



**UNIVERSITATEA POLITEHNICA
DIN BUCUREȘTI**



**Școala Doctorală de Electronică, Telecomunicații și
Tehnologia Informației**

Decizie nr. 961 din 16-11-2022

**REZUMAT TEZĂ
DE DOCTORAT**

Ing. Mihai BOLDEANU

**AUTOMATIC POLLEN CLASSIFICATION USING DEEP
LEARNING TECHNIQUES**

**CLASIFICARE AUTOMATĂ A POLENULUI FOLOSIND TEHNICI
DE INVĂȚARE AUTOMATĂ**

COMISIA DE DOCTORAT

Prof. Dr. Ing. Gheorghe BREZEANU Univ. Politehnica din București	Președinte
Prof. Dr. Ing. Corneliu BURILEANU Univ. Politehnica din București	Conducător de doctorat
Prof. Dr. Ing. Corneliu RUSU Univ. Politehnica din București	Referent
CS II Dr. Luminița MĂRMUREANU Institutul Național de Cercetare- Dezvoltare pentru Optoelectronică - INOE 2000	Referent
Conf. Dr. Ing. Horia CUCU Univ. Politehnica din București	Referent

BUCUREȘTI 2022

Cuprins

1	Introducere	1
1.1	Impactul polenului asupra calității aerului și a sănătății	1
1.2	Descrierea problemei	2
1.3	Motivația tezei	2
1.4	Scopul și obiectivele tezei	3
1.5	Structura tezei	4
2	Revizuire de literatură	5
2.1	Palinologie	5
2.2	Instrumente de monitorizare a polenului	5
2.3	Seturi de date cu polen	7
2.4	Clasificare automată a polenului	8
3	Cadru teoretic	9
3.1	Viziune computerizată	9
3.2	Algoritmi de clasificare	10
3.3	Algoritmi de segmentare a imaginilor	11
3.4	Căutare de hiper-parametri	12
3.5	Selectia, procesarea si augmentarea datelor	13
4	Utilizarea imaginilor microscopice in clasificarea de polen	14
4.1	Descrierea instrumentului	14
4.2	Crearea de seturi de date cu polen	15
4.3	Selecția arhitecturii	15
4.4	Rezultate	15
5	Date multi-modale de florescență si imprastiere utilizate in clasificarea polenului	18
5.1	Descrierea instrumentului	18
5.2	Crearea de seturi de date cu polen	19
5.3	Selecția arhitecturii	19
5.4	Rezultate	20

6	Modelarea si predicția concentrației de polen	22
6.1	Impactul umidității relative asupra sezonului de polen	23
6.2	Efectul temperaturii asupra sezonului de polen	23
6.3	Ploaia înghețată și impactul asupra vegetației	24
6.4	Zăpadă portocalie și transportul particulelor pe distanțe mari	24
7	Concluzii	25
7.1	Obiective și Rezultate	25
7.2	Contribuții originale	26
7.3	Lista cu articole publicate	27
7.4	Cercetări viitoare	28
	Bibliografie	30

Capitolul 1

Introducere

În această teză este abordată tema utilizarea învățării automate/ML pentru a dezvolta modele numerice capabile să facă clasificare automată de polen. Acest lucru este realizat folosind date ce provin de la analizoare de aerosoli automate și diverși algoritmi de învățare automată sau de învățare adâncă. Această tematică are multe aplicații posibile în domenii precum: agricultură, sănătate și monitorizarea schimbărilor climatice.

Polenul este o parte esențială a sistemului reproductiv al plantelor. El este compus din grăunțe fine, ce au forme și dimensiuni foarte variate în funcție de specia plantei, produse de structurile masculine ale plantelor. Această pudră este adesea transportată de insecte sau de vânt și odată ajunsă pe alte plante interacționează cu părțile feminine ale plantei și se produce fertilizarea.

Diferențele între dimensiunea și forma particulelor de polen este suficient de mare încât un expert poate identifica specia de plantă de la care provine polenul. Aproape toate angiospermele și gimnospermele pot fi identificate pe baza polenului, această arie de studiu este denumită palinologie și se referă la studiul polenului și al sporilor.

Acest capitol este organizat în următorul mod: secțiunea 1.1 prezintă impactul polenului asupra calității vieții umane; în secțiunea 1.2 sunt prezentate principalele obstacole în dezvoltarea de sisteme capabile să monitorizeze polen; secțiunea 1.3 prezintă motivația pentru această teză; secțiunea 1.4 subliniază scopul și obiectivele acestei teze și în final secțiunea 1.5 face o trecere în revistă a întregii lucrări.

1.1 Impactul polenului asupra calității aerului și a sănătății

Polenul este principalul vinovat în declanșarea alergiilor sezoniere la oameni. Popular, alergiile la polen este denumită "febra fânului" și oficial este denumită "rinită alergică sezonieră". În fiecare primăvară, vară și toamnă, plantele eliberează milioane de granule de polen pentru a își perpetua specia. Deși majoritatea polenului nu provoacă probleme, fiind transportat preponderent de insecte, există anumite specii de copaci, ierburi și buruieni ce se bazează pe vânt pentru transportul polenului. Aceste plante produc un

polen de dimensiuni mici, mai ușor și mai uscat pentru a fi ridicat de adierea vântului. Acest tip de polen ajunge în ochi, nas și plămâni și cauzează reacții alergice la cei sensibili.

În perioada recentă s-a observat o creștere a numărului de cazuri diagnosticate de alergii la polen. În Europa unele studii sugerează că numărul de persoane alergice se va dubla, de la 33 de milioane cât sunt în prezent, până la 77 de milioane în perioada 2040-2060 [30]. Deși încă nu este înțeles complet motivul acestei creșteri, anumite studii îl leagă de schimbările climatice [30] sau de gradul ridicat de poluare a aerului [52].

Ce este clar, este nevoia monitorizării mai atente a concentrațiilor de polen, în special a tipurilor de polen ce provoacă alergii.

1.2 Descrierea problemei

Există un număr de provocări ce fac dificilă monitorizarea polenului folosind echipamentele și tehnicile curente.

Prima problemă este legată de concentrația scăzută pe care polenul o are în atmosferă. Adesea concentrația poate fi de ordinul zecilor de particule pe metru cub. Este nevoie de eșantionarea unei cantități uriașe de aer pentru a surprinde momentul de început al sezonului de polen.

A doua problemă vine din relația complexă între apariția polenului și alți parametri atmosferici. Dacă o relație matematică ar fi identificată între apariția polenului și temperatură sau umiditate, monitorizarea pe suprafețe mari s-ar putea face urmărind acești parametri.

Ultima problemă majoră, în modul curent de monitorizare a polenului, este lipsa de automatizare. Majoritatea instrumentelor, ce pot fi folosite pentru monitorizarea polenului, se bazează pe tehnicieni atât pentru mentenanță cât și pentru numărarea efectivă a particulelor de polen. Acest aspect limitează zonele unde pot fi amplasate capcane de polen în orașe și zonele limitrofe.

Deși această teză nu poate rezolva toate dificultățile asociate monitorizării de polen, cu siguranță va ajuta prin prezentarea unor metode de standardizare și automatizare a colectării de date și a numărării particulelor de polen.

1.3 Motivația tezei

Plecând de la secțiunile anterioare, în care au fost discutate atât problemele în monitorizarea polenului cât și impactul asupra vieții umane, motivația tezei devine ușor de văzut. Dezvoltarea de sisteme automate, bazate pe inteligență artificială, capabile să detecteze, să identifice și să clasifice polen ar avea un impact pozitiv semnificativ atât pentru persoanele cu alergii dar și pentru palinologie în general prin o mai bună înțelegere a unui fenomen complex.

Dezvoltarea unor sisteme sau rețele, complet automate, de monitorizare a polenului, ar permite achiziția de date într-un mod ce ar permite atât alertarea rapidă a oamenilor cât și construirea de seturi de date standardizate, ce permit construirea modelelor de predicție. Aceste seturi de date ar fi mult mai complete față de ce este momentan înregistrat și ar oferi o oportunitate uriașă palinologiei. A doua motivație pentru această teză pleacă de la această oportunitate de a folosi sisteme de învățare automată pentru a extrage noi informații despre polen din seturi mari de date.

În timpul investigației, modului în care se face clasificarea polenului în momentul de față, am fost surprins de cât de mult este pus accentul pe oameni pentru clasificarea și numărarea particulelor de polen, folosind microscopul. Acest lucru m-a motivat să încerc să găsesc cât mai multe arii în care învățarea automată ar avea un impact, cum ar fi în achiziția datelor, în construirea seturilor de date, în identificare de arhitecturi potrivite pentru date complexe, dar și în găsirea de tehnici de augmentare a datelor și utilizarea lor eficientă pentru construirea și antrenarea modelelor numerice de clasificare.

1.4 Scopul și obiectivele tezei

Obiectivele acestei lucrări sunt:

Identificarea de instrumente științifice complet automate ce ar putea fi folosite pentru monitorizarea polenului. Acest lucru implică compararea diverselor sisteme existente cu scopul de a selecta câteva ce produc date suficient de calitative cu intervenție minimă din partea unui operator uman.

Identificarea sau crearea de seturi de date cu polen. Aceste seturi de date ar trebui să provină, ideal, de la cât mai multe instrumente de același tip. Datele ar trebui să fie standardizate și curățate pentru a permite utilizarea unei baterii cât mai largi de modele numerice pentru clasificarea de polen.

Dezvoltarea de metode de prelucrare a tuturor tipurilor de date ce pot fi folosite în caracterizarea polenului. Găsirea de metode ce pot folosi date ne-structurate (imagini din microscop, imagine de împrăștiere, spectre de fluorescență) pentru identificarea și clasificarea, până la nivel de specie. Aceste metode implică utilizarea de tehnici de feature selection și feature engineering.

Construirea și antrenarea unor modele numerice ce pot clasifica polenul, cu un grad de acuratețe cel puțin similar cu cel al experților umani. Deși acest obiectiv este ușor ambiguu, adesea în utilizarea de tehnici de învățare automată atingerea unei performanțe similare cu aceea a omului este un țel inițial, cu ideea de a îl depăși în stadiile ulterioare.

