



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Fondul Social European
POSDRU 2007-2013



Instrumente Structurale
2007-2013



MINISTERUL
EDUCAȚIEI
NAȚIONALE

OIPOSDRU



Universitatea Politehnică
din București

Universitatea „DUNĂREA DE JOS” din Galați



TEZĂ DE DOCTORAT

**METODE AVANSATE DE DETECȚIE ȘI IDENTIFICARE
A ORIGINII UNOR PRODUSE VEGETALE FOLOSITE IN
INDUSTRIA ALIMENTARĂ**

**ADVANCED METHODS OF DETECTION AND
IDENTIFICATION OF THE ORIGIN OF PLANT
PRODUCTS USED IN FOOD INDUSTRY**

**Doctorand,
Ing. Constantin Simona (Ghiniță Constantin)**

**Conducător științific,
prof. univ. dr. Mirela PRAISLER**

Seria I 4 : Inginerie industrială Nr. 22

2014

Studiile experimentale și dezvoltarea sistemului expert utilizat în identificarea / recunoașterea / detecția automată a originii unor produse vegetale folosite în industria alimentară au fost efectuate în:

- Centrul de Analize Fizico – Chimice, Morfo – Funcționale și Chemometrie, Facultatea de Științe și Mediu, Galați.
- Laboratorul de analize senzoriale, Facultatea de Știința și Ingineria Alimentelor.

Studiile experimentale descrise în această lucrare au fost efectuate în cadrul proiectului finanțat de Uniunea Europeană POSDRU/159/1.5/S/132397

Această lucrare a fost realizată cu suportul financiar al proiectului Programul Operational Sectorial Dezvoltarea Resurselor Umane, 2007-2013 al Ministerului Fondurilor Europene POSDRU / 159 / 1.5 / S / 132397.

Autorul este recunoscător pentru sprijinul financiar acordat de Uniunea Europeană, Guvernul României și Universitatea "Dunărea de Jos" din Galați.

Mulțumiri,

Incheierea procesului de elaborare a lucrării de doctorat, reprezintă o altă etapă în ceea ce privește formarea mea profesională și prezintă rezultatele obținute în cei trei ani de studii în cadrul școlii doctorale.

Teza de doctorat reprezintă o continuare a lucrării de disertație intitulată “Studiul fizico – chimic a unor compuși caracteristici produselor vegetale de interes alimentar”. Având în vedere importanța acestei teme și a efectului originii geografice asupra proprietăților produselor vegetale, am aprofundat studiul acestei teme.

La finalul acestei etape, gândurile și recunoștința mea se întreprind către cei ce m-au sprijinit, susținut și încurajat pe parcursul celor 3 ani de studiu, în vederea elaborării și finalizării tezei de doctorat.

În primul rând doresc să-i mulțumesc în mod deosebit coordonatorului științific, doamnei prof. dr. Mirela Praisler, pentru tot sprijinul pe care mi l-a acordat atât pe parcursul studiilor de master cât și de doctorat. Elaborarea și finalizarea acestei lucrări nu ar fi fost posibilă fără contribuția, profesionalismul academic, răbdarea și înțelegerea dânzei, care m-a încurajat și îndrumat permanent.

Mulțumesc foarte mult doamnei conf.dr.ing. Gabriela Iordăchescu, membru al Facultății de Știința și Ingineria Alimentelor, care m-a susținut, sfătuit și încurajat, fiind unul dintre membrii comisiei de îndrumare din perioada studiilor de doctorat. Îi mulțumesc și pentru contribuția deosebită în obținerea rezultatelor experimentale prezentate în teză.

Mulțumesc doamnei prof. dr. Luminița Moraru și domnului prof. dr. Lucian Georgescu, membri ai Facultății de Științe și Mediu, pentru analizele obiective și sfaturile date în calitate de membri ai Comisiei de îndrumare, cu ocazia prezentării referatelor științifice pe parcursul studiilor mele doctorale.

Mulțumesc conducerii și cadrelor didactice din cadrul Colegiului Tehnic “Costin D. Nenițescu”, Brăila, pentru susținerea acordată pe tot parcursul studiilor de doctorat.

Și nu în ultimul rând, vreau să mulțumesc în mod special familiei mele care m-a încurajat, susținut și sprijinit necondiționat.

Studiile experimentale și dezvoltarea metodelor avansate de detecție și identificare a originii unor produse vegetale folosite în industria alimentară au fost efectuate în cadrul Facultății de Științe și Mediu și în cadrul Facultății de Știința și Ingineria Alimentelor, din cadrul Universității “Dunărea de Jos” din Galați.

Galați, Noiembrie 2014

Ing. Ghiniță Constantin Simona

CUPRINS :

INTRODUCERE	6
Capitolul I - Aspecte teoretice privind produsele vegetale selecționate pentru caracterizarea și identificarea originii geografice a acestora	9
Capitolul II - Descrierea metodelor spectrale și multivariate folosite pentru detecția și identificarea originii geografice a produselor vegetale studiate	13
CAPITOLUL III - Contribuții proprii obținute în detecția și identificarea originii geografice a produselor vegetale folosite în industria alimentară prin analiza spectrală	16
III.1. Condiții experimentale	16
III.2. Identificarea ferestrelor spectrale UV-VIS care au putere de modelare și de discriminare a originii geografice	17
III.2.1. Mărar (<i>Anethum Graveolens</i>)	17
III.2.2. Leuștean (<i>Levisticum officinale</i>)	18
III.3. Analiza comparativa a puterii de discriminare a absorbțiilor UV-VIS în funcție de tipul vegetalelor	18
III.3.1. Mazăre (<i>Pisum sativum</i>)	19
III.3.2. Cimbru (<i>Satureja Hortensis</i>)	19
III.4. Discriminarea ciupercilor <i>Pleurotus (Pleurotus ostreatus)</i> și Champignon (<i>Agaricus bisporus</i>) prin analiză spectrală	20
III.5. Concluzii	21
Capitolul IV - Contribuții proprii obținute în detecția și identificarea originii geografice a produselor vegetale folosite în industria alimentară prin Analiza Componentelor Principale (PCA)	21
IV.1. Condiții experimentale	21
IV.2. Identificarea originii geografice a produselor vegetale folosite în industria alimentară prin PCA	22
IV.2.1. Mărar (<i>Anethum Graveolens</i>)	22
IV.2.2. Leuștean (<i>Levisticum officinale</i>)	23
IV.3. Identificarea principalelor caracteristici fizico-chimice care influențează calitatea produselor vegetale folosite în industria alimentară prin PCA	25
IV.3.1. Mazăre (<i>Pisum sativum</i>)	25
IV.3.2. Cimbru (<i>Satureja Hortensis</i>)	26
IV.4. Discriminarea ciupercilor <i>Pleurotus (Pleurotus ostreatus)</i> și Champignon (<i>Agaricus bisporus</i>) prin PCA	27
IV.5. Concluzii	28
Capitolul V - Contribuții proprii obținute în detecția și identificarea originii geografice a produselor vegetale folosite în industria alimentară prin Analiza Ierarhica a Clusterelor (HCA)	29
V.1. Condiții experimentale	30
V.2. Identificarea originii geografice a produselor vegetale folosite în industria alimentară prin HCA	30
V.2.1. Mărar (<i>Anethum Graveolens</i>)	30
V.2.1.1. Testarea sensibilității și selectivității sistemului de identificare	30
V.2.1.2. Evaluarea puterii de generalizare a sistemelor HCA pentru identificarea eficientă a originii geografice	32

V.2.1.3. Evaluarea numarului de zone geografice de origine ce pot fi testate concomitent prin HCA	33
V.2.2. Leuștean (<i>Levisticum officinale</i>)	34
V.3. Discriminarea ciupercilor <i>Pleurotus</i> (<i>Pleurotus ostreatus</i>) și Champignon (<i>Agaricus bisporus</i>) prin HCA	35
V.4. Concluzii	36
Capitolul VI - Contribuții proprii obținute în detecția și identificarea originii geografice a produselor vegetale folosite în industria alimentară prin Rețele Neuronale Artificiale (ANN)	37
VI.1. Condiții experimentale	37
VI.2. Identificarea originii geografice a produselor vegetale folosite în industria alimentară prin ANN	38
VI.2.1. Mărar (<i>Anethum Graveolens</i>)	38
VI.2.1.1. Instruirea sistemului ANN	38
VI.2.1.2. Validarea sistemului ANN	38
VI.3. Leuștean (<i>Levisticum Officinale</i>)	39
VI.3.1. Instruirea sistemului ANN	39
VI.4. Discriminarea ciupercilor <i>Pleurotus</i> (<i>Pleurotus ostreatus</i>) și Champignon (<i>Agaricus bisporus</i>) prin ANN	39
Capitolul VII - Concluzii generale, contribuții originale și perspective viitoare de cercetare	43
VII.1. Concluzii generale	48
VII.2. Contribuții originale	48
VII.3. Perspective viitoare de cercetare	49
LISTĂ LUCRĂRI PUBLICATE ȘI PREZENTATE	50

INTRODUCERE

Teza de doctorat reprezintă o continuare a lucrării de disertație intitulată “Studiul fizico – chimic a unor compuși caracteristici produselor vegetale de interes alimentar”. Am aprofundat studiul acestei teme având in vedere importanța efectului originii geografice asupra calității produselor vegetale folosite in industria alimentară. Rezultatele obținute s-au materializat prin diverse aplicații de inteligență artificială ce pot fi folosite ca instrumente automate de detecție și identificare a originii produselor vegetale, toate acestea fiind prezentate in prezenta lucrare.

Cunoașterea zonei geografice este importantă datorită faptului că proprietățile fizico-chimice ale produselor vegetale diferă in funcție de condițiile pedo-climatice in care acestea se dezvoltă. Multe proprietăți care asigură vegetalelor calitățile nutritive, inclusiv cele care influențează proprietățile senzoriale, depind de condițiile pedo-climatice din zona de origine [1]. Altfel spus, condițiile pedo-climatice favorabile potențează anumite proprietăți fizico-chimice și contribuie astfel la îmbunătățirea proprietăților senzoriale [2]. Atât proprietățile, cât și randamentul produselor vegetale sunt influențate de condițiile pedologice și condițiile climatice conexe (climatice, precipitații, temperaturile medii maxime și temperaturi medii minime, umiditate, înghețurile, etc.) din zona lor geografică de origine. Astfel, identificarea principalelor proprietăți fizico-chimice care definesc calitatea produselor vegetale in funcție de originea geografică a acestora este foarte importantă atât pentru industria alimentară cât și farmaceutică.

Calitatea nutrițională, gastronomică și/sau terapeutică a produselor vegetale poate suferi modificări semnificative in funcție de condițiile pedo-climatice specifice locului de recoltare [3]. Efectul acestor variabile asupra calității produselor poate fi decelat prin evaluarea caracteristicilor fizico-chimice ale probelor vegetale. Inșă această corelație este foarte complexă. Ca atare, pentru modelarea ei se impune o abordare multivariată, respectiv dezvoltarea unor aplicații chemometrice, de inteligență artificială, dezvoltate pentru identificarea automată a originii geografice a probelor vegetale. Acestea prezintă avantajul că, odată validate, pot constitui mijloace automate de identificare a originii probelor vegetale pe baza proprietăților fizico-chimice a acestora.

Principalele proprietăți fizico-chimice care pot fi influențate de originea geografică au fost determinate pentru următoarele produse vegetale: mărar (*Anethum graveolens*), leuștean (*Levisticum officinale*), ciuperci *Pleurotus* și ciuperci *Champignon* brun, cimbru (*Satureja hortensis*) și mazăre (*Pisum sativum*).

Prezenta lucrare prezintă rezultatele obținute cu ajutorul unor aplicații multivariate dezvoltate pentru detecția și identificarea originii unor produse vegetale folosite in industria alimentară. S-a urmărit dezvoltarea unor sisteme care sa realizeze in mod automat identificarea originii geografice a vegetalelor pe baza proprietăților fizico-chimice ale acestora. Metodele de inteligență artificială propuse permit nu numai detecția automată, dar și identificarea originii unor probe necunoscute.

Avantajele majore ale sistemelor dezvoltate și descrise în această teză sunt următoarele: a) automatizarea procesului de detecție permite recunoașterea probelor vegetale in funcție de zona lor de proveniență într-un timp foarte scurt (1-2 secunde); b) sensibilitatea și selectivitatea sistemelor de recunoaștere automată sunt excelente; c) detecția astfel realizată canalizează analistul spre analize ulterioare bine definite și în acest mod se economisesc resurse analitice și umane importante.

Aplicațiile au fost realizate cu ajutorul aplicațiilor software *The Unscrambler*, *Matlab 2013a* și *nnModel* (Neural Network Model). Metodele chemometrice au fost selectate astfel

încât să necesite resurse computaționale minime. Acestea trebuie să poată fi operate cu ajutorul unor unități de calcul adecvate instrumentelor de detecție portabile, respectiv unități ce nu necesită capacități de memorie deosebite.

Teza de doctorat cu titlul "*Metode avansate de detecție și identificare a originii unor produse vegetale folosite în industria alimentară*" este structurată în șapte capitole, în care sunt prezentate noțiunile teoretice și contribuțiile proprii obținute în dezvoltarea unor sisteme automate de detecție și identificare a originii unor produse vegetale. Lucrarea de față cuprinde o parte introductivă care descrie problema științifică, oportunitatea și motivația studierii acesteia, precum și modul în care au fost atinse obiectivele cercetării.

În capitolul I, sunt prezentate aspecte teoretice privind problematica și produsele vegetale studiate. Sunt sintetizate rezultatele unei cercetări bibliografice privind caracteristicile produselor vegetale selecționate pentru dezvoltarea aplicațiilor multivariate, respectiv mărar, leuștean, ciuperci, cimbru și mazărea.

În capitolul II sunt descrise aparatura principală și metodele multivariate folosite pentru detecția și identificarea originii produselor vegetale, respectiv analiza spectrală, PCA, HCA și ANN.

Capitolul III prezintă rezultatele proprii obținute în detecția și identificarea originii produselor vegetale folosite în industria alimentară prin analiza spectrală în cazul mărarului (*Anethum Graveolens*), ciupercilor *Pleurotus* (*Pleurotus ostreatus*) și Champignon (*Agaricus bisporus*), leușteanului (*Levisticum Officinale*), a boabelor de mazăre (*Pisum sativum*) și a cimbrului (*Satureja Hortensis*).

În Capitolul IV sunt descrise rezultatele proprii obținute în detecția și identificarea originii acestor produse vegetale printr-o metodă de recunoaștere a formelor nesupravegheată și non-ierarhică, respectiv PCA.

Capitolul V prezintă rezultatele proprii obținute în detecția și identificarea originii produselor vegetale folosite în industria alimentară printr-o metodă de recunoaștere a formelor nesupravegheată dar ierarhică, respectiv HCA.

În fine, capitolul VI descrie rezultatele proprii obținute în detecția și identificarea originii produselor vegetale folosite în industria alimentară prin ANN, care este o metodă de recunoaștere a formelor atât ierarhică cât și supravegheată.

În cel de-al VII capitol sunt prezentate concluziile generale, principalele contribuții originale și perspectivele viitoare de cercetare. Concluziile generale și contribuțiile originale sintetizează rezultatele obținute în dezvoltarea și optimizarea sistemelor automate de detecție și identificare a originii geografice a produselor vegetale analizate. Capitolul se încheie cu menționarea unor potențiale direcții de aplicare și dezvoltare a rezultatelor prezentate în cadrul acestei teze.

Identificarea originii geografice a produselor vegetale este importantă deoarece calitatea lor nutrițională, gastronomică și/sau terapeutică poate suferi modificări semnificative în funcție de condițiile pedo-climatice specifice locului de producție și recoltare.

Efectul proprietăților pedo-climatice asupra calității produselor poate fi decelat prin evaluarea caracteristicilor fizico-chimice ale probelor vegetale. Prin această lucrare am dorit să dezvoltăm sisteme multivariate care să permită detecția / identificarea / recunoașterea automată și eficiența originii produselor vegetale pe baza unor selecții de proprietăți fizico-chimice.

Scopul acestei cercetări a fost obținerea unor sisteme expert care să realizeze detecția și identificarea originii unor produse vegetale folosite în industria alimentară cu ajutorul unor metode chemometrice multivariate. A fost studiată eficiența unor sisteme dezvoltate cu metode

de recunoaștere a formelor nesupravegheate și supravegheate, precum și metode și non-ierarhice și ierarhice.

Investigația științifică a avut stabilite câteva obiective specifice, cum ar fi:

1) prezentarea generală a principalelor tehnici și metode de analiză multivariată utilizate în caracterizarea și identificarea originii produselor vegetale studiate;

2) dezvoltarea unor aplicații multivariate de detecție automată și identificare a originii produselor vegetale pe baza principalelor proprietăți fizico-chimice;

3) recunoașterea automată a originii unor probe necunoscute pe baza similarității proprietăților lor fizico-chimice cu cele ale vegetalelor și zonelor de origine modelate.

4) evaluarea eficienței sistemelor dezvoltate în ceea ce privește sensibilitatea și selectivitatea lor din punctul de vedere al recunoașterii originii probelor necunoscute.

5) evaluarea puterii de generalizare a rezultatelor obținute.

Bibliografie

- [1] J.M. Aguilera, D.W. Stanley, *Microstructural Principles of Food Processing and Engineering*. Springer, 1999, ISBN 0-8342-1256-0.
- [2] A. Davidson, *The Oxford Companion to Food*, 2nd ed. UK: Oxford University Press, 2006.
- [3] M.G. Harold, *On Food and Cooking: The Science and Lore of the Kitchen*, New York: Simon and Schuster, 2004, ISBN 0-684-80001-2.
- [4] P.A.G. Santos, A.C. Figueiredo, *Hairy root cultures of Anethum graveolens (dill): establishment, growth, time-course study of their essential oil and its comparison with parent plant oils*, doi:10.1023/A:1015653701265, *Biotechnology Letters* 24 (12): 1031-1036, 2002.
- [5] A. Smith, *Food Marketing in Oxford Encyclopedia of American Food and Drink*, New York: Oxford University Press, 2007.
- [6] P.G. Miles, S.T. Chang *Mushrooms: Cultivation, Nutritional Value, Medicinal Effect, and Environmental Impact*, Boca Raton, Florida: CRC Press, 2004, ISBN 0-8493-1043-1.
- [7] P. Harding, *Mushroom Miscellany*, HarperCollins. p. 149, 2008, ISBN 978-0-00-728464-1.
- [8] M.G. Pimenov, M.V. Leonov, *The Genera of the Umbelliferae*, Royal Botanic Gardens, Kew 1993, ISBN 0-947643-58-3
- [9] E.V. Nybe, M.Raj, K.V. Peter, *Spices: Vol.05. Horticulture Science Series*. New Delhi, New India Publishing. pp. 233, 234, 2007, ISBN 81-89422-44-8.
- [10] S. Jegtvig, *Peas, Nutrition*, About.com. Retrieved 2011-01-28, 2007.

CAPITOLUL I

ASPECTE TEORETICE PRIVIND PRODUSELE VEGETALE SELECȚIONATE PENTRU CARACTERIZAREA ȘI IDENTIFICAREA ORIGINII GEOGRAFICE A ACESTORA

I.1. Mărar (*Anethum graveolens*)

Mărarul aparține familiei *Apiaceae*, din genul *Anethum*, fiind cunoscut sub numele științific de *Anethum graveolens*. Este o plantă anuală, cu flori de culoare galbenă, care crește pe orice fel de teren bogat în umiditate. Partea aeriană a plantei și semințele recoltate înainte sau în timpul înfloririi se exploatează în scop fitoterapeutic. Mărarul este un condiment foarte aromat, originar din regiunile situate la est de Mediterană. Această mirodenie

este foarte folosită în Europa, dar și în India, unde frunzele uscate de mărar se folosesc atât la aromatizarea salatelor, fripturilor și sosurilor, cât și a lichiorurilor și dulcețurilor [1,2].



Figura I.1. Semințe de mărar (*Anethum graveolens*)

Sursa: <http://healthmeup.com/news-healthy-living/organic-seeds-healthy-ways-to-add-dill-seeds-to-your-diet/23844>

Un motiv pentru care cunoașterea originii geografice este importantă este faptul că atât calitatea nutrițională și gastronomică, cât și cea terapeutică pot suferi modificări semnificative în funcție de condițiile pedo-climatiche specifice locului de recoltare. Efectul acestor variabile asupra calității produselor poate fi decelat prin evaluarea caracteristicilor fizico-chimice ale probelor vegetale. Însă această corelație este foarte complexă [3,4]. Ca atare, pentru modelarea ei se impune o abordare multivariată, respectiv dezvoltarea unor aplicații chemometrice. Acestea prezintă și avantajul că, odată validate, pot constitui mijloace automate de identificare a originii probelor vegetale pe baza proprietăților lor fizico-chimice.

