

UNIVERSITATEA TEHNICĂ
"GHEORGHE ASACHI" DIN IAȘI

Facultatea de Inginerie Chimică și Protecția
Mediului



ALGORITMI DE INSPIRAȚIE BIOLOGICĂ PENTRU MODELAREA ȘI OPTIMIZAREA PROCESELOR DIN INGINERIA CHIMICĂ

- REZUMAT TEZĂ DE DOCTORAT -

Conducător de doctorat:

Prof. univ. dr. Silvia Curteanu

Doctorand:

Ioana-Cristina Butnariu

IAȘI - 2015

UNIVERSITATEA TEHNICĂ "GHEORGHE ASACHI" IAȘI
RECTORATUL

Către

Vă facem cunoscut că în ziua de **17 Decembrie 2015** la ora 10⁰⁰, în **Sala de Consiliu a Facultății de Inginerie Chimică și Protecția Mediului**, va avea loc susținerea publică a tezei de doctorat intitulată:

**ALGORITMI DE INSPIRAȚIE BIOLOGICĂ PENTRU MODELAREA ȘI
OPTIMIZAREA PROCESELOR DIN INGINERIA CHIMICĂ**

elaborată de dna **BUTNARIU IOANA-CRISTINA** în vederea conferirii titlului științific de doctor.

Comisia de doctorat este alcătuită din:

- | | |
|--|------------------------|
| 1. Prof.univ.dr.ing. Dan CAȘCAVAL | președinte |
| Universitatea Tehnică "Gheorghe Asachi" din Iași | |
| 2. Prof.univ.dr.ing. Silvia CURTEANU | conducător de doctorat |
| Universitatea Tehnică "Gheorghe Asachi" din Iași | |
| 4. CP1.dr.ing. Maria CAZACU | referent oficial |
| Institutul de chimie macromoleculară „Petru Poni” Iași | |
| 5. Conf.dr. Sabin BURAGA | referent oficial |
| Universitatea „Alexandru Ioan Cuza” din Iași | |
| 6. Prof.univ.dr.ing. Florin LEON | referent oficial |
| Universitatea Tehnică "Gheorghe Asachi" din Iași | |

Cu această ocazie vă invităm să participați la susținerea publică a tezei de doctorat.

RECTOR,

Prof. univ. dr. ing. **Ion GIURMA**



Secretar universitate,

Ing. Cristina Nagiț

Mulțumiri

Doresc să adresez sincere mulțumiri doamnei prof. univ. dr. ing. Silvia Curteanu pentru îndrumarea, sprijinul și ajutorul necondiționat acordat pe întreaga durată a cercetării și elaborării tezei de doctorat. Realizările mele și experiența acumulată în domeniul aplicațiilor inteligenței artificiale în modelarea și optimizarea proceselor din ingineria chimică nu ar fi fost posibile fără îndrumarea domniei sale.

De asemenea, doresc să îmi exprim aprecierea și deosebita recunoștință față de domnul dr. ing. Florin Leon pentru contribuția și indicațiile oferite. Mulțumesc tuturor colegilor pentru cadrul profesional creat dar mai ales Inei Drăgoi pentru sprijinul și indicațiile oferite.

Alese mulțumiri cadrelor didactice de la Facultatea de Informatică pentru pregătirea pe care mi-au oferit-o în timpul studiilor universitare.

Aș dori, de asemenea, să mulțumesc domnului prof. univ. dr. Nicu Curteanu de la Institutul de Informatică Teoretică care m-a îndemnat să urmez aceste studii doctorale.

Cele mai calde gânduri și mulțumiri sunt adresate familiei mele și prietenilor pentru înțelegerea, răbdarea și încurajările acordate în toți acești ani. De asemenea aș dori să mulțumesc unei persoane deosebite care m-a încurajat și mi-a oferit indicații prețioase de-a lungul ultimului an pentru finalizare studiilor.

Cuprins

1. INTRODUCERE	1
1.1 Obiectivele tezei.....	2
1.2 Structura tezei.....	3
4. FRAMEWORK PENTRU MODELAREA ȘI OPTIMIZAREA PROCESELOR DIN INGINERIA CHIMICĂ	8
4.1. Principii generale în dezvoltarea framework-ului.....	8
4.1.1. Metodologia de modelare cu SVM	8
4.1.2 Selectarea modelului SVM optimal.....	10
4.1.3. Variante ale framework-ului implementate	11
4.2. Varianta de implementare GS-SVC și aplicații.....	11
4.2.1. Identificarea proprietății de cristal lichid a unor compuși organici.....	12
4.2.2. Predicția rezistenței la coroziune a aliajelor de titan.....	15
4.3 Varianta de implementare GS-SVR și aplicații.....	17
4.3.1. Purificarea electrochimică a apelor uzate	18
4.3.2. Îndepărtarea metalelor grele din soluții apoase prin bioremediere	19
4.3.3. Extracția	21
5. ALGORITMI DE OPTIMIZARE DE INSPIRAȚIE BIOLOGICĂ PENTRU MODELARE PROCESELOR CHIMICE	23
5.1. Varianta de implementare DE-SVR și aplicații.....	23
5.1.1. Polimerizarea radicalică a stirenului	25
5.2. Varianta de implementare GSO-SVR și aplicații	27
5.2.1. Implementarea algoritmului GSO	27
5.3.2. Implementarea variantei GSO-SVR.....	30

5.3.3. Îndepărtarea metalelor grele prin biosorptie cu alge.....	33
5.3.4 Predicția coeficientului de partiționare în sisteme apoase bifazice	36
6. CONCLUZII FINALE.....	39
6.1 Concluzii referitoare la obiectivele tezei	39
6.2. Aspecte originale	45
6.3. Direcții de continuare a cercetărilor	46
PUBLICAȚII CE VIZEAZĂ OBIECTIVELE REZOLVATE ÎN TEZĂ.....	47
BIBLIOGRAFIE (SELECTIVĂ)	49

1. Introducere

În domeniul ingineriei chimice, problemele principale sunt reprezentate de modelarea și optimizarea proceselor chimice pe baza experimentelor. Majoritatea abordărilor clasice sunt bazate pe legile fizice și chimice care guvernează procesele, legi care sunt descrise de modele matematice aplicate la diferite nivele care variază de la macroscopic la molecular. Aceste modele sunt reprezentate, în general, de forme derivative ale unor funcții puternic neliniare care, de cele mai multe ori, sunt greu de rezolvat. Tehnicile și instrumentele software oferă probabil cel mai eficient mod de a reprezenta interacțiunile complexe care au loc în timpul reacțiilor, pentru fiecare process.

Domeniul Inteligenței Artificiale este definit generic ca un domeniu care se ocupă cu dezvoltarea de metode inteligente pentru maximizarea eficienței unui sistem dinamic și aplicarea acestora pentru învățarea și interpretarea automată a datelor experimentale. Există o multitudine de algoritmi eficienți aparținând inteligenței artificiale pot fi regăsiți în literatura de specialitate, mașinile cu suport vectorial (SVM) fiind una din numeroasele instrumente utilizate în învățarea automată.

Algoritmii de inspirație biologică sunt metode bazate pe modele simplificate ale structurilor biologice, acestea încercând să imite diferite procese pentru a crea proceduri de optimizare. Dintre numeroasele metode de inspirație biologică, printre cele mai cunoscute se afla algoritmii evolutivi (EA), inteligență de grup (colectivă) sau rețele neuronale.

În aceasta teză vom descrie și implementa astfel de algoritmi pentru modelarea și optimizarea proceselor chimice. Un aspect important care trebuie luat în considerare este faptul ca aceste metode ale inteligenței artificiale nu înlocuiesc metodele clasice, ci completează aria de metode, fiind considerate alternative viabile, în special în cazurile în care complexitatea și numărul ridicat de parametri pun probleme abordărilor clasice.

1.1 Obiectivele tezei

În cadrul tezei de doctorat „Algoritmi de inspirație biologică pentru modelarea și optimizarea proceselor din ingineria chimică”, **mașinile cu suport vectorial (SVM)** care reprezintă o tehnică de învățare automată, în combinație cu algoritmi de inspirație biologică, cum ar fi algoritmul evoluție diferențială (DE) și algoritmul glowworm swarm optimisation (GSO), sunt utilizați pentru modelarea și optimizarea diferitelor aspecte ale proceselor din ingineria chimică.

Mașinile cu suport vectorial stau la baza metodologiilor elaborate, acestea reprezentând modelul procesului, scopul principal fiind acela de a crea metode îmbunătățite care să optimizeze eficient nu numai modelul, dar și sistemul chimic.

Au fost selectate ca studii de caz o serie de procese chimice, fiecare dintre acestea fiind descrise de către un set complet de date experimentale. Cerințele fiecărei probleme sunt rezolvate pentru prima oară folosind mașinile cu suport vectorial în combinație cu diferiți algoritmi de optimizare. Astfel, studiile de caz considerate sunt:

- proprietatea de cristal lichid a unor compuși organici,
- rezistența la coroziune a aliajelor de titan,
- purificarea electrochimică a apelor uzate,
- polimerizarea radicalică a stirenului,
- îndepărtarea metalelor grele prin biosorptie cu alge
- extracția în sisteme apoaze bifazice
- extracția de polifenoli din scoarța de molid
- îndepărtarea metalelor grele din soluții apoase prin bioremediere

Principalele obiective propuse și rezolvate în cadrul tezei sunt legate de crearea unui framework general care să fie ușor de utilizat de inginerii chimiști. Acestea sunt:

- Dezvoltarea de metodologii bazate pe SVM în combinație cu GridSearch și aplicarea acestora pentru rezolvarea unor procese chimice complexe. Diferitele combinații realizate constau în: vect

- i) mașini cu suport vectorial pentru clasificare cu parametri neoptimizați sau optimizați folosind algoritmul GridSearch cu diverse nuclee (SVC, GS-SVC);
 - ii) mașini cu suport vectorial pentru regresie cu parametri neoptimizați sau optimizați folosind algoritmul GridSearch cu diverse nuclee (SVR, GS-SVR);
 - iii) compararea rezultatelor obținute cu aceste metode cu alte metode de învățare automată cum ar fi rețele neuronale (ANN), algoritmi de tip "eager" (C4.5, REPTree, Random forest) și algoritmi de tip "lazy" (k-nearest neighbour, NNGEP)..
- Crearea unei metodologii pentru modelare și optimizare folosind versiunea pentru regresie a SVM în combinație cu algoritmi de optimizare de inspirație biologică: evoluție diferențială (DE-SVR) și inteligență comportamentală de grup (GSO-SVM).
 - Dezvoltarea unui framework general concretizat în diverse combinații de algoritmi care:
 - i) poate fi ușor îmbunătățită prin introducerea de noi algoritmi; și
 - ii) poate fi utilizată de inginerii chimiști pentru rezolvarea unor probleme variate.

1.2 Structura tezei

Această teză este organizată astfel încât se prezintă gradual framework-ul general care cuprinde toți algoritmi și care stabilește cadrul comun al acestora, ordinea acestora fiind direct legată de creșterea complexității și adăugarea de noi elemente. La început sunt prezentate partea teoretică pentru fiecare componentă, mai exact mașinile cu suport vectorial și algoritmi de optimizare de inspirație biologică. De asemenea sunt suprinse aplicațiile acestor algoritmi în ingineria chimică. Sunt descrise mai apoi setările generale ale framework-ului, cum ar fi tipurile de mașini vectoriale folosite, varietățile de nuclee, funcția de fitness, metodele de evaluare a performanței și complexității. Se trece apoi la descrierea particularităților fiecărui tip de model SVM, combinat cu GS, sau DE, sau GSO, urmat de aplicațiile concrete experimentale care ilustrează și testează eficacitatea acestora. Pentru a determina performanțele algoritmilor într-un context mai general, pentru unele procese sunt testați mai mulți algoritmi, o serie de comparații indicând care este cea mai bună abordare în funcție de caracteristicile fiecărui proces. Pentru fiecare studiu de caz sunt discutate o serie de

particularități ca: structura bazei de date reprezentată de date experimentale sau obținute folosind simulatoare, adaptarea algoritmilor la modelare procesului, rezultatele obținute și interpretarea acestora.

Din punct de vedere structural, teza este alcătuită din cinci capitole dintre care Capitolul 1 reprezintă introducerea, Capitolele 2 și 3 conțin partea teoretică fundamentală a modelării și optimizării cu descrierea principiile de bază ale mașinilor cu suport vectorial, precum și ale algoritmilor evolutivi și algoritmilor inteligență de grup. Următoarele două capitole, Capitolul 4 și Capitolul 5 reprezintă contribuția originală a tezei. Fiecare capitol se încheie cu o secțiune de concluzii, ultimul capitol conținând concluziile finale prezentate din două puncte de vedere distincte: i) aspectele originale legate de dezvoltarea de noi algoritmi pentru rezolvarea diferitelor probleme specifice ingineriei chimice; și ii) procesele considerate și procedurile de modelare și optimizare aplicate.

Capitolul 1 prezintă a serie de aspecte generale, constituind o introducere în conținutul tezei. Sunt enumerate obiectivele principale ale tezei și este prezentată structura generală.

Capitolul 2 prezintă principalele elemente teoretice ale mașinilor cu suport vectorial și principiile de bază ale teoriei învățării statistice. Modele prezentate în această teză sunt bazate în mare parte pe SVM, și în consecință sunt prezentate detaliile teoretice ale acestui algoritm. De asemenea sunt prezentate aplicațiile acestora în ingineria chimică, care evidențiază capacitatea de generalizare și eficiența modelelor SVM în modelarea și optimizarea proceselor în domeniul ingineriei chimice.

Capitolul 3, al doilea capitol teoretic, detaliază principiile de bază din algoritmi evolutivi și algoritmi inteligență de grup, cu descrierea algoritmilor care vor fi folosiți în următoarele capitole. Deoarece algoritmul glowworm swarm intelligence (GSO) va fi implementat și folosit în cadrul framework-ului, sunt prezentate detaliile de implementare, avantaje și limitări, diferite setări și variante. De asemenea sunt prezentate aplicațiile existente ale algoritmilor evolutivi în domeniul ingineriei chimice.

Capitolul 4 este primul capitol care abordează contribuțiile originale ale tezei. Plecând de la o serie de considerente practice, sunt prezentate elementele de bază ale framework-ului

POM-SVM dezvoltat pentru a include toate metodologiile și algoritmi creați. Este descris fluxul general al aplicației, ca o combinație “*proces – mașini cu support vectorial – algoritm de optimizare*” și care reprezintă baza metodologiei propuse. Două tipuri de modele sunt dezvoltate cu acest framework, GS-SVC pentru clasificare cu aplicație în predicție proprietății de cristal lichid și GS-SVR pentru regresie, aplicat la alte două procese chimice:.

Capitolul 5 prezintă implementări mai complexe ale frameworkului POM-SVM, în care optimizatorii sunt algoritmi evolutivi, cum ar fi DE sau GSO. Prima parte descrie implementarea DE-SVM pentru modelarea unui proces complex de polimerizare. Partea a doua prezintă varianta GSO-SVM, care este inovativă atât din punct de vedere al implementării, cât și al aplicațiilor. Din punct de vedere al implementării, este dezvoltată o nouă implementare în Java a acestui algoritim. Din punct de vedere al aplicațiilor, combinația celor două modele GSO și SVM este nouă în ingineria chimică. Această variantă va fi aplicată în modelarea a două procese complexe. Ultima parte prezintă varianta GA-SVM folosită în optimizarea proceselor chimice.

Secțiunea **Concluzii** subliniază realizarea obiectivelor propuse din două puncte de vedere: *algoritmice* (tehnicile bazate pe instrumente de inteligență artificială dezvoltate și aplicate) și *proces* (problemele rezolvate, modelările și optimizările efectuate, precum și rezultatele obținute).

Elementele originale ale acestei teze sunt reprezentate de:

- 1) **dezvoltarea de metodologii noi, simple și eficiente**, bazate pe diverși algoritmi de învățare, individuali sau în combinații hibride;
- 2) **o implementare nouă a algoritmului GSO**, având un criteriu de convergență specific problemei de optimizare a parametrilor algoritmului SVM;
- 3) **aplicarea acestor metodologii la modelarea proceselor din ingineria chimică**, aceste probleme nefiind abordate până acum folosind aceste metode hibride de inspirație biologică;
- 4) **instrumente software ușor de utilizate** și care oferă flexibilitate prin diversele configurări.

