

LUCRARE DE LICENȚĂ

PREDICȚIA CONSTANTEI CHIMICE PKA FOLOSIND O REȚEA NEURONALĂ ȘI DETERMINAREA AUTOMATIZATĂ A ACESTEIA

Candidat: Cristian CRIȘAN

Coordonator științific: Prof.dr.habil.ing. Mihai-Victor MICEA

Sesiunea: Iunie 2025

REZUMAT

Această lucrare propune un model de predicție a unei constante chimice pKa a unor substanțe chimice, bazat pe rețele neuronale convoluționale aplicate pe grafuri moleculare. Reprezentarea compușilor chimici sub formă de grafuri permite o descriere obiectivă și detaliată a structurii moleculare și a interacțiunilor intramoleculare, aspect esențial pentru estimarea precisă a proprietăților chimice.

De asemenea, a fost creată o metodologie de titrare automatizată pentru determinarea valorilor pKa, care permite compararea predicțiilor modelului cu datele experimentale. Pentru a utiliza ușor a sistemului, o aplicație a fost dezvoltată pentru a controla întreg procesul automatizat.

Rezultatele experimentale indică o precizie ridicată a modelului propus, subliniind potențialul rețelelor neuronale bazate pe grafuri în domeniul chemoinformaticii. Proiectul propune un sistem complex, folosind predicții și o aplicație experimentală, reușind să facă o punte între cele două tipuri de determinări diferite conceptual. Acest mod de abordare poate facilita estimarea imediată și exactă a proprietăților chimice, cu aplicații relevante în industria farmaceutică, alimentară și agrochimică.

ABSTRACT

This study presents a predictive model for pKa, a fundamental parameter representing the acid strength, using graph neural networks (GNNs) applied to molecular graphs. Representing chemical compounds as graphs enables a detailed and objective description of molecular structure and intramolecular interactions, which is essential for the accurate prediction of chemical properties.

An automated titration system was also developed to experimentally determine the pKa values, allowing a direct comparison with model predictions. A user-friendly application was created to control the entire automated process, ensuring reproducibility and ease of use in both academic and industrial settings.

In today's context of greater dependence on artificial intelligence to speed chemical discovery and process optimization, this strategy complements the demand for quick, accurate, and comprehensible methods of computational chemistry and chemoinformatics. By using machine learning and experimental automation, the proposed framework contributes to the digital transformation of chemical research, with implications for more sustainable and efficient development in the pharmaceutical, food, and agrochemical industries.

Cuprins

1.	INTRODUCERE. PUNEREA PROBLEMEI	5
1.1.	CONTEXTUL ACTUAL	5
1.2.	NECESITATEA ȘI IMPORTANȚA LUCRĂRII	6
1.3.	OBIECTIVELE LUCRĂRII	7
1.4.	ABORDAREA PROCESULUI DE DEZVOLTARE	8
1.5.	STRUCTURA LUCRĂRII	9
2.	STUDIU BIBLIOGRAFIC	11
2.1.	METODE FOLOSITE PÂNĂ ACUM	11
2.2.	SISTEME UTILIZATE ÎN PREZENT	12
2.3.	PROVOCĂRI ȘI LIMITĂRI ALE SISTEMELOR ACTUALE	15
2.4.	IMPACTUL SISTEMELOR ACTUALE	16
2.5.	DIRECȚII VIITOARE	18
3.	FUNDAMENTARE TEORETICĂ	20
3.1.	REȚELE NEURONALE GRAFICE ÎN PREDICȚIA VALORII pK_a	20
3.2.	APLICAȚIILE ÎNVĂȚĂRII AUTOMATE ÎN CHEMOINFORMATICĂ	21
3.3.	PRINCIPII DE AUTOMATIZARE ȘI CONTROL AL PROCESELOR	24
3.4.	PRELUCRAREA NUMERICĂ A DATELOR EXPERIMENTALE	26
3.5.	VALIDAREA EXPERIMENTALĂ A MODELELOR PREDICTIVE	28
4.	SOLUȚIA PROPUȘĂ ȘI METODOLOGIA DE DEZVOLTARE	30
4.1.	DESCRIEREA GENERALĂ A SOLUȚIEI	30
4.2.	ARHITECTURA FUNCȚIONALĂ A SISTEMULUI PROPUȘ	31
4.3.	METODOLOGIA DE VALIDARE A REZULTATELOR	32
4.4.	ETAPELE DE DEZVOLTARE	33
5.	IMPLEMENTARE	35
5.1.	ALEGEREA SETULUI DE DATE UTILIZAT	35
5.2.	DEZVOLTAREA REȚELEI NEURONALE	36
5.3.	IMPLEMENTAREA APLICAȚIEI PREDICTORULUI	40
5.4.	DEZVOLTAREA SISTEMULUI AUTOMAT DE TITRARE	42
5.5.	IMPLEMENTAREA APLICAȚIEI DE CONTROL A TITRĂRII	47
6.	REZULTATE EXPERIMENTALE	49
6.1.	EVALUAREA PERFORMANȚEI MODELULUI DE PREDICȚIE	49
6.2.	DETERMINAREA EXPERIMENTALĂ A VALORII pK_a	51
6.3.	CONCORDANȚA PREDICȚIILOR CU DATELE EXPERIMENTALE	55
7.	CONCLUZII, CONTRIBUȚII ȘI DIRECȚII VIITOARE DE DEZVOLTARE	57
7.1.	CONCLUZII GENERALE	57
7.2.	CONTRIBUȚII	58
7.3.	DIRECȚII DE DEZVOLTARE	59
8.	BIBLIOGRAFIE	61

1. INTRODUCERE. PUNEREA PROBLEMEI

În contextul actual al digitalizării accelerate, științele exacte se intersectează tot mai frecvent cu domeniul tehnologiilor informatice, generând soluții inovatoare pentru probleme complexe. Chimie computațională, ingineria software și inteligența artificială formează un cadru interdisciplinar puternic, în care metodele clasice de cercetare pot fi extinse, optimizate și automatizate. Această lucrare își propune să valorifice avantajele acestui cadru modern, prin combinarea tehnologiilor hardware și software cu metode avansate de analiză a datelor, în scopul determinării și predicției unei proprietăți fundamentale: constanta de aciditate (pK_a).

Capitolul de față oferă o privire generală asupra contextului științific și tehnologic în care se înscrie această lucrare, evidențiind importanța pK_a -ului în chimie și necesitatea unor metode eficiente de determinare a acestuia. De asemenea, sunt prezentate direcțiile moderne de cercetare în domeniul învățării automate și automatizării aplicate în chimie, precum și motivația care stă la baza realizării proiectului.

1.1. CONTEXTUL ACTUAL

În ultimele decenii, dezvoltările rapide din domeniul inteligenței artificiale și, în mod special, al învățării automate au revoluționat modul în care problemele complexe sunt abordate în știință, tehnologie și inginerie. Aceste tehnologii au depășit faza experimentală și s-au integrat tot mai profund în procesele de cercetare și dezvoltare dintr-o gamă largă de domenii, inclusiv chimia computațională și chemoinformatica. Aceasta din urmă a devenit un instrument esențial pentru explorarea proprietăților moleculare, permițând analize predictive și optimizări ale compușilor chimici, reducând totodată nevoia de experimente laborioase și costisitoare.

Unul dintre parametrii fundamentali în chimia moleculelor cu caracter acid sau bazic este constanta de aciditate, notată pK_a . Această valoare exprimă tendința unei substanțe chimice de a elibera un proton într-o soluție apoasă, fiind o măsură esențială pentru descrierea comportamentului său acido-bazic. Înțelegerea valorii pK_a este crucială pentru o multitudine de aplicații științifice și tehnologice, de la proiectarea de medicamente eficiente și stabile, până la controlul reacțiilor chimice în procese industriale sau la evaluarea impactului unor substanțe asupra mediului. În biologie, pK_a -ul influențează reactivitatea și structura biomoleculelor, afectând activitatea enzimatică, formarea complexelor moleculare și transportul membranal.

Din punct de vedere tehnologic, metodele clasice de determinare a pK_a -ului, precum titrarea potențimetrică, spectrofotometria sau alte metode electrochimice, deși precise, implică o serie de limitări. Acestea presupun timp îndelungat de execuție, consum mare de reactivi, condiții stricte de laborator și, în general, o intervenție manuală constantă. În acest context, predicția computațională a valorilor pK_a , utilizând modele de învățare automată, a devenit o direcție strategică în cercetarea chimică modernă.

Integrarea metodelor de învățare automată în predicția pK_a oferă oportunitatea de a crea sisteme care pot învăța, pe baza unor baze de date experimentale, relații complexe între structura moleculară și comportamentul acid-bazic. Spre deosebire de metodele tradiționale de calcul bazate pe ecuații rigide sau pe simulări cuantice de înaltă complexitate, modelele de tip învățare automată pot generaliza comportamente și pot oferi estimări rapide și surprinzător de precise pentru substanțe noi, necunoscute anterior. În plus, aceste modele sunt capabile să utilizeze un set vast de descriori moleculari, cum ar fi electronegativitatea, distribuția sarcinilor parțiale sau masa atomică, oferind astfel o perspectivă multidimensională asupra relațiilor structură-proprietate.

Lucrarea de față își propune să valorifice potențialul acestor tehnologii, demonstrând modul în care cunoștințele dobândite în domeniul Ingineriei Calculatoarelor. Programarea, analiza de date, dezvoltarea software sau controlul hardware pot fi aplicate într-un context interdisciplinar. Concret, proiectul presupune dezvoltarea unui model de predicție a valorii pK_a pentru acizi organici monoprotici, utilizând o rețea neuronală artificială antrenată pe un set de date experimentale. Acest model este ulterior validat printr-un sistem experimental automatizat de titrare, construit folosind microcontrolerul Arduino, un electrod de pH, un amplificator operațional și un injectomat controlat software.

Prin această abordare integrată, se creează o punte reală între teoria învățării automate și aplicabilitatea practică în analiza chimică, oferind un exemplu convingător de cum pot fi utilizate instrumentele specifice informaticii și ingineriei calculatoarelor în sprijinul cercetării științifice. În plus, proiectul pune în valoare capacitatea acestor tehnologii de a contribui la accelerarea procesului de descoperire și înțelegere a proprietăților fundamentale ale substanțelor chimice, într-un mod mai rapid, mai eficient și mai accesibil decât metodele tradiționale.

1.2. NECESITATEA ȘI IMPORTANȚA LUCRĂRII

În situația progreselor tehnologice actuale, chimia ca știință nu mai este limitată la metodele experimentale clasice, ci se dezvoltă tot mai mult prin integrarea cu informatica, automatizarea și inteligența artificială. Această convergență a dus la apariția unor ramuri interdisciplinare precum chemoinformatica și chimia computațională, care permit modelarea, simularea și predicția comportamentului molecular prin metode numerice și algoritmi de învățare automată.

Necesitatea de a obține rapid și eficient informații despre proprietățile chimice fundamentale, cum este constanta de aciditate (pK_a), a determinat apariția unor metode alternative de determinare sau estimare, capabile să reducă timpul de analiză, consumul de reactivi și intervenția umană. În paralel, progresul în domeniul automatizării proceselor chimice, prin utilizarea microcontrolerelor și a sistemelor de achiziție și control, a permis reproducerea unor experimente clasice într-un mod standardizat, automat, reproductibil și accesibil.

În acest context, învățarea automată oferă o abordare modernă și eficientă pentru predicția valorilor pKa, bazată pe date existente și relații structurale complexe între molecule. Spre deosebire de metodele analitice rigide sau de simulările cuantice de mare complexitate, modelele de învățare automată pot fi antrenate pe baze de date experimentale și apoi utilizate pentru a face predicții rapide și generalizabile pentru molecule noi. Astfel, chemoinformatica devine un instrument indispensabil în etapa incipientă de cercetare și dezvoltare, permițând trierea rapidă a compușilor cu potențial interesant înainte de a trece la validarea experimentală.

1.3. OBIECTIVELE LUCRĂRII

Această lucrare propune o abordare interdisciplinară care îmbină metode avansate din ingineria calculatoarelor cu aplicații relevante în chimie analitică, având ca scop dezvoltarea unui sistem integrat pentru estimarea și validarea valorilor pKa ale unor acizi organici monoprotici. În centrul acestei lucrări stau două direcții complementare: predicția computațională a valorii pKa prin metode de învățare automată și determinarea experimentală a acesteia printr-un sistem automatizat de titrare. Prin această abordare, se urmărește nu doar rezolvarea unei probleme științifice punctuale, ci și demonstrarea modului în care cunoștințele din domeniul calculatoarelor pot contribui activ la dezvoltarea cercetării aplicate în științele naturale (în acest caz, chimia).

Predicția valorilor pKa utilizând rețele neuronale convoluționale

Un prim obiectiv major constă în dezvoltarea unui model computațional performant, capabil să estimeze valoarea constantei de aciditate pKa pe baza structurii moleculare.

Pentru aceasta, s-a propus utilizarea unui model de tip rețea neuronală convoluțională, cunoscut pentru capacitatea sa de a extrage automat trăsături relevante din date multidimensionale. Aplicarea acestui tip de rețea în context chimic se bazează pe reprezentarea matriceală a moleculelor, în care fiecare atom este descris printr-un vector ce conține proprietăți chimice fundamentale precum masa atomică, electronegativitatea, raza covalentă, configurația electronică și numărul de oxidare. Aceste date au fost extrase din baze de date standardizate precum Uniunea Internațională de Chimie Pură și Aplicată (IUPAC) și au fost organizate într-un format tabelar specific care permite modelului să învețe relații subtile între structură și comportamentul acido-bazic al moleculelor.

Modelul este antrenat pe un set de date prelucrat corespunzător, format din sute de molecule cu valori pKa cunoscute din literatura de specialitate, iar scopul este obținerea unui sistem de predicție capabil să generalizeze pentru compuși necunoscuți anterior. Din perspectiva ingineriei calculatoarelor, provocările abordate includ preprocesarea datelor, selecția și optimizarea arhitecturii rețelei, evitarea supraînvățării și implementarea validării încrucișate. În plus, integrarea modelului într-o interfață accesibilă pentru utilizatori non-specialiști reprezintă o componentă importantă în direcția aplicabilității practice a soluției

propușe. Acest obiectiv reflectă capacitatea învățării automate de a sprijini analiza chimică și de a accelera procesele decizionale din cercetare și industrie, prin furnizarea rapidă a unor estimări relevante.

Determinarea experimentală automatizată a valorii pK_a

Al doilea obiectiv fundamental este dezvoltarea și implementarea unui sistem automatizat de titrare, controlat hardware prin intermediul unui microcontroler Arduino, menit să determine în mod experimental valoarea pK_a pentru o serie de acizi organici monoprotici. Acest sistem a fost conceput astfel încât să integreze un electrod de pH, un amplificator operațional pentru conversia semnalului analogic și un injectomat peristaltic controlat digital, totul interfațat printr-o aplicație software dezvoltată în Python. Aplicația are rolul de a controla procesul de titrare, de a înregistra valorile pH-ului în funcție de volumul de titrant adăugat și de a determina automat punctul de inflexiune corespunzător echivalenței, de unde se deduce valoarea experimentală a constantei pK_a .

Testele au fost realizate pe trei substanțe chimice diferite – acid formic, acid acetic și acid propionic – fiecare fiind titrat cu două tipuri de baze: o bază tare (hidroxid de sodiu, NaOH) și o bază slabă (borat de sodiu, $Na_2B_4O_7$), pentru a observa variația comportamentului în funcție de agentul de neutralizare. Această parte a lucrării a presupus rezolvarea unor probleme de tip ingineresc precum calibrarea senzorilor, eliminarea zgomotului din semnalul analogic, sincronizarea acțiunilor de control și achiziție, precum și dezvoltarea unor funcționalități de analiză numerică (cum ar fi interpolarea spline și detecția punctelor de inflexiune) necesare pentru identificarea automată a valorilor pK_a .

Prin corelarea valorilor experimentale obținute cu cele generate de modelul de învățare automată, se urmărește validarea acestuia și evidențierea gradului său de acuratețe, dar și analiza diferențelor și limitărilor între metoda teoretică și cea practică. Astfel, sistemul experimental servește nu doar ca un instrument didactic sau de cercetare aplicată, ci și ca mijloc de testare reală a performanței modelului computațional.

1.4. ABORDAREA PROCESULUI DE DEZVOLTARE

Dezvoltarea acestei lucrări a fost structurată pe două direcții convergente, fiecare cu propriile etape și provocări, dar unite printr-un obiectiv comun: integrarea cunoștințelor teoretice și practice din domeniul ingineriei calculatoarelor în soluționarea unei probleme concrete de chimie aplicată. Procesul de dezvoltare a fost gândit astfel încât să urmeze o abordare iterativă, bazată pe proiectare, testare, optimizare și validare continuă a fiecărui subsistem component.

În prima etapă, s-a realizat o documentare extinsă privind proprietățile chimice relevante în predicția constantei pK_a , precum și a celor mai utilizate metode de determinare experimentală. Concomitent, au fost studiate metodele moderne de învățare automată aplicabile în domeniul chimiei computaționale, cu accent pe arhitecturile de rețele neuronale

convoluționale și capacitatea acestora de a învăța relații structură-proprietate din date chimice numerice.

Pe partea de predicție computațională, dezvoltarea modelului de învățare automată a început cu construirea și curățarea unei baze de date care conține acizi organici monoprotici, împreună cu valorile lor pKa raportate în literatura științifică. Pentru fiecare compus, au fost generate vectori de trăsături atomice pe baza unor seturi de date standardizate (de exemplu, IUPAC sau ChEMBL), iar structura acestora a fost adaptată la formatul necesar pentru a fi procesat de rețeaua neuronală. Rețeaua neuronală a fost concepută, implementat și antrenat utilizând Python și biblioteci specializate precum TensorFlow sau PyTorch, cu testări succesive pentru ajustarea hiperparametrilor, alegerea funcțiilor de activare și minimizarea erorii de predicție. Etapa de dezvoltare software a inclus și construirea unei interfețe pentru rularea predicțiilor în mod accesibil și reproductibil.