În vederea indentificării zonei de origine geografică a mărarului s-au aplicat metode de inteligență artificială, atât supravegheate cât și nesupravegheate, care au fost adaptate pentru recunoașterea automată a originii geografice a mărarului în funcție de o selecție a proprietăților sale fizico-chimice și spectrale. Rezultatele obținute în evaluarea sensibilității și selectivității metodelor aplicate recomandă utilizarea eficientă a acestor metode în vederea identificării provenienței geografice a probelor de mărar.

I.2. Ciuperci *Pleurotus* (*Pleurotus ostreatus*) și Champignon brune (*Agaricus bisporus*)

Ciupercile *Pleurotus* (*Pleurotus ostreatus*), cu un regim nutrițional lignicol-xilofag, fac parte din Ordinul *Agaricales*, la fel ca și ciuperca *Champignon* (*Agaricus bisporus*). Ciupercile cunoscute sub denumirea de *Champignon* apar în România preponderent prin cultura ciupercii *Agaricus Bisporus*. Aceasta este o ciupercă de cultură, de culoare albă sau crem, cu un regim termic crioofil, care necesită la fructificare temperaturi relativ scăzute, în domeniul 12°-18° C [5].

Cultivarea ciupercilor *Pleurotus* debutează la începutul secolului XX. În țara noastră primele culturi de *Pleurotus* au fost realizate începând cu anul 1973. *Pleurotus* este cunoscută ca o ciupercă prin excelență xilofagă, ce crește pe trunchiurile sau ciaturile putrezite ale arborilor. Bureții *Pleurotus* aparțin, din punct de vedere sistematic, aceleași clase și aceluiași ordin cu *Agaricus sp.*, însă fac parte din familii și genuri diferite. Deși diferă ca alcătuire exterioară, ele se aseamănă în ceea ce privește valoarea alimentară [6].



a)

b)

Figura I.2. Tipuri de ciuperci : a) Ciuperci *Champignon* b) Ciuperci *Pleurotus*

Sursa: <http://www.ciupercarie.ideideafaceri.ro/>

Atât afacerile la cheie cu ciuperci *Champignon*, cât și cele cu ciuperci *Pleurotus* sunt profitabile. Însă, în prezent, ciupercile *Champignon* sunt mai populare în rândul consumatorilor, bucurându-se de niveluri mai înalte de cerere față de ciupercile *Pleurotus*. De aceea, comercianții sunt uneori interesați de substituirea acestora. În vederea identificării și discriminării acestor două tipuri de ciuperci, au fost elaborate diverse metode multivariate pentru identificarea caracteristicilor lor [7].

I.3. Leuștean (*Levisticum officinale*)

Leușteanul (*Levisticum officinale*) este o plantă perenă care face parte din familia *Apiaceae*. Leușteanul este o plantă aromatică de origine mediteraneană, cunoscută și apreciată din antichitate, atât în scop culinar cât și medicinal. Este o plantă comestibilă utilizată adesea pentru a condimenta anumite supe sau mâncăruri. Din punct de vedere al aspectului, leușteanul este o plantă care poate atinge înălțimi de aproximativ 1,5 – 2 metri [8].



a)



b)

Figura I.3. Leuștean : a) Frunze de leuștean b) Flori de leuștean

Sursa: <http://en.wikipedia.org/wiki/Lovage>

Este cunoscut faptul că proprietățile care asigură vegetalelor calitățile menționate anterior depind de zona geografică de origine. De exemplu, condițiile pedo-climatiche favorabile potențază anumite proprietăți fizico-chimice și contribuie la îmbunătățirea proprietăților senzoriale. În consecință, dezvoltarea unor metode care să permită identificarea zonei de proveniență a diverselor tipuri de vegetale este importantă atât din punct de vedere gastronomic cât și nutrițional [9].

În vederea identificării și discriminării originii geografice a probelor de leuștean au fost elaborate diverse metode multivariate, bazate pe analiza proprietăților fizico-chimice ale probelor de leuștean prelevate din zonele geografice modelate.

I.4. Mazăre (*Pisum sativum*)

Mazărea are două motive pentru care să se considere vedetă: a fost planta pe care Gregor Mendel (fondatorul geneticii) a făcut, între anii 1856 și 1863, experiențele care au condus la elaborarea legilor eredității și a fost prima legumă ambalată în cutii de conserve. În țara noastră a fost adusă în secolul al XVII-lea [10].

Mazărea (*Pisum sativum*) este o plantă ierboasă, cultivată pentru comercializare datorită gustului său deosebit. Cultivarea mazării s-a extins foarte repede în multe țări, în unele părți ale lumii fiind considerată un aliment de bază, la fel ca orezul sau cartoful. Ca planta leguminoasă, cultivarea mazării este importantă și pentru refacerea structurii și fertilității solurilor [11].



a)

b)

Figura I.4. Mazăre : a) Flori de mazăre b) Boabe de mazăre

Sursa: <http://en.wikipedia.org/wiki/Pea>

Anumite proprietăți fizico-chimice, inclusiv cele care influențează proprietățile senzoriale, depind de condițiile pedo-climatice din zona de origine. Astfel, identificarea principalelor proprietăți fizico-chimice care influențează cel mai mult calitatea boabelor de mazăre are o importanță deosebită pentru caracterizarea acestui produs din punct de vedere nutrițional. Proprietățile care asigură vegetalelor calitățile nutritive depind de locul/solul de unde s-a făcut recoltarea. Altfel spus, proprietățile fizico-chimice ale mazării depind de proveniența acesteia, de locul de recoltare, climă și sol.

I.5. Cimbru (*Satureja hortensis*)

Cimbrul este un arbust peren cu viață lungă, care crește pe dealurile aride din preajma Mediteranei. Frunzele mici, verzi-cenușii, cu marginile curbate în interior, acoperă ramurile subțiri ivite din tulpinile lemnoase. Din luna mai până în august apar conuri de flori mici, albe sau rozalii. Gustul cald și înțepător și aroma bogată a frunzelor fac din cimbru un condiment popular pentru o mulțime de preparate culinare [12].

Părțile utile ale cimbrului sunt **frunzele și pețiolele florilor**, care se recoltează la începutul sezonului de înflorire și se usucă cu grijă, ferite de soare. Cimbrul uscat se folosește la prepararea de infuzii, pulberi pentru capsule și produse farmaceutice ca siropurile de tuse. Din ele se poate extrage și un ulei esențial, care intră în compoziția unor creme și unguente antiseptice. O plantă înrudită, cimbrul spaniol, este folosită și ea pentru calitățile antiseptice ale uleiului esențial [13].



Figura 1.5. Cimbrul : a) Flori de cimbru b) Cimbru proaspăt

Sursa: http://en.wikipedia.org/wiki/Summer_savory

Cimbrul (*Satureja hortensis*) este o plantă perenă din genul *Satureja*, familia labiatelor (*Lamiaceae*). Este o plantă scundă, care nu atinge mai mult de 20-30 centimetri în înălțime. Cimbrul preferă verile lungi, calde, secetoase, iar iarna supraviețuiește la temperaturi scăzute.

Cu toate acestea, randamentul cimbrului este influențat de condițiile pedologice și climatice conexe (clima, precipitații și temperaturi minime, medii, umiditate și înghețurile) din zona lor geografică de origine. Performanțele de creștere a acestor plante au fost înregistrate în diferite locuri și a fost determinat, în corelație, conținutul de ulei eteric. Analiza uleiurilor esențiale din *Satureja khuzestanica* obținută prin hidro-distilare, efectuată prin cromatografie lichidă de înaltă performanță (HPLC), indică faptul că există o relație pozitivă și liniară între înălțimea și conținutul carvacrol ca componentă majoră a cimbrului cu latitudine [14].

Astfel, identificarea principalelor proprietăți fizico-chimice care definesc calitatea cimbrului este foarte importantă atât pentru industria alimentară cât și farmaceutică. Astfel, sunt prezentate metode de inteligență artificială aplicate pentru identificarea celor mai importante variabile fizico-chimice care afectează calitatea probelor de cimbru, e.g. PCA. Distribuțiile scorurilor asociate primelor componente principale au fost analizate împreună cu distribuțiile ponderilor variabilelor, pentru a determina tipul de corelații între principalele proprietăți fizico-chimice care modelează sau discriminează probele în funcție de originea lor geografică. Rezultatele arată că acestea sunt umiditatea, conținutul de substanță uscată, de proteine și în mai mică măsură, de carbohidrați. Umiditatea probelor este puternic și negativ corelată cu ultimele trei variabile. Variabila principală care caracterizează similaritatea probelor de cimbru ce au aceeași origine geografică este conținutul de glutamat monosodic, compus asociat cu gustul delicios (umami).

Bibliografie

- [1] E. Teuscher, R. Anton, A. Lobstein, *Plantes aromatiques, epices, aromates, condiments et huiles essentielles*, Tec&Doc, Lavoisier, 2005.
- [2] J. Dyduch, *Koper ogrodowy (Anethum graveolens L.)*, Nasiennictwo, vol. 2, PWRiL, 2000.
- [3] J. Dyduch, *Koper ogrodowy (Anethum graveolens L.)*, Duczmal K.W., Halina Tucholska, PWRiL, 2000.
- [4] T.W. Bralewski, D. Szopińska, M. Morozowska, *Study for the evaluation of dill (Anethum graveolens L.) seeds*, Not. Bot. Hort. Agrobot. XXXIII, 2005.

- [5] S.T.Chang, P.G. Miles, *Mushrooms: Cultivation, nutritional value, medicinal effect, and environmental impact*, 2004, ISBN 0-8493-1043-1.
- [6] A.R. Bessette, A.E. Bessette, W. J. Neill, *Mushrooms of Cape Cod and the National Seashore*, Syracuse University Press, Syracuse, NY, pp.177, 2001.
- [7]<http://afacerilacheie.manager.ro/index.php?pag=articole&categorie=afaceriprofitabile&titlu=afaceri-la-cheie-privind-cultivarea-ciupercilor-champignon-sau-cultivarea-ciupercilor-pleurotus>
- [8] U. Stanescu, A. Miron, M. Hancianu, C. Aprotosoiaie *Plante medicinale de la A la Z* Monografii ale produselor de interes terapeutic, vol. II : J-Z, pag. 348-350, 2004.
- [9] Zs. Seregély, I. Novák, *Evaluation of the signal response of the electronic nose measured on oregano and lovage samples using different methods of multivariate analysis*, Acta Alimentaria, 2005.
- [10] I. Gherman, *Medicină alternativă tradițională, ghid practic*, Vestala, 2001.
- [11] F. Butle, *Relationship Marketing: Theory and Practice*, SAGE Publications Ltd., 1996.
- [12] P.M. Chisnall, *Marketing Research*, 6th ed. London: McGraw-Hill, 2001.
- [13] O. Pandia Olimpia, *Research Regarding the Effect of Fertilizers upon Maize Production and Quality*, Doctoral dissertation, 2006.
- [14] www.eukarya.ro, *Satureja hortensis*

CAPITOLUL II

DESCRIEREA METODELOR SPECTRALE ȘI MULTIVARIATE FOLOSITE PENTRU DETECȚIA ȘI IDENTIFICAREA ORIGINII PRODUSE VEGETALE STUDIAȚE

II.1. Spectrofotometria

Una dintre primele metode instrumentale apărute și utilizate frecvent în practica laboratoarelor de analize chimice moderne este metoda bazată pe absorbția luminii din domeniul vizibil (VIS). Se cunosc mai multe variante importante pentru această metodă: colorimetria, fotometria și spectrofotometria.

Colorimetria, una dintre tehnicile extrem de mult utilizate în practica analitică, reprezintă varianta în care intensitatea culorii probei se compară vizual sau instrumental, *în lumină albă*, cu un set de soluții etalon - preparate în condiții absolut identice cu proba. Aceasta este însă o metodă subiectivă și mai puțin selectivă, pentru că rezultatele depind mult de persoana care execută analiza. Se remarcă faptul că sensibilitatea maximă a ochiului omenesc atinge maximum pentru domeniul 550-560nm (domeniul culorii verzi), lucru important când compararea probei cu etalonul se face vizual. În această tehnică se pot realiza măsurători, prin comparație vizuală, chiar în eprubetă, la lumina zilei, rezultând analize chimice cu exactități mai slabe decât 1%. Cu cât există mai multe soluții etalon pentru comparație, cu atât metoda este mai exactă. Există și metode colorimetrice instrumentale, obiective. Dar acestea sunt tot mai puțin folosite, pentru ca există aparate ieftine care utilizează metode colorimetrice bazate pe reacții executate pe hârtie de filtru, pe substanțe aflate în stare adsorbită pe suporturi granulare - în cazul gazelor - și chiar pe reacții de culoare în soluții [1].

Fotometria și spectrofotometria măsoară instrumental lumina transmisă de o soluție *colorată* lucrând cu o sursă de lumină monocromatică. Când lumina incidentă este *filtrată*, prin filtre optice, având un spectru mai larg, realizăm o analiză de fotometrie. Când domeniul filtrat este mai îngust (utilizând monocromatoare), analizăm probele prin spectrofotometrie. În ultima

variantă, este posibilă fixarea mai precisă a lungimii de undă la care se lucrează. Cu ambele variante se poate chiar trasa un *spectru de absorbție*, adică o curbă obținută prin măsurarea semnalului în funcție de lungimea de undă a radiației incidente. În literatura de specialitate uneori se folosește pentru ambele metode și denumirea de *metodă colorimetrică* (sau chiar spectrocolorimetrică), ceea ce uneori poate crea confuzii [2].

În domeniul ultraviolet (UV), ochiul omenesc nu percepe radiația și ca atare se utilizează doar analiza prin *spectrofotometrie*. Întrucât principiile sunt identice, iar aparatele sunt în multe privințe similare în cele două domenii, în ultimul timp, în afară de aparatele dedicate domeniului VIS sau a celor pentru UV, se utilizează de multe ori un singur instrument pentru ambele intervale de lungimi de undă (spectrofotometre UV-VIZ). Construcția instrumentelor are în general două variante, respectiv *spectrofotometrele monocanal* (care folosesc cu un singur drum optic) și cele *comparative*, prevăzute cu două canale. În cazul spectrometrelor comparative, se compară proba etalon cu cea de analizat printr-o singură măsurătoare, utilizând două radiații care-și au originea în aceeași sursă (*coerente*) [3].

II.2. Analiza componentelor principale (PCA)

Analiza componentelor principale (PCA) este o procedură statistică care folosește o transformare ortogonală pentru a transforma un set de observații ale variabilelor posibil corelate într-un set de valori de variabile necorelate liniar numite componente principale (eng. *principal component*, PC). Numărul de PC este mai mic sau egal cu numărul de variabile originale. Această transformare este definită în așa fel încât prima componentă principală are cea mai mare posibilă variație (reprezintă mai mult de variabilitatea datelor în care este posibil). Fiecare componentă ulterioară are cea mai mare variație posibilă sub constrângerea că este ortogonală pe componenta precedentă (adică necorelată cu componentele precedente). PC sunt independente în cazul în care setul de date este distribuit normal. PCA este sensibilă la scalarea relativă a variabilelor originale [4,5].

În cazul în care în PCA se utilizează, ca input, o bază de date ce conține spectrele unor compuși cunoscuți, această metodă constă în găsirea acelor PC care modelează cel mai bine caracteristicile esențiale ale spectrelor compușilor din baza de date spectrale. Astfel, PCA extrage informații dintr-o cromatogramă sau dintr-un spectru pe două căi:

- a) numărul de componente semnificative este egal cu numărul de PC;
- b) fiecare PC este caracterizată de două informații: scorurile fiecărui spectru (eng. *scores*) și ponderea (eng. *loading*) fiecărei variabile (absorbții la o lungime de undă dată) [6].

II.3. Analiza ierarhica a clusterelor (HCA)

Analiza ierarhica a clusterelor (eng. Hierarchical Cluster Analysis, HCA) denumită și analiza de segmentare sau analiza taxonomică, creează grupuri sau cluster de date. Clusterelor sunt formate în așa fel încât obiectele din același cluster au proprietăți foarte asemănătoare, iar obiectele din grupe diferite au proprietăți distincte. HCA grupează probele prin crearea unui arbore de clasificare, denumit dendrograma. Acesta reprezintă o ierarhie cu nivele multiple, în care clusterelor de pe un nivel sunt unite cu cele din nivelul imediat următor în funcție de similaritatea acestora [7].

Gruparea ierarhică a datelor pe cluster are loc prin crearea unui arbore de clasificare denumit dendrograma. Arborele nu este un set unic de cluster, ci mai degrabă o ierarhie pe mai multe niveluri, în cazul în care grupurile de la un nivel sunt unite ca și cluster la nivelul

următor. Acest lucru permite să decidem nivelul de clusterizare cel mai potrivit pentru o aplicație data.

HCA se aplică atunci cand dorim să grupăm un set de obiecte în așa fel încât obiectele din același grup (denumit cluster) sunt mai asemănătoare (într-un sens sau altul) și mai deosebite de cele din celelalte grupuri (cluster). HCA este o metodă statistică de analiză a datelor prin explorare multivariată (eng. *data mining*). HCA este aplicată în multe domenii, ca mașină de învățare (eng. *learning machine*) sau ca metoda nesupravegheată de recunoaștere a formelor (eng. *pattern recognition*), pentru extragerea de informații [8].

II.4. Rețele neuronale artificiale (ANN)

Rețele neuronale artificiale (eng. *Artificial Neural Networks*, ANN) sunt modele de calcul inspirate din funcționarea sistemului nervos central (în special a creierului), care este capabil atât de învățare, cât și de folosirea cunoștințelor acumulate pentru rezolvarea unor situații noi (similare). ANN sunt prezentate în general ca sisteme interconectate de "neuroni" [9, 10].

De exemplu, o rețea neuronală dezvoltată pentru recunoașterea scrisului de mână este definit de un set de date de intrare cum ar fi pixelii unei imagini de intrare. După stabilirea ponderilor diverselor conexiuni, semnalul de intrare poate fi transformat de o funcție determinată de către proiectantul rețelei [80]. Semnalul este procesat mai departe prin activarea altor noduri care mimează neuronii. Acest proces se repetă până când, în final, este activat un anumit nod de ieșire din rețea. In acest exemplu, aceasta poate indica persoana care a scris textul. In mod analog, au fost dezvoltate multiple aplicații ANN pentru recunoașterea vocii [11].

Bibliografie

- [1] H. I. Nașcu, I. Jäntschi, *Chimie Analitică și Instrumentală*, Academic Pres & AcademicDirect, 2006
- [2] H. Nașcu, *Metode și Tehnici de Analiză Instrumentală*, U.T.PRES, 2003.
- [3] L. Roman, M. Bojiță, R. Săndulescu, *Validarea Metodelor de Analiză Instrumentală*, Medicală, 1998.
- [4] L. Smith , *Tutorial pentru Analiza Componentelor Principale*, 12-13, 2002.
- [5] I.T. Jolliffe, *Principal Component Analysis*, New York: Springer, 2002 , ISBN 0-387-95442-2.
- [6] I. Neamțu, P. G. Anoaică, *Aplicații ale radiațiilor electromagnetice în domeniul medical*, Medicală Universitară, 2006, ISBN 973-106-023-5
- [7] W.L. Martinez, A.R. Martinez, *Computational Statistics with MATLAB®*, Chapman & Hall/CRC Press, New York, 2002.
- [8] A.Genz, F. Bretz , *Comparison of Methods for the Computation of Multivariate Probabilities*, *Journal of Computational and Graphical Statistics*, vol. 11, no. 4, pp. 950–971, 2002.
- [9] H. Bhadeshia, *Neural Networks in Materials Science*, doi:10.2355/isijinternational.39.966, *ISIJ International* 39 (10): 966–979, 1999.
- [10] J. Hertz, R.G. Palmer, A.S. Krogh, *Introduction to the theory of neural computation*, Perseus Books, 1990, ISBN 0-201-51560-1
- [11] J.Lawrence, *Introduction to Neural Networks*, California Scientific Software Press, 1994, ISBN 1-883157-00-5

CAPITOLUL III

CONTRIBUȚII PROPRII OBȚINUTE ÎN DETECȚIA ȘI IDENTIFICAREA ORIGINII GEOGRAFICE A PRODUSELOR VEGETALE FOLOSITE ÎN INDUSTRIA ALIMENTARĂ PRIN ANALIZA SPECTRALĂ

Proprietățile nutritive ale vegetalelor depind de zona geografică de origine. Condiții pedo-climatice favorabile potențează și anumite proprietăți fizico-chimice care astfel contribuie la îmbunătățirea proprietăților senzoriale ale vegetalelor. În consecință, dezvoltarea unor metode care să permită identificarea zonei de proveniență a diverselor tipuri de vegetale, este importantă atât din punct de vedere gastronomic cât și nutrițional. În această lucrare prezentăm o aplicație de inteligență artificială dezvoltată pentru recunoașterea automată a originii geografice a leușteanului pe baza unei selecții de proprietăți fizico-chimice și spectrale.

