



(12) CERERE DE BREVET DE INVENȚIE

(21) Nr. cerere: a 2010 01327

(22) Data de depozit: 13.12.2010

(41) Data publicării cererii:
28.12.2012 BOPI nr. 12/2012

(71) Solicitant:
• UNIVERSITATEA DE MEDICINĂ ȘI FARMACIE "IULIU HAȚIEGANU" DIN CLUJ-NAPOCA, STR. EMIL ISAC NR.13, CLUJ-NAPOCA, CJ, RO;
• INSTITUTUL ONCOLOGIC "PROF. DR. ION CHIRICUTA" DIN CLUJ-NAPOCA, STR. REPUBLICII NR. 34-36, CLUJ-NAPOCA, CJ, RO;
• SOLUTIONS OF ARTIFICIAL INTELLIGENCE APPLICATIONS (ONG), STR. VLAHUȚA, BLOC LAMA C, AP. 45, CLUJ-NAPOCA, CJ, RO

(72) Inventatori:
• IRIMIE ALEXANDRU, STR. ADY ENDRE NR.40, CLUJ NAPOCA, CJ, RO;

• BĂLĂCESCU OVIDIU DANIEL, STR. LOIUS PASTEUR NR. 59, AP. 38, CLUJ-NAPOCA, CJ, RO;
• FLOARES GEORGE ALEXANDRU, STR. VLĂHUȚĂ, BLOC LAMA C, AP.45, CLUJ NAPOCA, CJ, RO;
• BERINDAN NEAGOE IOANA, BD. NICOLAE TITULESCU NR. 2, AP. 57, CLUJ-NAPOCA, CJ, RO

(74) Mandatar:
CABINET DE PROPRIETATE INDUSTRIALĂ CIUPAN CORNEL, STR. MESTECENILOR NR. 6, BL. 9E, AP. 2, CLUJ NAPOCA, JUDEȚUL CLUJ

(54) SISTEM INTELIGENT DE SUPORT AL DECIZIILOR CLINICE PENTRU DIAGNOSTICAREA CANCERULUI DE SÂN PRIN METODE NEINVAZIVE, BAZAT PE INTELIGENȚA ARTIFICIALĂ

(57) Rezumat:

Invenția se referă la un sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân prin metode neinvazive. Sistemul conform invenției este alcătuit dintr-un subsistem (A) de generare a unui model matematic, și dintr-un subsistem (B) de exploatare; primul subsistem (A) este, la rândul lui, alcătuit dintr-un bloc (1) de introducere a unor date i_{kj} , $k=1, \dots, 4$ și $j=1, \dots, 79$, constând din valoarea (i_{1j}) concentrației moleculei KGF pentru pacientul j , valoarea (i_{2j}) concentrației moleculei Angiopo2 pentru pacientul j , valoarea (i_{3j}) concentrației moleculei PDGF-bb pentru pacientul j , și vârsta (i_{4j}) pacientului j , dintr-un bloc (2) de împărțire aleatorie a setului de date de intrare în date de antrenare (a_{kj}), de testare (t_{kj}) și de validare (v_{kj}), dintr-un bloc (3) de procesare cu algoritmi genetici, care oferă un program selectat, care va fi validat într-un bloc (6), iar după validare, programul oferă un model matematic (5) ce va fi descris într-un limbaj de programare,

obținându-se un program care va fi utilizat la diagnosticarea cancerului de sân, în cadrul celui de-al doilea subsistem (B), de exploatare, care cuprinde un modul de calcul ce implementează modelul matematic (5), rezultatele obținute fiind afișate de un modul (10) de afișare.

