterioară, pierderea sau diminuarea mirosului, dureri faciale sau presiune în zona sinusurilor, cefalee, oboseală și dificultăți de respirație. Aceste simptome pot varia în intensitate și pot afecta semnificativ calitatea vieții pacienților, interferând cu somnul, respirația normală și activitățile zilnice.

”Calitatea vieții a fost restricționată de rinosinuzita cronică la 94% dintre pacienți înainte de operație și a fost clasificată ca fiind severă sau intolerabilă în 74% dintre

cazuri. Simptomele principale ale rinosinusitei cronice au fost obstrucția nazală în 92% dintre cazuri și scurgerea nazală posterioară în 87% dintre cazuri[2].”

Diagnosticul rinosinuzitei cronice cu polipi nazali se bazează pe istoricul simptomelor pacientului, examinarea fizică a cavităților nazale și sinusurilor, precum și pe imagistica medicală, cum ar fi tomografia computerizată (CT) sau imagistica prin rezonanță magnetică (IRM). Aceste investigații ajută la evaluarea gradului de obstrucție a sinusurilor și la identificarea polipilor.

Tratamentul implică o abordare multidisciplinară, care poate include tratament medicamentos și intervenții chirurgicale. Medicamentele utilizate includ corticosteroizi nazali pentru reducerea inflamației și a congestiei nazale, antibiotice pentru tratamentul infecțiilor asociate, medicamente antialergice și imunoterapie pentru controlul reacțiilor alergice. În cazurile severe sau rezistente la tratamentul medicamentos, se poate opta pentru intervenții chirurgicale, cum ar fi sinusectomia endoscopică și îndepărtarea polipilor nazali pentru deblocarea sinusurilor obstruate și restabilirea ventilării și drenajului sinusurilor.

Rinosinuzita cronică cu polipi nazali poate avea complicații semnificative care pot afecta starea de sănătate și calitatea vieții pacienților. Printre acestea se numără:

Infecții recurente: Inflamația cronică și obstrucția sinusurilor pot favoriza dezvoltarea infecțiilor bacteriene recurente. Aceste infecții pot cauza simptome severe, cum ar fi febră, durere facială intensă și secreție purulentă nazală.

Complicații oculare: Rinosinuzita cronică cu polipi nazali poate afecta sinusurile etmoidale, care sunt situate în apropierea ochilor. În cazurile severe și neglijate, inflamația și infecția din sinusurile etmoidale pot determina apariția complicațiilor oculare, cum ar fi inflamația pleoapelor, infecții ale ochilor, și chiar pierderea vederii.

Astm și alergii respiratorii: Pacienții sunt adesea predispuși la astm și alergii respiratorii. Această legătură se datorează inflamației cronice și răspunsului alergic asociat în căile respiratorii superioare și inferioare. Astmul poate agrava simptomele rinosinuzitei și poate necesita tratament suplimentar pentru a controla inflamația.

## **2.2 Din punct de vedere al datelor**

Din punct de vedere al datelor, se încearcă prezicerea valorii POSE a pacientului, pe baza analizelor postoperatorii efectuate, a rezultatelor extrase din polipii nazali sau a rezultatelor obținute după controalele postoperatorii. Pe parcursul experimentelor s-au folosit două modele de machine learning: Random Forest respectiv Decision Tree, și s-au analizat rezultatele comparativ. Totodată, s-au folosit atât modele de regresie cât și modele de clasificare.

Datele de intrare folosite reprezintă datele pacienților, date precum vârsta, sexul, dacă este sau nu fumător sau dacă are diverse alergii, dar și rezultate ale anali-

zilor precum miR 125. miR 125 este un microARN care ajută la reglarea activității genelor, influențând modul în care sunt produse proteinele în celule, având un rol important în evoluția postoperatorie a pacientului.

Output-ul, rezultatul prezis, îl reprezintă valoarea analizei POSE. În cazul în care vorbim despre regresie, valoarea prezisă reprezintă valoarea POSE, valoare între 0 și 32. În cazul clasificării, modelul va prezice clasa din care face parte valoarea POSE (0 - pentru o valoare între 0 și 8, 1 - pentru o valoare POSE între 8 și 16 și 2 - pentru o valoare POSE între 16 și 32).

# Capitolul 3

## Elemente teoretice

### 3.1 Machine Learning

#### 3.1.1 Scurtă introducere în Machine Learning

Machine learning are o istorie îndelungată, care a început în anii '50-'60, odată cu dezvoltarea primelor tehnici de învățare automată. În acea perioadă, cercetătorii au început să exploreze modul în care computerele pot fi programate să învețe din experiență și să-și îmbunătățească performanța în mod automat.

Unul dintre primii algoritmi de machine learning dezvoltat a fost Perceptronul, propus de Frank Rosenblatt în anii '50. Acesta a fost un model simplu de rețea neuronală care putea învăța să clasifice obiecte în funcție de caracteristicile lor.

În anii '70, s-au făcut progrese în teoria învățării și s-au dezvoltat metode de învățare supervizată și nesupervizată. În învățarea supervizată, modelele sunt antrenate cu exemple etichetate, în timp ce în învățarea nesupervizată, modelele descoperă singure pattern-uri și structuri în datele de intrare.

În anii '80 și '90, cercetările au continuat să se concentreze pe dezvoltarea algoritmilor de învățare automată, cum ar fi mașinile cu vectori suport (Support Vector Machines) și arborii de decizie. Aceste tehnici au fost utilizate pentru clasificarea datelor și rezolvarea problemelor de recunoaștere a modelelor.

Odată cu avansul tehnologic și creșterea exponențială a cantității de date disponibile, interesul pentru machine learning a crescut semnificativ în anii 2000 și după. S-au dezvoltat algoritmi mai avansați, cum ar fi rețelele neuronale profunde, care pot învăța și extrage caracteristici complexe din date.

În prezent, machine learning a devenit o ramură extrem de activă și promițătoare a inteligenței artificiale, cu aplicații într-o gamă largă de domenii, inclusiv medicină, finanțe, tehnologie, robotică și multe altele. Cercetările continuă să avanseze în dezvoltarea de modele mai precise și eficiente, precum și în explorarea unor noi direcții, cum ar fi învățarea automată prin înțelegerea limbajului natural sau

învățarea prin interacțiunea cu mediul înconjurător.

### 3.1.2 Machine Learning în medicină

Învățarea automată în domeniul sănătății aduce beneficii semnificative pentru industria medicală. Aceasta permite analiza și găsirea de modele și informații în datele medicale, ajutând la diagnosticare, dezvoltarea de tratamente noi, reducerea costurilor și îmbunătățirea calității îngrijirii pacienților. Algoritmii de învățare automată pot analiza imaginile medicale, recunoscând modele specifice bolilor și facilitând diagnosticul rapid și precis. De asemenea, pot fi folosiți în descoperirea de noi tratamente și medicamente, identificarea efectelor secundare și îmbunătățirea proceselor clinice.

### 3.1.3 Măsurarea performanței algoritmilor de ML

Performanța algoritmilor de machine learning poate fi măsurată folosind diferite metrici. În cazul regresiei, avem următoarele metrici:

- **MAE (%)** : valoarea medie absolută; cu cât este mai mică, cu atât mai exact este modelul.

$$MAE = \frac{1}{N} \sum_{i=0}^N |y_i - \hat{y}_i|$$

N- numărul de elemente

$y_i$ - valoarea actuală

$\hat{y}_i$ - valoarea prezisă

$|a|$ - modulul lui a

- **MSE (%)** : eroarea pătratică medie; cu cât este mai mică, cu atât mai exact este modelul.

$$MSE = \frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2$$

N- numărul de elemente

$y_i$ - valoarea actuală

$\hat{y}_i$ - valoarea prezisă

- $R^2$ : R pătrat- reprezintă o metrică pentru a măsura exactitatea modelului de regresie. Valorile cele mai bune sunt cele care tind spre 1.

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

N- numărul de elemente

$y_i$ - valoarea actuală

$\hat{y}_i$ - valoarea prezisă

$\bar{y}$ - media valorilor reale

În cazul clasificării, metricile urmărite sunt:

- **Acuratețe**: proporția predicțiilor corecte făcute de algoritm în raport cu totalul de instanțe; valorile ideale tind spre 1.

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

TP (true positive): valoare pozitivă, clasificată ca și o valoare pozitivă

TN (true negative): valoare negativă, clasificată ca și o valoare negativă

FP (false positive): valoare negativă, clasificată ca și o valoare pozitivă

FN (false negative): valoare pozitivă, clasificată ca și o valoare negativă

- **Precizia**: proporția de instanțe corect clasificate ca fiind pozitive din totalul instanțelor clasificate ca fiind pozitive

$$P = \frac{TP}{TP + FP}$$

- **Recall**: proporția de instanțe corect clasificate ca fiind pozitive din totalul de instanțe pozitive

$$R = \frac{TP}{TP + FN}$$

- **Matrice de confuzie**: oferă o reprezentare detaliată a performanței unui algoritm de clasificare, prin evidențierea numărului de predicții corecte și eronate



		Valori inițiale	
		Positive	Negative
Valori prezise	Positive	$TP$	$FP$
	Negative	$FN$	$TN$

## 3.2 Decision Tree

Un arbore de decizie este o metodă de învățare automată supervizată folosită pentru a lua decizii pe baza unor caracteristici sau attribute date.

”Această structură de date respectă structura unei organigrame în care fiecare nod se caută un prag pentru una din trăsături, pentru a împărții datele ajunse în nodul respectiv. Se pune o întrebare la fiecare nod legată de o trăsătură, și pe baza celui răspuns, datele se împart pe cele două ramuri”[4].

Arborele de decizie poate fi utilizat în probleme de clasificare și regresie.

⇒ clasificare: acesta poate fi folosit pentru a atribui exemplele în categorii sau clase diferite, în funcție de caracteristicile lor;

⇒ regresie: arborele de decizie poate fi folosit pentru a prezice o valoare numerică continuă, bazată pe caracteristici;

### 3.2.1 Parametri

Exista mai mulți parametri care pot influența performanța arborilor. Printre acestea, cei mai des folosiți sunt :

- Adâncimea maximă (*max\_depth*): specifică adâncimea maximă a arborelui. Limitarea adâncimii poate ajuta la prevenirea *overfitting*ului și la simplificarea modelului. Cu toate acestea, o adâncime prea mică poate duce la subajustare (*underfitting*).
- Numărul minim de exemple într-un nod (*min\_samples\_split*): specifică numărul minim de exemple necesare pentru a realiza o diviziune într-un nod. Dacă numărul de exemple este mai mic decât acest prag, diviziunea nu va fi făcută și acel nod devine o frunză.
- Numărul minim de exemple într-o frunză (*min\_samples\_leaf*): specifică numărul minim de exemple necesare pentru a forma o frunză în arbore.

- Impuritatea minimă (min\_impurity\_split): specifică impuritatea minimă necesară pentru a realiza o diviziune într-un nod. Diviziunile care conduc la o scădere a impurității sub acest prag nu vor fi realizate.

Parametrii specifici arborelui de decizie cu clasificare includ:

- Criteriul de decizie (criterion): Utilizează un criteriu precum indexul Gini sau entropia informațională pentru a măsura calitatea unei diviziuni în timpul construcției arborelui.
- Stratificarea (splitter): Specifică metoda utilizată pentru alegerea atributului și a valorii de diviziune la fiecare nod

Parametrii specifici arborelui de decizie cu regresie includ:

- Criteriul de decizie (criterion): Utilizează o funcție de criteriu precum eroarea medie pătratică (MSE) sau eroarea medie absolută (MAE) pentru a măsura discrepanța dintre valorile prezise și valorile reale.
- Strategia de selecție a diviziunii (splitter): Specifică metoda utilizată pentru alegerea atributului și a valorii de diviziune la fiecare nod.

### 3.2.2 Avantaje

Interpretabilitate: poate fi ușor interpretat și explicat, oferind o înțelegere clară a modului în care sunt luate deciziile.

Manevrabilitatea datelor: poate manipula seturi de date care conțin atât caracteristici numerice, cât și categorii de date. De asemenea, poate gestiona valori lipsă sau erori în date, fără a necesita prelucrare suplimentară.

Flexibilitate: poate fi utilizat în probleme de clasificare și regresie fiind capabil să modeleze relații complexe între variabile și să identifice interacțiuni între acestea.

Eficiență în timp de execuție: Construirea și evaluarea unui arbore de decizie sunt procese relativ rapide în comparație cu alte algoritme de învățare automată.

### 3.2.3 Dezavantaje

Tendința de overfitting: poate fi predispus la overfitting, ceea ce înseamnă că se poate potrivi foarte bine setului de date de antrenament, dar poate avea o performanță scăzută pe datele noi sau de testare. Overfittingul poate apărea atunci când arborii devine prea complex și memorează detalii nesemnificative ale setului de date de antrenament.

Problema selecției caracteristicilor: poate avea dificultăți în selecția și utilizarea optimă a caracteristicilor relevante.

Tendința de fragmentare: tinde să creeze diviziuni excesive și subarbori fragmentați atunci când se confruntă cu seturi de date complexe.

Distribuția inegală de date: poate avea dificultăți în gestionarea seturilor de date cu distribuții de clasă inegale. Acest lucru poate duce la o tendință de favorizare a clasei majore și la o performanță scăzută în clasificarea clasei minore.

### 3.2.4 Domenii de aplicabilitate

Clasificare de date: Arborele de decizie poate fi utilizat pentru a clasifica obiecte sau exemple în categorii sau clase diferite. De exemplu, poate fi utilizat pentru a clasifica e-mailurile în spam și non-spam, pentru a diagnostica boli în medicină sau pentru a prezice rezultatele creditelor în domeniul financiar.

Regresie: Arborele de decizie poate fi aplicat și în probleme de regresie, în care se dorește estimarea unei valori continue. De exemplu, poate fi utilizat pentru a prezice prețul unei locuințe pe baza caracteristicilor sale sau pentru a prognoza profitul unei companii.

Sisteme de recomandare: Arborele de decizie poate fi folosit în sistemele de recomandare pentru a face sugestii personalizate, precum produse, filme, cărți sau alte elemente relevante pentru fiecare utilizator în parte.

Diagnostic medical: Arborele de decizie poate fi aplicat în diagnosticul medical, ajutând la identificarea bolilor și a afecțiunilor pe baza simptomelor și a altor caracteristici clinice. Acesta poate oferi ghidare în alegerea testelor sau a tratamentului adecvat pentru pacienți. "Decision trees are a reliable and effective decision making technique that provide high classification accuracy with a simple representation of gathered knowledge. When using decision trees, the decision-making process itself can be easily validated by an expert. Because of these reasons decision trees are especially appropriate to support decision-making process in medicine"[3]

Analiza sentimentelor: poate fi utilizat în analiza sentimentelor sau în clasificarea textelor în funcție de tonalitatea emoțională. Prin evaluarea cuvintelor și contextului, arborele poate identifica dacă un text exprimă o emoție pozitivă, negativă sau neutră.

Prezicerea churn-ului: poate fi folosit în industria serviciilor pentru a prezice probabilitatea ca un client să renunțe la un serviciu sau să se retragă dintr-un abonament.

Analiza creditelor: poate fi utilizat pentru a evalua riscul creditelor și pentru a lua decizii cu privire la aprobarea sau respingerea cererilor de împrumut. Prin analiza caracteristicilor financiare și personale ale solicitantului, arborele poate estima probabilitatea de rambursare și poate lua decizii informate.

### 3.3 Random Forest

Random Forest (Pădurea aleatoare) este o metodă de învățare automată supervizată utilizată pentru clasificare și regresie. Este o tehnică de tip ansamblu care combină mai mulți arbori de decizie independenți pentru a obține o predicție mai precisă.

”Acesta constă dintr-o combinație de numeroși arbori decizionali pentru a grupa o pădure, deoarece modelul se rulează de mai multe ori în comparație cu utilizarea unui singur arbore decizional în care modelul poate fi rulat o singură dată. În esență, clasificatorul RF urmărește o tactică repetitivă în care un arbore este selectat la întâmplare din subsetul setului de date prezent. Acești arbori decizionali ar putea să nu fie cei mai buni în general, dar împreună pun în aplicare și clasifică destul de bine setul de date.” [5]

Pentru a face o predicție cu ajutorul unui Random Forest, fiecare arbore din pădure furnizează o predicție individuală și apoi se folosește votul majoritar (în cazul clasificării) sau media (în cazul regresiei) predicțiilor furnizate de toți arborii pentru a obține predicția finală.