Ultimul obiectiv al tezei constă în construirea unor biblioteci open-source în limbajul de programare Python ce ar permite o explorare și exploatare mult mai ușoară a metodelor selectate pe seturi noi de date.

1.5 Structura tezei

Teza are următoarea structură:

Capitolul doi este o revizie de literatură pentru studiul polenului în general, cu o privire aprofundată asupra seturilor de date de polen existente, a echipamentelor folosite în monitorizarea polenului și a istoriei tuturor încercărilor de clasificare automată de polen.

Capitolul trei prezintă cadrul teoretic al învățării automate și algoritmi folosiți în această lucrare. Acest capitol este structurat după principalele provocări ce trebuie rezolvate pentru a construi sisteme automate de monitorizare a polenului.

Capitolul patru este un studiu de caz orientat pe utilizarea de date de la un analizor de aerosoli automat bazat pe imagini. Acest capitol investighează utilizarea fotografierii microscopice pentru a captura informații despre polen și prezintă multiple abordări de a clasifica imagini cu polen.

Capitolul cinci este un studiu de caz orientat pe utilizarea de date de la un analizor de aerosoli automat bazat pe spectre de fluorescență. În acest capitol sunt utilizate seturi de date multi-modale pentru a face clasificare de polen.

Capitolul șase prezintă modul în care seturi mari de date, cu alți parametri atmosferici, pot fi utile în a face previziuni privind sezonul de polen.

În final, în capitolul șapte sunt trase concluziile și este prezentată o listă cu toate publicațiile rezultate în cadrul acestui doctorat.

Capitolul 2

Revizuire de literatură

Acest capitol abordează mai multe aspecte legate de clasificarea și monitorizarea polenului în general, printre care o descriere a palinologiei, o analiză a echipamentelor folosite pentru monitorizarea polenului, o punere în revistă a seturilor de date existente și în final o privire retrospectivă asupra tuturor încercărilor de clasificare automată de polen folosind tehnici de învățare automată.

Acest capitol este structurat după cum urmează: În secțiunea 2.1 este făcută o scurtă introducere în palinologie. Aceasta este urmată de o descriere a instrumentației folosite în analiza polenului în secțiunea 2.2. În secțiunea 2.3 sunt menționate seturile de date de polen disponibile public și în final în secțiunea 2.4 un istoric al tentativelor de clasificare de polen este prezentat.

2.1 Palinologie

Palinologia este știința ce se ocupă cu studiul particulelor mici, fie ele organice sau inorganice. Numele vine din limba greacă și se referă la studiul prafului. Metodele clasice folosite în palinologie se bazează pe achiziția de eșantioane din aer, din apă sau din sedimente de către un expert uman. Aceste eșantioane sunt analizate pentru a identifica indicii despre formele de viață, mediul și condițiile energetice ce le-au creat. Palinologia este o știință multi-disciplinară ce stă la intersecția științelor pământului (geologie) și științelor vieții (biologie). Odată cu creșterea interesului în utilizarea unor abordări automate sau asistate de calculator, palinologia începe să preia tehnici și din știința calculatorului și din domeniul învățării automate.

2.2 Instrumente de monitorizare a polenului

Instrumentația folosită în monitorizarea polenului își poate găsi originea în experimentele inițiale ale lui Blackley în secolul XIX [58]. Blackley suferea de ceea ce la vremea aceea

se numea febra fânului. El a început să investigheze cauza condiției sale în anul 1859 și primele rezultate au fost publicate în anul 1873 [58].

Primele tentative în capturarea de polen se bazau pe utilizarea unor lame de microscop acoperite cu glicerină. Aceste lamele erau expuse la aer și după o perioadă fixă de timp erau aduse sub lentila microscopului și analizate. Analiza constă în identificarea speciilor de plante ce au produs acel polen și numărarea tuturor particulelor. Această metodologie este foarte apropiată de modul de monitorizare a polenului în prezent, deși echipamentele sunt ușor mai complexe.

Prima capcană de polen care a adus standardizare este cea introdusă de Hirst în lucrarea "An Automatic Volumetric Spore Trap" în anul 1952[24].

Capcana lui Hirst [24] este descrisă ca un dispozitiv ce aspiră aer atmosferic printr-un orificiu, orientat în direcția din care bate vântul. Acest flux de aer este impactat de niște lamele acoperite cu Vaseline. Aceste lamele sunt mișcate cu o viteză de 2 milimetri pe oră, astfel diferite părți ale lamelei sunt expuse la momente diferite din zi. Acest sistem permite o mult mai bună eficiență în capturarea de particule datorită pompei de aer ce eșantionează volume mult mai mari de aer decât este posibil prin simpla expunere a lamelelor.

O îmbunătățire asupra capcanei lui Hirst a fost adusă de Kramer-Collins [28] prin prelungirea perioadei în care o capcană poate funcționa fără intervenție umană.

Capcanele Burkard [42] sunt o variantă mai recentă de instrument. În acest tip de dispozitiv aerul este aspirat printr-un orificiu la un flux de 10 litri pe minut și impactat pe o bandă acoperită cu adeziv. Acest mod de operare permite o mai bună rezoluție temporală.

O altă variantă recentă a capcanelor de polen este VPPS construit de Lanzoni. Acest tip de dispozitiv este foarte versatil datorită posibilității operării pentru perioade lungi de timp fără intervenție umană.

Capcanele Durham [45] reprezintă o abordare diferită de capturarea de polen. Ele se bazează pe metoda prezentată de Erdtman [20] și folosesc două discuri de plexiglas de aproximativ 22.5 cm în diametru. Între aceste discuri sunt plasate lamele de microscop acoperite cu Vaseline ca adeziv. Acest echipament se bazează pe vânt pentru a captura particule de polen și nu are părți în mișcare.

Din gama instrumentelor clasice de capturare de polen avem și capcana Rotorod [61]. Acest tip de dispozitiv folosește două brațe din plastic acoperite în adeziv ce sunt rotite la mare viteză în aer. Modul de funcționare este detaliat în [40]. GRIPST-2009 este un tip de capcană ce utilizează conceptul Rotorod. Aceste echipamente s-au dovedit foarte eficiente în capturarea de particule, cu dimensiunea până la 2 micrometri.

Mai departe vom continua cu echipamente ce au un mod de funcționare automat sau aproape automat.

Automatic-KH-3000 [55] este un numărător de particule construit cu scopul de a monitoriza polen folosind informația capturată prin împrăștierea unei raze laser. KH-

3000 este construit pentru a captura polen într-o coloană de aer. Această coloană de aer este ulterior trecută printr-o rază laser și impulsurile generate prin împrăștierea luminii laser sunt folosite pentru a număra particulele de polen.

Rapid-e de la Plair este un instrument ce poate analiza aerosoli, particulă cu particulă, în timp real. Complet automat, acest dispozitiv poate caracteriza particule cu dimensiuni între 0.5 și 100 de micrometri.

BAA-500 este un instrument ce funcționează similar cu capcanele manuale doar că își propune să automatizeze complet procesul. Dispozitivul face automat capturarea polenului, pregătirea eșantioanelor, detecția optică cu un microscop și o parte din procesarea digitală de imagine.

Swisens de la Poleno este un sistem complet automat de monitorizare a polenului construit pentru studii de lungă durată [49].

APS-300, dezvoltat de Pollen Sense LLC, este un sistem de imaging automat ce colectează particule de aerosoli, în timp real. Dispozitivul impactează aer încărcat cu aerosoli pe o bandă acoperită cu adeziv după care folosește un sistem automat de fotografiere pentru a obține imagini precise cu particulele capturate.

2.3 Seturi de date cu polen

Mai jos avem o listă cu seturile de date cu polen obținute din echipamente manuale:

- POLLEN13K [7].
- POLEN23E [23].
- POLLEN73S [5].
- POLLEN20L-det .[26]
- Artemisia pollen dataset [36].
- Cretan Pollen Dataset [56].
- Classifynder 46 [53].
- ABCPollen [29].

Următoarea listă prezintă seturile de date obținute pe echipamentul automat BAA-500:

- Data-set-15 [51]
- Data-set-31 [50]
- BAA-500 cropped dataset [?].

- Alternaria segmentation dataset[Citation needed].

În Tabelul 2.1 avem prezentate seturile de date publice obținute cu echipamentul Rapid-E:

Dataset Name	Number of Classes	Number of Samples	Minimum Number Samples per Class
SAU-SRB ^a	14	85 k	985
SAU-LI ^a	11	399 k	16,114
SAU-CH ^a	10	50 k	1,075
MARS ^b	13	105 k	3,020

Tabel 2.1 Overview of available datasets from Rapid-E devices. ^a from [57]; ^b from [9].

1

2.4 Clasificare automată a polenului

Această secțiune adună toate încercările de clasificare automată de polen. Toate aceste informații sunt sumarizate în Tabelul 2.2.

Classifier	Features used	Pollen taxa	Reported Accuracy	Reference
C-SVC (SVM), CST+BOW	BOW, color, shape and texture	23	64%	[23]
Linear discriminant classifier, CNN	Automatic feature extraction	23	97%	[53]
Multivariate statistical classification	Texture analysis	6	94%	[31]
NN, leave-one-out classifier	Texture features	4	100%	[32]
Mahalanobis distance	Color, shape features	30	77%	[12]
Mahalanobis distance	Color, shape, geometric features	30	88.25%	[11]
Regression trees	Morphometric features	3	n.a.	[35]
SVM	Invariant gray-scale features (3D)	26	92%	[48]
Linear discriminant analysis	Texture/shape	13, 4	100%	[33]
MDC	Fourier descriptors	3	90%	[47]
MDC (majority voting)	Texture features (FSM)	5	85%, 87.4%	[13]
Nearest Neighbor, SVM	Group integration	26, 7	96.9%, 99.7% (SVM)	[44]
Linear normal classifier	Shape, stat. gray-level, pore/colpus	3	97.20%	[15]
SVM/MDC/MLP	Texture and shape features	3	89%	[46]
Bayesian classifier	Invariant features (local jets)	3	83%	[41]
Adaptive Bayesian Combination, LLC	Texture features, LLT	7	90.58%	[63]
MLP	Geometric, shape, texture features	3	90%	[1]
SVM	3D discrete spherical features	26, 33	96.3%, 91.8%	[62]
Mahalanobis distance	Color, texture, optical spatial frequency	3, 40	77%	[25]
KNN, Gaussian, SVDD	Morphological, shape, textural, color	5	92.30%	[16]
Linear discriminant analysis	Morphological, statistical, three space-frequency	15	99.40%	[43]
CNN	Automatic feature extraction	30	90%	[18]
SVM, Random Forest, Logistic regression	Color, texture	23	79%	[3]
CNN+RNN	Automatic feature extraction	10	100%	[19]
CNN	Automatic feature extraction	5, 11	99.8%, 95.9%	[27]
CNN	Automatic feature extraction	11	99.75%	[21]
SIFT	Local key points (3D)	27, 33, 28	88.25%	[59]
Random Forest	Geometric, textural	6	88.24%	[37]
NN + CNN	Flourescence Spectroscopy	11,13,14	74%,77%,80%	[57]
CNN + LDC	Automatic feature extraction	46	97.86%	[53]
NN + CNN	Flourescence Spectroscopy	11,13,14	77%,80%,84%	[9]
RBF SVM, CNN	LBP, HOG, Automatic feature extraction	4	87%, 90%	[7]

Tabel 2.2 Table of the automatic pollen classification attempts.