Revendicări: 5

Figuri: 2

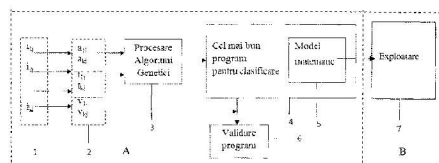
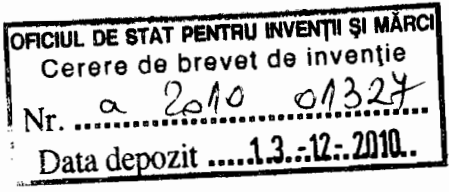


Fig. 1

Cu începere de la data publicării cererii de brevet, cererea asigură, în mod provizoriu, solicitantului, protecția conferită potrivit dispozițiilor art.32 din Legea nr.64/1991, cu excepția cazurilor în care cererea de brevet de invenție a fost respinsă, retrasă sau considerată ca fiind retrasă. Întinderea protecției conferite de cererea de brevet de invenție este determinată de revendicările conținute în cererea publicată în conformitate cu art.23 alin.(1) - (3).





Sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân prin metode neinvazive, bazat pe inteligență artificială

Invenția se referă la un sistem inteligent de suport al deciziilor clinice pentru stabilirea, cu acuratețe mergând până la 100%, a diagnosticului de cancer de sân, folosind metode neinvazive.

Utilizarea metodelor neinvazive reprezintă un avantaj al acestui sistem, în comparație cu alte tehnici propuse până în prezent pentru stabilirea diagnosticului de cancer de sân. Probele se recoltează doar din sângele pacienților, fără ca aceștia să fie supuși unei biopsii.

Problema pe care o rezolvă invenția propusă este de a oferi un sistem inteligent de diagnoză a cancerului de sân, pe cale neinvazivă, care oferă un grad mare de certitudine.

Sistemul inteligent de suport al deciziilor clinice pentru diagnosticul cancerului de sân prin metode neinvazive, bazat pe inteligență artificială, conform invenției, înlătură dezavantajele soluțiilor clasice, prin aceea că sistemul este alcătuit dintr-un subsistem A, de generare a modelului matematic pe baza unor algoritmi genetici, compus dintr-un bloc de introducere a datelor, dintr-un bloc de împărțire aleatorie a setului datelor de intrare în date de antrenare, de testare și de validare, dintr-un bloc de procesare cu algoritmi genetici, care oferă un program selectat care va fi validat într-un bloc de validare, programul validat generând un model matematic care va fi utilizat în subsistemul B, de exploatare în care se introduc concentrația moleculelor KGF și PDGF-bb și vârsta pacientului și care furnizează diagnosticul cu o acuratețe foarte ridicată.

S-a luat în considerare faptul că prezența unui cancer de sân în organism determină anumite anormalități biochimice. Modificările survenite în echilibrul dintre factorii angiogenici pozitivi și cei negativi conduc la angiogeneza, a carei promovare este facilitată de anormalitățile intervenite în organism, concomitent cu prevenirea apoptozei.

Astfel, analiza în laborator a probelor recoltate de la pacienți stabilește concentrația în sângele acestora a următoarelor molecule: Angiopo2, KGF și PDGF-bb.

Alegerea moleculelor s-a realizat conform unor criterii biomedical-oncologice și statistico-informatic, dintr-o multitudine de posibilități.

Sistemul propus prin această invenție presupune utilizarea valorilor unui grup de molecule, din cadrul a 8 tipuri de molecule analizate și a vârstei pacientului pentru a stabili, cu acuratețe mergând până la 100%, prin intermediul unui program ce descrie un model matematic pentru diagnosticul de cancer de sân sau normal (fără cancer) la pacient.

Dezvoltarea sistemului s-a realizat în două etape:

1. Efectuarea unui studiu pe un grup de 79 de pacienți și obținerea, prin aplicarea unei metode de inteligență artificială dovedită a fi alegerea optimă pentru problema propusă, a unui model matematic prin intermediul căruia se poate stabili diagnosticul de cancer de sân.
2. Implementarea modelului matematic într-un limbaj de programare cunoscut (C, Java, C#) pentru ca acesta să poată fi utilizat ca suport al deciziilor clinice în diagnosticarea cancerului de sân.

Cele două etape au presupus parcurgerea mai multor pași: selectarea grupului de pacienți și efectuarea analizelor de laborator; prelucrarea datelor prin tehnici de analiză statistică și procesarea lor prin diferite metode de inteligență artificială; obținerea formulei prin care se face clasificarea unui pacient, ca având cancer de sân sau nu, cu acuratețe foarte ridicată; implementarea formulei într-un limbaj de programare cunoscut.

