

**UNIVERSITATEA DE MEDICINĂ ȘI FARMACIE  
„CAROL DAVILA”, BUCUREȘTI**

**ȘCOALA DOCTORALĂ  
DOMENIUL FARMACIE**



***APLICAȚIILE RECUNOAȘTERII FORMELOR LA  
INTERPRETAREA MULTIVARIATĂ A DATELOR  
REZULTATE DIN CERCETAREA FARMACEUTICĂ***

**REZUMATUL TEZEI DE DOCTORAT**

**Conducător de doctorat:**

**PROF. UNIV. DR. NEGREȘ SIMONA**

**Student-doctorand:**

**VĂLEANU ANDREI**

**București**

**2018**

## Cuprins

Lista cu lucrările științifice publicate . . . . .	pagina 5
Lista cu abrevieri și simboluri . . . . .	pagina 7
Introducere și obiective . . . . .	pagina 9
<b>I. Partea generală . . . . .</b>	<b>pagina 12</b>
1. Recunoașterea de forme-generalități . . . . .	pagina 12
2. Recunoașterea nesupervizată de forme. Aplicații. . . . .	pagina 18
2.1. Analiza exploratorie a datelor (Exploratory Data Analysis-EDA) . . . . .	pagina 18
2.2. Analiza în componente principale (Principal Component Analysis-ACP) . . . . .	pagina 18
2.3. Analiza factorială (Factor Analysis-FA) . . . . .	pagina 21
2.4. Analiza de clusteri (Cluster Analysis-CA) . . . . .	pagina 21
2.4.1. Clusterizarea aglomerativă/ierarhică . . . . .	pagina 26
2.4.2. Clusterizările de tip K-means, K-medoids, FCM . . . . .	pagina 31
3. Recunoașterea supervizată de forme. Aplicații. . . . .	pagina 40
3.1. Regresia . . . . .	pagina 41
3.2. Clasificarea . . . . .	pagina 43
3.2.1. Tipuri de procese de clasificare și date de ieșire . . . . .	pagina 46
3.2.2. Exemple de clasificatori. Aplicații . . . . .	pagina 47
<b>II. Contribuții personale. . . . .</b>	<b>pagina 64</b>
4. Cercetări preliminare pentru utilizarea analizei de clusteri și a învățării supervizate în caracterizarea proteinelor serice umane oxidate cuantificate prin electroforeză capilară . . . . .	pagina 64
4.1. Introducere . . . . .	pagina 64
4.2. Material și metodă . . . . .	pagina 65
4.2.1. Probe . . . . .	pagina 65
4.2.2. Selectarea picurilor și clasificatorul kNN . . . . .	pagina 66
4.2.3. Implementarea metodei IC-kNNI pentru estimarea valorilor lipsă . . . . .	pagina 70
4.2.4. Evaluarea modelului de albumină umană carbonilată pe pacienți diabetici prin intermediul tehnicii K-means . . . . .	pagina 71
4.3. Rezultate . . . . .	pagina 73
4.3.1. Selectarea picurilor și clasificatorul kNN . . . . .	pagina 73
4.3.2. Rezultatele implementării IC-kNNI . . . . .	pagina 75
4.3.3. Evaluarea modelului de carbonilare proteică pe pacienți diabetici prin intermediul tehnicii K-means . . . . .	pagina 77
4.4. Discuții . . . . .	pagina 78
4.4.1. Modelul de albumină serică umană carbonilată . . . . .	pagina 78
4.4.2. Evaluarea modelului de carbonilare proteică pe pacienți diabetici . . . . .	pagina 80
4.5. Concluzii . . . . .	pagina 82

5. Utilizarea recunoașterii de forme în vederea analizării unor biomarkeri cu posibilă relevanță pentru evaluarea stadiilor retinopatiei diabetice . . . . .	pagina 84
5.1. Introducere . . . . .	pagina 84
5.2. Material și metodă . . . . .	pagina 85
5.2.1. Designul studiului . . . . .	pagina 85
5.2.2. Analiza statistică preliminară . . . . .	pagina 85
5.2.3. Analiza în componente principale (ACP) . . . . .	pagina 86
5.2.4. Analiza de clusteri-Învățare nesupervizată . . . . .	pagina 87
5.2.5. Evaluarea procesului de clusterizare . . . . .	pagina 95
5.3. Rezultate . . . . .	pagina 97
5.3.1. Analiza statistică preliminară . . . . .	pagina 97
5.3.2. Analiza în componente principale . . . . .	pagina 101
5.3.3. Evaluarea procesului de clusterizare . . . . .	pagina 103
5.3.4. Evaluarea cantitativă a acurateții procesului de clusterizare . . . . .	pagina 106
5.4. Discuții . . . . .	pagina 113
5.5. Concluzii . . . . .	pagina 118
6. Dezvoltarea unei aplicații de predicție personalizată a reacțiilor adverse în cazul pacienților cu polipatologie și polimedicație . . . . .	pagina 119
6.1. Introducere . . . . .	pagina 119
6.2. Material și metodă . . . . .	pagina 121
6.2.1. Selectarea surselor de informație . . . . .	pagina 121
6.2.2. Preprocesarea datelor . . . . .	pagina 126
6.2.3. Dezvoltarea sistemului de scoruri . . . . .	pagina 130
6.2.4. Validarea scorului de severitate . . . . .	pagina 134
6.2.5. Sistemul de interfață grafică și dezvoltarea aplicației . . . . .	pagina 136
6.3. Rezultate . . . . .	pagina 140
6.3.1. Exemplu de listă de RA armonizate conform MedDRA . . . . .	pagina 140
6.3.2. Exemple de patologii SMQ cu termeni LLT prezenți în grupare . . . . .	pagina 146
6.3.3. Exemplu de executare a aplicației de RA (pentru opțiunea generală) . . . . .	pagina 150
6.3.4. Exemplu de executare a aplicației de RA (pentru opțiunea multiplă) . . . . .	pagina 154
6.3.5. Exemplu de executare a aplicației de RA (pentru opțiunea specifică) . . . . .	pagina 155
6.3.6. Exemplu de executare a aplicației de RA (opțiunea generală pentru o prescripție electronică dintr-o farmacie de circuit deschis) . . . . .	pagina 156
6.3.7. Validarea scorului de severitate . . . . .	pagina 162
6.4. Discuții . . . . .	pagina 163
6.5. Concluzii . . . . .	pagina 179
7. Concluzii generale . . . . .	pagina 181
Bibliografie . . . . .	pagina 188
Anexe . . . . .	pagina 203

## **Mulțumiri**

Aș dori să adresez mulțumiri pentru realizarea acestei teze de doctorat doamnei CS II Ilie Mihaela care m-a ajutat să pun într-un mod relevant bazele cercetării desfășurate în cadrul studiilor doctorale. De asemenea, îi mulțumesc doamnei Prof. Dr. Negreș Simona care m-a preluat la începutul anului 2018, a contribuit la aducerea ideilor într-o formă matură și concretă și m-a ajutat să finalizez cu bine teza de doctorat.

## **Mențiune:**

Teza de doctorat a fost cofinanțată parțial din Fondul Social European prin Programul Operațional Sectorial pentru Dezvoltarea Resurselor Umane 2007-2013, proiect POSDRU „Susținerea excelenței în cercetarea științifică doctorală interdisciplinară în domeniile economic, medical și al științelor sociale”, **POSDRU/187/1.5/S/155463**, coordonator Academia de Studii Economice din București (<http://interdisciplinar.ase.ro/>).