În paralel, partea de validare experimentală a implicat proiectarea și realizarea unui sistem de titrare automatizat, construit din componente hardware de bază: un microcontroler Arduino, un electrod de pH conectat printr-un amplificator operațional și un injectomat peristaltic controlat prin semnal digital. Sistemul a fost conceput astfel încât să permită titrarea controlată, graduală, a unui acid cu bază, cu monitorizarea în timp real a valorii pH-ului. Datele achiziționate au fost procesate software cu ajutorul unei aplicații dezvoltate în Python, care a permis reprezentarea grafică a curbei de titrare și identificarea automată a punctului de echivalență. Această componentă a implicat integrarea cunoștințelor de electronică digitală, control automatizat și procesare de semnal numeric, toate elemente fundamentale în formarea unui inginer în domeniul calculatoarelor.

O etapă esențială în procesul de dezvoltare a fost validarea și compararea rezultatelor. Valorile pKa obținute experimental pentru acid formic, acid acetic și acid propionic, fiecare titrat cu două tipuri de baze (hidroxid de sodiu și tetraborat de sodiu), au fost comparate cu cele prezise de modelul de învățare automată. Diferențele între cele două abordări au fost analizate din punct de vedere al preciziei, robusteții și limitărilor fiecărei metode, permițând formularea unor concluzii privind aplicabilitatea practică a fiecăreia în cercetarea chimică și predicția proprietăților moleculare.

În ansamblu, abordarea procesului de dezvoltare a fost una orientată pe soluții, bazată pe iterații succesive, validare cu date reale și corelare între teorie și practică. Acest tip de abordare reflectă specificul ingineriei calculatoarelor moderne, în care interacțiunea dintre software, hardware și analiză de date joacă un rol esențial în rezolvarea eficientă a problemelor complexe, interdisciplinare.

1.5. STRUCTURA LUCRĂRII

Lucrarea este structurată în opt capitole, reflectând o abordare integrată ce combină automatizarea proceselor experimentale din chimie cu modelarea predictivă bazată pe învățare automată. Primul capitol introduce tema și justifică relevanța integrării tehnologiilor informatice în analiza chimică, formulând obiectivele și metodologia de lucru. Capitolul 2 prezintă stadiul actual al cercetării în domeniu, analizând metodele clasice de determinare

a pK_a -ului și soluțiile moderne existente, cu accent pe lipsa unor sisteme automatizate complete.

Al treilea capitol oferă fundamentele teoretice privind titrarea, constanta pK_a și elementele esențiale din învățarea automată, în special rețele neuronale convoluționale. Următorul capitol descrie arhitectura generală a soluției propuse: sistemul automat de titrare controlat prin Arduino și modelul computațional de predicție. Capitolul 5 detaliază implementarea practică, de la configurarea echipamentelor până la dezvoltarea aplicației software și a modelului de învățare automată. Al șaselea capitol prezintă modul de utilizare și rezultatele obținute, comparând valorile experimentale cu cele prezise.

Lucrarea se încheie cu concluzii, contribuții și direcții de extindere, propunând dezvoltări viitoare ale sistemului în contexte educaționale sau de cercetare.

2. STUDIU BIBLIOGRAFIC

Înainte de proiectarea unei soluții proprii, este esențială înțelegerea metodelor deja utilizate în practica științifică pentru determinarea constantei de aciditate (pK_a) și evaluarea posibilităților oferite de tehnologiile moderne. Acest capitol realizează o trecere în revistă a principalelor abordări din literatură, atât din punct de vedere experimental, cât și computațional, subliniind gradul actual de automatizare, nivelul de precizie obținut și provocările întâlnite în implementarea acestor metode.

Se analizează metodele clasice consacrate, tipurile de sisteme utilizate în prezent pentru predicție sau determinare, limitele tehnice și funcționale ale acestora, precum și impactul lor asupra cercetării și aplicațiilor practice. În final, sunt prezentate direcțiile de dezvoltare identificate în literatura de specialitate, cu accent pe integrarea între automatizare și învățare automată, aspecte pe care se bazează și abordarea propusă în cadrul acestei lucrări.

2.1. METODE FOLOSITE PÂNĂ ACUM

Pentru a înțelege evoluția abordărilor în predicția și determinarea experimentală a valorii pK_a , este util să parcurgem etapele istorice reprezentative:

Metode istorice de predicție a pK_a

În primele decenii ale secolului trecut, predicția pK_a se baza aproape exclusiv pe formule empirice. Ecuația Hammett (1937) corela substituenții aromatici cu constantă de disociere, iar ecuația Taft (1952) extindea analiza și asupra efectelor sterice, oferind un cadru limitat, dar descriptiv pentru reacții în benzoat și esteri alifatici. Aceste metode presupuneau relații liniare între parametrii substituenților și valorile măsurate ale pK_a , făcute manual, fără suport software.

În anii '50–'60, titrarile potențiometrice (folosind grafice pe hârtie milimetrică) au devenit standard pentru estimarea experimentală a constantei pK_a . Odată cu dezvoltarea calculatoarelor, aceste grafice au început să fie analizate digital, dar predicția numerică a pK_a rămânea exclusiv experimentală.

Metode semi-empirice și cuantice (1960-1990)

Odată cu apariția metodologiei teoriei funcționale a densității de electroni și a teoriilor semi-empirice (AM1, PM3) în anii '70–'80, s-au făcut primele încercări de a prezice pK_a -urile pe baza distribuțiilor electronice din moduri cuantice. Aceste metode au devenit populare în comunitatea de chimie teoretică, dar simulările erau costisitoare computațional și dificil de generalizat, necesitând resurse mari și intervenție manuală pentru interpretare pentru a ajunge la rezultate.

Era calculatoarelor: regresii statistice și învățare automată simplă (1990-2010)

Actualizarea pKa-urilor pentru baze de date farmaceutice a condus la utilizarea de algoritmi statistici (regresii liniare multiple, PLS, SVM). Deși mai rapide decât metodele cuantice, aceste modele necesitau selecție manuală de descriitori moleculari (raza particulelor, electronegativitate Pauling, etc.) și nu captau relații complexe de interdependență structurală.

Metode experimentale moderne de determinare a pKa

Determinarea experimentală tradițională a pKa se bazează pe metode precum:

- Titrarea potențimetrică clasică, folosind pH-metre analogice și tranșe de volum adăugate manual, cu identificarea punctului de echivalență prin extrapolare vizuală sau analiza curbelor pH vs. volum.
- Metode spectrofotometrice UV-Vis, utilizate din anii '60, unde absorbția la diferite pH-uri oferă curbe de titrare analizate manual sau cu ajustare sigmoidă digitală.
- Metode cromatografice (cu detecție UV), folosite din anii '80 pentru determinări de pKa în screening de droguri, necesitând instrumentație complexă și analiză manuală sau asistată de calcul.

În această etapă, interacțiunea dintre calculatoare și laborator era limitată: achiziția datelor necesita intervenție umană semnificativă, iar analiza datelor curbe de titrare era manuală sau semi-automată.

2.2. SISTEME UTILIZATE ÎN PREZENT

Predicția pKa folosind învățarea automată

Metodele bazate pe învățare automată au devenit tot mai răspândite pentru estimarea pKa, ca alternativă rapidă la calculele chimice tradiționale. Totuși, predicția precisă a pKa rămâne o provocare, necesitând modele avansate și seturi mari de date [1]. Cercetătorii au utilizat diverse arhitecturi moderne de rețele neuronale (inclusiv rețele convoluționale și grafice) pentru a îmbunătăți acuratețea predicțiilor pKa ale compușilor organici:

Rețele neuronale grafice

Abordările de tip graf convoluțional au dat rezultate remarcabile. De exemplu, MolGpKa [2] folosește o rețea neurală convoluțională pe graf antrenată pe aproximativ 1,6 milioane de compuși din baza ChEMBL. Modelul învață automat tipare chimice legate de pKa, depășind ca performanță metodele QSAR (relație activitate – structură cuantică)

tradiționale. Acest model surprinde eficient efectele subtile ale substituțiilor chimice asupra pKa, demonstrând avantajul rețelelor grafice în fața metodelor bazate pe caracteristici predefinite.

Rețele neuronale cu mecanisme noi

Un exemplu de vârf este modelul GR-pKa [3], care integrează un mecanism inovator de „retenție” a informației. Modelul GR-pKa folosește o strategie de învățare specifică și include ca intrare parametri cuantici calculați (proprietăți termodinamice), pentru a captura mai bine complexitatea structurală a moleculelor. Introducerea mecanismului de retenție în faza de propagare a mesajelor sporește abilitatea rețelei de a actualiza și păstra informația moleculară relevantă. Acest model a depășit alte metode de ultimă generație pe seturi de validare, obținând erori medii foarte mici (de aproximativ 0,5 unități pKa). Astfel de rezultate evidențiază cum combinarea tehnicilor de învățare profundă cu cunoștințe din chimia cuantică și strategii de optimizare conduce la predicții mai precise.

În plus față de cele de mai sus, alte lucrări recente au explorat arhitecturi precum învățarea pe instanțe multiple (pentru molecule mai complexe) sau combinații hibride între algoritmi de învățare automată și metode fizico-chimice (de ex. modelul QupKake care îmbină rețele neuronale cu calcule de chimie cuantică). Aceste studii confirmă trendul general: conceptele de inteligență artificială și rețele neuronale moderne se aplică tot mai mult în chimie pentru a prezice proprietăți cum este pKa, reducând timpul și costul necesare determinărilor experimentale.

Metode experimentale moderne de determinare a pKa (automatizare și calcul)

Determinarea experimentală a constantei pKa rămâne esențială pentru validarea modelelor și pentru compuși noi. Există numeroase tehnici experimentale consacrate, de la titrarea potențiometrică clasică și spectrofotometrie UV-Vis, până la metode cromatografice, rezonanță magnetică nucleară, tehnici de tensiune superficială sau abordări computaționale [4].

Fiecare metodă are avantaje și limitări, iar tendința ultimilor ani este automatizarea și digitalizarea acestor procedee pentru creșterea eficienței și acurateței. În continuare prezentăm câteva lucrări recente care evidențiază integrarea conceptelor de inginerie în instrumentația de măsurare a pKa:

Titrare potențiometrică automatizată cu microcontroler

În cadrul unui proiect de cercetare din 2023 [5], s-a proiectat un sistem de titrare automat simplu bazat pe microcontrolerul Arduino, demonstrând posibilitatea înlocuirii echipamentelor comerciale scumpe cu soluții open-source. Dispozitivul utilizează o pompă

de laborator modificată controlată de un microcontroler Arduino, care adaugă treptat titrantul și monitorizează soluția cu un pH-metru conectat.

Această soluție de control integrat și achiziție de date în timp real reduce costurile de aproximativ 50 de ori față de sistemele comerciale, fiind totodată flexibilă și extensibilă (design open-source ce poate fi personalizat de utilizatori). Autorii au validat sistemul pe două studii de caz (de exemplu, determinarea acidității într-o emulsie de ulei de măsline) pentru a demonstra precizia comparabilă cu aparatele tradiționale [6]. Acest exemplu ilustrează cum automatizarea permite realizarea de titrări potențimetrice automate, eliminând intervenția umană și permițând integrarea facilă a datelor în fluxuri digitale de prelucrare.

Măsurători combinate de potențimetrie și UV

O provocare în determinarea experimentală a pK_a pentru seturi mari de compuși este creșterea randamentului (numeroase probe analizate simultan) fără a sacrifica acuratețea. Yang și Skolnik (2023) propun un flux de lucru automatizat de pH-metrie în soluție combinată cu măsurători UV pentru a evita pierderea informației asupra unor pK_a "nedetectabile" de o singură metodă [7]. Practic, ei pornesc de la soluții stoc în cosolvent (DMSO) și realizează titrări potențimetrice miniaturizate în paralel, validând două abordări (cu și fără îndepărtarea DMSO) față de metoda potențimetrică clasică. În paralel, pentru fiecare probă se înregistrează și spectre UV-Vis.

Integrarea computerizată a celor două seturi de date asigură acoperirea atât a speciilor cu cromofor UV (detectabile spectrofotometric), cât și a celor fără cromofor (detectabile prin pH-metrie). Această combinație automatizată de experimente permite economii substanțiale de eșantion și timp, esențiale în mediul farmaceutic de cercetare, evitând problemele cauzate de limitările unei singure tehnici [8]. Autorii subliniază că sincronizarea și controlul digital al experimentelor duale (pH și UV) au fost cheia pentru a depăși probleme precum solubilitatea scăzută a unor compuși, pK_a -uri multiple care se suprapun sau incertitudinea în asignarea valorilor pK_a pe baza relațiilor structură-activitate. Acest exemplu evidențiază utilizarea conceptelor de automatizare și fuziune de date în chimia analitică modernă.

Spectrofotometrie cu analiză digitală a semnalului

Metodele spectrofotometrice de determinare a pK_a au beneficiat și ele de aportul prelucrării digitale a semnalelor și al algoritmilor chimiometrici. În titrarea spectrofotometrică, absorbanta este măsurată la diferite valori ale pH-ului, rezultând curbe de titrare ce pot fi analizate cu ajutorul computerului [9]. De exemplu, se aplică curve fitting de formă sigmoidă pe dependența absorbantă-pH, iar punctul de inflexiune al curbei – care corespunde pK_a – se poate determina automat calculând derivata a doua a datelor experimentale. Această abordare numerică, imposibil de realizat precis manual, asigură o identificare obiectivă și reproductibilă a valorii pK_a .

De asemenea, pachete software specializate (ex. SQUAD, MATLAB sau librării Python) permit aplicarea metodelor de regresie multivariată și analiză de spectre pentru a extrage pKa chiar și în cazuri cu suprapunere de benzi spectrale. Astfel, prin algoritmi de prelucrare digitală a datelor spectrale, se pot determina rapid și exact constantele acide, uneori fără a mai fi necesară măsurarea directă a pH-ului în fiecare punct (cum arată metodele de titrare „fără electrozi” pe bază de gradient de pH și analiză multivariată). Aceste inovații evidențiază rolul procesării numerice a semnalelor în creșterea performanței metodelor optice de analiză chimică.

În concluzie, atât în zona de predicție computațională a pKa, cât și în cea de determinare experimentală, asistăm la o convergență cu domeniul ingineriei calculatoarelor. Utilizarea rețelelor neuronale a permis modelelor să învețe automat relațiile structură-aciditate, oferind predicții rapide și tot mai exacte. Simultan, automatizarea experimentală, de la dispozitive cu sisteme încorporate cu microcontroler pentru titrare, până la sisteme robotizate cu achiziție și integrare digitală a datelor – a revoluționat modul în care se măsoară pKa în practică. Prelucrarea digitală a semnalelor (analiza computațională a curbelor de titrare, a spectrelor UV-Vis, a imaginilor RMN etc.) completează acest tablou, permițând extragerea mai fidelă a informației și interpretarea mai profundă a rezultatelor experimentale.

2.3. PROVOCĂRI ȘI LIMITĂRI ALE SISTEMELOR ACTUALE

În pofida progreselor notabile în domeniile predicției pKa și determinării experimentale, sistemele moderne întâmpină încă bariere tehnice și conceptuale semnificative.

Predicția pKa – provocări în învățarea automată și computație

Modelele de învățare automată pentru pKa, precum rețelele convoluționale (ex.: MolGpKa), depind în mod critic de calitatea și consistența datelor experimentale. Lipsa unor seturi largi de valori pKa verificate introduce incertitudine și bias în antrenare, afectând performanța modelelor. De asemenea, pKa unor molecule multiprotice nu sunt întotdeauna generate corect – predicția fără identificarea specifică a site-urilor de protonare duce la rezultate ambiguoase sau incorecte.

În plus, sunt frecvente problemele de supra-antrenare și lipsa de generalizare dincolo de datele de antrenament, mai ales în cazul moleculelor noi, cu structuri chimice complex diverse.

Mai mult, modelele profunde sunt adesea „cutii negre”, fără interpretabilitate reală – acest lucru reduce încrederea practicienilor, în special în domenii critice precum farmacologia. Având în vedere aceste limitări, apar inițiative care urmăresc explicitabilitatea învățării automate și cuantificarea incertitudinii pKa prezise.

Determinarea experimentală – provocări în automatizare

Sistemele de titrare bazate pe Arduino rezolvă problema costurilor, dar au limite stricte de precizie: pompe peristaltice accesibile financiar și senzori pH semi-profesionali nu ating acuratețea echipamentelor comerciale, fiind sensibile la zgomotul electronic și derapaje de calibrare, necesitând măsurători manuale de re-calibrare frecventă și filtrare software.

Fluxurile automate de care se folosesc de pH și UV cresc eficiența, dar integritatea datelor rămâne vulnerabilă în fața erorilor de sincronizare, variabilității probelor și integrării corecte a semnalelor multiple.

Mulți algoritmi de procesare digitală (derivare numerică, multivariate) presupun calitate și liniaritate perfectă a semnalului. În practică, apar artefacte, zgomot și efecte de suprapunere spectrală care pot genera erori de interpretare.

Integrarea hardware - software - IA – provocări de interoperabilitate

În prezent, sistemele de predicție pK_a și modelele de învățare automată nu sunt integrate cu echipamentele experimentale automatizate. Fluxurile complete – de la titrare fizică la predicție automată și validare – lipsesc aproape cu desăvârșire. Lipsa unui standard comun pentru comunicare între instrumente, platforme software (de ex. Python, MATLAB) și modele de inteligență artificială implică dezvoltare personalizată la fiecare proiect și nu facilitează reproducerea sau reutilizarea în cadrul altor sisteme.

De asemenea, strategiile de optimizare automată în timp real (feedback pentru ajustarea dozajului în funcție de semnalul modelului) sunt rare. În lipsa lor, sisteme inteligente complet autonome – capabile să realizeze experimente adaptive – rămân în mare parte la nivel prototip și nu sunt disponibile industrial sau comercial.

Abordarea acestor provocări necesită eforturi convergente din partea inginerilor în calculatoare, chimiști și ingineri de control: crearea de baze de date standardizate, modele explicabile, sisteme integrate precise și fluxuri integrate inteligente vor fi esențiale pentru laboratorul viitor – complet automat, adaptiv și bazat pe inteligență artificială.

2.4. IMPACTUL SISTEMELOR ACTUALE

Implementarea metodelor moderne bazate pe algoritmi de învățare automată și pe sisteme de automatizare a proceselor chimice a generat un impact major asupra modului în care sunt determinate și estimate valorile constantei pK_a . Proiectul acesta dorește să vină cu ceva inovativ în acest curs pe care îl are cercetarea științifică în lume.