Studiul s-a efectuat pe un grup de 79 de pacienți, împărțiți în două categorii: 58 de pacienți cu cancer de sân, ale căror probe s-au recoltat înainte de a fi supuși unei chimio-terapiei și 21 de pacienți fără cancer de sân (normal) care au reprezentat grupul de control.

Concentrația moleculelor Angiopo2, KGF și PDGF-bb s-a obținut prin evaluarea lor simultană folosind tehnologia FASTQuant.

Datele de la care s-a pornit în aplicarea metodelor de inteligență artificială au cuprins astfel valorile concentrației moleculelor - KGF, Angiopo2 și PDGF-bb – și vârsta celor 79 de pacienți, împărțiți în cele două clase: cu cancer de sân sau normal (fără cancer de sân).

Alegerea tehnicii de inteligență artificială s-a bazat pe evaluarea acuratetii cu care diversii algoritmi testați clasificau pacienții în cele două clase, folosind datele de intrare. Astfel, cele mai bune rezultate s-au obținut prin programare genetică, versiunea liniară (termenul nu se referă la obținerea unor modele liniare). Aceasta este o tehnică de învățare automată, supervizată, care presupune dezvoltarea unor programe folosind concepte ale evoluției biologice, implementate prin algoritmi genetici, o subclasă de algoritmi evolutivi.

Un pseudocod al algoritmului genetic utilizat ar putea fi prezentat în felul următor, ținând cont de notațiile:

P = populație inițială, formată din N cromozomi (programe, pentru programare genetică liniară)

P' = populație formată din n' cromozomi, aleși aleator din P , $n' < N$

p_k = cromozomi câștigatori selectați din P' , $k = 1, \dots, n'$, $n' < n'$ (parinti)

p_i = probabilitatea (rata) de încrucișare

p_m = probabilitatea (rata) de mutație

c_j = cromozomi obținuți prin încrucișarea cromozomilor p_k , $k = 1, \dots, n'$ și $j = 1, \dots, n' - n'$ (copii)

max = numărul maxim de generații (alegeri de subpopulații P') pe care algoritmul le construiește

$soluție$ = reține cel mai bun cromozom determinat în cadrul unei generații

$eroare$ = eroarea returnată de funcția de performanță pentru cel mai bun cromozom din generație

$const$ = eroarea soluției nu se mai îmbunătățește (devine constantă) după un anumit număr de generații

Algoritm Genetic(P, N, n', n', p_i, p_m)

// se initializeaza populatia P cu N cromozomi generati aleator

Initializare(P);

contor = 0;

Repeta

//se aleg din P, n' cromozomi care formeaza populatia P' și care participa la competitie

Alegere_Aleatoare(*P, P'*);

contor = *contor* + 1;

//evaluarea cromozomilor din P' cu ajutorul unei functii de performanta (fitness)

Evaluare(*P'*);

//se selecteaza din P', conform unui algoritm de selectie, un numar n'' de cromozomi castigatori ai competitiei

Selectare(*P', p_k, k = 1, ..., n''*);

//se incruciseaza cromozomii p_k, k = 1, ..., n'', denumiti parinti, conform unei probabilitati de incrucisare pi și se obtin noi cromozomi, denumiti copii c_j, j = 1, ..., n'-n''

Incruisare(*p_k, k = 1, ..., n'', c_j, j = 1, ..., n'-n''*);

//copii obtinuti prin incrucisare pot suferi mutatie cu o probabilitate pm

Mutatie(*c_j, j = 1, ..., n'-n''*);

//acesti copii inlocuiesc cromozomii necastigatori în P'

Inlocuire(*c_j, j = 1, ..., n'-n'', P'*);

// se reevalueaza populatia P', și se retine în solutie cromozomul care indeplineste cel mai bine cerinta problemei

Atribuire(*solutie, cel_mai_bun_cromozom(Evaluare(P'))*);

Pana cand ($contor = max$) sau ($eroare(solutie) = const$)

//returneaza cromozomul care a indeplinit cel mai bine cerintele problemei, fie ca eroarea evaluarii acestuia nu s-a mai modificat prin generarea unor noi cromozomi, fie ca s-a atins numarul maxim de generatii

Returneaza(*solutie*);

Sfarsit Algoritm Genetic.