## Lista cu lucrările științifice publicate

### A. Articole publicate în reviste de specialitate

#### Prim autor:

1. **Văleanu, A.**, Ilie, M., Dima, I. și Purdel, C. (2016). K Nearest Neighbours analysis of human serum carbonyl proteins using capillary electrophoregrams. *Romanian Journal of Biophysics*, 26(1), pp. 001-010 (disponibil online la: <https://www.rjb.ro/k-nearest-neighbours-analysis-of-human-serum-carbonyl-proteins-using-capillary-electrophoregrams/> ), indexare **B+**.

2. **Văleanu, A.**, Margină, D.M., Grădinaru, D., Ilie, M., Dima, I.E., Purdel, C.N. și Dănciulescu-Miulescu, R. (2016). Diabetic retinopathy and inflammation: a comparative statistical study on relevant blood serum parameters. În: C. Serafinceanu, O. Negoită, V. Elian (Eds.), *InterDIAB 2016 Book Series, International Conference on Interdisciplinary Management of Diabetes Mellitus and its Complications: „Diabetes Mellitus as Cardiovascular Disease”*, Editura Niculescu. București, România, pp. 490-499, ISSN: 2393-3488, indexare **ISI**.

#### Al 3-lea autor:

Dima, I., Purdel, C., **Văleanu, A.**, Moldoveanu, G., Ilie, M. și Margină, D. (2018). Capillary Electrophoresis For The Evaluation Of The Carbonylation Pattern In Type 2 Diabetes Mellitus. Which Are The Premises? În: C. Serafinceanu, O. Negoită, V. Elian (Eds.), *InterDIAB 2018 Book Series, 4<sup>th</sup> International Conference on Interdisciplinary Management of Diabetes Mellitus and its Complications: „Surgical Crossroads with Diabetes Mellitus”*, Editura Niculescu. București, România, pp. 268-276, ISSN: 2393-3488.

### B. Lucrări prezentate la manifestări științifice naționale

**Văleanu A.**, Margină D., Grădinaru D., Ilie M., Evaluarea comparativă a unor pacienți diabetici prin intermediul analizei în componente principale a unor biomarkeri relevanți pentru retinopatia diabetică, poster prezentat la Congresul Național de Farmacie din România, ediția a XVI-a, 28 septembrie-1 octombrie 2016, București, România

### C. Lucrări prezentate la manifestări științifice internaționale

1. **Văleanu A.**, Dima I., Purdel C., Ilie M., Capillary electrophoresis pattern of carbonylated human serum albumin using cluster analysis, prezentare orală la “First

Euroregional Conference for PhD students and young researchers in biomedicine”, 27-28 martie 2015, Timișoara, România

2. **Văleanu A.**, Purdel C., Dima I., Margina D., Ilie M., The pattern of carbonylated human serum albumin using cluster analysis, poster prezentat la Congresul “51-st Congress of the European Societies of Toxicology”, Eurotox 2015, 13-16 septembrie, Porto, Portugalia, rezumat publicat în Toxicology Letters, 238S, 2015, S169, DOI: 10.1016/j.toxlet.2015.08.575

3. **Văleanu A.**, Dima I., Purdel C., Ilie M., Margină D., Dănciulescu-Miulescu R., Ionescu-Tîrgoviște C., The electrophoretic pattern of carbonyl proteins in human serum albumin using k nearest neighbours method, poster prezentat la 4-th International Symposium on Adipobiology and Adipopharmacology, 28-31 octombrie 2015, București, România

4. **Văleanu A.**, Margină D.M., Grădinaru D., Ilie M., Dima I.E., Purdel C.N., Dănciulescu-Miulescu R., Diabetic retinopathy and inflammation: a comparative statistical study on relevant blood serum parameters, prezentare orală la Interdiab 2016, 3-5 martie, București, România.

5. **Văleanu A.**, Margină D., Grădinaru D., Ilie M., A fuzzy c-means and k-means clustering analysis on relevant diabetic retinopathy biomarkers, poster prezentat la Congresul “52-nd Congress of the European Societies of Toxicology”, Eurotox 2016, 4-7 septembrie, Sevilha, Spania, rezumat publicat în Toxicology Letters, 258, 2016, S117, DOI: 10.1016/j.toxlet.2016.06.1476

6. **Văleanu A.**, Damian C., Ilie M., Computer application for a patient-tailored adverse drug reaction prediction in poly medication, poster prezentat la “6th FIP Pharmaceutical Sciences World Congress” (PSWC 2017), 21-24 mai 2017, Stockholm, Suedia, rezumat disponibil online la:  
[http://www.fip.org/menu\\_sitemap?page=abstracts&action=generatePdf&item=18151](http://www.fip.org/menu_sitemap?page=abstracts&action=generatePdf&item=18151)

## Introducere și obiective

Recunoașterea de forme reprezintă un tip de analiză de date care implică sortarea, gruparea, clasificarea și/sau reorganizarea datelor disponibile cu aducerea acestora într-o formă ușor de prelucrat și/sau analiza efectivă a acestora (James, 2013). Recunoașterea de forme câștigă tot mai mult teren în ultima perioadă în științele farmaceutice și medicale, datorită dimensionalității și complexității din ce în ce mai ridicate ale bazelor de date rezultate din cercetarea efectuată în aceste domenii.

Astfel, este necesară dezvoltarea și aplicarea unor tehnici noi de analiză, dar și a unor combinații ale unor metode deja existente, în scopul exploatării acestor seturi de date prezente în număr extrem de mare, cu evidențierea relevanței acestora și obținerea de noi ipoteze și rezultate cu potențial clinic superior (James, 2013).

Ținând cont de aspectele menționate, principalul obiectiv al tezei de doctorat a fost reprezentat de implementarea și evaluarea unor astfel de metode de recunoaștere de forme în vederea analizei mai multor seturi de date multivariate rezultate din diverse ramuri ale cercetării farmaceutice. Scopurile principale ale aplicării tehnicilor menționate au fost reprezentate de:

- i) preprocesarea informațiilor disponibile în seturile de date neprelucrate
- ii) prelucrarea propriu-zisă a seturilor de date.
- iii) extragerea informațiilor cu relevanță clinică prin prisma obiectivelor generale urmărite în cadrul cercetării farmaceutice.
- iv) evaluarea propriu-zisă a algoritmilor testați, cu evidențierea capacității acestora de a analiza cu o precizie ridicată seturile de date obținute.

Capitolele 1, 2 și 3 au constituit partea generală a tezei elaborate, în schimb ce capitolele 4, 5 și 6 au constituit partea de contribuții personale ale studentului-doctorand. Capitolul 4 a utilizat date referitoare la proteinele carbonilate cuantificate prin electroforeză capilară, capitolul 5 a urmărit stabilirea unei legături între diverși biomarkeri serici și stadiile retinopatiei diabetice iar obiectivul capitolului 6 a fost reprezentat de dezvoltarea unei aplicații de predicție personalizată a reacțiilor adverse în cazul pacienților cu polimedicație.