Numărul mare de proprietăți fizico-chimice ce caracterizează comportarea complexă a produselor vegetale poate impune o abordare multivariată pentru a identifica originea acestora. Acest obiectiv poate fi atins prin elaborarea unei aplicații de inteligență artificială care să fie capabilă să detecteze eficient compușii pe baza proprietăților lor care au cea mai mare putere de modelare și / sau discriminare.

Dintre aceste aplicații, sistemele de management a bazelor de date pot servi numai la detecția probelor identice cu cele din baza de date de instruire. Pe de altă parte, aplicațiile de recunoaștere a formelor pot atribui identitatea de clasă (originea) unei probe necunoscute cu proprietăți *similare* (și nu numai identice) celor modelate.

III.1. Condiții experimentale

Analiza spectrală s-a realizat cu un spectrofotometru UV-VIZ Cintral. Extractele pentru analiza spectrală au ținut cont de solubilitatea componentelor responsabili pentru gustul umami (delicios) a produselor [1,2].

Extractele au fost obținute urmărind protocolul propus de Taylor, Hershey, Levine, Coy și Olivelle (1981):

5 g produs a fost supus extracției cu 25 mL apă deionizată.

Suspensia a fost adusă la fierbere 1 minut, răcită și centrifugată până la 11,800 g (3000 rot/min – 15 minute).

Extracția a fost repetată cu 20 mL apă deionizată.

Spectrele au fost înregistrate între 190,3387 și 1099,992 nm, cu un pas de digitizare de 0,426829 nm. S-a observat faptul că sub absorbțiile din domeniul 190 – 250 nm sunt afectate puternic de zgomotul spectral, probabil datorită stabilității mai reduse a emisiei sursei spectrometrului în această domeniu. Din acest motiv, aceste absorbții au fost eliminate din studiu, continuându-se analiza numai cu absorbțiile rămase. Eliminarea variabilelor spectrale care nu au putere de discriminare conduce la evitarea informației redundante și la o procesare multivariată a datelor mult mai rapidă.

III.2. Identificarea ferestrelor spectrale UV-VIS care au putere de modelare și de discriminare a originii geografice

III.2.1. Mărar (*Anethum Graveolens*)

Pentru a putea estima variația caracteristicilor fizico-chimice atât din punctul de vedere al tipului de vegetale, cât și a variației datorate influenței condițiilor pedo-climatice specifice diverselor zone geografice producătoare a acestor vegetale, au fost analizate și codificate câte

10 probe din următoarele categorii de mărar : mărar din județul Brăila (cod MBRx), județul Brașov (cod MBVx), județul Vrancea (cod MVNx) și din județul Cahul, Republica Moldova (cod MBAx), unde $x = 1, \dots, 10$ [3].

Pentru a decide dacă variațiile inter-categorie sunt relevante, au fost calculate spectrele medii pentru cele 4 tipuri de mărar. Figura III.1. arată că spectrele medii ale probelor de mărar din județele Brăila, Brașov, Vrancea și din județul Cahul, Republica Moldova sunt practic identice în domeniul 550-1100 nm. În această fereastră spectrală absorbțiile sunt deosebit de stabile, atât ca lungime de undă cât și ca intensitate a absorbției.

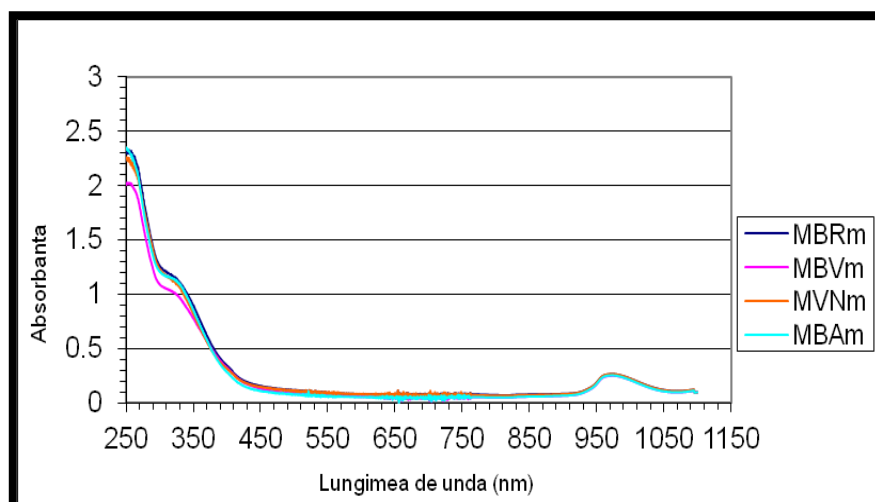


Figura III.1. - Spectrele UV-VIS medii ale probelor de mărar de diferite proveniențe geografice:
a) județul Brăila (cod spectru MBRm); b) județul Brașov (cod spectru MBVm); c) județul Vrancea (cod spectru MVNm); d) județul Cahul, Republica Moldova (cod spectru MBAm).

Pe de altă parte, domeniul spectral 250-550 nm are putere de modelare și de discriminare a diverselor origini geografice studiate pentru probele de mărar. În mod deosebit, mărarul din Brașov poate fi decelat de celelalte trei tipuri de mărar în special datorită absorbțiilor din domeniul 250-350 nm, prezentând absorbții semnificativ mai reduse decât mărarul din Brăila, Vrancea sau Cahul. În concluzie, domeniul 250-550 nm poate indica proveniența geografică a probelor de mărar pentru arii largi de proveniență, datorită faptului că variațiile inter-categorie sunt statistic semnificative. Restul spectrului UV-VIZ pare să poată fi neglijat.

III.2.2. Leuștean (*Levisticum officinale*)

Pentru a putea estima variația caracteristicilor fizico-chimice atât din punctul de vedere al tipului de vegetală, cât și a variației datorate influenței condițiilor pedo-climatice specifice diverselor zone geografice de origine, au fost analizate spectrele UV-VIS înregistrate pentru 10 probe de leuștean din județul Brașov (cod proba LBVx) și din județul Brăila (cod proba LBRx) – unde $x = 1, \dots, 10$.

Pentru a decide dacă variațiile inter-categorie sunt relevante, au fost calculate spectrele medii pentru probele de leuștean din Brăila și din Brașov (vezi Figura III.2.). Acestea confirmă faptul că în domeniul 550-1100 nm benzile de absorbție sunt deosebit de stabile, atât ca lungime de undă cât și ca intensitate. Acest domeniu spectral poate fi deci folosit pentru caracterizarea spectrală a tipului de vegetală în sine.

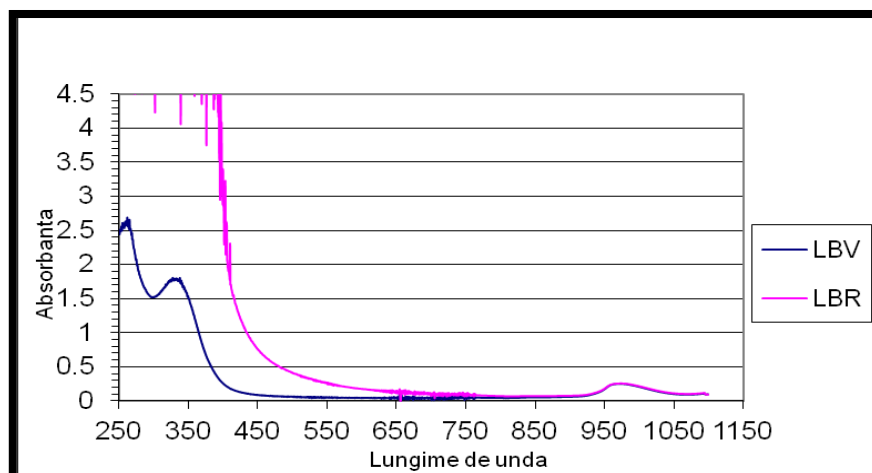


Figura. III.2. Spectrele medii UV-VIS ale probelor de leuștean de diferite proveniențe geografice: a) județul Brăila (cod spectru LBR); b) județul Brașov (cod spectru LBV)

Pe de altă parte, spectrele medii prezintă diferențe semnificative în domeniul spectral 250 - 550 nm, unde probele de leuștean de Brăila prezintă absorbții considerabil mai intense decât probele de leuștean de Brașov. În concluzie, absorbțiile din domeniul spectral 250 - 550 nm pot fi folosite pentru decelarea provenienței geografice, atât în cazul mărarului cât și a leușteanului, datorită variațiilor semnificative inter-categorie.

III.3. Analiza comparativă a puterii de discriminare a absorbțiilor UV-VIS în funcție de tipul vegetalelor

Rezultatele obținute în cazul mărarului și a leușteanului a generat următoarele întrebări: a) în ce măsură se poate generaliza folosirea absorbțiilor din domeniul 250 - 550 nm pentru discriminarea originii altor tipuri de vegetale; b) în ce măsură folosirea absorbțiilor din domeniul 250 - 550 nm este suficientă pentru discriminarea originii geografice unei game largi de tipuri de vegetale. În acest sens, s-a realizat analiza spectrală și pentru boabe de mazare, respectiv cimbru.

III.3.1. Mazăre (*Pisum sativum*)

Figura III.3. prezintă spectrele UV-VIZ înregistrate pentru 10 probe de boabe de mazăre (cod proba MDMx, x = 1, 2, ...10). În cazul acestei vegetale, absorbțiile sunt stabile, atât ca număr de undă cât și ca intensitate a absorbțiilor, într-un domeniu mai larg, respectiv în regiunea 450-1100 nm. Variații de intensitate apar în spectrele probelor de mazăre în domeniul 250-450 nm. Se observă că, în comparație cu spectrele UV-VIS ale probelor de mărar și de leuștean, spectrele boabelor de mazăre au o variație intra-categorie mult mai redusă.

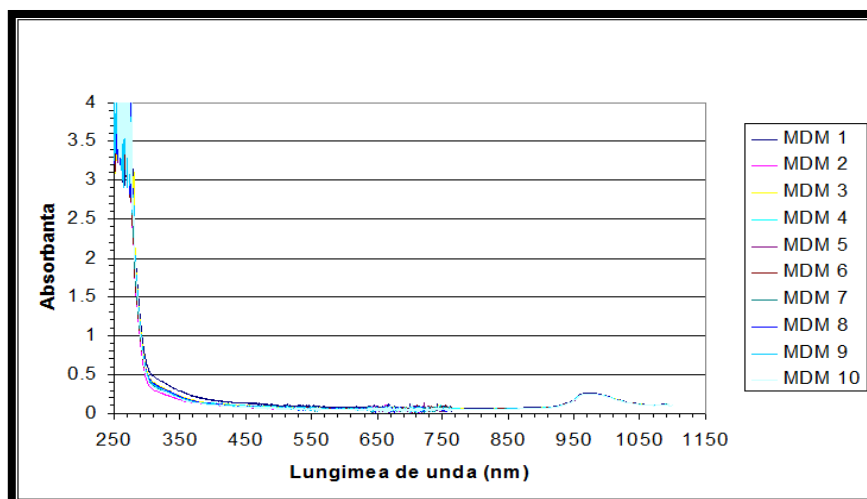


Figura III.3. Spectrele UV-VIS ale probelor de boabe de mazăre (cod proba MDMx, x =1,..., 10).

III.3.2. Cimbru (*Satureja Hortensis*)

Spectrele UV-VIS unui număr de 10 probe de cimbru originar din județul Galați (cod probe CGLx, x = 1, 2, ..., 10) sunt prezentate in figura III.4.

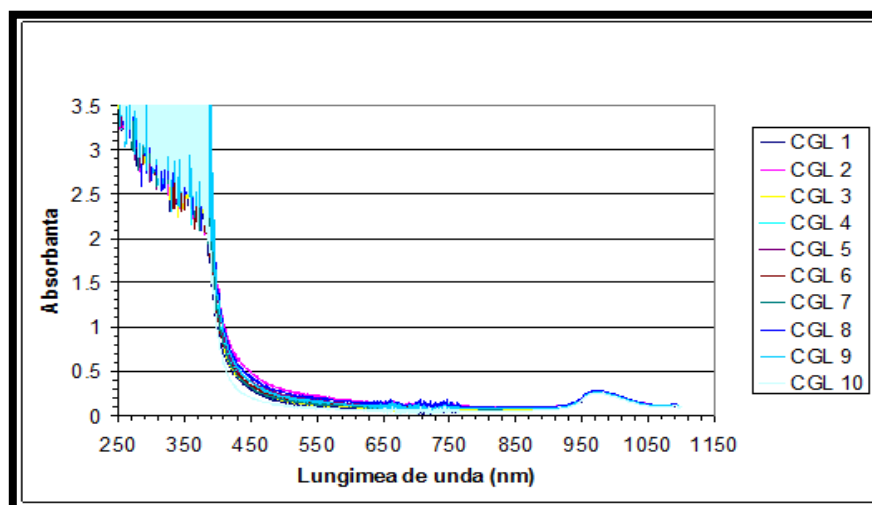


Figura. III.4. Spectrele UV-VIS ale probelor de cimbru (cod proba CGLx, x =1,..., 10).

Se observă că și in cazul cimbrului absorbțiile din regiunea 550-1100 nm sunt extrem de stabile, atât ca număr de undă cât și ca intensitate, spectrele diferitelor probe de cimbru fiind practic identice. Intensitatea benzilor de absorbție are în general o variație intra-categorie redusă, unele diferențe fiind înregistrate în regiunea 400-500 nm.

Ca și in cazul probelor de boabe de mazare, spectrele probelor de cimbru nu sunt foarte specifice in regiunea 250-550 nm. In această regiune nu apar benzi de absorbție bine definite, absorbțiile din spectrele probelor de cimbru deosebindu-se de cele ale boabelor de mazăre, de exemplu, doar prin faptul că intensitatea absorbțiilor crește exponențial începând cu lungimi de undă mai mari, respectiv de la 550 nm (in loc de 450 nm).

III.4. Discriminarea ciupercilor *Pleurotus* (*Pleurotus ostreatus*) și *Champignon* (*Agaricus bisporus*) prin analiză spectrală

Analiza spectrală a fost aplicată și pentru a verifica in ce măsură fereastra spectrală găsită ca adecvată pentru discriminarea originii geografice a unor vegetale este relevanta și pentru discriminarea unor vegetale ce aparțin aceluiași ordin. Studiul de caz a fost realizat pentru ciuperci *Pleurotus* (*Pleurotus ostreatus*) și *Champignon* brun (*Agaricus bisporus*). Bureții *Pleurotus* aparțin, din punct de vedere sistematic, aceleași clase și aceluiași ordin cu *Agaricus* sp., însă fac parte din familii și genuri diferite. Deși diferă ca alcătuire exterioară, ele se aseamănă in ceea ce privește valoarea alimentară [4,5].

In acest sens, au fost analizate spectrele UV-VIS inregistrate pentru 10 probe de ciuperci *Pleurotus* (cod CPLx) și 10 probe ciuperci *Champignon* brune (cod CBRx) – unde x =1,...,10.

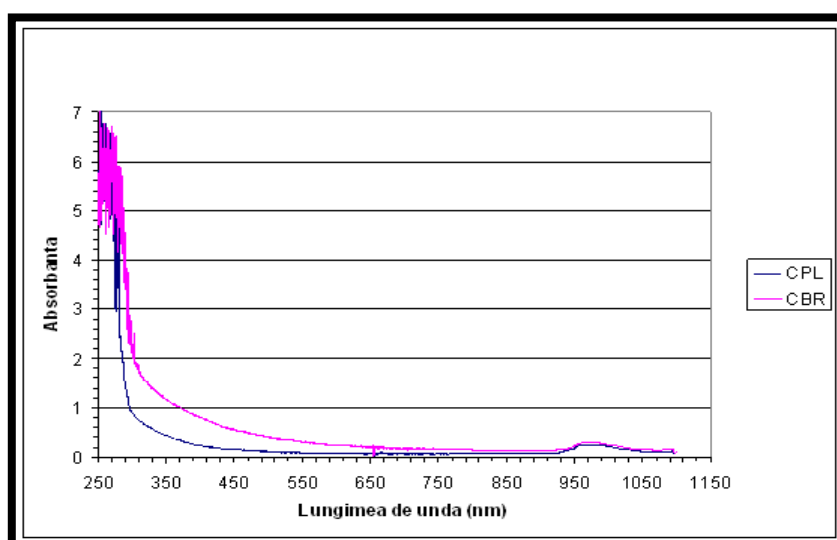


Figura III.5. Spectrele UV-VIS medii a ciupercilor *Pleurotus* (cod spectru CPL) și *Champignon* brune (cod spectru CBR)

Spectrele ciupercilor *Champignon* brune sunt discriminate de ciupercile *Pleurotus* prin absorbțiile din domeniul 250-550 nm, după cum se poate observă in Figura III.5. Variațiile intra-categorie se manifestă mai ales sub 550 nm, fiind mai importante in cazul ciupercilor *Champignon* brune decât in cazul *Pleurotus*. Pe de altă parte, banda de absorbție de la 960 nm este extrem de stabilă atât in cazul ciupercilor *Pleurotus*, cât și a celor *Champignon* brune.

III.5. Concluzii

Se poate trage deci concluzia că absorbțiile spectrale din fereastra spectrală UV-VIS cuprinsă intre 250-550 nm are putere de discriminare atât in cazul in care de dorește deosebirea probelor de vegetale din același ordin, cât și a celor din ordine diferite. Absorbțiile din acest domeniu spectral sunt adecvate inclusiv pentru decelarea originii geografice a aceleiași vegetală.

Totuși, nu toate vegetalele prezintă in acest domeniu benzi de absorbție bine definite. De asemenea, unele vegetale prezintă variații intra-categorie destul de insemnate in această fereastră spectrală. Ca atare, pentru identificarea automata eficientă a originii geografice a

vegetalelor se impune caracterizarea acestora printr-un număr mai mare de proprietăți fizico-chimice.

Pe de altă parte, așa cum s-a menționat anterior, spectrele UV-VIS au fost înregistrate cu un pas de digitizare de 0,426829 nm. Acest lucru înseamnă că și dacă baza de date folosită pentru identificarea originii geografice a vegetalelor ar fi formată numai cu absorbțiile din domeniul 250-550 nm, deja s-ar lucra cu peste 600 de variabile. În concluzie, dezvoltarea unui sistem automat de recunoaștere a originii geografice a diverselor vegetale trebuie să se bazeze pe metode chemometrice multivariate.

Bibliografie

- [1] P.J. Werbos, *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, 1975.
- [2] M. Devgun, *Research Journal of Pharmacy and Technology*, 3(3), 2010
- [3] G. Iordăchescu *Umami taste and consumer perception*, Annals of the Dunarea de Jos University of Galati, Fascicle VI – Food Technology, II (XXXI), 2008, 51-56.
- [4] I. Tudor, *Champignon – Tehnologia de cultură*, Lucman, 2010.
- [5] I. Tudor, *Pleurotus – Tehnologia de cultură*, Nomina Lex, 2010.

CAPITOLUL IV

CONTRIBUȚII PROPRII OBȚINUTE ÎN DETECȚIA ȘI IDENTIFICAREA ORIGINII GEOGRAFICE A PRODUSELOR VEGETALE FOLOSITE ÎN INDUSTRIA ALIMENTARĂ PRIN ANALIZA COMPONENTELOR PRINCIPALE (PCA)

Scopul analizei exploratorii realizată prin PCA este identificarea clusterelor formate (dacă e cazul) și a principalelor variabile care contribuie la modelarea diverselor tipuri de vegetale (variație intra-categorie) sau la discriminarea probelor din același tip de vegetală care însă provin din diverse zone geografice (variație inter-categorie).

IV.1. Condiții experimentale

Pentru a evalua posibilitatea de a identifica automat originea diverselor vegetale analizate, s-a construit câte o bază de date de instruire hibridă specifică fiecărei vegetală în parte. În fiecare bază de date au fost introduse atât absorbțiile UV-VIS din domeniul spectral identificat ca având cea mai mare putere de discriminare (absorbțiile spectrale din domeniul 250-550 nm), cât și valorile determinate pentru principalele caracteristici fizico-chimice ce definesc calitatea vegetalelor, respectiv: umiditatea (cod variabilă - MOIST), conținutul de substanță uscată (cod variabilă - DRYM), carbohidrați (cod variabilă - CARB), proteine (cod variabilă - PROT), grăsimi (cod variabilă - FAT), cenușă (cod variabilă - ASH) și de glutamat monosodic (cod variabilă - MSG). Astfel, baza de date de instruire dezvoltată pentru fiecare tip de vegetală include un număr de 711 variabile.

Sistemele PCA au fost dezvoltate cu ajutorul aplicației software *The Unscrambler* [1].