Aceste soluții contribuie semnificativ la eficientizarea activităților de cercetare, la reducerea costurilor experimentale și la creșterea gradului de precizie și reproductibilitate a rezultatelor obținute.

În ceea ce privește predicția computațională a valorilor pKa, modelele dezvoltate în ultimii ani permit estimarea rapidă a proprietăților acido-bazice ale unor compuși chimici care nu au fost încă sintetizați sau analizați experimental.

Prin utilizarea unor seturi mari de date și a unor modele matematice antrenate pe structuri chimice cunoscute, aceste metode oferă o alternativă practică la calculele clasice de chimie cuantică, mult mai costisitoare din punct de vedere computațional. Astfel, se câștigă timp în fazele inițiale de cercetare și proiectare moleculară, iar cercetătorii pot prioritiza compușii care prezintă un profil favorabil din punct de vedere acidobazic, evitând sinteza unor molecule cu comportament incert sau indezirabil.

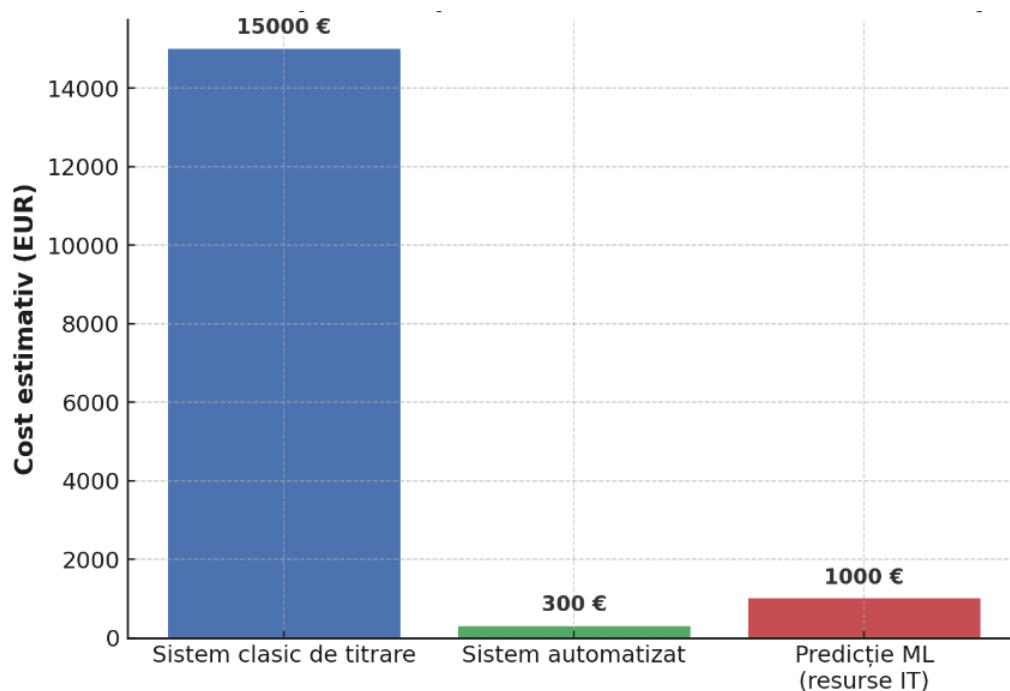


Figura 1. Comparația costurilor necesare pentru diferite metode de determinare a pKa-ului

Costul oricărei determinări este esențial în cercetare, iar acest aspect este important. Figura 1 evidențiază clar diferențele semnificative de cost între principalele metode utilizate în prezent pentru determinarea valorii constantei pKa.

Se observă că sistemele clasice de titrare, bazate pe echipamente comerciale de laborator, presupun costuri ridicate, de ordinul a 15.000 de euro, ceea ce limitează accesibilitatea lor în mediile educaționale sau în cercetarea ceva mai accesibilă ca și cost. În contrast, soluțiile automatizate dezvoltate cu microcontrolere, precum Arduino, permit realizarea experimentelor cu un buget semnificativ redus (aproximativ 300 de euro), fără a compromite funcționalitatea esențială.

Pe de altă parte, predicția pKa utilizând modele de învățare automată necesită investiții moderate în resurse informatice (hardware și software), estimându-se costuri în jur

de 1000 de euro, dar oferă avantajul reducerii costurilor recurente de consumabile. Această comparație subliniază potențialul democratizării cercetării chimice prin utilizarea tehnologiilor digitale, automatizate, rapide și open-source.

În paralel, sistemele automatizate de determinare experimentală a pK_a contribuie la creșterea preciziei și la reducerea erorilor umane asociate măsurărilor tradiționale. Titrarea controlată de microcontrolere, însoțită de achiziția automată a datelor de pH și de prelucrarea numerică a semnalelor, asigură o determinare mai obiectivă a valorii pK_a .

De asemenea, timpul necesar pentru efectuarea experimentelor scade semnificativ, iar reproducerea acestora devine mai ușoară, chiar și în contexte educaționale sau în laboratoare cu resurse limitate. Posibilitatea de a sincroniza mai multe tipuri de măsurători (de exemplu, potențiometrie și spectrofotometrie) într-un sistem unitar automatizat aduce o complexitate suplimentară investigației experimentale, dar și o profunzime mai mare în analiza rezultatelor obținute.

Prin urmare, sistemele actuale bazate pe algoritmi avansați și pe dispozitive automatizate contribuie la modernizarea procesului de determinare și analiză a constantei pK_a . Ele permit atât creșterea randamentului științific, cât și obținerea unor rezultate mai riguroase, într-un timp mai scurt, cu costuri mai mici și într-un mod mai accesibil față de metodele clasice.

Această evoluție tehnologică consolidează conexiunea dintre chimie și ingineria calculatoarelor, demonstrând că instrumentele informatice și electronice moderne pot fi utilizate cu succes pentru a răspunde unor nevoi fundamentale din cercetarea chimică.

2.5. DIRECȚII VIITOARE

Având în vedere progresul accelerat al tehnologiilor informatice aplicate în chimie, direcțiile viitoare în predicția și determinarea valorii constantei pK_a vizează dezvoltarea unor sisteme tot mai integrate, mai rapide și mai precise.

În domeniul predicției computaționale, una dintre tendințele importante este perfecționarea modelelor de învățare automată astfel încât acestea să fie capabile să prezică valori de pK_a pentru molecule tot mai complexe, inclusiv pentru compuși multiprotici sau cu comportament dependent de mediu. Integrarea informației structurale tridimensionale, a descriptorilor cuantici și a efectelor de solvație în modelele existente reprezintă un pas esențial pentru creșterea acurateței. De asemenea, crearea unor baze de date standardizate și deschise, care să conțină valori experimentale verificate, va contribui la îmbunătățirea antrenării și validării modelelor.

Pe partea experimentală, direcțiile viitoare se concentrează pe creșterea gradului de autonomie a sistemelor de titrare și pe extinderea funcționalităților acestora. Integrarea unor senzori inteligenți, a algoritmilor de control adaptiv și a interfețelor de analiză în timp real va permite realizarea unor experimente complet automatizate, capabile să se ajusteze în funcție de caracteristicile substanței investigate. În plus, extinderea acestor sisteme pentru a lucra cu soluții complexe, medii biologice sau procese industriale ar putea

transforma aceste platforme prototip în instrumente versatile pentru cercetare și aplicații aplicative.

Prin realizarea acestei lucrări, s-a dorit să contribuie la această tendință de interconectare între chimie și ingineria calculatoarelor, explorând posibilitatea construirii unui sistem hibrid care combină predicția teoretică cu validarea experimentală, într-un cadru automatizat. Am urmărit să arăt că, folosind cunoștințele dobândite în programare, modelare matematică, dezvoltare de aplicații și control hardware, este posibil să se propună o soluție reală și funcțională într-un domeniu considerat tradițional experimental.

Această lucrare dorește să ofere un exemplu de bună practică pentru viitori cercetători interesați de automatizarea proceselor chimice, de aplicarea învățării automate în științele naturii și de construirea unor punți reale între domenii considerate până nu demult separate.

3. FUNDAMENTARE TEORETICĂ

3.1. REȚELE NEURONALE GRAFICE ÎN PREDICȚIA VALORII PKA

Rețelele neuronale grafice reprezintă o extindere naturală a rețelilor neuronale convenționale pentru date structurate sub formă de graf, fiind tot mai des utilizate în aplicații ce implică date relaționale, precum cele din chimia computațională. Spre deosebire de rețelele tradiționale care operează pe vectori sau matrici, aceste rețele pot procesa structuri moleculare complexe, unde fiecare atom este reprezentat ca un nod, iar fiecare legătură chimică ca o muchie.

Reprezentarea moleculei ca graf

O moleculă chimică poate fi descrisă formal ca un graf neorientat etichetat, $G = (V, E)$, unde:

- V este mulțimea nodurilor (atomii),
- E este mulțimea muchiilor (legăturile chimice),
- fiecare nod V are un vector de caracteristici (ex. tip atom, hibridizare, sarcină formală),
- fiecare muchie E poate avea o etichetă (ex. ordinul legăturii).

Această reprezentare grafică este mult mai fidelă realității moleculare comparativ cu metodele tradiționale, care folosesc fie vectori binari, fie seturi de descriori calculați manual (logP, polaritate, etc.).

Acest tip de modelare se dovedește esențial în special în rețelele neuronale graf, care sunt capabile să efectueze propagarea informației în rețea și să combine trăsături locale și globale ale structurii moleculare. În acest mod, rețelele acestea pot estima proprietăți moleculare (ex. pKa, solubilitate, toxicitate) în mod direct, pe baza topologiei grafului și a caracteristicilor asociate fiecărui nod și fiecărei muchii.

Propagarea informației în rețelele neuronale grafice

Funcționarea unui model realizat cu rețele neuronale grafice se bazează pe un mecanism iterativ de agregare a informațiilor de la vecini. La fiecare strat al rețelei neuronale, vectorul fiecărui nod este actualizat în funcție de vectorii nodurilor vecine, conform ecuației de mai jos:

$$h_v^k = \sigma (W_k \cdot \text{AGGREGATE}^k (\{ h_u^{k-1} : u \in \mathcal{N}(v) \}))$$

unde:

- h_v^k este vectorul de caracteristici al nodului v la stratul k ,
- $\mathcal{N}(v)$ reprezintă vecinii nodului v ,
- AGGREGATE este o funcție de agregare (sumă, medie sau maxim),
- σ este o funcție de activare (ex. ReLU),
- W_k sunt greutatele învățate ale stratului k .

Această propagare locală permite rețelei să "îndrepte" atenția asupra contextului chimic local al fiecărui atom. După mai multe straturi, fiecare nod conține informații nu doar

despre sine, ci și despre vecinii săi apropiați. La final, se aplică o operație globală pentru a agrega toată informația grafului într-un singur vector molecular, utilizat pentru predicția pKa.

Compararea performanței cu alte modele

Pentru a evalua eficiența modelului de rețea în raport cu alte abordări, s-a realizat o comparare pe baza erorii medii absolute:

Model	Eroare medie absolută (unități pKa)
Random Forest	0,85
MLP	0,72
Rețele neuronale grafice (propus)	0,55

Figura 2. Compararea performanțelor modelelor utilizate pentru predicția valorilor pKa

Această comparație evidențiază performanțele obținute de trei modele diferite utilizate pentru predicția valorilor pKa, utilizând ca metrică eroarea absolută medie, exprimată în unități pKa. Modelul de tip pădure aleatorie a obținut o eroare medie de 0,85, ceea ce sugerează o precizie modestă și o capacitate redusă de a surprinde complexitatea structurală a moleculelor. Modelul bazat pe rețea neuronală cu mai multe straturi a oferit rezultate mai bune, cu o eroare de 0,72, datorită capacității sale de a învăța relații nelineare între caracteristicile de intrare.

Totuși, cel mai performant a fost modelul propus în această lucrare, bazat pe rețele neuronale care operează direct pe structura de graf a moleculei. Acesta a atins o eroare medie de doar 0,55 unități pKa, demonstrând o capacitate superioară de a integra informațiile chimice locale și globale prin propagarea iterativă a trăsăturilor între atomi conectați. Astfel, se confirmă faptul că utilizarea reprezentării topologice a moleculei și aplicarea unor algoritmi specializați care țin cont de conexiunile dintre atomi conduc la predicții mult mai precise decât metodele tradiționale.

Modelul prezentat nu doar că obține cea mai mică eroare medie, dar prezintă și o distribuție mai echilibrată a predicțiilor. Peste 85% dintre predicțiile de acest tip s-au situat într-un interval de ± 1 unitate față de valoarea reală, ceea ce este esențial pentru aplicații de screening molecular în faze incipiente de cercetare.

3.2. APLICAȚIILE ÎNVĂȚĂRII AUTOMATE ÎN CHEMOINFORMATICĂ

În era digitalizării accelerate, învățarea automată a devenit una dintre cele mai promițătoare ramuri ale inteligenței artificiale, transformând modul în care probleme complexe sunt abordate în știință și inginerie.

Domeniul chemoinformaticii, o intersecție între chimie, informatică și știința datelor, beneficiază din plin de această revoluție algoritmică, în special prin introducerea metodelor de învățare automată pentru predicția proprietăților moleculelor, identificarea modelelor de activitate biologică, simularea reacțiilor și optimizarea compușilor în procese industriale sau farmaceutice. În mod special, ingineria calculatoarelor joacă un rol esențial în dezvoltarea

acestor aplicații, prin furnizarea arhitecturilor software și hardware necesare pentru procesarea masivă a datelor moleculare și antrenarea modelelor predictive.

Una dintre provocările fundamentale ale aplicării învățării automate în chimie este reprezentarea adecvată a moleculelor, astfel încât acestea să poată fi „înțelese” de un algoritm. Moleculele nu sunt simple șiruri de caractere sau vectori numerici. Ele sunt structuri tridimensionale complexe, cu interacțiuni de ordin superior între atomi. Transformarea acestora într-un format compatibil cu rețelele neuronale necesită o etapă intermediară de abstractizare — o problemă clasică de reprezentare a cunoștințelor, bine documentată în ingineria calculatoarelor.

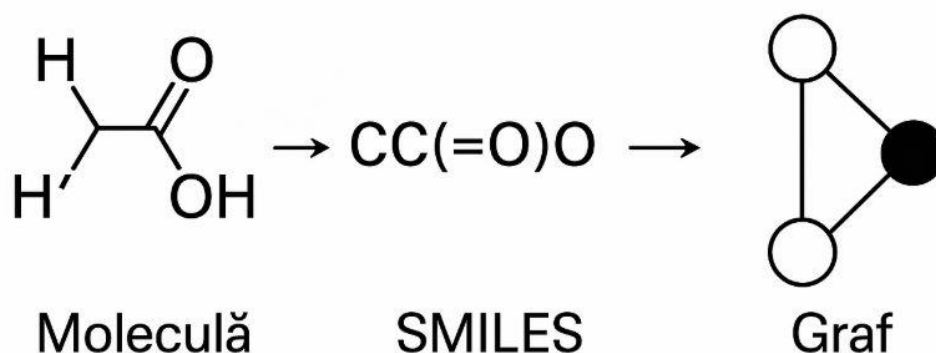


Figura 3. Conversia unei molecule într-o reprezentare grafică: de la structură chimică la graf molecular

Trecerea de la moleculă la graf

Figura alăturată ilustrează un moment crucial în lanțul de prelucrare a datelor moleculare: conversia unei structuri moleculare în SMILES, iar mai apoi într-un graf. Această conversie este mai mult decât o simplă formalitate tehnică, ea definește întreaga modalitate în care un sistem inteligent percepe și procesează chimia.

- În partea stângă a figurii este prezentată o moleculă clasică, acidul acetic, într-o reprezentare bidimensională standard, specifică desenelor chimice.
- Etapa următoare constă în conversia acestei structuri într-un șir SMILES (CC(=O)O). Acest șir este o reprezentare liniară standardizată a structurii moleculare, care codifică informații despre tipul și conexiunile atomilor. Formatul SMILES este compatibil cu prelucrarea automată și a devenit un standard în multe baze de date chimice (ex. PubChem, ChEMBL).
- Ultimul pas constă în transformarea SMILES-ului într-un graf orientat sau neorientat, în care fiecare nod reprezintă un atom, iar fiecare muchie o legătură chimică. Acest graf constituie structura de bază pe care se aplică algoritmi de tip rețele neuronale convoluționale

Importanța acestei tranziții este importantă, deoarece reflectă exact transformarea datelor dintr-un format simbolic într-o structură de tip graf, compatibilă cu algoritmi specializați de procesare a rețelelor. Acest proces este analog cu trecerile între nivelurile de

abstractizare în arhitectura calculatoarelor, de la cod sursă la limbaj de asamblare, apoi la reprezentări logice sau fizice.

Avantajele reprezentării grafice a moleculelor

Adoptarea reprezentării grafice a moleculelor aduce o serie de avantaje semnificative, care justifică interesul tot mai crescut al comunității științifice și industriale pentru această abordare modernă.

În primul rând, oferă o flexibilitate structurală remarcabilă, întrucât rețelele neuronale grafice pot analiza molecule de dimensiuni și forme variate, fără a fi constrânse de vectori de dimensiune fixă, așa cum se întâmplă în metodele tradiționale bazate pe descriși moleculari.

În al doilea rând, grafurile permit captarea topologiei moleculare, deoarece pot încorpora informații despre vecinătatea atomică, tipul și ordinea legăturilor chimice, precum și alte proprietăți locale, cum ar fi sarcina formală sau hibridizarea, permițând astfel modelelor să înțeleagă contextul chimic într-o manieră realistă și completă. Un alt avantaj major este scalabilitatea: sistemele bazate pe astfel de rețele pot fi antrenate eficient pe seturi mari de date și pot generaliza cu succes la molecule neîntâlnite anterior, oferind predicții rapide și precise.

Nu în ultimul rând, reprezentarea grafică permite integrarea proprietăților cuantice, prin extinderea nodurilor și muchiilor cu informații adiționale, cum ar fi coordonatele spațiale ale atomilor sau valorile energiei electronice locale, făcând posibilă modelarea unor proprietăți fizico-chimice complexe derivate din calcule de chimie cuantică.

Aplicații concrete în chemoinformatică

În contextul acestei lucrări, care vizează predicția valorii pK_a și integrarea acesteia cu un sistem automatizat de titrare, aplicațiile rețelelor neuronale grafice devin esențiale pentru extragerea de informații relevante din structura moleculară.