## I. Partea generală

### 1. Recunoașterea de forme – generalități

În prezent, recunoașterea de forme (recunoaștere de model sau Pattern recognition, așa cum este aceasta denumită în mod oficial în limba engleză) cuprinde un set de metode de analiză de date al căror scop primar este de a reorganiza, grupa, sorta și/sau clasifica informația prelucrată, în vederea evidențierii anumitor caracteristici și/sau regularități în setul de date. Nu în ultimul rând, se recurge la identificarea datelor relevante pentru scopul analizei respective. Nevoia de a dezvolta astfel de tehnici cu aplicabilitate extinsă a pornit de la dezvoltarea continuă a informației prezente în diferitele domenii științifice (Brereton, 2003).

### 2. Recunoașterea nesupervizată de forme. Aplicații.

Recunoașterea nesupervizată de forme presupune construirea unui model din date necunoscute, a căror regularitate nu a fost probată. Recunoașterea nesupervizată poate fi împărțită în mai multe ramuri, în funcție de structura setului de date, scopul analizei și algoritmiile utilizate. Cele mai utilizate tehnici sunt analiza în componente principale și analiza de clusteri (Brereton, 2003). Capitolul detaliază aceste tehnici și prezintă pe larg aplicațiile lor în cercetarea din domeniile medical și farmaceutic.

**Analiza în componente principale** este utilizată în special în domenii în care se obțin cantități extrem de mari de date, care sunt imposibil de analizat ca atare (exemplu: cromatografie, spectrometrie, cristalografie, etc.). În acest sens, tehnica are ca primă etapă **reducerea dimensionalității setului de date**, cu eliminarea acelor care nu prezintă, din punct de vedere statistic, semnificație pentru modelul investigat (James, 2013).

**Analiza de clusteri** este cel mai des citată ca ramură principală a recunoașterii nesupervizate de forme. CA presupune gruparea (sortarea, ierarhizarea) datelor analizate, ținând cont de diferențele dintre ele în privința parametrilor cercetați, având ca rezultat crearea uneia sau mai multor clase cu aceleași proprietăți (Brereton, 2003; James, 2013). Există mai multe tipuri principale de CA, în speță clusterizarea aglomerativă/ierarhică, clusterizarea bimodală și clusterizările K-means, K-medoids și Fuzzy c-means (FCM). Dintre aplicațiile cu impact în

cercetarea farmaceutică este de menționat realizarea unui sistem de predicție sistematică a reacțiilor adverse la medicamentele de uz uman (Ngufor, 2015).

### **3. Recunoașterea supervizată de forme. Aplicații.**

Recunoașterea supervizată de forme, spre deosebire de cea nesupervizată, presupune existența unui set de date deja cunoscut, pentru care sunt cunoscute valorile variabilelor de interes. Acesta va constitui setul de antrenament și va fi utilizat pentru estimarea valorilor necunoscute din cadrul setului de testare. Din punct de vedere al scopului pe care o astfel de analiză îl urmărește, sunt de menționat două tipuri principale, în speță *regresia*, respectiv *clasificarea datelor necunoscute* (James, 2013).

Tehnicile de clasificare prezintă o importanță mai mare decât cele de regresie, în special datorită implementării pe scară largă a acestora în cadrul algoritmilor de recunoaștere și învățare automată (machine learning) (James, 2013).

În funcție de tipul de algoritm utilizat, există mai multe tipuri de metode de clasificare: regresia logistică, tehnica celor mai apropiați k vecini (k Nearest Neighbors, kNN), metoda arborilor aleatori (Random Forests), rețelele neurale (Neural networks), clasificatorii bayesieni (Naive Bayes), clasificatorii de tip vectori suport (Support Vector Classifier), clasificatorii de tip ansamblu (Ensemble Classifiers) (Sharma, 2013). Capitolul detaliază metodele respective și descrie aplicațiile importante ale acestora în cercetarea din cadrul științelor farmaceutice și medicale.

Tehnicile de clasificare prezintă numeroase aplicații în domeniul farmaceutic, dintre care se pot aminti analiza datelor de farmacovigilență (Sarker, 2015), modelarea farmacocinetică și farmacodinamică (Sutariya, 2013) sau metode performante de diagnostic medical, cum ar fi cele axate pe înregistrări medicale electronice (Electronic Health Records) (DuBrava, 2017).

## II. Contribuții personale

### 4. Cercetări preliminare pentru utilizarea analizei de clusteri și a învățării supervizate în caracterizarea proteinelor serice umane oxidate cuantificate prin electroforeză capilară

#### 4.1. Introducere

Ținând cont de importanța majoră pe care o prezintă proteinele carbonilate în stresul oxidativ (Uttara, 2009), obiectivul prezentului studiu a fost reprezentat de utilizarea unor tehnici specifice de recunoaștere de forme (analiza de clusteri, clasificatorul kNN) în vederea stabilirii unui model (pattern) în electroforegramele proteinelor carbonilate obținute din abumina serică umană și din probele prelucrate din serul unor pacienți diabetici (Văleanu, 2016a).

#### 4.2. Material și metodă

În vederea construirii setului de date de analizat, au fost utilizate rezultatele experimentale obținute în urma aplicării tehnicii de electroforeză capilară pe două tipuri de probe: albumină umană standardizată (5 probe) și probe de ser obținute de la 6 pacienți diabetici.

În vederea implementării algoritmilor de recunoaștere de forme, timpii de retenție ai picurilor selectate din cadrul probelor de albumină umană au fost utilizați pentru a construi mai multe clase de picuri (clusteri), astfel încât fiecare clasă să conțină picuri cât mai asemănătoare. A fost utilizată ca măsură de disimilaritate radicalul sumei normalizate a diferențelor distanțelor euclidiene pătratice (dintre fiecare pereche de picuri dintr-o clasă) (DDEP). Clasele obținute la 214 nm, respectiv 365 nm, au constituit setul de antrenament. Acesta a fost utilizat pentru clasificarea prin intermediul kNN a picurilor electroforetice ale celor 6 pacienți diabetici (Văleanu, 2016a). După clasificare, a fost propus un algoritm care să stabilească dacă electroforegramele pacienților corespund sau nu modelului (pattern-ului) de albumină umană carbonilată.

Pentru a evalua modul în care timpii de retenție ai picurilor electroforetice ar putea urma un anumit model care să caracterizeze procesul de carbonilare proteică pentru pacienții diabetici, au fost utilizate probe recoltate de la 12 pacienți diabetici diferiți (diabet zaharat tip II). Valorile timpilor de retenție obținuți au fost grupate în 10 clase prin intermediul unui algoritm de analiză

de clusteri de tip K-means, cu distanța euclidiană pătratică ca măsură de disimilaritate. După efectuarea grupării, fiecare clasă de picuri formată a fost caracterizată prin intermediul mai multor parametri (Dima, 2018).

### **4.3. Rezultate**

#### **Clasificatorul kNN**

Rezultatele procesului de grupare a picurilor electroforetice selectate au indicat o bună separare a grupurilor, cu un minim al DDEP de 0,0474 și un maxim de 0,7248. Din punct de vedere al clasificării picurilor electroforetice ale pacienților diabetici, un număr total de 115 picuri au fost clasificate la 214 nm și 365 nm prin intermediul kNN.

#### **Evaluarea modelului de carbonilare proteică pe pacienți diabetici prin intermediul tehnicii K-means**

Un număr total de 100 de timpi de retenție au fost grupați în cei 10 clusteri. Valorile coeficienților de variație în clasele obținute au fost cuprinse între 1,06 și 21,56%, cu o singură valoare de peste 10%.