#### 3.3.1 Parametri

- Numărul de arbori (`n_estimators`): reprezintă numărul de arbori de decizie independenți din pădure. Cu cât este mai mare acest număr, cu atât va fi mai complex modelul și cu atât va fi nevoie de mai mult timp de antrenare. Cu toate acestea, creșterea numărului de arbori poate duce la o performanță mai bună a modelului, dar doar până la un anumit punct, după care rezultatele stagnează sau se poate observa *overfitting*-ul.
- Adâncimea arborelui (`max_depth`): reprezintă adâncimea maximă a fiecărui arbore din pădure.
- Numărul minim de exemple într-un nod (`min_samples_split`): Acesta reprezintă numărul minim de exemple necesare într-un nod pentru a permite bifurcarea (`split`) acestuia.
- Numărul minim de exemple într-o frunză (`min_samples_leaf`): Acest parametru specifică numărul minim de exemple necesare într-o frunză a arborelui.
- Numărul maxim de caracteristici (`max_features`): Acest parametru specifică numărul maxim de caracteristici luate în considerare pentru separare la fiecare nod.

## 3.4 Avantaje

**Precizie și performanță:** Random Forest este cunoscut pentru precizia sa în clasificare și regresie. Prin combinarea mai multor arbori de decizie, fiecare cu propriile sale previziuni, modelul poate reduce erorile arborilor, obținând astfel o precizie mai mare.

**Reducerea overfitting-ului:** Random Forest are capacitatea de a controla overfitting-ul, deoarece utilizează tehnicile de agregare și alegeri aleatorii ale caracteristicilor. Prin construirea mai multor arbori independenți, modelul poate generaliza mai bine și reduce riscul de overfitting.

**Gestionarea caracteristicilor:** Random Forest poate evalua importanța caracteristicilor în procesul de învățare automată. Utilizând importanța caracteristicilor, puteți identifica caracteristicile relevante și elimina cele redundante sau irelevante, ceea ce poate duce la îmbunătățirea performanței modelului.

**Scalabilitate și eficiență:** Random Forest poate fi antrenat și folosit eficient pe seturi de date mari.

**Flexibilitate:** Random Forest poate fi utilizat pentru atât probleme de clasificare, cât și de regresie. Este capabil să gestioneze atât clase de date, cât și valori continue, făcându-l versatil în diferite domenii și aplicații.

### 3.4.1 Dezavantaje

**Dificultatea interpretării:** Arborele de decizie individual din cadrul Random Forest poate fi relativ ușor de interpretat, dar când vine vorba de un ansamblu de arbori, interpretarea rezultatelor devine mai dificilă.

**Consumul de resurse:** Random Forest implică construirea și antrenarea mai multor arbori de decizie în paralel. Acest lucru necesită o cantitate semnificativă de resurse și de memorie, mai ales în cazul seturilor de date mari sau complexe. Procesul de predicție poate fi, de asemenea, mai lent decât în cazul unui singur arbore de decizie.

### 3.4.2 Domenii de aplicabilitate

**Clasificare medicală:** poate fi utilizat pentru a clasifica pacienții în funcție de simptome, analize de laborator sau imagistică medicală, ajutând la diagnosticul bolilor sau a afecțiunilor medicale.

**Analiza imaginilor și recunoașterea obiectelor:** poate fi aplicat în recunoașterea obiectelor, clasificarea imaginilor sau segmentarea acestora în diferite regiuni.

**Detectarea fraudei:** poate fi folosit pentru a identifica modele și anomalii în tranzacțiile financiare, ajutând la detectarea fraudelor bancare sau comerciale.

Prezicerea comportamentului consumatorilor: poate fi utilizat pentru a prezice preferințele și comportamentul consumatorilor, ajutând astfel la strategiile de marketing personalizate și la luarea deciziilor comerciale.

Sisteme de recomandare: poate fi aplicat în dezvoltarea de sisteme de recomandare personalizate, utilizând caracteristicile și istoricul utilizatorilor pentru a oferi sugestii și recomandări relevante.

Analiza de credit și evaluarea riscului: poate fi aplicat în evaluarea riscului în instituțiile financiare, inclusiv acordarea de credite și estimarea scorului de creditare pentru clienți.

# Capitolul 4

## Studiu de caz

### 4.1 Baza de date, preprocesare și normalizare

Datele folosite au fost prelevate de la 86 de pacienți. Fiecare pacient a fost evaluat preoperator, fiind adăugate analizele obținute din materialele prelevate în timpul operației, precum și analizele efectuate la controlul de 6 luni, un an și un an și jumătate (pentru pacienții care au efectuat deja cel de-al treilea control). Pentru că baza de date pentru 18 luni nu este completă, nu s-au luat în considerare datele respective, lucrându-se doar cu rezultatele inițiale, de la 6 luni și de la 12 luni. La fiecare din cele trei controale s-au adăugat rezultatele analizelor POSE, SNOT și HPQ-9.

Baza de date a fost analizată pentru a urmări distribuția procentuală a datelor. În figura 4.1 este prezentată distribuția pacienților în funcție de cele patru caracteristici. În tabelul 4.1 sunt descrise caracteristicile pentru pacienții din baza de date.

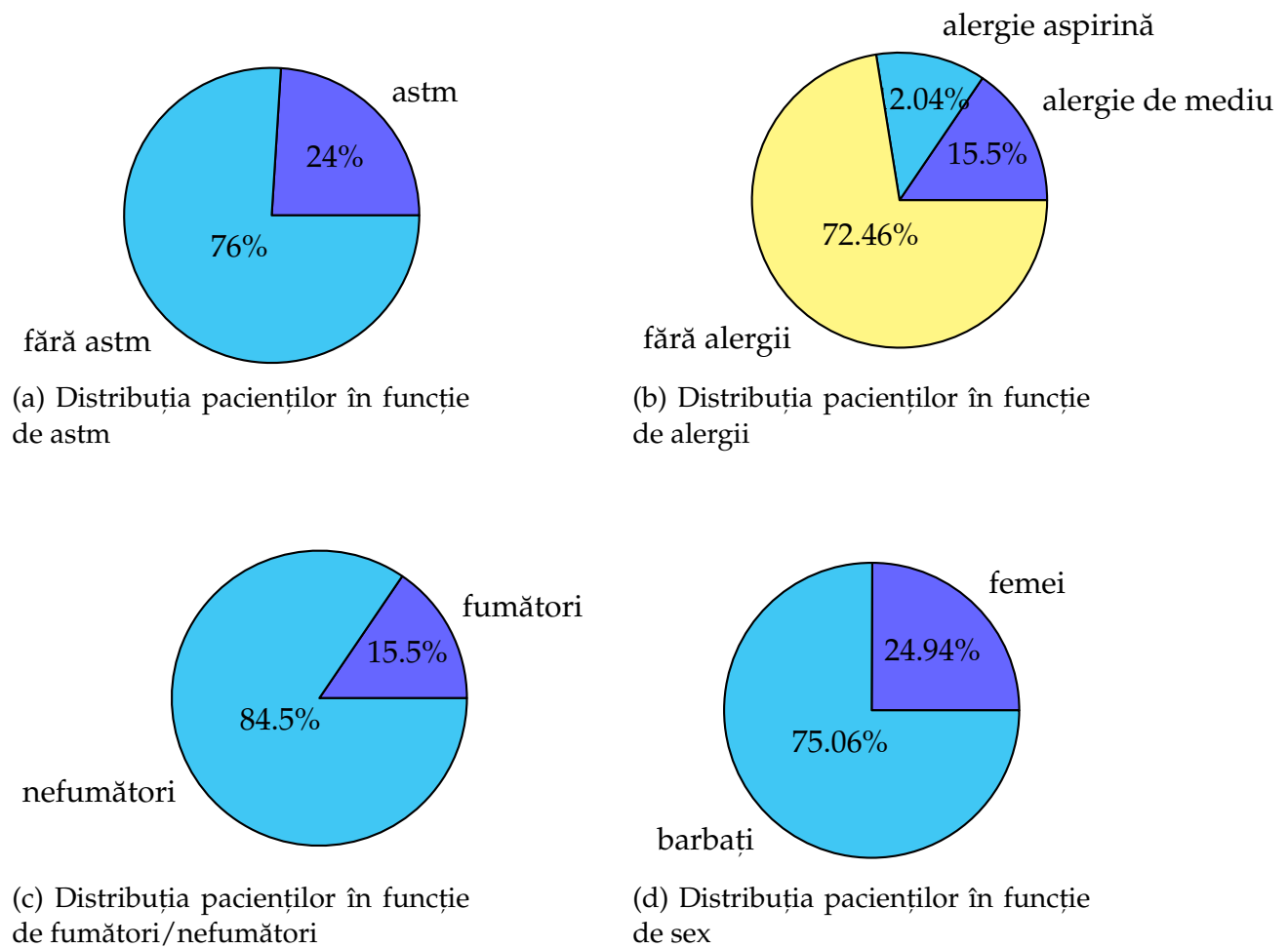


Figura 4.1: Distribuția pacienților în baza de date, după cele patru caracteristici



nume	valoare	descriere
<b>vârstă</b>	număr întreg	vârsta pacientului
<b>data operației</b>	data	data operației
<b>marker astm</b>	0 sau 1	Prezența unui astm bronșic (1-da, 0-nu)- astmul bronșic se asociază frecvent cu polipoza nazală și determină de cele mai multe ori afecțiuni mai agresive
<b>alergie la aspirină</b>	0, 1 sau 2	Prezența unei alergii de mediu(1) sau a alergiei/intoleranței la aspirină (2)- asocierea dintre polipoză, astm și intoleranța la aspirină determina subtipul cel mai agresiv de CRSwNP
<b>eoziinofilele (EO)</b>	număr real (procent)	o valoare crescută a acestora a fost, de asemenea, asociată unei forme mai agresive de boală. A fost demonstrată o boală mai agresivă, cu refacerea rapidă a polipilor după intervenția chirurgicală, la pacienții cu hipereozinofilie
<b>fumător</b>	0 sau 1	0-nu, 1-da
<b>sex</b>	F sau M	
<b>SNOT</b>	număr întreg (0-100)	scor privind calitatea vieții pacientului cu polipoză nazală, efectuat de pacient înaintea intervenției chirurgicale, care se corelează cu statusul bolii 0-10 - lipsa unei probleme 11-40 - problemă moderată 41-69 - moderat spre sever 70-100 sever spre foarte grav

Continuă pe următoarea pagină

<b>HPQ-9</b>	număr întreg (0-27)	preoperator a fost efectuat acest chestionar HPQ-9 pentru evaluarea stării psihice a pacientului și a gradului de depresie, cu rezultate interpretabile astfel: 0-4 fără depresie 5-9 depresie ușoară 10-14 depresie moderată 15-19 depresie moderat-severă 20-27 depresie severă
<b>Lund-Mackay</b>	număr întreg (0-12);	reprezintă un scor de evaluare a polipozei nazale pe examinare CT- între 0 și 2 fiecărui sinus nazal (absența (0), prezența parțială(1) sau ocuparea în totalitate(2) a sinusurilor de către polipii nazali) Scorul maxim este de 12 pentru fiecare fosă nazală
<b>scor endoscopic</b>	număr întreg (0-3)	acest scor evaluează dimensiunea polipilor în fosa nazală cu ajutorul unei camere video, 0- fără polipi nazali 1- polipi mici 2- polipi de dimensiuni intermediare 3- polipi care ocupă toată fosa nazală este evaluată fiecare fosă nazală
<b>miR 125</b>	float	microARN determinat din polipii obținuți de la pacient în timpul intervenției chirurgicale
<b>miR 203</b>	float	microARN determinat din polipii obținuți de la pacient în timpul intervenției chirurgicale
<b>age marker</b>	bool	parametru introdus ulterior, adevărat dacă pacientul are peste 64 de ani, altfel fals

Tabela 4.1: Descrierea caracteristicilor pentru baza de date

S-a urmărit și nivelul de corelare dintre date, pe baza matricei de corelație.

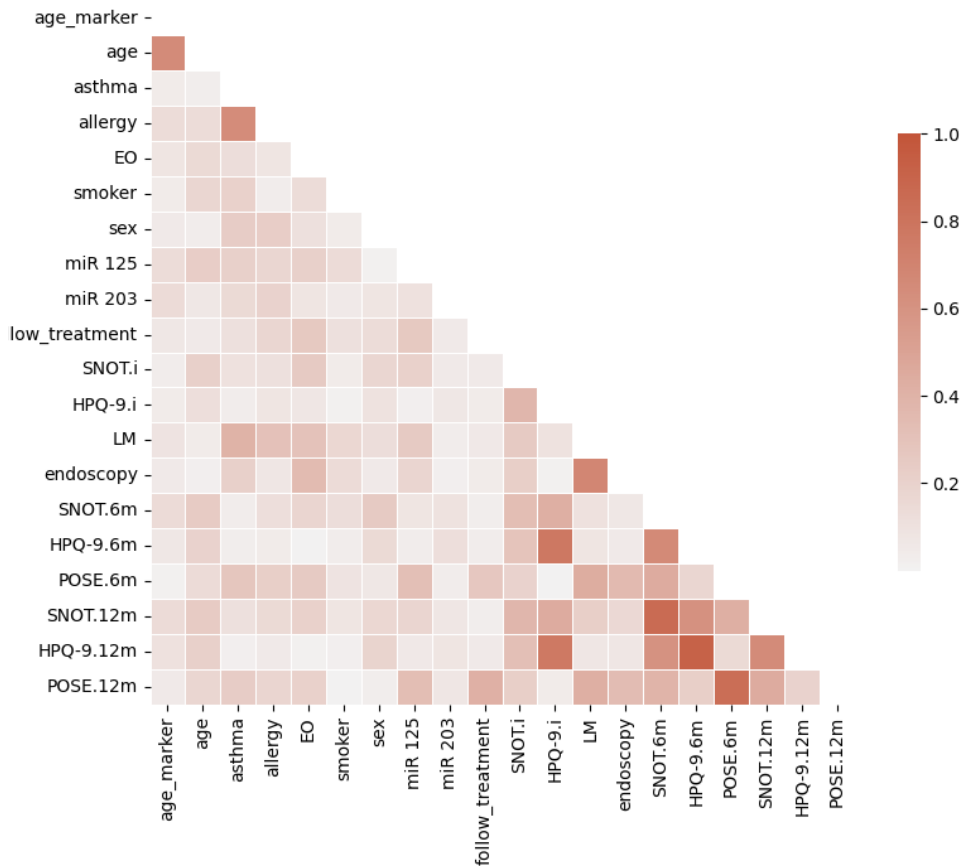


Figura 4.2: Matrice de corelație date inițiale

Se poate observa corelarea dintre HPQ-9 și SNOT pentru fiecare perioadă de observație, evidențiindu-se legătura dintre acestea (dintre scorul care măsoară nivelul depresiei și scorul care măsoară calitatea vieții). Totodată, se poate observa și corelația în timp a scorurilor SNOT și HPQ-9, concluzionând că îmbunătățirea calității vieții necesită timp după operație. Scorul HPQ-9 nu a mai fost luat în considerare pentru ultimele 2 experimente.

Totodată, este evidentă corelarea vârstei cu markerul de vârstă, marker introdus artificial pentru experimente.

Un alt scor care se corelează în timp este POSE, evidențiind faptul că evoluția bolii este constantă în timp, iar dacă un pacient suferă de o boală agresivă la controlul de 6 luni, este puțin probabil ca aceasta să se amelioreze la controlul de 12 luni. Totodată, corelarea în timp a valorilor face posibilă prezicerea scorului POSE de la

12 luni doar din datele inițiale.