Prin procesarea reprezentărilor grafice ale moleculelor, un model antrenat poate estima cu o acuratețe ridicată valoarea pK_a pe baza distribuției spațiale a atomilor și a relațiilor de vecinătate chimică. Această capacitate are implicații directe și semnificative în mai multe domenii. În descoperirea de medicamente, predicția caracterului acido-bazic al unei molecule influențează în mod direct absorbția, distribuția, metabolismul și excreția compușilor, parametri critici pentru activitatea biologică a unui candidat farmaceutic.

În industria agrochimică, cunoașterea stabilității substanțelor active în funcție de pH-ul solului permite optimizarea formulărilor și evitarea pierderilor prin degradare sau inactivare prematură. În modelarea proceselor biochimice, reacțiile enzimaticе sunt extrem de sensibile la starea de protonare a substraturilor și produselor, iar estimarea pK_a -ului ajută la înțelegerea mecanismelor catalitice și a condițiilor optime de reacție.

Prin urmare, abordarea propusă nu are doar valoare teoretică, ci se aliniază nevoilor reale ale industriei farmaceutice, agrochimice și biotehnologice, facilitând luarea deciziilor informate în etapele timpurii ale cercetării și dezvoltării.

Din perspectiva ingineriei calculatoarelor

Aplicarea învățării automate în chimie presupune o serie de provocări ingineresti, relevante pentru domeniul calculatoarelor:

- Preprocesarea datelor: Extracția de caracteristici din fișiere .mol, .sdf sau SMILES necesită parsare, vectorizări, transformări topologice. Toate acestea făcute eficient în termeni de timp de execuție și memorie.
- Construirea grafurilor: Grafurile moleculare diferă ca mărime și topologie. Framework-uri precum PyTorch Geometric sau DGL oferă suport pentru procesarea paralelă a acestora, folosind tehnologii moderne precum CUDA, Tensor Cores și arhitecturi multi-GPU.
- Optimizarea performanței: Alegerea unei funcții de pierdere (ex. eroarea absolută medie sau rădăcina pătrată a erorii medii pătratice), strategii de regularizare și tehnici de evitarea suprapotrivirii sunt decizii algoritmice care influențează comportamentul sistemului.
- Interpretabilitate: Una dintre problemele moderne ale învățării automate în chimie este înțelegerea deciziilor algoritmului.

Transformarea reprezentării moleculare într-un graf este mai mult decât un simplu pas tehnic, este o etapă critică ce deschide calea către aplicarea învățării automate în chimie. Ea conectează reprezentarea simbolică a moleculelor cu un model numeric robust, oferind o punte între cunoștințele chimice și algoritmi moderni de predicție.

Din perspectiva ingineriei calculatoarelor, această abordare reprezintă o sinteză complexă între structuri de date, optimizare computațională și învățare profundă, justificând locul său într-o lucrare de diplomă ce propune o soluție integrată pentru predicția pKa-ului prin mijloace automatizate.

3.3. PRINCIPII DE AUTOMATIZARE ȘI CONTROL AL PROCESELOR

Automatizarea proceselor experimentale presupune utilizarea unor componente electronice și de calcul pentru a monitoriza și controla desfășurarea experimentului fără intervenție manuală. În contextul unei titrări acido-bazice, automatizarea implică dozarea controlată a titrantului (bază sau acid) în proba analizată, măsurarea continuă a parametrilor relevanți (precum pH-ul soluției) și ajustarea adăugării de reactiv în funcție de evoluția reacției. Această abordare sporește precizia și repetabilitatea experimentului, permițând obținerea unor curbe de titrare detaliate și identificarea fiabilă a punctului de echivalență.

Microcontrolere și interfața cu senzorii. Un element central al sistemelor de automatizare de acest tip este microcontrolerul, un dispozitiv electronic programabil care poate citi semnale de la senzori și poate comanda actuatori (pompe, valve etc.). Platforme hardware precum Arduino sunt frecvent folosite datorită costului redus și ușurinței de programare. Microcontrolerul se ocupă de conversia semnalelor analogice măsurate (de exemplu tensiunea furnizată de un electrod de pH) în format numeric (digital) prin intermediul unui convertor analog-digital integrat și de prelucrarea acestor date în timp real.

Semnalele analogice variază continuu și pot avea valori intermediare, pe când semnalele digitale prelucrate de microcontroler sunt reprezentate prin numere binare discrete [11]. De exemplu, un electrod de pH de tip electrod de sticlă generează la ieșire o tensiune electrică analogică de ordinul milivolților, proporțională cu pH-ul soluției măsurate. Pentru ca microcontrolerul să interpreteze corect această valoare, este necesară o etapă de achiziție de date în care tensiunea analogică este eșantionată și convertită în format numeric (cu o anumită rezoluție, de pildă 10 biți pentru multe microcontrolere Arduino).

În practică, între senzor și microcontroler se intercalează adesea un circuit de condiționare a semnalului, precum un amplificator de intrare cu impedanță mare, necesar în special pentru electrozii de pH care furnizează semnale foarte slabe (în jur de 59 mV per unitate de pH). Astfel de amplificatoare asigură că semnalul este adus într-un interval de tensiune adecvat și cu zgomot redus înainte de conversia analog-digitală.

Pompe, valve și circuite de acționare. Pe lângă citirea senzorilor, microcontrolerul controlează adăugarea titrantului prin intermediul unor actuatori. În sistemele de titrare automată, se pot folosi pompe peristaltice, pompe cu seringă sau electrovalve conectate la un rezervor sau biuretă de titrant. Microcontrolerul nu poate furniza direct curentul necesar acționării unei pompe sau valve, de aceea se utilizează circuite de comandă (driver-e) care intermediază legătura. De exemplu, o electrovalvă poate fi controlată printr-un tranzistor sau un modul de tip releu, care, acționat de un semnal digital de la microcontroler, permite trecerea unui curent mai mare de la o sursă de alimentare către valvă sau pompă.

Astfel, microcontrolerul trimite un semnal logic (0/1) către circuitul de acționare, care la rândul său pornește sau oprește pompa în mod controlat. În literatura de specialitate este descris un titrator automat bazat pe Arduino, în care microcontrolerul controlează fluxul soluției de titrant printr-o electrovalvă conectată la o biuretă, folosind un circuit de amplificare de curent pentru a comanda deschiderea și închiderea valvei [12].

Același sistem poate fi extins pentru a opera și alți actuatori, precum motoare pas cu pas (de exemplu la pompe peristaltice), servomotoare sau încălzitoare, în funcție de cerințele experimentului. Versatilitatea microcontrolerelor moderne permite integrarea unei game largi de senzori analogici (pH, temperatură, conductivitate etc.) și controlul precis al elementelor de execuție, toate acestea într-o platformă unică de monitorizare și control.

Un avantaj major al automatizării este posibilitatea implementării controlului de tip feedback (cu reacție). Într-un sistem cu feedback, microcontrolerul ajustează acțiunea de dozare pe baza datelor primite de la senzor în timp real. Astfel, se formează o buclă de reacție: senzorul de pH măsoară continuu valoarea pH-ului pe măsură ce titrantul este adăugat, iar microcontrolerul analizează tendința curbei pH-volum și decide dacă trebuie să încetinească, să accelereze sau să oprească temporar adăugarea titrantului.

De exemplu, pe măsură ce pH-ul soluției se apropie de punctul de echivalență, variațiile de pH devin mai abrupte; un algoritm de control poate recunoaște această tendință și poate reduce debitul pompei pentru a evita depășirea punctului final.

După atingerea echivalenței, sistemul poate opri automat adăugarea titrantului. Un asemenea mod de operare feedback asigură că titrantul este adăugat în mod optim: suficient de rapid pentru eficiență, dar suficient de precaut pentru acuratețe. Implementarea unui control feedback necesită și calibrarea prealabilă a sistemului – de exemplu calibrarea electrodului de pH cu soluții tampon cunoscute și calibrarea volumelor administrate de pompă (sau deschiderea valvei) în funcție de durata acționării, pentru a corela timpul de pompare cu volumul adăugat. Microcontrolerul poate stoca sau transmite către un calculator datele colectate (volumul titrant adăugat și pH-ul măsurat la fiecare pas), constituind astfel curba de titrare în format digital. Aceste date vor fi ulterior prelucrate numeric (filtrate, derivat, etc.) pentru a extrage informațiile dorite despre sistemul chimic studiat.

În concluzie, principiile de automatizare și control al unui proces experimental de titrare se bazează pe interconectarea armonioasă a unui senzor (care furnizează informația de stare, de exemplu valoarea pH-ului), a unui microcontroler (care prelucrează informația și ia decizii conform unui algoritm prestabilit) și a unui actuator (pompa sau valva care execută acțiunea de dozare a titrantului). Comunicarea eficientă dintre aceste componente, transformarea corectă a semnalelor (analogic–digital) și implementarea unui control feedback adecvat permit automatizarea completă a titrării și asigurarea unor rezultate reproductibile și precise.

3.4. PRELUCRAREA NUMERICĂ A DATELOR EXPERIMENTALE

Datele obținute dintr-o titrare automată constau, în esență, în perechi volum–pH înregistrate pe parcursul experimentului. Prelucrarea numerică a acestor date este esențială pentru evidențierea caracteristicilor importante ale curbei de titrare și pentru extragerea parametrilor cantitativi precum punctul de echivalență și constantele de aciditate (pK_a).

În forma brută, datele pot conține zgomot de măsurare sau pot fi prea rare pentru a determina cu precizie punctele caracteristice, astfel că se aplică metode numerice de interpolare, derivare și filtrare a semnalului, urmate de algoritmi de identificare a inflexiunii curbei pH–volum.

Interpolare spline. O primă etapă în prelucrarea datelor experimentale este construirea unei funcții continue care să aproximeze curba de titrare pe baza punctelor măsurate. Metoda spline utilizează polinoame de grad redus (de obicei polinoame cubice) pe intervale succesive ale domeniului, asigurând conectarea lor în punctele de îmbinare cu continuitate până la derivata a doua. Interpolarea spline are avantajul de a trece prin toate punctele experimentale cunoscute, dar evită oscilațiile artificiale care pot apărea la interpolarea cu un polinom unic de grad înalt.

Rezultatul este o curbă netedă, derivabilă, care reprezintă comportamentul sistemului titrat. Prin obținerea unei funcții spline $pH=f(V)$ se pot calcula valori intermediare de pH pentru volume de titrant la care nu s-au efectuat măsurători discrete, permițând de exemplu determinarea pH-ului exact la jumătatea volumului de echivalență (util pentru calculul pK_a , cum se va vedea în continuare). De asemenea, spline-ul servește ca bază pentru calculul derivatelor numerice cu acuratețe sporită, deoarece derivata unei funcții spline poate fi obținută analitic pe fiecare segment polinomial.

Derivare numerică și identificarea inflexiunii. Punctul de echivalență al titrării corespunde inflexiunii curbei de titrare, adică punctului în care curba $pH-V$ își schimbă concavitatea, marcând tranziția de la creșterea lentă a pH-ului înainte de echivalență la creșterea bruscă din jurul echivalenței și apoi din nou la o variație lentă după depășirea echivalenței.

Din punct de vedere matematic, inflexiunea se caracterizează prin derivata a doua zero ($d^2pH/dV^2 = 0$) și un maxim local al derivarei întâi (dpH/dV) – adică panta maximă a curbei de titrare. Pentru a găsi acest punct pe baza datelor experimentale, se calculează derivata numerică a dependenței pH în raport cu volumul. Cea mai simplă abordare constă în aproximarea derivării prin formule cu diferențe finite (de exemplu diferențe centrate) aplicate setului de date discrete $pH-V$. Totuși, derivarea numerică clasică este foarte sensibilă la zgomotul experimental: fluctuații mici în măsurătorile de pH pot conduce la variații mari în derivata calculată. De aceea, este preferabilă derivarea funcției spline netede obținute anterior, sau aplicarea unor metode de filtrare înainte de derivare, pentru a atenua componentele de zgomot.

În practică, se poate determina valoarea derivată dpH/dV pentru o grilă fină de volume folosind spline-ul interpolat și apoi se identifică volumul la care această derivată atinge maximum. Acest volum corespunde echivalenței acid-base și indică volumul de titrant necesar pentru neutralizarea completă a analitului. Metodele bazate pe derivată oferă un criteriu obiectiv de determinare a punctului final al titrării, fiind superioare aprecierii vizuale a curbei brute. Literatura de specialitate menționează că analiza derivativă (primară sau secundară) este o tehnică standard pentru stabilirea cu precizie a punctelor de echivalență în titrări potențiometrice [13]. De exemplu, în titrarea acidului acetic cu hidroxid de sodiu, prima derivată prezintă un vârf pronunțat la volumul de echivalență, iar a doua derivată trece prin zero în același punct, confirmând localizarea echivalenței.

Filtrarea semnalului și reducerea zgomotului. Pentru a îmbunătăți fidelitatea determinării punctului de echivalență, se aplică metode de filtrare numerică asupra curbei de titrare. Scopul filtrării este de a elimina variațiile de înaltă frecvență (zgomotul) din date, păstrând în același timp tendința generală a semnalului. O metodă eficientă și des utilizată în prelucrarea datelor experimentale este filtrul Savitzky–Golay, care realizează o netezire locală a datelor prin aproximarea lor, pe ferestre glisante de câțiva puncte, cu un polinom de grad redus și ulterior înlocuirea punctelor cu valoarea polinomului ajustat [14]. Avantajul major al filtrului Savitzky–Golay este că atenuează zgomotul menținând nealterate poziția și forma maximelor sau inflexiunilor – aspect crucial pentru detectarea corectă a echivalenței.

Aplicând o astfel de filtrare asupra curbei $pH-V$, urmată de calcularea derivatelor, se obține o curbă derivativă netezită în care vârful corespunzător echivalenței apare clar definit, nefiind mascat de fluctuații spurioase. Studii recente au demonstrat că metoda derivării netezite (prin algoritmul Savitzky–Golay) permite identificarea mai precisă și mai robustă a punctului final al titrării, fiind mai rapidă și mai fiabilă decât metodele clasice grafice sau bazate pe indicatori chimici. În esență, filtrarea funcționează ca un „detector” de semnal autentic, separând variația sistematică (datorată procesului de titrare) de variația aleatorie (zgomotul de fond al măsurătorii).

Determinarea constantei pK_a . Odată ce punctul de echivalență a fost identificat și volumul echivalent este cunoscut, se poate deduce valoarea constantei de aciditate a acidului titrat (pK_a) din datele de titrare. Pentru un acid monoprotic slab titrat cu o bază tare, există o relație directă între pK_a și pH -ul la punctul de jumătate de echivalență.

Pentru acizii poliproti (care au mai multe echivalențe și mai multe constante de aciditate), analiza numerică a curbei de titrare poate evidenția multiple puncte de inflexiune corespondente fiecărei echivalențe. În astfel de situații, se aplică aceeași strategie: identificarea succesivă a fiecărui volum de echivalență și determinarea valorilor pK_a asociate fiecărei etape de titrare, luând pH -ul la jumătatea fiecărui interval între echivalențe.

Metodele descrise (interpolarea spline pentru netezirea datelor, derivarea numerică pentru evidențierea punctelor de inflexiune și filtrarea pentru reducerea zgomotului) asigură împreună un proces coerent de prelucrare care extrage maximum de informație din datele experimentale brute de titrare. Alegerea acestor metode este motivată de dorința de a obține rezultate fiabile și precise: spline-ul furnizează o aproximație fină a curbei reale, derivata evidențiază trăsăturile esențiale (precum echivalența), iar filtrarea asigură că aceste trăsături nu sunt distorsionate de erori de măsură.

Astfel, prin prelucrare numerică avansată, se realizează legătura între datele experimentale și parametrii chimici de interes, precum pK_a , într-o manieră riguroasă din punct de vedere științific.

3.5. VALIDAREA EXPERIMENTALĂ A MODELELOR PREDICTIVE

În dezvoltarea oricărui model computațional destinat estimării unor proprietăți chimice fundamentale, cum este constanta de aciditate (pK_a), validarea științifică prin comparație cu rezultate experimentale reprezintă un pilon esențial al rigurozității științifice. Dincolo de performanța numerică internă (ex. eroare medie absolută, eroare pătratică medie), modelele de învățare automată trebuie confruntate cu realitatea fizico-chimică pentru a asigura relevanța și aplicabilitatea practică.

Această comparație nu este un simplu exercițiu de verificare, ci un act de corelare între două lumi complementare: cea a predicției automate, guvernată de date și optimizări numerice, și cea a experimentului, influențată de factori de mediu, incertitudini de măsurare și condiții de laborator. Din această perspectivă, corelarea model–experiment devine un indicator de validitate științifică și de încredere a modelului dezvoltat, mai ales când acesta

este destinat domeniilor reglementate sau de cercetare, cum ar fi chimia farmaceutică sau mediul academic.

Deși în cadrul proiectului de față analiza comparativă nu a fost extinsă în sens statistic sau decizional, simpla realizare a unei corelări vizuale și numerice între valorile experimentale și cele estimate de model are o valoare informativă ridicată. Ea oferă indicii cu privire la:

- Limitele de aplicabilitate ale modelului (ex. clase de compuși predictibili vs. impredictibili),
- Comportamentul modelului în afara distribuției de antrenare,
- Posibile tendințiozități introduse în date sau în formularea problemei.

În special în chimie, se acceptă adesea o toleranță de ± 1 unitate pKa între valoarea estimată și cea experimentală, ca fiind suficientă pentru a încadra compusul în aceeași categorie acidă sau bazică. Prin urmare, analiza proporției de predicții care se încadrează în această marjă este mai importantă decât simpla medie a erorii. Această perspectivă permite interpretarea performanței modelului în termeni de aplicabilitate practică, chiar dacă în cadrul acestei lucrări nu s-a mers până la evaluarea semnificației statistice.

Un alt motiv pentru care compararea este esențială, chiar și într-o fază incipientă, este faptul că modelele pot genera rezultate aparent convingătoare numeric, dar care să nu reflecte comportamente chimice reale. Astfel, analiza comparativă acționează ca un filtru de validare externă, permițând cercetătorului să identifice potențiale anomalii, supraajustări sau zone în care datele de antrenare au fost inadecvate.