Un astfel de algoritm genetic, în cazul programării genetice liniare, va utiliza datele introduse de utilizator pentru a antrena generații de programe și pentru a returna ca soluție programul care îndeplinește cel mai bine cerințele problemei, sau altfel spus, a cărui eroare determinată de funcția de performanță este minimă (la baza erorii aflându-se diferența dintre diagnosticul real și cel dat de program).

Pentru aplicarea tehnicii de programare genetică liniară, datele (perechile input-output) s-au împărțit în trei seturi: de antrenare (33%), de testare (33%) și de validare (33%). Inițializarea parametrilor, folosiți în procesul de construire și selectare a programelor optime la un moment dat, s-a efectuat cu valori alese în mod adecvat, în urma a numeroase experimente. Codul celui mai bun program a fost salvat în limbajele de programare C, Java, și C#. Setul de antrenare și cel de testare sunt folosite pentru antrenarea și selectarea programelor, astfel încât rata clasificării corecte să fie cât mai mare. Setul de validare intervine în evaluarea performanței unui program construit și returnat ca fiind cel mai bun dintr-o populație de programe, când algoritmul genetic și-a terminat execuția.

Datele corespunzătoare celor 79 de pacienți au fost împărțite în mod egal și aleator în cele trei seturi de date: de antrenare (27 pacienți), de testare (26 pacienți) și de validare (26 pacienți), iar acuratețea cu care cel mai bun program a clasificat setul de validare a fost de 100%.

Figura 1 reprezintă structura sistemului, necesară pentru obținerea prin programare genetică liniară a celui mai bun program, a cărui rată de clasificare, acceptată în etapa de validare, este de 100%.

Sistemul se compune dintr-un subsistem A, de generare a modelului matematic, și un subsistem B, de exploatare. Subsistemul A este alcătuit dintr-un bloc 1, de introducere a datelor, dintr-un bloc 2, de împărțire aleatorie a setului datelor de intrare în date de antrenare, de testare și de validare, dintr-un bloc 3, de procesare cu algoritmi genetici, care oferă un program selectat și considerat ca fiind cel mai bun. Programul va fi validat, într-un bloc 6. După validare programul oferă un model matematic 5, care va conține formula de diagnosticare. Modelul matematic este transferat subsistemului B, de exploatare.

Datele de intrare care se introduc în blocul 1 sunt notate cu $i_{kj}, k=1, \dots, 4$ și $j=1, \dots, 79$ și reprezintă:

i_{1j} = valoarea concentrației moleculei KGF pentru pacientul j ;

i_{2j} = valoarea concentrației moleculei Angiopo2 pentru pacientul j ;

i_{3j} = valoarea concentrației moleculei PDGF-bb pentru pacientul j ;

i_{4j} = vârsta pacientului j ;

$o_j, j=1, \dots, 79$, reprezintă diagnosticul pentru pacientul j , cu valoarea "1", dacă pacientul are cancer de sân și "0", în caz contrar.

Blocul 2 are rolul de împărțire a datelor de intrare, în mod aleator, în cele trei seturi de date de antrenare, testare și validare, după cum urmează:

a_{kj} = datele de intrare pentru setul de antrenare, $k=1, \dots, 4$ și $j=1, \dots, 27$;

t_{kj} = datele de intrare pentru setul de testare, $k=1, \dots, 4$ și $j=1, \dots, 26$;

v_{kj} = datele de intrare pentru setul de validare, $k=1, \dots, 4$ și $j=1, \dots, 26$;