Capitolul 3

Cadru teoretic

Acest capitol prezintă toți algoritmi și toate metodologiile de procesare de date folosite, în cadrul acestei teze, și are rolul de glosar pentru acestea. Deși descrierile algoritmilor în acest capitol sunt generice, pe alocuri apar completări pentru a detalia cum se aplică în contextul clasificării de polen.

În secțiunea 3.1 este prezentată tematica de Computer Vision cu obiective și evoluție. În secțiunea 3.2 sunt prezentați algoritmi de clasificare. Secțiunea 3.3 este despre proceduri de segmentare de imagine, necesare în detecția și identificarea polenului din imagini. Secțiunea 3.4 prezintă metodologii de îmbunătățire a performanței algoritmilor de clasificare. În final, în secțiunea 3.5 atenția este concentrată pe metodologii de procesare a datelor cu scopul de a le face mai ușor de folosit în antrenarea modelelor numerice.

3.1 Viziune computerizată

Când vorbim despre Computer Vision, ne referim la abilitatea unui calculator sau a unui sistem computațional de a extrage informații din imagini digitale, din fluxuri video sau din alte tipuri de date nestructurate cum ar fi date obținute din sisteme LIDAR, RADAR sau SODAR. Din perspectiva inginerescă, Computer Vision are ca scop imitarea capabilităților sistemului vizual uman de către un computer.

Câteva dintre sarcinile studiate în domeniul de Computer Vision sunt:

- Recunoașterea de obiecte sau clasificarea - se referă la analizarea unei imagini sau a unui flux video cu scopul de a găsi obiecte dintr-o listă predefinită de clase. Această sarcină poate fi extinsă și la identificarea locației obiectului în imagine.
- Identificarea - este sarcina de a identifica o instanță dintr-o clasă. Ca exemplu putem vorbi de identificarea feței unei persoane pentru a permite accesul.

- Detectia - se referă la analiza unei imagini pentru depistarea îndeplinirii unor condiții învățate. Ca exemplu avem detectarea țesutului anormal în imaging medical sau detectarea vehiculelor la un sistem automat de taxare pe autostradă.

În următoarea secțiune vom prezenta câțiva algoritmi ce au fost folosiți pentru clasificare de polen.

3.2 Algoritmi de clasificare

Această secțiune investighează algoritmi de clasificare, cu accent pe cei care au fost utilizați în această lucrare.

În prima parte sunt prezentați algoritmi clasici de învățare automată după care în a doua parte sunt prezentate și câteva arhitecturi de învățare adâncă.

Analiza discriminantă quadratică sau analiza discriminantă normală este o metodă folosită în statistică și în alte domenii pentru găsirea de combinații liniare de caracteristici, ce pot fi folosite pentru separarea unor obiecte în clase. Combinația rezultată poate fi apoi folosită ca un clasificator liniar.

Analiza discriminantă quadratică este o metodă similară cu analiza discriminantă quadratică, diferența constă în renunțarea la presupunerea că măsurătorile fiecărei clase au o distribuție normală. În QDA nu se consideră că matricile de covarianța sunt identice între clase.

Support vector machines sunt un tip de model de clasificare supervizat ce poate fi folosit atât pentru clasificare, cât și pentru regresii. SVMs sunt o metodă robustă de predicție ce se bazează pe identificarea unui hiper-plan ce separă elementele claselor diferite. Acesta clasă de algoritmi funcționează plasând exemplele de antrenare într-un spațiu înalt dimensional în așa fel încât distanța dintre clase este maximizată.

Clasificatori Bayes Naiv sunt o familie de modele probabilistice bazate integral pe teorema lui Bayes cu presupunerea că variabilele de input sunt independente între ele și normal distribuite.

Arborii de decizie sunt un tip de model foarte populari în multe arii de cercetare. Această abordare folosește o structură în formă de arbore ce poate fi folosită pentru a ajunge de la observații la niște clase predefinite. În această structură de arbore, frunzele reprezintă etichetele de clasă și ramurile reprezintă combinația de caracteristici care duc către acele clase.

Random forest sau păduri aleatoare sunt un tip de model obținut prin construirea mai multor arbori de decizie. Acest tip de model este adesea numit un meta-model deoarece el este compus dintr-un număr mare de modele mai simple. Acest tip de meta-model poate fi folosit atât pentru regresie cât și pentru clasificare. În cazul clasificării răspunsul pădurii este clasa selectată de majoritatea arborilor de decizie. Avantajul unui astfel de meta-model constă în faptul că evită problema overfitting-ului ce este adesea întâlnită la arbori de decizie.

Perceptronul este un algoritm de învățare supervizată ce poate fi folosit atât pentru clasificare cât și pentru regresii. Acest tip de model are ca sarcină găsirea unui prag între două clase folosind o combinație liniară a caracteristicilor de input. Perceptronul este cea mai simplă variantă a unei rețele neurale.

Între algoritmi clasici de machine learning și cei de deep learning nu există o diferență clară. De obicei, diferențierea se face pe baza capacității unui model de a învăța pe seturi de date foarte mari. Performanța modelelor numerice bazate pe algoritmi clasici nu se scalează odată cu creșterea seturilor de date de antrenare peste un anumit prag, pe de altă parte modelele numerice de tip deep learning sunt capabile să își îmbunătățească continuu performanța cu creșterea volumului de date.

Apropierea dintre cele două tipuri de algoritmi este evidentă în cazul rețelelor neurale. Perceptronul multi-strat (MLP) este o extensie a perceptronului clasic. Diferența dintre cele două abordări constă în neliniaritatea introdusă în MLP prin folosirea unor funcții de activare. De asemenea, MLP are unul sau mai multe straturi ascunse între stratul de intrare și cel de ieșire.

MLP este o familie de rețele neurale artificiale de tip feed-forward. Acest tip de model are conexiunile dintre straturi dense sau complet conectate, practic fiecare neuron dintr-un strat este influențat de toți neuronii din stratul anterior. MLP este feed-forward, pentru că în timpul folosirii datele "curg" de la stratul de intrare către stratul de ieșire fără a avea conexiuni recursive, straturile de la final nu influențează straturile de la începutul rețelei. MLP este adesea numele folosit generic pentru orice rețea neurală artificială, dar definiția strictă se referă doar la acele modele construite prin cascada mai multor perceptroni ce folosesc de asemenea o funcție de activare liniară sau neliniară.

Când vine vorba de utilizarea rețelelor neurale în rezolvarea de sarcini de Computer vision, algoritmi de tip MLP își arată limitările. Pentru astfel de sarcini rețelele neurale convoluționale (CNN) sunt de obicei varianta mai utilă. CNN-urile au fost construite în mod particular în abordarea datelor nestructurate cum sunt imaginile, fluxurile video sau audio sau alte tipuri de date 2D/3D. Avantajul principal al acestui tip de rețea este o arhitectură bazată pe reutilizarea valorilor neuronilor. Acest lucru fiind realizat prin utilizarea unor kerneli de convoluție ce sunt aplicați peste datele de intrare într-un mod succesiv. Aceste filtre digitale sunt construite în timpul antrenării modelului.

3.3 Algoritmi de segmentare a imaginilor

În această secțiune sunt prezentați algoritmi utilizați în segmentarea de imagini. Când vorbim de segmentarea unor date 2D sau 3D ne referim la trei posibilități în funcție de ce informații sunt extrase din date. Segmentarea semantică se referă la identificarea tuturor pixelilor dintr-o imagine care fac parte dintr-un obiect, ce aparține unui set de clase predefinite. Ca exemplu putem avea un algoritm ce identifică toți pixelii ce reprezintă cerul sau pământul într-o imagine, sau toți pixelii ce reprezintă vegetație într-o imagine

din satelit. Segmentarea de instanță se referă la identificarea tuturor pixelilor unei instanțe obiect. O astfel de segmentare diferențiază între mai multe instanțe de obiecte, dar nu diferențiază după clasă. Ultimul tip de segmentare se referă la segmentarea panoptică. Acest mod de lucru combină celelalte moduri de segmentare, astfel se obțin informații despre fiecare instanță, dar și despre clasa căreia îi aparține acea instanță.

Una dintre cele mai simple metode de segmentare de imagine constă în alegerea unor praguri pentru diferențierea pixelilor ce formează o imagine. Cu astfel de metode se pot transforma imagini cu nuanțe de gri în imagini alb negru sau în imagini binare. Partea dificilă într-o astfel de abordare este alegerea pragului corect pentru o anumită sarcină. Câteva din metodele utilizate în selectarea valorii de prag sunt metoda entropiei maxime, metoda histogramei echilibrate, metoda maximizării varianței și chiar metoda de clusterizare k-means.

Detectarea muchiilor este un alt tip de segmentare bine-cunoscut în procesarea digitală de imagine. Obiectele, în imagini, sunt adesea mărginite de muchii foarte pronunțate, cu un contrast puternic. Această diferență în intensitate, la marginea dintre obiecte, poate fi folosită ca o metodă de segmentare a imaginilor. Detectarea muchiilor este de obicei primul pas în algoritmi mai complecși de segmentare ce folosesc multiple operații morfologice asupra unei imagini pentru a obține segmentări bune.