Metodologiile dezvoltate au un caracter general și pot fi ușor adaptate și aplicate altor procese și sisteme din domeniul ingineriei chimice.

Cercetarea realizată în cadrul acestei teze deschide noi perspective și posibilități de a modela eficient procese chimice complexe, cu aplicații practice în tehnica experimentală (economii de materiale, energie, timp) și industrială (eficientizarea controlului optimal). Interacțiunea cu instrumentele software dezvoltate este destul de flexibilă astfel încât inginerul chimist să obțină cu ușurință predicții ori de câte ori este necesar și în orice condiții experimentale.

Figura 1.1 ilustrează structura tezei, marcând clar cele două niveluri pe care a fost construită (algoritmii de modelare/optimizare și procesele din domeniul ingineriei chimice).

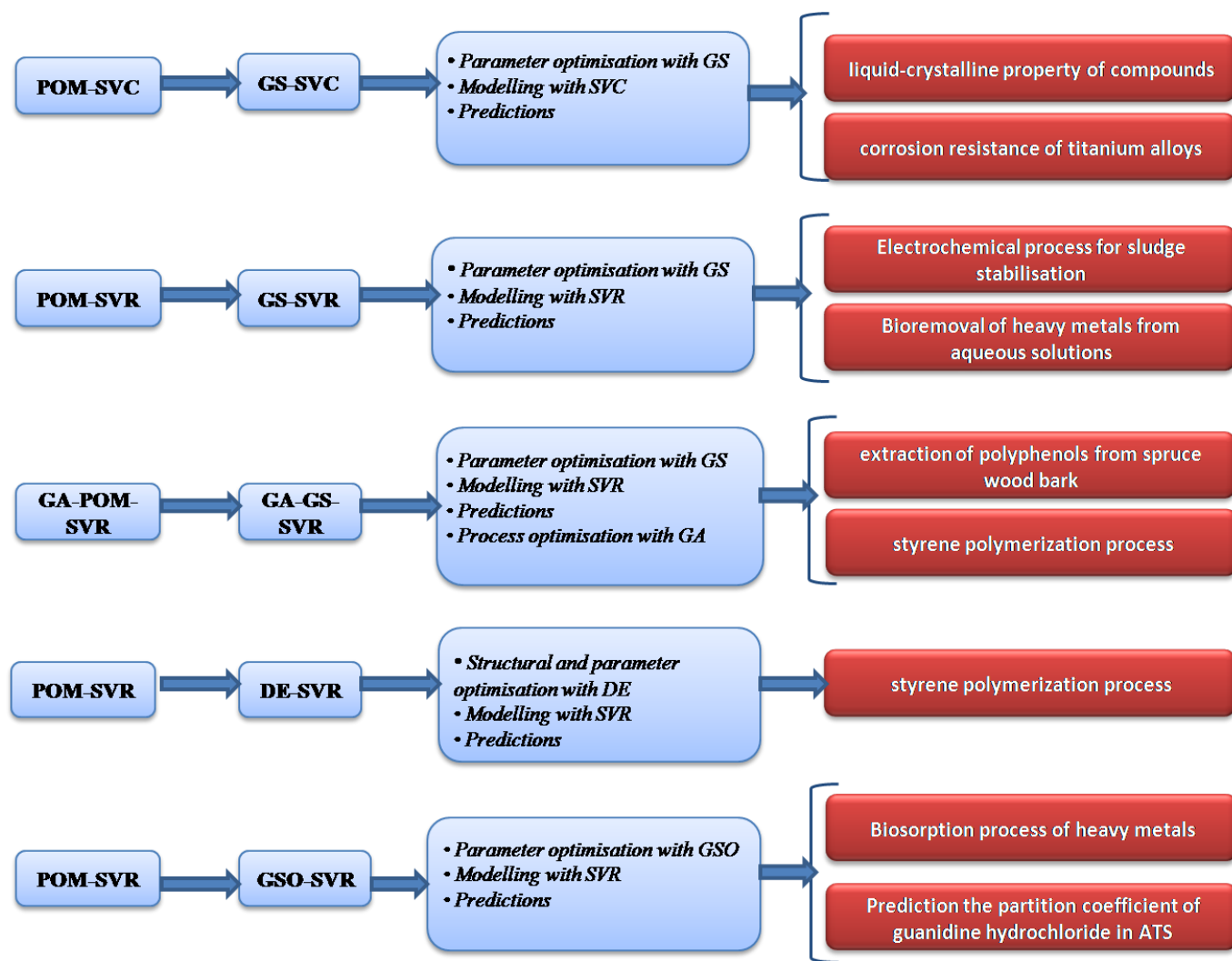


Figura 1.1. Structura tezei

4. Framework pentru modelarea și optimizarea proceselor din ingineria chimică

Framework-ul utilizat în această teză a fost dezvoltat pentru modelarea, predicția, clasificarea, și optimizarea sistemelor din ingineria chimică. Pentru a lucra cu o gamă largă de procese și a avea rezultate optime, framework-ul se bazează pe mașini cu suport vectorial care acționează ca un model general datorită capacității de aproximator universal și pe un algoritm de optimizare a parametrilor (GS, DE sau GSO).

4.1. Principii generale în dezvoltarea framework-ului

Framework-ul propus, numit POM-SVM (denumire derivată din termenii model de optimizare și mașini cu suport vectorial) este unul general, putând rezolva diferite aspecte ale unui sau mai multor procese. Modelare proceselor se realizează cu algoritmul SVM, care are capacitatea de aproximator universal, în combinație cu un algoritm de optimizare a parametrilor modelului SVM (GS, DE sau GSO). Acești algoritmi de inspirație biologică au fost selectați pe baza adaptabilității și capacității de generalizare. Metodele de optimizare primesc ca și intrare un număr de parametri împreună cu intervalele de valori și returnează valorile optime ale acestor parametri în spațiul de căutare dat. În acest framework, algoritmi de optimizare au rolul de a selecta modelul SVM optimal, prin evaluarea diferitelor setări posibile de parametri pe baza funcției de fitness a modelului.

O schemă conceptuală a framework-ului este prezentată în Figura 4.1.

4.1.1. Metodologia de modelare cu SVM

Metodologia are doi pași de bază: pregătirea datelor de intrare și mai apoi procedura de învățare. În primul pas, datele experimentale sunt împărțite aleator în diferite proporții pentru antrenare/testare cu păstrarea unei distribuții normale a datelor de ieșire. Metodologia

de învățare constă în antrenarea unui model SVM pe datele de antrenare și evaluarea acestuia pe datele de testare.

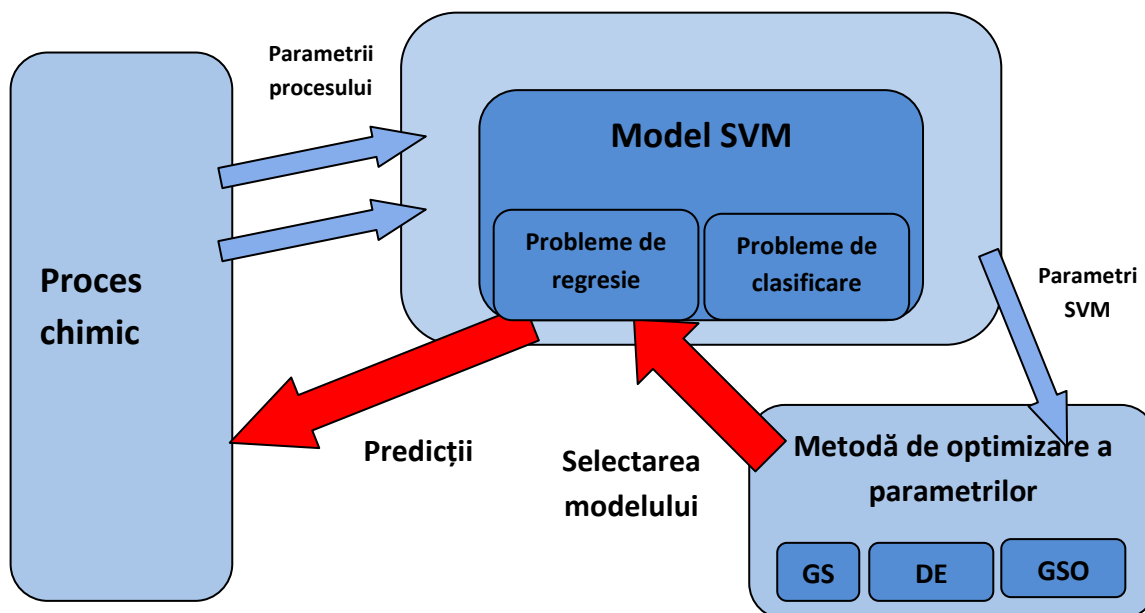


Figura 4.1. Schema generală a frameworkului

Implementarea SVM folosită în timpul experimentelor prezentate în teză este cea oferită de librăria LIBSVM (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). Această versiune de implementare SVM a fost dezvoltată de [Chang and Lin, 2011] pe baza algoritmului Optimizare Minimală Secvențială (SMO). Au fost utilizate în experimente trei tipuri de nuclee disponibile în LIBSVM: liniar, polynomial și Gaussian (RBF). Parametrii de configurare pentru aceste modele SVM sunt parametru cost C și parametri specifici nucleelor.

O performanță crescută a mașinilor cu suport vectorial poate fi obținută prin rezolvarea a două aspecte: optimizarea parametrilor și normalizarea datelor de intrare neprelucrate. Deoarece datele de intrare sunt foarte variate, metode de normalizare și standardizare sunt aplicate folosind ecuația 4.1. Normalizarea reprezintă o conversie liniară de scalare care asignează aceeași valoare absolută valorii relative corespunzătoare [Leeghim et al., 2008].

$$x_{normalized} = 2 \frac{x - Min}{Max - Min} - 1 \quad (4.1)$$

unde x reprezintă valoarea curentă din datele de intrare, Min este valoare minimă a datelor neprelucrate și Max este valoarea maximă. Această ecuație va scala liniar valoarea fiecărui atribut din datele de intrare neprelucrate în intervalul $[-1,1]$.

Avantajele acestei metode constau atât în minimizarea erorilor de estimare, dar și în reducerea timpului de procesare necesar antrenării unui model de către mașinile cu support vectorial [Chang and Lin, 2011].

4.1.2 Selectarea modelului SVM optimal

Performanța modelelor SVM este influențată în primul rând de selectarea unui nucleu corespunzător, dar mai ales alegerea parametrilor optimali specifici modelului. Probleme diferite au parametri optimali specifici care trebuie aflați. În consecință, selectarea modelului potrivit și a setărilor optimale este o cerință importantă în modelare SVM.

O abordare direct constă în căutarea în spațiul parametrilor folosind o abordare exhaustivă de tip grilă, numită GS (grid search) [Popov & Sautin, 2008]. Algoritmul GS antrenează modelul SVM prin încercarea unui număr predefinit de combinații de valori ai parametrilor și selectează pe cei care obțin valoarea maximă pentru funcția de fitness.

LIBSVM oferă o implementare de tip grid-search pentru optimizarea parametrilor care poate căuta valori optime pentru maxim doi parametri concomitent [Hsu et al., 2010]. Utilizatorul oferă intervalul de valori pentru fiecare parametru și acest instrument returnează valorile optimale conform cu acuratețea de validare încrucișată. Scriptul de rulare este adaptat pentru nucleul Gaussian care depinde doar de doi parametri: parametrul de cost C și parametrul de nucleu γ . Acest script a fost adaptat să ruleze și pentru nucleul linear (care depinde doar de parametrul de cost) și cu cel polinomial, care depinde de trei parametri de nucleu γ , d , and r , plus parametru de cost C . În acest caz, căutarea valorilor optimale pentru C și γ sunt evaluate folosind valoarea inițială pentru parametrul r , în timp ce parametrul d care desemnează gradul polinomului are câteva valori prestabilite ($d = 2, 3$ sau 4).

Algoritmii mai complecși folosiți în optimizare sunt Evoluție Diferențială (DE) și Glowworm Swarm Optimisation (GSO), care aparțin categoriei algoritmilor de inspirație

biologică. În literatura recentă, acești algoritmi de optimizare biologici au început să fie folosiți ca și parametri de optimizare pentru modelele SVM, datorită convergenței rapide și a abilităților mai bune de căutare globală [Lu et al., 2009; Ren and Bai, 2010; Jiang et al., 2013]. GSO este un algoritm relativ nou, care nu a fost încă folosit la capacitatea sa maximă. În Capitolul 5 vom descrie propria implementare a acestui algoritm, și sunt descrise două aplicații de optimizare a parametrilor modelelor SVM în ingineria chimică.

O varietate de modele, rezultate din diverse combinații de algoritmi au fost testate pe studiile de caz prezentate pentru a stabili influența parametrilor SVM și pentru a alege varianta optimă a modelului pentru probleme specifice de clasificare și regresie.

4.1.3. Variante ale framework-ului implementate

Framework-ul propus are două variante de modelare: POM-SVC pentru probleme de clasificare și POM-SVR pentru probleme de regresie. Trei algoritmi de optimizare a parametrilor sunt utilizați: grid-search (GS), evoluție diferențială (DE) și inteligență colectivă - glowworm swarm optimisation (GSO) și un algoritm de optimizare a proceselor: algoritmi genetici (GA). În această teză s-au utilizat următoarele implementări ale frameworkului general, aplicate modelării proceselor chimice: GS-SVC, GS-SVR, DE-SVR, GSO-SVR, and GA-SVR. Primele două variante și aplicațiile lor sunt descrise în acest capitol, iar celelalte sunt detaliate în capitolul 5.

4.2. Varianta de implementare GS-SVC și aplicații

Problema de clasificare este un caz special de învățare, în care ieșirea este un set fix de etichete. Problema este de a construi un model bazat pe datele de intrare experimentale care va asigura corect etichete pentru datele noi, obținând astfel un model cu o capacitate bună de generalizare. Procedura standard de învățare constă în împărțirea datelor aleatoare în date de antrenare și date de test. Ipoteza este apoi învățată pe baza setului de antrenare, iar eroarea de clasificare se estimează pe datele de testare. Pentru a obține o evaluare cât mai reală a modelului, împărțirea datelor se face de mai multe ori și este raportată eroarea medie.

Performanța clasificatorilor SVM este estimată prin calcularea *acurateței* modelelor, care este o metodă statistică definită ca procentul de rezultate estimate corect din numărul total de date testate:

$$Accuracy = \frac{\#correctly_predicted_data}{\#total_testing_data} \times 100 \quad (4.2)$$

4.2.1. Identificarea proprietății de cristal lichid a unor compuși organici

Clasificarea compușilor organici pe baza proprietății lor de cristal lichid (LC) este o problemă complexă care poate fi rezolvată folosind abordări diferite. Pentru modelarea cu SVM, a fost utilizată o bază de date (371 în total) care include o mare varietate de compuși bis-aromatici (-ph-ph-) și azo-aromatici (-ph-n = n-ph-). S-a realizat o analiză de tip structură-proprietate în care proprietatea de cristal lichid a fost corelată cu proprietățile molecular ale acestor compuși: lungime parte rigidă (Lrig), lungime parte flexibilă (Lflex), masa molecular (M) și factorul de simetrie (S). Varianta de implementare GS-SVC a fost folosită pentru clasificarea compușilor, rezultatele diverselor împărțiri a datelor și cu diverse configurații e nuclee sunt afișate în Tabelul 4.1. Pentru a se evidenția importanța optimizării parametrilor modelelor, modelul hibrid cu parametri optimizați cu GS este comparat cu varianta simplă de clasificare SVC. Se observă ca în general nucleele complexe, cum sunt cele polinomiale și Gaussiene, obțin rezultate mult mai bune când sunt optimizate.

Tabelul 4.1. Acuratețea obținută de modelele GS-SVC și SVC cu diverse nuclee.