IV.2. Identificarea originii geografice a produselor vegetale folosite în industria alimentară prin PCA

IV.2.1. Mărar (Anethum Graveolens)

Pentru a putea estima variația caracteristicilor fizico-chimice atât din punctul de vedere al tipului de vegetale, cât și a variației datorate influenței condițiilor pedo-climatice specifice diverselor zone geografice producătoare, au fost determinate proprietățile fizico-chimice menționate pentru câte 10 probe de mărar din județul Brăila (cod MBRx), județul Brașov (cod MBVx), județul Vrancea (cod MVNx) și din județul Cahul, Republica Moldova (cod MBAx), unde $x = 1, \dots, 10$. Aceste valori au fost incluse, împreună cu absorbțiile UV-VIS din domeniul 250-550 nm, într-o bază de date hibridă care a fost apoi folosită pentru instruirea PCA. Altfel spus, în cazul probelor de mărar, baza de date de intrare folosită pentru PCA formează o matrice cu 711 x 40 intrări.

Cele mai bune rezultate în discriminarea probelor de mărar în funcție de originea lor geografică a fost obținută cu ajutorul primelor două PC, care sunt caracterizate de o variantă explicată cumulată de 82%. Distribuția scorurilor probelor în reprezentarea PC1 și PC2 indică formarea a 2 cluster (vezi Figura IV.1.). Un cluster este format de probele de mărar din județul Brașov (cod proba MBVx), caracterizat de scoruri PC1 și PC2 pozitive, de valoare medie. Al doilea cluster este asociat probelor de mărar din județul Brăila (cod probe MBRx), Vrancea (cod probe MVNx) și din Cahul (cod probe MBAx), care sunt caracterizate de scoruri PC1 negative și PC2 pozitive.

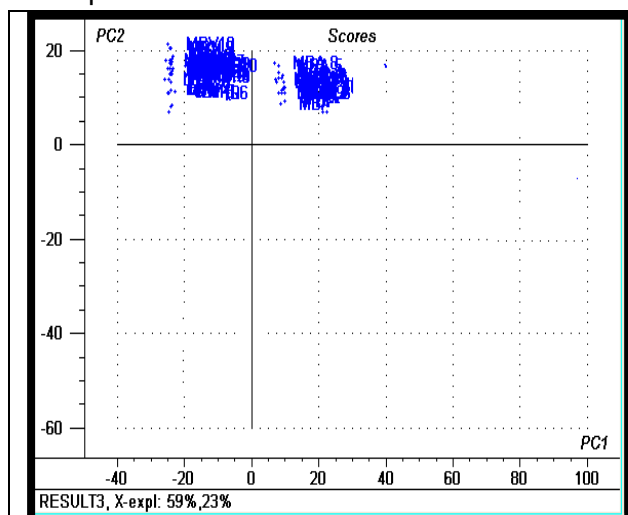


Figura IV.1. – Distribuția scorurilor probelor în reprezentarea PC2 vs. PC1 în cazul probelor de mărar din județul Brașov, Brăila, Vrancea și Cahul (Republica Moldova).

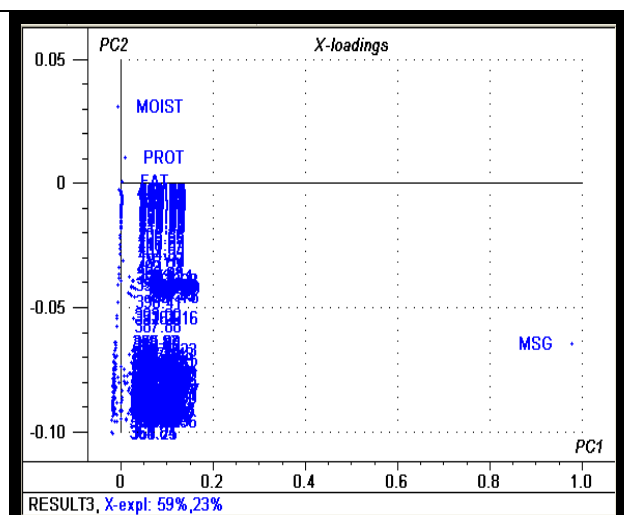


Figura IV.2. – Distribuția ponderilor variabilelor în reprezentarea PC2 vs. PC1 în cazul probelor de mărar din județul Brașov, Brăila, Vrancea și Cahul (Republica Moldova).

Principalele variabile care conduc la formarea acestor cluster au fost identificate analizând distribuția ponderilor variabilelor în reprezentarea PC2 vs. PC1. Figura IV.2. arată că variabilele umiditate (MOIST), conținutul de proteine (PROT) și de grăsime (FAT) sunt responsabile pentru generarea scorurilor PC2 pozitive. Absorbțiile spectrale (mai ales cele din jurul lungimii de undă 360 nm) contribuie la scăderea scorurilor PC2.

În concluzie, cum ambele cluster sunt caracterizate de scoruri PC2 pozitive, aceste variabile au putere de modelare, dar nu au putere de discriminare. Principala variabilă cu putere de discriminare a celor două cluster este conținutul de glutamat monosodic (MSG), care contribuie la discriminarea celor două cluster în sensul generării de scoruri PC1 pozitive.

IV.2.2. Leuștean (*Levisticum Officinale*)

Analiza exploratorie realizată prin PCA permite identificarea clusterelor formate și a principalelor variabile care contribuie la caracterizarea diverselor tipuri de leuștean în funcție de zona geografică de origine [1]. Analiza distribuției scorurilor probelor de leuștean (vezi figura IV.4.) în reprezentarea PC2 vs. PC1 indică formarea a două cluster foarte bine definite, asociate probelor în funcție de originea lor geografică. Astfel, probele de leuștean de Brașov (cod probe LBVx, x = 1, 2, ..., 10.) formează un cluster situat în cadranul II, punctele asociate fiind caracterizate de scoruri PC1 negative și PC2 pozitive.

Pe de altă parte, probele din leuștean de Brăila (cod probe LBRx, x = 1, 2, ..., 10) generează un cluster situat în cadranul III, caracterizat de scoruri PC1 și PC2 negative. Discriminarea se datorează mai ales PC2, care explică numai 23% din varianta explicată a datelor. Spre deosebire de această componentă principală, PC1 nu discriminează semnificativ cele două cluster: scorurile tuturor probelor de leuștean au valori apropiate ale scorurilor PC1, indiferent de originea lor, deși PC1 explică 59% din varianta datelor.

Distribuția ponderilor variabilelor în reprezentarea PC2 vs. PC1 (vezi figura IV.3. și figura IV.4) indică faptul că umiditatea (MOIST), conținutul de proteine (PROT) și grăsimi (FAT) sunt în mod deosebit responsabile pentru formarea clusterului caracterizat de scoruri PC2 pozitive, adică a celui specific probelor de leuștean din județul Brașov.

În distribuția scorurilor probelor în reprezentarea PC3 vs. PC1 (vezi figura IV.5. și figura IV.6) punctele asociate probelor de leuștean se contopesc practic într-un singur nor, i.e. un cluster caracterizat de o dispersie relativ mare. Unele probe de leuștean de Brăila (LBRx) sunt caracterizate de scoruri PC1 și PC3 negative, punctele lor asociate fiind plasate în cadranul III. Punctele asociate restului de probe de leuștean de Brăila formează un nor împreună cu punctele asociate probelor de leuștean de Brașov (LBVx), fiind caracterizate de scoruri PC1 negative și PC3 pozitive.

Intr-adevar, distribuția scorurilor probelor în reprezentarea PC4 vs. PC1 (vezi figura IV.7) pune în evidență în mod deosebit variabilitatea crescută a proprietăților probelor de leuștean din Brăila. Reprezentarea ponderilor variabilelor în reprezentarea PC4 vs. PC1 (vezi figura IV.8) confirmă că acesta se datorează mai ales diferențelor manifestate în comportarea spectrală. Pondere PC4 foarte mică asociată conținutului de glutamat monosodic (MSG) indică faptul că diferențele dintre probele de leuștean evidențiate de PC4 sunt independente de această variabilă. Altfel spus, diferențele evidențiate de PC4 în proprietățile probelor nu afectează calitățile probelor din punct de vedere al gustului delicios (umami).

Aceleași informații sunt furnizate și de distribuția scorurilor probelor și a ponderilor variabilelor în reprezentarea PC5 vs. PC1 (vezi figura IV.9 și figura IV.10.).

În concluzie, s-a evaluat prin PCA fezabilitatea identificării originii probelor de leuștean pe baza unei selecții de proprietăți fizico-chimice și a comportării spectrale a acestora.

PCA a pus în evidență faptul că cea mai bună discriminare a probelor de leuștean poate fi realizată prin analiza distribuției scorurilor probelor în reprezentarea PC2 vs. PC1. În această reprezentare, probele de leuștean din județul Brăila și cele din județul Brașov formează două clustere foarte bine definite. Variabilele cu cea mai mare putere de modelare și/sau de discriminare sunt umiditatea (MOIST), conținutul de proteine (PROT) și grăsimi (FAT), respectiv absorbțiile spectrale din jurul lungimii de undă de 360 nm.

IV.3. Identificarea principalelor caracteristici fizico-chimice care influențează calitatea produselor vegetale folosite în industria alimentară prin PCA

IV.3.1. Mazăre (*Pisum sativum*)

Pentru aplicarea PCA, s-a construit o bază de date de instruire hibridă, în care au fost introduse atât spectrele UV-VIZ (absorbțiile spectrale din domeniul 250-550 nm) ale probelor de boabe de mazăre (cod proba MDMx, x = 1, 2, ..., 10) cât și valorile determinate pentru umiditate (cod variabilă - MOIST), substanță uscată (cod variabilă - DRYM), carbohidrați (cod variabilă - CARB), proteine (cod variabilă - PROT), grăsimi (cod variabilă - FAT), cenușă (cod variabilă - ASH), glutamat monosodic (cod variabilă - MSG) [1,2].

PCA a fost realizată cu primele trei PC, următoarele PC explicând variante de sub 2%. Analizând varianta reziduală X în funcție de influența probelor (eng. *leverage*) asupra modelului PCA (vezi figura IV.11), se poate observa că în setul de date nu există probe care să constituie probe cu o comportare semnificativ diferită (eng. *out-lier*), toate probele de mazăre fiind caracterizate de valori mici ale variantei reziduale X și a influenței [3].

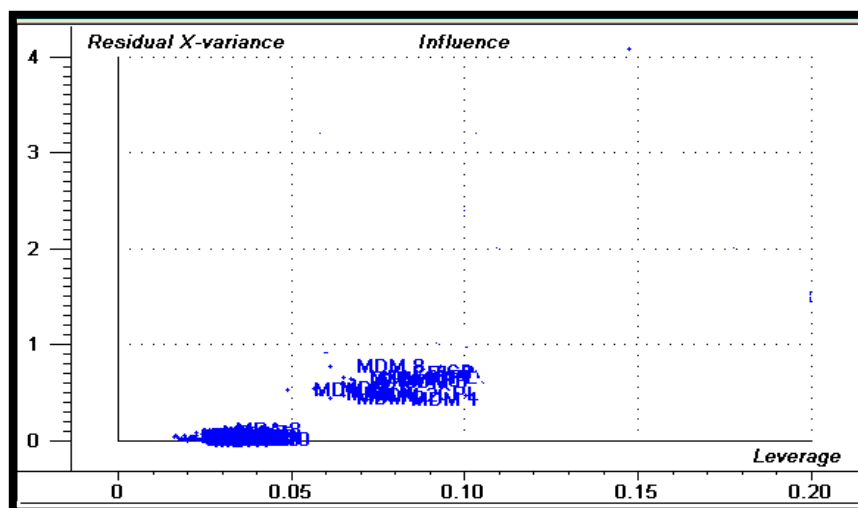


Figura IV.11. Analiza variantei reziduale X și a influenței caracteristicilor probelor de mazăre asupra modelării PCA (cod proba MDMx, x = 1, 2, ..., 10)

Boabele de mazăre cu un grad mai ridicat de umiditate sunt caracterizate de un conținut mai scăzut de substanță uscată (DRYM), proteine (PROT) și carbohidrați (CARB). Pe de altă parte, conținutul de glutamat monosodic (MSG) este practic independent de aceste proprietăți fizico-chimice (MOIST, DRYM, PROT și CARB).

IV.3.2. Cimbru (*Satureja Hortensis*)

Principalele proprietăți fizico-chimice care generează diferențe în calitatea probelor de cimbru au fost identificate în mod obiectiv prin PCA [1,4]. Distribuția scorurilor probelor au fost analizate împreună cu cele ale ponderilor variabilelor pentru a estima dimensiunea variațiilor intra-clasa (regiune de cultivare) și a determina tipul de corelații între principalele proprietăți fizico-chimice ale probelor de cimbru (cod proba CGLx, x = 1, 2, ..., 10) cu aceeași origine geografică (județul Galați).

În acest sens, s-a construit o bază de date hibridă, prin concatenarea absorbțiilor spectrale din regiunea 250-550 nm din spectrele UV-VIS ale probelor de cimbru, umiditatea acestora (cod variabilă - MOIST), conținutul de substanță uscată (cod variabilă - DRYM), glucide (cod variabilă - CARB), proteine (cod variabilă - PROT), grăsime (cod variabilă - FAT), cenușă (cod variabilă - ASH) și de glutamat monosodic (cod variabilă - MSG) [3].

Analiza variantei reziduală X în funcție de influența probelor (eng. *leverage*) asupra modelului PCA (vezi figura IV.12) confirmă faptul că nici o probă de cimbru nu este un out-liear [82]. Totuși, se observă că probele de cimbru sunt caracterizate de o variație a valorilor variantei reziduale X și ale influenței probelor asupra modelului PCA mult mai mare decât în cazul probelor de boabe de mazăre. Astfel, pentru unele probe de cimbru, varianta reziduală X este de până la de 3 ori mai mare decât cea a probelor de boabe de mazăre. Același lucru este valabil și pentru influența probelor asupra modelului PCA obținut, care variază între 0.1 și 0.18, unele probe de cimbru având astfel o influență de aproape 2 ori mai mare asupra modelului decât în cazul probelor de boabe de mazăre [5].

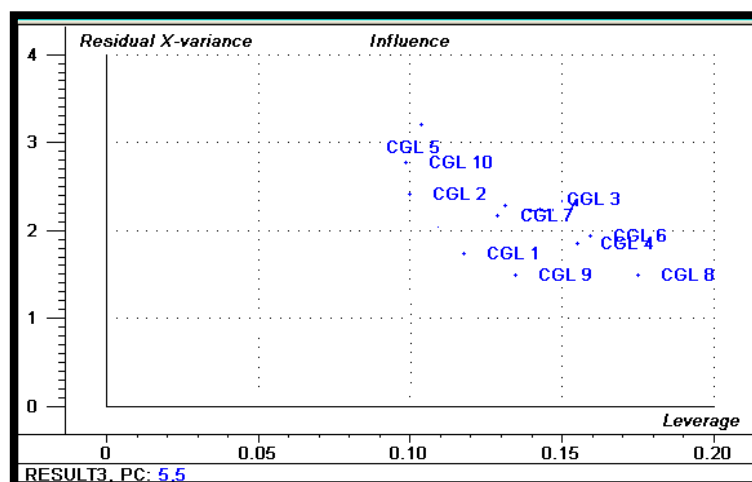


Figura IV.12. Analiza variantei reziduale X și a influenței caracteristicilor probelor de cimbru asupra modelării PCA (cod proba CGLx, x = 1, 2, ..., 10).

PCA a indicat că baza de date de instruire poate fi comprimată într-un nou hiperspațiu, definit de numai 4 variabile latente independente, primele patru PC cumulând o variație explicată de 89%. Distribuțiile ponderilor variabilelor indică și faptul că dispersia relativ mai mare a valorii scorurilor asociate probelor de cimbru se datorează în principal variațiilor înregistrate în intensitatea benzilor de absorbție prezente în spectrele probelor de cimbru în regiunea 400-500 nm.

IV.4. Discriminarea ciupercilor *Pleurotus (Pleurotus ostreatus)* și Champignon (*Agaricus bisporus*) prin PCA

Elaborarea unui sistem expert dedicat identificării tipului de ciuperci *Pleurotus (Pleurotus ostreatus)* și Champignon (*Agaricus bisporus*), prezintă o importanță deosebită pentru depistarea fraudelor din comerț [4]. Folosirea unui astfel de sistem expert este o modalitate de detecție eficientă și accesibilă (ieftină și ușor de utilizat) pentru screening-ul produselor vegetale.

Analiza exploratorie are rolul identificării acelor variabile care contribuie in mod semnificativ la modelarea similarității și la discriminarea inter-categorie (tipul de ciuperca) sau intra-categorie (aceiași tip de ciupercă). Analiza are și rolul identificării variabilelor care nu au putere de discriminare. Eliminarea acestora din baza de date supusă calculului chemometric conduce la evitarea informației redundante și deci o procesare multivariată a datelor mult mai rapidă [6].

Pentru a evalua posibilitatea de a identifica automat natura ciupercilor analizate, s-a construit o bază de date de instruire hibridă, in care au fost introduse atât spectrele UV-VIZ (absorbțiile spectrale din domeniul 250-550 nm) ale probelor cât și valorile determinate pentru umiditate (cod variabila - MOIST), conținut de substanță uscată (cod variabila - DRYM), carbohidrați (cod variabila - CARB), proteine (cod variabila - PROT), grăsimi (cod variabila - FAT), cenușă (cod variabila - ASH) și de glutamat monosodic (cod variabila - MSG). Altfel spus, baza de date de instruire formează o matrice cu 711 x 20 intrări.

Pentru determinarea numărului de PC cu ajutorul cărora este util să se efectueze PCA, s-a analizat dinamica variantei explicate cumulate. Figura IV.13 arată că aceasta crește până când atinge un palier (o valoare practic constantă). Într-adevăr, după cum se observă în tabelul VI.2., varianta explicată cumulată crește rapid datorită primelor două componente principale (PC), fiecare PC ulterior contribuind cu un procent din ce în ce mai mic [7].

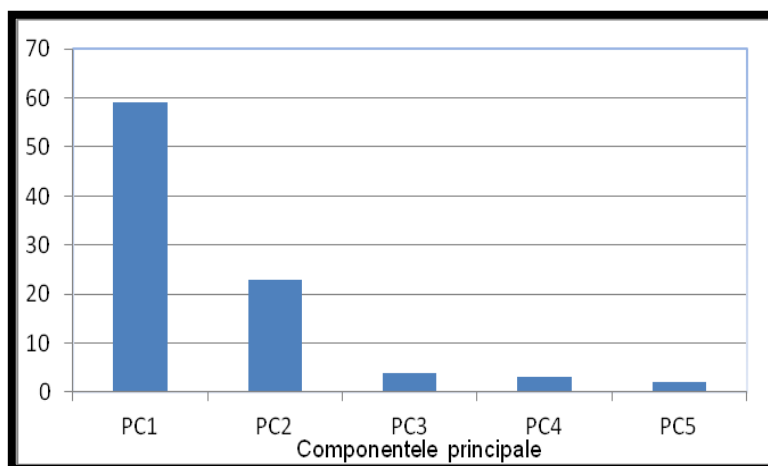


Figura IV.13. Varianta explicată a componentelor principale in cazul PCA dezvoltată pentru ciuperci *Pleurotus (Pleurotus ostreatus)* și Champignon (*Agaricus bisporus*)

In concluzie, PCA indică faptul că hiperspațiul poate fi comprimat la un spațiu definit de cel mult 5 PC. Dinamica variantei explicate cumulate este prezentată in Tabelul IV.2.

Tabelul IV.1. Dinamica variantei explicate absolute și a celei cumulate pentru primele PC in cazul PCA dezvoltata pentru ciuperci *Pleurotus* (*Pleurotus ostreatus*) și *Champignon* (*Agaricus bisporus*)

Componenta principală	PC1	PC2	PC3	PC4	PC5
Varianta explicată (%)	59	23	4	3	2
Varianta explicată cumulată (%)	59	82	86	89	91

In concluzie, rezultatele arată că PCA este o metodă adecvată pentru identificarea și discriminarea eficientă a ciupercilor *Pleurotus* (*Pleurotus ostreatus*) de ciupercile *Champignon* brun (*Agaricus bisporus*). Analiza spectrelor UV-VIZ au dat o primă indicație asupra faptului că ciupercile *Champignon* brune pot fi discriminate de ciupercile *Pleurotus* prin absorbțiile din domeniul 250-450 nm. Acestea au fost incluse intr-o bază de date hibridă, care concatenează aceste proprietăți spectrale și proprietăți fizico-chimice precum umiditatea, cenușa, conținutul de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic.

IV.5. Concluzii

PCA permite identificarea originii geografice a probelor de mărar, atribuind acestora in mod automat identitatea de clasă, pe baza unei selecții de proprietăți fizico-chimice și a comportării spectrale a acestora.