### **4.4. Discuții**

#### **Modelul de albumină serică umană carbonilată**

Numai câteva picuri de pacienți au corespuns în întregime modelului: 5 la lungimea de undă de 214 nm și 4 la lungimea de undă de 365 nm. Alte picuri au corespuns modelului de albumină serică umană carbonilată numai din punct de vedere al timpului de retenție: 4 la 214 nm, respectiv un singur pic la 365 nm. Timpul de retenție a fost parametrul care a condus la cel mai ridicat grad de similaritate între setul de antrenament și setul de testare, dovedind astfel o mai bună asemănare calitativă și o mai slabă asemănare cantitativă.

#### **Evaluarea modelului de carbonilare proteică pe pacienți diabetici**

Rezultatele prezentate evidențiază gradul redus de suprapunere în privința timpilor de retenție pentru probele obținute din serul pacienților diabetici. Cu toate acestea, este de remarcat faptul că algoritmul de tip K-means care a fost implementat a condus la obținerea unor clase de

picuri compacte, care nu se interesează, fiind astfel îndeplinite condițiile unui proces relevant de clusterizare (Brereton, 2003).

#### **4.5. Concluzii**

Picurile electroforegramelor capilare obținute pot fi utilizate în vederea construirii unui model al albuminei serice umane oxidate. Modelul a fost creat utilizând analiza de clusteri și tehnica kNN ca algoritmi, timpul de retenție ca variabilă și distanța euclidiană ca măsură de disimilaritate. Din punct de vedere al construirii unui model de carbonilare proteică pentru pacienții suferind de diabet zaharat, s-a remarcat un grad redus de suprapunere a electroforegramelor acestora. Cu toate acestea, grupurile obținute în urma implementării tehnicii K-means au prezentat un caracter extrem de compact, care dovedește existența unui model de carbonilare proteică, care ar putea fi optimizat pe viitor prin includerea informațiilor referitoare la medicația și comorbiditățile pacienților diabetici.

### **5. Utilizarea recunoașterii de forme în vederea analizării unor biomarkeri cu posibilă relevanță pentru evaluarea stadiilor retinopatiei diabetice**

#### **5.1. Introducere**

Analiza literaturii de specialitate nu a condus la evidențierea vreunui studiu care să evalueze utilizarea recunoașterii de forme pentru corelarea parametrilor serici cu stadiile retinopatiei diabetice (ex: retinopatia neproliferativă și retinopatia proliferativă). Ținând cont de acest aspect și de prevalența extrem de mare a retinopatiei diabetice, obiectivul studiului a fost reprezentat de găsirea unei legături între astfel de parametri serici și diversele faze ale patologiei, prin intermediul mai multor tehnici de analiză de clusteri (K-means, K-medoids, Fuzzy c-means), precum și prin analiza în componente principale.

#### **5.2. Material și metodă**

Studiul a presupus analiza unei baze de date conținând valorile a 8 parametri serici determinați pentru 72 de pacienți diabetici, incluși pe criteriu clinic în 3 clase: fără retinopatie (FR) (20 de pacienți), cu retinopatie neproliferativă (RNP) (28 de pacienți) și cu retinopatie proliferativă (RP) (24 de pacienți). Biomarkerii analizați au fost glicemia (Gli), colesterolul total (CT), trigliceridele (TG), molecula de adeziune intercelulară 1 (ICAM-1), factorul vascular de

creștere endotelială (VEGF), factorul de necroză tumorală alfa (TNF- $\alpha$ ), malonildialdehida (MDA) și LDL-ul oxidat (LDL-ox).

Într-o primă etapă, pentru a analiza structura intrinsecă a fiecărei clase clinice (FR, RNP, RP), a fost necesară realizarea unei analize statistice preliminare (James, 2013). Apoi, analiza în componente principale (ACP) a fost implementată pentru a evalua posibilitatea de reducere a dimensionalității obținută prin intermediul acestei tehnici de recunoaștere de forme de la 8 variabile la numai 2 componente principale (CP) relevante (James, 2013).

### **Analiza de clusteri – Învățare nesupervizată**

În vederea implementării analizei de clusteri, au fost create 3 cazuri, în funcție de parametrii aleși: *Cazul 1* (Gli și VEGF), *Cazul 2* (TNF- $\alpha$  și VEGF) și *Cazul 3* (toți cei 8 parametri). Cei trei algoritmi de analiză de clusteri (*K-means*, *K-medoids*, *Fuzzy c-means - FCM*) au fost implementați pentru fiecare dintre cele trei cazuri (Yin, 2014; Arora, 2016). Pacienții au fost grupați în scopul obținerii unor clusteri cu pacienți asemănători din punct de vedere al valorilor biomarkerilor. Au fost obținuți 3 până la 15 clusteri.

### **Evaluarea procesului de clusterizare**

O modalitate extrem de importantă de a stabili legătura dintre clasificarea clinică și grupurile matematice a avut la bază atribuirea clasă clinică – cluster. Astfel, pentru fiecare implementare a analizei de clusteri, au fost extrase două măsuri de evaluare, puritatea (P) și indicele Rand (Rand index - RI).

Puritatea a funcționat ca un vot majoritar și poate fi considerată ca fiind o formă de clasificare a clusterilor în interiorul claselor clinice. Pe de altă parte, indicele Rand (RI) s-a bazat pe legătura dintre perechile de pacienți aparținând aceleiași clasificări sau a două clase clinice diferite. În acest sens, două tipuri de măsuri au fost calculate, numărul de perechi adevărat pozitive (AP) și numărul de perechi adevărat negative (AN) iar suma acestora a fost împărțită la numărul total de perechi de pacienți (Manning, 2008).

### **5.3. Rezultate**

#### **Analiza statistică preliminară și ACP**

Cele mai importante rezultate ale analizei statistice preliminare s-au referit la coeficientul de variație (RSD%), care a evidențiat o variabilitate minimă în interiorul fiecărei clase clinice pentru TNF- $\alpha$  (15,17%-18,37%) și VEGF (9,83%-13,96%) (Văleanu, 2016b). De asemenea, analiza varianței CP a arătat faptul că primele 2 CP au explicat 76,4% din cantitatea totală de varianță, în timp ce primele 3 CP au contribuit cumulativ la 88,64% din varianța totală.

#### **Evaluarea cantitativă a acurateții procesului de clusterizare**

Evaluarea cantitativă a acurateții procesului de clusterizare a reprezentat cea mai importantă parte a evaluării implementării algoritmilor de recunoaștere de forme, aceasta evidențiind legătura dintre grupurile matematice și clasele clinice în termeni de eficiență. Tabelul 5.1 prezintă intervalele de valori obținute pentru P și RI. Sunt prezentate atât intervalele de valori obținute pentru  $K = 3$  ( $P_3$ ,  $RI_3$ ), cât și intervalele de valori maxime ale P și RI ( $P_{max}$ ,  $RI_{max}$ ) (valoarea maximă obținută pentru fiecare implementare a analizei de clusteri). În plus, sunt redate cazurile pentru care valorile optime ale  $P_3$ ,  $RI_3$ ,  $P_{max}$  și  $RI_{max}$  au fost atinse, precum și numărul de clusteri pentru care rezultatul optim a fost atins ( $K_{max}$  – valabil pentru  $P_{max}$  și  $RI_{max}$ ).