Nucleu	Model	33%	20%	10%
Linear	SVC	84,8	87	82,1
	GS-SVC	82,3	85,7	89,7
Polinomial	SVC	83,8	85,7	87
	GS-SVC	94,9	94,8	94,9

Gaussian	SVC	82,3	83,1	82,1
	GS-SVC	91,5	88,3	94,9

În plus, pentru a arăta impactul optimizării parametrilor SVM, figurile 4.1 și 4.2 arată cum variază acuratețea modelului odată cu variația celor doi parametri ai nucleului Gaussian(RBF) C și γ care sunt optimizați. Sunt folosite atât datele de antrenare, cât și pentru datele de testare. Linia punctată marchează acuratețea obținută folosind valoare implicită a parametrilor. Se observă o îmbunătățire vizibilă a acurateței mai ales pe datele de antrenare, care de obicei sunt mai numeroase și au un impact mai clar.

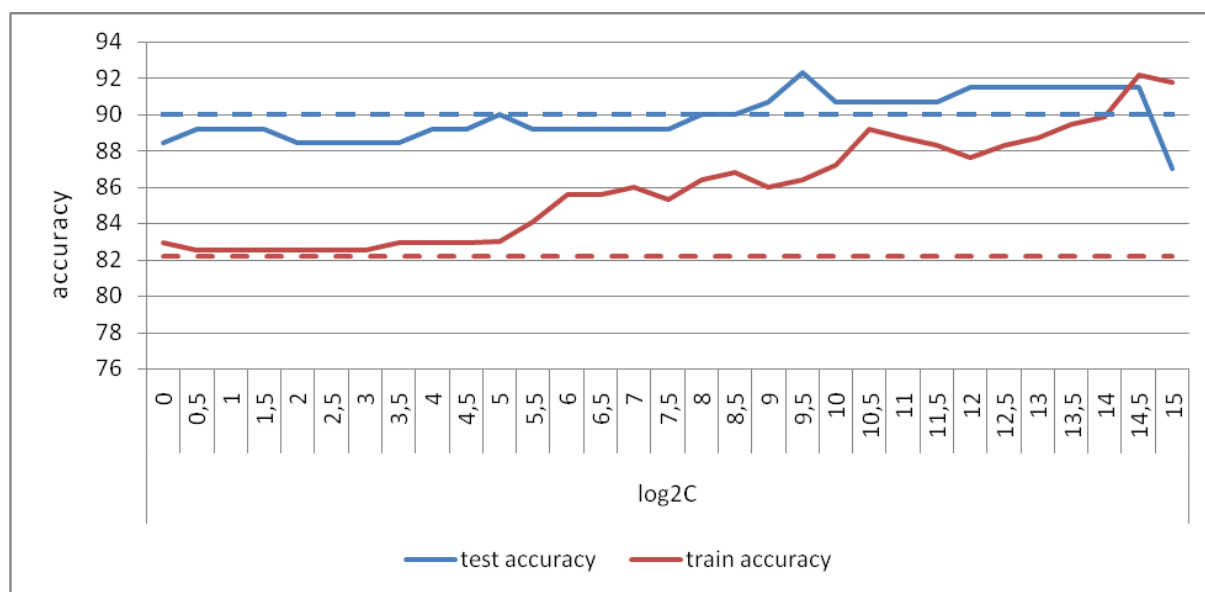


Figura 4.2. Acuratețea obținută când parmetrul C este variabil pentru nucleul RBF.

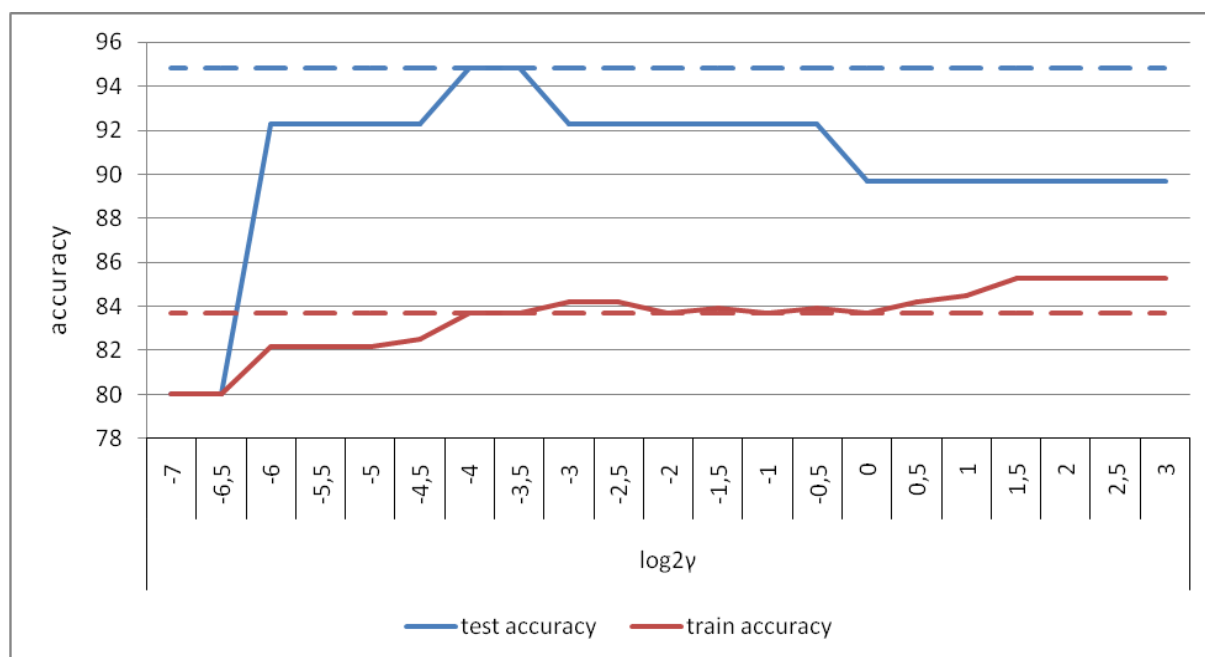


Figura 4.3. Acuratețea obținută când parmetrul γ este variabil pentru nucleul RBF.

În scopul testării algoritmului aplicat (GS-SVC) și în scopul comparării rezultatelor cu cele dintr-un studiu anterior în care s-a folosit modelarea cu rețele neuronale, a fost folosită clasa 2 de compuși ca în Leon *et al.* (2010). Procentele utilizate pentru divizarea bazei de date au fost aceleași pentru ambele metode: 90% pentru datele de antrenare și 10% pentru datele de testare. Tabelul 4.2 arată comparația directă între datele de test estimate cu cele două modele și sunt comparate cu cel experimental. Se observă că modelul SVM obține rezultate mai apropiate de valorile obținute experimental.

Tabelul 4.2. Comparația directă dintre predicțiile modelul SVM și modelul ANN.

Lrig	Lflex	S	M	LC experimental	LC neural network	LC SVM
10.0	17.6	0.108	397	0	1	0
8.6	10.6	0.191	327	0	0	0
10.0	15.8	0.112	369	1	0	1
10.0	12.5	0.129	340	0	0	0
10.0	26.6	0.086	495	0	0	0
9.7	8.9	0.142	291	1	1	1
10.0	18.9	0.109	411	0	0	0
10.0	15.1	0.121	369	0	0	0

Lrig	Lflex	S	M	LC experimental	LC neural network	LC SVM
10.0	29.0	0.079	551	0	0	0
10.0	20.9	0.092	481	0	0	0
7.2	19.3	0.118	445	0	0	0
10.0	20.2	0.101	425	0	0	1
10.0	15.6	0.128	397	0	1	0
10.0	20.2	0.103	425	0	0	0
10.0	18.9	0.107	411	0	0	0
10.0	29.1	0.081	579	0	0	0
10.7	2.6	0.196	291	0	0	0
10.0	20.2	0.107	425	0	0	0
10.0	18.9	0.109	411	0	0	0
7.2	21.8	0.107	473	0	0	0
8.6	14.4	0.185	355	0	0	0
10.0	10.0	0.145	312	0	0	0
10.0	20.8	0.152	525	0	0	0
10.0	19.0	0.130	483	0	0	0
7.2	11.6	0.147	294	0	0	0

Comparația acurateții obținute de modelului hybrid GS-SVC cu alți algoritmi de clasificare este afișată în Tabelul 4.3.

Tabelul 4.3. Comparația modelului SVM cu alți clasificatori.

Split	33%	20%	10%
k-nearest neighbor	85.3	88.5	90.8
C4.5	83.7	86	87.9
Random forest	84.5	90.6	95.3
REP Tree	81.4	87.2	91.6
NNGEP	85.2	91	95.5
GS-SVC	93	93.5	97.4

4.2.2. Predicția rezistenței la coroziune a aliajelor de titan

În acest experiment s-a investigat comportamentul electrochimic a două aliaje de titan ZrTi și NiTiNb în salivă artificială, cu diverse valori ale pH-ului și ale concentrației de NaF, au adăugare de proteină albumină. Rezistența la coroziune s-a evaluat prin măsurarea rezistenței de polarizare. Modelarea rezistenței la coroziune a celor două tipuri de aliaje s-a făcut

mașinile cu suport vectorial, în varianta GS-SVC. Rezultatele de clasificare obținute pentru cele trei tipuri de nuclee, prezentate în Tabelul 4.4, indică o acuratețe medie de peste 90%, ceea ce arată utilitatea practică a modelelor GS-SVC pe partea experimentală.

Tabelul 4.4. Acuratețea obținută de modelul GS-SVC pentru cele două aliaje.

Dataset	Kernel	33%	20%	10%
ZrTi	Linear	87.5	91.6	91.6
	RBF	92.5	100	100
	Polynomial	87.5	87.5	100
NiTiNb	Linear	86.9	90.5	85.7
	RBF	89.9	90.5	85.7
	Polynomial	91.3	85.7	90.5

Modelul hibrid GS-SVC obține o acuratețe superioare față de modelul SVC neoptimizat, și această creștere a performanței este surprinsă în Figura 4.3 pentru nucleul de tip Gaussian (RBF). Concluzia acestui experiment este că modelele hibride obțin rezultate superioare cu nuclee complexe de tip Gaussian(RBF) sau polinomial, în timp ce modelele neoptimizate obțin rezultate bune cu nucleul linear.

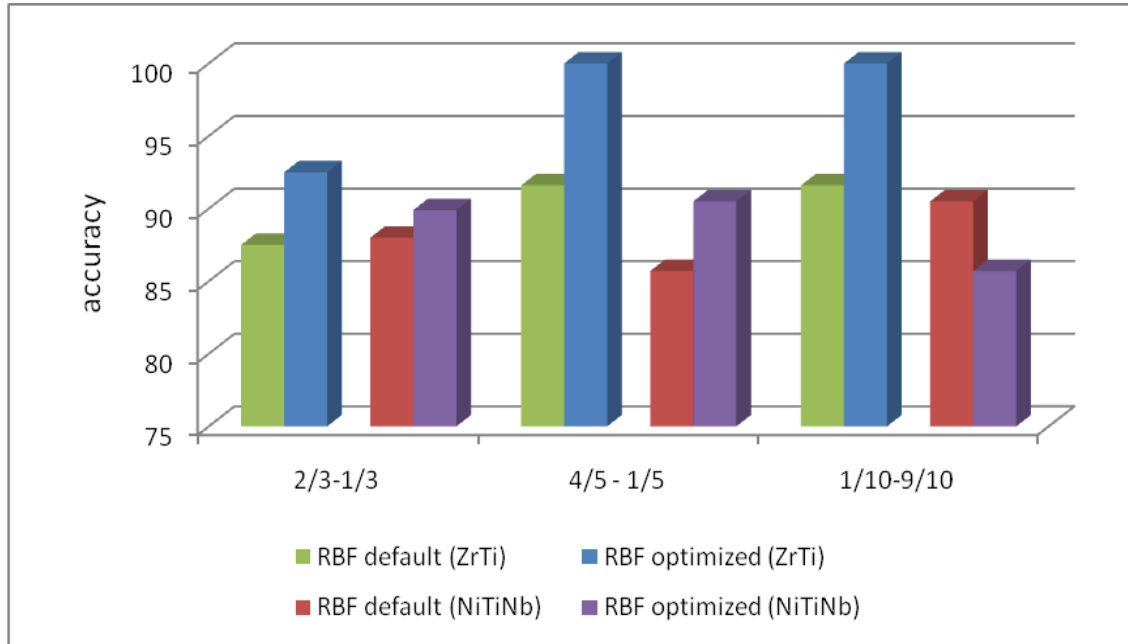


Figura 4.4. Acuratețea obținută pentru cele două aliaje de nucleul RBF cu modelul SVC optimizat și neoptimizat.

4.3 Varianta de implementare GS-SVR și aplicații

Această variantă de implementare este aplicată pentru modelarea problemele de regresie, în care ieșirea este o funcție continuă. În experimentele de regresie s-a folosit varianta SVR din librăria LibSVM, cu cele două implementări ϵ -SVR and μ -SVR, care folosesc parametri de penalizare diferiți: ϵ și μ și nucleul Gaussian (RBF) care s-a dovedit a fi cel mai potrivit pentru modelare.

Indicii de performanță pentru evaluarea problemelor de regresie sunt eroarea medie pătrată și coeficientul de corelare:

$$MSE = \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2 \quad (4.3)$$

$$r^2 = \frac{(N \sum_{i=1}^N f(x_i) y_i - \sum_{i=1}^N f(x_i) \sum_{i=1}^N y_i)^2}{(N \sum_{i=1}^N f(x_i)^2 - (\sum_{i=1}^N f(x_i))^2)(N \sum_{i=1}^N y_i^2 - (\sum_{i=1}^N y_i)^2)} \quad (4.4)$$

unde $f(x_i)$ reprezintă valorile de ieșire ale modelului de regresie SVM, iar y_i reprezintă valorile experimentale. MSE, denumit și riscul de predicție, măsoară eroarea dintre estimările SVM și valorile reale ale funcției obiectiv și reflectă puterea de generalizare. Coeficientul de corelare măsoară cât de apropiate sunt estimările de valorile experimentale.

4.3.1. Purificarea electrochimică a apelor uzate

În cazul procesului modelat, parametrii de intrare sunt reprezentați de: intensitatea curentului, timpul operațiunii, pH-ul, tipul electrodului, precum și concentrațiile inițiale a următorilor parametri: coli form (TC), fical coli form (FC), cererea de oxigen biologic (COD), electroconductivitatea (EC) și cantitatea totală solidă dizolvată (TDS). Există mai mulți parametri de ieșire, reprezentați de concentrațiile finale ale parametrilor TC, FC, COD, EC și TDS.

S-a folosit varianta de implementare GS-SVR, fiind antrenat câte un model SVM antrenat pentru fiecare variabilă de ieșire. Rezultatele obținute cu varianta GS-SVR în care parametrii modelele SVR: C , γ și ε / μ au fost optimizate cu GS sunt afișate în Tabelul 4.5. Pe lângă modele SVR au fost dezvoltate și modele ANN pe același set de date pentru a compara performanța celor două modele. Rezultatele arată ca modelele GS-SVR obțin rezultate asemănătoare cu modelele ANN pe datele de antrenare, dar mai bune pe datele de testare, având o capacitate de generalizare mai bună.

Tabelul 4.5. Rezultatele obținute de modelele GS-SVR și ANN pentru modelarea procesului de electroliză.

Model	Output variable	Model type	Epochs	Training		Testing	
				MSE	r^2	MSE	r^2
1	COD	MLP(7:4:3) ANN_1	400	0.003	0.980	0.225	0.959
		ε -SVR SVR_1a	-	0.01	0.992	0.039	0.977
	EC	MLP(7:4:3) ANN_1	400	0.005	0.988	0.366	0.987
		μ -SVR SVR_1b	-	0.03	0.990	0.09	0.990

Model	Output variable	Model type	Epochs	Training		Testing	
				MSE	r ²	MSE	r ²
	TDS	MLP(7:4:3) ANN_1	400	0.003	0.985	0.45	0.971
		μ-SVR SVR_1c	-	0.007	0.990	0.012	0.988
2	TC	MLP(5:4:4:1)	3000	0.491	0.999	0.608	0.568
		μ-SVR	-	0.1	0.98	0.2	0.92
3	FC	MLP(5:15:1)	1000	0.51	1	0.377	0.531
		μ-SVR	-	0.01	0.98	0.03	0.932

4.3.2. Îndepărtarea metalelor grele din soluții apoase prin bioremediere

Acest experiment cuprinde două etape: mai întâi modelarea cu varianta framework-ului GS-SVR urmată de etapa de optimizare a procesului folosind algoritmul GA.