Etichetarea componentelor conectate este un tip de algoritm ce este util în segmentarea de instanță a unor imagini. Pentru a putea folosi un astfel de algoritm, imaginile sunt întâi pre-procesate pentru a obține imagini binare. După acest pas etichetarea componentelor conectate este folosită pentru a separa diverse instanțe ale celor două clase ale măștii binare.

A doua categorie de algoritmi de segmentare sunt aceia bazați pe utilizarea de rețele neuronale convoluționale. Avantajul major al unei astfel de abordări este că poate fi aplicat pe imagini fără a fi necesară o parte de pre-procesare a imaginii. Practic modelul numeric învățat care este cea mai bună combinație de praguri valorice și de filtre de convoluție necesare pentru a segmenta o imagine.

Cele mai folosite arhitecturi sunt cele de tipul rețea complet convoluționară. Acest tip de rețea folosește doar straturi convoluționare și produce ca rezultat o mască de aceeași dimensiune spațială cu imaginea de la intrare. Practic rețeaua face o clasificare pentru fiecare pixel al imaginii.

Un tip particular de rețele complet convoluționare este U-net-ul. Acest tip de model a fost construit pentru segmentarea imaginilor medicale cu celule și țesuturi.

3.4 Căutare de hiper-parametri

În această secțiune sunt prezentate câteva strategii ce pot fi folosite pentru îmbunătățirea performanței unor modele numerice prin selectarea corectă a parametrilor în funcție de sarcina de clasificare.

Generic, hiper-parametrii reprezintă o configurație a modelului ce trebuie aleasă de utilizator înainte de antrenarea modelului. Acești parametri controlează toate aspectele unui model și se pot referi de la numărul de straturi ascunse sau de neuroni a unui MLP, la numărul și dimensiunea filtrelor de convoluție la un CNN, la numărul de arbori într-un Random Forest sau la tipul de metrică folosită într-un SVM. Deoarece acești parametri impactează puternic performanța modelului final este necesară o selecție atentă pentru a obține o soluție suficient de bună.

Există multiple moduri de selectare a hiper-parametrilor pentru o anumită problemă de clasificare sau regresie. Problema cea mai mare este adesea întâlnită la modele complexe, cu număr mare de hiper-parametri, unde nu se poate face verificarea tuturor configurațiilor posibile pentru a garanta că a fost identificată soluția optimă. Dacă avem 3 hiper-parametri cu 10 valori posibile fiecare, pentru a parcurge tot spațiul de posibilități este nevoie de 1000 de experimente. În realitate, 1000 de experimente ar fi o situație acceptabilă, problema reală apare când avem zeci de hiper-parametri ce pot avea valori continue.

Metoda de căutare cea mai simplă este aceea de căutare în grilă. Pentru fiecare hiper-parametru alegem un subset de valori discrete și testăm toate combinațiile dintre ele. Acest tip de abordare merge doar pentru un număr relativ mic de hiper-parametri.

Căutarea aleatoare este similară cu aceea de căutare în grilă, dar configurațiile sunt alese aleator dintr-un spațiu de posibilități. Acest tip de căutare, deși nu are nicio garanție a rezultatului, are tendința de a găsi o soluție suficient de bună mai rapid.

O variantă mai bună decât căutarea oarbă de hiper-parametri este utilizarea unei metode ghidate. O astfel de metodă este Optimizarea Bayesiană. Această abordare încearcă să caute în spațiul de posibilitate a hiper-parametrilor configurații care sunt foarte diferite de cele cunoscute inițial, iar după un timp începe să caute în jurul soluțiilor celor mai bune deja identificate pentru a găsi maxime locale.

3.5 Selectia, procesarea si augmentarea datelor

Curățarea datelor se referă la pașii de pre-procesare a datelor ce ajută la omogenizarea unui set de date. Poate implica eliminarea de valori duplicate, corectarea de imagini, structurarea datelor pentru a le face mai accesibile unor algoritmi.

Adesea după curățare se folosesc statisticile setului de date pentru a scala valorile de input. Acest lucru se face pentru a echilibra impactul fiecărei variabile de intrare asupra rezultatului final.

Un alt pas foarte important atunci când sunt folosiți algoritmi clasici de învățare automată este acela de reducere dimensională. Acest pas se referă la proiectarea datelor inițiale într-un spațiu dimensional redus pentru a concentra informația prezentă în date.

Capitolul 4

Utilizarea imaginilor microscopice în clasificarea de polen

În acest capitol sunt prezentate rezultatele de clasificare folosind imagini microscopice cu polen obținute de la un instrument numit BAA-500. Acest tip de dispozitiv automat de monitorizare a particulelor este construit pentru a imita modul de funcționare a capcanelor de polen Hirst. Diferența majoră constă în faptul că majoritatea pașilor care erau făcuți de un tehnician sunt acum automatizați.

În acest capitol sunt discutate multiple sarcini asociate cu cea de clasificare de polen precum detecția de particule în imagini, segmentarea semantică de imagini, utilizarea de modele convoluționare pre-antrenate pentru îmbunătățirea clasificării, identificarea de metode de augmentare a seturilor de date și de metode de antrenare eficientă a unor modele.

Capitolul are următoarea structură:

Secțiunea 4.1 prezintă setup-ul experimental folosit pentru obținerea de date de la un dispozitiv de tipul BAA-500, datele brute și modul de pre-procesare a datelor.

În secțiunea 4.2 sunt introduse două seturi de date noi ce au fost dezvoltate în cadrul acestei lucrări.

În secțiunea 4.3 sunt detaliați pașii de procesare a datelor brute și tipul modelelor numerice folosite pentru clasificarea de polen.

Rezultatele sunt prezentate și discutate în secțiunea 4.4.

4.1 Descrierea instrumentului

BAA-500 este un dispozitiv capabil să facă analiza particulelor din aer sau din alte medii lichide complet automat. În monitorizarea polenului este utilizat ca un sistem complet ce automatizează toți pașii ce erau anterior efectuați de om în alte capcane de polen manuale. Acest dispozitiv este utilizat în exterior și este ținut într-un container cu aer condiționat pentru a fi protejat de intemperii.

4.2 Crearea de seturi de date cu polen

În cadrul acestei lucrări au fost construite două seturi de date cu polen folosind un BAA-500. Aceste seturi de date au fost construite pentru a aborda două sarcini diferite. Prima sarcină este aceea de clasificare de particule de polen. Al doilea task se referă la clasificarea de polen folosind datele brute de la echipament, acest lucru necesitând o segmentare preliminară.

Primul set de date este compus din imagini cu polen din 19 clase. Pentru a putea antrena modele de învățare adâncă pe acest set de date numărul de exemple este peste 40 de mii. Acest set de date a fost construit utilizând sistemele existente ale rețelei ePIN cu o validare suplimentară efectuată de doi experți.

Al doilea set de date a fost construit manual pentru a rezolva sarcina de segmentare a imaginilor cu polen. Pentru acest set de date au fost construite măști cu informații despre clasă pentru aproximativ 4 mii de imagini.

Aceste două seturi de date au fost utilizate ulterior pentru antrenarea tuturor modelelor prezentate în secțiunile ulterioare.

4.3 Selecția arhitecturii

În această secțiune sunt prezentate tipurile de modele alese pentru clasificarea de imagini cu polen, dar și pentru segmentarea imaginilor microscopice cu polen.

Pentru clasificare de polen au fost folosite atât modele numerice de învățare automată clasice, cât și modele mai complexe de tip convoluționar sau de învățare adâncă. Algoritmii clasici includ Linear SVC, RBF SVC, Random Forest, ADABOOST, MLP și pentru a putea aplica acest tip de algoritmi pe date nestructurate s-au folosit ca algoritmi de pre-procesare HOG și LBP.

În cazul algoritmilor de tip CNN au fost alese un număr de arhitecturi foarte cunoscute cum ar fi: VGG-16, VGG-19, ResNet50, InceptionV3, Xception, DenseNet201. Aceste arhitecturi au fost alese pentru că există și variante ale lor pre-antrenate pe seturi mari de imagini, și astfel s-au putut face comparații între modele antrenate de la zero și modele care erau deja adaptate pentru clasificare de imagini.

Al doilea task abordat a fost acela de segmentare semantică a imaginilor microscopice cu polen. Pentru acest task au fost folosite arhitecturi complet convoluționare și mai exact variații ale U-net-ului.

4.4 Rezultate

Rezultatele abordării clasice de clasificare de polen pot fi văzute în Tabelul 4.1. Algoritmii selectați au avut ca punct de plecare rezultatele obținute în [7].

Pentru acest set de experimente datele au fost re-scalate de la 0–255 la valori cuprinse între 0 și 1. După acest pas dimensiunea datelor de intrare au fost normalizate la 360×360 . Peste acest set de date uniform au fost aplicați algoritmi histogram of gradients (HOG) și local binary pattern (LBP) pentru a transforma informația utilă într-o formă ușor de folosit de algoritmi clasici de clasificare. Modelele folosite sunt Support Vector Classifiers (SVCs), Random Forest (RF), Decision Trees cu Adaboost ensemble și Multi-Layer Perceptron (MLP). Toate aceste modele au fost construite utilizând biblioteca sci-kit learn.

Tabel 4.1 Classification F1 score for classical ML approaches.

MODEL TYPE/FEAT.ENG.	HOG	LBP
LINEAR SVC	0.46	0.46
RBF SVC	0.29	0.48
RANDOM FOREST	0.46	0.53
ADABOOST	0.41	0.52
MLP	0.47	0.62

Experimentele ce folosesc arhitecturi de tip CNN pot fi împărțite în două categorii: i) cele ce utilizează modele pre-antrenate pe setul de date ImageNet și ii) cele antrenate de la zero.

Când au fost folosite modele pre-antrenate datele de intrare au fost tratate ca și cum ar fi color prin duplicarea datelor pe cele 3 canale de culoare RGB. Modele pre-antrenate au fost adaptate pentru clasificare de polen prin înlocuirea părții dense a rețelelor cu un strat de tip Global Pooling și cu strat de tip soft-max utilizat pentru clasificare.