Tabel 5.1. Sumarizarea rezultatelor obținute pentru P și RI

<b>Măsura de evaluare</b>	<b>Intervalul valorilor (K = 3-15 clusteri)</b>	<b>Valoare maximă</b>	<b>K<sub>max</sub></b>	<b>Algoritmul pentru care a fost atinsă valoarea maximă</b>	<b>Cazul de selecție de biomarkeri (pentru valoarea maximă)</b>
<b>P<sub>3</sub></b>	0,4179-0,5	<b>0,5</b>	3	K-medoids	<i>Cazul 1</i>
<b>RI<sub>3</sub></b>	0,5459-0,5845	<b>0,5845</b>	3	K-medoids	<i>Cazul 1</i>
<b>P<sub>max</sub></b>	0,5555-0,6866 (K = 7-15)	<b>0,6866</b>	14	K-medoids	<i>Cazul 3</i>
<b>RI<sub>max</sub></b>	0,6381 și 0,6721 (K = 8-15)	<b>0,6721</b>	14	K-medoids	<i>Cazul 3</i>

#### 5.4. Discuții

Variabilitatea mai mică obținută pentru VEGF și TNF- $\alpha$  în cadrul claselor clinice a condus la stabilirea cazurilor în care au fost selectate perechi de biomarkeri (*Cazurile 1 și 2*). Rezultatele obținute în urma implementării tehnicii ACP indică un procentaj relativ bun de explicare a varianței setului de date, cu un minim de 75% atunci când au fost selectate cel puțin 2 componente principale.

Evaluarea cantitativă a acurateții procesului de clusterizare a revelat anumite legături între valorile obținute pentru măsurile utilizate (P și RI) și numărul de clusteri (K) pentru care acestea au fost obținute, cât și parametrii biochimici selectați. În general, evaluarea suprapunerii directe dintre clasele clinice și grupurile matematice (realizată prin P<sub>3</sub> și RI<sub>3</sub>) a condus la rezultate mai slabe față de evaluarea similarităților locale (cuantificată prin P<sub>max</sub> și RI<sub>max</sub>). Ambele evaluări au prezentat valori optime pentru algoritmul de analiză de clusteri K-medoids. Suprapunerea directă a fost cel mai bine evidențiată în *Cazul 1*, în schimb ce similaritățile locale au fost

revelate în mod optim atunci când toți cei 8 parametri serici au fost luați în considerare (*Cazul 3*).

### **5.5. Concluzii**

Studiul actual prezintă avantajul combinării mai multor tehnici de analiză de clusteri (K-means, K-medoids, Fuzzy c-means), dar și a analizei în componente principale în vederea efectuării unei analize complexe a legăturii dintre valorile unor parametri biochimici specifici și stadiile retinopatiei diabetice. Comparația dintre grupurile matematice formate și cele 3 clase clinice (FR, RNP, RP), implementată prin calcularea a doi indicatori specifici, P și RI, a evidențiat faptul că suprapunerea cea mai bună dintre clusteri și clasele clinice a fost obținută atunci când au fost selectați glicemia și VEGF, în schimb ce similaritățile locale au fost cel mai bine evidențiate considerând toți cei 8 parametri biochimici. Studii viitoare ar putea urmări mărirea numărului de pacienți din cadrul bazei de date, dar și includerea altor factori fiziologici și medicamentoși relevanți care ar putea contribui la dezvoltarea unui model relevant de predicție a retinopatiei diabetice.

## **6. Dezvoltarea unei aplicații de predicție personalizată a reacțiilor adverse în cazul pacienților cu polipatologie și polimedicatie**

### **6.1. Introducere**

În prezent, reacțiile adverse (RA) reprezintă un factor important care trebuie avut în vedere înainte de inițierea unei terapii medicamentoase, precum și în cursul acesteia. Pentru cuantificarea corespunzătoare a acestora, este necesară cunoașterea detaliată a factorilor de risc care contribuie la dezvoltarea RA, cum ar fi vârsta, patologia multiplă, polimedicatia, posologia, polimorfismul genetic, prezența insuficienței hepatice (IH) sau renale (IR) (Alomar, 2014). Așadar, scopul prezentului studiu a fost reprezentat de realizarea unei aplicații de predicție personalizată a reacțiilor adverse.

### **6.2. Material și metodă**

Aplicația a fost construită utilizând informațiile extrase din RCP-urile a 16 medicamente utilizate pentru a trata boli cronice: atorvastatină, rosuvastatină, enalapril, perindopril, metoprolol, indapamidă, metformin, omeprazol, amlodipină, trimetazidină, pentoxifilină,

candesartan, clopidogrel, fenofibrat, tamsulosin și betahistină. În plus, au mai fost utilizate baza de date Eudravigilance și MedDRA (Medical Dictionary of Regulatory Activities).

### **Dezvoltarea sistemului de scoruri**

Seturile de date rezultate au fost înglobate în sistemul de scoruri pe baza căruia aplicația a fost construită. Acest sistem a constat în 2 părți: frecvență și severitate, ambele fiind exprimate sub formă de probabilități (Learn. Poisson Bin. Distributions, 2015). Scorul de frecvență a fost calculat pe baza informațiilor din RCP-urile produselor conținând DCI-urile selectate iar scorul de severitate a utilizat detaliile din Eudravigilance. Ambele scoruri au fost calculate pentru fiecare RA unică în parte.

Construirea scorului de severitate a reprezentat partea personalizată a aplicației de reacții adverse prezentate, întrucât a ținut cont de schema terapeutică, precum și de vârsta, sexul și patologiiile relevante ale pacientului (IR, IH, insuficiență cardiacă (IC) și diabet zaharat (DZ)). Într-o primă etapă, pe baza acestor 4 factori, dar și a RA analizate (predictori, input), a fost realizată o predicție a rezultatului și a spitalizării (valori prezise, output). Ambele predicții reflectă probabilitatea de apariție a unui eveniment negativ (rezultat nefavorabil sau spitalizare). Regresia logistică a fost clasificatorul utilizat în acest sens (Drotar, 2014). În final, din predicția rezultatului și a spitalizării, a fost calculat un scor combinat care evidențiază severitatea RA analizate.

Atât scorul de frecvență, cât și cel de severitate au reprezentat o măsură importantă a ierarhizării reacțiilor adverse. Practic, pentru o anumită medicație și pentru un pacient cu anumite caracteristici, RA pot fi sortate în funcție de frecvență sau severitate, în funcție de preferința utilizatorului și sunt listate în funcție de indicele lor în ordinea descrescătoare a scorului solicitat. După ce scorul de frecvență și de severitate au fost obținute, a fost creat și un scor total, calculat per pacient și care cuantifică riscul total de a dezvolta RA, luând în considerare frecvența și severitatea pentru toate RA listate, precum și un scor calculat pentru fiecare SOC în parte (System Organ Class, clasificarea pe aparate și sisteme), care evidențiază riscul asociat fiecărui grup de RA.

## **Validarea scorului de severitate**

Pentru evaluarea performanțelor aplicației personalizate de RA, a fost concepută o validare a modelului de predicție a severității, pe baza a 54663 de pacienți din Eudravigilance cu un număr de 228 RA unice raportate. Au fost calculați mai mulți indicatori statistici: acuratețea, precizia, sensibilitatea, scorul F1 (pentru predicția clasei de rezultat și spitalizare), scorul Brier și scorul ROC AUC (Area under the Receiver Characteristic Operating Curve) (pentru estimarea probabilității de apariție a unui eveniment sever asociat RA) (Drotar, 2014).