Experimentul constă în studiul îndepărtării cadmiului Cd(II) din soluțiile apoase prin bioremediere cu fungi. Au fost considerați următorii parametri de intrare: pH-ul soluției inițiale, concentrația sorbent, concentrația inițială de ioni metalici de cadmiu, timpul de contact și temperatura soluției. În urma experimentelor s-au obținut 190 de date pentru modelare.

Acuratețea de modelare cu GS-SVR este dovedită de eroarea mică (mai puțin de 5%) și coeficient de corelare mare (0.89), ilustrat grafic în Figura 4.5. Numărul mic de parametri de optimizat și rezultatele bune dovedesc eficiența modelul GS-SVR pentru modelarea acestui proces.

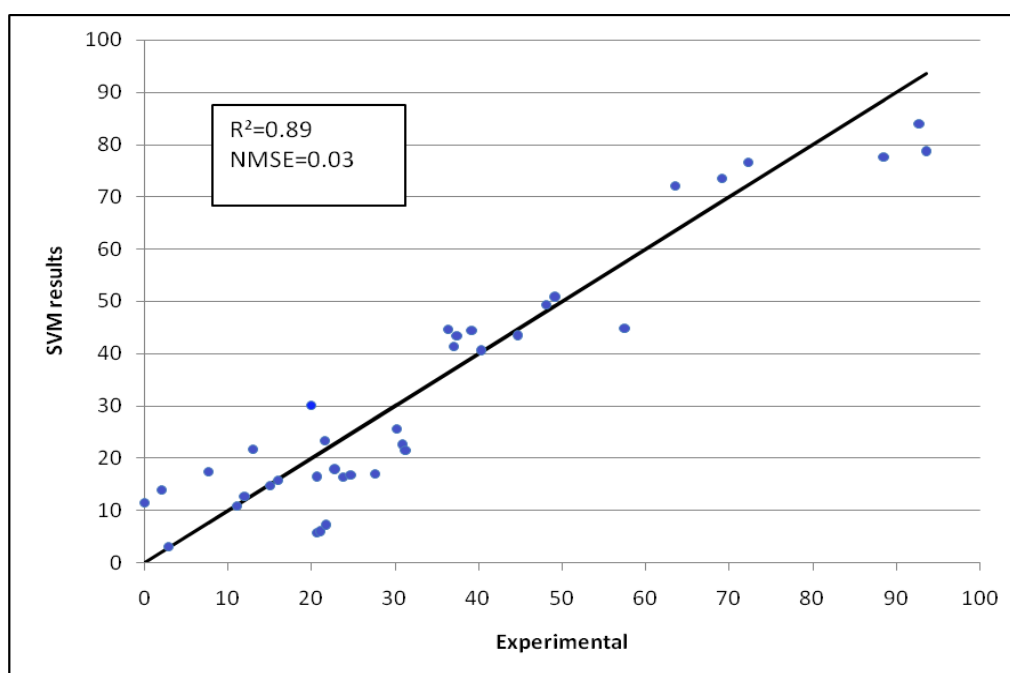


Figura 4.5. Corelarea dintre rezultatele experimentale și cele obținute cu GS-SVR.

Următoarea etapă constă în optimizarea procesului care se realizează prin algoritmului GA ca instrument de optimizare a modelul SVR obținut în prima etapă. Procedura de optimizare GA-SVM determină valorile optime ale variabilelor de decizie, reprezentate ca și intrare în algoritmul GA, care conduc la eficiența maximă a procesului modelat, care va fi funcția obiectiv determinată de modelul SVM.

Tabelul 4.6. Valorile optime ale variabilelor de decizie ale procesului obținute de GA

No.	Population dimension	Number of generations	Tournament dimension	pH	Biomass dosage	Time	Temp.	Initial conc.	Efficiency
1	20	20	2	5.85	8.41	3380	48	36.69	90.45
2	50	20	2	5.63	7.87	3867	46.5	30.60	94.98
3	100	100	2	6.00	8.01	3848	46.5	28.14	98.33
4	100	100	4	6.00	8.00	3833	46.5	26.11	98.91

Tabelul 4.6 arată valorile optime obținute cu diverse configurații ale algoritmului GA: numărul de generații, dimensiunea populației, metoda de selectare, precum și eficiența obținută cu ajutorul modelului SVM.

4.3.3. Extracția de polifenoli din scoarța de molid

În acest experiment s-a modelat procesul de extracție de polifenoli din plante folosind unde electromagnetice. S-a urmărit influența parametrilor precum concentrația de etanol în soluția apoasă, timpul de contact și temperatura asupra cantității totale de polifenoli (TPC) extrasă. Modelare acestui proces s-a realizat cu modelul SVR, care a fost optimizat cu GS și validare încrucișată, obținându-se un model optimal de tip ϵ -SVR, cu nucleul RBF și parametrii $C=10000$ și $\gamma=1$. Figura 4.6 arată o bună corelare între datele experimentale și cele calculate de modelul GS-SVR.

A doua parte experimentală constă în găsirea valorilor optime pentru parametrii de intrare ai procesului. S-au folosit algoritmi genetici pentru optimizare, combinat cu modelul SVR obținut anterior. Funcția de fitness este dată de ieșirea modelului SVR și reprezintă concentrația maximă de polifenoli.

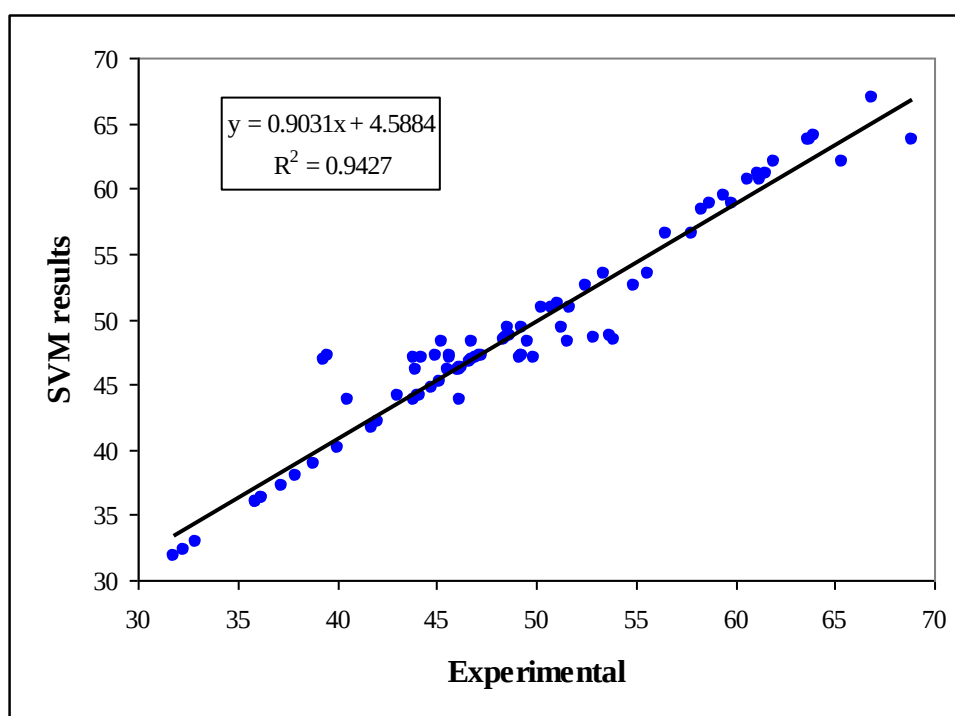


Figura 4.6. Rezultatele SVM și cele experimentale obținute la extracția de polifenoli.

Parametrii de control pentru modelul GA: dimensiunea inițială a populației (pop_dim), numărul de generații (gen_no), probabilitatea de cross-over (cross_prob) și probabilitatea de mutație (mut_prob) au fost determinați experimental prin încercări. Rezultatele obținute pentru diverse setări sunt prezentate în tabelul 4.7. Pentru problemele 2, 3 și 4 au fost impuse restricții adiționale de timp sau temperatură. A fost prezentată o metodă rapidă și eficientă de modelare a procesului de extracție de polifenoli, cele mai bune rezultate fiind obținute folosind o concentrație inițială de 50% de etanol, un timp de extracție de 3 minute și o temperatură de 60 de grade.

Tabelul 4.7. Rezultatele de optimizare obținute pentru patru probleme de optimizare

No. crt.	GA parameters	Problem 1	Problem 2	Problem 3	Problem 4
1	pop_dim = 20 gen_no = 20 cross_prob = 0.95 mut_prob = 0.05	T = 42.51 EtOH = 42.86 t = 25.48 TPC = 48.17	T = 48.77 EtOH = 49.61 t = 4.20 TPC = 48.39	T = 41.98 EtOH = 68.23 t = 16.66 TPC = 48.17	T = 31.04 EtOH = 65.21 t = 7.72 TPC = 48.17
2	pop_dim = 50 gen_no = 20 cross_prob = 0.95 mut_prob = 0.05	T = 49.88 EtOH = 70.52 t = 30.12 TPC = 51.38	T = 60.32 EtOH = 50.11 t = 4.28 TPC = 50.09	T = 41.46 EtOH = 40.81 t = 85.37 TPC = 48.17	T = 44.67 EtOH = 43.11 t = 14.78 TPC = 48.17
3	pop_dim = 100 gen_no = 20 cross_prob = 0.95 mut_prob = 0.05	T = 60.14 EtOH = 70.31 t = 29.55 TPC = 57.29	T = 59.97 EtOH = 49.84 t = 2.99 TPC = 58.07	T = 39.88 EtOH = 49.48 t = 54.83 TPC = 48.39	T = 41.33 EtOH = 70.65 t = 10.34 TPC = 48.17
4	pop_dim = 100 gen_no = 100 cross_prob = 0.95 mut_prob = 0.05	T = 60.01 EtOH = 50.01 t = 30.1 TPC = 63.81	T = 59.95 EtOH = 49.86 t = 3.01 TPC = 58.12	T = 37.36 EtOH = 63.64 t = 44.63 TPC = 48.17	T = 33.40 EtOH = 36.98 t = 11.20 TPC = 48.17
5	pop_dim = 50 gen_no = 150 cross_prob = 0.95 mut_prob = 0.05	T = 59.98 EtOH = 49.99 t = 35.01 TPC = 63.69	T = 60.01 EtOH = 49.91 t = 2.99 TPC = 58.25	T = 39.93 EtOH = 49.96 t = 55.16 TPC = 48.47	T = 43.04 EtOH = 36.23 t = 3.04 TPC = 48.17

În acest capitol au fost prezentate implementări ale framework-ului de tip GS-SVC și GS-SVR și aplicațiile acestora la modelarea proceselor chimice. În următorul capitol sunt prezentate implementări ale framework-ului folosind metode de inspirație biologică combinate cu modele SVR.

5. Algoritmi de optimizare de inspirație biologică pentru modelare proceselor chimice

În acest capitol sunt descrise variante ale framework-ului POM-SVM în care algoritmi de optimizare de inspirație biologică, și anume algoritmul de evoluție diferențială DE-SVM și algoritmul Glowworm Swarm Optimisation (GSO). Acești algoritmi sunt bazați pe populații de indivizi care caută în paralel în spațiul de căutare. Datorită căutării în paralel, acești algoritmi sunt foarte eficienți în găsirea punctului optimal global, cu toate că aceștia prezintă probleme precum convergență prematură sau precizie slabă, mai ales în cazul unor probleme complexe.

5.1. Varianta de implementare DE-SVR și aplicații

În această implementare, modelarea se face cu varianta de regresie a modelului SVM, iar optimizarea cu o variantă auto-adaptivă a algoritmului DE, în care parametrii de control ai algoritmului sunt incluși în procedura de optimizare. Evoluția populației de indivizi se face prin trei operații de bază: mutație, recombinare și selecție, până când este îndeplinită o condiție de terminare. Schema de funcționare a implementării DE-SVR este prezentată în figura 5.1.

La fiecare iterație se evaluează un model SVR cu parametrii din pasul curent, pentru a calcula funcția de fitness a fiecărui cromozom. După terminarea iterațiilor, cromozomul optimal indică setările finale cu care se va construi modelul SVR optimal. Performanța acestui model se va evalua cu eroarea medie pătrată și coeficientul de corelare.

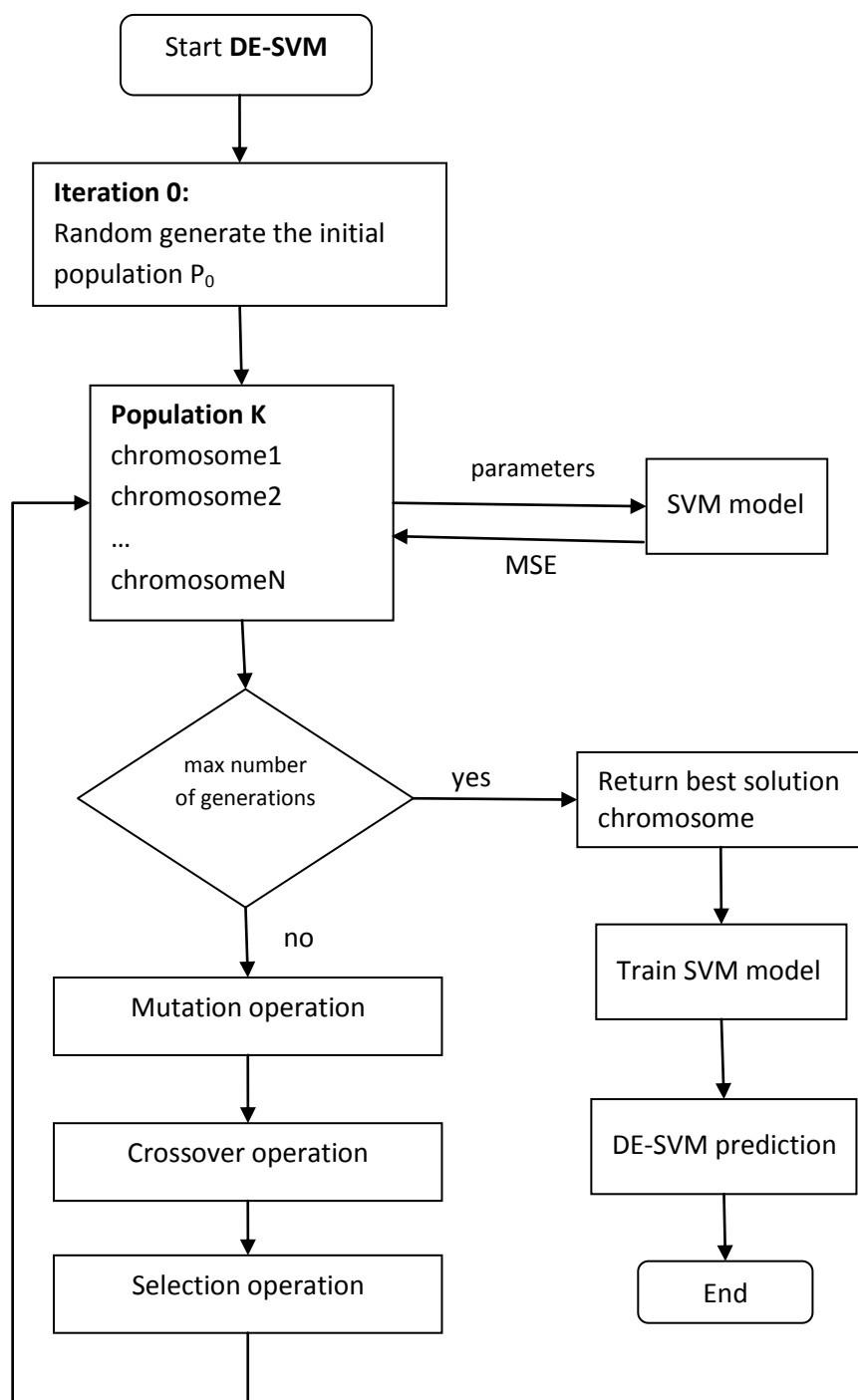


Figura 5.1. Schema de funcționare a procedurii DE-SVR

5.1.1. Polimerizarea radicalică a stirenului

Pentru procesul de polimerizare prin radicali liberi a stirenului – a fost utilizat un model matematic complet, bazat pe ecuații de conservare aplicate elementelor din amestecul de reacție, iar pentru rezolvare s-a folosit metoda momentelor de distribuție a concentrațiilor (Curteanu, 2003). Pe baza acestui model, a fost generată o bază de date completă formată din 3494 exemplare (simulare pe modelul cinetic). Mașinile cu suport vectorial optimizate prin metodologia DE-SVR (Tabelul 5.1) au fost folosite pentru predicția conversiei monomerului și a maselor moleculare, în funcție de concentrația de inițiator, temperatură și timp de reacție. În plus, s-a folosit pentru comparație rezultate furnizate de modelele cu suport vectorial determinate de varianta GS-SVR. O comparație vizuală a rezultatelor obținute de celor două modele pentru parametrii de ieșire se poate vedea în figurile 5.2 și 5.3.