Rezultatele pot fi observate în Tabelul 4.2

Tabel 4.2 Classification F1 score for deep architectures.

ARCHITECTURE USED	PRE-TRAINED ON IMAGENET	FULLY TRAIN ON POLLEN
VGG-16	0.82	0.90
VGG-19	0.80	0.92
RESNET50	0.59	0.90
INCEPTIONV3	0.85	0.93
XCEPTION	0.86	0.93
DENSENET201	0.87	0.92

Modelele de tip U-net nu pot fi folosite direct în sarcina de clasificare de imagine. Folosind o metodă de agregare a clasificărilor tuturor pixelilor unei imagine putem reduce output-ul unui U-net la o singură clasă. Astfel, se pot compara rezultatele cu cele anterioare. Folosind această abordare, U-net-urile au avut un scor F1 de clasificare pe setul de test de 0.95, depășind cu mult toate celelate arhitecturi. Pentru că un U-net este

antrenat pe un task mai dificil el are adesea performanță mai bună pe un task mai simplu cum este cel de clasificare.

În Tabelul 4.3 pot fi văzute rezultatele rețelelor de tip U-net pe sarcina de segmentare.

Tabel 4.3 Classification class mean unweighted class IoU for U-net and variants at different widths.

U-NET VARIANT	W-4	W-8	W-16	W-32
FCN	0.61	0.72	0.71	0.79
U-NET	0.67	0.81	0.85	0.86
U-NET(ADD)	0.65	0.79	0.85	0.88
FCN + RES	0.64	0.78	0.82	0.82
U-NET + RES	0.66	0.78	0.83	0.87
U-NET(ADD) + RES	0.66	0.82	0.83	0.88

Capitolul 5

Date multi-modale de fluorescență și împrăștiere utilizate în clasificarea polenului

În acest capitol sunt prezentate rezultatele clasificării de polen folosind date multi-modale de fluorescență de la un dispozitiv Rapid-E. Acest capitol este structurat astfel:

În secțiunea 5.1 este prezentat instrumentul și modul în care poate fi folosit pentru a obține date de bună calitate.

Secțiunea 5.2 prezintă setup-ul experimental folosit pentru construirea de noi seturi de date, cât și toți pașii de pre-procesare necesari pentru a curăța și a augmenta seturile de date.

Secțiunea 5.3 descrie tipul de modele numerice ce pot fi folosite pentru a clasifica polen utilizând spectre de fluorescență. De asemenea, în această secțiune sunt prezentate moduri în care se poate face o căutare de hiper-parametri într-un mod eficient pentru a îmbunătăți performanța modelelor de clasificare.

Secțiunea 5.4 compară toate rezultatele de clasificare pe toate seturile de date publice obținute pe echipamente de tip Rapid-E.

5.1 Descrierea instrumentului

Rapid-E-ul este un analizor de aerosoli automat dezvoltat de Plair. Acest dispozitiv este amplasat în exterior pentru a putea aspira aer ce conține aerosoli. Particulele capturate în acest flux de aer sunt analizate folosind mai multe surse laser și mai mulți detectori. Datele obținute din acest dispozitiv pot caracteriza particule folosind spectre de fluorescență și imagini de împrăștiere. Datele brute obținute de dispozitiv sunt accesibile printr-o conexiune la internet ceea ce permite ca dispozitivul să fie amplasat în locuri relativ izolate. Echipamentul dispune de un container cu aer condiționat ce permite utilizarea dispozitivului tot timpul anului, indiferent de condițiile atmosferice.

În modul normal de funcționare, Rapid-e poate caracteriza până la 10 mii de particule pe minut, dar în realitate nu sunt depășite 2-3 mii de măsurători pe minut. Acest lucru este cauzat de fluxul relativ mic de aer și a concentrației de polen scăzute din aer.

Pentru a diferenția polenul de alți aerosoli prezenți în aer Rapid-E folosește mai multe filtrări.

5.2 Crearea de seturi de date cu polen

Pentru a construi seturi de date cu polen ce pot fi folosite pentru antrenarea de modele de clasificare, este necesară mutarea echipamentului din exterior într-un laborator pentru un mai bun control al expunerii la aerosoli.

În această secțiune sunt prezentate mai multe moduri de a crea seturi de date.

Primul setup experimental folosit constă în utilizarea unui flux constant de gaz inert, N_2 , pentru a ridica polen de pe un eșantion de plante, similar cum vântul transportă polen. Acest mod de eșantionare a polenului a avut rezultate limitate datorită celorlalți contaminanți ce sunt prezenți pe florile plantelor.

A doua tentativă s-a bazat pe utilizarea directă a unei cantități mai mari de polen colectat manual. Astfel s-a putut asigura o cantitate mare de eșantioane pentru multe specii de plante, cât și o calitate mai mare a datelor colectate.

Rezultatele acestor experimente de construire de seturi de date folosind un dispozitiv Rapid-E au fost descrise în detaliu și în lucrarea de conferință [10].

5.3 Selecția arhitecturii

În orice încercare de învățare automată adesea este cel mai bine să se pornească de la cele mai simple tipuri de modele și să fie crescută complexitatea lor doar dacă este necesar.

Motivele principale pentru o astfel de abordare sunt: Pornind de la modele simple pot fi mai ușor diagnosticate probleme legate de calitatea setului de date. De asemenea, modelele simple pot fi folosite ca referință pentru un minim de performanță pe care o putem aștepta de la modelele mai complexe. Și în final este posibil ca un model relativ simplu să aibă o acuratețe de clasificare suficient de bună pentru a nu fi nevoie să încercăm și modele mai complexe.

Când discutăm de clasificarea polenului folosind date de la un Rapid-E avem două tipuri de modele: cele care necesită reducere dimensională a datelor, modele clasice de învățare automată, și cele ce pot fi utilizate direct pe datele brute, modele de învățare adâncă. Datele brute sunt compuse din o imagine de împrăștiere și două tipuri de reprezentări ale spectrului de fluorescență.

Modelele simple utilizate pe acest tip de date sunt: Naive Bayes (GNB), Quadratic Discriminant Analysis (QDA), Decision Trees(DT), și multilayer perceptron (MLP).

Aceste modele au fost folosite după ce mai întâi s-au aplicat metode de reducere dimensională cum ar fi Principal Component Analysis (PCA), Independent Component Analysis(ICA), Gaussian Random Projection (GRP) sau Sparse Random Projection(SRP).

Modelele de învățare adâncă folosite au avut la baza arhitecturi de tip CNN, deoarece acestea permit procesarea directă a datelor 2D și 3D. Modelul utilizat este compus din 3 ramuri pentru cele 3 tipuri de date brute. Aceste ramuri se întâlnesc într-un trunchi comun ce are ca rol combinarea datelor și clasificarea finală.

Fiecare ramură convoluționară este compusă din multiple blocuri convoluționare, care la rândul lor sunt compuse din straturi de convoluție, de pooling și de neliniaritate.

5.4 Rezultate

În această secțiune rezultatele tuturor modelelor sunt prezentate pe multiple seturi de date.

Rezultatele prezentate în această secțiune au fost obținute după ce a fost efectuată o căutare de hiper-parametri pentru fiecare tip de model.

Rezultatele modelelor clasice de învățare automată sunt prezentate în Tabelul 5.1.

Algo	Config/Dataset	RO	LI	SRB	CH
GNB	Best Conf.	31%	32%	34%	54%
	Ensemble	31%	32%	35%	54%
QDA	Best Conf.	44%	48%	53%	60%
	Ensemble	44%	48%	53%	61%
DT	Best Conf.	42%	51%	52%	63%
	Ensemble	47%	56%	57%	69%
	Rand. Forest	42%	55%	54%	63%
MLP	Best Conf.	63%	71%	67%	73%
	Ensemble	66%	74%	70%	77%

Tabel 5.1 Classification accuracy for all classical models.

În Tabelul 5.2, se pot vedea rezultatele modelelor de tip CNN. Acest tabel prezintă rezultatele inițiale și după o căutare de hiper-parametri pentru 4 seturi de date. Se poate observa că îmbunătățirea datorată alegerii hiper-parametrilor nu este egală pe toate seturile de date. Acest lucru poate fi explicat prin faptul că fiecare set de date conține tipuri diferite de polen cu un grad diferit de dificultate în clasificare.

Creșterea performanței de clasificare poate fi analizat și pe fiecare tip de date de input. Examinând ramura utilizată pentru imaginea de împrăștiere observăm o scădere de la 42% la 38% a erorii medii. Pentru spectrul de fluorescență se observă o scădere medie a erorii de la 44% la 40%. Pentru modelul complet avem următoarele scăderi ale erorilor: de la 23% la 20%, pentru setul de date SAU-SRB; de 12%, de la 16% la 14%, pentru SAU-LI; de 13%, de la 15% la 13%, pentru SAU-CH; și de 20%, de la 24% la 19%, pentru setul de date MARS.

Tabel 5.2 CNN Model performance after hyper-parameter tuning (accuracy).

Data-Set	SAU-SRB	SAU-LI	SAU-CH	MARS
Baseline ^a	74%	73%	80%	-
Initial architecture ^b	77%	84%	85%	76%
Best architecture	80%	86%	87%	81%
Scattering image only	58%	62%	61%	64%
Fluorescence spectrum only	56%	72%	66%	58%
Lifetime signal only	68%	59%	72%	41%

^a Results from [57]; ^b results from [8].

Capitolul 6

Modelarea si predicția concentrației de polen

Spre deosebire de capitolele anterioare, ce abordau sarcina de clasificare de polen, în acest capitol sunt abordate metode de predicție a concentrație de polen din aer în funcție de alți parametri atmosferici mai ușor de monitorizat.

O posibilă soluție pentru persoanele ce suferă de alergii ar fi dezvoltarea unor sisteme ce pot face prognoze legate de concentrațiile de polen din aer. Dacă aceste prognoze ar fi suficient de precise și ar da informații până la nivelul genului sau a speciilor de plante care urmează să elibereze polen, persoanele alergice ar putea începe preventiv tratamentul cu antihistaminice și astfel ar minimiza impactul.