Pentru a obține programul cu cea mai bună rată de clasificare în cele două categorii (cu cancer de sân sau cu diagnostic normal) datele de antrenare și cele de testare sunt procesate prin algoritmi de programare genetică liniară, implementați în blocul 3, de procesare. Cel mai bun program



rezultat din prelucrarea genetică este transferat într-un bloc 4, de calcul, după care urmează o procedură de validare într-un bloc 6, numit bloc de validare a programului. Blocul 6 realizează validarea programului prin verificarea acurateții de clasificare a pacienților care se face pe setul de date de validare, date care nu au fost utilizate în antrenarea și selectarea programelor. Datele din setul de validare se oferă ca date de intrare, iar programul oferă, pe baza algoritmului, un diagnostic "calculat" pentru fiecare pacient. Diagnosticul "calculat" este comparat cu diagnosticul cunoscut al pacientului din setul de validare. Programul este acceptat numai dacă oferă un diagnostic corect pentru toți pacienții din setul de validare, în caz contrar sistemul este reinițializat și se generează un nou program.

Obținerea celui mai bun program presupune operații succesive de încrucișare și mutație, eventual replicări ale programelor existente într-o populație. În cazul programării genetice liniare, aceste modificări au loc la nivelul instrucțiunilor din program, existând astfel posibilitatea ca unele variabile de intrare să nu se mai regăsească în codul celui mai bun program.

Instrucțiunile cuprinse în program descriu un model matematic (bloc 5) care presupune: o funcție matematică care depinde de toate sau numai de o parte din variabilele considerate în analiză și expresii logice cu ajutorul cărora se compară valoarea funcției cu o valoare "prag". În funcție de rezultatul comparației, programul returnează valoarea 1, dacă pacientul ale cărui date au fost prelucrate de funcție are cancer de sân sau valoarea 0, în caz contrar.

În cazul datelor furnizate de studiul propus în cadrul invenției, funcția matematică este descrisă prin relația:

$$f[Varsta, KGF, PDGF-bb] = \left\{ \left[Varsta - \frac{\sqrt{KGF}}{PDGF-bb} \times (-0.9487254619598389) \right] \times KGF \right\}^2 / KGF$$

Se compară valoarea calculată a funcției f cu valoarea „prag”.

Dacă $f[Varsta, KGF, PDGF-bb] \geq \text{valoare_prag}$, atunci programul returnează valoarea 1, în caz contrar returnează valoarea 0.

Se observă utilizarea ca și argumente ale funcției doar a variabilelor: *Varsta*, *KGF* și *PDGF-bb*, molecula *Angiopo2* nefiind suficient de relevantă pentru a fi păstrată în cadrul celui mai bun



program. Cel mai bun program este cel pentru care funcția f are o valoare care să ducă la returnarea valorii 1 de către program în cazul pacientului care are cancer de sân și 0 în caz contrar, când pacientul aparține clasei „normal”.

Acuratețea de 100% a clasificării realizate de cel mai bun program obținut prin programare genetică, a justificat crearea sistemului prin implementarea modelului matematic într-un limbaj de programare astfel încât acesta să poată fi utilizat ca suport al deciziilor clinice pentru diagnosticarea cancerului de sân.

Astfel, sistemul propune aplicarea unui model matematic pentru diagnosticarea unui pacient ca având cancer de sân sau nu folosind metode neinvazive de recoltare a probelor de la acesta.

Implementarea sistemului în practica medicală presupune parcurgerea următoarelor etape:

1. Recoltarea de la pacient a probelor de sânge
2. Determinarea concentrației moleculelor: KGF și PDGF-bb.
3. Introducerea valorilor celor două molecule și a vârstei pacientului în program
4. Interpretarea rezultatului returnat de program cu acuratețe mergând până la 100%:
 - 1, dacă pacientul are cancer de sân și 0, în caz contrar.

Se presupune că pasul 3 include și operațiunea de procesare a datelor de intrare pe care medicul le-a furnizat programului care descrie modelul matematic.

Subsistemul B, de exploatare (figura 2) se compune din blocul 7 care conține un bloc 8, de calcul, care, la rândul lui conține modelul matematic de la blocul 5, descris într-un limbaj de programare (C, Java sau C#). Programul din blocul 8 se bazează pe formula matematică prin care se procesează datele, pe expresii logice de evaluare a rezultatelor și de comparare a acestora cu valoarea “prag”. Datele de intrare pentru stabilirea diagnosticului unui pacient se introduc într-un bloc 9, de introducere a datelor, iar diagnosticul rezultat este afișat pe un dispozitiv de ieșire (de exemplu, un ecran) 10.