Tabelul 5.1. Rezultatele de modelare obținute pentru procesul de polimerizare

Output variable	Model type	Training		Testing	
		MSE	r ²	MSE	r ²
x	DE-SVR	0.0085	0.9714	0.0075	0.9656
	GS-SVR	0.004	0.96	0.0045	0.93
Mn	DE-SVR	0.0010	0.9936	0.0014	0.9767
	GS-SVR	0.09	0.997	0.27	0.981

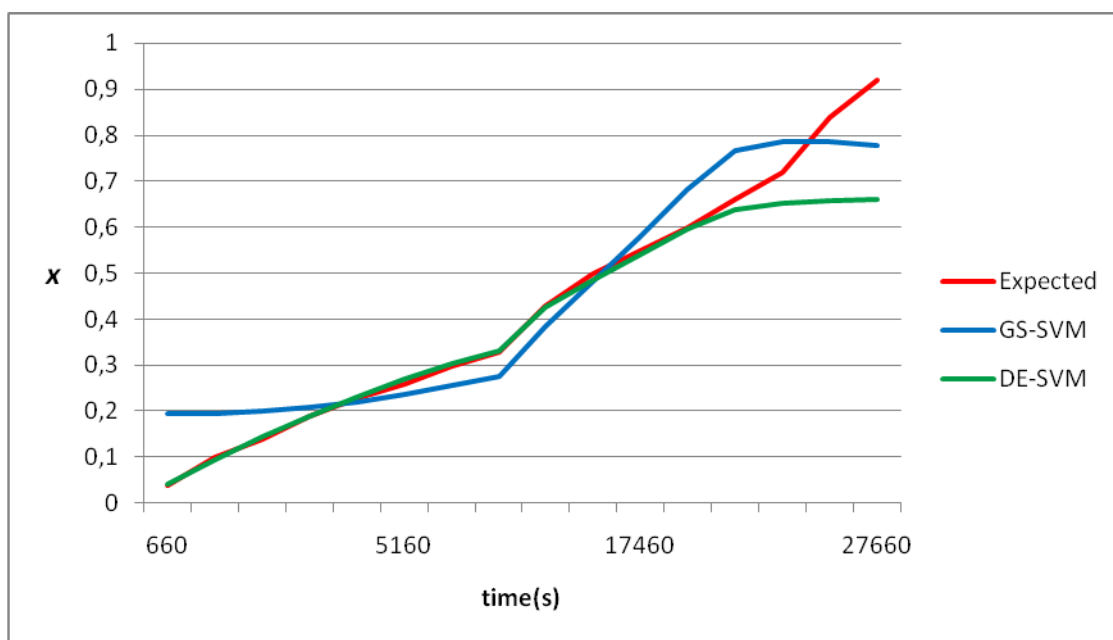


Figura 5.2. Comparația dintre predicția conversiei monomerului (variabila x) cu modelele GS-SVM și DE-SVM și datele experimentale când parametrii procesului sunt $T = 368K$ și $I_0 = 10 \text{ mol/l}$.

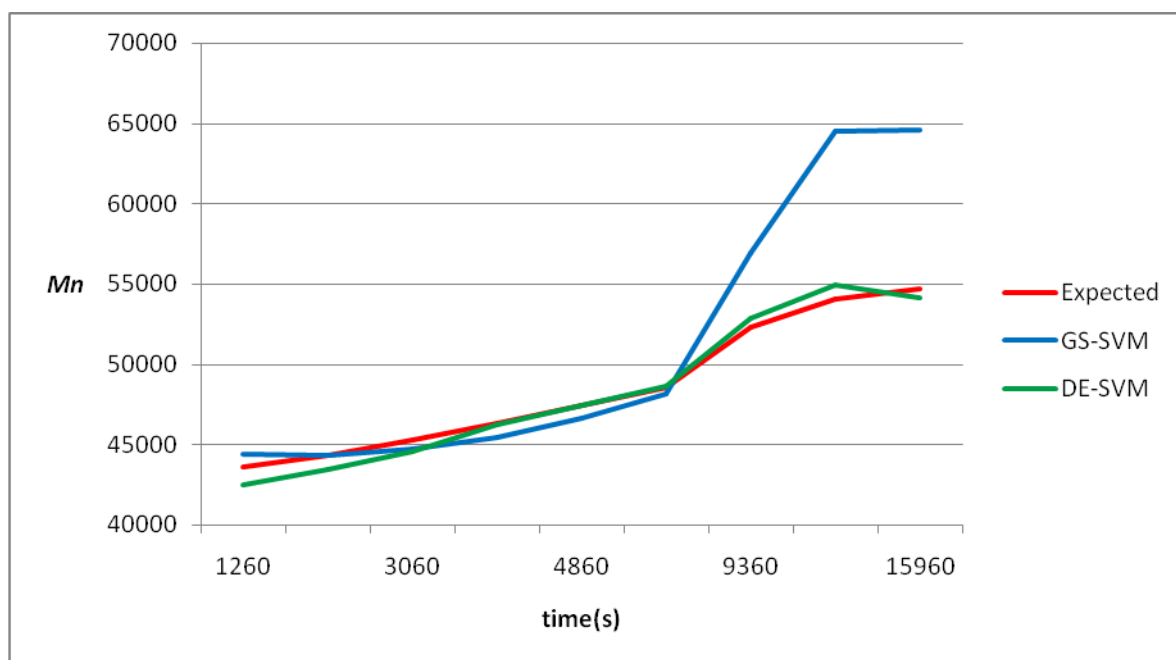


Figura 5.3. Comparația dintre predicția masei molecular (variabila M_n) cu modelele GS-SVM și DE-SVM și datele experimentale când parametrii procesului sunt $T = 383K$ and $I_0 = 20 \text{ mol/l}$ (valoare inițială a inițiatorului).

5.2. Varianta de implementare GSO-SVR și aplicații

Algoritmul GSO reprezintă un algoritm de optimizare de inspirație biologică bazat pe populații de indivizi care imită comportamentul inteligent de grup al licuricilor care sunt atrași de indivizii cu bioluminescența cea mai intensă. Indivizii se vor grupa în jurul valorilor extreme locale sau globale.

Varianta GSO-SVR este bazată pe o implementare proprie a algoritmului GSO în JAVA, deoarece acesta nu este inclus în pachetele software standard. Implementarea este ușor combinată cu varianta SVR implementată în pachetul LibSVM.

5.2.1. Implementarea algoritmului GSO

Principalii pași ai algoritmului sunt descriși în Figura 5.5. Inițial, populația de N agenți este împrăștiată aleator în spațiul de căutare. Spațiului de căutare este definit de valorile variabilelor de intrare ale funcției obiectiv J . Starea unui agent i la timpul t poate fi definit astfel: o poziție în spațiul de căutare $x_i(t)$, o valoare a luminozității $l_i(t)$, și o rază de vizibilitate $r_d^i(t)$. La fiecare iterație, starea unui agent se schimbă în cei trei pași indicați mai jos: actualizarea luminozității, schimbarea locației și modificarea razei de vizibilitate.

În faza de actualizare a luminozității (*Luciferin-update phase*), fiecare agent își calculează luminozitatea în funcție de noua poziție. La început toți agenții au aceeași valoare a luminozității, dar pe măsură ce aceștia se mișcă în spațiul de căutare, vor actualiza luminozitatea în funcție de valoarea funcției obiectiv în acea poziție, după formula

$$l_i(t) = \max\{0, (1 - \rho)l_i(t-1) + \gamma J_i(t)\} \quad (5.1)$$

unde $l_i(t)$ reprezintă valoarea luminozității asociată agentului i la timpul t , ρ este o constantă de scădere $0 < \rho < 1$, γ este o constantă de îmbunătățire, iar $J_i(t)$ reprezintă valoarea funcției obiectiv pentru locația agentului i la timpul t . În faza de schimbare a locației (*Movement*

phase), fiecare agent se mișcă spre agentul aflat în raza de vizibilitate cu cea mai mare luminozitate. Selectare agentului țintă se face printr-un mecanism de probabilități bazat pe valoarea luminozității, după formula:

$$P_{ij}(t) = \frac{l_j(t) - l_i(t)}{\sum_{k \in N_i(t)} l_k(t) - l_j(t)} \quad (5.2)$$

unde $j \in N_i(t)$, $N_i(t) = \{j : d_{ij}(t) < r_d^i(t), l_i(t) < l_j(t)\}$ reprezintă setul de agenți aflați în vecinătatea agentului i la timpul t , $d_{ij}(t)$ reprezintă distanța euclidiană dintre agenții i și j la timpul t și $r_d^i(t)$ reprezintă raza de vizibilitate a agentului i la timpul t . Ecuația de mișcare a agențului i către agentul j este:

$$x_i(t+1) = x_i(t) + s \frac{x_j(t) - x_i(t)}{\|x_j(t) - x_i(t)\|} \quad (5.3)$$

unde s este o constantă ce reprezintă dimensiunea pasului.

Ultimul pas este modificarea razei de vizibilitate (*neighborhood range update*), în care fiecare agent își modifică raza de vizibilitate folosind ecuația:

$$r_d^i(t) = \min\{r_s, \max\{0, r_d^i(t-1) + \beta(n_t - |N_i(t-1)|)\}\} \quad (5.4)$$

unde β este o constantă și n_t este un parametru folosit pentru a controla numărul de vecini.

Algoritmul GSO definește o rază de vizibilitate dinamică, $0 < r_d^i < r_s$, unde r_s este o constant ce definește valoarea maximă. Motivația folosirii unei raze dinamice și nu fixă este că agenții folosesc doar informații locale pentru a decide schimbarea poziției. Agenții se vor grupa în jurul punctelor de extrem local sau final, numărul de extreme depinzând și de setările constantelor, de exemplu a valorii maxime a razei de vizibilitate. În cazuri extreme aceasta poate determina mișcarea tuturor agenților spre punctul optimal global. De aceea folosirea unei raze de vizibilitate dinamice poate duce la detectarea tuturor punctelor de optim local sau global. Pentru a evalua performanța algoritmului, două noțiuni sunt introduse: numărul de iterații pentru convergență, și numărul de valori optime găsite. Spunem că un punct de optim a fost găsit dacă cel puțin trei agenți se află la o distanță minimală față de acel punct.

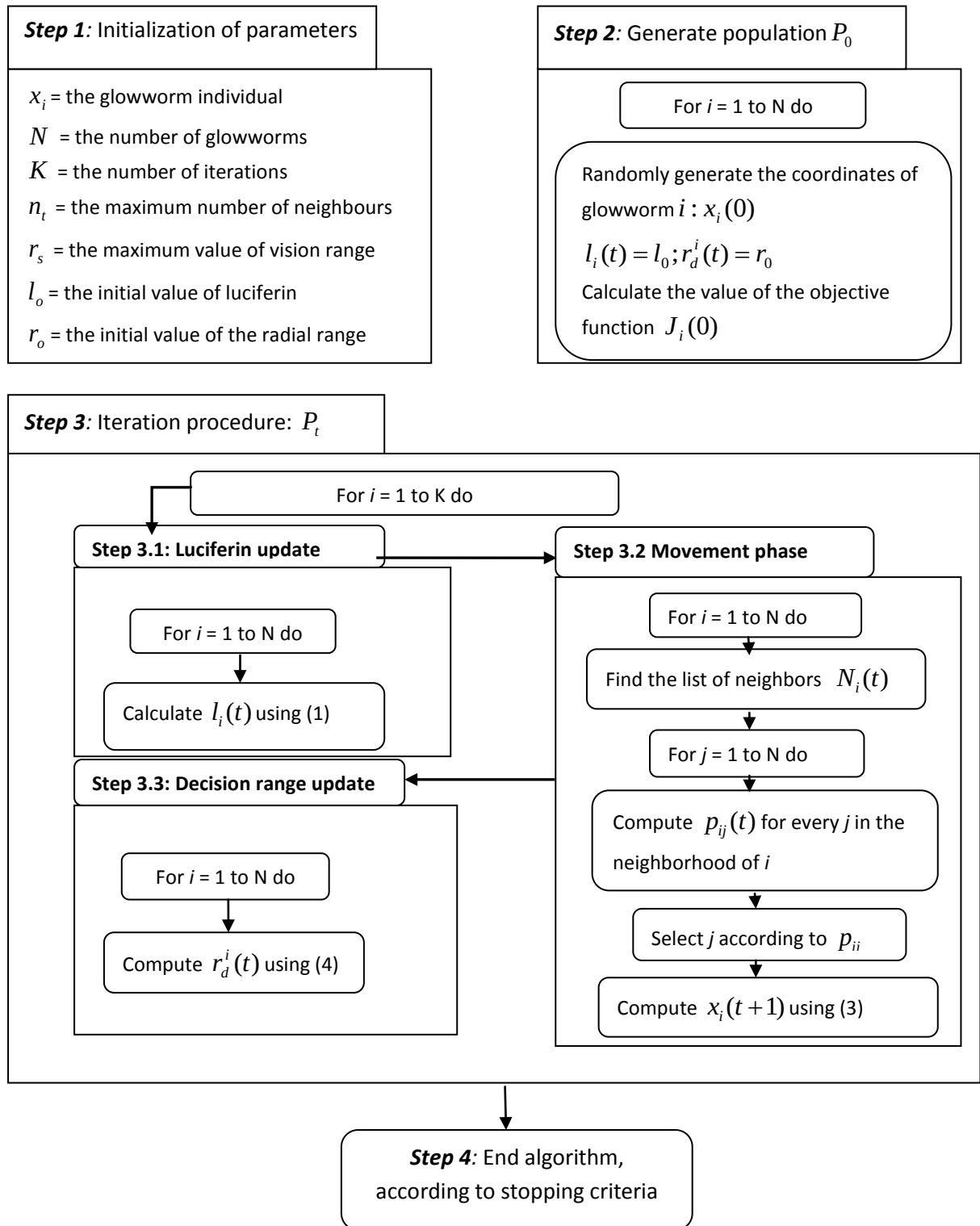


Figura 5.4. Principalii pași în algoritmul de optimizare GSO

5.3.2. Implementarea variantei GSO-SVR

În această variantă de implementare, GSO are rolul de a optimiza parametrii algoritmului SVR. Variabilele de intrare în algoritm sunt parametrii SVM care afectează capacitatea de generalizare a acestuia: parametrul de cost C , care reprezintă o balanță între capacitatea de generalizare și erorile la antrenare, precum și parametrii nucleului. Fiecare tip de nucleu are proprii parametri. În cazul nucleului Gaussian (RBF) există un singur parametru de optimizat, folosit în experimentele din următoarele secțiuni, parametrul γ . Parametrii SVM definesc spațiul de căutare al algoritmului GSO.