In cazul sistemului dezvoltat pentru identificarea originii geografice a probelor de leuștean, rezultatele confirmă utilitatea folosirii bazelor de date hibride și indică faptul că originea geografică a unei probe de origine necunoscută poate fi recunoscută in funcție de scorurile sale in reprezentarea PC1 și PC2. Altfel spus, metoda chemometrică propusă permite automatizarea identificării originii unei probe de leuștean pe baza proprietăților spectrale și fizico-chimice menționate.

Studiul realizat a indicat faptul că in PCA este o metodă chemometrică ce poate fi aplicată si pentru identificarea principalelor caracteristici fizico-chimice care pot afecta calitatea vegetalelor cu aceeași origine geografică. In cazul in care se aplică PCA pentru identificarea originii geografice a unui produs vegetal, discriminarea se bazează pe variația inter-clasă, care este mai mare decât cea intra-clasa.

PCA a indicat faptul că similaritatea boabelor de mazăre este dată in principal de conținutul de glutamat monosodic (MSG), compus care asigură gustul delicios (UMAMI) al unui produs vegetal. Conținutul de glutamat monosodic (MSG) este aproape constant pentru boabele de mazăre având aceeași origine geografică, fiind practic independent de variația celorlalte caracteristici fizico-chimice studiate prin această metodă.

In cazul cimbrului, analiza distribuției scorurilor probelor și a ponderilor variabilelor, in reprezentările obținute pentru primele trei PC, a indicat că principalele proprietăți fizico-chimice care influențează în mod semnificativ calitatea cimbrului sunt umiditatea, conținutul de substanță uscată, conținutul de proteine și în mai mică măsură, conținutul de carbohidrați. Umiditatea este puternic și negativ corelată cu ultimele trei variabile.

Variabila principală care asigură similitudinea dintre probele de cimbru având aceeași origine geografică este conținutul de glutamat monosodic, care generează gustul delicios

(umami). Conținutul monoglutamat de sodiu (MSG) este practic independent de restul variabilelor luate in studiu.

In cazul identificării și discriminării ciupercilor *Pleurotus* (*Pleurotus ostreatus*) de ciupercile Champignon brun (*Agaricus bisporus*), PCA a condus la comprimarea spațiului 711-dimensional la 3 dimensiuni, date de primele trei PC, ceea ce permite atribuirea identității de clasa cu resurse de calcul mult reduce.

Bibliografie

- [1] G. Saporta, N. Niang N. *Principal component analysis: application to statistical process control*. In: Govaert G, Data Analysis, John Wiley & Sons, 2009.
- [2] T.L. Pownall, C.C.Udenigwe, R.E. Aluko, *Amino acid composition and antioxidant properties of pea seed (*Pisum sativum* L.) enzymatic protein hydrolysate fractions*, doi:10.1021/jf904456r. PMID 20359226, Journal of Agricultural and Food Chemistry 58 (8): 4712–4718, 2010.
- [3] S. C. Ghinita, M. Praisler, G. Iordăchescu, *PCA identification of the main physico-chemical properties influencing the quality of pea beans (*Pisum sativum*)*, Abstract book of the Second International Conference on Natural and Anthropic Risks, 2014, pp.122, ISSN 2360-4018 <http://icnar.ub.ro/>
- [4] P. M.Chisnall , Marketing Research, 6th ed. London: McGraw-Hill, 2001.
- [5] S.C. Ghinita, M. Praisler, G. Iordăchescu, „*Identification of the Main Physico-Chemical Properties Influencing the Quality of Thyme (*Satureja Hortensis*)*”, Journal of Engineering Studies and Research, vol.20, no. 2, 2014, 33-38. <http://pubs.ub.ro/?pg=revues&rev=jesr>
- [6] L. Wade. *Identifying Mushrooms*, Peninsula Mycological Circle. Retrieved, 2012.
- [7] S. Constantin, M. Praisler, G. Iordăchescu, *Chemometrical method designed for the automatic identification of mushrooms*, Annals of "Dunarea de Jos" University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II, Year V (XXXVI) 2013, Nos. 1-2 , p. 32 – 39, ISSN 2067-2071, (revista cotata B+, indexata in: Cambridge Scientific Abstracts (CSA), EBSCO).

CAPITOLUL V

CONTRIBUȚII PROPRII OBȚINUTE ÎN DETECȚIA ȘI IDENTIFICAREA ORIGINII PRODUSELOR VEGETALE FOLOSITE ÎN INDUSTRIA ALIMENTARĂ PRIN ANALIZA IERARHICĂ A CLUSTERELOR (HCA)

Cunoașterea originii geografice a produselor vegetale este importantă deoarece calitatea lor nutrițională, gastronomică și/sau terapeutică poate suferi modificări semnificative în funcție de condițiile pedo-climatice specifice locului de recoltare. Efectul acestor variabile asupra calității produselor poate fi decelat prin evaluarea caracteristicilor fizico-chimice ale probelor vegetale. Inșă această corelație este foarte complexă.

Ca atare, pentru modelarea ei se impune o abordare multivariată, respectiv dezvoltarea unor aplicații chemometrice. Acestea prezintă și avantajul că, odată validate, pot constitui mijloace automate de identificare a originii probelor vegetale pe baza proprietăților lor fizico-chimice [1]. In acest capitol sunt prezentate o serie de sisteme automate care au fost dezvoltate pentru recunoașterea originii geografice a vegetalelor printr-o metodă multivariată ierarhica și nesupravegheată, respectiv Analiza Ierarhică a Clusterelor (eng. *Hierarchical Cluster Analysis*,

HCA). Metoda aplicată în această lucrare pentru modelarea similarității este cea de clustering aglomerativ.

V.1. Condiții experimentale

Aplicațiile HCA prezentate în acest capitol au fost dezvoltate cu ajutorul programului *Matlab 2013a*. Pentru modelarea similarității și disimilarității dintre fiecare pereche de două probe din setul de date, distanța multivariată dintre două probe a fost calculată pe baza distanței Euclidiene. Algoritmul de calcul folosit pentru determinarea arborilor de clasificare, denumiți dendrograme, este cel de clustering aglomerativ.

Pentru gruparea probelor în arborele de clasificare, perechile de probe care sunt unite în funcție de similaritatea lor, adică sunt cuplate acele probe pentru care distanța Euclidiană este minimă. Pentru realizarea aplicației descrisă prin această metodă, cuplarea a fost realizată folosind funcția de unire medie (eng. *average linkage function*) [2].

Această funcție folosește informația generată la primul pas de unire pentru a determina proximitatea fiecărei probe față de celelalte luate în calcul. Se formează astfel perechi de probe ce formează clustere binare [3]. Clusterelor nou formate sunt apoi grupate în clustere mai mari, până când se formează întregul arbore de clasificare. Acesta este reprezentat sub forma unei dendrograme, care reprezintă pe abscisa codul probelor și pe ordonată disimilaritatea (distanța Euclidiană) dintre probe [4].

V.2. Identificarea originii geografice a produselor vegetale folosite în industria alimentară prin HCA

V.2.1. Mărar (*Anethum Graveolens*)

Cunoașterea zonei geografice este importantă datorită faptului că proprietățile fizico-chimice ale mărarului diferă în funcție de condițiile pedo-climatice în care se dezvoltă. Prin această metodă prezentăm o aplicație multivariată de recunoaștere a originii geografice a probelor de mărar (*Anethum graveolens*), respectiv analiza hierarhică a clusterelor [1].

V.2.1.1. Testarea sensibilității și selectivității sistemului de identificare

În această secțiune prezentăm o aplicație de inteligență artificială dezvoltată pentru identificarea automată a originii geografice a probelor de mărar cu ajutorul HCA. Pentru modelarea influenței condițiilor pedo-climatice specifice diverselor zone geografice producătoare asupra caracteristicilor fizico-chimice a probelor de mărar, au fost analizate câte 10 probe provenind din județul Brăila și din Cahul (Republica Moldova). Un număr de 9 probe din fiecare grup de origine a fost folosit pentru instruirea sistemului de recunoaștere a originii geografice pe baza proprietăților lor fizico-chimice, restul fiind folosite pentru testarea aplicației din punctul de vedere al sensibilității. O probă de mărar provenind din județul Vrancea și alta recoltată din județul Brașov, au fost de asemenea folosite pentru testarea selectivității aplicației.

Figura V.1. prezintă rezultatele obținute pentru discriminarea acestor probe de mărar, cele recoltate în județul Brăila având codul de probe 1-9, iar cele originare din Cahul având codurile 10-18. Se observă că discriminarea celor două clase de mărar este foarte eficientă, având în vedere că arborele de clasificare este format din două ramuri foarte bine delimitate, fiecare ramură fiind asociată uneia din zonele geografice considerate. Ramura din dreapta este

formată din probele de mărar din județul Brăila (probele 1-9), iar cea din stânga este asociată probelor de mărar din Cahul (probele 10-18).

Distanța Euclidiană dintre cele două clusteruri asociate acestor ramuri, reprezentată pe ordonata dendrogramei, este o măsură a disimilarității celor două clase de probe de mărar. Se observă că aceasta este mult mai mare decât distanța dintre oricare pereche de probe de mărar aparținând uneia dintre ramuri. Mai precis, figura V.1. arată ca distanța dintre ramurile asociate probelor de mărar originare din județul Brăila și cea specifică probelor de mărar recoltate din Cahul este de aproape 12 ori mai mare decât distanța maximă dintre probele cu aceeași proveniență geografică [5].

Altfel spus, variația multivariată inter-clasă este mult mai mare decât cea intra-clasă. În consecință, în dendrograma obținută nu se înregistrează nici o probă clasificată incorect. Putem trage deci concluzia că metoda aleasă (HCA) este foarte adecvată discriminării probelor de mărar din județul Brăila de cele din Cahul.

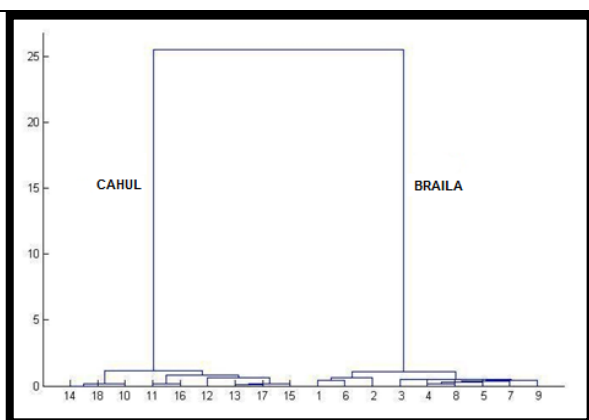


Figura V.1. Dendrograma determinată pentru discriminarea probelor de mărar construită cu câte 9 probe provenind din județul Brăila (cod proba 1-9) de cele recoltate din Cahul (cod proba 10-18).

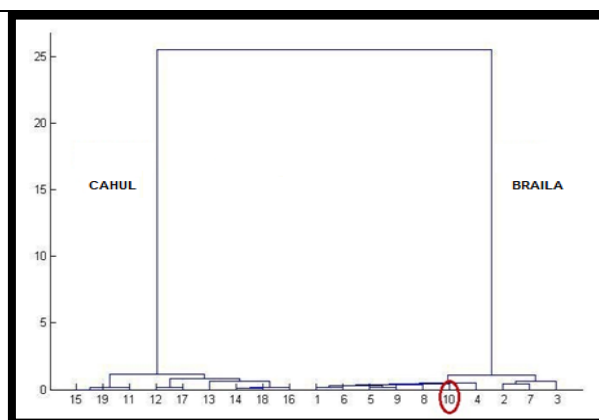


Figura V.2. Identificarea originii geografice a unei probe necunoscute de mărar (cod proba 10) care în fapt provine din județul Brăila, prin evaluarea similarității proprietăților fizico-chimice ale acesteia cu cele ale probelor de mărar cu aceeași origine (cod probe 1-9), respectiv cu cele ale celor provenind din Cahul (cod probe 11-19).

Pentru identificarea originii geografice a unei probe necunoscute, se determină aceleași proprietăți fizico-chimice ca și pentru probele de mărar introduse în baza de date de instruire. Apoi se determină noua dendrogramă, luând în calcul cele 19 probe, respectiv cele 18 din baza de date de instruire și proba necunoscută. Această procedură este exemplificată în figura V.2., care prezintă rezultatele obținute pentru testarea sensibilității aplicației, proba necunoscută (cod proba 10) testată provenind de fapt din județul Brăila. Probele de mărar din baza de date de instruire provenind din județul Brăila au codul de probe 1-9, iar cele recoltate din Cahul au codul 11-19. Se observă că aplicația identifică corect originea probei necunoscute de mărar, proba 10 fiind corect atribuită ramurii arborelui de clasificare specific probelor originare din județul Brăila. Aceeași sensibilitate a fost obținută și în cazul analizei corectitudinii clasificării unei probe necunoscute care în fapt este originară din Cahul.

Evaluarea eficienței aplicației de inteligență artificială din punctul de vedere al selectivității a fost realizată prin testarea unor probe de mărar care provin din alte zone

geografice decât cele modelate. Astfel, au fost testate probe de mărar provenind din județul Vrancea (România), respectiv din județul Brașov (România).

In concluzie, rezultatele obținute pentru aplicația HCA indică fezabilitatea identificării originii geografice a probelor de mărar in acest mod. Eficiența sistemului este remarcabilă atât din punctul de vedere al sensibilității cât și a selectivității.

Rezultatele obținute în evaluarea sensibilității și selectivității aplicației o recomandă ca o metodă automată foarte eficientă pentru identificarea provenienței geografice a unei probe necunoscute de mărar.

V.2.1.2. Evaluarea puterii de generalizare a sistemelor HCA pentru identificarea eficientă a originii geografice

Pentru evaluarea puterii de generalizare a sistemelor HCA pentru identificarea originii geografice a probelor de mărar, s-a efectuat un studiu comparativ al dendrogramei construite pentru diverse combinații de seturi de probe de mărar provenind din zone geografice diferite. In acest sens, au fost construite dendrograme folosind caracteristicile a câte 10 probe din județele Brăila, Vrancea, Brașov (România) și din Cahul (Republica Moldova). Astfel, au fost determinate pentru cele 40 probe de mărar următoarele proprietăți fizico-chimice: umiditatea, cenușa, conținutul de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic [5].

Dendrograma obținută pentru 10 probe de mărar din județul Brăila (cod probe 1-10) și 10 probe de mărar din Cahul (cod probe 11-20) este prezentată în figura V.3. Se observă că aceasta asigură o discriminare foarte clară a celor 2 ramuri formate pentru probele analizate în funcție de originea geografică a acestora. Distanța dintre ramurile corespunzătoare probelor de mărar din județul Brăila și a celor din Cahul este de aproape 20 de ori mai mare decât distanța maximă dintre probele cu aceeași proveniență geografică.

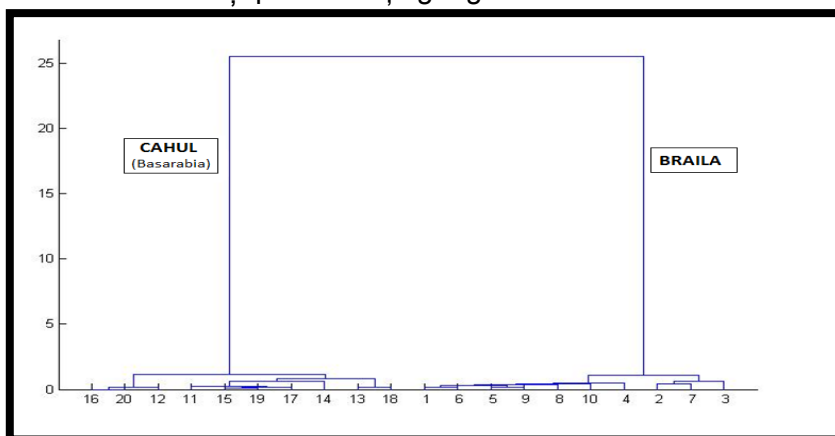


Figura V.3. Dendrograma determinată pentru discriminarea probelor de mărar din județul Brăila (cod probe 1-10) și din Cahul (cod probe 11-20).

Metoda de detecție și identificare a zonei geografice este caracterizată de o sensibilitate și o selectivitate deosebită: in dendrogramă nu există nici o falsă pozitivă în niciuna dintre cele 2 ramuri. Acest lucru este remarcabil, având în vedere ca județul Brăila este relativ apropiat din punct geografic de Cahul, deci condițiile pedo-climatice nu diferă foarte mult.

Aceleași performanțe în recunoașterea originii geografice sunt obținute și pentru probe de mărar provenind din alte zone relativ apropiate, respectiv din județul Brăila și din județul Vrancea. In cazul discriminării probelor de mărar ce provin din județul Brașov și din județul Vrancea performanțele sistemului de discriminare sunt mai reduse. Metoda de clustering

aglomerativ asigură și recunoașterea eficientă a originii probelor de mărar ce provin din zone geografice caracterizate de condiții pedo-climatice mai diferite, respectiv din județul Vrancea și din Cahul. Aceeași senzitivitate și selectivitate se înregistrează și în recunoașterea originii geografice a probelor de mărar recoltate din zone geografice care au condiții pedo-climatice foarte deosebite, respectiv din județul Brașov și din Cahul. În schimb, pentru a deosebi probele de mărar din județul Brăila și din județul Brașov probabilitate este mai scăzută în recunoaștere corectă a originii geografice a unei probe de mărar necunoscută.

În concluzie, eficiența sistemului HCA de recunoaștere a originii probelor de mărar nu poate fi generalizată pentru orice pereche de zone geografice. Puterea de modelare și de discriminare a dendrogramelor trebuie verificată pentru fiecare combinație de zone geografice de origine a probelor de mărar care trebuiesc modelate.

V.2.1.3. Evaluarea numărului de zone geografice de origine ce pot fi testate concomitent prin HCA

Rezultatele prezentate în capitolul anterior arată că sistemul de clasificare realizat prin PCA cu ajutorul aplicației software *The Unscrambler* poate testa originea unei probe de mărar în mod concomitent pentru toate zonele geografice (clase) luate în studiu, respectiv județul Brăila, județul Brașov, județul Vrancea și Cahul, pe baza unei selecții de proprietăți fizico-chimice și a comportării spectrale a acestora. Discriminarea modelată pe baza influenței condițiilor pedo-climatice asupra proprietăților fizico-chimice și spectrale poate fi decelată prin PCA numai pentru arii geografice largi (e.g. Moldova vs. Transilvania), pentru care există diferențe semnificative de altitudine și nivel al precipitațiilor.

Pe de altă parte, rezultatele prezentate în secțiunea anterioară par să indice că folosirea sistemelor HCA dezvoltate prin clustering aglomerativ permit o discriminare mai eficientă a probelor de mărar în funcție de originea lor geografică. În această secțiune sunt prezentate rezultatele obținute în cazul creșterii progresive a numărului de zone geografice testate concomitent, cu maximă eficiență (senzitivitate și selectivitate), cu ajutorul aplicației software *Matlab 2013a* [5].

Dendrograma construită pentru probele de mărar recoltate din județul Brăila, din județul Brașov și din Cahul este formată din 3 ramuri distincte fără a exista vreo falsă pozitivă în nici una dintre cele 3 ramuri formate și putem trage concluzia că sistemul multivariat HCA poate fi folosit în mod eficient și în cazul în care testăm concomitent originea unei probe de mărar necunoscute pentru cele trei zone geografice modelate.

Pentru a putea evalua eficiența sistemului HCA construit pentru patru zone geografice, dendrograma aferentă probelor de mărar provenite din toate zonele studiate a fost mai întâi construită cu ajutorul a numai 5 probe provenind din fiecare zonă modelată, respectiv din județul Brăila, județul Brașov, județul Vrancea, și din Cahul. În acest caz, se observă formarea a numai trei ramuri distincte. În aceste ramuri nu se înregistrează nici o falsă pozitivă. În schimb, dendrograma construită cu numai câte 5 probe de mărar nu decelează probele de mărar din județul Vrancea de cele din județul Brașov, acestea formând o singură ramură în dendrograma.

Pe de altă parte, eficiența mai scăzută se poate datora numărului mai redus (5) de probe utilizate pentru modelarea caracteristicilor probelor de mărar provenite din fiecare zonă geografică. De aceea s-a verificat în ce măsură eficiență poate fi îmbunătățită prin mărirea numărului de probe de mărar folosite în baza de date de instruire din fiecare zonă geografică. Dendrograma aferentă probelor de mărar construită cu câte 7 probe de mărar provenite din toate cele patru zone studiate nu poate distinge disimilaritatea probelor de mărar din județul Vrancea de cele din județul Brașov.

Dendrograma obținută crescând numărul de probe folosite pentru modelare la câte 8 din fiecare din cele 4 zone geografice studiate, adică 32, rezultă că nici in acest caz probele de mărar din județul Vrancea nu sunt decelate de cele din județul Brașov. In schimb, pe abscisa dendrogramei mai sunt vizibile doar 30 de probe, 2 din probele din Cahul depășind capacitatea de reprezentare grafica a aplicației software *Matlab 2013a*.