Prognozele de polen încearcă să prezică un număr de parametri printre care: data de început a sezonului de polen pentru o anumite specie, perioada de vârf și finalul sezonului, dar și concentrațiile zilnice dacă există suficiente date.

Astfel de prognoze pot fi efectuate cu două tipuri de modele numerice, cele bazate doar pe observații și cele ce folosesc modele fenologice.

Modele bazate doar pe observații nu introduc presupuneri despre relația dintre concentrația de polen și alți parametri atmosferici. Această categorie de modele include modele de regresie, modelarea seriilor de timp și toate tipurile de modele ce folosesc învățare automată și învățare adâncă.

Modelele fenologice încearcă să modeleze întregul circuit de viață al plantelor. Aceste modele folosesc parametri atmosferici împreună cu observații despre distribuția spațială a plantelor pentru a face predicții despre începutul sezonului de polen și despre concentrații zilnice de polen. Modelele fenologice sunt modele bazate pe proces, deoarece ele sunt construite folosind un număr de presupuneri despre răspunsul fiziologic al plantelor la modificări în mediul înconjurător.

În acest capitol mai mulți parametri atmosferici ce impactează sezonul de polen sunt discutați, împreună cu metode ce pot fi folosite pentru analiza datelor pe perioade mari de timp. Acești parametri atmosferici sunt importanți deoarece ei sunt mult mai ușor

de monitorizat și dacă pot fi folosiți în predicția de polen ei ar ușura mult modul de monitorizare a polenului.

Multiple variabile independente au fost folosite de-a lungul timpului pentru a prezice concentrația zilnică de polen, printre acestea putem menționa: temperatura medie, minimă și maximă pentru o zi, cantitatea de precipitații, umiditatea relativă, numărul de ore lumină, viteza vântului împreună cu direcția și persistența. (Stach et al. 2008 ; Smith and Emberlin 2005 ; Iglesias et al. 2007 ; Sánchez-Mesa et al. 2002 ; Rodriguez Rajo et al. 2004, 2005 ; Mendez et al. 2005 ; Toro et al. 1998; Goldberg et al. 1988)

Acest capitol are următoarea structură: Secțiunea 6.1 prezintă modul în care se poate monitoriza continuu umiditatea relativă și cantitatea totală de apă din atmosferă. Umiditatea relativă având un rol important pentru a prognoza începutul sezonului de polen [4], [6], [17], [64]. Secțiunea 6.2 prezintă o analiză a temperaturii și a unor indici derivați pe perioade lungi de timp. În această secțiune analizăm trendul crescător al temperaturii la scară mare. În secțiunea 6.3 este discutat impactul ploilor înghețate asupra vegetației și cum pot fi acestea prezise. În final în secțiunea 6.4 este prezentat un studiu de caz ce se referă la formarea fulgilor de zăpadă în jurul particulelor transportate pe distanțe mari.

6.1 Impactul umidității relative asupra sezonului de polen

Deși apa este un element esențial pentru creșterea și dezvoltarea plantelor, un nivel prea ridicat de precipitații sau de umiditate relativă poate avea un impact negativ asupra ciclului reproductiv al plantelor [4], [6]. Pentru dezvoltarea de modele de predicție de polen este foarte important ca apa din atmosferă să poată fi monitorizată precis și continuu.

În această secțiune sunt prezentate câteva metode de monitorizare continuă a cantității de apă din atmosferă folosind un fotometru solar Cimel. Acesta este amplasat la o stație de monitorizare din sud-estul Europei și este colocat cu un radiometru cu microunde. Datele de la cele două sisteme au fost comparate cu date de referință de la radiosonde lansate în apropiere pe perioada 2007-2017.

6.2 Efectul temperaturii asupra sezonului de polen

Între temperatura și începutul sezonului de polenizare este o legătură foarte puternică [22], [39], [54], [14]. În plus, climatul unei regiuni are o influență puternică asupra distribuției spațiale a plantelor.

Din aceste motive este importantă analiza trendului temperaturii. Impactul modificărilor temperaturii planetare are atât efectul de a perturba sezonul normal de polen

cât și a cauza o modificare a distribuției spațiale a plantelor, cu specii invazive care se răspândesc dinspre sud înspre nord.

În această secțiune sunt prezentate metode de analiză a stresului termic asupra bioclimatologiei pe arii extinse și a impactului pe care stresul termic îl are asupra ecosistemelor. Pentru astfel de analize sunt necesare seturi de date mari, și din perspective temporale, cât și spațiale. Un astfel de set de date este Universal Thermal Climate Index (UTCI) obținut din datele de reanaliză ERA5-HEAT. Deși acest index este utilizat pentru a studia impactul temperaturii asupra vieții umane el poate fi extins și la efectul asupra ecosistemului.

6.3 Ploaia înghețată și impactul asupra vegetației

Ploaia înghețată și înghețul au un efect devastator asupra plantelor [34], [60]. În fiecare toamnă, iarnă și primăvară, multe plante sunt în pericol de a fi rănite de vremea rece. În funcție de specia de plante această deteriorare poate fi cauzată de un simplu îngheț sau, în cazuri extreme, de evenimente de ploaie înghețată.

În această secțiune este prezentată o analiză a condițiilor atmosferice necesare pentru formarea ploii înghețate [2]. Această analiză este utilă încât poate fi ulterior folosită pentru a putea prezice evenimente viitoare. De asemenea, evenimentele de ploaie înghețată, deși sunt rare, au un impact puternic asupra vegetației și acest lucru trebuie considerat atunci când sunt construite modele de predicție de polen.

6.4 Zapadă portocalie și transportul particulelor pe distanțe mari

Această secțiune prezintă un fenomen neobișnuit observat deasupra României în dimineața zilei de 23 martie 2018. Evenimentul a fost observat în sud-estul țării și a fost caracterizat ca o ninsoare portocalie. Evenimentul a fost prezentat pe larg în mass-media locală și multe întrebări au fost ridicate asupra cauzei și originii prafului portocaliu în stratul de zăpadă.

Acest eveniment de intruziune de praf din Sahara este unul comun pentru România și Europa, dar un astfel de eveniment în timpul unei perioade cu temperaturi negative se întâmplă extrem de rar. Intruziunile de praf saharean sunt adesea întâlnite primăvara în toată Europa, dar nu suntacompaniate de ninsori.

În studiul, [38], o analiză a condițiilor sinoptice a fost efectuată pentru a găsi originea intruziunii. De asemenea, metodologii de segmentare și analiză a imaginilor microscopice SEM/LEM a fost prezentată în acest studiu. Aceste metode de segmentare sunt similare cu cele aplicate în cazul analizei imaginilor cu polen.

Capitolul 7

Concluzii

Acest capitol are rolul de a sumariza toate rezultatele acestei lucrări de doctorat. În secțiunile următoare sunt prezentate rezultatele în comparație cu obiectivele tezei și sunt enumerate contribuțiile individuale și publicațiile rezultate în urma acestei lucrări. În ultima secțiune sunt prezentate posibile direcții de cercetări viitoare.

7.1 Obiective și Rezultate

Această secțiune prezintă rezultatele obținute în timpul elaborării tezei. Aceste rezultate sunt grupate în funcție de capitolul în care au fost prezentate.

Capitolul 4 este o analiză profundă a modelelor de învățare automată ce pot fi folosite pentru clasificare și segmentare de imagini microscopice cu polen, folosind date de la un instrument de tip BAA-500.

- (a) Construirea de noi seturi de date validate de experți folosind un echipament BAA-500.
- (b) Dezvoltarea de modele de clasificare de polen folosind învățare automată clasică.
- (c) Dezvoltarea de modele de clasificare de polen folosind învățare adâncă și modele pre-antrenate pe alte sarcini de Computer Vision.
- (d) Construirea de modele de segmentare a imaginilor microscopice de polen.
- (e) Dezvoltarea unei noi abordări de augmentare a datelor pentru o utilizare eficientă a datelor de antrenare.
- (f) Construirea de seturi de date dedicate sarcinii de segmentare pentru anumite specii de spori alergeni.
- (g) Validarea rezultatelor modelelor printr-o comparație realizată în cadrul unei campanii de măsurători cu multiple echipamente colocate.

Capitolul 5 este o analiză profundă a modelelor de învățare automată ce pot fi folosite pentru clasificare de polen, folosind date multi-modale de la un instrument de tip Rapid-E.

- (a) Construirea de noi seturi de date validate de experți folosind un echipament Rapid-E.
- (b) Dezvoltarea de modele de clasificare de polen folosind învățare automată clasică.
- (c) Dezvoltarea de modele de clasificare cu performanță de SOTA pe toate seturile de date publice de la echipamente Rapid-E.
- (d) Dezvoltarea de metode de augmentare a datelor pentru îmbunătățirea performanței.
- (e) Identificarea de metode de căutare eficientă a hiper-parametrilor pentru arhitecturi de mari dimensiuni.

Capitolul 6 are ca obiectiv identificarea și studierea unor parametri atmosferici ce pot fi ulterior folosiți pentru a face prognoze de polen.

- (a) Realizarea analizei duratei de zi-lumină și de acoperire noroasă pentru a determina impactul acestora asupra măsurătorilor obținute din fotometre solare. Acest factor de corecție este util în determinarea cantității de apă din atmosferă.
- (b) Realizarea analizei indicelui Universal Thermal Climate Index (UTCI) pentru a identifica trendul de temperatură la nivel European pentru ultimii 40 de ani.
- (c) Realizarea analizei datelor de radiosondaj pentru perioada 1980-2000 pentru a identifica toate evenimentele de ploaie înghețată care s-au petrecut în România. Folosind această metodologie au putut fi izolate condițiile ce cauzează ploi înghețate. Acest lucru fiind util de integrat într-un model de predicție de polen.
- (d) Dezvoltarea unei metodologii de segmentare și fuzionare a imaginilor de microscopie de tip SEM/LEM pentru a putea vizualiza și identifica particule de dimensiuni mici transportate pe distanțe mari.

7.2 Contribuții originale

1. Îmbunătățit performanța de clasificare pe date de la echipamente de tip Rapid-E [1].
2. Construirea de noi seturi de date folosind un Rapid-E poziționat la sud de București [2].