Pașii de implementare ai algoritmului GSO-SVM sunt redați în figura 5.5. Procedura GSO-SVM începe prin inițializarea parametrilor GSO cum ar fi: dimensiunea populației, valoarea inițială a luminozității, raza de vizibilitate inițială. Apoi se inițializează aleator pozițiile agenților în spațiul de căutare. Pasul al doilea este evaluarea funcției de fitness pentru fiecare agent, pe baza valorii MSE a algoritmului SVM calculată la poziția curentă. Pasul iterativ constă în stabilirea agenților locali cu luminozitate maximă prin compararea funcției fitness a fiecărui agent cu agenții aflați în vecinătate. Urmează schimbarea poziției agenților către cei selectați și actualizarea luminozității și a razei de vizibilitate. Algoritmul se încheie când numărul maxim de iterații este atins sau nu mai există schimbări de poziții în populație. Aceste mișcări permite agenților să se împartă în grupuri disjuncte în jurul pozițiilor optime. După convergență, poziția sau pozițiile globale sunt determinate, iar acestea vor fi folosite pentru antrenarea modelului SVM optimal final.

Funcția obiectiv (de fitness) pentru optimizarea GSO este bazată pe eroarea pătratică medie MSE:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - f(x_i/\phi_i(C, \gamma)))^2 = F(C, \gamma) \quad (5.4)$$

unde (C, γ) reprezintă parametrii SVM. Pentru un set de antrenare, MSE reprezintă o funcție $F(C, \gamma)$ având ca argumente parametrii C și γ , care reflectă capacitatea de generalizare a procedurii GSO-SVM framework.

Pentru evaluarea performanței algoritmului de optimizare GSO se ia în considerare în primul rând convergența algoritmului (găsierea cel puțin a unei valori optimale), dar și viteza de convergență, care este definită ca numărul (unic) de apeluri la modulul SVM necesare pentru convergența algoritmului.

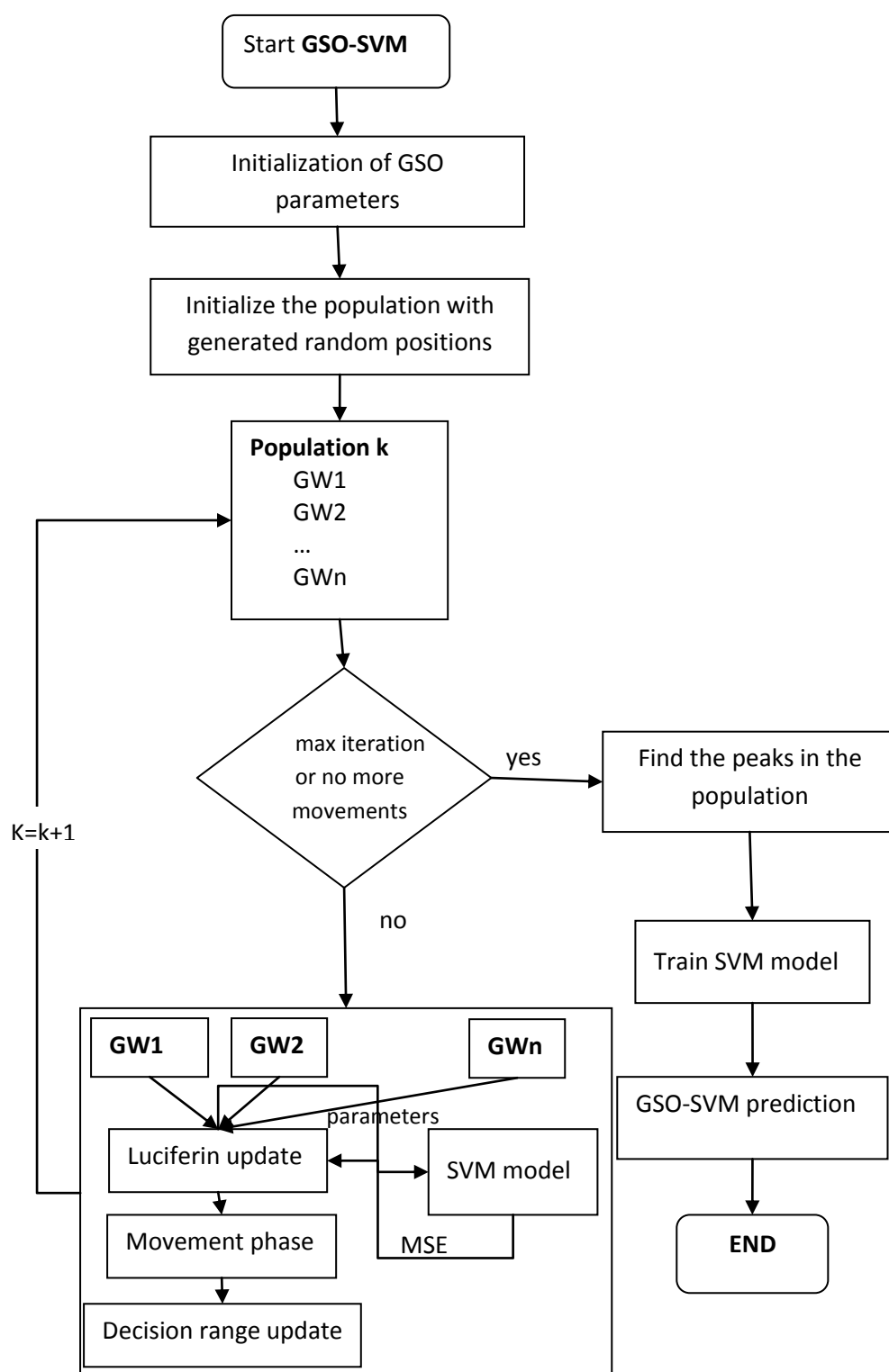


Figura 5.5. Implementarea modelului GSO-SVM

5.3.3. Îndepărtarea metalelor grele prin biosorpție cu alge

În acest experiment, în urma procedurii de biosorpție a trei tipuri de metale cu ajutorul algelor, s-au format trei seturi de date, corepunzătoare fiecărui metal: setul de date Cd(II) cu 50 de exemplare, setul Co(II) cu 48 de exemplare, și setul Pb(II) cu 47 de exemplare. Sunt cinci variabile de intrare în modelul GSO-SVM: pH soluție inițială, concentrație sorbent, concentrație inițială ioni metalici (cadmiu, cobalt, plumb), timp de contact, temperatură de lucru. Variabila de ieșire este reprezentată de cantitatea de ioni (exprimată în procente) absorbiți pentru fiecare metal. Datele au fost împărțite aleator în 80% date de antrenare, 10% de validare și 10% de testare.

După setarea constantelor GSO, procedura GSO-SVM stabilește valorile posibile ale parametrilor SVM pentru a define spațiul de căutare: $(2^{-5}, 2^{15})$ pentru C și $(2^{-5}, 2^3)$ pentru γ , așa cum se recomandă în librăria LibSVM. Astfel se formează un spațiu de căutare bidimensional care poate fi ușor exploatat prin combinarea valorilor incrementate exponențial pentru C și γ , obținând un spațiu de căutare $[-5, 15] \times [-5, 3]$ prin folosirea valorilor logaritmice.

Valorile optime pentru C și γ obținute pe datele experimentale de antrenare și test, precum și eroarea obținută de procedura GSO-SVM, precum și coeficientul de corelare (r^2) sunt raportate în tabelele 5.2, 5.3 și 5.4. Pentru comparație sunt afișate și rezultatele obținute de varianta GS-SVM. Rezultatele returnate de cele două variante sunt similare, valorile parametrilor fiind apropiate, la fel și eroarea și coeficientul de corelare. Aceasta arată că algoritmul de optimizare GSO obține rezultate foarte bune, comparabile cu un algoritm exhaustiv de căutare. Avantajul GSO constă în găsirea mult mai rapidă a acestor valori optime pentru parametrii SVM. Aceasta se reflectă în viteza de convergență.

Conform cu [Krishnanand and Ghose \[2009\]](#), valoarea luminozității agenților co-locați la punctul optimal X_i au aceeași valoare de fitness I_i . Acest rezultat este valabil și în cazul experimentelor efectuate, așa cum se poate vizualiza în figura 5.6. În acest caz, la iterația 160, valorile luminozităților tuturor agenților converg către valoarea luminozității agentului aflat în punctul optimal.

Tabelul 5.2. Rezultatele obținute de GSO-SVM și GS-SVM pentru setul **Cd(II)**

Algorithm	Data	Parameters C: γ	MSE validation	r^2 validation	MSE test	r^2 test	Number of SV
GSO	split1	3451: 0.35	26.6	0.78	3	0.99	35
	split2	258: 0.75	14.97	0.91	2.66	0.96	33
	split3	1515: 0.26	10.5	0.947	35.2	0.82	34
	Average	-	17.35	0.879	13.62	0.923	34
GS	split1	3566: 0.35	26.6	0.78	3	0.99	35
	split2	239: 0.81	15.2	0.91	3.3	0.95	33
	split3	1911: 0.233	10.8	0.94	32.8	0.83	35
	Average	-	17.53	0.87	13.03	0.923	

Table 5.3. Rezultatele obținute de GSO-SVM și GS-SVM pentru setul **Co(II)**

Co(II)	Model	MSE validation	r^2 validation	MSE test	r^2 test	C	γ	Number of SV
split1	GS	0.956	0.95	6.3	0.94	19.7	0.287	23
	GSO	0.95	0.95	6.3	0.94	19.5	0.29	23
split2	GS	4.479	0.97	5.2	0.976	128	0.66	30
	GSO	4.4	0.972	6.1	0.973	217	0.48	29
split3	GS	2.18	0.991	16.5	0.94	1260.7	0.66	35
	GSO	2.16	0.992	18.4	0.93	1337	0.71	36
average	GS	2.538	0.97	9.33	0.952	-	-	-
	GSO	2.5	0.971	10.26	0.947	-	-	-

Table 5.4. Rezultatele obținute de GSO-SVM și GS-SVM pentru setul **Pb(II)**

Pb(II)	Model	MSE validation	r^2 validation	MSE test	r^2 test	C	γ	Number of SV
split1	GS	5.5	0.996	64.8	0.97	55.7	0.287	26
	GSO	5.2	0.996	59.5	0.976	68	0.252	26
split2	GS	0.139	0.999	0.578	0.999	8198	0.66	35
	GSO	0.139	0.999	0.578	0.999	8743.5	0.66	35
split3	GS	14	0.96	4	0.95	1260.7	0.8	35
	GSO	14.3	0.96	5.37	0.93	8416.5	0.61	35
average	GS	6.646	0.985	23.12	0.973	-	-	-
	GSO	6.546	0.985	21.81	0.968	-	-	-

S-a comparat viteza de convergență a modelelor GSO-SVM cu modelul GS-SVM, care prezintă un număr invariabil de combinații posibile ale valorilor parametrilor: 67 de valori posibile pentru C și 50 pentru γ , rezultând un total de 3350 de perechi distincte (C, γ) pentru care valoarea MSE trebuie calculată. Tabelul 5.7 prezintă rata de convergență a

modelului GSO-SVM pentru toate seturile de date, calculată ca totalitatea apelurilor la procedura SVM pentru calcularea funcției fitness.

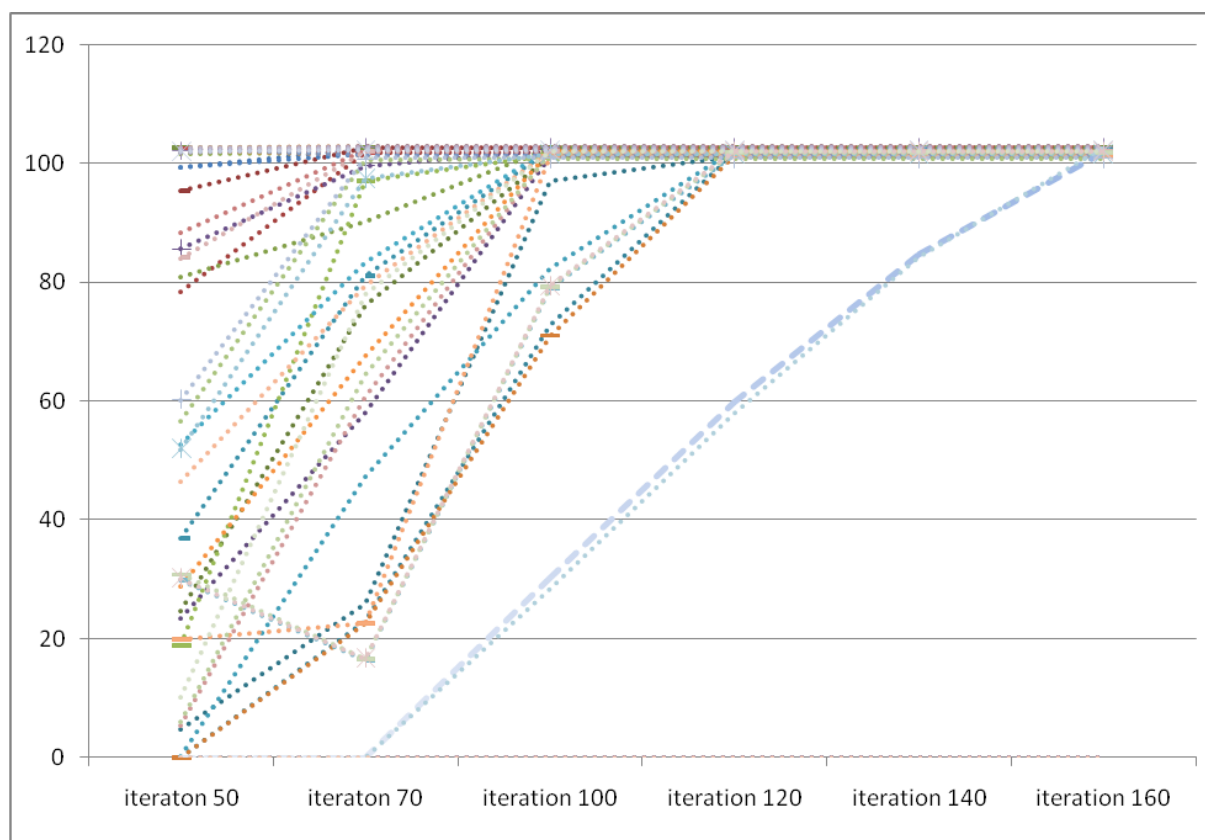


Figura 5.6. Convergența luminozității agenților de-a lungul iterațiilor în modelul GSO-SVM pentru experimentele cu setul Co(II)

Sunt prezentate rezultatele experimentelor cu dimensiunile pașilor de 0.1, 0.2 sau 0.3, cu populații de dimensiuni variabile și numărul minim de iterații până sunt îndeplinite criteriile de convergență.

Table 5.7. Rata de convergență pentru algoritmul GSO-SVM pentru cele trei seturi de date

Model	Data	No. of glowworms	Step size	Convergence rate	No. of iterations to convergence
GSO-SVM	Cd(II) – split1	70	0.3	790	60
	Cd(II) – split2	50	0.1	2020	70
	Cd(II) – split3	50	0.2	755	50
		70	0.3	1285	60
	Co(II) – split1	50	0.3	901	25
	Co(II) – split2	50	0.3	740	40
	Co(II) – split3	50	0.3	1239	65
	Pb(II) – split1	60	0.3	1079	35
	Pb(II) – split2	70	0.3	874	30
	Pb(II) – split3	50	0.3	823	30
GS-SVM	All datasets	-	0.3	3350	-

5.3.4 Predicția coeficientului de partiționare în sisteme apoase bifazice

În acest experiment, modelul GSO-SVM este utilizat pentru predicția coeficientului de partiționare al clorhidratul de guanină folosit ca solvent într-un sistem apos bifazic. Intrările algoritmului sunt procentul de fosfat din soluție, care variază între 0 și 50 (% w/w), concentrația clorhidratului de guanină, care variază între 10 și 30 (% w/w), și pH-ul soluției care poate avea valorile: 7.2, 9.1, și 10.8. Ieșirea o reprezintă coeficientul de partiționare al clorhidratului de guanină.