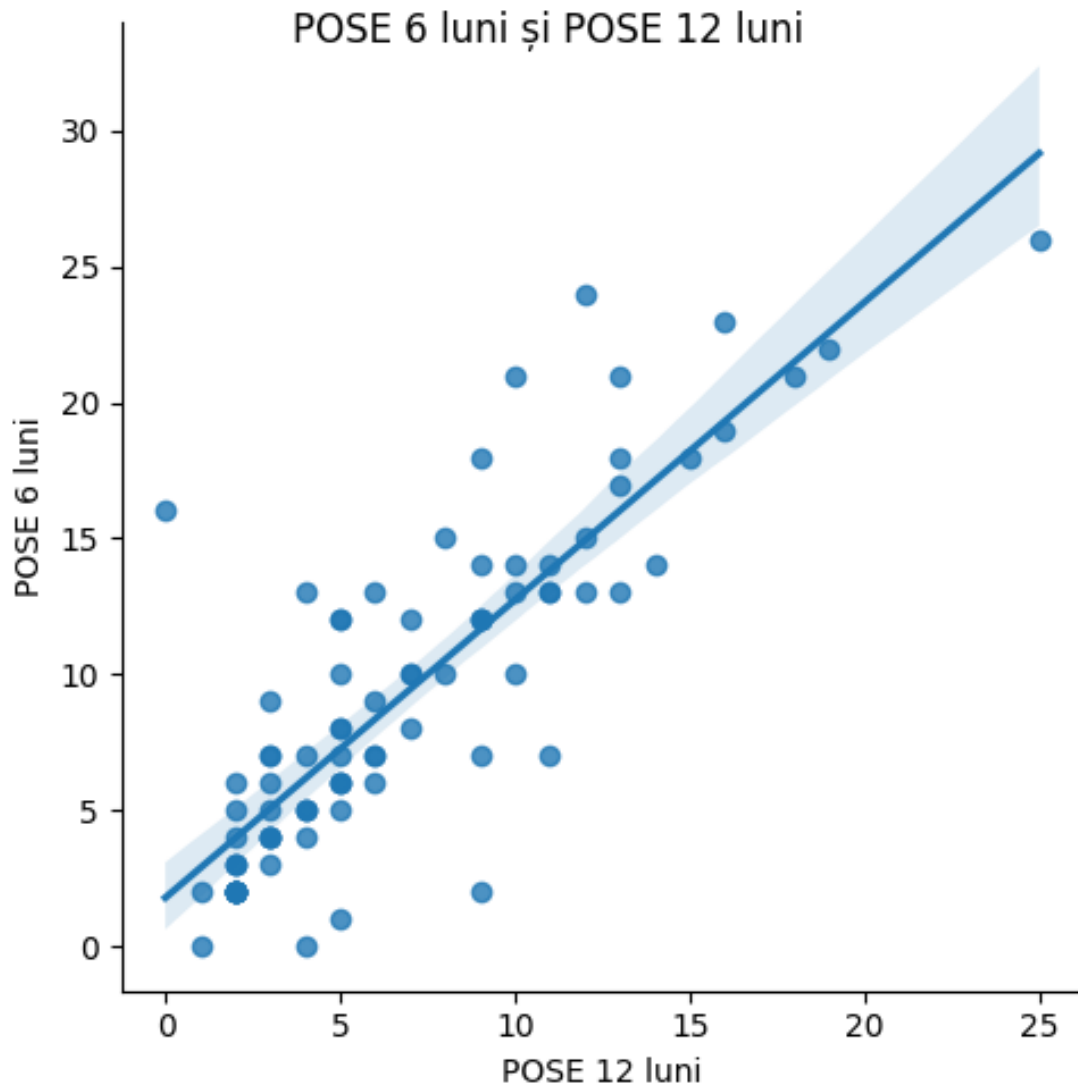


Figura 4.3: Corelarea dintre POSE 6 luni și POSE 12 luni

Datele inițiale au fost preprocesate, făcându-se conversia la valori numerice, astfel:

- pentru datele de tipul adevarat/fals și feminin/masculin s-au primit valori booleene (adevarat, feminin - 1; fals/masculin - 0)
- s-a introdus parametrul age\_marker, astfel: 1 pentru pacienții de peste 64 de ani, 0 pentru cei sub 64 de ani
- datele au fost separate pe fose nazale (stânga, dreapta)
- valorile eozinofilelor au fost transpuse în 0, 1 (0 dacă ambele valori sunt în parametri normali, 1 dacă nu)

## 4.2 Validare experimentală

### 4.2.1 Experiment I

În primul experiment s-a folosit regresia, fără a se adăuga markerul de vârstă. Fiecare fosă nazală a fost tratată separat, astfel încât și rezultatele sunt exprimate pentru fiecare fosă nazală. Totodată, analizele miR 125 și miR 203 nu au fost incluse.

Selecția pacienților pentru setul de antrenament și pentru setul de testare a fost făcută aleator (26 de pacienți pentru testare și 60 pentru antrenare).

Pentru acest experiment s-a exclus, ca și caracteristică a datelor (feature), data operației, pentru că nu aduce valoare informațională din punct de vedere medical. Apoi s-a încercat excluderea, pe rând, a celor 2 caracteristici ale datelor care nu influențează, din punct de vedere medical, evoluția bolii, în vederea obținerii unui model viabil. Ca și parametri ai arborelui de decizie, s-au folosit `min_samples_leaf = 4` și `max_depth = 7`. Rezultatele acestor experimente, conform metricilor urmărite, sunt redată în tabela 4.2

S-au obținut așadar 4 modele care au fost aplicate pe setul de test. În tabela 4.3 se pot observa rezultatele prezise.

Cavitate Nazală	Perioada	Date excluse	$MSE$	$MAE$	$R^2$
stânga	6 luni	-	<b>0.0243</b>	<b>0.1217</b>	<b>-0.2373</b>
		vârsta	0.0254	0.1262	-0.2964
		sex	0.0243	0.1217	-0.2373
		vârsta, sex	0.0254	0.1262	-0.2964
	12 luni	-	0.0297	0.1337	0.1903
		<b>vârsta</b>	<b>0.0147</b>	<b>0.0921</b>	<b>0.5988</b>
		sex	0.0224	0.1131	0.5742
		vârsta, sex	0.0154	0.0921	0.5743
dreapta	6 luni	-	<b>0.0271</b>	<b>0.1215</b>	<b>0.1919</b>
		vârsta	0.0281	0.1265	0.1769
		sex	0.0271	0.1215	0.1919
		vârsta, sex	0.0281	0.1265	0.1769
	12 luni	-	0.0217	0.1098	0.4761
		vârsta	0.0205	0.1022	0.4956
		<b>sex</b>	<b>0.0203</b>	<b>0.0992</b>	<b>0.5061</b>
		vârsta, sex	0.0207	0.1082	0.4906

Tabela 4.2: Rezultate experiment 1

#	6 luni								12 luni							
	stânga				dreapta				stânga				dreapta			
	O	N	PN	P	O	N	PN	P	O	N	PN	P	O	N	PN	P
1	2	0.125	0.1339	2	2	0.125	0.1125	2	0	0.0	0.1875	3	0	0.0	0.0625	1
8	2	0.125	0.4125	7	2	0.125	0.2916	5	7	0.4375	0.2291	4	6	0.375	0.0	0
13	1	0.0625	0.4125	7	1	0.0625	0.2916	5	1	0.0625	0.1406	2	1	0.0625	0.0625	1
14	4	0.25	0.2152	3	7	0.4375	0.5312	8	3	0.1875	0.21875	3	4	0.25	0.375	6
15	2	0.125	0.0937	1	4	0.25	0.2812	4	8	0.5	0.2291	4	5	0.3125	0.1812	3
16	3	0.1875	0.0892	1	3	0.1875	0.2321	3	4	0.25	0.125	2	3	0.1875	0.1812	3
26	1	0.0625	0.1339	2	1	0.0625	0.0625	1	3	0.1875	0.0625	1	3	0.1875	0.0625	1
31	6	0.375	0.2083	3	7	0.4375	0.2321	4	9	0.5625	0.4375	7	9	0.5625	0.5625	9
37	1	0.0625	0.2152	3	1	0.0625	0.2375	4	1	0.0625	0.1406	2	2	0.125	0.1812	3
41	2	0.125	0.0937	2	1	0.0625	0.2321	4	2	0.125	0.1718	3	2	0.125	0.0892	1
44	5	0.3125	0.0892	1	6	0.375	0.2321	3	7	0.4375	0.4375	7	7	0.4375	0.375	6
53	4	0.25	0.4125	7	5	0.3125	0.4218	7	6	0.375	0.5	8	6	0.375	0.7031	11
70	3	0.1875	0.4125	7	4	0.25	0.2375	4	5	0.3125	0.2291	4	7	0.4375	0.0	0
71	1	0.0625	0.2152	3	1	0.0625	0.1517	2	2	0.125	0.0625	1	1	0.0625	0.0625	1
73	1	0.0625	0.0937	1	1	0.0625	0.15	2	1	0.0625	0.1406	2	1	0.0625	0.1812	3
80	11	0.6875	0.4125	7	14	0.875	0.2916	5	12	0.4375	0.6666	11	14	0.875	0.7031	11
81	3	0.1875	0.0937	2	5	0.3125	0.15	2	7	0.75	0.3125	5	8	0.5	0.375	6
83	2	0.125	0.0937	2	1	0.0625	0.0625	1	3	0.1875	0.1718	3	1	0.0625	0.0892	1
84	2	0.125	0.1875	3	2	0.125	0.1875	3	2	0.125	0.125	2	3	0.1875	0.1812	3
85	3	0.1875	0.1339	2	3	0.1875	0.2321	4	4	0.25	0.1718	3	3	0.1875	0.0892	1
86	1	0.0625	0.2083	3	1	0.0625	0.2321	4	1	0.0625	0.125	2	1	0.0625	0.1812	3
89	1	0.0625	0.0892	1	1	0.0625	0.1517	2	2	0.125	0.125	2	3	0.1875	0.0892	1
92	1	0.0625	0.0937	1	1	0.0625	0.15	2	1	0.0625	0.1406	2	1	0.0625	0.1812	3
94	6	0.375	0.0937	2	6	0.375	0.2375	4	8	0.5	0.4375	7	7	0.4375	0.4375	7
97	5	0.3125	0.2083	3	4	0.25	0.0625	1	1	0.0625	0.4375	7	1	0.0625	0.0892	1
100	2	0.125	0.0937	2	1	0.0625	0.125	2	4	0.25	0.3125	5	5	0.3125	0.1812	3

Tabela 4.3: Rezultate experiment 1 - valorile POSE prezise pentru setul de test, O – valori originale, N – valori normalizate, PN – valorile normalizate ale predicției și P – valori prezise

### 4.2.2 Experiment II

În cel de-al doilea experiment, caracteristicile pentru cele două fose nazale au fost însumate. Totodată, s-a adăugat markerul de vârstă, care indică dacă pacientul are sau nu peste 64 de ani.

Au fost luate în considerare și analizele miR 125 și miR 203, iar împărțirea pacienților în seturi de antrenament și test nu a mai fost făcută aleator, ci s-a ținut cont de distribuția procentuală a acestora în baza de date inițială.

Ca și parametri, s-au păstrat cei din primul experiment, și anume `min.samples_leaf = 4` și `max_depth = 7`.

La fel ca și la primul experiment, data operației a fost exclusă, iar apoi s-a încercat eliminarea, pe rand, a caracteristicilor datelor (features) care nu aduc valoare informațională. Modelele obținute au fost aplicate atât pe setul de test, cât și pe setul de antrenament.

În tabelul 4.4 și tabelul 4.5 se pot observa rezultatele experimentelor pentru 6 luni și 12 luni.



Perioada	Date Exclude	MSE	MAE	$R^2$
6 luni	vârsta, marker_vârsta, sex, miR 125, miR 203	0.0250	0.1310	-0.0017
	vârsta, marker_vârsta, sex, miR 125	0.0249	0.1292	0.0050
	vârsta, marker_vârsta, sex, miR 203	0.0267	0.1413	-0.0696
	vârsta, marker_vârsta, sex	0.0258	0.1341	-0.0335
	vârsta, marker_vârsta, miR 125, miR 203	0.0252	0.1322	-0.0070
	vârsta, marker_vârsta, miR 125	0.0258	0.1315	-0.0341
	vârsta, marker_vârsta, miR 203	0.0269	0.1419	-0.0771
	vârsta, marker_vârsta	0.0269	0.1419	-0.0784
	vârsta, sex, miR 125, miR 203	0.0250	0.1310	-0.0017
	vârsta, sex, miR 125	0.0249	0.1292	0.0050
	vârsta, sex, miR 203	0.0260	0.1347	-0.0410
	vârsta, sex	0.0269	0.1419	-0.0771
	vârsta, miR 125, miR 203	0.0252	0.1322	-0.0070
	vârsta, miR 125	0.0258	0.1315	-0.0341
	vârsta, miR 203	0.0267	0.1413	-0.0696
	vârsta	0.0271	0.1428	-0.0842
	marker_vârsta, sex, miR 125, miR 203	0.0249	0.1275	0.0022
	marker_vârsta, sex, miR 125	0.0215	0.1178	0.1410
	marker_vârsta, sex, miR 203	0.0269	0.1398	-0.0780
	marker_vârsta, sex	0.0278	0.1458	-0.1137
	marker_vârsta, miR 125, miR 203	0.0249	0.1275	0.0022
	marker_vârsta, miR 125	0.0212	0.1161	0.1503
	marker_vârsta, miR 203	0.0269	0.1398	-0.0780
	marker_vârsta	0.0279	0.1461	-0.1158
	sex, miR 125, miR 203	0.0249	0.1275	0.0022

	<b>sex, miR 125</b>	<b>0.0212</b>	<b>0.1161</b>	<b>0.1503</b>
	sex, miR 203	0.0267	0.1392	-0.0705
	sex	0.0279	0.1458	-0.1150
	miR 125, miR 203	0.0249	0.1275	0.0022
	miR 125	0.0215	0.1178	0.1410
	miR 203	0.0284	0.1479	-0.1380
	-	0.0279	0.1458	-0.1150

Tabela 4.4: Rezultate experiment 2 pentru setul de test - 6 luni

Perioada	Date Exclude	MSE	MAE	$R^2$
12 luni	vârstă, marker_vârstă, sex, miR 125, miR 203	0.0157	0.0946	0.5602
	vârstă, marker_vârstă, sex, miR 125	0.0155	0.0937	0.5640
	vârstă, marker_vârstă, sex, miR 203	0.0156	0.0956	0.5625
	vârstă, marker_vârstă, sex	0.0156	0.0956	0.5625
	vârstă, marker_vârstă, miR 125, miR 203	0.0157	0.0955	0.5596
	vârstă, marker_vârstă, miR 125	0.0157	0.0955	0.5596
	vârstă, marker_vârstă, miR 203	0.0155	0.0947	0.5631
	vârstă, marker_vârstă	0.0156	0.0956	0.5625
	vârstă, sex, miR 125, miR 203	0.0157	0.0955	0.5596
	vârstă, sex, miR 125	0.0157	0.0955	0.5596
	vârstă, sex, miR 203	0.0156	0.0956	0.5625
	vârstă, sex	0.0154	0.0938	0.5669
	vârstă, miR 125, miR 203	0.0157	0.0946	0.5602
	vârstă, miR 125	0.0155	0.0937	0.5640
	vârstă, miR 203	0.0155	0.0947	0.5631
	vârstă	0.0156	0.0956	0.5625

marker_vârstă, sex, miR 125, miR 203	0.0183	0.1011	0.4869
marker_vârstă, sex, miR 125	0.0170	0.0968	0.5229
marker_vârstă, sex, miR 203	0.0153	0.0925	0.5691
marker_vârstă, sex	0.0160	0.0972	0.5492
marker_vârstă, miR 125, miR 203	0.0179	0.0984	0.4960
marker_vârstă, miR 125	0.0184	0.1023	0.4834
marker_vârstă, miR 203	0.0160	0.0972	0.5492
marker_vârstă	0.0153	0.0925	0.5691
sex, miR 125, miR 203	0.0179	0.0984	0.4960
sex, miR 125	0.0170	0.0959	0.5235
sex, miR 203	0.0154	0.0934	0.5685
<b>sex</b>	<b>0.0152</b>	<b>0.0916</b>	<b>0.5729</b>
miR 125, miR 203	0.0180	0.0993	0.4954
miR 125	0.0170	0.0959	0.5235
miR 203	0.0160	0.0972	0.5492
-	0.0154	0.0934	0.5685