3. Îmbunătățiri în acuratețea clasificării pe date Rapid-E folosind proceduri avansate de căutare de hiper-parametri [3].
4. Dezvoltarea de modele de clasificare cu necesar redus de putere de calcul pentru date de la Rapid-E [4].
5. Segmentare și clasificare de polen în imagini obținute de la echipamente de tip BAA-500 [5].
6. Construirea de seturi de date noi folosind echipamente de tip BAA-500 [5].
7. Dezvoltarea de metodologie de antrenare folosind date sintetice pentru a antrena modele pentru segmentare de imagine [5].
8. Construirea de modele capabile să segmenteze și să identifice anumite specii de spori alergeni în timp real pe date de la BAA-500 [6].
9. Analiza duratei de zi-lumină și de acoperie noroasă [7].
10. Vizualizarea și metodologia de analiză a seturilor mari de date cu informații despre temperatură [8].
11. Dezvoltarea metodologiei de analiză, de vizualizare și softurile necesare pentru a studia date de radiosondaj [9].
12. Dezvoltarea metodologiei de segmentare de imagine și de vizualizare pentru date provenite de la microscop electronic [10].

7.3 Lista cu articole publicate

1. **Boldeanu, M.**, Cucu, H., Burileanu, C., and Mărmureanu, L. Automatic pollen classification using convolutional neural networks. In 2021 44th International Conference on Telecommunications and Signal Processing (TSP), pages 130–133.
2. **Boldeanu, M.**, Marin, C., Ene, D., Mărmureanu, L., Cucu, H., and Burileanu, C. Mars: the first romanian pollen dataset using a rapid-e particle analyzer. In 2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pages 145–150.
3. **Boldeanu, M.**; Cucu, H.; Burileanu, C.; Mărmureanu, L. Multi-Input Convolutional Neural Networks for Automatic Pollen Classification. Appl. Sci. 2021, 11, 11707. <https://doi.org/10.3390/app112411707>
4. **Boldeanu, M.**; Cucu, H.; Burileanu, C.; Mărmureanu, L. Pollen Classification using Classical Machine Learning Algorithms on Fluorescence and Scattering Imaging, Buletinul Polithenicii (Not yet Published)

5. **Boldeanu, M.+**; González-Alonso, M.+; Cucu, H.;Burileanu, C.; Maya-Manzano, J. M. and Buters, J. T. M., "Automatic Pollen Classification and Segmentation Using U-Nets and Synthetic Data," in IEEE Access, vol. 10, pp. 73675-73684, 2022, doi: 10.1109/ACCESS.2022.3189012. ¹
6. González-Alonso, M.+; **Boldeanu, M.+**; Koritnik, T.; Gonçalves, J. ; Belzner, L. ; Stemmler, T. ; Gebauer, R. ; Grewling L.; Tummon, F.; Maya-Manzano, J. M. ;Ariño, A.H.; Schmidt-Weber, C.; and Buters, J. T. M., Alternaria spore exposure in Bavaria, Germany, measured by using artificial intelligence algorithms in a network of BAA500 automatic pollen monitors., Scientific Reports (**Not yet Published**)²
7. Fragkos, K., Antonescu, B., Giles, D. M., Ene, D., **Boldeanu, M.**, Efstathiou, G. A., Belegante, L., and Nicolae, D.: Assessment of the total precipitable water from a sun photometer, microwave radiometer and radiosondes at a continental site in southeastern Europe, Atmos. Meas. Tech., 12, 1979–1997, <https://doi.org/10.5194/amt-12-1979-2019>, 2019.
8. Antonescu, B., Mărmureanu, L., Vasilescu, J., Marin, C., Andrei, S., **Boldeanu, M.**, Ene, D., and Țilea, A. (2021). A 41-year bioclimatology of thermal stress in europe. International Journal of Climatology, 41(7):3934–3952.
9. Andrei, S., Antonescu, B., **Boldeanu, M.**, Mărmureanu, L., Marin, C. A., Vasilescu, J., and Ene, D. (2019). An exceptional case of freezing rain in Bucharest (Romania). Atmosphere, 10(11).
10. Mărmureanu, L., Marin, C. A., Andrei, S., Antonescu, B., Ene, D., **Boldeanu, M.**, Vasilescu, J., Vițelaru, C., Cadar, O., and Levei, E. (2019). Orange snow—a saharan dust intrusion over romania during winter conditions. Remote Sensing, 11(21).

7.4 Cercetări viitoare

Această secțiune prezintă câteva direcții de cercetare ce ar putea fi urmate în continuarea acestei teze.

Pentru date provenite de la Rapid-E:

Identificare de arhitecturi ce permit o mai bună generalizare a modelelor antrenate pe date de la dispozitive multiple.

Dezvoltarea de seturi de date folosind mai multe surse și mai multe specii de polen.

¹Authors with + contributed equally.

²Authors with + contributed equally.

O validare a rezultatelor prin organizarea unor campanii de măsurători cu multiple echipamente co-locate.

Pentru date provenind de la BAA-500:

Dezvoltarea de modele ce pot face segmentare panoptică.

Îmbunătățirea generării de date artificiale pentru a crește eficiența antrenării de modele.

Construirea de seturi de date pentru sarcina de segmentare cu un număr mare de clase.

Bibliografie

- [1] Allen, G. P., Hodgson, R. M., Marsland, S. R., and Flenley, J. R. (2008). Machine vision for automated optical recognition and classification of pollen grains or other singulated microscopic objects. In *2008 15th International Conference on Mechatronics and Machine Vision in Practice*, pages 221–226.
- [2] Andrei, S., Antonescu, B., Boldeanu, M., Marmureanu, L., Marin, C. A., Vasilescu, J., and Ene, D. (2019). An exceptional case of freezing rain in bucharest (romania). *Atmosphere*, 10(11).
- [3] Arias, D. G., Mussel Cirne, M. V., Chire, J. E., and Pedrini, H. (2017). Classification of pollen grain images based on an ensemble of classifiers. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 234–240.
- [4] Aronne, G. (1999). Effects of relative humidity and temperature stress on pollen viability of *cistus incanus* and *myrtus communis*. *Grana*, 38(6):364–367.
- [5] Astolfi, G., Gonçalves, A. B., Menezes, G. V., Borges, F. S. B., Astolfi, A. C. M. N., Matsubara, E. T., Alvarez, M., and Pistori, H. (2020). Pollen73s: An image dataset for pollen grains classification. *Ecological Informatics*, 60:101165.
- [6] Barnes, C., Pacheco, F., Landuyt, J., Hu, F., and Portnoy, J. (2001). The effect of temperature, relative humidity and rainfall on airborne ragweed pollen concentrations. *Aerobiologia*, 17:61–68.
- [7] Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., and Siniscalco, C. (2020). Pollen13k: A large scale microscope pollen grain image dataset. *CoRR*, abs/2007.04690.
- [8] Boldeanu, M., Cucu, H., Burileanu, C., and Mărmureanu, L. (2021a). Automatic pollen classification using convolutional neural networks. In *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, pages 130–133.
- [9] Boldeanu, M., Cucu, H., Burileanu, C., and Mărmureanu, L. (2021b). Multi-input convolutional neural networks for automatic pollen classification. *Applied Sciences*, 11(24).
- [10] Boldeanu, M., Marin, C., Ene, D., Marmureanu, L., Cucu, H., and Burileanu, C. (2021c). Mars: the first romanian pollen dataset using a rapid-e particle analyzer. In *2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, pages 145–150.
- [11] Bonton, P., Boucher, A., Thonnat, M., Tomczak, R., Hidalgo, P. J., Belmonte, J., and Galan, C. (2011). Colour image in 2d and 3d microscopy for the automation of pollen rate measurement. *Image Analysis & Stereology*, 21(1):25–30.

- [12] Boucher, A., Hidalgo, P. J., Thonnat, M., Belmonte, J., Galan, C., Bonton, P., and Tomczak, R. (2002). Development of a semi-automatic system for pollen recognition. *Aerobiologia*, 18(3):195–201.
- [13] Carrión, P., Cernadas, E., Gálvez, J. F., Damián, M., and de Sá-Otero, P. (2004). Classification of honeybee pollen using a multiscale texture filtering scheme. *Machine Vision and Applications*, 15(4):186–193.
- [14] Chaturvedi, P., Wiese, A. J., Ghatak, A., Drábková, L. Z., Weckwerth, W., and Honys, D. (2021). Heat stress response mechanisms in pollen development. *New Phytologist*, 231(2):571–585.
- [15] Chen, C., Hendriks, E. A., Duin, R. P. W., Reiber, J. H. C., Hiemstra, P. S., de Weger, L. A., and Stoel, B. C. (2006). Feasibility study on automated recognition of allergenic pollen: grass, birch and mugwort. *Aerobiologia*, 22(4):275–284.
- [16] Chica, M. (2012). Authentication of bee pollen grains in bright-field microscopy by combining one-class classification techniques and image processing. *Microscopy Research and Technique*, 75(11):1475–1485.
- [17] Dąbrowska-Zapart, K., Chłopek, K., and Niedźwiedź, T. (2018). The impact of meteorological conditions on the concentration of alder pollen in sosnowiec (poland) in the years 1997-2017. *Aerobiologia*, 34(4):469–485. 30532345[pmid].
- [18] Daood, A., Ribeiro, E., and Bush, M. (2016). Pollen grain recognition using deep learning. In Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., and Isenberg, T., editors, *Advances in Visual Computing*, pages 321–330, Cham. Springer International Publishing.
- [19] Daood, A. I., Ribeiro, E., and Bush, M. B. (2018). Sequential recognition of pollen grain z-stacks by combining cnn and rnn. In *FLAIRS Conference*.
- [20] Erdtman, G. (1952). Pollen morphology and plant taxonomy. *Geologiska Foreningen i Stockholm Forhandlingar*, 74(4):526–527.
- [21] Gallardo-Caballero, R., García-Orellana, C. J., García-Manso, A., González-Velasco, H. M., Tormo-Molina, R., and Macías-Macías, M. (2019). Precise pollen grain detection in bright field microscopy using deep learning techniques. *Sensors*, 19(16).
- [22] Gehrig, R. (2006). The influence of the hot and dry summer 2003 on the pollen season in switzerland. *Aerobiologia*, 22(1):27–34.
- [23] Gonçalves, A. B., Souza, J. S., Silva, G. G. d., Cereda, M. P., Pott, A., Naka, M. H., and Pistori, H. (2016). Feature extraction and machine learning for the classification of brazilian savannah pollen grains. *PLOS ONE*, 11(6):1–20.
- [24] HIRST, J. M. (1952). AN AUTOMATIC VOLUMETRIC SPORE TRAP. *Annals of Applied Biology*, 39(2):257–265.
- [25] Holt, K., Allen, G., Hodgson, R., Marsland, S., and Flenley, J. (2011). Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany and Palynology*, 167(3):175–183.