În urma experimentului s-a format un set de 149 de date, dintre care 75% sunt folosite pentru antrenare și 25% pentru testare. În faza de optimizare, modelul GSO-SVR selectează parametrilor optimali ai modelului SVR de către GSO. În acest caz spațiul de căutare pentru parametrilor (C , γ) este (2^{-5} , 2^{10}) pentru C și (2^{-3} , 2^3) pentru γ . Evaluarea funcției de fitness a fiecărui agent în faza de antrenare se face prin validarea încrucișată multi-strat. Stabilirea dimensiunii populației și a numărului de iterații se face manual prin încercări repetate. În experimentele curente nu s-au folosi tmai mult de 30 de agenți și 30 de iterații. La final se vor selecta parametrilor care reprezintă coordonatele agenților cu valoare de fitness maximală. Cu aceștia se antrenează modelul final și se evaluează pe datele de test. Rezultatele obtinute sunt afișate în tabelul 5.8. Pentru comparație, rezultatele obținute cu varianta GS-SVR a frameworkului, dar și cu două variante DE-NN sunt raportate. Evaluarea se face cu eroarea medie pătratică (MSE) și eroarea medie (ARE).

Tabelul 5.8. Rezultatele cele mai bune obținute cu modelele GSO-SVM, GS-SVM și
DE-NN, DE-NN2

Algorithm	Parameters	Training		Testing	
		ARE %	MSE	ARE %	MSE
GSO-SVM	C=32.87, $\gamma=0.0678$	1,17	2.21E-04	1.35	2.6E-04
	C=4, $\gamma=4$	0.65	1.37E-04	1.55	4.14E-04
	C=0.45, $\gamma=0.04$	1.26	2.56E-04, $r^2=0.89$	1.436	2.88E-04, $r^2=0.85$
	C= 0.27, $\gamma=$ 0.59	1.08	2.02E-04, $r^2=0.91$	1.34	2.85E-04, $r^2=0.85$
	C=4, $\gamma=0.12$	1.17	2.21E-04, $r^2=0.9$	1.36	2.68E-04, $r^2=0.86$
GS-SVM	Step=1 C=0.5, $\gamma=0.031$	1.28	2.57, $r^2=0.89$	1.46	2.89, $r^2=0.85$
	Step=0.5 C=4, $\gamma=0.044$	1.24	2.46, $r^2=0.89$	1.45	3.03, $r^2=0.85$
DE-NN	3:19:1	0.9528	1.69E-04	1.1923	2.21E-04
DE-NN-2	3:18:1	0.9345	1.58E-04	1.3134	2.63E-04

Se observă că modelul optimal GSO-SVM (**C= 0.27, $\gamma= 0.59$**) obține rezultate superioare modelului optimal GS-SVM (**C=4, $\gamma=0.044$**) atât în faza de antrenare, cât și la testare. Aceasta se poate observa analizând valorile metricelor MSE și ARE din tabelul de mai sus. Diferența dintre ele nu este semnificativă, dar ținând cont și de faptul că modelul GS-SVM este de câteva ori mai costisitor, folosirea algoritmului biologic GSO pentru selectarea modelului optimal SVM este recomandată. De asemenea GSO-SVM este comparabil cu DE-NN, rezultatele lor fiind similare, însă acesta din urmă presupune o complexitate mai mare a algoritmilor implicați, dar și o flexibilitate mai mare, ceea ce poate face diferența la selectarea modelului potrivit. Aceste concluzii nu sunt general valabile, dar specifice modelării procesului de identificare a coeficientului de partiționare a clorhidratului de guanină. În figurile 5.7 și 5.8 sunt comparate punctual rezultatele obținute pe datele de test a celor trei modele în comparație cu rezultatele experimentale.

Acuratețea predicțiilor modelului GSO-SVM în modelarea celor două procese chimice chimice prezentate îl recomandă ca instrument în ingineria proceselor chimice, fiind de ajutor

în înlocuirea, sau cel puțin asistarea experimentelor. Studii de caz prezentate arată generalitatea framework-ului POM-SVM utilizând atât modele combinatoriale cât și algoritmi de inspirație biologic, fiind utile în modelarea altor procese, cu șanse mari de a obține rezultate optime.

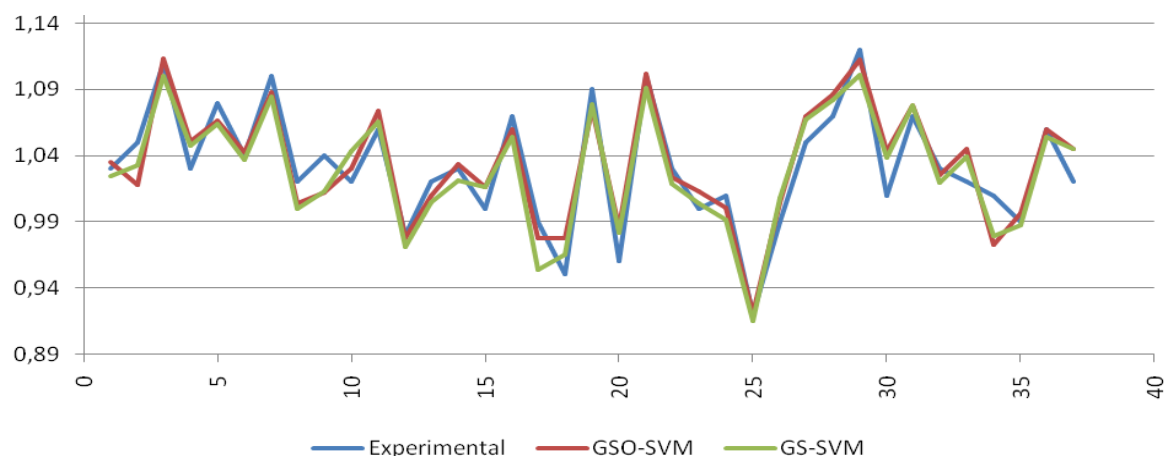


Figura 5.7. Comparația între datele experimentale și predicțiile modelelor GSO-SVM și GS-SVM în faza de testare.

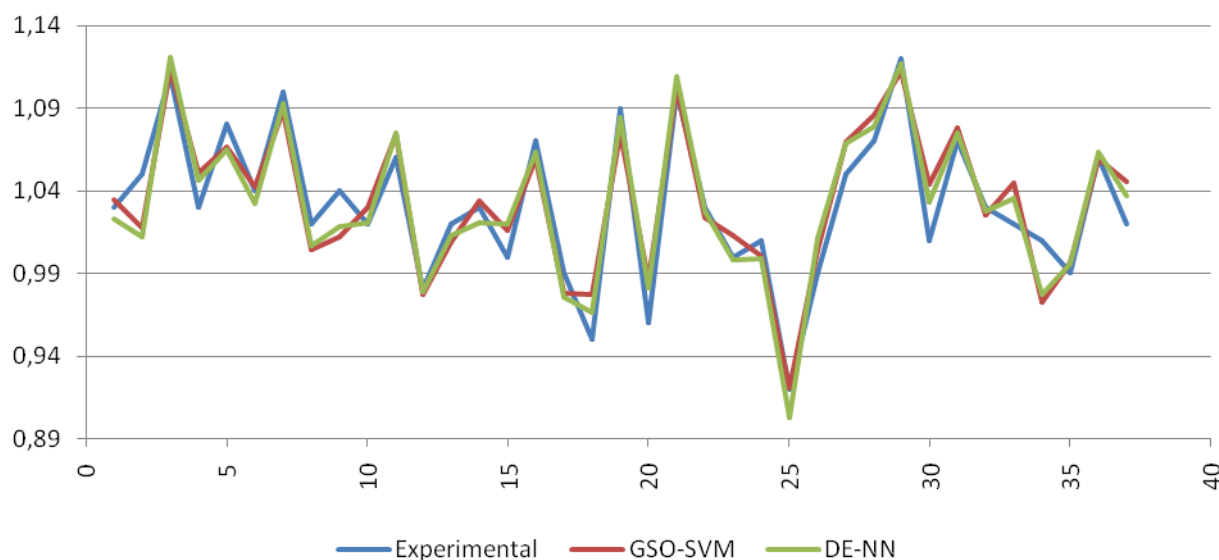


Figura 5.7. Comparația între datele experimentale și predicțiile modelelor GSO-SVM și DE-NN în faza de testare.

6. Concluzii finale

6.1 Concluzii referitoare la obiectivele tezei

Mașinile cu suport vectorial reprezintă unul din instrumentele de bază folosite în ingineria chimică, după metoda celor mai mici pătrate și rețele neuronale. În această teză este propus un framework general POM-SVM care include multiple tehnici de optimizare ale inteligenței artificiale pentru dezvoltarea de modele eficiente și procese chimice optime. Tehnicile generale folosite în această teză sunt descrise în continuare.

Mașinile cu suport vectorial (SVM) sunt aplicate în modelarea, predicția și clasificarea diferitelor procese din ingineria chimică.

- **Grid Search (GS)** folosit ca un algoritm de optimizare a parametrilor, în combinație cu modelul SVM pentru clasificare sau regresie.
- **Differential evolution (DE)**, care aparține clasei algoritmilor de inspirație biologică, folosit ca o metodă avansată pentru selectarea modelului SVM optimal.
- **Algoritmii genetici (GA)**, reprezintă o clasă de algoritmi evolutivi de inspirație biologică, utilizați pentru optimizarea proceselor chimice, după optimizarea procesului SVM.
- **Glowworm swarm optimisation (GSO)**, un algoritm tot de inspirație biologică, care aparține clasei algoritmilor de inteligență de grup, este aplicat pentru optimizarea avansată a parametrilor modelelor SVM.
- **Algoritmi hibridi.** Prin combinarea algoritmilor SVM, GS, GA, DE și GSO cu diverse configurații, se obțin modele hibride noi și eficiente bazate pe modele kernel. Variantele de hibridizare utilizate și implementate în această teză (GS-SVC, GS-SVR, GA-SVR, DE-SVR, GSO-SVR) rețin caracteristicile fiecărui algoritm, și sunt implementate încât să fie reduse la minimum limitările și dezavantajele acestora. Aceste modele hibride sunt folosite cu scopul de a obține modele optimizate, dar și pentru optimizarea proceselor. În aceste combinații, algoritmul SVM are rolul de

modelarea a proceselor, iar ceilalți algoritmi sunt folosiți în scopul de optimizatori de model sau process.

Prin folosirea acestor instrumente în diverse combinații și abordări, următoarele obiective sunt îndeplinite:

➤ ***Găsirea unui principiu general pentru optimizarea modelelor bazate pe mașinile cu support vectorial.*** Toate metodologiile prezentate în această teză sunt bazate pe același principiu de utilizare a unui algoritm de optimizare pentru a determina parametrii optimali și funcția kernel potrivită, pentru a modela procesele chimice. Deși idea de a utiliza un algoritm de căutare pentru găsirea unui model SVM optimal nu este nouă, implementările existente în ingineria chimică sunt dependente de problemele abordate. În această teză, toate implementările propuse au un caracter general și sunt pot fi aplicate modelării oricărui process, atât timp cât sunt respectate următoarele condiții asupra datelor experimentale: existența unei relații de dependență între datele de intrare și cele de ieșire, și convergența uniformă a datelor existente în spațiului de căutare al parametrilor. În implementare modelelor prezentate sunt folosite cele două tipuri de mașini cu support vectorial: clasificare și regresie, în funcție de tipul de problemă abordată. Alt aspect constă în tipul de nucleu folosit și a parametrilor incluși în procesul de optimizare (parametrul de cost și parametrul γ).

➤ ***Utilizarea variantei SVM de clasificare pentru problem de găsire de șabloane.*** Această metodologie este folosită pentru problem cum ar fi discriminarea între aliajele de titan rezistente și nerezistente folosite ca material dental în aplicații biomedicale sau identificarea proprietății de cristal lichid a unor compuși organic bazat pe o analiză structurală și geometrică a proprietăților moleculare.

➤ ***Utilizarea variantei SVM de regresie pentru problemele de aproximare a funcției obiectiv.*** Această metodologie este utilizată în problem precum cuantificarea îndepărtării metalelor din mediu prin biosorpție, sau scăderea cantității de compuși organic din reziduuri activate print-o metodă de electrooxidare.

➤ ***Analiza impactului diverselor funcții nucleu asupra performanței modelului SVM.*** S-a realizat a analiză detaliată asupra impactului funcțiilor nucleu asupra capcătății de modelare SVM pentru două problem de clasificare. În timp ce funcția nucleu lineară este

varianta cea mai simplă, ea oferă rezultate bune într-un timp foarte scurt, fără să necesite optimizarea parametrilor nucleu sau multe date de antrenare. Celelalte două tipuri de nuclee sunt mai complexe, trebuie optimizate, dar posedă capabilități de generalizare avansate. Experimentele au arătat că aceste nuclee pot obține performanță ridicată, fiind capabile să transforme datele nelineare într-un spațiu de dimensiuni de mari dimensiuni unde datele pot fi separate. Aceste tipuri de funcții nucleu necesită un timp mai lung de antrenare și o cantitate mai mare de date pentru antrenare și optimizarea parametrilor.

➤ **Îmbunătățirea performanței modelelor SVM prin optimizarea parametrilor.** Toate modelele SVM implementate și utilizate au fost optimizate utilizând diverse strategii și algoritmi. O comparație detaliată a fost făcută în experimentele de modelare a proceselor chimice între modele SVM cu parametrii implicați și modelele cu parametri optimizați. În toate studiile de caz s-a constatat o îmbunătățire semnificativă a capacității de generalizare a modelelor optimizate, dovedindu-se importanța optimizării în modelarea SVM.

➤ **Agregarea tuturor metodologiilor într-un framework general și flexibil.** Fiecare metodologie prezentată în această teză este bazată pe principiul general de optimizare stability și aceste aspecte comune sunt grupate unitar în framework-ul POM-SVM. Acest cadru comun oferă flexibilitate, avantajul constând în paleta largă de probleme care pot fi rezolvate, rezultate bune, ușurința în utilizare. Pentru testarea performanței framework-ului, dar și a variantelor modelelor SVM optimizate sunt folosite diverse metrici: eroarea pătrată medie, coeficientul de corelare, numărul de iterații până la convergență sau dimensiunea populației

➤ **Aplicarea tuturor metodologiilor de modelare și optimizare pentru diverse procese din ingineria chimică** pentru a identifica proprietățile specifice a produselor chimice obținute. O serie de probleme, cum ar fi modelarea, predicția, clasificarea și optimizarea au fost rezolvate cu ajutorul framework-ului POM-SVM pentru șapte studii de caz din ingineria chimică. Astfel, instrumentele vechi și incomplete pentru rezolvarea problemelor din ingineria chimică (de obicei instrumente fenomenologice) sunt înlocuite cu un singur instrument complex. Studiile de caz descrise în această teză sunt:

✚ Identificarea proprietății de cristal lichid a compușilor organici;

✚ Predicția rezistenței la coroziune a aliajelor de titan;

- ✚ Metoda de electro-oxidare aplicată pentru tratarea rezidurilor active;
- ✚ Îndepărtarea metalelor grele din soluții apoase prin bioremediere
- ✚ Extracția de polifenoli din scoarța de molid
- ✚ Polimerizarea radicalică a stirenului
- ✚ Îndepărtarea metalelor grele prin biosorptie cu alge
- ✚ Modelarea procesului de predicție a coeficientului de partiție în sisteme apoase bifazice

Fiecare proces a fost modelat folosind cel puțin una din metodologiile framework-ului POM-SVM. Inițial se face o scurtă descriere a caracteristicilor fiecărui proces, precum și a modului de creare a datelor pentru simulare. Mai apoi, în funcție de aspectele considerate, rezultatele au fost analizate și discutate. Aspectul original constă în asocierea dintre proces (problemă) – metodologie care este nouă și neîntâlnită în literatură.

✚ Identificarea proprietății de cristal lichid a compușilor organici. Primul studiu este reprezentat de identificarea proprietății de cristal lichid a unor compuși organici prin analiza descriptorilor structurali și geometrici ai moleculelor. Proprietatea de cristal lichid (LC) este una din proprietățile cele mai interesante a compușilor organici, materialele aflate în această stare având două caracteristici legate de ordine și mobilitate. Un aspect foarte important este cuantificarea proprietății de cristal lichid și obținerea celor mai importanți descriptori structurali care influențează în mod semnificativ această proprietate. Au fost selectați descriptori moleculari structurali, calculați prin modelarea moleculară: lungime parte rigidă, lungime parte flexibilă, masa moleculară și factorul de simetrie.