Datele de intrare necesare pentru stabilirea diagnosticului unui pacient sunt:



I_1 = concentrația moleculei KGF pentru pacient.

I_2 = concentrația moleculei PDGF-bb pentru pacient.

I_3 = vârsta pacientului.

În concluzie, invenția se referă la un sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân, bazat pe implementarea unui model matematic. Acest model permite clasificarea unui pacient ca având cancer de sân sau nu, utilizând informații ca: vârsta pacientului și concentrația moleculelor angiogenice KGF și PDGF-bb. Probele preluate de la pacient sunt obținute prin metode neinvazive, un factor de maximă importanță în asigurarea unui proces de investigare cât mai puțin traumatizant pentru pacient.



REVENDICĂRI

1. Sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân prin metode neinvazive, bazat pe inteligența artificială, **caracterizat prin aceea că**, sistemul este alcătuit dintr-un subsistem (A), de generare a modelului matematic și dintr-un subsistem (B), de exploatare, subsistemul (A) constă dintr-un bloc (1), de introducere a datelor, dintr-un bloc (2), de împărțire aleatorie a setului datelor de intrare în date de antrenare, de testare și de validare, dintr-un bloc (3), de procesare cu algoritmi genetici, care oferă un program selectat, care va fi validat într-un bloc (6), iar după validare programul oferă un model matematic (5), care va fi descris într-un limbaj de programare obținându-se un program care se utilizează la diagnosticarea cancerului de sân, în cadrul subsistemului (B), program care conține un modul de introducere a datelor pacientului într-un bloc (9), un modul de calcul care implementează modelul matematic (5) în blocul (8) și un modul (10), de afișare a rezultatului.
2. Sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân prin metode neinvazive bazat pe inteligența artificială, conform revendicării 1, caracterizat prin aceea că, modelul matematic utilizează funcția f descrisă prin relația:

$$f[Varsta, KGF, PDGF-bb] = \left\{ \left[Varsta - \frac{\sqrt{KGF}}{PDGF-bb} \times (-0.9487254619598389) \right] \times KGF \right\}^2 / KGF,$$

a cărei valoare, calculată pentru datele unui pacient, se compară cu o valoare „prag” , o valoare a funcției f care respectă condiția $f[Varsta, KGF, PDGF-bb] \geq \text{valoarea „prag”}$ indicând prezența cancerului de sân la pacientul diagnosticat, iar în caz contrar, indicând absența acestei boli.

3. Sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân prin metode neinvazive bazat pe inteligența artificială, conform revendicării 2, caracterizat prin aceea că, argumentele funcției f sunt: KGF, care reprezintă concentrația moleculei KGF, PDGF-bb, care reprezintă concentrația moleculei PDGF-bb și variabila „Varsta” reprezentând vârsta pacientului.
4. Sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân prin metode neinvazive, bazat pe inteligența artificială, conform revendicărilor 1 și 2, caracterizat

prin aceea că, setul de date analizat și utilizat pentru generarea modelului matematic prin procesare cu algoritmi genetici, împărțit în date de antrenare, testare și validare se prezintă sub forma de vectori, iar un vector corespunzător datelor unui pacient cuprinde:

i_{1j} = valoarea concentrației moleculei KGF pentru pacientul j;

i_{2j} = valoarea concentrației moleculei Angiopo2 pentru pacientul j;

i_{3j} = valoarea concentrației moleculei PDGF-bb pentru pacientul j;

i_{4j} = vârsta pacientului j;

o_j = diagnosticul (output) pentru pacientul j.

5. Sistem inteligent de suport al deciziilor clinice pentru diagnosticarea cancerului de sân prin metode neinvazive, bazat pe inteligență artificială, conform revendicărilor 1 și 2, caracterizat prin aceea că, datele introduse în modulul (9) al subsistemului de exploatare (B), date necesare pentru investigarea unui nou pacient constau în: concentrația moleculei KGF pentru pacient (I_1), concentrația moleculei PDGF-bb pentru pacient (I_2) și vârsta pacientului (I_3).