## **Sistemul de interfață grafică și dezvoltarea aplicației**

Dezvoltarea propriu-zisă a aplicației s-a bazat pe construirea unei interfețe grafice (în cadrul limbajului de programare Python, versiunea 3.5.2), utilizată pentru introducerea datelor care vor fi prelucrate (vârstă, sex, medicamente, patologii) și prezentarea rezultatelor (datele de ieșire) (GUI with Tk, 2017). După procesarea datelor și calcularea scorurilor, aplicația prezintă ca date de ieșire liste sortate de RA, cu sau fără detalii pe termenii unici, în funcție de opțiunea aleasă. Aplicația poate fi executată cu 3 opțiuni principale, cea generală, când se iau în considerare toate RA din polimedicație, cea multiplă, atunci când numai anumite RA sunt luate în considerare și cea specifică, care permite vizualizarea detaliilor pe RA selectată. Fig. 6.1 prezintă în mod schematic opțiunile și datele de ieșire ale aplicației.

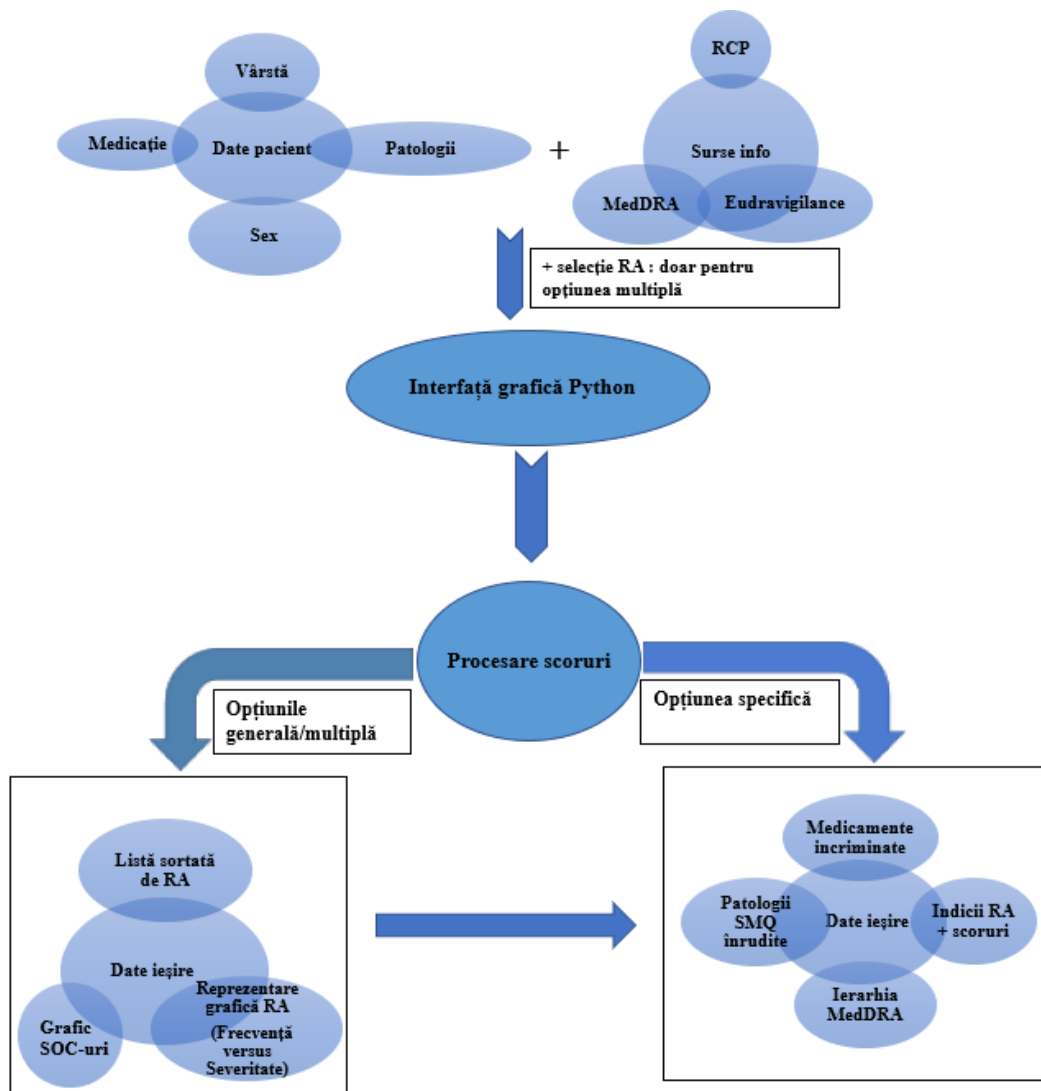


Fig. 6.1. Opțiunile și datele de ieșire ale aplicației personalizate de RA

### 6.3. Rezultate

#### Exemplu de executare a aplicației de RA (pentru opțiunea generală)

Figura 6.2 prezintă lista sortată de RA în urma executării aplicației pentru un pacient din baza de date Eudravigilance. Este vorba de o persoană de sex feminin, în vârstă de 70 de ani, aflată în tratament cu candesartan, enalapril, metoprolol, omeprazol și rosuvastatină. Pacienta a dezvoltat 4 RA pe parcursul tratamentului: angină pectorală (Angina pectoris), bradicardie (Bradycardia), tuse (Cough) și oboseală (Fatigue). Este prezentată lista sortată în funcție de

severitate. În cadrul tezei de doctorat sunt prezentate și alte opțiuni și exemplificări ale aplicației, relevante din punct de vedere clinic și al recunoașterii de forme.

ADR Rank	Adverse reaction	Number of drugs	Frequency	Severity
1	Alveolitis allergic	1	0.00055	0.9464
2	Bilateral pulmonary infiltrates	1	0.00055	0.9367
3	Hepatitis	5	0.001265	0.9362
4	Stevens Johnson syndrome	3	0.00066	0.9242
5	Hyponatraemia	3	0.006102	0.9239
6	Ileus	1	0.0055	0.9235
7	Pancreatitis	2	0.006047	0.9189
8	Retroperitoneal fibrosis	1	5.5e-05	0.916
9	Thrombocytopenia	4	0.001704	0.9074
10	Agranulocytosis	3	0.00066	0.9039
11	Eosinophilic pneumonia	1	0.00055	0.9035
12	Toxic epidermal necrolysis	2	0.000605	0.9008
13	Neuropathy peripheral	1	5.5e-05	0.8959
14	Polyneuropathy	1	5.5e-05	0.8922
15	Orthostatic hypotension	2	0.060198	0.8917
16	Hepatic enzymes increased	4	0.006648	0.8902
17	Confusion	3	0.006102	0.8823
18	Pancytopenia	2	0.000605	0.8805
19	Vomiting	3	0.111887	0.88
20	Cholestasis	1	0.00055	0.8798

Total Score	ADR details	SOC plot	ADR 2D plot
0.9505			

Fig. 6.3. Lista sortată cu primele 20 de RA din punct de vedere al severității pentru cazul prezentat. Sunt prezentate: indicele de severitate a fiecărei RA (ADR Rank), numărul medicamentelor incriminate în producerea RA respective (Number of drugs), scorul de frecvență (Frequency) și cel de severitate (Severity).