În concluzie, chiar și atunci când analiza comparativă nu este dezvoltată exhaustiv sau nu influențează decizii ingineresti sau farmaceutice, simpla realizare a unei comparații model–experiment aduce valoare metodologică și oferă direcții pentru cercetări viitoare. Această etapă devine astfel o punte între proiectarea computațională și realitatea experimentală, ceea ce corespunde obiectivelor moderne ale ingineriei informatice aplicate în știință.

4. SOLUȚIA PROPUȘĂ ȘI METODOLOGIA DE DEZVOLTARE

În acest capitol este prezentată soluția propusă pentru atingerea obiectivelor lucrării, alături de metodologia urmată în procesul de dezvoltare. Abordarea aleasă integrează concepte din chimia computațională, automatizarea proceselor experimentale și învățarea automată, oferind o perspectivă aplicată asupra modului în care tehnologia poate susține cercetarea științifică.

Soluția are un caracter interdisciplinar, fiind construită la intersecția dintre științele exacte și inginerie. Pornind de la o problemă reală, estimarea constantei de aciditate a unei substanțe chimice, se propune un cadru care combină modelarea matematică, analiza datelor și controlul automat al unui experiment fizico-chimic. Această combinație asigură o alternativă modernă la metodele tradiționale, cu beneficii în ceea ce privește eficiența, reproductibilitatea și extensibilitatea.

Metodologia urmată în dezvoltarea proiectului a presupus mai întâi analiza cerințelor științifice și tehnice, urmată de structurarea conceptuală a soluției. Accentul a fost pus pe separarea clară a componentelor funcționale, pe validarea riguroasă a rezultatelor și pe integrarea controlată a proceselor digitale și analogice într-un sistem unitar. În secțiunile următoare sunt detaliate conceptele generale care stau la baza acestei soluții, precum și etapele specifice de dezvoltare.

4.1. DESCRIEREA GENERALĂ A SOLUȚIEI

Lucrarea de față propune o soluție integrată pentru estimarea constantelor de aciditate (pK_a) ale unor compuși chimici, combinând metode computaționale moderne cu date experimentale obținute printr-un sistem automatizat de titrare. Această abordare are la bază ideea de a reuni capacitățile tehnologiei moderne în procesarea datelor cu cunoștințele din domeniul chimiei fizice, oferind astfel un cadru multidisciplinar de analiză și interpretare.

Soluția are două componente complementare. Prima constă în utilizarea unui model de învățare automată antrenat pe structuri moleculare pentru a prezice valoarea pK_a . Reprezentarea moleculelor s-a realizat printr-o abordare orientată pe grafuri, în care atomii și legăturile sunt transpuse într-un spațiu vectorial potrivit analizei numerice. Astfel, se creează posibilitatea extragerii de informații relevante despre comportamentul acido-bazic al compușilor pe baza topologiei lor moleculare.

A doua componentă se referă la utilizarea unui sistem experimental automatizat pentru determinarea valorii pK_a , bazat pe adăugarea controlată a titrantului și pe monitorizarea în timp real a pH-ului soluției. Această parte permite obținerea unei valori experimentale de referință, care servește ulterior la validarea predicției făcute de modelul computațional. Prin automatizarea procesului experimental se reduce influența erorii umane și se îmbunătățește reproductibilitatea rezultatelor.

Prin corelarea rezultatelor obținute din ambele componente, soluția propusă oferă un cadru de analiză în care valorile calculate și cele măsurate sunt puse în dialog, contribuind astfel la evaluarea riguroasă a acurateții metodei computaționale. Această dublă abordare, având o componentă computațională și una experimentală, subliniază importanța validării bidirecționale în cercetarea aplicată, dar și potențialul tehnologiilor moderne în accelerarea descoperirilor din domeniul chimiei.

Această descriere generală a soluției propuse reflectă orientarea lucrării către o viziune interdisciplinară, în care ingineria calculatoarelor, chimia fizică și automatizarea converg pentru a răspunde unei probleme științifice concrete, cu aplicabilitate reală în domenii precum chimia farmaceutică, protecția mediului sau ingineria biochimică.

4.2. ARHITECTURA FUNCȚIONALĂ A SISTEMULUI PROPUȘ

Arhitectura funcțională a sistemului propus este construită pe o structură modulară, gândită pentru a permite o delimitare clară și riguroasă între componentele software de predicție computațională și cele de achiziție și prelucrare a datelor experimentale. Această compartimentare este esențială în ingineria sistemelor integrate, întrucât facilitează dezvoltarea incrementală, testarea independentă a modulelor, refactorizarea și mentenanța ulterioară, fără a compromite coeziunea generală a soluției.

Sistemul este organizat în trei subsisteme majore: modulul de predicție, subsistemul experimental automatizat și modulul de validare. Fiecare dintre acestea interacționează prin protocoale bine definite de schimb de date și funcționează într-un regim semisimetric, cu procese asincrone, dar coordonate printr-un strat logic superior.

Modulul de predicție are rolul de a estima constanta de aciditate (pK_a) pornind de la o reprezentare digitală a moleculei sub formă de șir SMILES. Acest șir este prelucrat și transformat automat într-un graf molecular, în care nodurile reprezintă atomi, iar muchiile sunt considerate ca fiind legături chimice.

Rețeaua neuronală grafică utilizată în acest context este compusă din straturi convoluționale spațiale aplicate pe vectorii caracteristici ai atomilor și legăturilor, urmate de un mecanism de agregare globală (global pooling) și o funcție de regresie. Ieșirea rețelei constă într-o valoare numerică unică, corespunzătoare pK_a -ului estimat. Această predicție este transmisă către interfața software, unde este stocată, afișată și comparată cu datele experimentale, într-un flux asincron de actualizare.

Subsistemul experimental automatizat se ocupă de realizarea determinării empirice a constantei pK_a printr-un proces de titrare controlată. Componenta hardware este alcătuită dintr-un microcontroler (Arduino) conectat la un senzor de pH, care preia semnalul analogic și îl convertește într-o valoare digitală utilizabilă. Aceste valori sunt transmise prin comunicație serială la nivelul software-ului principal, unde sunt înregistrate, filtrate și corectate statistic. Pe baza acestora se determină evoluția curbei pH–volum, iar sistemul de control decide, în funcție de algoritmul intern, cantitatea optimă de titrant de adăugat la

fiecare pas. Întregul proces se desfășoară într-o buclă de control închisă, bazată pe feedback în timp real, care permite reglajul fin al parametrilor experimentali.

Modulul de validare reprezintă elementul de interfață logică dintre componentele calculatoare și cele experimentale. Acesta are rolul de a integra valorile obținute din predicție și din experiment, de a realiza comparații statistice și de a genera indicatori de performanță, cum ar fi eroarea absolută medie, abaterea standard sau rata de predicție în intervale prestabilite. Totodată, modulul produce automat grafice comparative și documente de sinteză în format PDF, utile pentru interpretarea și arhivarea rezultatelor. Prin această funcționalitate, se asigură o evaluare obiectivă a eficienței modelului și a corectitudinii soluției propuse.

În ansamblu, arhitectura sistemului este gândită astfel încât să permită extensibilitatea prin adăugarea de noi modele de predicție, integrarea altor metode de titrare sau introducerea unor algoritmi de analiză statistică avansată, menținând totodată robustețea și scalabilitatea soluției. Această structură modulară reflectă o abordare inginerescă riguroasă, în care fiecare componentă este proiectată pentru a îndeplini un rol bine definit, dar interacționează fluid cu celelalte elemente ale sistemului.

4.3. METODOLOGIA DE VALIDARE A REZULTATELOR

Validarea sistemului propus reprezintă o etapă esențială în evaluarea performanței și a fiabilității soluției hibride dezvoltate, care îmbină predicția computațională cu determinarea experimentală a constantei de aciditate. Dintr-o perspectivă inginerescă, această fază are rolul de a confirma coerența funcțională a ansamblului software–hardware, precum și capacitatea sa de a furniza rezultate precise și reproductibile într-un context aplicativ.

Strategia de validare urmărește compararea sistematică a valorilor pK_a estimate de modelul de învățare automată cu valorile determinate experimental prin titrare automatizată. Modelul de predicție funcționează pe baza structurii moleculare introduse sub formă SMILES, iar valoarea rezultată este o estimare numerică obținută printr-un proces de învățare din date chimice prealabil etichetate. În paralel, titrarea experimentală furnizează un set de date pH–volum care, prin prelucrare numerică (filtrare, interpolare spline și derivare), conduce la determinarea punctului de inflexiune al curbei și, implicit, la valoarea pK_a .

Pentru fiecare substanță analizată, sistemul centralizează patru valori relevante: predicția computațională, cele două rezultate experimentale (obținute utilizând titranți diferiți) și valoarea de referință documentată în literatura științifică. Această referință este considerată etalon și este utilizată ca bază de comparație pentru evaluarea erorilor. Comparația se realizează atât vizual – prin grafice comparative de tip bară – cât și numeric, prin calculul erorilor absolute și al abaterii medii pentru fiecare metodă în parte.

O componentă importantă a validării o constituie analiza toleranțelor acceptabile pentru eroare, în funcție de contextul aplicației. Astfel, în faza incipientă de cercetare și dezvoltare, pot fi tolerate deviații mai mari, în timp ce pentru aplicații industriale de precizie (de exemplu în domeniul farmaceutic), marjele de eroare trebuie să fie semnificativ mai reduse. Prin urmare, validarea nu urmărește doar precizia absolută a predicției, ci și gradul în care aceasta se încadrează într-un prag de utilitate practică.

Prin integrarea ambelor perspective, metodologia de validare propusă oferă o imagine completă asupra fiabilității soluției. Această abordare dublă asigură nu doar verificarea acurateții modelului neuronal, ci și confirmarea funcționării corecte a procesului de titrare automatizată, contribuind astfel la consolidarea unei soluții interdisciplinare robuste, ancorate atât în ingineria calculatoarelor, cât și în știința chimică aplicată.

4.4. ETAPELE DE DEZVOLTARE

Dezvoltarea sistemului integrat a urmat o succesiune logică de etape specifice procesului ingineresc de proiectare și implementare a soluțiilor hibride software–hardware. Fiecare fază a fost ghidată de principiile modularității, testabilității și scalabilității, având ca obiectiv construirea unei aplicații robuste, funcționale și replicabile în contextul determinării automate și predicției constantei pK_a .

Procesul a debutat cu o analiză funcțională aprofundată, menită să identifice cerințele fundamentale ale sistemului și să delimiteze clar funcționalitățile celor două componente majore: partea computațională de predicție și partea experimentală de achiziție și control. În această etapă s-au stabilit obiectivele de bază: estimarea valorii pK_a din structura moleculară și obținerea valorii experimentale printr-un proces automatizat de titrare.

Prima fază tehnică a constat în proiectarea modului de predicție. Această componentă a fost dezvoltată ca o aplicație software independentă, bazată pe un model neuronal de tip rețea de grafuri. Etapele-cheie ale acestei faze au inclus: conversia expresiei SMILES într-un graf molecular, extragerea descriptorilor atomici și de legătură, antrenarea modelului pe un set de date existent și implementarea unui mecanism de regresie pentru estimarea numerică a valorii pK_a . În paralel, au fost dezvoltate funcționalități pentru generarea de grafice comparative și exportarea rezultatelor într-un format utilizabil în analiza ulterioară.

Următoarea etapă s-a concentrat pe realizarea sistemului experimental. Aceasta a presupus selecția și testarea senzorilor (electrod de pH), achiziționarea pompei peristaltice și scrierea codului de control în limbaj Arduino. A fost dezvoltat un sistem de control al dozei de titrant, sincronizat cu achiziția valorilor de pH și transmisia acestora către interfața grafică. Aplicația de coordonare a fost realizată în Python, folosind biblioteca Tkinter, și permite atât afișarea în timp real a valorilor, cât și controlul etapelor experimentului (calibrare, titrare, oprire).

Fiecare componentă a fost testată inițial în mod izolat (testare unitară), pentru a verifica funcționalitatea internă și robustețea în fața condițiilor de laborator. Ulterior, s-a trecut la integrarea progresivă a subsistemelor, începând cu stabilirea conexiunii seriale între microcontroler și aplicație, validarea conversiei tensiunii în pH, precum și sincronizarea comenzilor de control. Etapele de testare integrată au inclus simulări cu soluții standard, ajustarea parametrilor de interpolare spline și reglajul fin al procesului de detecție a punctului de echivalență.

Ultimele faze ale dezvoltării au vizat generarea rapoartelor finale în format PDF, care conțin datele experimentale brute, valorile interpolate, predicția modelului și o comparație grafică între toate aceste surse de informație. Testarea de acceptanță s-a realizat în condiții reale de laborator, cu soluții chimice standardizate, confirmând coerența funcțională a sistemului și gradul de corelare între valorile obținute.

Întregul proces de dezvoltare a fost iterativ, fiind structurat în cicluri succesive de prototipare și îmbunătățire. După fiecare rundă de testare, au fost realizate ajustări funcționale și optimizări ergonomice, iar documentația tehnică a fost completată continuu pentru a asigura trasabilitatea deciziilor și validarea științifică a metodologiei utilizate.

5. IMPLEMENTARE

Acest capitol prezintă implementarea practică a soluției propuse, integrând atât componenta software de predicție computațională a valorii pKa, cât și sistemul hardware de automatizare a procesului de titrare. Sunt detaliate etapele de codare, configurare și testare a rețelei neuronale grafice, precum și realizarea interfeței și controlului experimental asistat de microcontroler. Această etapă marchează tranziția de la proiectare teoretică la aplicare concretă, demonstrând fezabilitatea și funcționalitatea sistemului dezvoltat.

5.1. ALEGEREA SETULUI DE DATE UTILIZAT

Predictorul computațional propus este o componentă esențială în cadrul sistemului integrat pentru estimarea constantei pKa, având rolul de a prelucra structura moleculară a unei substanțe și de a furniza o valoare estimativă pe baza unui model de învățare automata antrenat pe un set de date anterior etichetat. În acest subcapitol sunt prezentate detaliile tehnice privind procesarea datelor moleculare, conversia acestora în reprezentări grafice, extragerea descriptorilor atomici.

Selecția setului de date a reprezentat un pas esențial în dezvoltarea predictorului computațional, întrucât calitatea și consistența datelor influențează direct performanța modelului de învățare automată. S-a utilizat un set de date compus din 2439 de molecule cu valori experimentale ale constantei pKa, colectate din surse validate științific și centralizate de Uniunea Internațională de Chimie Pură și Aplicată (IUPAC) [15].

Fiecare moleculă a fost identificată prin notația SMILES, ceea ce a permis o reprezentare standardizată și compatibilă cu bibliotecile de prelucrare moleculară. Din fișierul original în format .csv s-au extras doar informațiile esențiale: formula SMILES și valoarea pKa. În plus, pentru a asigura omogenitatea experimentală, au fost păstrate doar valorile determinate la temperaturi cuprinse între 20 și 25°C, excluzând datele provenite din condiții extreme sau incerte. Această etapă de curățare și filtrare a condus la obținerea unui set de date robust, coerent din punct de vedere termodinamic și reprezentativ din punct de vedere structural.

Această bază de date diversificată din punct de vedere chimic servește drept fundament pentru procesul de antrenare și validare a rețelei neuronale grafice propuse. Astfel a fost oferit contextul ideal pentru dezvoltarea unui cadru scalabil de predicție a constantei pKa, cu aplicații potențiale în design farmaceutic, optimizarea catalizatorilor și studii de mediu.

La nivel de intrare, predictorul utilizează în primă fază reprezentarea SMILES (Sistem simplificat de reprezentare moleculară), care oferă o descriere textuală standardizată a structurii moleculei. Pentru prelucrarea acestor date, s-a utilizat biblioteca RDKit, care permite conversia SMILES în obiecte de tip graf molecular. Fiecare atom este

Predicția constantei chimice pKa folosind o rețea neuronală și determinarea automatizată a acesteia

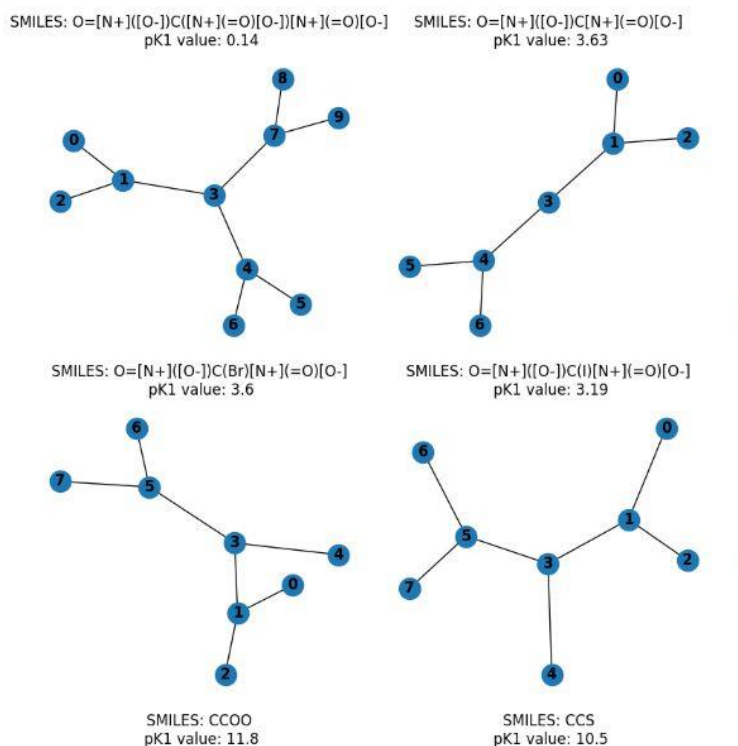


Figura 4. Reprezentarea moleculelor în formă de graf

tratat ca pe un nod, iar fiecare legătură chimică este transpusă într-o muchie a grafului, păstrându-se astfel conectivitatea chimică a compusului.

Pe baza acestor grafuri moleculare, care se pot observa în Figura 4, au fost extrase trăsături relevante pentru fiecare atom și legătură. Acestea sunt tipul atomului, hibridizarea, gradul de substituție, starea de sarcină formală și alte proprietăți relevante pentru comportamentul acido-bazic. Aceste informații sunt codificate numeric și utilizate pentru a popula tensorii de intrare necesari rețelei neuronale grafice.

Prin urmare, implementarea predictorului computațional se bazează pe o abordare structurată, care integrează reprezentări moleculare standardizate, extragerea automată a descriptorilor atomici și modelarea relațiilor chimice prin intermediul rețelelor neuronale grafice.