Dar aspectul cel mai important este faptul că in acest caz, apare un număr important de false positive.

In concluzie, numărul zonelor geografice de origine ce pot fi testate concomitent, in mod eficient, pentru clasificarea probelor de mărar este de 3 clase. Discriminarea probelor de mărar ce provin din 4 zone diferite este compromisă datorită falselor pozitive ce se regăsesc în toate ramurile dendrogramei formate.

V.2.2. Leuștean (*Levisticum officinale*)

Pentru realizarea sistemului HCA de recunoaștere automată a originii geografice a probelor de leuștean au fost analizate probe provenite din județul Brăila și din județul Brașov. Toate probele au fost caracterizate din punct de vedere al următoarelor proprietăți fizico-chimice: umiditate, cenușa, conținut de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic. Sistemul HCA dezvoltat pentru probele de leuștean a fost dezvoltat tot cu ajutorul programului *Matlab 2013a*, folosind același algoritm ca și in cazul probelor de mărar, respectiv cel de clustering aglomerativ [6,7].

Analiza eficienței cu care se poate atribui identitatea zonei geografice de origine a fost realizată cu ajutorul dendrogramelor determinate pentru probele de leuștean din Brașov și din județul Brăila. In dendrograma obținută pentru 10 probe de leuștean din județul Brașov și 10 probe de leuștean din Brăila, apare un număr important de false pozitive.

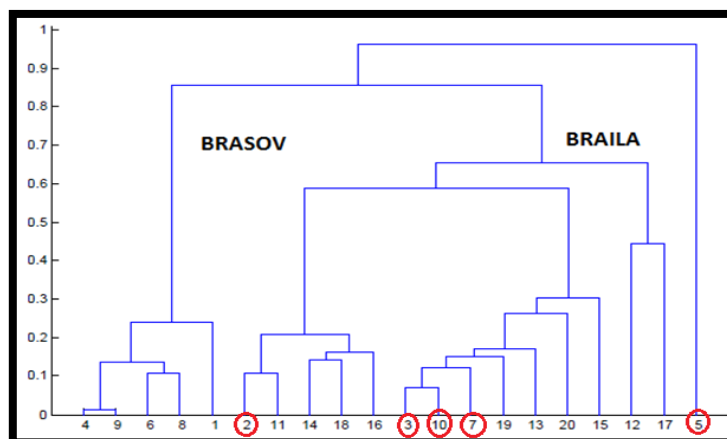


Figura V.4. Dendrograma determinată pentru discriminarea probelor de leuștean din județul Brașov (cod probe 1-10) și județul Brăila (cod probe 11-20)

In concluzie, HCA nu poate genera un sistem eficient de recunoaștere a originii geografice. Distanța dintre ramurile corespunzătoare probelor de Brașov, este de 0.8 de ori mai mare decât distanța maximă dintre probele cu aceeași proveniență geografică, distanță relativ mică, pentru a putea face discriminarea probelor de leuștean conform originii de proveniență.

V.3. Discriminarea ciupercilor *Pleurotus* (*Pleurotus ostreatus*) și *Champignon* (*Agaricus bisporus*) prin HCA

Probele de ciuperci *Pleurotus* și *Champignon* au fost studiate pentru a evalua în ce măsură HCA poate constitui o metodă eficientă pentru discriminarea lor automată pe baza aceleiași proprietăți fizico-chimice determinate și pentru restul vegetalelor studiate (mărar și leuștean), respectiv umiditate, cenușa, conținutul de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic. Cele două clase de probe au fost modelate prin metoda HCA, aplicând algoritmul de clustering aglomerativ. Aplicația a fost dezvoltată cu ajutorul programului *Matlab 2013a*.

Arborele de clasificare (dendrograma) a fost obținută prin calculul distanței Euclidiene dintre fiecare pereche de probe. Funcția folosită pentru cuplarea probelor în perechi (clustere binare) și apoi în clustere mai mari a probelor a fost funcția de unire medie (eng. *average linkage function*) [7].

Dendrograma obținută prin HCA pentru 10 probe de ciuperci *Pleurotus* și 10 probe de ciuperci *Champignon* asigură o discriminare foarte clară a două ramuri, fiecare ramură fiind asociată uneia din clasele de ciuperci analizate.

Arborele de clasificare asigură o recunoaștere automată foarte eficientă a tipului de ciuperci, având în vedere că nu există probe incorect clasificate în niciuna din ramuri. Altfel spus, aceasta metodă de recunoaștere a tipului de ciuperci este atât sensibilă cât și selectivă.

Acest lucru ne permite să tragem concluzia că proprietățile fizico-chimice considerate sunt relevante pentru scopul propus și adecvate pentru metoda de recunoaștere a formelor nesupravegheată folosită pentru dezvoltarea aplicației de clasificare, respectiv HCA.

Având în vedere puterea mare de discriminare a HCA observată în cazul ciupercilor *Pleurotus* și a celor *Champignon*, am verificat în ce măsură se poate realiza o discriminare eficientă a celor două clase de ciuperci folosind un număr mai mic de probe pentru modelare. Dacă sensibilitatea și selectivitatea sistemului se păstrează și în acest caz, atunci aplicația poate fi dezvoltată pentru un număr mai mare de categorii de ciuperci aparținând unor familii sau genuri diferite.

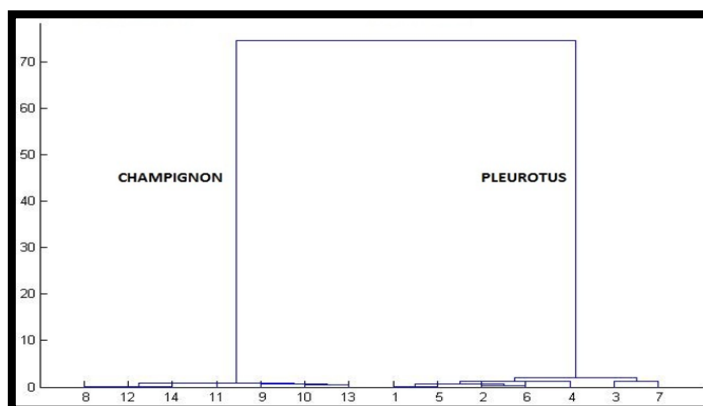


Figura V.5. Dendrograma determinată pentru discriminarea probelor de ciuperci *Pleurotus* (cod probe 1 - 7) și a probelor de ciuperci *Champignon* (cod probe 8 - 14) folosind pentru modelare câte 7 probe de ciuperci din fiecare clasă

O măsură a eficienței discriminării este disimilaritatea maximă dintre cele două clase modelate, care în cazul dendrogramei construite cu câte 10 probe din fiecare clasă are valoarea 74.7. Construind dendrogramele luând în calcul un număr din ce în ce mai mic de probe, s-a obținut aceeași valoare a disimilarității maxime până la un număr de câte 7 probe din cele 2

clase modelate. În cazul în care se construiește dendrograma cu câte 6, 5 sau 4 probe, disimilaritatea devine egală cu 75.88. În cazul în care numărul probelor scade la câte 3 probe din cele 2 clase, disimilaritatea maximă este 76.78. Se poate trage deci concluzia ca variația disimilarității maxime este minoră și deci numărul de probe de ciuperci luate în calcul poate fi scăzut în funcție de numărul de clase de ciuperci care trebuie discriminate automat.

V.4. Concluzii

Performanțele sistemului HCA dezvoltat pentru recunoașterea originii geografice a probelor de mărar a fost evaluată mai întâi, ca studiu de caz, analizând dendrograma determinată pentru probe care provin din județul Brăila (România) și din Cahul (Republica Moldova). Testele efectuate indică faptul că aplicația este caracterizată de o eficiență deosebită. Astfel, clasificarea probelor aparținând claselor de mărar specifice zonelor geografice modelate are loc cu o sensibilitate deosebită. Totodată, aplicația este caracterizată de o selectivitate remarcabilă, distingând clar disimilaritatea probelor de mărar originare din alte zone geografice decât cele modelate. Mai mult, atunci când proba testată nu aparține nici uneia din clasele modelate, aplicația indică care dintre aceste clase este cea mai similară cu proba necunoscută din punctul de vedere al proprietăților fizico-chimice caracteristice probelor de mărar. Extinzând studiul pentru diverse combinații de zone geografice, rezultatele au indicat că HCA este o tehnică foarte adecvată pentru detecția automată a zonei de origine a probelor de mărar. Analiza dendrogramelor a indicat eficiența (sensibilitatea și selectivitatea) deosebită a acestei metode de inteligență artificială în detecția zonei de proveniență geografică a probelor de mărar. Aceeași eficiență deosebită se înregistrează în cazul în testării concomitente a trei din zone luate în studiu. Totuși, numărul zonelor de origine testate concomitent nu poate fi crescut oricât. În cazul testării concomitente a tuturor celor patru zone modelate, eficiența sistemului de clasificare a scăzut semnificativ. Pentru a putea trage concluzii comparative, sistemul HCA pentru identificarea provenienței geografice a probelor de leuștean a fost dezvoltat în același mod ca și în cazul probelor de mărar. Astfel, dendrogramele au fost obținute prin clustering aglomerativ, cu ajutorul aplicației software *Matlab 2013a*, pe baza unui număr de șapte caracteristici fizico-chimice ale acestor produse vegetale. Aceste proprietăți au fost determinate pentru probe de leuștean ce provin din două regiuni geografice, respectiv județul Brașov și județul Brăila.

Rezultatele obținute indică faptul că aplicația multivariată nu este eficientă pentru discriminarea probelor de leuștean provenite din aceste două regiuni, numărul falselor pozitive fiind semnificativ. În cazul probelor de leuștean, PCA este mai potrivită decât HCA. În plus, PCA ne poate indica și care sunt variabilele care au o putere mare de modelare și/sau discriminare și care dintre acestea pot fi eliminate din calcul pentru că au o contribuție minoră în atribuirea identității de clasă.

HCA s-a dovedit o metodă automată foarte eficientă și în cazul discriminării probelor de ciuperci *Pleurotus* și *Champignon* care aparțin, din punct de vedere sistematic, aceleași clase și aceluiași ordin cu *Agaricus sp.*, însă fac parte din familii și genuri diferite. Pentru a putea trage concluzii comparative, dendrogramele au fost obținute și în acest caz cu ajutorul aplicației software (*Matlab 2013a*). Dendrogramele au fost obținute prin același algoritm (clustering aglomerativ) și pe baza aceluiași caracteristici fizico-chimice. Aceste proprietăți au fost determinate pentru un număr de 20 de probe de ciuperci ce aparțin regnurilor *Pleurotus* și *Champignon*. Analiza dendrogramelor a indicat eficiența deosebită a acestei metode de inteligență artificială în discriminarea ciupercilor *Pleurotus* de ciupercile *Champignon*. Din acest punct de vedere, metoda clustering aglomerativ se dovedește a fi foarte performantă pentru

identificarea tipului probelor vegetale studiate și totodată pentru discriminarea eficientă a acestora. Numărul mic de probe care sunt necesare pentru modelarea și discriminarea celor două tipuri de ciuperci indică faptul că, în acest caz, se poate dezvolta un sistem automat HCA de discriminare a unui număr important de clase de vegetale care pot fi testate concomitent.

Bibliografie

- [1] Y. Freund, *A more robust boosting algorithm*. arXiv:0905.2138v1, 2009.
- [2] J. Friedman, *Greedy function approximation: A gradient boosting machine*. Annals of Statistics, Vol. 29, No. 5, pp. 1189–1232, 2001.
- [3] J. Friedman, T. Hastie, R. Tibshirani, *Additive logistic regression: A statistical view of boosting*. Annals of Statistics, Vol. 28, No. 2, pp. 337–407, 2000.
- [4] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, second edition. Springer, New York, 2008.
- [5] S. C. Ghinita, M. Praisler, G. Iordăchescu, *Automatic detection of the geographical origin of dill (Anethum graveolens) based on hierarchical cluster analysis*, Abstract book of the Second International Conference on Natural and Anthropogenic Risks - ICNAR 2014, 4 - 7 June, 2014, Bacău, Romania, pp.123, ISSN 2360-4018 <http://icnar.ub.ro/>
- [6] V. Kecman, T.M. Huang, M. Vogt, *Iterative Single Data Algorithm for Training Kernel Machines from Huge Data Sets: Theory and Performance*, In *Support Vector Machines: Theory and Applications*, Edited by Lipo Wang, 255–274. Berlin: Springer-Verlag, 2005.
- [7] C. Seiffert, T. Khoshgoftaar, J. Hulse, A. Napolitano, *RUSBoost: Improving classification performance when training data is skewed*, 19th International Conference on Pattern Recognition, pp. 1–4, 2008.

CAPITOLUL VI

Contribuții proprii obținute în detecția și identificarea originii produselor vegetale folosite în industria alimentară prin rețele neuronale artificiale (ANN)

VI.1. Condiții experimentale

Pentru dezvoltarea unui sistem ANN care să realizeze eficient detecția și identificarea automată a originii produselor vegetale studiate (mărar, leuștean și ciuperci), pe baza principalelor lor proprietăți fizico-chimice (umiditate, cenușă, conținut de substanță uscată, carbohidrați, proteine, grăsimi și glutamat monosodic), s-a folosit aplicația software *nnModel* (Neural Network Model) [1].

Pentru găsirea soluției optime pentru sistemele ANN construite, au fost studiate diverse arhitecturi ale rețelei artificiale. Astfel, numărul maxim de neuroni introduși în stratul ascuns a fost variat de la 1 la 6. De precizat că numărul maxim de neuroni din acest strat nu poate fi mai mare de 6, conform formulei $n-1$, unde n reprezintă numărul de variabile ($n = 7$ proprietăți fizico-chimice) din baza de date de instruire. [2,3]. De asemenea, numărul maxim de cicluri de antrenare a fost variat între 1000-100000. Modalitatea selectată pentru adăugarea de straturi ascunse a fost incrementarea automată [4].

VI.2. Identificarea originii geografice a produselor vegetale folosite în industria alimentară prin ANN

VI.2.1. Mărar (*Anethum Graveolens*)

VI.2.1.1. Instruirea sistemului ANN

Cunoșterea originii geografice este importantă pentru industria alimentară din punct de vedere nutrițional, gastronomic și fizico-chimic. Proprietățile fizico-chimice ale produselor vegetale variază în funcție de calitatea solului, precipitații, sezon, specie, temperatură, etc.

În cele ce urmează prezentăm performanțele unui sistem ANN dezvoltat pentru recunoaștere automată a zonei geografice de origine a probelor de mărar (*Anethum graveolens*), pe baza următoarelor proprietăți fizico-chimice: umiditatea (cod variabilă V1), conținutul de substanță uscată (V2), carbohidrați (V3), proteine (V4), grăsimi (V5), cenușă (V6) și de glutamat monosodic (V7) [4].

În prima instanță, a fost estimată puterea de discriminare a fiecărei variabile, prin intermediul amplitudinii variației și semnificația statistică a diferenței dintre valorile medii ale proprietăților fizico-chimice determinate pentru probele de mărar originare din cele patru zone studiate. Variabilele identificate ca fiind caracterizate de o putere de discriminare semnificativă au fost selectate pentru formarea bazei de date de instruire a ANN.

Rezultatele obținute demonstrează că influența condițiilor pedo-climatice asupra proprietăților fizico-chimice este într-adevăr complexă și deci poate fi pusă în valoare pentru a identifica originea geografică a probelor de mărar numai printr-o abordare multivariată.

Pentru evaluarea eficienței sistemului ANN construit pe baza celor șapte proprietăți fizico-chimice menționate, au fost testate un număr de 10 probe de mărar (cod probe 1-10) provenite din județul Brăila, 10 (cod proba 11-20) din județul Brașov, 10 (cod proba 21-30) din județul Vrancea și 10 (cod proba 31-40) din Cahul.

Astfel, pentru identificarea originii geografice a unei probe, se introduc în sistemul ANN valorile determinate pentru proprietățile fizico-chimice (V1-V7). La interogare, sistemului ANN calculează valoarea de ieșire a probei analizată (V8). Această se compară cu valorile de ieșire alocate fiecărei zone geografice. Cu cât valoarea de ieșire a unei probe este mai apropiată de valoarea de ieșire alocată unei zone anume, cu atât probabilitatea identificării este mai mare.

VI.2.1.2. Validarea sistemului ANN

Având în vedere rezultatele obținute în cazul testării celor 40 de probe de mărar, s-a construit un nou sistem ANN, care a fost dezvoltat, pentru păstrarea echilibrului între numărul de probe folosite pentru caracterizarea fiecărei zone geografice, introducând în baza de date de instruire câte 9 probe de mărar din fiecare zonă modelată. Setul de probe de validare a fost deci constituit în acest caz din 4 probe de mărar, respectiv câte una din fiecare zonă de origine studiată.

Având în vedere eficiența remarcabilă a sistemului ANN în recunoașterea automată a originii geografice a probelor de mărar și în această configurație, a fost evaluată posibilitatea de a micșora și mai mult volumul bazei de date de instruire.

VI.3. Leuștean (*Levisticum Officinale*)

VI.3.1. Instruirea sistemului ANN

În această secțiune prezentăm rezultatele obținute în cazul unei aplicații ANN dezvoltată pentru recunoașterea automată a originii geografice a probelor de leuștean. Dezvoltarea bazei de date de instruire a sistemului ANN s-a realizat prin analiza a 10 probe provenite din județul Brașov și 10 din județul Brăila, pentru care s-au determinat aceleași proprietăți fizico-chimice ca și în cazul probelor de măr, respectiv umiditatea (cod variabilă V1), conținutul de substanță uscată (V2), carbohidrați (V3), proteine (V4), grăsimi (V5), cenușa (V6) și conținutul de glutamat monosodic (V7).

Puterea de discriminare a fiecărei variabile a fost estimată prin intermediul amplitudinii variației și semnificația statistică a diferenței dintre valorile medii ale proprietăților fizico-chimice determinate pentru probele de leuștean originare din cele două județe studiate.

Având în vedere rezultatele obținute, putem trage concluzia că variabilele cu cea mai mare putere de discriminare a originii geografice a probelor de leuștean sunt: umiditatea, conținutul de substanță uscată, carbohidrați, proteine și de glutamat monosodic.

Baza de date formată cu proprietățile fizico-chimice măsurate pentru cele 20 de probe de leuștean a fost folosită pentru instruirea unui sistem ANN format din trei straturi, respectiv un strat de intrare cu 7 noduri (corespunzătoare variabilelor selectate), un strat intermediar și un strat de ieșire cu două noduri (corespunzătoare celor două clase de leuștean modelate, respectiv leuștean din județul Brașov și din județul Brăila). Aplicația a fost realizată cu ajutorul aplicației software *NnModel*.

Altfel spus, originea geografică a probelor este identificată corect în proporție de 100% iar atribuirea originii este caracterizată de o probabilitate foarte mare.

VI.4. Discriminarea ciupercilor *Pleurotus (Pleurotus ostreatus)* și *Champignon (Agaricus bisporus)* prin ANN

În această secțiune este prezentată o aplicație ANN dezvoltată pentru identificarea și discriminarea ciupercilor *Pleurotus (Pleurotus ostreatus)* și a ciupercilor *Champignon (Agaricus bisporus)*. Sistemul ANN funcționează pe aceleași proprietăți fizico-chimice ca și sistemele ANN dezvoltate pentru măr și leuștean, respectiv: umiditatea (cod variabilă V1), conținutul de substanță uscată (V2), carbohidrați (V3), proteine (V4), grăsimi (V5), cenușă (V6) și de glutamat monosodic (V7). Acestea au fost determinate pentru 10 probe de ciuperci *Pleurotus* și 10 de ciuperci *Champignon* brune. Sistemul ANN a fost constituit tot cu ajutorul aplicației software *NnModel* (Network neuronal Model).

Pentru a putea identifica variabilele caracterizate de o putere de discriminare semnificativă, s-a studiat mai întâi variația caracteristicilor fizico-chimice atât din punctul de vedere al tipului de vegetale, cât și în cadrul aceluiași tip de ciuperci.

VI.5. Concluzii

Pentru dezvoltarea sistemului ANN dedicat identificării automate a originii **probelor de măr** au fost folosite un număr de 40 de probe, respectiv câte 10 probe din județul Brăila, județul Brașov, județul Vrancea și din Cahul (Republica Moldova). Instruirea sistemului s-a realizat pe baza unui număr de șapte caracteristici fizico-chimice determinate pentru aceste probe de măr, respectiv umiditatea, cenușa, conținutul de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic. Capacitatea acestor variabile de a discrimina

probelor de mărar în funcție de proveniența lor geografică a fost evaluată prin intermediul amplitudinii variației lor și a semnificației statistice a diferenței dintre valorile lor medii.