### Validarea scorului de severitate

Tabelul 6.1 prezintă rezultatele obținute în urma validării scorului de severitate. Rezultatele au fost considerate satisfăcătoare, ținând cont de informațiile disponibile în baza de date Eudravigilance. Sunt prezentate rezultatele obținute pe setul de testare, pentru evidențierea performanței modelului pe cazuri necunoscute.

Tabel 6.1. Rezultatele validării scorului de severitate, realizată pe datele a 54663 de pacienți din baza de date Eudravigilance

<b>Rezultatele validării scorului de severitate</b>		
<b>Măsura de validare</b>	<b>Predicția clasei de rezultat</b>	<b>Predicția clasei de spitalizare</b>
Acuratețe	60,03%	65,07%
Precizie	56,59%	69,24%
Sensibilitate	58,75%	64,32%
Scor F1	57,61%	66,66%
<b>Măsura de validare</b>	<b>Predicția Pr<sub>rezultat</sub> (clasa 1)</b>	<b>Predicția Pr<sub>spitalizare</sub> (clasa 1)</b>
Scor Brier	23,38%	21,84%
Scor ROC AUC *	64,15%	70,51%

\*Cea mai importantă măsură de validare.

#### **6.4. Discuții**

Sistemul de scoruri al aplicației de RA prezentate permite o diferențiere a reacțiilor adverse care pot apărea pentru un anumit pacient cu o anumită polimedicație. Modul de dezvoltare a aplicației asigură extragerea și sortarea celor mai relevante RA atât din punct de vedere al frecvenței, cât și al severității, în funcție de preferințele utilizatorului.

Ținând cont de exemplificarea oferită și de aspectele menționate, se poate considera că opțiunea generală a aplicației personalizate de RA poate fi utilizată pentru vizualizarea profilului complet de RA al unui anumit pacient cu niște caracteristici specifice și care primește un anumit tratament medicamentos, având avantajul de a permite extragerea celor mai relevante reacții adverse care pot apărea, alături de importanța acestora, oferită prin afișarea indicilor corespunzători, a numărului de medicamente cu relație de cauzalitate, precum și a valorilor celor două tipuri de scoruri. După analiza listei afișate, medicul sau farmacistul clinician se pot orienta înspre vizualizarea unor detalii specifice pentru efectele adverse considerate ca fiind relevante din punct de vedere clinic.

De asemenea, este de menționat faptul că o importanță deosebită în cadrul interpretării rezultatelor generate de program o au cele două analize grafice care pot fi realizate în urma vizualizării celor două tipuri de reprezentări: cea bidimensională a RA și cea de tip coloană a scorurilor SOC. Aceste grafice își pot dovedi utilitatea prin efectuarea unei comparații directe între profilul de RA a două sau mai multe combinații medicamentoase utilizate pentru a trata aceleași afecțiuni sau între doi pacienți care primesc același tratament farmacologic.

Validarea modelului de predicție a severității RA a condus la valori maxime de 70,51% în privința estimării probabilității de apariție a unui eveniment sever asociat RA, rezultat obținut prin calcularea scorului ROC AUC. Valorile obținute au fost considerate satisfăcătoare ținând cont de variabilele predictive prezente în Eudravigilance. Astfel, este de menționat că limitările principale ale aplicației dezvoltate au fost reprezentate de faptul că nu s-a putut ține cont de anumiți factori relevanți, cum ar fi polimorfismul genetic, rasa sau posologia, interacțiunile farmacocinetice fiind alt factor important care nu a fost inclus în cadrul programului.

## **6.5. Concluzii**

Studiul prezentat și-a propus și a reușit să creeze o aplicație personalizată de reacții adverse pentru pacienții cu polipatologie și polimedicatie, care îi permite utilizatorului să analizeze cele mai relevante RA care ar putea apărea pentru un anumit pacient cu un anumit tratament farmacologic. Modelul a fost construit prin combinarea mai multor surse de informație, în speță RCP-urile actualizate ale medicamentelor selectate, dicționarul MedDRA și setul de date Eudravigilance furnizat de către Agenția Europeană a Medicamentului. Pe baza informațiilor detaliate cuprinse în cele trei surse selectate, două tipuri de scoruri au fost generate pentru fiecare RA posibilă, unul de frecvență și altul de severitate, primul bazat pe incidența reacției adverse analizate iar cel de-al doilea pe probabilitatea de apariție a unui eveniment sever în cazul în care RA apare. Sistemul de scoruri care a fost creat, alături de diversele opțiuni ale aplicației permit analiza reacțiilor adverse din mai multe puncte de vedere, incluzând o interpretare grafică și o cuantificare prin intermediul indicilor a listei obținute. Programul ar putea fi utilizat de către specialiștii în domeniul sănătății în cadrul unei predicții personalizate de RA, dar și pentru monitorizarea siguranței tratamentului urmat de pacient și pentru optimizarea terapiei medicamentoase.

## 7. Concluzii generale

Teza de doctorat elaborată și-a propus și a reușit să prezinte contribuțiile personale ale doctorandului în privința multiplelor aplicații pe care recunoașterea de forme le poate prezenta în analiza multivariată a datelor rezultate din cercetarea farmaceutică. Seturile de date cu complexitate ridicată au făcut justificată utilizarea algoritmilor de recunoaștere de forme în vederea construirii unor modele teroretice din date necunoscute, cum a fost cazul analizei nesupervizate, însă și pentru evidențierea acurateții de clasificare și a capacității predictive a informației procesate, în situația în care învățarea supervizată a fost selectată în vederea prelucrării datelor obținute.

Partea originală a tezei de doctorat a cuprins aplicarea mai multor tehnici de recunoaștere nesupervizată și supervizată de forme în vederea prelucrării și analizei mai multor seturi de date rezultate din cercetarea farmaceutică. În majoritatea cazurilor, tehnicile de tip nesupervizat au fost utilizate în combinație cu cele de tip supervizat în cadrul aceluiași capitol, demonstrând astfel atât complexitatea informațiilor analizate, cât și utilitatea implementării în combinație a mai multor astfel de metode.

Astfel, Capitolul 4 a implementat atât o analiză de clusteri de tip K-means, cât și o tehnică de tip kNN în vederea caracterizării albuminei umane carbonilate cuantificate prin electroforeză capilară. Capitolul 5 a utilizat cu precădere o recunoaștere nesupervizată, prin implementarea algoritmilor K-means, K-medoids și FCM în vederea evaluării unor parametri biochimici relevanți pentru caracterizarea stadiilor retinopatiei diabetice. Capitolul 6 a utilizat metode de analiză de date, precum și tehnici de programare și algoritmi de recunoaștere și învățare automată în vederea dezvoltării unei aplicații de predicție personalizată a RA în cazul pacienților cu poliatologie și polimedicatie.

Una dintre limitările importante ale tezei de doctorat a fost reprezentată de către scorurile de sub 70% care au fost obținute în privința validării algoritmilor implementați, atât în cazul celor de recunoaștere nesupervizată (Capitolul 5), cât și în cazul celor de analiză supervizată (Capitolul 6).