Tabela 4.5: Rezultate experiment 2 pentru setul de test - 12 luni

#	6 luni				12 luni			
	O	N	PN	P	O	N	PN	P
1	4	0.1250	0.1161	2	0	0.0000	0.0885	2
8	4	0.1250	0.1178	2	12	0.4062	0.2969	8
9	2	0.0938	0.2578	8	2	0.0938	0.0885	2
13	2	0.0625	0.1178	2	2	0.0625	0.0625	2
15	6	0.1875	0.2500	8	12	0.4062	0.3672	10
18	12	0.4062	0.2734	8	16	0.5312	0.4922	14
19	2	0.0625	0.0677	2	2	0.0938	0.1187	2
20	12	0.3750	0.2578	8	12	0.4062	0.4922	14
22	2	0.0625	0.1625	4	2	0.0625	0.0625	2
26	2	0.0625	0.1161	2	6	0.1875	0.1187	2
28	8	0.2812	0.2500	8	18	0.5625	0.4531	14
29	4	0.1562	0.1178	2	12	0.3750	0.2969	8
30	4	0.1562	0.1161	2	10	0.3125	0.2109	6
31	12	0.4062	0.2969	8	18	0.5625	0.4922	14
34	8	0.2500	0.2344	6	10	0.3125	0.2109	6
35	10	0.3438	0.2344	6	12	0.4062	0.4531	14
37	2	0.0625	0.1178	2	2	0.0938	0.1187	2
41	2	0.0938	0.2969	8	4	0.1250	0.0885	2
42	8	0.2812	0.1625	4	14	0.4375	0.4531	14
44	10	0.3438	0.2578	8	14	0.4375	0.3672	10
45	6	0.2188	0.2344	6	8	0.2500	0.2109	6
47	4	0.1250	0.1178	2	4	0.1250	0.2000	6
49	4	0.1562	0.4323	12	8	0.2500	0.2109	6
56	6	0.2188	0.2578	8	10	0.3125	0.3672	10

57	4	0.1562	0.1625	4	0	0.0312	0.0885	2
59	14	0.4688	0.4323	12	18	0.5625	0.6510	20
63	2	0.0938	0.1161	2	4	0.1562	0.0885	2
64	2	0.0938	0.2344	6	6	0.2188	0.2000	6
65	2	0.0625	0.1178	2	4	0.1250	0.0859	2
66	2	0.0938	0.0677	2	4	0.1250	0.2109	6
69	12	0.4062	0.4323	12	12	0.4062	0.6510	20
70	6	0.2188	0.1178	2	12	0.3750	0.3750	12
71	2	0.0625	0.0677	2	2	0.0938	0.0859	2
72	10	0.3125	0.2734	8	12	0.4062	0.4531	14
73	2	0.0625	0.1178	2	2	0.0625	0.0625	2
74	4	0.1250	0.1178	2	4	0.1562	0.2000	6
75	4	0.1250	0.2344	6	6	0.2188	0.2000	6
79	6	0.2188	0.1625	4	10	0.3125	0.3672	10
80	24	0.7812	0.4323	12	26	0.8125	0.6510	20
81	8	0.2500	0.1178	2	14	0.4688	0.3750	12
82	2	0.0625	0.0677	2	2	0.0625	0.0625	2
83	2	0.0938	0.1625	4	4	0.1250	0.0885	2
84	4	0.1250	0.1161	2	4	0.1562	0.2109	6
85	6	0.1875	0.1161	2	6	0.2188	0.3750	12
86	2	0.0625	0.0677	2	2	0.0625	0.0859	2
87	2	0.0625	0.1178	2	2	0.0625	0.0859	2
88	2	0.0938	0.2734	8	6	0.2188	0.2969	8
89	2	0.0625	0.1161	2	4	0.1562	0.1187	2
91	2	0.0625	0.0677	2	2	0.0625	0.1187	2
92	2	0.0625	0.1178	2	2	0.0625	0.0625	2

94	12	0.3750	0.2500	8	14	0.4688	0.4922	14
95	12	0.3750	0.2344	6	24	0.7500	0.6510	20
97	8	0.2812	0.2734	8	2	0.0625	0.2109	6
98	10	0.3125	0.2500	8	14	0.4375	0.3750	12
99	16	0.5000	0.4323	12	22	0.7188	0.6510	20
100	2	0.0938	0.2500	8	8	0.2812	0.2000	6
101	4	0.1562	0.1178	2	6	0.1875	0.2969	8
103	8	0.2812	0.4323	12	12	0.3750	0.3750	12
104	12	0.4062	0.2969	8	20	0.6562	0.6510	20
108	8	0.2812	0.2969	8	6	0.2188	0.2109	6

Tabela 4.8: Rezultate experiment 2 pentru setul de antrenament, O – valori originale, N – valori normalizate, PN – valorile normalizate ale predicției și P –valori prezise

#	6 luni				12 luni			
	O	N	PN	P	O	N	PN	P
3	4	0.1562	0.4323	12	8	0.2500	0.2109	6
5	0	0.0000	0.1625	4	0	0.0000	0.0625	2
7	4	0.1562	0.4323	12	6	0.1875	0.2109	6
10	2	0.0625	0.1178	2	2	0.0625	0.0625	2
11	4	0.1562	0.1178	2	12	0.3750	0.2969	8
12	0	0.0312	0.0677	2	2	0.0625	0.1187	2
14	10	0.3438	0.2578	8	6	0.2188	0.3672	10
16	6	0.1875	0.1178	2	6	0.2188	0.3750	12
21	0	0.0000	0.2578	8	16	0.5000	0.1187	2
32	8	0.2812	0.2734	8	12	0.3750	0.4531	14
33	2	0.0938	0.0677	2	6	0.1875	0.0885	2
36	18	0.5938	0.4323	12	22	0.6875	0.6510	20
38	10	0.3438	0.1178	2	12	0.4062	0.4531	14
43	4	0.1562	0.1625	4	6	0.2188	0.0885	2
48	6	0.1875	0.2344	6	6	0.1875	0.3750	12
50	4	0.1250	0.2734	8	4	0.1562	0.2000	6
51	14	0.4375	0.2969	8	14	0.4375	0.6510	20
53	8	0.2812	0.2344	6	12	0.3750	0.3750	12
54	4	0.1562	0.1161	2	6	0.1875	0.0885	2
55	4	0.1562	0.1161	2	4	0.1562	0.0885	2
60	18	0.5625	0.2578	8	20	0.6562	0.6510	20
67	16	0.5000	0.4323	12	18	0.5938	0.6510	20
77	2	0.0938	0.2734	8	4	0.1250	0.0885	2
106	6	0.1875	0.2969	8	8	0.2812	0.2109	6
107	10	0.3125	0.4323	12	20	0.6562	0.4531	14
109	10	0.3125	0.4323	12	10	0.3125	0.3750	12

Tabela 4.7: Rezultate experiment 2 pentru setul de test, O – valori originale, N – valori normalizate, PN – valorile normalizate ale predicției și P – valori prezise

### 4.2.3 Experiment III

Pentru că s-a observat că rezultatele folosind Random Forest sunt mult mai bune față de Decision Tree, pentru acest experiment s-au efectuat experimente folosind doar Random Forest.

În urma analizelor matricei de corelație, s-a observat corelarea firească dintre HPQ-9 și SNOT. Așadar, s-a încercat efectuarea experimentelor fără HPQ-9 și fără ambele valori. În timp ce experimentele efectuate de mine, fără ambele valori, au dus la rezultate puțin mai slabe, experimentele efectuate de Adel au dus la concluzia că parametrul HPQ-9 se poate exclude din experimentele viitoare.

În tabela 4.9 și tabela 4.10 se pot observa rezultatele experimentelor, iar tabelele 4.11 și 4.12 prezintă valorile POSE prezise.

6 luni	dropped columns	antrenament.mse	antrenament.r2	antrenament.mae	test.mse	test.r2	test.mae
	HPQ-9, SNOT, vârstă,	0.0066	0.6875	0.0587	0.0184	0.2673	0.1111
	HPQ-9, SNOT, marker_vârstă,	0.0061	0.7142	0.0573	0.0174	0.3044	0.1101
	HPQ-9, SNOT, sex,	0.0061	0.7113	0.0574	0.017	0.3213	0.1067
	HPQ-9, SNOT, miR 125,	0.0082	0.6118	0.0729	0.0165	0.3414	0.0962
	HPQ-9, SNOT, miR 203,	0.0065	0.6959	0.0575	0.0177	0.2936	0.1064
	HPQ-9, SNOT, vârstă, marker_vârstă,	0.0063	0.7031	0.0579	0.0165	0.3412	0.1035
	HPQ-9, SNOT, vârstă, sex,	0.006	0.7163	0.0544	0.0166	0.3383	0.1061
	HPQ-9, SNOT, vârstă, miR 125,	0.0079	0.6258	0.0736	0.0171	0.3164	0.1034
	HPQ-9, SNOT, vârstă, miR 203,	0.0062	0.7072	0.0576	0.0172	0.3144	0.1078
	HPQ-9, SNOT, marker_vârstă, sex,	0.0062	0.7089	0.0568	0.0166	0.3365	0.1015
	HPQ-9, SNOT, marker_vârstă, miR 125,	0.0086	0.5953	0.0744	0.0166	0.3361	0.1024
	HPQ-9, SNOT, marker_vârstă, miR 203,	0.0061	0.7132	0.0541	0.0167	0.3331	0.1037
	HPQ-9, SNOT, sex, miR 125,	0.0082	0.6137	0.0721	0.0162	0.354	0.0988
	HPQ-9, SNOT, sex, miR 203,	0.0058	0.7278	0.0535	0.0174	0.3043	0.1059
	HPQ-9, SNOT, miR 125, miR 203,	0.0084	0.6031	0.0746	0.0161	0.3592	0.0966
	HPQ-9, SNOT, vârstă, marker_vârstă, sex,	0.0064	0.6978	0.0563	0.0189	0.245	0.1132
	HPQ-9, SNOT, vârstă, marker_vârstă, miR 125,	0.0086	0.593	0.0751	0.0175	0.3017	0.1028
	HPQ-9, SNOT, vârstă, marker_vârstă, miR 203,	0.0066	0.6887	0.0567	0.0166	0.3367	0.1032
	HPQ-9, SNOT, vârstă, sex, miR 125,	0.0084	0.6066	0.0739	0.017	0.3199	0.1026
	HPQ-9, SNOT, vârstă, sex, miR 203,	0.0062	0.7081	0.0559	0.0159	0.3656	0.0975
	HPQ-9, SNOT, vârstă, miR 125, miR 203,	0.0091	0.5722	0.0772	0.0177	0.2926	0.1066
	HPQ-9, SNOT, marker_vârstă, sex, miR 125,	0.0082	0.6163	0.0704	0.0161	0.3584	0.098
	HPQ-9, SNOT, marker_vârstă, sex, miR 203,	0.0061	0.7138	0.055	0.0167	0.3349	0.1025
	HPQ-9, SNOT, marker_vârstă, miR 125, miR 203,	0.009	0.5771	0.0769	0.0156	0.3787	0.0963
	HPQ-9, SNOT, sex, miR 125, miR 203,	0.0087	0.5915	0.0757	0.0172	0.3116	0.1025
	HPQ-9, SNOT, vârstă, marker_vârstă, sex, miR 125,	0.0081	0.6198	0.0732	0.0168	0.3279	0.1001
	HPQ-9, SNOT, vârstă, marker_vârstă, sex, miR 203,	0.0064	0.7002	0.0577	0.0161	0.3595	0.1025
	HPQ-9, SNOT, vârstă, marker_vârstă, miR 125, miR 203,	0.0088	0.5876	0.0776	0.018	0.2825	0.1041
	HPQ-9, SNOT, vârstă, sex, miR 125, miR 203,	0.0089	0.5807	0.0766	0.0167	0.332	0.101
	HPQ-9, SNOT, marker_vârstă, sex, miR 125, miR 203,	0.0086	0.5934	0.0743	0.0164	0.3453	0.1001
	HPQ-9, SNOT, vârstă, marker_vârstă, sex, miR 125, miR 203,	0.0092	0.5681	0.0789	0.018	0.2831	0.1044
	HPQ-9, SNOT	0.0062	0.7082	0.0575	0.0163	0.3507	0.104

Tabela 4.9: Rezultate experiment 3 pentru setul de test - RF - 6 luni fără SNOT și HPQ-9



12 luni	dropped columns	antrenament.mse	antrenament.r2	antrenament.mae	test.mse	test.r2	test.mae
	HPQ-9, SNOT, vârstă,	0.0053	0.8648	0.0512	0.0136	0.6184	0.082
	HPQ-9, SNOT, marker_vârstă,	0.0056	0.8569	0.0536	0.0137	0.6175	0.0802
	HPQ-9, SNOT, sex,	0.005	0.8742	0.0519	0.013	0.6353	0.0801
	HPQ-9, SNOT, miR 125,	0.0057	0.8545	0.0536	0.0138	0.6136	0.0844
	HPQ-9, SNOT, miR 203,	0.0056	0.8576	0.0531	0.0128	0.6417	0.0769
	HPQ-9, SNOT, vârstă, marker_vârstă,	0.0062	0.841	0.0541	0.0123	0.6569	0.0735
	HPQ-9, SNOT, vârstă, sex,	0.0061	0.8444	0.0554	0.0131	0.6331	0.0804
	HPQ-9, SNOT, vârstă, miR 125,	0.0052	0.8667	0.0525	0.0153	0.5718	0.0882
	HPQ-9, SNOT, vârstă, miR 203,	0.0055	0.8601	0.0551	0.0126	0.6473	0.0793
	HPQ-9, SNOT, marker_vârstă, sex,	0.0058	0.8525	0.0535	0.0134	0.6254	0.079
	HPQ-9, SNOT, marker_vârstă, miR 125,	0.0056	0.857	0.0542	0.014	0.6074	0.0825
	HPQ-9, SNOT, marker_vârstă, miR 203,	0.0055	0.8608	0.052	0.0132	0.6315	0.0752
	HPQ-9, SNOT, sex, miR 125,	0.0053	0.8656	0.0528	0.0142	0.6026	0.0877
	HPQ-9, SNOT, sex, miR 203,	0.0058	0.8527	0.0528	0.0124	0.6534	0.0774
	HPQ-9, SNOT, miR 125, miR 203,	0.0061	0.8439	0.0545	0.0141	0.6054	0.0835
	HPQ-9, SNOT, vârstă, marker_vârstă, sex,	0.0058	0.853	0.0542	0.0141	0.6065	0.0851
	HPQ-9, SNOT, vârstă, marker_vârstă, miR 125,	0.006	0.8469	0.0551	0.0138	0.6125	0.083
	HPQ-9, SNOT, vârstă, marker_vârstă, miR 203,	0.0055	0.8614	0.0544	0.0131	0.634	0.08
	HPQ-9, SNOT, vârstă, sex, miR 125,	0.0057	0.8542	0.0549	0.0138	0.613	0.0824
	HPQ-9, SNOT, vârstă, sex, miR 203,	0.0057	0.8562	0.0518	0.0129	0.6384	0.0747
	HPQ-9, SNOT, vârstă, miR 125, miR 203,	0.0062	0.8428	0.0561	0.0137	0.6168	0.0805
	HPQ-9, SNOT, marker_vârstă, sex, miR 125,	0.0051	0.8704	0.0524	0.013	0.6353	0.0816
	HPQ-9, SNOT, marker_vârstă, sex, miR 203,	0.0055	0.8613	0.0523	0.0141	0.606	0.0808
	HPQ-9, SNOT, marker_vârstă, miR 125, miR 203,	0.0059	0.8508	0.0559	0.0131	0.6326	0.0789
	HPQ-9, SNOT, sex, miR 125, miR 203,	0.0054	0.8621	0.0534	0.0137	0.6159	0.0827
	HPQ-9, SNOT, vârstă, marker_vârstă, sex, miR 125,	0.0057	0.8544	0.0544	0.0139	0.6107	0.0816
	HPQ-9, SNOT, vârstă, marker_vârstă, sex, miR 203,	0.0056	0.8569	0.0545	0.0125	0.649	0.0757
	HPQ-9, SNOT, vârstă, marker_vârstă, miR 125, miR 203,	0.0059	0.8511	0.0558	0.0136	0.6187	0.0814
	HPQ-9, SNOT, vârstă, sex, miR 125, miR 203,	0.0058	0.8517	0.0543	0.0138	0.6137	0.08
	HPQ-9, SNOT, marker_vârstă, sex, miR 125, miR 203,	0.0059	0.8503	0.0549	0.0139	0.6109	0.0834
	HPQ-9, SNOT, vârstă, marker_vârstă, sex, miR 125, miR 203,	0.0061	0.8435	0.0553	0.0132	0.6317	0.0801
	HPQ-9, SNOT	0.0056	0.8568	0.0552	0.0126	0.6475	0.0796

Tabela 4.10: Rezultate experiment 3 pentru setul de test - RF - 12 luni fără SNOT și HPQ-9