Sistemul ANN a fost dezvoltat folosind aplicația software *NnModel* (Neural Network Model). Rezultatele au indicat faptul că atunci când în baza de date de instruire se introduc proprietățile fizico-chimice câte 9 probe de mărar din fiecare zonă geografică modelată, sistemul ANN este caracterizat de o eficiență (sensitivitate și selectivitate) remarcabilă. Testele de validare a sistemelor ANN astfel construite au indicat o rată de clasificare de 100%, originea tuturor probelor de mărar testate fiind corect identificată. În concluzie, pentru obținerea unui sistem ANN care să execute cu eficiența maximă identificarea automată a originii geografice a probelor de mărar, este necesar ca instruirea acestuia să se realizeze cu ajutorul unei baze de date care să conțină cel puțin 9 probe de mărar din fiecare zonă modelată.

Dacă instruirea se realizează introducând în baza de date câte 8 sau 7 probe de mărar din fiecare zonă, sistemul prezintă aceeași eficiență deosebită numai în recunoașterea originii probelor ce provin din zone în care condițiile pedo-climatice diferă semnificativ (e.g. zona de șes vs. zona de deal). În acest caz însă, eficiența în discriminarea probelor ce provin din zone geografice similare este mai scăzută (e.g. județul Brașov și județul Vrancea). Dacă însă în baza de date de instruire se includ doar câte 6 probe de mărar, atunci eficiența sistemului ANN scade și în cazul identificării originii geografice a unor probe provenind din zone cu condiții pedo-climatice relativ diferite (e.g. județul Brăila și județul Brașov).

Această comportare este explicabilă, având în vedere faptul că și în cazul dezvoltării sistemelor HCA de recunoaștere a originii probelor de mărar s-au obținut rezultate analoage. Dendograma obținută în condiții similare cu sistemul ANN, respectiv cu o bază de date de instruire formată cu aceleași 7 proprietăți fizico-chimice măsurate pentru câte 8 probe de mărar din fiecare zonă modelată, nu poate deosebi probele de mărar provenind din județul Brașov de cele din județul Vrancea. Trebuie însă subliniat faptul că și în aceste condiții, în care sistemul ANN generează un număr de false pozitive, acesta este caracterizat de o rată de clasificare corectă semnificativ mai mare decât cea specifică sistemului HCA.

Aceeași dendrogramă confirmă și faptul că proprietățile fizico-chimice a probelor de mărar din județele Brașov și Vrancea sunt mai similare cu cele ale probelor ce provin din județul Brăila decât cu cele ale probelor din Cahul. Aceste rezultate explică modul în care eficiența sistemului ANN scade atunci când numărul de probe de mărar introduse în baza de date de instruire scade de la 7-8 la 6 probe.

Un alt aspect important este faptul că dezvoltarea sistemului ANN nu este afectată de limitarea întâlnită în cazul sistemelor HCA construite cu ajutorul aplicației software *Matlab*, respectiv a numărului de probe ce pot fi afișate pe abscisa dendrogramei. Altfel spus, în cazul sistemelor ANN, atât numărul probelor de mărar folosite pentru modelarea identității lor geografice, cât și numărul zonelor ce pot fi testate concomitent poate fi crescut oricât. De asemenea, numărul probelor necunoscute ce pot fi testate simultan poate fi oricât de mare. Din acest punct de vedere, se poate trage concluzia cel mai operativ mod de identificare automată a originii geografice a probelor de mărar este folosirea unui sistem ANN.

Pe de altă parte, rezultatele prezentate în capitolele anterioare arată faptul că sistemul HCA este mai performant în recunoașterea originii geografice a probelor de mărar decât sistemul PCA. După cum s-a menționat, discriminarea originii geografice a probelor de mărar cu ajutorul PCA este eficientă numai pentru zone geografice foarte diferite din punct de vedere pedo-climatic (e.g. județul Brașov vs. județul Brăila sau Cahul), dar nu și între județul Brăila și Cahul sau Vrancea. Este important de subliniat din nou faptul că aceste rezultate au fost obținute deși informația inclusă în baza de date de intrare a sistemului PCA a fost mult mai bogată (atât cele

7 proprietăți fizico-chimice cât și o selecție de proprietăți spectrale) decât in cazul sistemelor HCA sau ANN.

Putem trage deci concluzia că dacă HCA realizată prin clustering aglomerativ este mai adecvată pentru recunoașterea automată a originii geografice a probelor de mărar decât PCA, sistemul optim ANN este mai eficient și mai operativ decât ambele metode menționate.

Pe de altă parte, nu trebuie neglijat faptul că PCA, spre deosebire de HCA și ANN, poate indica, prin intermediul distribuției ponderilor variabilelor, care sunt variabilele care contribuie cel mai mult la modelarea fiecărei zone geografice, astfel încât cele redundante să poată fi eliminate și astfel să se obțină un sistem de calcul mai suplu și mai ieftin. Altfel spus, că și in cazul HCA, este utilă înainte de construcția unui sistem ANN să se realizeze o analiză exploratorie prin PCA.

In baza de date de instruire folosită pentru construcția sistemului ANN construit pentru identificarea automată a probelor de **leuștean** au fost incluse aceleași proprietăți fizico-chimice ca și in cazul probelor de mărar, respectiv umiditatea, cenușa, conținutul de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic. Probele de leuștean analizate provin din județul Brașov și județul Brăila. Identificarea proprietăților fizico-chimice care sunt influențate semnificativ de zona geografică de origine datorită influenței condițiilor pedo-climatice a fost realizată prin evaluarea amplitudinii variației și a semnificației statistice a diferenței dintre mediile obținute pentru aceste variabile in cazul celor două clase de probe de leuștean.

Aplicația a fost dezvoltată cu ajutorul programului *nnModel* (Neural Network Model). Dacă sistemul ANN este dezvoltat incluzând in baza de date de instruire proprietățile fizico-chimice a cel puțin câte 9 probe de leuștean din fiecare din cele două zone geografice modelate, atunci se obține o rată de clasificare corectă foarte bună. In concluzie, metoda propusă poate fi folosită cu succes atât pentru verificarea automată a corectitudinii etichetării leușteanului comercializat, cât și pentru studii de analiza senzorială.

Scăderea numărului de probe de leuștean incluse in baza de date de instruire conduce la o scădere semnificativă a sensibilității sistemului in recunoașterea originii probelor provenite din județul Brăila și a selectivității sistemului in recunoașterea probelor de leuștean originare din județul Brașov. Această comportare este similară celei deduse cu ocazia validării sistemului ANN dezvoltat pentru identificarea originii probelor de mărar. Sistemele ANN par deci potrivite pentru identificarea originii unei game largi de vegetale.

Se observă faptul că rezultatele obținute cu sistemul ANN dezvoltat pentru identificarea originii probelor de leuștean este mult mai performant decât cel construit pe baza HCA. Performanțele sistemului ANN construit cu câte 9 probe de leuștean din fiecare zona modelată par a fi comparabile, cu cele ale sistemului PCA dezvoltat pentru această vegetală.

Atât ciupercile Champignon cât și ciupercile Pleurotus prezintă o importanță deosebită din punct de vedere alimentar. Proprietățile acestora sunt de mare interes pentru consumatori, industria alimentară, farmaceutică și terapeutică. Cerințele acestora sunt foarte specifice, fiind legate de diete, rețete, toleranță sau intoleranță față de anumiți compuși.

Sistemul ANN dezvoltat pentru identificarea și discriminarea **ciupercilor Pleurotus (*Pleurotus ostreatus*) și a ciupercilor Champignon (*Agaricus bisporus*)** este capabil să recunoască genul probelor de ciuperci cu mare acuratețe, chiar și in cazul in care numărul de probe folosit pentru instruire este relativ scăzut. Acesta funcționează pe baza aceluiași proprietăți fizico-chimice ca și sistemele ANN dezvoltate pentru mărar și leuștean, respectiv: umiditatea, cenușa, conținutul de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic. Aplicația a fost dezvoltată cu ajutorul programului *nnModel* (Neural Network Model).

Studiul comparativ al variației proprietăților fizico-chimice folosite pentru construirea bazei de date de instruire a sistemului ANN a indicat faptul că, din punctul de vedere al nutrienților necesari organismului uman (carbohidrați, proteine și grăsimi), ciupercile caracterizate de o umiditate mai mare au și un conținut mai mare de carbohidrați. Pe de altă parte, ciupercile cu un conținut de substanța uscată mai mare (respectiv umiditate mai mică) au un conținut de proteine și grăsimi mai mare. Astfel, ciupercile Champignon au o umiditate și un conținut de carbohidrați mai mare, generând și o cantitate de cenușă mai mare. Ciupercile Pleurotus au un conținut mai bogat în proteine, grăsimi și glutamat monosodic, compus care le și conferă astfel un gust mai delicios (umami). Rezultatele obținute la validarea sistemului ANN construit pentru discriminarea celor două genuri de ciuperci indică faptul că folosirea unui astfel de sistem expert este o modalitate de detecție eficientă și accesibilă (ieftină și ușor de utilizat) pentru screening-ul produselor vegetale.

Bibliografie

- [1] I. Nikola K. Kasabov, *Foundation of Neural Networks, Fuzzy systems, and Knowledge Engineering*, Cambridge, The MIT Press, 1998.
- [116] B. Widrow, M.A. Lehr, *30 Years of adaptive neural networks: Perceptron, Madaline, and Backpropagation*, Proceedings of the IEEE, vol. 78 no.9, 1990.
- [2] D. Anderson, G. McNeill, *Artificial neural networks technology*, Kaman Sciences corporation, 1992.
- [3] D. Dumitrescu, H. Costin, *Rețele neurale, teorie si aplicatii*, Teora, 1996.
- [4] S.C. Ghinita, M. Praisler, G. Iordăchescu, „Identification of the origin of vegetables based on Artificial Neural Networks”, Annals of ”Dunarea de Jos” University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II, Year VI (XXXVII) 2014, No. 1, p. 110 – 115. ISSN 2067-2071, http://www.phys.ugal.ro/Annals_Fascicle_2

CAPITOLUL VII

Concluzii generale, contribuții originale și perspective viitoare de cercetare

VII.1. Concluzii generale

Cercetarea realizată în cadrul tezei de doctorat, *Metode avansate de detecție și identificare a originii unor produse vegetale folosite în industria alimentară*, este un studiu comparativ realizat pentru identificarea metodei chemometrice optime care să permită obținerea unui sistem expert care să permită detecția și identificarea originii geografice a unei serii de vegetale pe baza principalelor lor proprietăți fizico-chimice și spectrale.

Alegerea acestei teme de cercetare este motivată de importanța identificării originii geografice a produselor vegetale pentru industria alimentară și farmaceutică. Condițiile pedoclimatice specifice locului de recoltare pot influența semnificativ calitatea nutrițională, gastronomică și/sau terapeutică a vegetalelor. Efectul acestor variabile asupra calității produselor poate fi decelat prin evaluarea caracteristicilor fizico-chimice și spectrale ale probelor vegetale.

Studiul a început printr-o analiză spectrală care a urmărit identificarea ferestrelor spectrale din spectrul UV-VIS care au cea mai mare putere de discriminare a diverselor tipuri de vegetale analizate, respectiv mărar (*Anethum Graveolens*), leuștean (*Levisticum officinale*), mazăre (*Pisum sativum*), cimbru (*Satureja Hortensis*), ciuperci Pleurotus (*Pleurotus ostreatus*) și Champignon (*Agaricus bisporus*). Rezultatele au indicat că absorbțiile din domeniul 250-550 nm sunt cele mai adecvate pentru decelarea originii geografice a vegetalelor studiate. Aceasta fereastră spectrală s-a dovedit a avea cea mai mare putere de discriminare atât în cazul în care de dorește deosebirea probelor de vegetale din același ordin, cât și a celor din ordine diferite.

Analiza comparativă a comportării spectrale a condus la următoarele concluzii: a) nu toate vegetalele prezintă în acest domeniu benzi de absorbție bine definite; b) unele vegetale prezintă variații intra-categorie destul de însemnate. Ca atare, pentru identificarea automată eficiența a originii geografice a vegetalelor se impune caracterizarea acestora vegetalelor studiate și prin alte proprietăți fizico-chimice.

O altă concluzie importantă a analizei spectrale este legată de caracterul metodelor de prelucrare a datelor experimentale care pot conduce la dezvoltarea unor sisteme automate de identificare a originii geografice a vegetalelor. Absorbțiile UV-VIS sunt înregistrate cu un pas de digitizare de 0,426829 nm. Ca atare, chiar dacă baza de date folosită pentru identificarea originii geografice a vegetalelor ar fi formată numai cu absorbțiile din domeniul 250-550 nm, deja s-ar lucra cu peste 600 de variabile. În concluzie, dezvoltarea unui sistem automat de recunoaștere a originii vegetalelor trebuie să se bazeze pe metode multivariate (chemometrice).

Ca atare, studiul a fost continuat printr-o analiză exploratorie realizată prin PCA. Aceasta este o metodă de inteligență artificială nesupravegheată și non-ierarhică. Ca atare, PCA poate indica, prin intermediul distribuțiilor scorurilor, în ce măsură probele formează clustere *in mod natural*, adică grupări de probe cu proprietăți foarte similare. Evaluarea eficienței unui sistem automat de identificare sau detecție este apreciată în primul rând prin sensibilitatea și selectivitatea acestuia. În cazul PCA, densitatea acestor clustere este o măsură foarte bună a sensibilității sistemului. Totodată, distribuțiile scorurilor pot indica cât de bună poate fi discriminarea diverselor grupuri formate (cât de depărtate sunt clusterelor unele de altele).

În plus, PCA poate indica, prin intermediul distribuțiilor ponderilor variabilelor, care sunt variabilele care contribuie cel mai mult la formarea diverselor clustere. Acest aspect este important nu numai pentru a înțelege din ce puncte de vedere sunt similare probele ce formează un cluster dat, dar și care sunt proprietățile care deosebesc cel mai mult diversele

clustere. Altfel spus, indica variabilele cu cea mai mare putere de modelare și pe cele cu cea mai mare putere de discriminare. Odată identificate aceste variabile, restul pot fi eliminate din baza de date de instruire, pentru ca informația pe care o aduc în sistem este ne semnificativă (redundantă). Altfel spus, pe lângă o reducere a dimensionalității informației incluse în baza de date, PCA permite și o selecție a variabilelor necesare pentru constituirea bazei de date de instruire cu variabilele minim necesare pentru atingerea scopului propus. Acest aspect este foarte important, pentru că reduce timpul, numărul și deci costul analizelor necesare pentru formarea bazei de date sau pentru testarea unei probe necunoscute. Sistemele PCA prezentate în această lucrare au fost dezvoltate folosind aplicația software *The Unscrambler*.

Primele analize exploratorii PCA au fost realizate pentru evaluarea posibilității de a identifica originea geografică a probelor de mărar și a celor de leuștean pe baza unei selecții de proprietăți fizico-chimice (umiditate, cenușă, conținutul de substanța uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic) și a comportării spectrale (absorbțiile din domeniul 250-550 nm) a acestora. În cazul mărărului, distribuția scorurilor a indicat formarea a două clustere, unul specific probelor de mărar din județul Brașov și altul asociat celor din județele Brăila, Vrancea și Cahul. Deci influența condițiilor pedo-climatice asupra caracteristicilor probelor de mărar poate fi decelată prin PCA numai pentru arii geografice largi (e.g. Moldova vs. Transilvania), pentru care există diferențe semnificative de altitudine și nivel al precipitațiilor. Variabilele cu cea mai mare putere de modelare sunt umiditatea, conținutul de proteine și de grăsime, precum și absorbțiile spectrale (mai ales cele din jurul lungimii de undă 360 nm). Variabila cu cea mai mare putere de discriminare a celor două clustere este conținutul de glutamat monosodic. Altfel spus, mărarul cu gustul delicios cel mai pregnant este cel originar din Brașov. În concluzie, metoda propusă poate fi folosită și pentru studii de analiză senzorială.

Analiza performanțelor sistemului PCA dezvoltat pentru identificarea originii geografice a probelor de leuștean a confirmat utilitatea folosirii bazelor de date hibride pentru obținerea unui sistem PCA performant dedicat recunoșterii automate a originii geografice. Conținutul de glutamat monosodic este una dintre cele mai importante variabile pentru discriminarea probelor de leuștean după originea lor geografică. Această variabilă contribuie semnificativ la formarea clusterului asociat probelor de leuștean de Brăila, ceea ce indică faptul că acestea au un gust delicios mai pronunțat, respectiv o valoare gastronomică mai ridicată decât leușteanul de Brașov. În consecință, PCA poate fi folosită și pentru automatizarea aplicațiilor de analiză senzorială.

Trebuie însă reținut faptul că variabilele care au cea mai mare putere de modelare și / sau discriminare identificate în cazul leușteanului diferă de cele aferente probelor de mărar. Ca atare, arhitectura sistemelor PCA dezvoltate pentru identificarea originii geografice a vegetalelor nu poate fi generalizată. Totodată trebuie remarcat faptul că, în ambele cazuri, influența condițiilor pedo-climatice poate fi decelată de sistemele PCA numai pentru arii geografice largi, pentru care există diferențe semnificative de altitudine și nivel al precipitațiilor.

Studiul realizat a indicat faptul că în PCA este o metodă chemometrică ce poate fi aplicată cu succes și pentru identificarea principalelor caracteristici fizico-chimice care pot afecta calitatea vegetalelor cu aceeași origine geografică. PCA a indicat faptul că similaritatea boabelor de mazăre este dată în principal de conținutul de glutamat monosodic, care este aproape constant pentru probele cu aceeași origine geografică. Proprietățile fizico-chimice care pot conduce la variații semnificative în calitatea boabelor de mazăre sunt umiditatea, conținutul de substanță uscată, proteine și, într-o măsură mai mică, de carbohidrați. Umiditatea este puternic corelată negativ cu ultimele trei proprietăți.

În cazul cimbrului, PCA a indicat că principalele proprietăți fizico-chimice care influențează în mod semnificativ calitatea probelor sunt umiditatea, conținutul de substanță uscată, proteine și în mai mică măsură, carbohidrați. Umiditatea este puternic și negativ

corelată cu ultimele trei variabile. Similaritatea probelor de cimbru este conținutul de glutamat monosodic, care este practic independent de restul variabilelor luate în studiu.

PCA s-a dovedit foarte utilă și pentru discriminarea genului de ciuperci. Studiul a fost realizat pentru discriminarea ciupercilor *Pleurotus* de cele *Champignon* brune. Acestea aparțin, din punct de vedere sistematic, aceleași clase și aceluiași ordin, însă fac parte din familii și genuri diferite. Sistemul PCA construit s-a dovedit foarte eficient. PCA a condus la comprimarea hiperspațiului inițial de 711 variabile la numai 3 dimensiuni (PC), ceea ce permite atribuirea identității de clasă cu resurse de calcul mult reduse.

Trebuie subliniat că aceste performanțe au fost obținute printr-o metoda de recunoaștere a formelor nesupravegheată, în baza de date de instruire nefiind incluse informații cu privire la originea sau genul probelor. Având în vedere că PCA a confirmat fezabilitatea dezvoltării unor sisteme multivariate dedicate identificării originii geografice a produselor vegetale, studiul a fost continuat cu HCA. Aceasta este tot o metodă nesupravegheată, clusterizarea având loc natural, ca și în cazul PCA. Spre deosebire însă de această tehnică însă, HCA este o metodă ierarhică. Ea indică, prin intermediul dendrogramelor, ierarhia probelor din punctul de vedere al (di)similarității. Algoritmul folosit a fost cel de clustering aglomerativ. Bazele de date de instruire folosite pentru determinarea dendrogramelor pentru diversele vegetale conțin următoarele variabile (proprietăți fizico-chimice), determinate pentru o serie de probe de mărar de origine cunoscută: umiditate, cenușa, substanță uscată, conținutul de carbohidrați, proteine, grăsimi și de glutamatul monosodic. Sistemele HCA prezentate în această lucrare au fost dezvoltate folosind aplicația software *Matlab 2013*.

Analiza performanțelor sistemului HCA dezvoltat pentru recunoașterea originii geografice a probelor de mărar a fost evaluată mai întâi, ca studiu de caz, analizând dendrograma determinată pentru probe care provin din județul Brăila (România) și din Cahul (Republica Moldova). Aplicația PCA clasifică foarte eficient probele de mărar, diferențiind clar originea lor geografică. Mai mult, atunci când proba testată nu aparține nici uneia din clasele modelate, aplicația indică care dintre aceste clase este cea mai similară cu proba necunoscută din punctul de vedere al proprietăților fizico-chimice caracteristice probelor de mărar. Procedura de testare este simplă și rapidă, procesarea și afișarea rezultatului durând în jur de o secundă.