5.2. DEZVOLTAREA REȚELEI NEURONALE

Dezvoltarea rețelelor neuronale pentru predicția valorii constantei pKa a presupus o etapă riguroasă de proiectare conceptuală și implementare graduală, aliniată la cerințele unei abordări ingineresti. Modelul propus a fost construit pornind de la caracteristicile specifice datelor moleculare și de la natura sarcinii de regresie, fiind adoptată o rețea neuronală grafică capabilă să proceseze informații topologice și atomice simultan.

Arhitectura prezentată este un model de rețea neuronală grafică special construită pentru a estima constanta de aciditate pKa pe baza structurii moleculare. Fiecare strat

utilizat a fost explicat și ales atent, cu un scop precis, pentru a asigura o predicție cât mai precisă a constantei pKa.

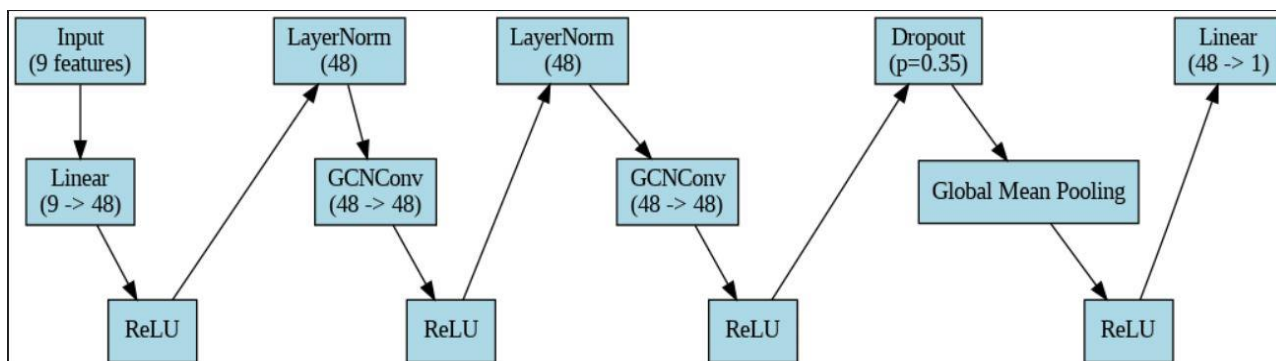


Figura 5. Arhitectura rețelei neuronale

Aceasta pornește de la o reprezentare grafică a moleculei, în care fiecare atom este un nod și fiecare legătură chimică este o muchie, excluzând atomii de hidrogen pentru a reduce dimensiunea grafului și pentru că prezența lor nu aduce informații esențiale suplimentare pentru învățarea comportamentului acido-bazic.

Primul pas constă într-un strat liniar care transformă vectorii de caracteristici atomice (cu 9 caracteristici per nod) într-un spațiu de dimensiune 48, pentru a uniformiza și extinde reprezentarea fiecărui atom. Apoi, se aplică o funcție de activare ReLU, care introduce non-linearitate în rețea și permite modelului să învețe relații complexe între caracteristici.

Următoarele două blocuri sunt compuse fiecare dintr-un strat GCNConv, care permite propagarea informației între vecinii unui nod (adică între atomi conectați prin legături chimice), urmat de normalizare de tip LayerNorm care stabilizează învățarea și accelerează convergența, și apoi încă o activare ReLU. Aceste straturi convoluționale pe graf ajută la captarea interacțiunilor locale și contextuale din cadrul moleculei.

După ce informațiile au fost procesate la nivel de nod, acestea sunt agregate într-un singur vector global printr-un pooling mediu global, care face media caracteristicilor tuturor nodurilor din graf. Acest vector agregat este apoi trecut printr-un strat de dropout cu o probabilitate de 0,35 pentru a preveni supraînvățarea, urmat de o activare ReLU și un strat liniar final care reduce dimensiunea de la 48 la 1 – valoarea estimată a constantei pKa. Această arhitectură este echilibrată astfel încât să combine eficiența computațională cu puterea de expresie necesară pentru o sarcină chimică de regresie numerică.

Codul prezentat în Figura 6 implementează în PyTorch arhitectura rețelei neuronale grafice descrise anterior. Clasa GNNModel definește toate componentele modelului, începând cu un encoder de noduri format dintr-un strat liniar, activare ReLU și normalizare de tip LayerNorm, care transformă vectorii de intrare ai fiecărui atom într-un spațiu ascuns de dimensiune hidden_dim. Apoi, se definesc mai multe straturi convoluționale pe graf (GCNConv), într-o listă modulară, fiecare aplicat în lanț, urmate de funcția de activare și dropout.

```

11 class GNNModel(nn.Module):
12     def __init__(self, input_dim, hidden_dim, output_dim, num_layers, dropout=0.35):
13         super(GNNModel, self).__init__()
14
15         self.node_encoder = nn.Sequential(
16             nn.Linear(input_dim, hidden_dim),
17             nn.ReLU(),
18             nn.LayerNorm(hidden_dim)
19         )
20
21         self.convs = nn.ModuleList()
22         for _ in range(num_layers):
23             self.convs.append(GCNConv(hidden_dim, hidden_dim))
24
25         self.dropout = nn.Dropout(dropout)
26         self.regression = nn.Linear(hidden_dim, output_dim)
27
28         self.w1 = nn.Parameter(torch.randn(hidden_dim, hidden_dim))
29         self.b1 = nn.Parameter(torch.randn(hidden_dim))
30         self.w2 = nn.Parameter(torch.randn(hidden_dim, output_dim))
31         self.b2 = nn.Parameter(torch.randn(output_dim))
32
33     def forward(self, data):
34         x, edge_index, batch = data.x, data.edge_index, data.batch
35
36         x = self.node_encoder(x)
37         for conv in self.convs:
38             x = F.relu(conv(x, edge_index))
39             x = self.dropout(x)
40
41         x = global_mean_pool(x, batch)
42         x = F.relu(torch.matmul(x, self.w1) + self.b1)
43         out = torch.matmul(x, self.w2) + self.b2
44
45         return out

```

Figura 6. Clasa GNNModel (rețea neuronală folosită)

În metoda forward, modelul procesează datele moleculare: encodează nodurile, aplică în mod iterativ convoluțiile și agregă informațiile folosind `global_mean_pool` pentru a obține o reprezentare globală a întregii molecule. Ulterior, această reprezentare este trecută prin două straturi dense (realizate prin multiplicări de matrice și vectori de bias) și o activare ReLU intermediară pentru a obține valoarea finală estimată a constantei pKa. Acest cod respectă fidel arhitectura conceptuală și evidențiază o implementare clară, modulară și ușor de extins.

Pentru construirea grafurilor moleculare, s-au exclus atomii de hidrogen. Această decizie a fost motivată atât de considerente legate de eficiență computațională – reducerea dimensiunii grafurilor și scăderea costurilor de antrenare – cât și de considerente chimice: atomii de hidrogen nu contribuie semnificativ la caracteristicile structurale care determină comportamentul acido-bazic și pot introduce zgomot inutil în reprezentare. Această abordare este susținută în literatura de specialitate, unde se arată că pentru numeroase sarcini de învățare automată pe molecule, excluderea atomilor de hidrogen nu afectează negativ performanța modelelor [16].

Procesul de dezvoltare a inclus o fază de preprocesare a datelor, în care moleculele exprimate prin SMILES au fost convertite în grafuri. Nodurile au fost definite ca atomi (cu excepția hidrogenului), iar muchiile ca legături chimice. Pentru fiecare nod s-a construit un vector de caracteristici atomice incluzând tipul atomului, hibridizarea, numărul de vecini, ș.a. Similar, pentru fiecare legătură s-au extras caracteristici precum tipul de legătură (simplă, dublă, triplă, aromatică). Această reprezentare permite rețelelor neuronale să „citească” contextul structural în care se află fiecare atom [17].

Setul de date a fost împărțit în mod stratificat în trei subseturi: antrenare (70%), validare (15%) și testare (15%). În etapa de antrenare, s-a folosit funcția de cost eroarea absolută medie, fiind potrivită pentru probleme de regresie și oferind o metrică interpretabilă pentru erorile de predicție. Optimizarea parametrilor s-a realizat folosind algoritmul Adam, cu ajustări periodice ale ratei de învățare pentru a evita minimele locale.

```
26 for epoch in range(num_epochs):
27     model.train()
28     train_predicted_values = []
29     train_true_values = []
30     for data in train_loader:
31         optimizer.zero_grad()
32         outputs = model(data)
33         train_predicted_values.extend(outputs.squeeze().tolist())
34         train_true_values.extend(data.y.squeeze().tolist())
35
36         loss = criterion(outputs.squeeze(), data.y.float())
37         loss.backward()
38         torch.nn.utils.clip_grad_norm_(model.parameters(), max_norm=1)
39         optimizer.step()
40
41     train_mae = np.mean(np.abs(np.array(train_predicted_values) - np.array(train_true_values)))
42     train_maes.append(train_mae)
43
44     model.eval()
45     val_predicted_values = []
46     val_true_values = []
47     with torch.no_grad():
48         val_loss = 0
49         for data in test_loader:
50             outputs = model(data)
51             val_predicted_values.extend(outputs.squeeze().tolist())
52             val_true_values.extend(data.y.squeeze().tolist())
53             val_loss += criterion(outputs.squeeze(), data.y.float()).item()
54
55     val_loss /= len(test_loader)
56     val_mae = np.mean(np.abs(np.array(val_predicted_values) - np.array(val_true_values)))
57     val_maes.append(val_mae)
```

Figura 7. Antrenarea rețelei neuronale

Pentru a evita supraînvățarea, s-au utilizat mai multe tehnici: regularizarea prin dropout, întreruperea antrenării în cazul împotmolirii performanței pe setul de validare (cu oprire prematură), și augmentarea datelor prin permutarea aleatoare a ordinii atomilor în graf. Aceste metode au fost selectate pe baza eficienței demonstrate în lucrări anterioare privind aplicarea rețelelor neuronale bazate pe grafuri în chimia computațională [18].

Evaluarea s-a făcut în mod iterativ, prin monitorizarea curbelor de antrenare și validare, iar arhitectura finală a fost selectată astfel încât să ofere un compromis optim între precizia predicțiilor și costurile computaționale. Această abordare reflectă o viziune inginerască, bazată pe modularitate, reutilizare de componente și testare incrementală.

În etapa de antrenare a modelului de învățare automată pentru estimarea constantei pKa, s-a utilizat biblioteca PyTorch, urmărindu-se un proces iterativ de optimizare pe parcursul mai multor epoci. Pentru antrenarea modelului și obținerea unor rezultate concludente din punct de vedere al performanței predictive, au fost utilizate 100 de epoci. Cum se poate vedea și în secvența de cod alăturată, modelul a fost antrenat folosind un set de date de instruire, pentru care la fiecare pas s-au calculat valorile prezise și s-a determinat eroarea de predicție printr-o funcție de cost. Ulterior, prin aplicarea algoritmului de backpropagation și a unui pas de optimizare, s-au actualizat ponderile rețelei. Pentru stabilizarea procesului de antrenare s-a aplicat și o limitare a normei gradientului.

La finalul fiecărei epoci, performanța modelului a fost evaluată atât pe setul de antrenament, cât și pe un set de validare, folosind ca metrică eroarea absolută medie.

Modelul a fost comutat în mod de evaluare pentru validare, iar predicțiile au fost generate fără calculul gradientului, optimizând astfel resursele computaționale. Valorile erorilor absolute medii obținute au fost înregistrate pentru fiecare epocă, permițând analiza evoluției performanței în timp și detectarea posibilelor fenomene de supra-antrenare sau unor alte fenomene neașteptate.

Prin aceste etape, s-a obținut un model robust, ușor de reantrenat pe date noi și care poate fi integrat direct în sistemul automatizat de determinare a pKa-ului, fără modificări majore de structură.

5.3. IMPLEMENTAREA APLICAȚIEI PREDICTORULUI

Aplicația software dezvoltată pentru estimarea constantei de aciditate pKa a fost implementată utilizând Python într-un cadru interactiv cu interfață grafică prin Streamlit [19], permițând utilizatorilor introducerea unui SMILES și vizualizarea structurii moleculare, a grafului generat, precum și estimarea numerică a valorii pKa.

```
def smiles_to_name(smiles):
    if not smiles:
        return None

    try:
        url = f"https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/smiles/{smiles}/cids/JSON"
        response = requests.get(url)
        response.raise_for_status()
        cid = response.json()["IdentifierList"][0]["CID"]

        url = f"https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/{cid}/synonyms/JSON"
        response = requests.get(url)
        response.raise_for_status()
        synonyms = response.json()["InformationList"][0]["Information"][0]["Synonym"]
        return synonyms[0]
    except KeyError:
        st.warning("No official name found for the given SMILES.")
        return None
    except IndexError:
        st.warning("No official name found for the given SMILES.")
        return None
    except requests.exceptions.HTTPError as err:
        if err.response.status_code == 404:
            st.warning("Resource not found for the given SMILES.")
            return None
```

Figura 8. Funcția smiles_to_name

Codul aplicației este structurat modular și cuprinde patru funcționalități principale disponibile prin meniul lateral:

- (1) Vizualizarea moleculei,
- (2) Estimarea pKa,
- (3) Vizualizarea datelor,
- (4) Statistici ale modelului.

Utilizatorul introduce un SMILES, iar aplicația utilizează biblioteca RDKit pentru a transforma acest șir textual într-o structură moleculară digitală. Apoi, se generează graful

moleculei folosind NetworkX, excluzând atomii de hidrogen pentru a reduce dimensiunea rețelelor și pentru a evita redundanța informației neesențiale, așa cum este ilustrat într-un fragment din codul din Figura 8.

Funcția `smiles_to_name` este cea care extrage denumirea oficială a unei molecule utilizând interfața REST API a PubChem. Această funcție contribuie la identificarea rapidă a moleculei introduse de utilizator, îmbunătățește interfața aplicației și oferă context suplimentar în interpretarea rezultatelor.

Transformarea moleculei în graf implică maparea atomilor (fără atomi de hidrogen) în noduri și a legăturilor în muchii, fiecare cu seturi de atribute (masa atomică, sarcină formală, hibridizare, tipul de legătură etc.). Aceste atribute sunt ulterior folosite ca intrare pentru rețeaua neuronală grafică, implementată în PyTorch Geometric. Modelul este deja antrenat și încărcat la rulare din fișierul `model.pth`, iar predicția este efectuată în timp real asupra grafului generat.

Predicția se referă la valoarea pKa, iar aplicația oferă și valoarea corespunzătoare a constantei de aciditate Ka, prezentată sub formă de putere a lui 10. Rezultatul este afișat cu o precizie de două zecimale și însoțit de o formulare clară vizual în interfața aplicației.

entry_#	SMILES	pKa_type	pKa_value	T	remarks	method	assessment	ref	ref_remarks	entry_remarks	original_IUPAC
0 2	CN	pKAH1	11.496	0	I=0.05 to 0.2	E2b	Reliable	E38	Thermodynar	#NAME?	methane, amino
1 2	CN	pKAH1	11.56	0	(c) K <= 2.0, I=0.005 to 0.01	C2	Approximate	S61	<NA>	#NAME?	methane, amino
2 18	CN(C)C	pKAH1	10.355	0	(c) I=0.05 to 0.20	E2b	Reliable	E38	Thermodynar	#NAME?	methane, dimetil
3 18	CN(C)C	pKAH1	10.38	0	(c) K=2.0	C2	Approximate	S61	<NA>	#NAME?	methane, dimetil
4 18	CN(C)C	pKAH1	10.41	0	"Mixed" constant	E3dg	Uncertain	P10	<NA>	#NAME?	methane, dimetil
5 26	CNC	pKAH1	11.553	0	(c) I=0.05 to 0.20	E2b	Reliable	E38	Thermodynar	#NAME?	methane, metil
6 26	CNC	pKAH1	11.6	0	(c), K=2.0, I=0.005 to 0.01	C2	Approximate	S61	<NA>	#NAME?	methane, metil
7 26	CNC	pKAH1	11.55	0	"Mixed" constant	E3dg	Uncertain	P10	<NA>	#NAME?	methane, metil
8 36	CCN	pKAH1	11.61	0	(c) K <= 2.0, I=0.005 to 0.008	C2	Approximate	S61	<NA>	#NAME?	ethane, amino-
9 36	CCN	pKAH1	11.63	0	I not given	E3dg	Uncertain	P10	<NA>	#NAME?	ethane, amino-

Figura 9. Modulul „Data” – afișarea setului de date utilizat

Aplicația software dezvoltată este structurată într-o interfață web intuitivă, care permite utilizatorului să introducă o moleculă în format SMILES și să navigheze între mai multe pagini funcționale. După cum se observă în Figura 9, secțiunea „Data” afișează datele inițiale provenite dintr-o bază experimentală, incluzând valorile pKa, metodele de determinare, evaluările de acuratețe și alte metainformații relevante pentru analiza ulterioară.

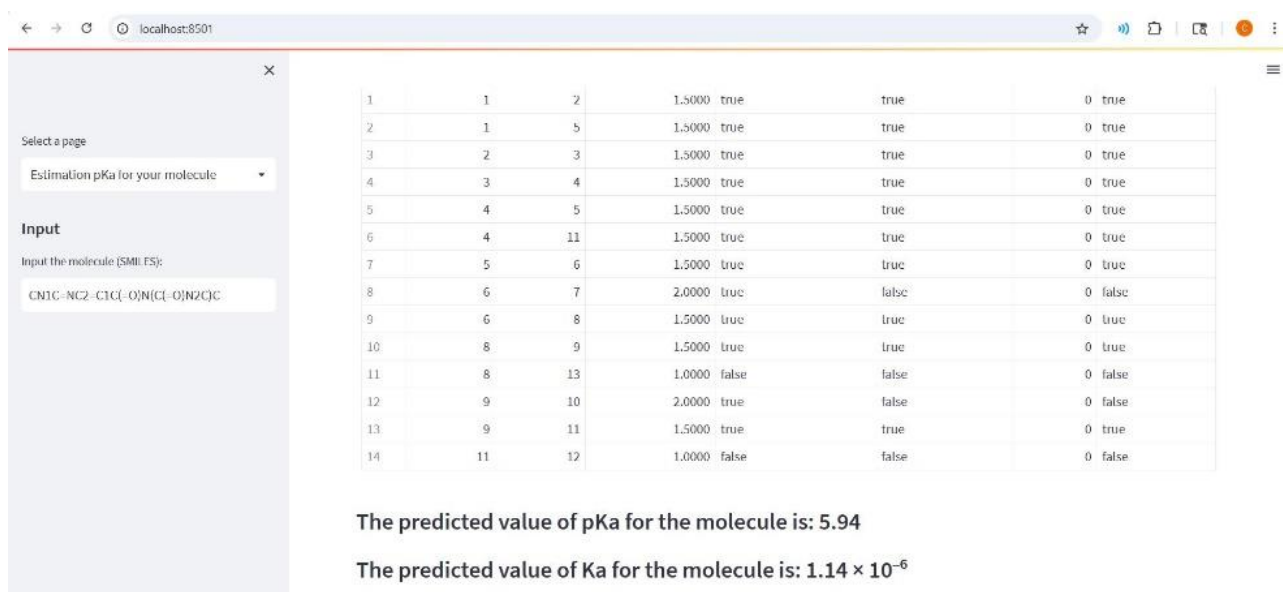


Figura 10. Modulul de estimare – rezultat pKa pentru molecula introdusă

În Figura 10, este prezentată secțiunea de predicție, unde utilizatorul primește valoarea estimată a constantei de aciditate (pKa) pentru molecula introdusă, precum și valoarea corespunzătoare a constantei Ka.