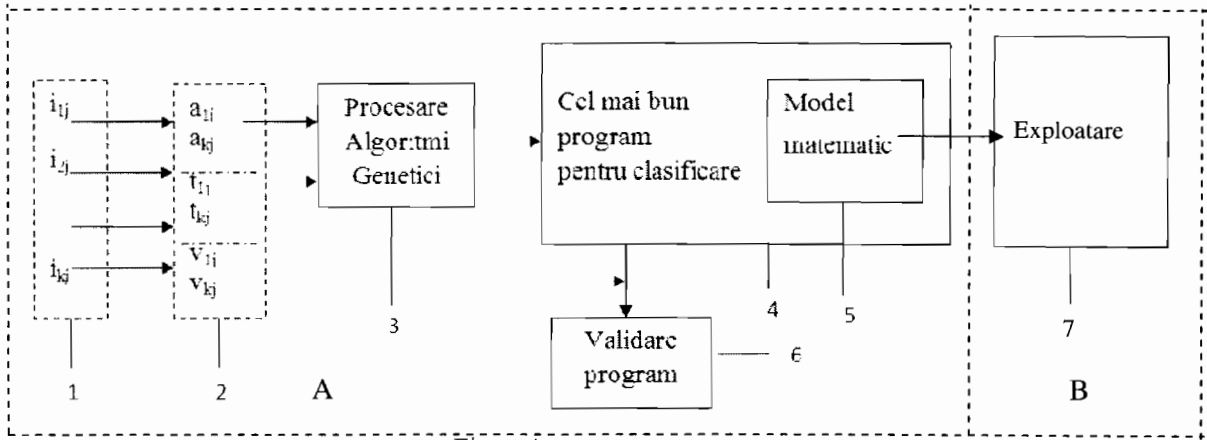


Figura 1

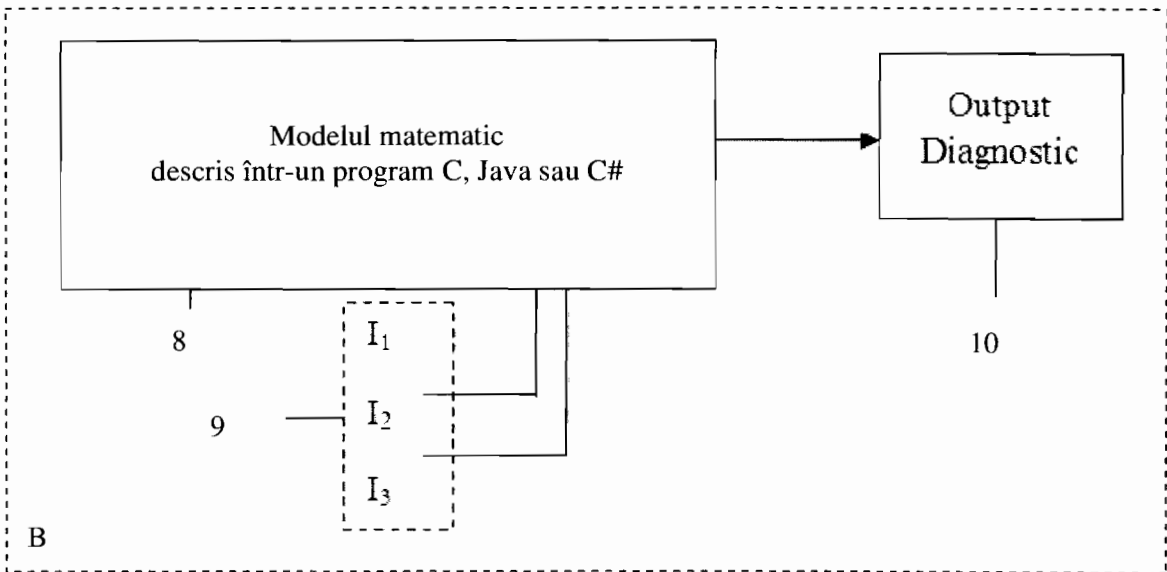


Figura 2