Ideile și rezultatele originale prezentate sunt:

- i) Utilizarea în combinație a analizei de clusteri și a recunoașterii supervizate în vederea construirii unui model de albumină umană carbonilată cuantificată prin intermediul electroforezei capilare (Capitolul 4)
- ii) Implementarea unui algoritm complex de stabilire a apartenenței electroforegramelor pacienților diabetici la modelul de carbonilare proteică, prin utilizarea mai multor parametri definitorii: timpul de retenție, aria picului, înălțimea picului, putând fi făcute astfel supoziții referitoare la nivelul de stres oxidativ (Capitolul 4)
- iii) Utilizarea recunoașterii nesupervizate de forme în vederea stabilirii legăturii dintre valorile anumitor parametri biochimici și stadiile retinopatiei diabetice – rezultatele obținute pot conduce în studii viitoare la instituirea unui model simplu și eficace de diagnostic și stadializare a retinopatiei diabetice (Capitolul 5)
- iv) Stabilirea unor legături complexe între grupurile obținute în urma implementării analizei de clusteri și clasificarea clinică disponibilă în urma controlului oftalmologic la care au fost supuși pacienții diabetici (Capitolul 5)
- v) Îmbogățirea instrumentelor de predicție personalizată a reacțiilor adverse disponibile la momentul actual în sănătate cu o aplicație complexă și simplă de utilizat (Capitolul 6)
- vi) Punerea la punct a unui sistem de scoruri care permite analiza individuală și stratificarea reacțiilor adverse în cazul unui pacient cu anumite caracteristici și un anumit tratament medicamentos (Capitolul 6)
- vii) Caracterizarea reacțiilor adverse atât din punct de vedere al frecvenței, cât și al severității, sub aspectul analizei separate a celor 2 parametri, precum și al generării unui scor final care să permită stratificarea pacienților din punct de vedere al profilului farmacotoxicologic personalizat (Capitolul 6)

În concluzie, teza de doctorat expusă și-a propus și a reușit implementarea unor tehnici de recunoaștere de forme și a unor combinații ale acestora în vederea analizei unor seturi de date multivariate rezultate din cercetarea farmaceutică. Capitolele prezentate au demonstrat abilitatea metodelor aplicate de a caracteriza în mod coerent și precis informațiile disponibile, în pofida diferențelor seturilor de date în privința dimensionalității, subdomeniului din cadrul științelor farmaceutice sau a scopului final care a fost urmărit. Algoritmii care au fost detaliați au contribuit la construirea unor modele complexe, capabile să extragă informații esențiale cu impact major în cercetarea de specialitate.

## Bibliografie selectivă

- Alomar, M. (2014). Factors affecting the development of adverse drug reactions (Review article). *Saudi Pharmaceutical Journal*, 22(2), pp.83-94.
- Arora, P., Deepali, D. și Varshney, S. (2016). Analysis of K-Means and K-Medoids Algorithm For Big Data. *Procedia Computer Science*, 78, pp.507-512.
- Brereton, R. (2003). *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Chichester: Wiley, capitol: *Pattern Recognition*, pp. 183-269. ISBN: 9780471489771.
- Daskalakis, C., Diakonikolas, I. și Servedio, R.A., (2015). *Learning Poisson Binomial Distributions*. [online] Disponibil la: <http://www.iliasdiakonikolas.org/papers/pbds-learning.pdf> [Accesat pe 17.06.2018].
- Dima, I., Purdel, C., **Văleanu, A.**, Moldoveanu, G., Ilie, M. și Margină, D. (2018). Capillary Electrophoresis For The Evaluation Of The Carbonylation Pattern In Type 2 Diabetes Mellitus. Which Are The Premises? În: *C. Serafinceanu, O. Negoită, V. Elian (Eds.), InterDIAB 2018 Book Series, 4<sup>th</sup> International Conference on Interdisciplinary Management of Diabetes Mellitus and its Complications: „Surgical Crossroads with Diabetes Mellitus”, Editura Niculescu*. București, România, pp. 268-276.
- Drotar, P. și Smekal, Z. (2014). Comparative study of machine learning techniques for supervised classification of biomedical data. *Acta Electrotechnica et Informatica*, 14(3), pp. 5–10. DOI: 10.15546/aei-2014-0021.
- DuBrava, S., Mardekian, J., Sadosky, A., Bienen, E.J., Parsons, B., Hopps, M. și Markman, J. (2017). Using Random Forest Models to Identify Correlates of a Diabetic Peripheral Neuropathy Diagnosis from Electronic Health Record Data. *Pain Medicine*, 18(1), pp. 107–115.
- James, G., Witten, D., Hastie, T. și Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. *Springer Texts in Statistics*, 103. Springer Science+Business Media New York. DOI 10.1007/978-1-4614-7138-7.
- Manning, C., Raghavan, P. și Schütze, H. (2008). Evaluation of clustering. În: *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, pp.356-360, ISBN: 0521865719.
- Ngufor, C. și Wojtusiak, J. (2015). A Systematic Prediction of Adverse Drug Reactions Using Pre-clinical Drug Characteristics and Spontaneous Reports. In: *P. Balakrishnan, J. Srivatsava, W.T. Fu, S. Harabagiu, F. Wang (Eds.), Proceedings of 2015 IEEE International Conference*

on *Healthcare Informatics (ICHI 2015)*, Dallas, Texas, SUA, pp. 76-8.  
DOI: [10.1109/ICHI.2015.16](https://doi.org/10.1109/ICHI.2015.16).

Python Software Foundation, (2017). *Graphical User Interfaces with Tk* [online] Disponibil la: <https://docs.python.org/3/library/tk.html> [Accesat pe 17.06.2018].

Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T. și Gonzalez, G. (2015). Utilizing Social Media Data for Pharmacovigilance: A Review. *Journal of Biomedical Informatics*, 54, pp. 202-212.

Sharma, P. și Kaur, M. (2013). Classification in Pattern Recognition: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(4), pp. 298-306.

Sutariya, V., Groshev, A., Sadana, P., Bhatia, D. și Pathak, Y. (2013). Artificial Neural Network in Drug Delivery and Pharmaceutical Research. *The Open Bioinformatics Journal*, 7(1), pp. 49-62.

Uttara, B., Singh, A., Zamboni, P. și Mahajan, R. (2009). Oxidative Stress and Neurodegenerative Diseases: A Review of Upstream and Downstream Antioxidant Therapeutic Options. *Current Neuropharmacology*, 7(1), pp.65-74.

**Văleanu, A., Ilie, M., Dima, I. și Purdel, C. (2016a).** K Nearest Neighbours analysis of human serum carbonyl proteins using capillary electrophoregrams. *Romanian Journal of Biophysics*, 26(1), pp. 001-010.

**Văleanu, A., Margină, D.M., Grădinaru, D., Ilie, M., Dima, I.E., Purdel, C.N. și Dănciulescu-Miulescu, R. (2016b).** Diabetic retinopathy and inflammation: a comparative statistical study on relevant blood serum parameters. În: *C. Serafinceanu, O. Negoită, V. Elian (Eds.), InterDIAB 2016 Book Series, International Conference on Interdisciplinary Management of Diabetes Mellitus and its Complications: „Diabetes Mellitus as Cardiovascular Disease”, Editura Niculescu. București, România, pp. 490-499.*

Yin, J., Sun, H., Yang, J. și Guo, Q. (2014). Comparison of K-Means and Fuzzy c-Means Algorithm Performance for Automated Determination of the Arterial Input Function. *PLoS ONE*, 9(2), p.e85884.