#	6 luni		12 luni	
	O	P	O	P
3	5	10	8	7
5	1	4	0	2
7	5	11	6	7
10	2	4	2	4
11	5	4	12	7
12	1	7	2	4
14	11	6	7	13
16	6	3	7	9
21	0	8	16	4
32	9	8	12	11
33	3	4	6	4
36	19	12	22	17
38	11	9	13	15
43	5	7	7	7
48	6	4	6	7
50	4	5	5	5
51	14	9	14	17
53	9	6	12	13
54	5	3	6	6
55	5	5	5	6
60	18	11	21	19
67	16	10	19	18
77	3	5	4	4
106	6	8	9	9
107	10	13	21	15
109	10	9	10	14

Tabela 4.11: Rezultate experiment 3- valorile POSE prezise pentru setul de test, folosind RF, O – valori originale și P – valori prezise

#	6 luni		12 luni	
	O	P	O	P
1	4	4	0	3
8	4	4	13	8
9	3	6	3	4
13	2	4	2	3
15	6	5	13	9
18	13	7	17	16
19	2	3	3	3
20	12	10	13	15
22	2	3	2	2
26	2	3	6	3
28	9	8	18	13
29	5	7	12	9
30	5	4	10	6
31	13	10	18	17
34	8	7	10	9
35	11	9	13	13
37	2	3	3	3
41	3	4	4	4
42	9	7	14	12
44	11	10	14	12
45	7	7	8	10
47	4	3	4	5
49	5	10	8	7

56	7	6	10	10
57	5	5	1	4
59	15	13	18	19
63	3	4	5	4
64	3	4	7	5
65	2	3	4	3
66	3	5	4	5
69	13	11	13	16
70	7	6	12	11
71	2	4	3	3
72	10	7	13	13
73	2	3	2	3
74	4	4	5	6
75	4	4	7	6
79	7	7	10	9
80	25	13	26	19
81	8	5	15	11
82	2	3	2	2
83	3	4	4	4
84	4	5	5	5
85	6	4	7	8
86	2	3	2	2
87	2	3	2	2
88	3	5	7	5
89	2	3	5	3
91	2	5	2	3

92	2	3	2	2
94	12	9	15	16
95	12	8	24	17
97	9	7	2	8
98	10	6	14	13
99	16	12	23	19
100	3	4	9	5
101	5	4	6	7
103	9	10	12	12
104	13	11	21	18
108	9	7	7	10

Tabela 4.12: Rezultate experiment 3- valorile POSE prezise pentru setul de antrenament, folosind RF, O – valori originale și P – valori prezise

#### 4.2.4 Experiment IV

În cel de-al patrulea experiment, s-a încercat trecerea de la regresie la clasificare. Rezultatele analizei POSE au fost împărțite în 3 categorii:

- pentru o valoare POSE între 0 și 8: 0
- pentru o valoare POSE între 8 și 16: 1
- pentru o valoare POSE între 16 și 32: 2

În prima variantă a acestui experiment, s-a încercat prezicerea clasei din care face parte POSE pentru 6 luni. Apoi, s-a încercat prezicerea POSE-ului la 12 luni, folosind doar datele inițiale (cele folosite și pentru prezicerea POSE 6 luni). În ultima variantă a experimentului s-a încercat prezicerea POSE-ului de 12 luni folosind datele inițiale și POSE-ul de la 6 luni.

Așadar, pentru prezicerea POSE 6 luni avem o acuratețe de 88,33% pentru setul de antrenament și 73,37% pentru setul de test.

		preziceri		
		control	control parțial	recidivă
actual	control	35	2	0
	control parțial	1	18	0
	recidivă	1	3	0

Tabela 4.13: experiment 4 - matrice de confuzie pentru setul de antrenament, pentru a prezice POSE la 6 luni din datele inițiale

		preziceri		
		control	control parțial	recidivă
actual	control	16	2	0
	control parțial	4	3	0
	recidivă	0	1	0

Tabela 4.14: experiment 4 - matrice de confuzie pentru setul de test, pentru a prezice POSE la 6 luni din datele inițiale

caracteristică	scor de importanță
miR125	0.31
LM	0.21
miR203	0.20
SNOT	0.17
follow_treatment	0.06
EO	0.05
age_marker	0.00
asthma	0.00
allergy	0.00
smoker	0.00
sex	0.00
endoscopy	0.00

Tabela 4.15: Importanța caracteristicilor în prezicerea POSE 6 luni, folosind datele inițiale

Pentru prezicerea POSE-ului de 12 luni, cu datele inițiale, avem o acuratețe de 88.33% pentru setul de antrenament și 76.92% pentru setul de test.

		preziceri		
		control	control parțial	recidivă
actual	control	28	0	0
	control parțial	4	17	0
	recidivă	2	1	8

Tabela 4.16: experiment 4 - matrice de confuzie pentru setul de antrenament, pentru a prezice POSE la 12 luni din datele inițiale

		preziceri		
		control	control parțial	recidivă
actual	control	15	1	0
	control parțial	2	4	2
	recidivă	0	1	1

Tabela 4.17: experiment 4 - matrice de confuzie pentru setul de test, pentru a prezice POSE la 12 luni din datele inițiale

caracteristică	scor de importanță
miR125	0.47
endoscopy	0.16
miR203	0.13
follow_treatment	0.09
LM	0.07
allergy	0.05
SNOT	0.03
asthma	0.00
EO	0.00
smoker	0.00
sex	0.00

Tabela 4.18: Importanța caracteristicilor în prezicerea POSE 12 luni, folosind datele inițiale

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
1	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
Continuă pe următoarea pagină										



#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
8	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
13	0	1	0.00	1.00	0.00	0	1	0.00	1.00	0.00
14	1	1	0.25	0.50	0.25	0	0	1.00	0.00	0.00
15	0	0	1.00	0.00	0.00	1	2	0.00	0.00	1.00
16	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
26	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
31	1	1	0.00	1.00	0.00	2	2	0.00	0.00	1.00
37	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
41	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
44	1	0	1.00	0.00	0.00	1	0	1.00	0.00	0.00
53	1	0	0.75	0.00	0.25	1	2	0.00	0.00	1.00
70	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
71	0	0	1.00	0.00	0.00	0	0	0.50	0.00	0.50
73	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
80	2	1	0.00	0.50	0.50	2	1	0.00	1.00	0.00
81	1	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
83	0	0	1.00	0.00	0.00	0	0	0.50	0.50	0.00
84	0	1	0.00	1.00	0.00	0	0	0.33	0.33	0.33
85	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
86	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
89	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
92	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
94	1	1	0.33	0.66	0.00	1	1	0.00	1.00	0.00
97	1	0	1.00	0.00	0.00	0	0	0.50	0.00	0.50

Continuă pe următoarea pagină

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
100	0	0	1.00	0.00	0.00	1	0	1.00	0.00	0.00

Tabela 4.19: valorile POSE prezise pentru setul de test O – valoare originală și P – valoare prezisă

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
3	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
5	0	0	1.00	0.00	0.00	0	0	0.50	0.50	0.00
7	0	0	1.00	0.00	0.00	0	0	0.33	0.33	0.33
9	0	0	1.00	0.00	0.00	0	0	0.50	0.00	0.50
10	0	0	1.00	0.00	0.00	0	0	0.50	0.50	0.00
11	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
12	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
18	1	1	0.00	1.00	0.00	2	2	0.00	0.00	1.00
19	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
20	1	1	0.00	1.00	0.00	1	1	0.00	1.00	0.00
21	0	0	1.00	0.00	0.00	2	2	0.00	0.00	1.00
22	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
28	1	1	0.00	1.00	0.00	2	2	0.00	0.00	1.00
29	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
30	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
32	1	1	0.00	1.00	0.00	1	0	0.33	0.33	0.33
33	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
34	1	1	0.00	1.00	0.00	1	0	0.50	0.50	0.00

Continuă pe următoarea pagină

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
35	1	1	0.33	0.66	0.00	1	1	0.00	1.00	0.00
36	2	0	0.75	0.00	0.25	2	2	0.00	0.00	1.00
38	1	0	0.66	0.33	0.00	1	1	0.00	0.66	0.33
42	1	1	0.25	0.50	0.25	1	1	0.00	1.00	0.00
43	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
45	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
47	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
48	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
49	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
50	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
51	1	1	0.00	1.00	0.00	1	1	0.00	1.00	0.00
54	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
55	0	0	0.75	0.00	0.25	0	0	1.00	0.00	0.00
56	0	0	1.00	0.00	0.00	1	1	0.00	1.00	0.00
57	0	0	0.75	0.00	0.25	0	0	1.00	0.00	0.00
59	1	1	0.00	1.00	0.00	2	0	0.33	0.33	0.33
60	2	1	0.00	0.50	0.50	2	2	0.00	0.00	1.00
63	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
64	0	1	0.33	0.66	0.00	0	0	1.00	0.00	0.00
65	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
66	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
67	2	1	0.25	0.50	0.25	2	0	0.50	0.00	0.50
69	1	1	0.00	0.50	0.50	1	1	0.00	0.66	0.33
72	1	1	0.00	1.00	0.00	1	0	0.50	0.50	0.00

Continuă pe următoarea pagină

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
74	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
75	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
77	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
79	0	0	1.00	0.00	0.00	1	0	0.50	0.50	0.00
82	0	1	0.25	0.50	0.25	0	0	1.00	0.00	0.00
87	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
88	0	0	1.00	0.00	0.00	0	0	1.00	0.00	0.00
91	0	0	0.66	0.33	0.00	0	0	0.50	0.50	0.00
95	1	1	0.00	1.00	0.00	2	2	0.00	0.00	1.00
98	1	1	0.33	0.66	0.00	1	1	0.00	1.00	0.00
99	2	1	0.00	0.50	0.50	2	1	0.00	0.66	0.33
101	0	0	0.75	0.00	0.25	0	0	1.00	0.00	0.00
103	1	1	0.00	1.00	0.00	1	1	0.00	1.00	0.00
104	1	1	0.00	0.50	0.50	2	2	0.00	0.00	1.00
106	0	0	0.66	0.33	0.00	1	1	0.00	1.00	0.00
107	1	1	0.00	1.00	0.00	2	2	0.00	0.00	1.00
108	1	1	0.25	0.50	0.25	0	0	1.00	0.00	0.00
109	1	1	0.00	1.00	0.00	1	1	0.00	1.00	0.00

Tabela 4.20: valorile POSE prezise pentru setul de antrenament O – valoare originală și P – valoare prezisă

Pentru prezicerea POSE-ului de 12 luni, cu datele inițiale și valoarea POSE de la 6 luni, avem o acuratețe de 91.66% pentru setul de antrenament și 76.92% pentru setul de test.

		preziceri		
		control	control parțial	recidivă
actual	control	28	0	0
	control parțial	3	18	0
	recidivă	1	1	9

Tabela 4.21: experiment 4 - matrice de confuzie pentru setul de antrenament, pentru a prezice POSE la 12 luni din datele inițiale și valoarea POSE de la 6 luni

		preziceri		
		control	control parțial	recidivă
actual	control	14	1	1
	control parțial	3	5	0
	recidivă	0	1	1

Tabela 4.22: experiment 4 - matrice de confuzie pentru setul de test, pentru a prezice POSE la 12 luni din datele inițiale și valoarea POSE de la 6 luni

caracteristică	scor de importanță
6months.POSE	0.62
result.miR125	0.10
general.age	0.09
general.sex	0.08
result.miR203	0.07
general.allergy	0.02
0months.SNOT	0.01
general.age_marker	0.00
general.asthma	0.00
general.EO	0.00
general.smoker	0.00
Continuă pe următoarea pagină	

caracteristică	scor de importanță
0months.follow_treatment	0.00
0months.LM	0.00
0months.endoscopy	0.00

Tabela 4.23: Ța caracteristicilor în prezicerea POSE 12 luni, folosind datele inițiale și valoarea POSE de la 6 luni

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
1	0	0	1.00	0.00	0.00
8	1	0	1.00	0.00	0.00
13	0	0	1.00	0.00	0.00
14	0	1	0.00	1.00	0.00
15	1	0	1.00	0.00	0.00
16	0	0	0.75	0.25	0.00
26	0	0	1.00	0.00	0.00
31	2	1	0.00	1.00	0.00
37	0	0	1.00	0.00	0.00
41	0	0	0.50	0.00	0.50
44	1	1	0.00	1.00	0.00
53	1	1	0.00	1.00	0.00
70	1	1	0.00	1.00	0.00
71	0	0	0.75	0.25	0.00
73	0	0	1.00	0.00	0.00
80	2	2	0.00	0.00	1.00

Continuă pe următoarea pagină

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
81	1	1	0.00	1.00	0.00
83	0	0	1.00	0.00	0.00
84	0	0	1.00	0.00	0.00
85	0	0	0.75	0.25	0.00
86	0	0	0.50	0.00	0.50
89	0	0	0.75	0.25	0.00
92	0	0	1.00	0.00	0.00
94	1	1	0.00	0.50	0.50
97	0	2	0.00	0.00	1.00
100	1	0	1.00	0.00	0.00

Tabela 4.24: valorile POSE prezise pentru setul de test O – valoare originală și P – valoare prezisă

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
3	1	1	0.00	1.00	0.00
5	0	0	1.00	0.00	0.00
7	0	0	0.50	0.50	0.00
9	0	0	0.50	0.00	0.50
10	0	0	1.00	0.00	0.00
11	1	0	0.75	0.25	0.00
12	0	0	1.00	0.00	0.00
18	2	1	0.00	0.50	0.50
19	0	0	0.75	0.25	0.00

Continuă pe următoarea pagină

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
20	1	1	0.00	1.00	0.00
21	2	0	0.50	0.00	0.50
22	0	0	1.00	0.00	0.00
28	2	2	0.00	0.00	1.00
29	1	1	0.00	1.00	0.00
30	1	1	0.00	1.00	0.00
32	1	1	0.00	1.00	0.00
33	0	0	1.00	0.00	0.00
34	1	1	0.00	1.00	0.00
35	1	1	0.00	1.00	0.00
36	2	2	0.00	0.00	1.00
38	1	1	0.00	1.00	0.00
42	1	1	0.00	1.00	0.00
43	0	0	1.00	0.00	0.00
45	1	1	0.00	1.00	0.00
47	0	0	1.00	0.00	0.00
48	0	0	1.00	0.00	0.00
49	1	0	0.50	0.50	0.00
50	0	0	1.00	0.00	0.00
51	1	1	0.00	1.00	0.00
54	0	0	1.00	0.00	0.00
55	0	0	0.75	0.25	0.00
56	1	1	0.00	1.00	0.00
57	0	0	1.00	0.00	0.00

Continuă pe următoarea pagină



#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
59	2	2	0.00	0.00	1.00
60	2	2	0.00	0.00	1.00
63	0	0	1.00	0.00	0.00
64	0	0	1.00	0.00	0.00
65	0	0	1.00	0.00	0.00
66	0	0	1.00	0.00	0.00
67	2	2	0.00	0.00	1.00
69	1	0	0.50	0.50	0.00
72	1	1	0.00	1.00	0.00
74	0	0	1.00	0.00	0.00
75	0	0	1.00	0.00	0.00
77	0	0	1.00	0.00	0.00
79	1	1	0.00	1.00	0.00
82	0	0	1.00	0.00	0.00
87	0	0	0.75	0.25	0.00
88	0	0	1.00	0.00	0.00
91	0	0	1.00	0.00	0.00
95	2	2	0.00	0.00	1.00
98	1	1	0.00	1.00	0.00
99	2	2	0.00	0.00	1.00
101	0	0	1.00	0.00	0.00
103	1	1	0.00	0.50	0.50
104	2	2	0.00	0.00	1.00
106	1	1	0.00	1.00	0.00

Continuă pe următoarea pagină

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
107	2	2	0.00	0.00	1.00
108	0	0	0.50	0.50	0.00
109	1	1	0.00	1.00	0.00

Tabela 4.25: valorile POSE prezise pentru setul de antrenament O – valoare originală și P – valoare prezisă

### 4.2.5 Experiment V

În cel de-al cincilea experiment, s-a folosit doar Random Forest, deoarece rezultatele obținute cu acest clasificator au fost mult mai bune. De această dată, s-a încercat prezicerea POSE-urilor excluzând analizele miR 125 și miR 203, fiind analize prelevate din polipii extrași în timpul operațiilor. S-au urmărit, comparativ, cele trei variante de rezultate. Cele mai bune rezultate s-au obținut atunci când nu s-a folosit miR 203, obținând o acuratețe de 91,66% pentru setul de antrenament, și 80,76% pentru setul de test.