Este remarcabil faptul că aceste performanțe au fost obținute numai pe baza unui număr relativ redus (7) de variabile folosite pentru caracterizarea probelor de mărar. Rezultatele sunt cu atât mai meritorii, cu cât zonele luate în studiu sunt relativ foarte apropiate din punct de vedere geografic. PCA distinge originea probelor de mărar numai pentru zone geografice mult mai îndepărtate, pentru care condițiile pedo-climatice diferă mult mai mult (e.g. județul Brăila și județul Brașov, România). Mai mult, PCA recunoașterea a fost posibilă numai în condițiile în care baza de date de instruire folosită pentru aplicația PCA conține, pe lângă proprietățile fizico-chimice folosite pentru HCA, și o selecție de caracteristici spectrale. Altfel spus, HCA este mai potrivită pentru identificarea originii geografice a probelor decât PCA. HCA însă nu permite o selecție a variabilelor, respectiv nu poate indica care sunt variabilele care au o putere mare de modelare și/sau discriminare. Ca atare, se recomandă efectuarea PCA anterior construirii aplicațiilor HCA.

Studiul a fost extins pentru diverse combinații de zone geografice, testând probe de mărar ce provin din județul Brăila, județul Brașov, județul Vrancea și Cahul. În cazul testării originii luând în considerație concomitent câte două zone geografice, algoritmul folosit conduce la o eficacitate deosebit de bună, inclusiv pentru zone geografice relativ apropiate, în care condițiile pedo-climatice sunt similare. Din acest punct de vedere, HCA bazată de clustering aglomerativ se dovedește a fi o metoda mai performantă în detecția originii geografice a probelor de mărar decât PCA.

Aceeași eficiență deosebită se înregistrează în cazul în testării concomitente a trei din zonele luate în studiu. Totuși, numărul zonelor de origine testate concomitent nu poate fi crescut oricât. În cazul testării concomitente a tuturor celor patru zone modelate, eficiența sistemului de clasificare a scăzut semnificativ. O soluție pentru îmbunătățirea capacității de discriminare ar fi creșterea numărului de probe folosite pentru modelarea claselor de probe de mărar de diverse proveniențe geografice. Acest lucru însă nu este permis de capacitatea de ilustrare grafică a aplicației software folosită pentru determinarea dendrogramelor, i.e. *Matlab 2013a*.

Eficiența discriminării sistemelor dezvoltate prin HCA este cu atât mai apreciabilă cu cât discriminarea originii geografice a probelor de mărar cu ajutorul PCA este eficientă numai pentru zone geografice foarte diferite din punct de vedere pedo-climatic (e.g. probe de mărar din județul Brașov vs. probe din județul Brăila sau Cahul, dar nu și între probe din județul Brăila vs. probe din Cahul sau din județul Vrancea). Ca atare, putem trage concluzia că HCA realizată prin clustering aglomerativ este mai adecvată pentru recunoașterea automată a originii geografice a probelor de mărar decât PCA. Această concluzie este susținută și de faptul că sistemul PCA a fost realizat aplicând un număr mult mai mare de variabile (proprietățile fizico-chimice folosite în acest studiu și proprietățile spectrale) ce caracterizează probele de mărar.

Rezultatele au indicat însă faptul că discriminarea HCA nu poate fi realizată eficient pentru absolut orice combinație de zone geografice. În concluzie, deși este de așteptat ca algoritmul propus să poată fi folosit cu succes și în decelarea altor origini geografice sau a zonei de proveniență și a altor vegetale, combinațiile de zone și numărul zonelor ce pot fi testate concomitent trebuie verificate cu atenție.

Pentru a putea trage concluzii comparative, sistemul HCA pentru identificarea provenienței geografice a probelor de leuștean a fost dezvoltat în același mod ca și în cazul probelor de mărar. Probele de leuștean provin din două regiuni geografice, respectiv județul Brașov și județul Brăila. Rezultatele obținute indică faptul că aplicația multivariată nu este eficientă. Acest rezultat ne indică faptul că proprietățile fizico-chimice ale leușteanului din Brașov și Brăila sunt foarte asemănătoare, informația inclusă în baza de date de instruire fiind insuficientă pentru obținerea unei puteri de discriminare care să asigure o clasificare eficientă. În consecință, nu se poate generaliza o configurație HCA care să funcționeze cu aceeași eficiență în cazul tuturor tipurilor de vegetale.

Pe de altă parte, rezultatele obținute în cazul sistemului PCA de recunoaștere automată a originii probelor de leuștean, prezentate anterior pentru aceleași zone geografice studiate (județul Brașov și județul Brăila), au indicat că acesta poate distinge eficient proveniența acestora. Ca atare, în cazul probelor de leuștean se recomandă folosirea sistemului de clasificare construit cu ajutorul PCA. Trebuie subliniat faptul că sistemul PCA dedicat identificării originii probelor de leuștean a fost dezvoltat cu ajutorul unei baze de date hibride, care conține, atât proprietățile fizico-chimice folosite pentru sistemul HCA cât și absorbțiile din domeniul spectral 250-550 nm. Ori PCA a pus în evidență faptul că, în cazul probelor de leuștean, absorbțiile spectrale din jurul lungimii de undă de 360 nm se numără printre variabilele cu cea mai mare putere de modelare și/sau de discriminare (pe lângă umiditate, conținutul de proteine și grăsimi). În concluzie, folosirea unei baze de date hibride pare deci crucială pentru identificarea originii geografice a probelor de leuștean.

HCA s-a dovedit o metodă automată foarte eficientă și în cazul discriminării genului de vegetală, i.e. a probelor de ciuperci *Pleurotus* și *Champignon*. Numărul mic de probe care sunt necesare pentru modelarea și discriminarea celor două tipuri de ciuperci indică faptul că, în acest caz, se poate dezvolta un sistem automat HCA de discriminare a unui număr important de clase (genuri) de vegetale care pot fi testate concomitent.

Toate rezultatele menționate anterior au fost obținute cu metode de inteligență artificială nesupravegheate. Deși nu au fost incluse date despre originea sau genul probelor în baza de date de instruire a sistemelor de recunoaștere automată, atât PCA, cât și HCA au confirmat fezabilitatea dezvoltării unor sisteme multivariate dedicate identificării originii geografice a produselor vegetale. S-au obținut rezultate bune atât în cazul PCA, care este o metodă de recunoaștere a formelor non-ierarhice, cât și în cazul HCA, care este o metodă ierarhică. Având în vedere că atât PCA cât și HCA sunt metode de recunoaștere a formelor nesupravegheate, s-a continuat studiul cu ANN, care este o metoda de recunoaștere a formelor supravegheată și ierarhică. Bazele de date de instruire ce au fost folosite pentru dezvoltarea sistemelor ANN pentru diversele vegetale conțin următoarele variabile (proprietăți fizico-chimice): umiditate, cenușa, substanță uscată, conținutul de carbohidrați, proteine, grăsimi și de glutamatul monosodic. Sistemele ANN prezentate au fost dezvoltate folosind aplicația software *NnModel* (Neural Network Model).

Pentru construirea bazei de date de instruire și de validare a sistemului ANN dedicat identificării automate a originii probelor de mărar, au fost folosite un număr de 40 de probe, respectiv câte 10 probe din județul Brăila, județul Brașov, județul Vrancea și din Cahul (Republica Moldova). Rezultatele au indicat faptul că atunci când în baza de date de instruire se introduc proprietățile fizico-chimice a câte 9 probe de mărar din fiecare zonă geografică modelată, sistemul ANN este caracterizat de o eficiență (sensitivitate și selectivitate) remarcabilă, originea tuturor probelor de mărar testate fiind corect identificată. Dacă instruirea se realizează introducând în baza de date câte 8 sau 7 probe de mărar din fiecare zonă, sistemul prezintă aceeași eficiență deosebită numai în recunoașterea originii probelor ce provin din zone în care condițiile pedo-climatice diferă semnificativ (e.g. zona de șes vs. zona de deal). În acest caz însă, eficiența în discriminarea probelor ce provin din zone geografice similare este mai scăzută (e.g. județul Brașov și județul Vrancea).

Această comportare este explicabilă, având în vedere faptul că și în cazul dezvoltării sistemelor HCA de recunoaștere a originii probelor de mărar s-au obținut rezultate analoage. Trebuie însă subliniat faptul că sistemul ANN este caracterizat de o rată de clasificare corectă semnificativ mai mare decât cea specifică sistemului HCA. În plus, dezvoltarea sistemului ANN nu este afectată de limitarea întâlnită în cazul sistemelor HCA construite cu ajutorul aplicației software *Matlab*, respectiv a numărului de probe ce pot fi afișate pe abscisa dendrogramei. În cazul sistemelor ANN, atât numărul probelor de mărar folosite pentru modelarea identității lor geografice, cât și numărul zonelor ce pot fi testate concomitent poate fi crescut oricât. De asemenea, numărul probelor necunoscute ce pot fi testate simultan poate fi oricât de mare.

Este important de subliniat din nou faptul că aceste rezultate au fost obținute deși informația inclusă în baza de date de intrare a sistemului PCA a fost mult mai bogată (atât cele 7 proprietăți fizico-chimice cât și o selecție de proprietăți spectrale) decât în cazul sistemelor HCA sau ANN. Pe de altă parte, sistemul HCA este mai performant în recunoașterea originii geografice a probelor de mărar decât sistemul PCA. Se poate trage concluzia că cel mai operativ mod de identificare automată a originii geografice a probelor de mărar este folosirea unui sistem ANN. Însă nu trebuie neglijat faptul că PCA, spre deosebire de HCA și ANN, poate indica variabilele care contribuie cel mai mult la modelarea fiecărei zone geografice, astfel încât cele redundante să poată fi eliminate și astfel să se obțină un sistem de calcul mai suplu și mai ieftin. Altfel spus, ca și în cazul HCA, este util ca înainte de construcția unui sistem ANN să se realizeze o analiză exploratorie prin PCA.

În baza de date de instruire folosită pentru construcția sistemului ANN construit pentru identificarea automată a probelor de **leuștean** au fost incluse aceleași proprietăți fizico-chimice ca și în cazul probelor de mărar, respectiv umiditatea, cenușa, conținutul de substanță uscată, carbohidrați, proteine, grăsimi și de glutamat monosodic. Probele de leuștean analizate provin

din județul Brașov și județul Brăila. Dacă sistemul ANN este dezvoltat incluzând în baza de date de instruire proprietățile fizico-chimice a cel puțin câte 9 probe de leuștean din fiecare din cele două zone geografice modelate, atunci se obține o rată de clasificare corectă foarte bună. În concluzie, metoda propusă poate fi folosită cu succes pentru identificarea originii probelor de leuștean. Mai mult, sistemul poate fi folosit cu succes și pentru studii de analiza senzorială.

Rezultatele obținute cu sistemul ANN dezvoltat pentru identificarea originii probelor de leuștean sunt mult mai bune decât cele obținute cu sistemul HCA. Performanțele sistemului ANN construit cu câte 9 probe de leuștean din fiecare zonă modelată par a fi comparabile, cu cele ale sistemului PCA dezvoltat pentru această vegetală. Sistemele ANN par deci a fi cele mai eficiente pentru identificarea originii unei game largi de vegetale.

Sistemul ANN dezvoltat pentru identificarea și discriminarea ciupercilor *Pleurotus (Pleurotus ostreatus)* și a ciupercilor Champignon (*Agaricus bisporus*) este capabil să recunoască genul probelor de ciuperci cu mare acuratețe, chiar și în cazul în care numărul de probe folosit pentru instruire este relativ scăzut. Rezultatele obținute la validarea sistemului ANN indică faptul că folosirea unui astfel de sistem expert este o modalitate de detecție eficientă și accesibilă (ieftină și ușor de utilizat) pentru screening-ul produselor vegetale.

În concluzie, metodele chemometrice multivariate propuse permit automatizarea identificării originii probelor vegetale studiate pe baza proprietăților spectrale și /sau proprietăților lor fizico-chimice (umiditate, substanță uscată, carbohidrați, proteine, grăsimi, cenușă, glutamat monosodic).

Printre avantajele și performanțele sistemelor automate PCA, HCA sau ANN dezvoltate se numără: automatizarea procesului de detecție și identificare a originii de proveniență a produselor vegetale analizate într-un timp foarte scurt (secunde); operarea detecției cu o sensibilitate și o selectivitate maximă; performanțele sunt obținute numai pe baza unui număr relativ redus de variabile; algoritmi folosiți conduc la o eficacitate deosebit de bună inclusiv pentru zone geografice relativ apropiate, în care condițiile pedo-climatice sunt similare; rezultatele clasificării automate permit orientarea analistului spre analize ulterioare bine definite, rezultând astfel o economie de resurse analitice și umane importante.

VII.2. CONTRIBUȚII ORIGINALE

În studiile realizate în cadrul tezei de doctorat, *Metode avansate de detecție și identificare a originii unor produse vegetale folosite în industria alimentară*, au fost aduse următoarele contribuții personale:

- studiile au fost realizate comparativ, printr-o abordare modernă (chemometrică) și interdisciplinară, focalizate pe optimizarea continuă a sistemelor de detecție obținute;

- realizarea unui amplu set de studii în vederea identificării celor mai eficiente metode multivariate pentru detecția și indentificarea originii geografice a produselor vegetale folosite în industria alimentară;

- au fost evaluate atât metode de recunoaștere a formelor nesupravegheate (PCA și HCA), cât și supravegheate (ANN);

- au fost evaluate atât metode de recunoaștere a formelor non-ierarhice (PCA), cât și ierarhice (HCA, ANN);

- sistemele automate PCA, HCA și ANN au fost evaluate din punctul de vedere al eficienței (senzitivității și selectivității) identificării și detecției mai multor tipuri de vegetale;

- sistemele automate PCA, HCA și ANN au fost evaluate din punctul de vedere al numărului de zone geografice ce pot fi modelate și discriminate eficient;

- sistemele automate PCA, HCA și ANN au fost evaluate din punctul de vedere al numărului de zone geografice ce pot fi testate simultan;

sistemele automate PCA, HCA și ANN au fost evaluate din punctul de vedere al numărului de probe necunoscute ce pot fi testate simultan;

sistemele automate PCA, HCA și ANN au fost evaluate din punctul de vedere al răspunsului dat în cazul unei probe de vegetală ce nu provine din una din zonele geografice modelate.

Cercetarea științifică din teză este desfășurată în cadrul proiectului finanțat de Uniunea Europeană, Proiect POSDRU 132397- Excelență în cercetare prin burse doctorale și postdoctorale (ExcelDOC).

VII.3. PERSPECTIVE VIITOARE DE CERCETARE

Rezultatele obținute deschid calea spre următoarele direcții de cercetare:

- ✚ diversificarea metodelor multivariate utilizate în detecția și identificarea originii produselor vegetale folosite în industria alimentară;

- ✚ testarea metodelor de inteligență artificială care au fost găsite ca fiind cele mai adecvate pentru identificarea automată a originii geografice și pentru alte tipuri de produse vegetale;

- ✚ optimizarea identificării / detecției sistemelor PCA, HCA și ANN din punctul de vedere al selectivității și sensibilității prin îmbogățirea bazelor de date de instruire utilizate de sistemele de detecție.

LISTĂ LUCRĂRI PUBLICATE ȘI PREZENTATE

Articole publicate în reviste de specialitate cotate ISI

1. S. Constantin Ghinita, M. Praisler, G. Iordăchescu „Automatic detection of the geographical origin of dill (*Anethum graveolens*) based on hierarchical cluster analysis”, depus spre publicare în ***Environmental Engineering and Management Journal*** (revista cotată ISI, factor de impact 1,117 pentru 2013)

Articole publicate în reviste incluse în baze de date internaționale cotate CNSIS (B+)

1. S. Constantin, M. Praisler, G. Iordăchescu, *Chemometrical method designed for the automatic identification of mushrooms*, Annals of "Dunarea de Jos" University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II, Year V (XXXVI) 2013, No. 1-2, p. 32 – 39.
ISSN 2067-2071, (revista cotată B+, indexată în: Cambridge Scientific Abstracts (CSA), EBSCO)
2. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, "Agglomerative clustering system designed for the detection of the geographical origin of dill", Annals of "Dunarea de Jos" University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II, Year VI (XXXVII) 2014, No. 1, p. 80 – 85.
ISSN 2067-2071, http://www.phys.ugal.ro/Annals_Fascicle_2 (revista cotată B+, indexată în: Cambridge Scientific Abstracts (CSA), EBSCO).
3. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, "Automatic identification of vegetables based on hierarchical cluster analysis: mushrooms" Annals of "Dunarea de Jos" University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II, Year VI (XXXVII) 2014, No. 1, p. 86 – 91.
ISSN 2067-2071, http://www.phys.ugal.ro/Annals_Fascicle_2 (revista cotată B+, indexată în: Cambridge Scientific Abstracts (CSA), EBSCO)
4. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, „Identification of the origin of vegetables based on Artificial Neural Networks”, Annals of "Dunarea de Jos" University of Galati, Mathematics, Physics, Theoretical Mechanics, Fascicle II, Year VI (XXXVII) 2014, No. 1, p. 110 – 115.
ISSN 2067-2071, http://www.phys.ugal.ro/Annals_Fascicle_2 (revista cotată B+, indexată în: Cambridge Scientific Abstracts (CSA), EBSCO)
5. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, „Identification of the Main Physico-Chemical Properties Influencing the Quality of Thyme (*Satureja Hortensis*)”, Journal of Engineering Studies and Research, vol. 20, no. 2, 2014, 1-6.
ISSN 2068-7559. <http://pubs.ub.ro/?pg=revues&rev=jesr> (revista indexată în: ProQuest CSA, VINITI, EBSCO, Index Copernicus, Academic Journals Database, Directory of Research Journal Indexing, Directory of Open Access Journals DOAJ)
6. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, "PCA identification of the main physico-chemical properties influencing the quality of pea beans (*Pisum sativum*)", depusă spre publicare la Journal of Engineering Studies and Research, <http://pubs.ub.ro/?pg=revues&rev=jesr> (revista indexată în: ProQuest CSA, VINITI,

Participari la conferințe internaționale

1. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, *Identification of the origin of vegetable products: dill (Anethum graveolens)*, Proceedings of the 6th International Conference of Management and Industrial Engineering, ICMIE 2013, October 31 – November 2, 2013, Bucharest, Romania, pp. 173-180, ISSN-L 2344-0937, ISSN 2344-0937. <http://www.icmie.ro/>
2. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, *Identification of the origin of vegetable products: lovage (Levisticum officinale)*, Proceedings of the 6th International Conference of Management and Industrial Engineering, ICMIE 2013, October 31 – November 2, 2013, Bucharest, Romania, pp. 245-252, ISSN-L 2344-0937, ISSN 2344-0937 <http://www.icmie.ro/>
3. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, *PCA identification of the main physico-chemical properties influencing the quality of pea beans (Pisum sativum)*, Abstract book of the Second International Conference on Natural and Anthropogenic Risks - ICNAR 2014, 4 - 7 June, 2014, Bacau, Romania, pp.122, ISSN 2360-4018 <http://icnar.ub.ro/>
4. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, *Automatic detection of the geographical origin of dill (Anethum graveolens) based on hierarchical cluster analysis*, Abstract book of the Second International Conference on Natural and Anthropogenic Risks - ICNAR 2014, 4 - 7 June, 2014, Bacau, Romania, pp.123, ISSN 2360-4018 <http://icnar.ub.ro/>

Participari la conferințe naționale

1. S. Constantin, M. Praisler, G. Iordăchescu, *Chemometrical method designed for the automatic identification of mushrooms*, Book of abstracts, CSSD-UDJG 2013, First Edition, 16-17 May 2013, Galati, Romania.
2. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, *„Agglomerative clustering system designed for the detection of the geographical origin of dill”*, Book of abstracts, CSSD-UDJG 2014, Second Edition, 15-16 May 2014, Galati, Romania. <http://www.cssd-udjg.ugal.ro/index.php?edition=2014&lang=ro>
3. S. Constantin Ghinita, M. Praisler, G. Iordăchescu, *„Automatic identification of vegetables based on hierarchical cluster analysis: mushrooms”*, Book of abstracts, CSSD-UDJG 2014, Second Edition, 15-16 May 2014, Galati, Romania. <http://www.cssd-udjg.ugal.ro/index.php?edition=2014&lang=ro>

FONDUL SOCIAL EUROPEAN

Programul Operational Sectorial pentru Dezvoltarea Resurselor Umane 2007-2013

INVESTEȘTE IN OAMENI !

Titlul proiectului : “Exelență in cercetare prin burse doctorale și postdoctorale - ExcelDoc”

**Axa prioritară 1 : “Educație și formare profesională in sprijinul creșterii economice și dezvoltării
societății bazate pe cunoaștere”**

Domeniul major de intervenție 1.5. : Programe doctorale și postdoctorale in sprijinul cercetării

Numărul de identificare al contractului : POSDRU/159/1.5/S/132397

Editor : Universitatea “DUNAREA DE JOS” din Galați

Data publicării : 2014

www.fonduri-ue.ro

www.fonduri-ue.ro

**Conținutul acestui material nu reprezintă in mod obligatoriu poziția oficială a Uniunii Europene
sau a Guvernului României**