Pentru a identifica proprietatea de cristal lichid au fost propuse mai multe modele SVM de clasificare. O analiză detaliată arată influența funcțiilor kernel și a parametrilor asupra performanței modelelor. S-au comparat rezultatele obținute cu modelul GS-SVM cu alte modele de clasificare, precum DE-NN, algoritmi de învățare de tip lazy sau eager pentru a evalua capacitatea de generalizare a fiecărui model. În faza de testare, rezultatele obținute de GS-SVM sunt dintre cele mai bune, eroare fiind de 7%, în timp ce la celelalte modele eroarea variază într 7-15%. Deși rezultatele generale sunt similare, fiecare metodă are avantaje.

Modelul SVM combinat cu GridSearch poate fi considerat o alternativă importantă de modelare având în vedere acuratețea predicțiilor.

✚ **Predicția rezistenței la coroziune a aliajelor de titan.** Al doilea studiu de caz este reprezentat de discriminarea dintre aliaje de titan rezistente și nerezistente folosite ca și materiale dentare în aplicațiile biomedicale. Rezistența de coroziune a fost evaluată cantitativ de rezistența de polarizare. Datele experimentale pentru cele două aliaje, ZrTi și NiTiNb, au fost obținute în diverse condiții de lucru, cu variația pH-ului, a timpului de contact, a concentrațiilor de albumină și NaF, dar și a compoziției chimice a materialului dentar metalic. Diverse tipuri de modele de clasificare SVM au fost testate pentru diverse funcții nucleu și diverși parametri. S-a observat că nucleul linear este recomandat de folosit datorită simplității și timpului scurt de antrenare, nefind necesară optimizarea parametrilor, dar se observă că folosirea SVM în combinație cu GridSearch pentru optimizarea parametrilor și cu funcții nucleu mai complexe se obține acuratețe maximală.

✚ **Metoda de electro-oxidare aplicată pentru tratarea rezidurilor active.** O metodă de electrooxidare a fost folosită pentru a reduce compușii organici și a înlătura micro-organismele în rezidurile active. Au fost măsurate și folosite ca variabile de intrare concentrațiile inițiale ale parametrilor: coli form (TC) fical coli form (FC), cererea de oxigen biologic (COD), electroconductivitatea (EC) și cantitatea totală solidă dizolvată (TDS), iar ca variabilă de ieșire concentrațiile finale.

Varianta de regresie a mașinilor cu suport vectorial este utilizată pentru modelarea procesului și predicția parametrilor de ieșire. Un dezavantaj al modelului SVM este că acesta poate face predicții pentru o singură variabilă de ieșire, în timp ce algoritmi precum rețelele neuronale pot modela multiple variabile cu un singur model. Astfel s-au construit câte un model pentru fiecare variabilă de ieșire. Aceste modele au fost optimizate cu GridSearch, obținând în final predicții bune, având un coeficient de corelare mare ($R^2 > 0.92$) și o eroare de predicție mică ($MSE < 0.6$). Aceste rezultate sunt comparabile cu cele obținute de modele neuronale optimizate, evaluate pe același set de date. Diferențele mici între valorile experimentale și cele obținute de SVM arată capacitatea de generalizare a acestor modele.

✚ **Îndepărtarea metalelor grele din soluții apoase prin bioremediere.** Influența majoră pe care o are eliminarea metalelor grele din apele uzate asupra mediului și a sănătății

umane reprezintă motivația acestui studiu de caz. O metodologie bazată pe variant hibridă GS-SVM este folosită pentru a obține un model optimal pentru procesul considerat. Următoare etapă constă în optimizarea procesului și determinarea condițiilor optimale de lucru care duc către eficiența maximă prin folosirea algoritmul GA în combinație cu algoritmul optimizat GS-SVM.

✚ **Extracția de polifenoli din scoarța de molid. Utilizând metoda de extracției cu ajutorul microundelor au fost făcute experimente** pentru investigarea impactul unor parametri precum concentrația de etanol, timp și temperatură asupra cantității totale de polifenoli (TPC) extrase din scoarța de molid. Creșterea temperaturii poate îmbunătăți eliberarea de compuși din plante, și implicit a cantității de polifenoli. De asemenea experimentele indică că indicatorul TPC este influențat de timpul de extracție, odată cu creșterea timpului crește și TPC.

Un model SVM pentru regresie a fost folosit pentru a aproxima datele experimentale. A fost folosită funcția nucleu RBF, iar parametrii au fost optimizați cu gridSearch folosind validarea multi-strat. S-a obținut un coeficientul de determinare $R^2 = 0.9427$ care indică o corelare bună între datele de predicție și cele experimentale.

✚ **Polimerizarea radicalică a stirenului.** Următorul studiu de caz este reprezentat de modelarea unui process complex de polimerizare a stirenului folosind un model SVM regresie combinat cu un algoritm DE pentru optimizarea modelului. Rezultatele pun în evidență sensibilitatea performanței modelului SVM la selectarea parametrilor și arată îmbunătățirea rezultatelor în urma optimizării cu DE. De asemenea este folosit în modelare un model neuronal NN de asemenea optimizat parametric și structural cu DE, obținându-se rezultate similare. În concluzie se concluzionează ca hibridizare aduce performanțe semnificativ îmbunătățite pentru ambele tipuri de modele, atât în ce privește acuratețea cât și particularitățile procesului.

✚ **Îndepărtarea metalelor grele prin biosorpție cu alge.** În acest studiu de caz s-a aplicat variant GSO-SVR pentru a modela legătura între eficiența procesului și pH-ul soluției inițiale, concentrație sorbent, concentrație inițială ioni metalici (cadmiu, cobalt, plumb), timp de contact, temperatură de lucru. Capacitatea de bioabsorbție a algelor a fost relaționată cu condițiile de lucru.

Printre motivațiile folosirii metodei GSO în combinație cu SVM este flexibilitatea algoritmului de adaptare la problema optimizării parametrilor. De asemenea comparația rezultatelor cu metoda exhaustivă de căutare GridSearch arată avantajele GSO: oferirea de multiple soluții la problema de optimizare și convergență rapidă, ceea ce înseamnă timp redus de căutare. Rezultatele bune obținute (eroare mai mică de 5%) precum și ușurința în configurare și utilizare, recomandă această tehnică hibridă GSO-SVR pentru modelare și optimizare.

 **Modelarea procesului de predicție a coeficientului de partiție în sisteme apoase bifazice.** Ultimul studiu de caz se referă la predicția coeficientului de partiție al clorhidratului de guanină în sisteme apoaze bifazice. Modelarea procesului s-a făcut cu trei variante: GS-SVM, GSO-SVM și DE-NN. Modele hibride care folosesc optimizatori avansați de inspirație biologică au obținut rezultate asemănătoare și superioare modelului optimizat cu GridSearch (GS). În concluzie, optimizarea cu algoritmi avansați oferă o mai bună acuratețe de modelare în acest studiu de caz.

6.2. Aspecte originale

Aplicarea instrumentelor de inteligență artificială pentru rezolvarea diverselor aspecte a proceselor din ingineria chimică este în stare incipientă, majoritatea cercetărilor bazându-se pe abordările clasice. Teza curentă propune un framework general conținând diferite metodologii pentru modelare, predicție, clasificare și optimizare. Aceste metodologii sunt abordări noi bazate pe diverse combinații hibride optimizator-mașini cu suport vectorial. Contribuțiile originale ale tezei sunt următoarele:

➤ Dezvoltarea unei metodologii generale pentru modelarea mașinilor cu support vectorial cu parametri optimizați bazați pe optimizatori globali cum ar fi Grid Search, Differential Evolution and Glowworm Swarm Optimization. Caracteristicile specifice ale metodologiei constau în: i) stabilirea spațiului de căutare prin detectarea marginilor pentru toți meta-parametrii SVM: parametrul de cost și parametrii funcției nucleu; ii) abordarea atât a problemelor de regresie cât și a celor de clasificare, în funcție de natura studiile experimentale; iii) folosirea unei funcții fitness simple bazate pe acuratețe sau MSE în faza de antrenare.

➤ Dezvoltarea unei implementări noi a algoritmului GSO în limbajul JAVA, adaptat la optimizarea parametrilor SVM și prin modificarea criteriilor de convergență și de identificare a parametrilor optimali.

➤ Combinarea tuturor metodologiilor într-un framework nou pentru modelare cu SVM a diferitelor tipuri de problem în ingineria chimică. Acesta este un avantaj, în special pentru inginerul chimist care se poate concentra pe problemele specific procesului și nu pe mecanismele interioare ale algoritmului. În același timp framework-ul prezintă și implementări software originale.

6.3. Direcții de continuare a cercetărilor

Cercetările efectuate în această teză se pot continua pe trei direcții mari: i) dezvoltarea de modele de regresie SVM cu ieșiri multiple; ii) noi tipuri de nuclee; iii) optimizatori noi și îmbunătățirea celor existenți.

➤ În legătură cu modele de regresie SVM, acestea sunt folosite în teză în variant de implementare LibSVM, cu o singură ieșire. Cazurile în care ieșirile proceselor erau multiple, s-a construit un model pentru fiecare variabilă de ieșire, independente unele de altele. Extensia variantei SVR cu o ieșire poate fi făcută considerând relația non-lineară între trăsături dar și între variabilele de ieșire. Studiile existente arată că această abordare este mai eficientă deoarece consider și corelarea dintre variabilele de ieșire, care de obicei sunt inter-dependente.

➤ O trăsătură de bază în modelarea cu mașini cu support vectorial este selecția unei funcții nucleu potrivită. Framework-ul general a folosit nucleele cel mai frecvent utilizate în literatură, găsite în pachetul LibSVM. Există și funcții nucleu avansate alternative, cum ar fi nuclee de tip arbore, nuclee neuronale, nucleu anova, nucleu spline care pot fi mai adecvate pentru anumite probleme. De exemplu, la identificarea proprietății de cristal lichid a compușilor s-a folosit descriptori structurali ai molecule. Acești descriptori sunt modelați cu funcții nucleu standard pentru a măsura similiaritatea dintre doi compuși organici. Pentru

acest tip de modelare, unii cercetători recomandă folosirea nucleelor de tip arbore, deoarece aceste funcții pot reprezenta toate detaliile structurale ale moleculelor folosind o reprezentare complexă de tip graf.

➤ O altă îmbunătățire a framework-ului poate fi adăugarea unui nou nivel de optimizare peste cel de optimizare a modelului, în scopul găsirii condițiilor optime de lucru a proceselor. Acest pas are ca intrare un model SVM optimizat care este returnat de framework, și determină valorile optime ale datelor de intrare care conduc la eficiența maximă. Există deja două exemple de optimizare de proces în teză prin folosirea algoritmului GA combinat cu GS-SVR, dar această etapă nu este inclusă în framework-ul general.

Sunt necesare efectuarea de experimente în vederea stabilirii modului de îmbunătățire a optimizatorului GSO în cadrul framework-ului prin automatizarea selecției parametrilor, prin testare unor funcții de fitness alternative, prin folosirea paralelismului sau căutarea graduală pentru reducerea timpului de procesare. De asemenea se pot efectua teste cu GSO cu diverse configurații ale modelului SVM, cu diverse nuclee (s-a folosit doar cel Gaussian) și diverși parametri. Algoritmul GSO s-a dovedit eficient în identificarea punctelor de optim global, după cum arată rezultatele experimentale efectuate în această teză. Totuși pentru anumite date algoritmul prezintă problem de convergență, ca în cazul procesului de biosorpție a metalelor grele, datorită complexității interacțiunilor datelor și din acest motiv s-au efectuat un număr mare de teste pentru a obține rezultate bune. În consecință este nevoie de o procedură automată de găsire a setărilor GSO optime pentru fiecare problem în parte.

Publicații ce vizează obiectivele rezolvate în teză

Butnariu C., Lisa C, Leon F, Curteanu S. Prediction of liquid-crystalline property using support vector machine classification, *Journal of Chemometrics* 2013, 27:179–188. *ISI factor 1.80.*

Curteanu S, Godini K, Piuleac C, Azarian G, Rahmani A, **Butnariu C**. Electro-Oxidation Method Applied for Activated Sludge Treatment: Experiment and Simulation Based on Supervised Machine Learning Methods, *Industrial & Engineering Chemistry Research* 2014, 53(12):4902-4912. *ISI factor 2.235*.

Curteanu S, Dragoi E-N, Leon F, **Butnariu C**. Artificial intelligence modelling methodologies applied to a polymerization process, *Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH)* 2014, 43-49.

Curteanu S, **Butnariu C**, Bulgariu L. Modelling of a heavy metals biosorption using Glowworm Swarm Optimisation – Support Vector Machines procedure; *to be submitted to Chemical Engineering Communication*.

Total ISI factor: **4.035**

Bibliografie (selectivă)

Burbidge R, Trotter M, Holden S, Buxton B. Drug Design by Machine Learning: Support Vector Machines for Pharmaceutical Data Analysis. *Computers & Chemistry* 2001; 26:5-14.

Butnariu C, Lisa C, Leon F, Curteanu S. Prediction of liquid-crystalline property using support vector machine classification, *Journal of Chemometrics* 2013, 27: 179–188.

Cartwright H, Curteanu S. Neural networks applied in chemistry. II. Neuro-evolutionary techniques in process modeling and optimization. *Industrial & Engineering Chemistry Research* 2013.

Chan CH, Yusoff R, Ngoh GC, Kung F. Microwave-assisted extractions of active ingredients from plants, *Journal of Chromatography* 2011, A 1218(37): 6213-6225.

Chang C-C, Lin C-J. LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2011, 2(3):27.

Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. *Machine learning* 2002, 46(1-3):131-159.

Chelaru R, Suditu D, Mareci D, Bolat G, Cimpoiu N, Leon F, Curteanu S. Prediction of corrosion resistance of some dental metallic materials applying artificial neural networks and regression based models. *JOM* 2015; 67(4): 767-774.

Cherkassky V, Mulier F. *Learning from Data: Concepts Theory and Methods*, Wiley, New York, 1998.

Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression, *Neural networks* 2004, 17(1): 113-126.

Curteanu S, Dragoi E-N, Leon F, Butnariu C. Artificial intelligence modelling methodologies applied to a polymerization process, *Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH)* 2014a, p 43-49.

Curteanu S, Godini K, Piuleac C, Azarian G, Rahmani A, Butnariu C. Electro-Oxidation Method Applied for Activated Sludge Treatment: Experiment and Simulation Based on Supervised Machine Learning Methods, *Industrial & Engineering Chemistry Research* 2014b, 53 (12): 4902-4912.

Drăgoi EN, Curteanu S, Galaction AI, Cascaval D. Optimization methodology based on neural networks and self-adaptive differential evolution algorithm applied to an aerobic fermentation process, *Applied Soft Computing* 2013, 13(1): 222-238.

Hlihor R M, Diaconu M, Leon F, Curteanu S, Gavrilescu M. Experimental analysis and mathematical prediction of Cd(II) removal by biosorption using support vector machines and genetic algorithms, *New biotechnology* 2014.

Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. Technical report 2010, *Department of Computer Science, National Taiwan University*.

Ivanciuc O. Applications of Support Vector Machines in Chemistry. *Reviews in Computational Chemistry*, 2007. 23: 291-400.

Krishnanand KN, Ghose D. Glowworm swarm based optimization algorithm form multimodal functions with collective robotic applications. *Multiagent and Grid Systems* 2006a, 2(3): 209–222.

Krishnanand KN, Amruth P, Guruprasad MH, Bidargaddi SV, Ghose D. Glowworm-inspired robot swarm for simultaneous taxis toward multiple radiation sources. In *Proceedings of IEEE international conference on robotics and automation 2006b*, p. 958–963.

Leon F, Lisa C, Curteanu S. Prediction of the liquid-crystalline property using different classification methods, *Molecular Crystals and Liquid Crystals* 2010, 518:129-148.

Li H, Liang Y, Xu Q. Support vector machines and its applications in chemistry. *Chemometrics and Intelligent Laboratory Systems* 2009, 95: 188-198.

Vapnik V. *The Nature of Statistical Learning Theory*, Springer: New York, 1995.