		preziceri		
		control	control parțial	recidivă
actual	control	37	0	0
	control parțial	4	15	0
	recidivă	0	1	3

Tabela 4.26: experiment 5 - matrice de confuzie pentru setul de antrenament, pentru a prezice POSE la 6 luni din datele inițiale, fără miR 203

		preziceri		
		control	control parțial	recidivă
actual	control	18	0	0
	control parțial	4	3	0
	recidivă	0	1	0

Tabela 4.27: experiment 5 - matrice de confuzie pentru setul de test, pentru a prezice POSE la 6 luni din datele inițiale, fără miR 203

caracteristică	scor de importanță
0months.LM	0.20
result.miR125	0.19
general.age	0.14
0months.SNOT	0.13
Continuă pe următoarea pagină	

caracteristică	scor de importanță
0months.followtreatment	0.08
general.allergy	0.06
0months.endoscopy	0.05
general.EO	0.04
general.sex	0.03
general.asthma	0.02
general.agemarker	0.02
general.smoker	0.02

Tabela 4.28: Importanța caracteristicilor în prezicerea POSE 6 luni, folosind datele inițiale, mai puțin miR 203

Pentru prezicerea de la 12 luni cu datele inițiale, dar fără valoarea miR 203, am obținut 95% acuratețe pentru setul de antrenament și 73.07% pentru setul de test.

		preziceri		
		control	control parțial	recidivă
actual	control	28	0	0
	control parțial	1	20	0
	recidivă	1	1	9

Tabela 4.29: experiment 5 - matrice de confuzie pentru setul de antrenament, pentru a prezice POSE la 12 luni din datele inițiale, fără miR 203

		preziceri		
		control	control parțial	recidivă
actual	control	15	1	0
	control parțial	4	4	0
	recidivă	0	2	0

Tabela 4.30: experiment 5 - matrice de confuzie pentru setul de test, pentru a prezice POSE la 12 luni din datele inițiale, fără miR 203

caracteristică	scor de importanță
result.miR125	0.23
0months.SNOT	0.17
0months.endoscopy	0.14
0months.LM	0.13
general.age	0.11
0months.followtreatment	0.06
general.allergy	0.05
general.asthma	0.03
general.EO	0.03
general.smoker	0.02
general.sex	0.02
general.agemarker	0.01

Tabela 4.31: Importanța caracteristicilor în prezicerea POSE 12 luni, folosind datele inițiale, mai puțin miR 203

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
1	0	0	0.69	0.29	0.02	0	0	0.57	0.37	0.06
8	0	0	0.69	0.27	0.04	1	1	0.37	0.50	0.13
13	0	0	0.59	0.37	0.04	0	1	0.31	0.45	0.23
14	1	1	0.41	0.41	0.18	0	0	0.40	0.31	0.29
15	0	0	0.72	0.26	0.02	1	0	0.54	0.31	0.16
16	0	0	0.58	0.40	0.02	0	0	0.49	0.35	0.16

Continuă pe următoarea pagină

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
26	0	0	0.72	0.28	0.00	0	0	0.71	0.23	0.06
31	1	1	0.45	0.51	0.04	2	1	0.32	0.41	0.27
37	0	0	0.74	0.26	0.00	0	0	0.59	0.35	0.06
41	0	0	0.69	0.25	0.07	0	0	0.71	0.19	0.10
44	1	0	0.64	0.30	0.06	1	1	0.30	0.41	0.29
53	1	0	0.57	0.27	0.16	1	0	0.43	0.30	0.27
70	0	0	0.70	0.26	0.04	1	1	0.38	0.49	0.13
71	0	0	0.64	0.34	0.02	0	0	0.46	0.42	0.12
73	0	0	0.83	0.17	0.00	0	0	0.82	0.11	0.07
80	2	1	0.26	0.57	0.18	2	1	0.14	0.47	0.39
81	1	0	0.64	0.34	0.02	1	0	0.65	0.29	0.06
83	0	0	0.83	0.16	0.01	0	0	0.76	0.23	0.02
84	0	0	0.51	0.46	0.03	0	0	0.45	0.40	0.15
85	0	0	0.91	0.09	0.00	0	0	0.71	0.25	0.03
86	0	0	0.80	0.15	0.04	0	0	0.69	0.21	0.10
89	0	0	0.77	0.17	0.06	0	0	0.61	0.25	0.14
92	0	0	0.59	0.41	0.00	0	0	0.73	0.19	0.09
94	1	1	0.45	0.55	0.00	1	1	0.47	0.49	0.05
97	1	0	0.79	0.21	0.01	0	0	0.55	0.42	0.03
100	0	0	0.79	0.20	0.01	1	0	0.80	0.19	0.01

Tabela 4.32: experiment 5 - valorile POSE prezise din datele inițiale, fără miR 203, pentru setul de test. O – valoare originală și P – valoare prezisă

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
3	0	0	0.65	0.27	0.08	1	1	0.19	0.69	0.12
5	0	0	0.83	0.17	0.01	0	0	0.76	0.24	0.00
7	0	0	0.57	0.34	0.09	0	0	0.52	0.30	0.18
9	0	0	0.79	0.17	0.04	0	0	0.61	0.20	0.19
10	0	0	0.83	0.17	0.00	0	0	0.75	0.20	0.05
11	0	0	0.90	0.08	0.02	1	1	0.31	0.54	0.15
12	0	0	0.82	0.15	0.04	0	0	0.59	0.34	0.08
18	1	1	0.44	0.54	0.02	2	0	0.40	0.31	0.28
19	0	0	0.88	0.09	0.03	0	0	0.58	0.33	0.09
20	1	1	0.39	0.58	0.03	1	1	0.31	0.52	0.17
21	0	0	0.70	0.18	0.12	2	2	0.24	0.26	0.50
22	0	0	0.90	0.09	0.01	0	0	0.77	0.20	0.03
28	1	1	0.43	0.56	0.01	2	1	0.22	0.47	0.30
29	0	0	0.60	0.38	0.02	1	1	0.16	0.68	0.16
30	0	0	0.89	0.11	0.00	1	1	0.43	0.50	0.07
32	1	1	0.33	0.59	0.08	1	1	0.30	0.55	0.15
33	0	0	0.81	0.12	0.06	0	0	0.71	0.14	0.15
34	1	0	0.50	0.50	0.01	1	1	0.29	0.54	0.17
35	1	0	0.56	0.44	0.00	1	1	0.47	0.50	0.03
36	2	2	0.29	0.33	0.38	2	2	0.29	0.30	0.41
38	1	1	0.32	0.65	0.03	1	1	0.18	0.57	0.25
42	1	0	0.50	0.41	0.09	1	1	0.27	0.62	0.10
43	0	0	0.82	0.17	0.00	0	0	0.53	0.34	0.13
45	0	0	0.78	0.21	0.00	1	1	0.25	0.63	0.12

Continuă pe următoarea pagină

#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
47	0	0	0.85	0.15	0.00	0	0	0.70	0.27	0.03
48	0	0	0.83	0.17	0.00	0	0	0.80	0.14	0.06
49	0	0	0.59	0.33	0.07	1	1	0.28	0.46	0.26
50	0	0	0.91	0.09	0.01	0	0	0.87	0.11	0.03
51	1	1	0.30	0.62	0.07	1	1	0.17	0.64	0.20
54	0	0	0.94	0.06	0.00	0	0	0.81	0.17	0.02
55	0	0	0.82	0.10	0.08	0	0	0.66	0.19	0.14
56	0	0	0.80	0.14	0.06	1	1	0.27	0.55	0.17
57	0	0	0.75	0.18	0.07	0	0	0.79	0.14	0.07
59	1	1	0.17	0.71	0.12	2	2	0.23	0.22	0.55
60	2	1	0.23	0.38	0.38	2	2	0.14	0.23	0.64
63	0	0	0.78	0.20	0.02	0	0	0.75	0.19	0.05
64	0	0	0.82	0.18	0.01	0	0	0.76	0.18	0.06
65	0	0	0.78	0.22	0.01	0	0	0.77	0.16	0.06
66	0	0	0.90	0.10	0.01	0	0	0.60	0.35	0.05
67	2	2	0.22	0.38	0.40	2	2	0.20	0.26	0.54
69	1	1	0.21	0.56	0.23	1	1	0.17	0.45	0.38
72	1	1	0.46	0.50	0.04	1	1	0.41	0.49	0.10
74	0	0	0.85	0.14	0.01	0	0	0.84	0.12	0.04
75	0	0	0.86	0.13	0.01	0	0	0.88	0.07	0.05
77	0	0	0.92	0.08	0.00	0	0	0.87	0.08	0.05
79	0	0	0.81	0.17	0.02	1	1	0.28	0.66	0.06
82	0	0	0.80	0.14	0.06	0	0	0.61	0.27	0.12
87	0	0	0.81	0.17	0.02	0	0	0.62	0.28	0.10

Continuă pe următoarea pagină



#	6 luni					12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
88	0	0	0.91	0.09	0.00	0	0	0.92	0.06	0.02
91	0	0	0.84	0.15	0.01	0	0	0.53	0.40	0.07
95	1	1	0.28	0.64	0.08	2	2	0.33	0.25	0.42
98	1	0	0.56	0.43	0.01	1	0	0.59	0.37	0.04
99	2	2	0.22	0.37	0.42	2	2	0.14	0.40	0.46
101	0	0	0.77	0.19	0.04	0	0	0.79	0.10	0.11
103	1	1	0.23	0.72	0.06	1	1	0.12	0.67	0.21
104	1	1	0.06	0.73	0.21	2	2	0.10	0.35	0.54
106	0	0	0.68	0.25	0.07	1	1	0.23	0.59	0.18
107	1	1	0.25	0.66	0.09	2	2	0.14	0.34	0.52
108	1	1	0.32	0.60	0.08	0	0	0.55	0.23	0.22
109	1	1	0.23	0.67	0.10	1	1	0.15	0.60	0.25

Tabela 4.33: experiment 5 - valorile POSE prezise din datele inițiale, fără miR 203, pentru setul de antrenament. O – valoare originală și P – valoare prezisă

Pentru a prezice valoarea POSE de la 12 luni, folosind datele inițiale și valoarea POSE de la 6 luni, însă fără analiza miR 203, clasificatorul a obținut 98,33% acuratețe pentru setul de antrenament și 76,92% pentru setul de test.

		preziceri		
		control	control parțial	recidivă
actual	control	28	0	0
	control parțial	0	21	0
	recidivă	0	1	10

Tabela 4.34: experiment 5 - matrice de confuzie pentru setul de antrenament, pentru a prezice POSE la 12 luni din datele inițiale și valoarea POSE de la 6 luni, fără miR 203

		preziceri		
		control	control parțial	recidivă
actual	control	15	1	0
	control parțial	4	4	0
	recidivă	0	1	1

Tabela 4.35: experiment 5 - matrice de confuzie pentru setul de test, pentru a prezice POSE la 12 luni din datele inițiale și valoarea POSE de la 6 luni, fără miR 203

caracteristică	scor de importanță
6months.POSE	0.23
result.miR125	0.18
0months.LM	0.12
0months.endoscopy	0.09
general.age	0.08
0months.followtreatment	0.06
0months.HPQ-9	0.06
0months.SNOT	0.05
general.allergy	0.05
general.sex	0.03
general.asthma	0.03
general.EO	0.01
general.smoker	0.00
general.agemarker	0.00

Tabela 4.36: Importanța caracteristicilor în prezicerea POSE 12 luni, folosind datele inițiale și valoarea POSE de la 6 luni, mai puțin miR 203

	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
1	0	0	0.58	0.36	0.06
8	1	1	0.39	0.45	0.16
13	0	1	0.28	0.53	0.19
14	0	0	0.38	0.33	0.30
15	1	0	0.57	0.29	0.14
16	0	1	0.39	0.42	0.19
26	0	0	0.61	0.32	0.07
31	2	1	0.20	0.48	0.32
37	0	1	0.44	0.46	0.10
41	0	0	0.74	0.13	0.13
44	1	1	0.20	0.55	0.25
53	1	0	0.46	0.34	0.20
70	1	1	0.40	0.43	0.17
71	0	1	0.37	0.54	0.09
73	0	0	0.76	0.16	0.08
80	2	1	0.17	0.54	0.28
81	1	0	0.53	0.36	0.11
83	0	0	0.75	0.23	0.02
84	0	0	0.49	0.35	0.16
85	0	0	0.73	0.25	0.02
86	0	0	0.70	0.22	0.08
89	0	0	0.67	0.20	0.13
92	0	0	0.64	0.20	0.17
94	1	1	0.39	0.51	0.10
97	0	0	0.58	0.39	0.03

Continuă pe următoarea pagină

	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
100	1	0	0.76	0.22	0.02

Tabela 4.37: experiment 5 - valorile POSE prezise din datele inițiale și valoarea POSE de la 6 luni, fără miR 203, pentru setul de test. O – valoare originală și P – valoare prezisă

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
3	1	1	0.13	0.79	0.08
5	0	0	0.71	0.29	0.00
7	0	0	0.58	0.32	0.11
9	0	0	0.68	0.20	0.11
10	0	0	0.71	0.19	0.10
11	1	1	0.29	0.60	0.11
12	0	0	0.60	0.31	0.09
18	2	0	0.41	0.32	0.27
19	0	0	0.62	0.31	0.07
20	1	1	0.31	0.56	0.13
21	2	2	0.20	0.23	0.56
22	0	0	0.75	0.22	0.03
28	2	1	0.17	0.61	0.22
29	1	1	0.13	0.69	0.18
30	1	0	0.50	0.46	0.04
32	1	1	0.30	0.59	0.11
33	0	0	0.72	0.10	0.18
34	1	1	0.22	0.56	0.22

Continuă pe următoarea pagină

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
35	1	1	0.45	0.50	0.05
36	2	2	0.32	0.28	0.40
38	1	1	0.21	0.56	0.22
42	1	1	0.30	0.55	0.15
43	0	0	0.58	0.30	0.12
45	1	1	0.22	0.68	0.10
47	0	0	0.72	0.26	0.02
48	0	0	0.83	0.12	0.05
49	1	1	0.21	0.47	0.32
50	0	0	0.82	0.17	0.02
51	1	1	0.22	0.54	0.24
54	0	0	0.83	0.13	0.04
55	0	0	0.68	0.24	0.08
56	1	1	0.19	0.67	0.14
57	0	0	0.78	0.14	0.08
59	2	2	0.19	0.20	0.61
60	2	2	0.11	0.23	0.66
63	0	0	0.75	0.25	0.01
64	0	0	0.82	0.16	0.02
65	0	0	0.73	0.13	0.14
66	0	0	0.62	0.36	0.02
67	2	2	0.22	0.23	0.55
69	1	1	0.21	0.49	0.29
72	1	0	0.50	0.40	0.10

Continuă pe următoarea pagină

#	12 luni				
	O	P	$p(\text{control})$	$p(\text{partial})$	$p(\text{recidivă})$
74	0	0	0.80	0.18	0.02
75	0	0	0.86	0.11	0.03
77	0	0	0.79	0.13	0.09
79	1	1	0.35	0.59	0.06
82	0	0	0.62	0.22	0.16
87	0	0	0.59	0.33	0.09
88	0	0	0.94	0.04	0.02
91	0	0	0.57	0.33	0.10
95	2	2	0.33	0.26	0.41
98	1	0	0.65	0.31	0.05
99	2	2	0.22	0.38	0.40
101	0	0	0.84	0.10	0.05
103	1	1	0.10	0.73	0.17
104	2	2	0.10	0.31	0.59
106	1	1	0.26	0.50	0.24
107	2	2	0.19	0.37	0.45
108	0	0	0.53	0.24	0.23
109	1	1	0.26	0.46	0.28

Tabela 4.38: experiment 5 - valorile POSE prezise din datele inițiale și valoarea POSE de la 6 luni, fără miR 203, pentru setul de antrenament. O – valoare originală și P – valoare prezisă

## 4.3 Dezvoltare sistem PosePredict

### 4.3.1 Funcționalități:

Pentru a prezenta aplicabilitatea experimentelor efectuate, am realizat o aplicație web destinat medicilor. Site-ul PosePredict are ca scop furnizarea unui mediu accesibil pentru medici, prin care aceștia să poată folosi modelele generate.

După accesarea site-ului, utilizatorul trebuie să se autentifice, fie prin utilizarea unui cont preexistent, fie prin crearea unui nou. După autentificare, utilizatorul ajunge la pagina principală. Aici este prezentată o listă a tuturor pacienților pe care îi are adăugați, ordonați după ID. În plus, există un buton prin care utilizatorul poate adăuga noi pacienți.