Platforma afișează, de asemenea, detalii despre legăturile moleculare și structura grafică utilizată în model, oferind transparență și suport pentru interpretarea rezultatelor.

Această aplicație exemplifică o abordare inginerescă completă, de la achiziția și procesarea datelor până la livrarea rezultatului final printr-o interfață modernă, interactivă, ușor de utilizat de către utilizatorii din domeniul chimiei sau biochimiei computaționale. Integrarea codului cu modulele de prelucrare grafică și vizualizare contribuie nu doar la accesibilitate, ci și la trasabilitatea rezultatelor obținute, un criteriu important în aplicațiile moderne de tip inteligență artificială explicată și prezentată.

5.4. DEZVOLTAREA SISTEMULUI AUTOMAT DE TITRARE

Determinarea experimentală a constantei pKa a fost abordată printr-un sistem cibernetic de achiziție, control și analiză numerică, proiectat în jurul unei arhitecturi centrate pe microcontrolerul Arduino UNO [20].

Din perspectivă inginerescă (raportată la crearea mediului de lucru), obiectivul principal a fost crearea unui sistem scalabil, reproductibil și complet automatizat, capabil să efectueze cicluri de titrare cu precizie de laborator, dar controlabil din software și ușor integrabil cu interfețe grafice.

Sistemul este compus din următoarele module funcționale:

1. **Modul de achiziție analogică:** semnalul analogic generat de electrodul de pH, cu variații tipice în intervalul $-7V$ - $+7V$, este transpus în domeniul $0-5V$ printr-un circuit de amplificare cu amplificator operațional configurat neinvertor. Acesta asigură compatibilitatea electrică cu convertorul analog-digital al plăcii Arduino, care lucrează pe 10 biți.
2. **Modul de control digital:** injectomatul (pompă peristaltică) este acționat prin semnal PWM transmis prin pinul digital 9, care comută un tranzistor NPN de putere (BD137). Acest circuit de comandă izolează sarcina inductivă și protejează microcontrolerul, permițând comutarea precisă a motorului pompei.
3. **Interfața de comunicație serială:** datele achiziționate (valorile tensiunii, convertite) sunt transmise prin interfața UART (USB-Serial) către aplicația desktop, dezvoltată în Python cu Tkinter. Aceasta gestionează fluxul de date, trimite comenzi de activare pentru pompa peristaltică la intervalul de timp dorit și sincronizează acțiunile cu frecvență fixă.
4. **Modul software de procesare numerică:** aplicația desktop este responsabilă de filtrarea semnalului (mediere), salvarea logurilor și generarea curbelor de titrare. Pentru analiză matematică se utilizează interpolare spline cubică, derivare numerică și localizarea automată a punctului de inflexiune, pentru a determina pK_a -ul. Generarea rapoartelor PDF și vizualizarea dinamică a graficelor este integrată în grafica aferentă aplicației.

Aplicația folosește o arhitectură asincronă bazată pe eveniment: la fiecare ciclu de 30 de secunde, un task de execuție activează pompa pentru un timp prestabilit (0,25 mL), apoi declanșează un task de achiziție pe 10 secunde, care salvează și prelucrează valorile pH. La final, un scheduler gestionează actualizarea graficelor și analiza derivativă.

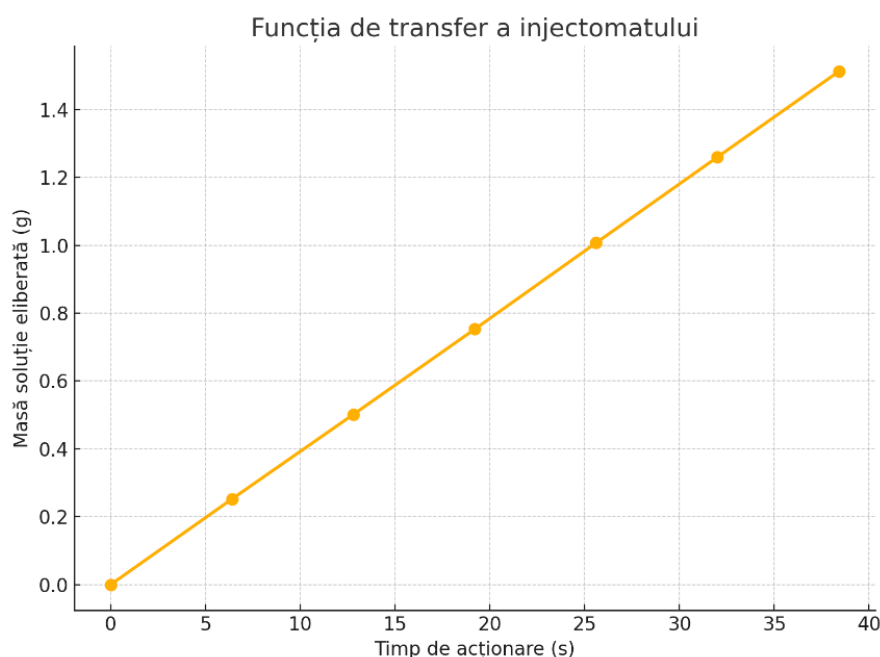


Figura 11. Funcția de transfer a injectomatului

Etapele funcționale se pot descrie astfel:

- **Inițializare:** sistemul este calibrat folosind 3 soluții standard; se determină panta și ordonata de origine a funcției de conversie $V \rightarrow \text{pH}$.
- **Calibrare injectomat:** se construiește funcția de transfer masă \rightarrow timp de injecție (Figura 11).
- **Execuție ciclică:** sistemul efectuează secvențe: dozare \rightarrow așteptare \rightarrow achiziție pH \rightarrow transmitere date.
- **Analiză finală:** spline cubică + derivare numerică \rightarrow identificare punct inflexiune \rightarrow calcul pKa.

Pentru procesarea semnalului generat de electrodul de pH, s-a proiectat un circuit analogic de amplificare, utilizând un amplificator operațional configurat în mod neînversor (Figura 12).

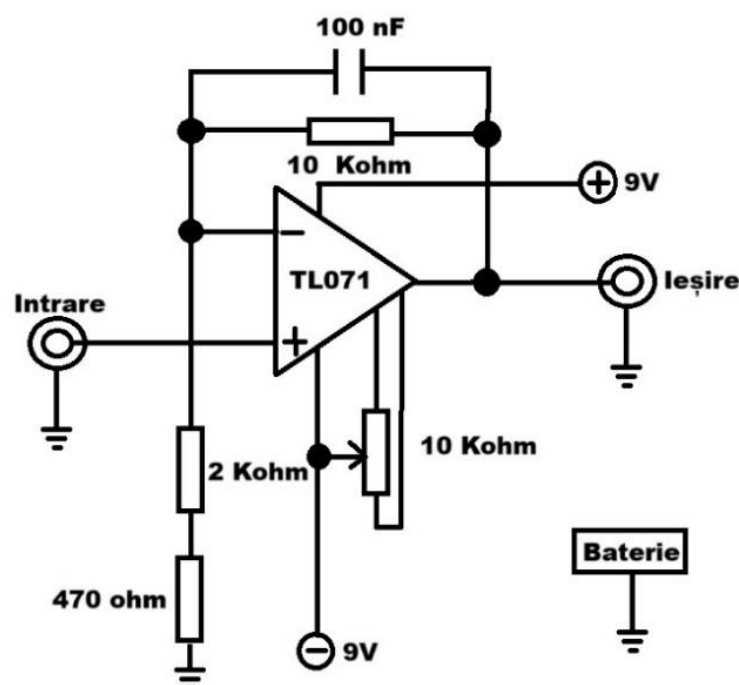


Figura 12. Schema de amplificare a semnalului de la electrodul de pH

Din perspectivă inginerescă, acest circuit are rolul de a crește amplitudinea semnalului inițial, astfel încât să fie compatibil cu intervalul de tensiune acceptat de convertorul analog-digital încorporat în microcontrolerul Arduino.

Transpunerea semnalului în domeniul 0-5V permite o digitalizare eficientă și precisă, esențială pentru interpretarea ulterioară a datelor în contextul unui sistem integrat.

În contextul calculatoarelor, semnalul analogic amplificat devine un input valid pentru placa Arduino, unde este prelucrat numeric după conversia analog-digitală efectuată

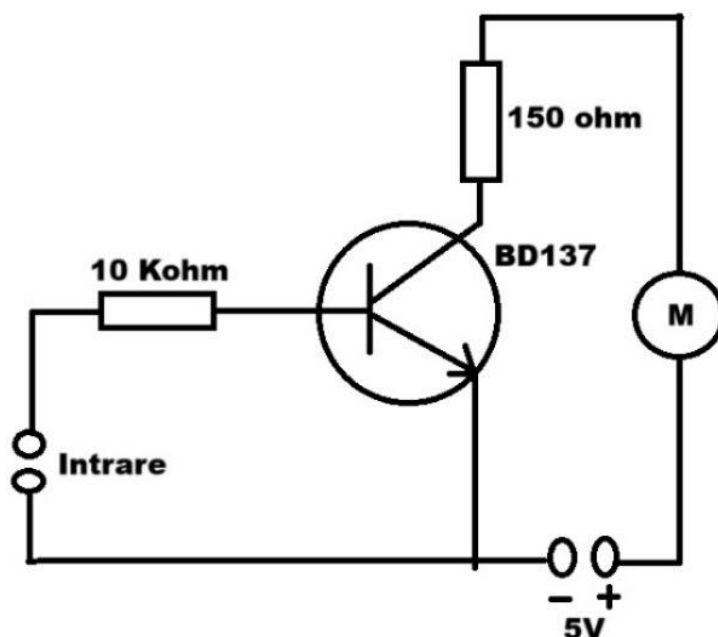


Figura 13. Schema de comandă a injectomatului cu tranzistorul BD137

la pinul A0. Valoarea numerică rezultată este introdusă în logica de control, unde algoritmul implementează o decizie automatizată bazată pe dinamica pH-ului monitorizat în timp real.

Pentru realizarea dozării automate a titrantului, s-a conceput un sistem de acționare electrică (Figura 13 și Figura 14), care conectează nivelul logic al microcontrolerului cu mecanismul fizic de injectare.

Motorul injectomatului este acționat printr-un semnal generat pe pinul digital 9 al plăcii Arduino, care comandă un tranzistor de tip BD137 configurat ca întrerupător electronic.

Din punct de vedere hardware, această arhitectură permite izolarea etajului de putere față de logica de comandă și oferă un control robust și precis al motorului, esențial pentru aplicarea unui feedback bazat pe pH.

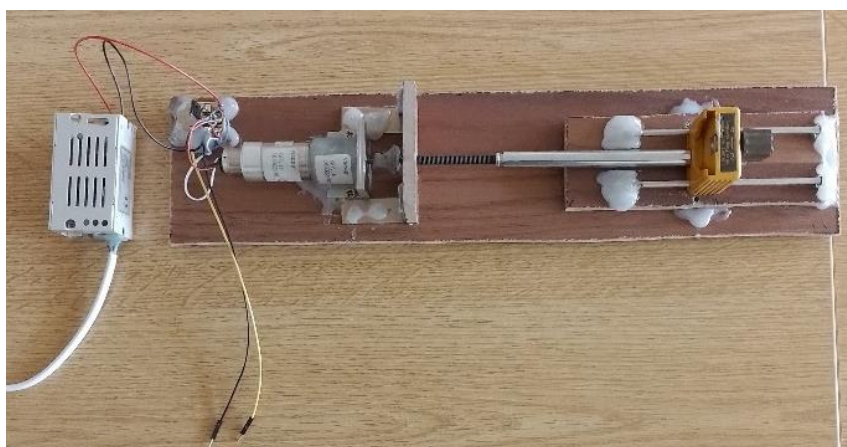


Figura 14. Injectomatul propriu-zis utilizat pentru adăugarea titrantului

Pentru verificarea semnalului analogic amplificat, s-a inclus în sistem un voltmetru digital, conform configurației prezentate în Figura 15. Această verificare în timp real permite calibrarea eficientă a circuitului și oferă redundanță în procesul de achiziție a datelor.



Figura 15. Vedere de ansamblu a montajului pH-metrului digital

Figura 16 prezintă integrarea completă a sistemului pH-metric automatizat. Arhitectura integrată rezultată include module de achiziție analogică, amplificare, conversie și control al actuatoarei, toate gestionate prin intermediul unui microcontroler.

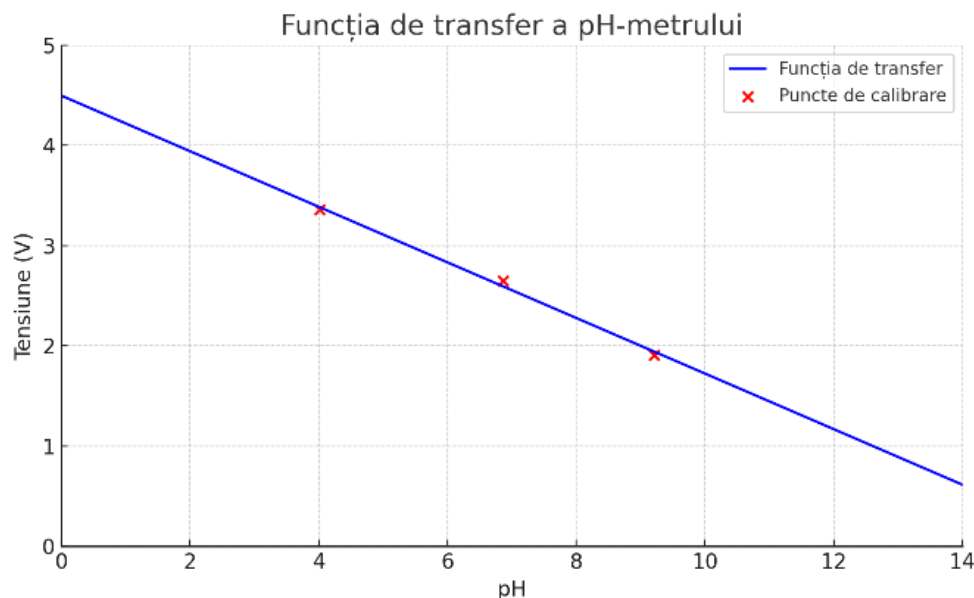
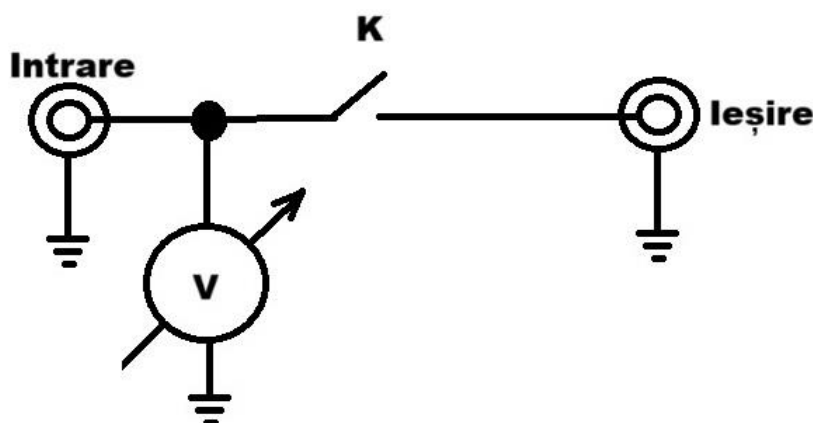


Figura 16. Funcția de transfer a pH-metrului

Sistemul astfel implementat asigură desfășurarea unei titrări automate cu monitorizare continuă și control adaptiv al dozării titrantului, reprezentând o aplicație concretă a conceptelor din electronică aplicată și sisteme de control în ingineria calculatoarelor.

Prin măsurarea tensiunii corespunzătoare fiecărui pH s-a trasat dreapta de etalonare a sistemului ($pH = a \cdot T + b$), unde T este tensiunea măsurată. Această funcție de transfer a permis convertirea în timp real a tensiunii în valoare de pH, pe toată durata titrării.



**Figura 17. Monitorizarea ieșirii semnalului cu voltmetru digital
(înainte de intrarea în Arduino)**

Funcția de transfer a pH-metrului (Figura 17) descrie dependența dintre tensiunea de ieșire a amplificatorului și valoarea pH-ului soluției măsurate. Aceasta este necesară pentru a putea converti semnalul electric măsurat (în volți) în valoarea chimică dorită (pH).

Dezvoltarea acestui sistem a implicat integrarea unor biblioteci software precum: Matplotlib [20] (vizualizare), NumPy [22] și SciPy [23] (analiză numerică), Scikit-learn pentru calculul scorului de regresie [24], Tkinter [25] (interfață), PySerial (comunicare serială), și FPDF [26] (raportare).

Hardware-ul a fost programat în Arduino IDE folosind limbajul C++ specific platformei. Arhitectura generală poate fi ușor extinsă prin adăugarea de senzori suplimentari sau integrarea cu aplicații mobile ori baze de date externe pentru trasabilitate și controlare.

5.5. IMPLEMENTAREA APLICAȚIEI DE CONTROL A TITRĂRII

Aplicația software dezvoltată integrează controlul hardware și reprezentarea grafică a procesului de titrare printr-o interfață grafică realizată cu biblioteca Tkinter din Python.