- [26] Khanzhina, N., Filchenkov, A., Minaeva, N., Novoselova, L., Petukhov, M., Khari-sova, I., Pinaeva, J., Zamorin, G., Putin, E., Zamyatina, E., et al. (2022). Combating data incompetence in pollen images detection and classification for pollinosis prevention. *Computers in biology and medicine*, 140:105064.
- [27] Khanzhina, N., Putin, E., Filchenkov, A., and Zamyatina, E. (2018). Pollen grain recognition using convolutional neural network. In *ESANN*.
- [28] Kramer, C. L. and Pady, S. M. (1966). A new 24-hour spore sampler. *Phytopathology*, 56(5):517–520.
- [29] Kubera, E., Kubik-Komar, A., Piotrowska-Weryszko, K., and Skrzypiec, M. (2021). Deep learning methods for improving pollen monitoring. *Sensors (Basel)*, 21(10).
- [30] Lake, I. R., Jones, N. R., Agnew, M., Goodess, C. M., Giorgi, F., Hamaoui-Laguel, L., Semenov, M. A., Solmon, F., Storkey, J., Vautard, R., and Epstein, M. M. (2017). Climate change and future pollen allergy in europe. *Environmental health perspectives*, 125(3):385–391. EHP173[PII].
- [31] Langford, M., Taylor, G., and Flenley, J. (1990). Computerized identification of pollen grains by texture analysis. *Review of Palaeobotany and Palynology*, 64(1):197–203. The Proceedings of the 7th International Palynological Congress (Part I).
- [32] Li, P. and Flenley, J. R. (1999). Pollen texture identification using neural networks. *Grana*, 38(1):59–64.
- [33] Li, P., Treloar, W., Flenley, J., and Empson, L. (2004). Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. *Journal of Quaternary Science: Published for the Quaternary Research Association*, 19(8):755–762.
- [34] Lind, L., Nilsson, C., and Weber, C. (2014). Effects of ice and floods on vegetation in streams in cold regions: implications for climate change. *Ecology and evolution*, 4(21):4173–4184. 25505542[pmid].
- [35] Lindbladh, M., O'Connor, R., and Jacobson, G. L. (2002). Morphometric analysis of pollen grains for paleoecological studies: classification of picea from eastern north america. *American Journal of Botany*, 89(9):1459–1467.
- [36] Lu, L.-L., Jiao, B.-H., Qin, F., Xie, G., Lu, K.-Q., Li, J.-F., Sun, B., Li, M., Ferguson, D. K., Gao, T.-G., Yao, Y.-F., and Wang, Y.-F. (2022). *Artemisia* pollen dataset for exploring the potential ecological indicators in deep time. *Earth System Science Data Discussions*, 2022:1–48.
- [37] Manikis, G. C., Marias, K., Alissandrakis, E., Perrotto, L., Savvidaki, E., and Vidakis, N. (2019). Pollen grain classification using geometrical and textural features. In *2019 IEEE International Conference on Imaging Systems and Techniques (IST)*, pages 1–6.
- [38] Marmureanu, L., Marin, C. A., Andrei, S., Antonescu, B., Ene, D., Boldeanu, M., Vasilescu, J., Vițelaru, C., Cadar, O., and Levei, E. (2019). Orange snow—a saharan dust intrusion over romania during winter conditions. *Remote Sensing*, 11(21).
- [39] Natali, F., Cecchi, L., Torrigiani Malaspina, T., Barbano, F., and Orlandini, S. (2013). Impact of 2003 heat waves on aerobiological indices of allergenic herbaceous family pollen season in tuscany (italy). *Aerobiologia*, 29(3):399–406.

- [40] Noll, K. E. (1970). A rotary inertial impactor for sampling giant particles in the atmosphere. *Atmospheric Environment (1967)*, 4(1):9–19.
- [41] Ranzato, M., Taylor, P., House, J., Flagan, R., LeCun, Y., and Perona, P. (2007). Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters*, 28(1):31–39.
- [42] Razmovski, V., O’meara, T., Hjelmroos, M., Marks, G., and Tovey, E. (1998). Adhesive tapes as capturing surfaces in burkard sampling. *Grana*, 37(5):305–310.
- [43] Redondo, R., Bueno, G., Chung, F., Nava, R., Víctor Marcos, J., Cristóbal, G., Rodríguez, T., Gonzalez-Porto, A., Pardo, C., Déniz, O., and Escalante-Ramírez, B. (2015). Pollen segmentation and feature evaluation for automatic classification in bright-field microscopy. *Computers and Electronics in Agriculture*, 110:56–69.
- [44] Reisert, M. and Burkhardt, H. (2006). Invariant features for 3d-data based on group integration using directional information and spherical harmonic expansion. In *18th International Conference on Pattern Recognition (ICPR’06)*, volume 4, pages 206–209.
- [45] Rodinkova, V., Kremenska, L., Palamarchuk, O., Motruk, I., Alexandrova, E., Dudarenko, O., Vakolyuk, L., and Yermishev, O. (2018). Seasonal changes in plant pollen concentrations over recent years in vinnitsya, central ukraine. *Acta Agrobotanica*, 71.
- [46] Rodriguez-Damian, M., Cernadas, E., Formella, A., Fernandez-Delgado, M., and Sa-Otero, P. D. (2006). Automatic detection and classification of grains of pollen based on shape and texture. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 36(4):531–542.
- [47] Rodriguez-Damian, M., Cernadas, E., Formella, A., and Sa-Otero, R. (2004). Pollen classification using brightness-based and shape-based descriptors. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 212–215 Vol.2.
- [48] Ronneberger, O., Schultz, E., and Burkhardt, H. (2002). Automated pollen recognition using 3d volume images from fluorescence microscopy. *Aerobiologia*, 18(2):107–115.
- [49] Sauvageat, E., Zeder, Y., Auderset, K., Calpini, B., Clot, B., Crouzy, B., Konzelmann, T., Lieberherr, G., Tummon, F., and Vasilatou, K. (2020). Real-time pollen monitoring using digital holography. *Atmospheric Measurement Techniques*, 13:1539–1550.
- [50] Schaefer, J., Milling, M., Schuller, B. W., Bauer, B., Brunner, J. O., Traidl-Hoffmann, C., and Damialis, A. (2021). Towards automatic airborne pollen monitoring: From commercial devices to operational by mitigating class-imbalance in a deep learning approach. *Science of The Total Environment*, 796:148932.
- [51] Schiele, J., Rabe, F., Schmitt, M., Glaser, M., Häring, F., Brunner, J. O., Bauer, B., Schuller, B., Traidl-Hoffmann, C., and Damialis, A. (2019). Automated classification of airborne pollen using neural networks. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4474–4478.

- [52] Sedghy, F., Varasteh, A.-R., Sankian, M., and Moghadam, M. (2018). Interaction between air pollutants and pollen grains: The role on the rising trend in allergy. *Reports of biochemistry & molecular biology*, 6(2):219–224. 29766006[pmid].
- [53] Sevillano, V., Holt, K., and Aznarte, J. L. (2020). Precise automatic classification of 46 different pollen types with convolutional neural networks. *PLOS ONE*, 15(6):1–15.
- [54] Smith, L. M. (2019). The Heat Is On: Maize Pollen Development after a Heat Wave. *Plant Physiology*, 181(2):387–388.
- [55] Takahashi, Y., Kawashima, S., Fujita, T., Ito, C., Togashi, R., and Takeda, H. (2001). [comparison between real-time pollen monitor KH-3000 and burkard sampler]. *Arerugi*, 50(12):1136–1142.
- [56] Tsiknakis, N., Savvidaki, E., Manikis, G. C., Gotsiou, P., Remoundou, I., Marias, K., Alissandrakis, E., and Vidakis, N. (2022). Pollen grain classification based on ensemble transfer learning on the cretan pollen dataset. *Plants*, 11(7).
- [57] Šaulienė, I., Šukienė, L., Daunys, G., Valiulis, G., Vaitkevičius, L., Matavulj, P., Brdar, S., Panic, M., Sikoparija, B., Clot, B., Crouzy, B., and Sofiev, M. (2019). Automatic pollen recognition with the rapid-e particle counter: the first-level procedure, experience and next steps. *Atmospheric Measurement Techniques*, 12(6):3435–3452.
- [58] Waite, K. J. (1995). Blackley and the development of hay fever as a disease of civilization in the nineteenth century. *Medical history*, 39(2):186–196. 7739297[pmid].
- [59] Wang, Z., Bao, W., Lin, D., and Wang, Z. (2019). A local feature descriptor based on sift for 3d pollen image recognition. *IEEE Access*, 7:152658–152666.
- [60] Wassan, S., Xi, C., Jhanjhi, N., and Binte-Imran, L. (2021). Effect of frost on plants, leaves, and forecast of frost events using convolutional neural networks. *International Journal of Distributed Sensor Networks*, 17(10):15501477211053777.
- [61] Wood, G. S. (1961). Sampling apparatus and method.
- [62] Xie, Y. and OhEigartaigh, M. (2010). 3d discrete spherical fourier descriptors based on surface curvature voxels for pollen classification. In *2010 WASE International Conference on Information Engineering*, volume 1, pages 207–211.
- [63] Yaslan, Y. and Cataltepe, Z. (2008). Co-training with adaptive bayesian classifier combination. In *2008 23rd International Symposium on Computer and Information Sciences*, pages 1–4.
- [64] Zemmer, F., Dahl, Å., and Galán, C. (2022). The duration and severity of the allergenic pollen season in istanbul, and the role of meteorological factors. *Aerobiologia*.