Pentru a adăuga un pacient nou, este necesară introducerea datelor sale de identificare, analizelor preoperatorii, precum și a datei operației și a rezultatelor analizelor pentru polipii nazali. După validarea datelor, utilizatorul este redirecționat către pagina de vizualizare a tuturor pacienților, iar tabelul de predicții POSE este actualizat.

Tabelul de predicții POSE conține valorile POSE prezise pentru 6 și 12 luni, în funcție de datele disponibile în acel moment.

Atât la crearea unui pacient nou, cât și atunci când se modifică sau se adaugă informații unui pacient existent, datele introduse trebuie să respecte scala analizei respective. De exemplu, pentru scorul SNOT sunt acceptate doar valori cuprinse în intervalul 0-100.

Pentru fiecare pacient, există opțiunea de a accesa mai multe detalii. Pe lângă informațiile generale introduse la crearea pacientului, este disponibilă și secțiunea de 6 luni, iar pe viitor vor fi adăugate și secțiunile de 12 luni și 18 luni. Aici, medicul poate adăuga rezultatele controalelor efectuate la 6, respectiv 12 sau 18 luni. După adăugarea acestor rezultate, tabelul de predicții POSE este actualizat. De asemenea, în această pagină pot fi vizualizate și predicțiile valorii POSE.

În cazul în care pacientul în cauză nu are introduse datele de la 6 luni, se vor efectua atât predicții pentru 6 luni, cât și pentru 12 luni, pe baza valorilor inițiale. Dacă pacientul are datele de la controlul de 6 luni introduse, predicția pentru 12 luni va ține cont și de acestea.

În aceeași pagină există și opțiunea de editare, care duce utilizatorul către pagina de editare, unde toate datele introduse despre un pacient pot fi modificate. La salvarea noilor modificări, predicția POSE va fi recalculată conform datelor actualizate.

### 4.3.2 Proiectare:

#### Workflow

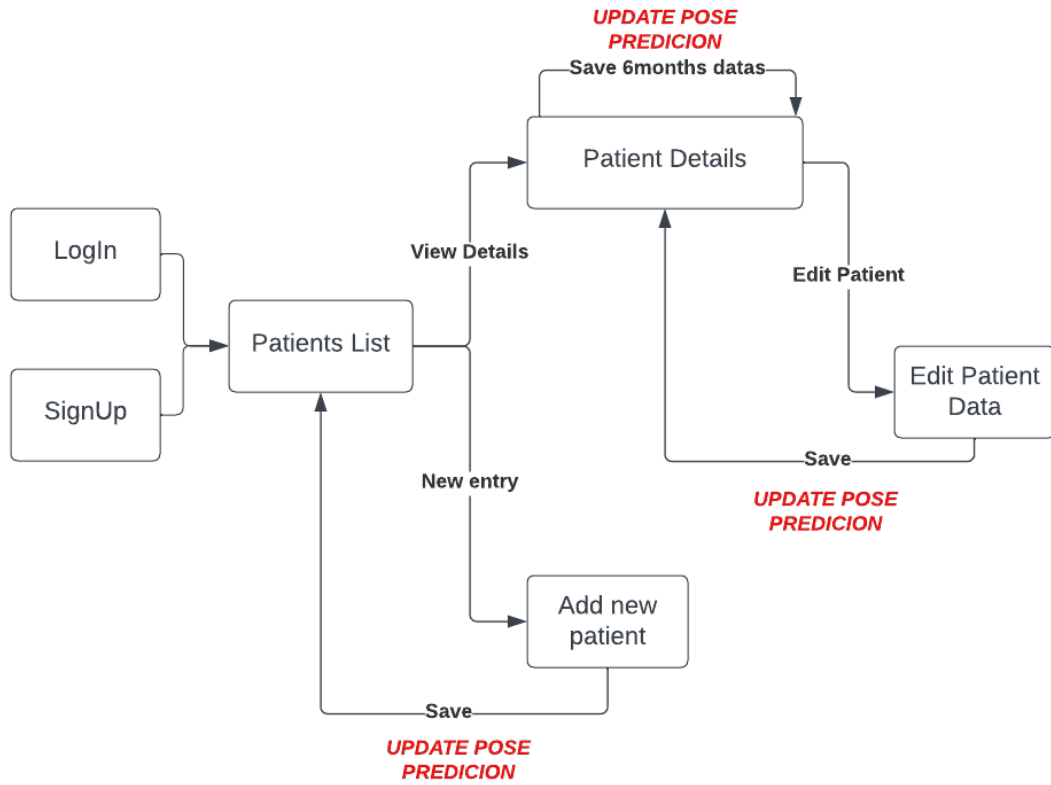


Figura 4.4: Workflow PosePredict website



## Baza de date

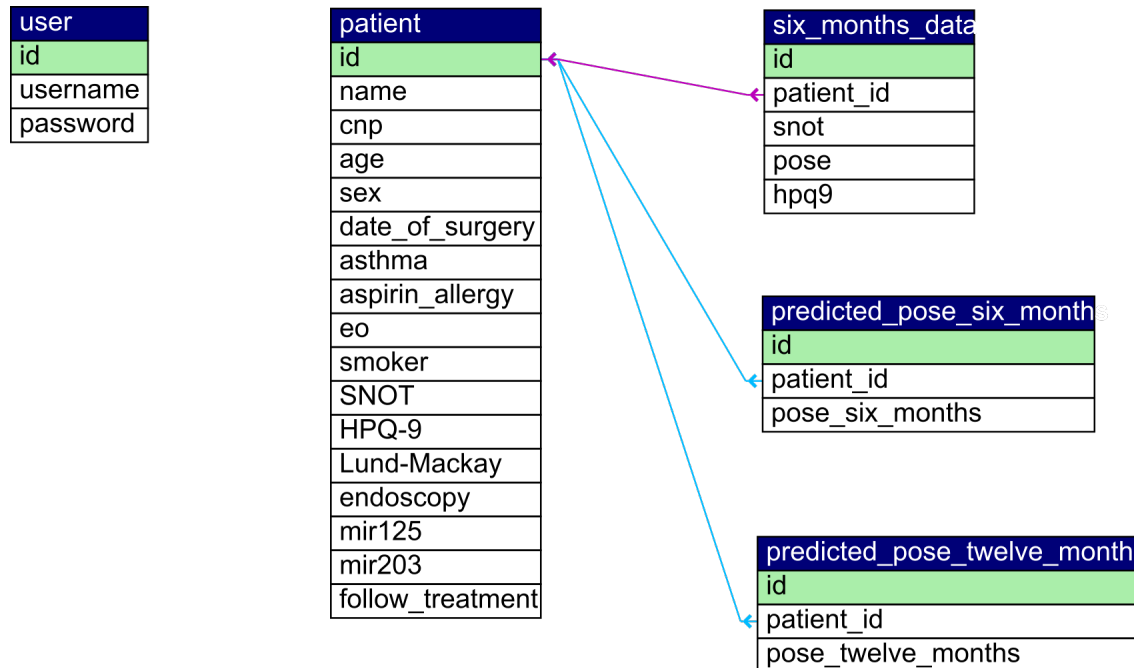


Figura 4.5: baza de date

### 4.3.3 Implementare

#### Tehnologii implicate

Site-ul a fost creat folosind o combinație de limbaje și tehnologii web pentru a oferi o experiență interactivă și atractivă utilizatorilor. Limbajele de programare PHP, HTML și CSS au fost utilizate în procesul de dezvoltare.

PHP a fost ales pentru a asigura funcționalități dinamice și prelucrarea datelor în timp real, în timp ce HTML a fost utilizat pentru structurarea conținutului și CSS pentru stilizarea și formatarea paginilor web.

Pentru a integra modelele, au fost utilizate și scripturi de Python. Python este un limbaj de programare popular și versatil, adesea folosit în domeniul dezvoltării web pentru manipularea datelor, inteligență artificială și alte sarcini complexe.

”Cu accentul pus pe ușurința în citire, Python este un limbaj de programare interpretat de nivel înalt, recunoscut pentru ușurința de a fi învățat, dar totodată capabil să utilizeze puterea limbajelor de programare de nivel sistemic atunci când este necesar. În afară de beneficiile aduse de limbaj în sine, disponibilitatea instrumentelor și a bibliotecilor fac din Python o alegere deosebit de atractivă pentru sarcinile de lucru din domeniul științei datelor, a învățării automate și al calculului științific[6].

”

Baza de date a site-ului a fost găzduită folosind phpMyAdmin, o aplicație web care facilitează administrarea bazelor de date MySQL. MySQL este un sistem de gestiune a bazelor de date relaționale utilizat frecvent în dezvoltarea aplicațiilor web. Utilizarea phpMyAdmin permite gestionarea eficientă a datelor, inclusiv stocarea, modificarea și interogarea lor.

Pentru a testa și a vizualiza site-ul, acesta a fost postat pe localhost prin intermediul USBWebserver. USBWebserver este o aplicație software care permite configurarea unui server web local pe un dispozitiv, precum un computer personal. Acesta facilitează rularea și testarea site-ului.

#### 4.3.4 Implementare efectivă

- `signUp.html`: Pagina conține formularul pentru înregistrarea utilizatorilor. Formularul include câmpuri de introducere pentru numele de utilizator, parolă și confirmarea parolei, împreună cu un buton de salvare și un buton pentru a reveni la pagina de autentificare.
- `index.php`: acest script creează pagina de autentificare. Formularul de autentificare conține câmpuri pentru numele de utilizator și parolă, un buton de trimitere către pagina de înregistrare și un buton de logare. Atunci când utilizatorul completează formularul și îl trimite, datele sunt procesate de scriptul "`process.php`" utilizând metoda POST, iar în cazul în care datele sunt valide, utilizatorul ajunge la pagina principală.
- `process.php`: se ocupă de procesarea autentificării și înregistrării utilizatorilor. Scriptul se conectează la baza de date MySQL și efectuează operațiile necesare.
- `main.php`: Acest script afișează lista de pacienți preluată din bază de date. Folosind tehnologia PDO, scriptul se conectează la baza de date și extrage informațiile necesare despre pacienți prin intermediul unei interogări SELECT. Apoi, datele sunt prezentate într-un tabel HTML, iar în cazul în care nu există pacienți în baza de date, se afișează un mesaj corespunzător. De asemenea, scriptul oferă și un buton pentru opțiunea de adăugarea unui pacient nou.
- `add_patient.php`: Acest script gestionează adăugarea unui pacient nou prin procesarea datelor trimise prin intermediul formularului și adăugarea lor în baza de date. Apoi, prin intermediul unei funcții externe, scriptul actualizează tabelul de predicții, "`predicted_pose`", cu ajutorul ID-ului. Formularul de adăugare a pacientului conține câmpuri precum numele, vârsta, sexul, istoricul medical și informațiile despre tratament.

- `edit_patient.php`: Aceasta pagina permite vizualizarea și editarea datelor unui pacient existent. Utilizând tehnologia PDO, scriptul se conectează la baza de date MySQL și extrage informațiile pacientului pe baza ID-ului acestuia transmis în URL. Apoi, datele sunt afișate într-un formular pentru editare, sau adăugare (în cazul în care nu au fost adăugate valorile pentru 6 luni).
- `save_patient.php`: Acest script PHP actualizează informațiile pacientului în baza de date MySQL, pe baza datelor trimise prin formular și al ID-ului.
- `save_six_months.php`: Scriptul primește scorul SNOT, POSE și HPQ-9 și le actualizează în tabel, actualizând și scorul POSE prezis.
- `script.js`: Acest cod JavaScript verifică dacă datele introduse pentru email respectă structura standard, afișând o bifă în caz pozitiv sau un "X" în caz contrar.
- `model.py`: Acest script Python încarcă modelul de DT și îl utilizează pentru a face predicția valorii POSE la 6 luni, pe baza setului de date preluate dintr-un fișier JSON (datele pacientului pentru care se face predicția). Înainte de a face predicțiile, scriptul efectuează preprocesarea datelor: le normalizează și elimină coloanele în plus. Apoi, rezultatele sunt afișate într-un alt JSON.
- `model_12m.py`: încarcă modelul corespunzător de DT și îl utilizează pentru a face predicția valorii POSE la un an, având și valoarea POSE de la 6 luni.
- `model_12m_from_preop.py`: încarcă modelul de DT și îl utilizează pentru a face predicția valorii POSE la un an, fără a avea valorile de la controlul de 6 luni.

#### 4.3.5 Modelele de predicție bazate pe Machine Learning folosite

Pentru acest site web s-au folosit 3 modele de Decision Tree: unul care prezice valoarea POSE la 6 luni din datele inițiale, unul care prezice valoarea POSE la 12 luni, tot din datele inițiale, și un model care prezice valoarea POSE la 12 luni ținând cont de datele inițiale și valorile de la controlul de 6 luni.

**POSE pentru 6 luni prezis din datele inițiale**

Acest model a fost obținut folosindu-se următorii parametri:

- `random_state=0`
- `max_depth=10`
- `min_samples_split=4`
- `min_samples_leaf=2`

	MAE (%)	RMSE (%)	$R^2$
<b>antrenament</b>	0.9073	0.0413	0.0031
<b>test</b>	0.1460	0.1335	0.0360

Tabela 4.39: Rezultate model prezicere POSE 6 luni

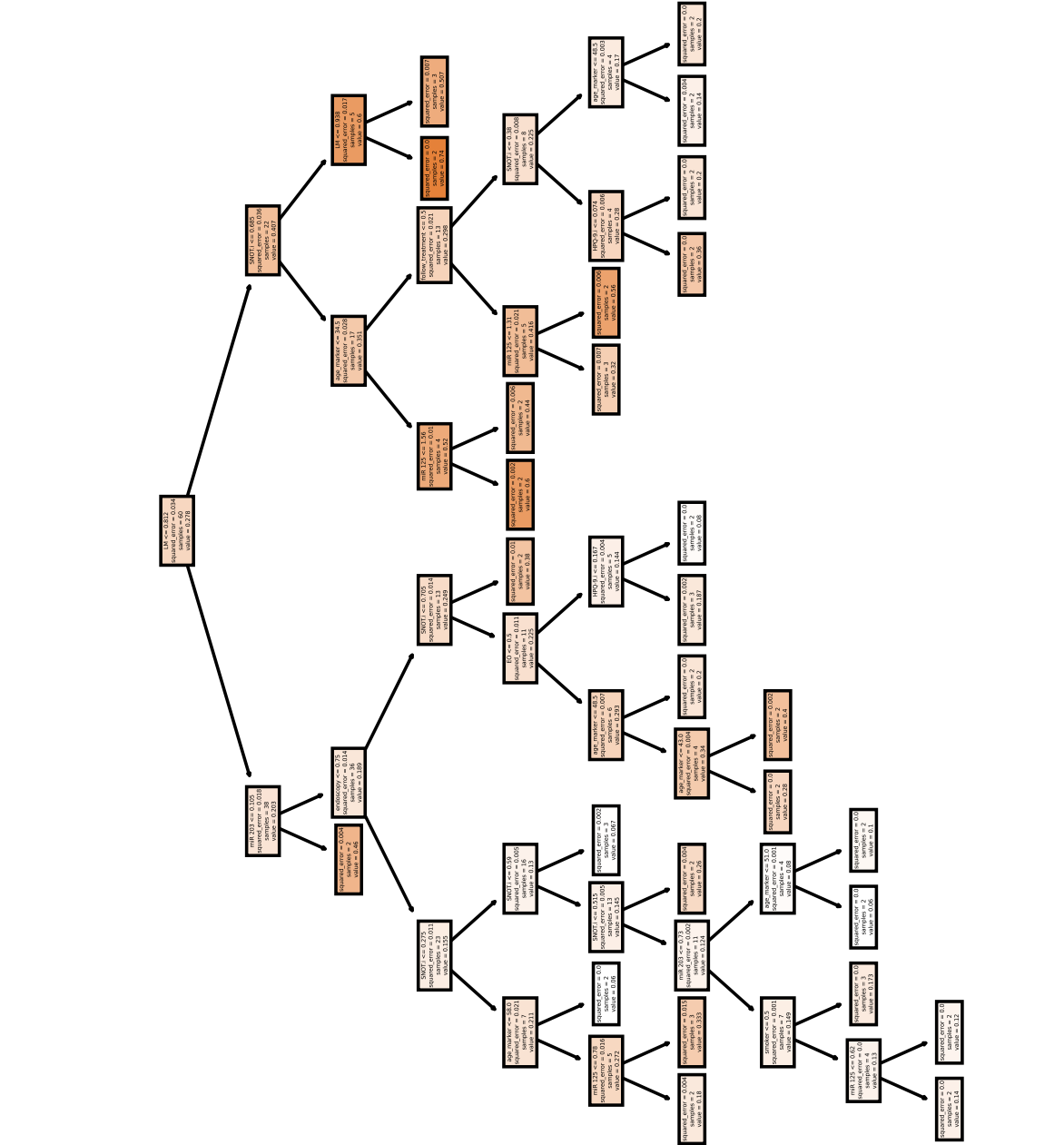


Figura 4.6: Arborele de decizie pentru modelul de 6 luni

**Pose pentru 12 luni prezis din datele inițiale**

Acest model a fost obținut folosindu-se următorii parametri:

- `random_state=0`
- `max_depth=5`
- `min_samples_split=2`
- `min_samples_leaf=3`

	MAE (%)	RMSE (%)	$R^2$
<b>antrenament</b>	0.7985	0.0820	0.0114
<b>test</b>	0.1197	0.1713	0.0496

Tabela 4.40: Rezultate model prezicere POSE 12 luni, din datele inițiale

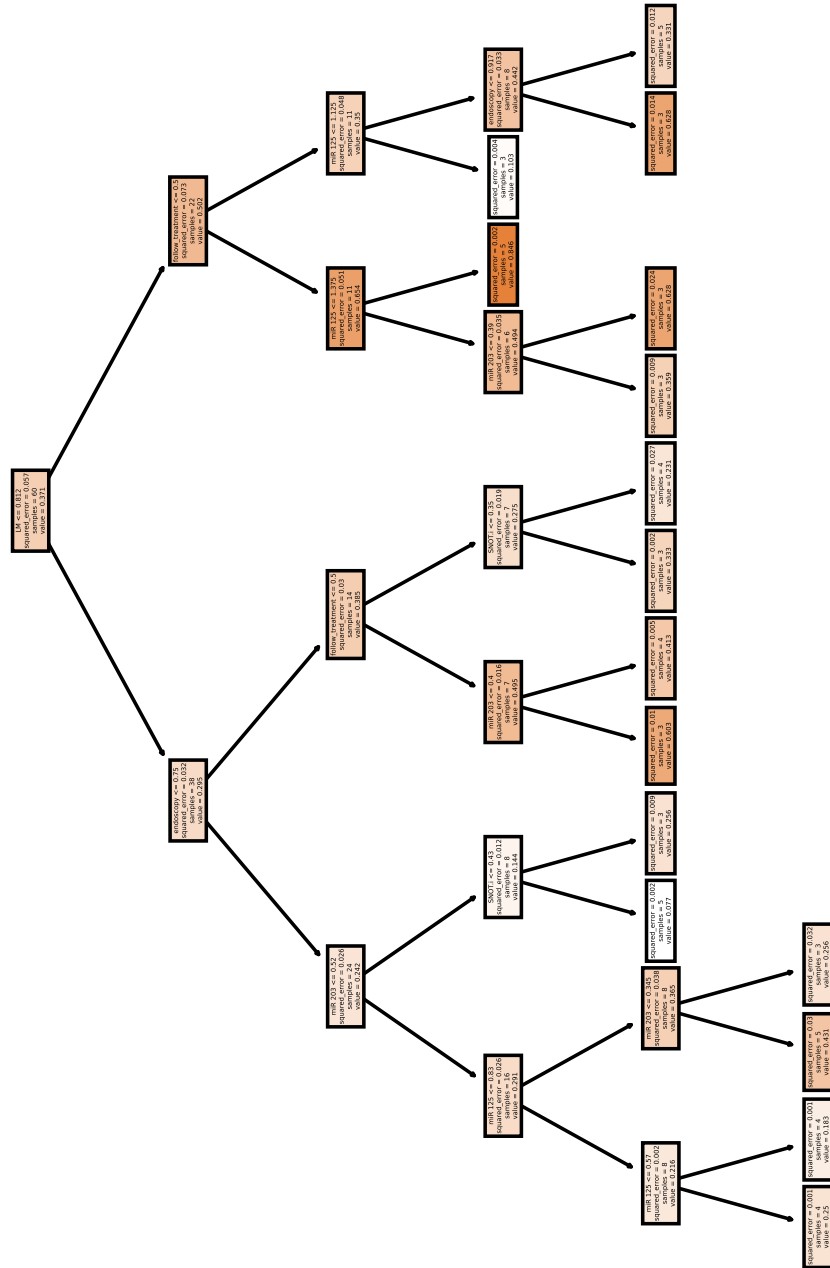


Figura 4.7: Arborele de decizie pentru modelul de 12 luni, din datele inițiale

**Pose pentru 12 luni prezis din datele inițiale și datele de la controlul de 6 luni**

Acest model a fost obținut folosindu-se următorii parametri:

- `random_state=0`
- `max_depth=10`
- `min_samples_split=4`
- `min_samples_leaf=2`

	MAE (%)	RMSE (%)	$R^2$
<b>antrenament</b>	0.9173	0.0391	0.0047
<b>test</b>	0.5812	0.1166	0.0236

Tabela 4.41: Rezultate model prezicere POSE 12 luni, din datele inițiale și datele de la controlul de 6 luni



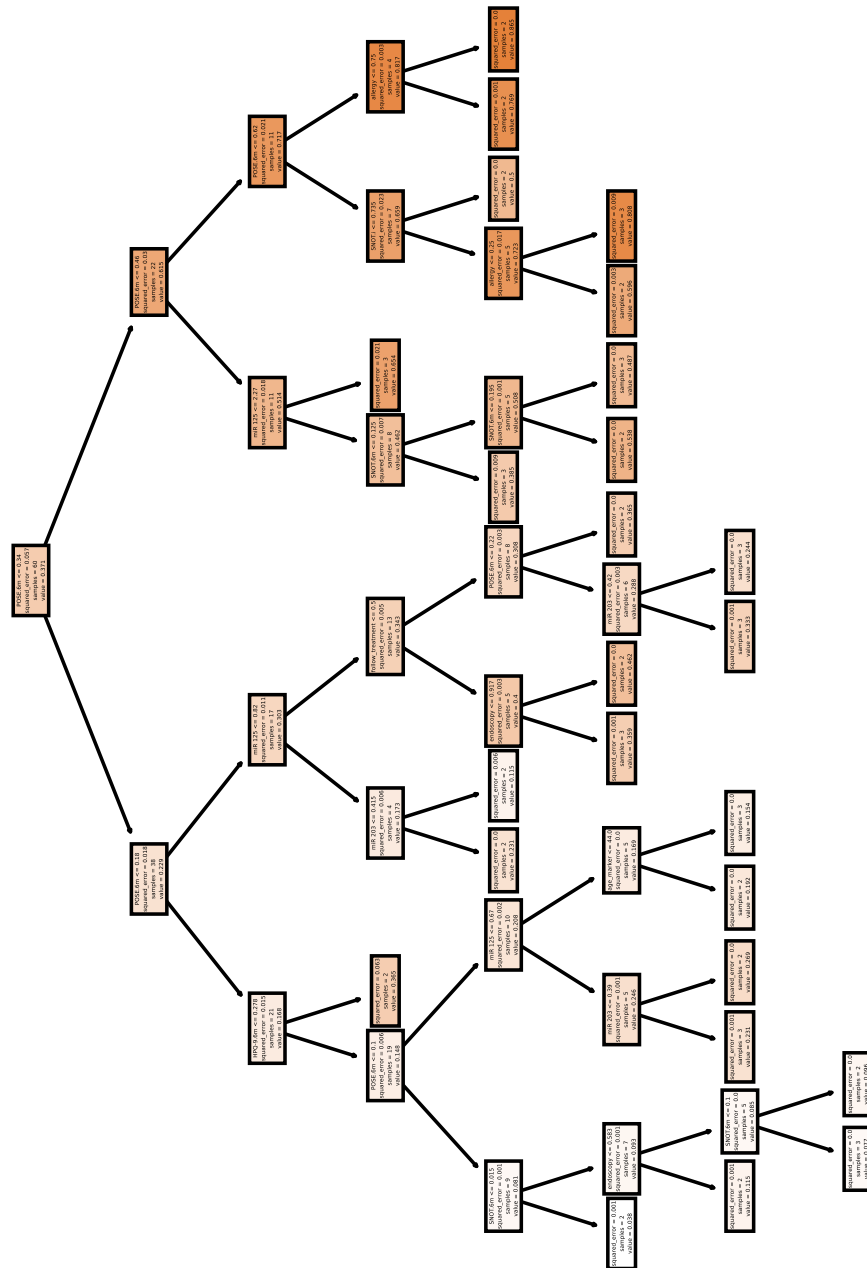


Figura 4.8: Arborele de decizie pentru modelul de 12 luni, din datele inițiale și datele de la controlul de 6 luni

### 4.3.6 Validare și testare

Testarea unui site web presupune evaluarea și validarea mai multor aspecte legate de funcționalitate, utilizabilitate, securitatea datelor, performanță, integrare și compatibilitatea unei aplicații web înainte de a fi lansată către utilizatorii finali.

Principalele tipuri de testare sunt:

- **Testarea funcționalității:** În acest tip de testare sunt verificate scenariile pe care utilizatorii finali (clienții) ar trebui să le urmeze și se asigură că atât scenariile pozitive, cât și cele negative sunt testate. Scopul principal este să se verifice dacă site-ul își îndeplinește scopul pentru care a fost creat.
- **Testarea utilizabilității :** În acest tip de testare se evaluează cât de ușor este navigarea între paginile site-ului, utilizarea meniurilor și vizibilitatea informațiilor prezentate. De asemenea, se verifică cât de intuitivă este utilizarea site-ului.
- **Testarea securității:** Se verifică aspecte precum accesul neautorizat, gestionarea sesiunilor și a datelor, verificarea token-urilor de acces și validarea certificatelor SSL și a redirect-urilor site-ului.
- **Testarea performanței:** Se asigură că aplicația răspunde într-un timp optim, chiar și în cazul unor căutări extinse sau unui volum mare de date. Se testează comportamentul aplicației în momentul în care este supusă unui volum mare de cereri sau utilizatorilor simultani.
- **Testarea compatibilității:** Această testare se asigură că site-ul funcționează optim pe diferite configurații, cum ar fi diverse browsere (Chrome, Firefox, Edge, Safari, Opera), sisteme de operare (MacOS, Linux, Windows, Android, iOS) și dispozitive (laptopuri, telefoane mobile, tablete). Testele de compatibilitate se concentrează pe afișarea corectă a elementelor site-ului în funcție de rezoluția ecranului și de caracteristicile specifice ale fiecărei configurații.
- **Testarea integrării (Integration testing)** Se concentrează pe asigurarea unei comunicări optime între componentele aplicației, inclusiv aplicația în sine, serverul web și baza de date. Se verifică corectitudinea gestionării cererilor din partea aplicației către serverul web și baza de date, precum și a răspunsurilor primite.

În cadrul aplicației, am testat diverse scenarii de utilizare pentru a mă asigura de buna funcționare a acestora și am încercat să mă asigur că aplicația răspunde tuturor nevoilor pentru care a fost proiectată. Totodată, am adăugat diverse aserțiuni prin care să asigur funcționarea optimă a site-ului, precum:

- `assert($pdo !== false , 'Failed to connect to the database');`

verifică dacă conexiunea cu baza de date a fost făcută corect

- `assert($stmt->rowCount() !== false , 'Failed to execute the SQL statement');`

verifică dacă scriptul SQL a fost executat corespunzător

- `assert(function_exists('updatePredictedPose') ,  
'updatePredictedPose function not found');`

verifică dacă funcția care urmează să fie apelat există

- `assert(isset($patientId) , 'Patient ID is missing');`

această linie se află înaintea unui script SQL prin care se selectează un pacient pe baza id-ului, așadar, verifică înaintea executării comenzii dacă id-ul a fost setat corespunzător

- `assert($patients !== false , 'Failed to fetch patients');`

această aserțiune verifică dacă lista pacienților a fost preluată corect.

- `assert($updatePatientStmt->rowCount() > 0 , 'Failed to update patient data');`

aserțiunea verifică dacă update-ul pacientului a fost făcut corespunzător

Totodată, pentru a asigura consistența bazei de date, utilizatorul nu poate introduce un format de date diferit față de cel regăsit în baza de date, iar formularele nu pot fi salvate cât timp câmpurile obligatorii din baza de date nu sunt completate.

De exemplu, scorul SNOT reprezintă o valoare necesară în baza de date, asigurându-ne că e prezentă prin "required". Valoarea SNOT trebuie să fie cuprinsă între 1 și 100, astfel încât câmpul corespunzător valorii SNOT are setat un minim = 1 și un maxim = 100.

```
<label for="SNOT">SNOT:</label>  
  <input type="number" name="SNOT" id="SNOT"  
    required min="1" max="100"><br>
```

# Capitolul 5

## Concluzii

Rinosinuzita cronică cu polipi nazali este o afecțiune inflamatorie a sinusurilor nazale și a căilor respiratorii superioare, caracterizată prin inflamație persistentă și formarea de polipi nazali în cavitatea nazală și în sinusuri.

Tratamentul pentru această afecțiune poate include terapia medicamentoasă sau poate consta în eliminarea polipilor prin intervenție chirurgicală. Apariția polipilor, atât înainte de operație, cât și după, are o cauză necunoscută.

Scopul acestor experimente este de a încerca să aproximeze valorile POSE după operația de extragere a polipilor, atât pentru 6 luni postoperatorii, cât și pentru 12 luni postoperatorii. Pentru a realiza experimentele, s-a folosit Decision Tree și Random Forest, atât pentru clasificare, cât și pentru regresie.

Rezultatele obținute au indicat o precizie de cel puțin 80% pentru a prezice valorile POSE. S-a remarcat, de fiecare dată, o performanță mult mai bună a Random Forest, în detrimentul Decision Tree. Totodată, experimentele în care s-a folosit clasificarea au avut rezultate mai bune.

Așadar, se remarcă viabilitatea folosirii algoritmilor de machine learning în problema rinosinuzitei cronice cu polipi nazali, susținând potențialul pe care utilizarea ML îl are în medicină.

# Bibliografie

- [1] Dr. Bogdan Mocanu, Polipoza nazală, <https://drbogdanmocanu.ro/afectiuni-si-tratamente/polipoza-nazala/>,.
- [2] Damm, M., Quante, G., Jungehuelsing, M., Stennert, E. (2002). Impact of functional endoscopic sinus surgery on symptoms and quality of life in chronic rhinosinusitis. *The Laryngoscope*, 112(2), 310-315.
- [3] Bogdan,P,[https://www.researchgate.net/publication/364929961\\_Studiul\\_asupra\\_algoritmilor\\_supervizati\\_folositi\\_in\\_aplicatii\\_de\\_Machine\\_Learning](https://www.researchgate.net/publication/364929961_Studiul_asupra_algoritmilor_supervizati_folositi_in_aplicatii_de_Machine_Learning), Raport de cercetare - Studiul asupra algoritmilor supervizați folosiți în aplicații de Machine Learning.
- [4] Podgorelec, V., Kokol, P., Stiglic, B., Rozman, I. (2002). Decision trees: an overview and their use in medicine. *Journal of medical systems*, 26, 445-463.
- [5] Paraschiv, E., Ovreiu, E. (2020). Tehnici bazate pe Machine Learning pentru îmbunătățirea depistării cancerului de sân. *Romanian Journal of Information Technology and Automatic Control*, 30(2), 67-80.
- [6] Raschka, S., Patterson, J., Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, 11(4), 193.
- [7] Wu, Q., Chen, J., Deng, H., Ren, Y., Sun, Y., Wang, W., ... Yang, Q. (2020). Expert-level diagnosis of nasal polyps using deep learning on whole-slide imaging. *Journal of Allergy and Clinical Immunology*, 145(2), 698-701.
- [8] Liu, I., Stein, E., Tan, B., Richter, C. P. , Using Machine Learning to Recognize Chronic Rhinosinusitis, *Annual Biomedical Research Conference for Minority Students (ABRCMS)*, 2022.
- [9] Ay, B., Turker, C., Emre, E., Ay, K., Aydin, G. (2022). Automated classification of nasal polyps in endoscopy video-frames using handcrafted and CNN features. *Computers in Biology and Medicine* Volume 147.