Această aplicație acționează ca un punct central de legătură între microcontrolerul Arduino, responsabil de achiziția datelor de la senzorul de pH și de controlul pompei de titrare, și utilizatorul uman, care poate urmări vizual progresul experimentului și poate genera rapoarte automate.

Comunicarea cu sistemul hardware se realizează prin intermediul portului serial, utilizând biblioteca serial. Pentru fiecare etapă a procesului, aplicația trimite comenzi corespunzătoare spre Arduino („Calibrare1”, „Titrare” etc.) și primește citiri analogice de la senzor, pe care le convertește în valori de tensiune (0-5V) și apoi, printr-o ecuație de calibrare determinată empiric, în valori de pH.

Interfața grafică a fost gândită modular, fiecare fază a experimentului fiind asociată cu butoane și etichete corespunzătoare. Etapele sunt organizate secvențial: calibrare prin trei soluții standard (de exemplu, pH 4,01, 6,86 și 9,21), determinarea dreptei de etalonare, începerea titrării și analiza numerică a datelor pentru identificarea punctului de inflexiune, de unde se extrage valoarea pK_a .

Elementul grafic esențial în aplicație este integrarea modulelor de afișare grafică Matplotlib în Tkinter prin FigureCanvasTkAgg. Astfel, utilizatorul poate vizualiza în timp real curba de calibrare și curba de titrare.

Această integrare nu doar că oferă feedback vizual instantaneu, dar și salvează automat graficele în fișiere PNG, care sunt ulterior inserate într-un document PDF generat dinamic cu biblioteca FPDF. Acestea au fost implementate conform Figurii 18.

```
# Afișam plot-ul în interfața Tkinter
canvas = FigureCanvasTkAgg(fig, master=frame)
canvas.draw()
canvas.get_tk_widget().pack()
canvas.print_figure('calibrare.png', dpi=300)
```

Figura 18. Implementarea afișării graficului

Acest fragment de cod prezentat arată cum se afișează graficul în interfața grafică și cum se salvează automat imaginea pentru raportul experimental. Interoperabilitatea dintre componentele software (Tkinter, Matplotlib, FPDF) și hardware (Arduino, senzor pH, motor dozator) demonstrează o integrare completă între sisteme de control și interfețe grafice, specifică aplicațiilor moderne de inginerie mecatronică și automatizare de laborator.

6. REZULTATE EXPERIMENTALE

În acest capitol sunt prezentate rezultatele obținute atât prin evaluarea modelului de predicție a valorii pK_a , cât și prin determinările experimentale realizate prin titrare automată. Se urmărește validarea și compararea abordării integrate, care combină metode computaționale și experimente fizico-chimice, evidențiind performanța modelului și corelarea cu valorile reale.

6.1. EVALUAREA PERFORMANȚEI MODELULUI DE PREDICȚIE

Evaluarea performanței sistemului de predicție a valorii constantei de aciditate (pK_a) a fost realizată pe baza unei analize riguroase a erorilor, metodelor de validare și metricilor de performanță, utilizate frecvent în domeniul inteligenței artificiale aplicate în chimie computațională.

Modelul de rețea neuronală grafică a fost implementat în Python folosind biblioteca PyTorch Geometric, iar datele moleculare au fost procesate din formate SMILES în grafuri moleculare, prin conversie și extragerea caracteristicilor atomice. Evaluarea performanței se bazează pe analiza rezultatelor obținute pentru setul de testare, după antrenarea modelului pe datele experimentale cunoscute.

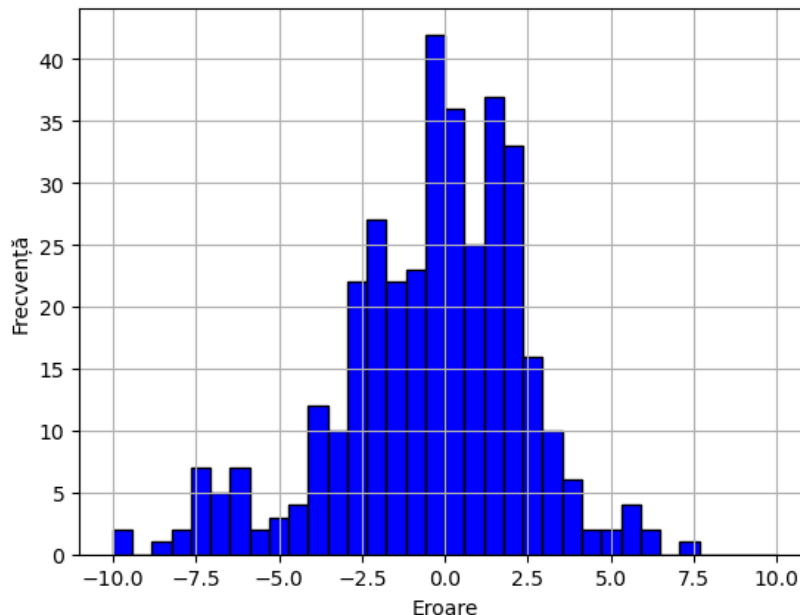


Figura 19. Histograma distribuției erorilor modelului

Histograma distribuției erorilor de predicție din Figura 19, exprimată în unități pK_a , care indică o erorare centrată în jurul valorii zero, ceea ce reflectă o precizie ridicată în estimarea valorilor. Acest aspect este relevant din punct de vedere ingineresc deoarece arată că modelul nu introduce bias sistematic în estimări.

Distribuția este aproximativ gaussiană, cu o dispersie redusă, ceea ce sugerează o capacitate a modelului de a generaliza dincolo de datele de antrenament, comportament așteptat în arhitecturi robuste de inteligență artificială. De asemenea, prelucrarea acestor erori a fost realizată cu ajutorul bibliotecii NumPy, prin compararea valorilor reale cu cele prezise pentru fiecare moleculă.

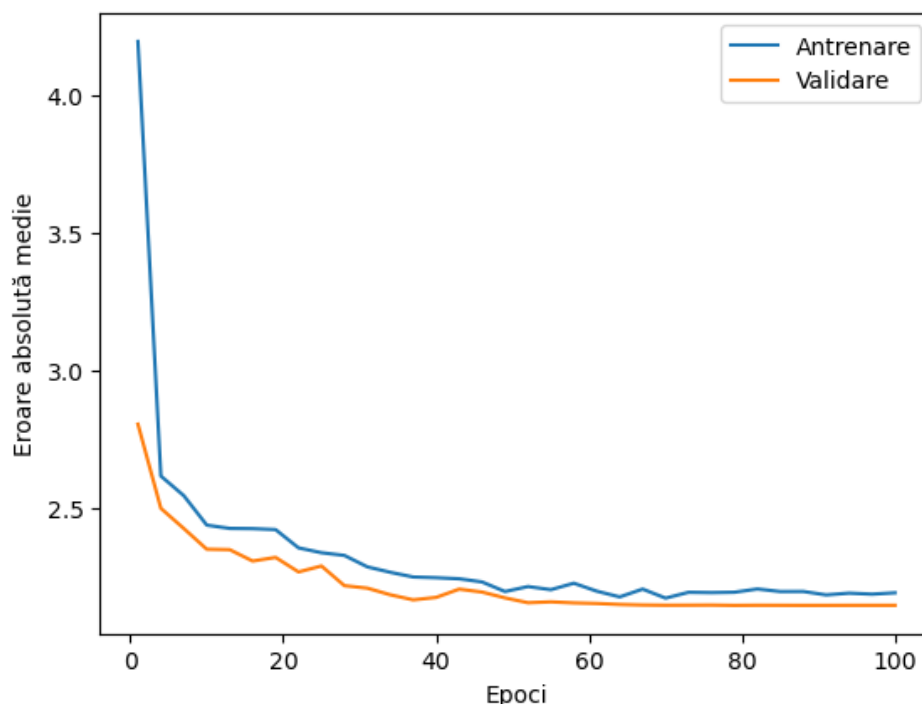


Figura 20. Evoluția erorii absolute medii în funcție de epocile de antrenare

În Figura 20 se evidențiază evoluția erorii absolute medii în timpul procesului de antrenare, pentru seturile de antrenament și validare. Scăderea rapidă a valorilor erorilor în primele epoci confirmă capacitatea modelului de a învăța relația structurală dintre caracteristicile moleculare și valoarea pKa.

După aproximativ 50 de epoci, valorile erorilor se stabilizează, ceea ce indică atingerea unui minim local al funcției de cost, fără apariția supraantrenării. Această stabilitate este semnul unui model bine calibrat, cu hiperparametri selectați adecvat, precum rata de învățare, numărul de straturi ale rețelei neuronale și dimensiunea vectorilor de caracteristici.

Analiza toleranței, prezentată în Figura 21, oferă o perspectivă aplicată asupra performanței modelului, fiind un indicator important pentru implementarea în scenarii reale. Graficul arată proporția de predicții care se încadrează într-o anumită marjă de toleranță față de valorile experimentale. Se observă că pentru o toleranță de ± 1 unitate pKa, puțin peste 30% dintre predicții se află în jurul valorilor reale, ceea ce este semnificativ având în vedere complexitatea sarcinii de regresie non-liniară.

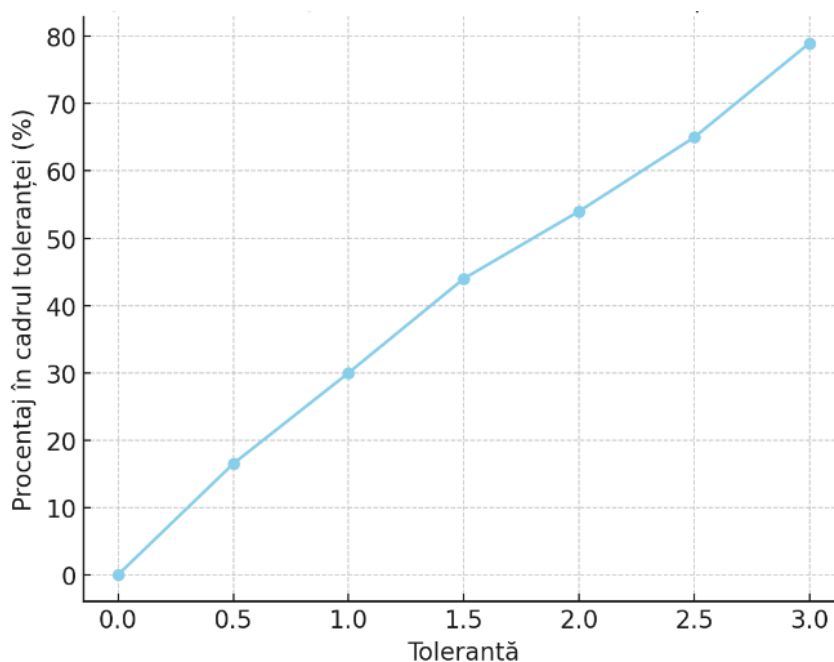


Figura 21. Analiza gradului de acceptabilitate a predicțiilor în funcție de marja de eroare admisă

Această proporție crește progresiv, atingând aproape 80% la o toleranță de ± 3 . Acest tip de analiză este asociat cu noțiunea de funcții de pierdere robustă, ce minimizează efectele outlierilor asupra antrenamentului. Totodată, interpretarea toleranței poate fi asociată unei distribuții de probabilitate a erorii condiționate de datele de intrare, caracteristică modelelor bayesiene sau a celor care includ incertitudinea în estimare.

Pe tot parcursul dezvoltării modelului, procesele de prelucrare a datelor, normalizare, fragmentare a seturilor (antrenare, testare, validare) și vizualizare au fost automatizate printr-un pipeline software modular, compatibil cu integrarea în aplicații mai complexe de tip platformă de chemoinformatică. Codul a fost structurat pentru a permite reînvățare ușoară pe alte seturi de date, adaptarea arhitecturii și salvarea modelului final în format serializat (torch.save), pentru reutilizare ulterioară.

În concluzie, sistemul de predicție dezvoltat în acest proiect demonstrează o performanță ridicată, susținută de o analiză cantitativă riguroasă și de vizualizări interpretabile, din perspectiva ingineriei calculatoarelor. Acesta poate fi considerat un model de referință pentru estimarea valorilor pKa în contextul inteligenței artificiale aplicate chimiei, cu potențial de extindere în sisteme mai mari de decizie asistată sau laborator virtual.

6.2. DETERMINAREA EXPERIMENTALĂ A VALORII PKA

Determinarea valorii pKa a fost realizată prin analiza experimentală a curbelor de titrare pentru trei acizi carboxilici cu structură diferită: acid formic, acid acetic și acid propanoic. Fiecare dintre aceștia a fost titrat cu două soluții de bază: o bază puternică,

hidroxidul de sodiu (NaOH), și o bază slabă, anume boratul de sodiu ($B_4O_7^{2-}$), pentru a investiga comportamentul sistemului de titrare în condiții diferite de forță bazică.

Acid formic

Pentru acidul formic, curbele de titrare sunt prezentate în Figura 22 (titrare cu borat de sodiu) și Figura 23 (titrare cu NaOH). Acest acid prezintă cel mai mic pK_a dintre cei analizați, datorită naturii sale puternic polarizate și absenței unui grup metil care să stabilizeze forma nedisociată.

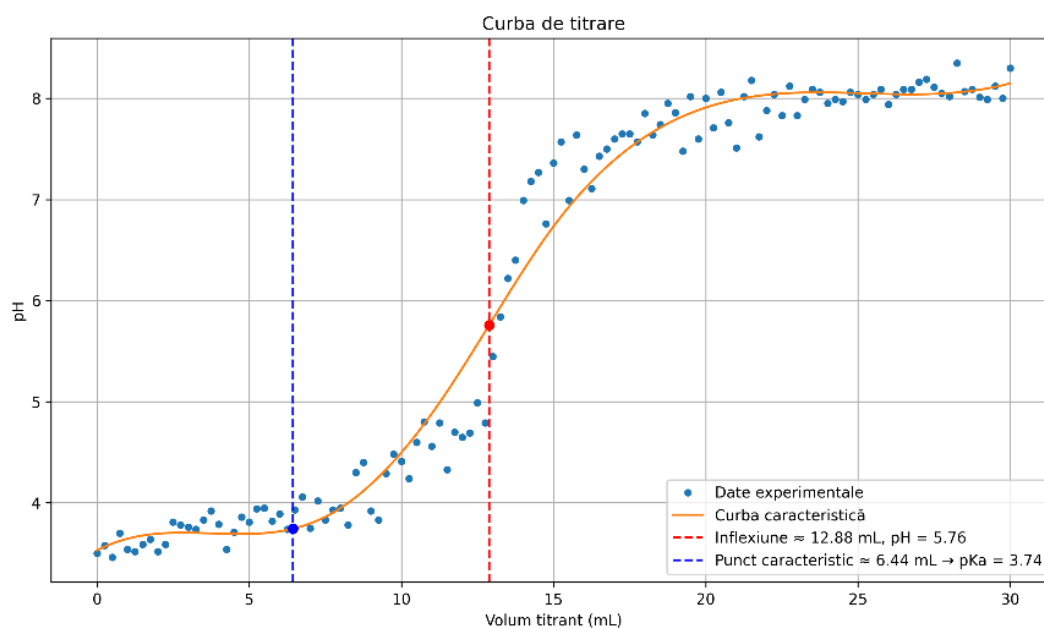


Figura 23. Curba de titrare a acidului formic cu boratul de sodiu

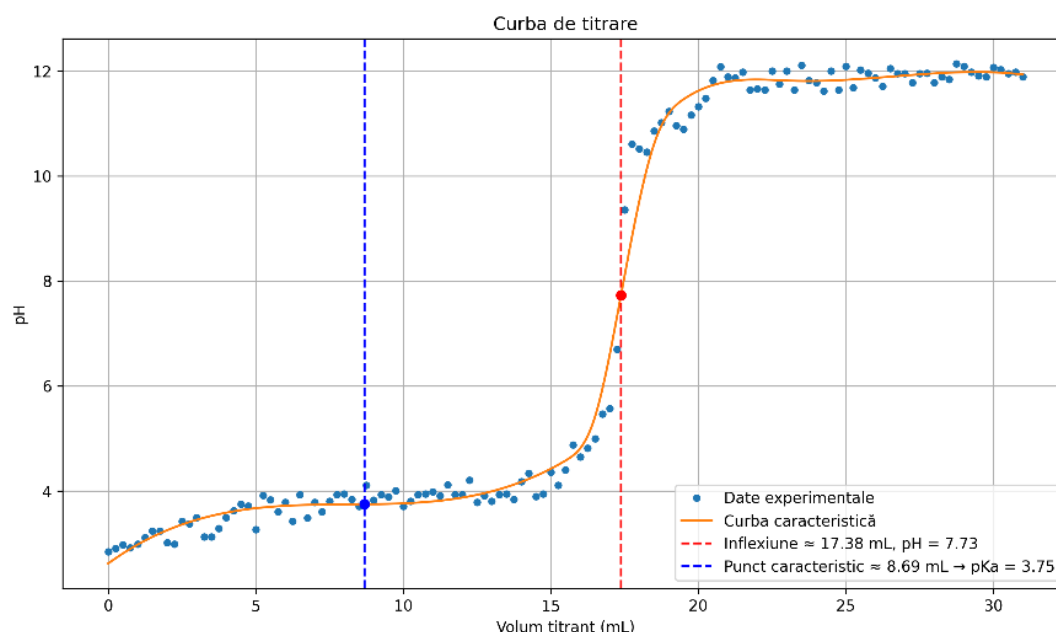


Figura 22. Curba de titrare a acidului formic cu hidroxidul de sodiu

În ambele experimente, punctul de inflexiune al curbei, identificat numeric prin derivata a doua, indică un volum de titrant în jur de 12,9 mL pentru titrarea cu NaOH și de aproximativ 17,4 mL pentru titrarea cu borat. Valoarea pK_a se determină la jumătatea determinată experimental ca fiind în jur de 3,74 în titrarea cu NaOH și 3,75 în cazul titrării cu borat, diferențele fiind minime, ceea ce confirmă reproductibilitatea metodei.

Acid acetic

Acidul acetic, analizat în Figura 24 (titrare cu borat) și Figura 25 (titrare cu NaOH), este un acid slab monocarboxilic cu un grup metil adiacent, care are un efect inductiv slab electronodonator, ușor stabilizator al formei protonate.

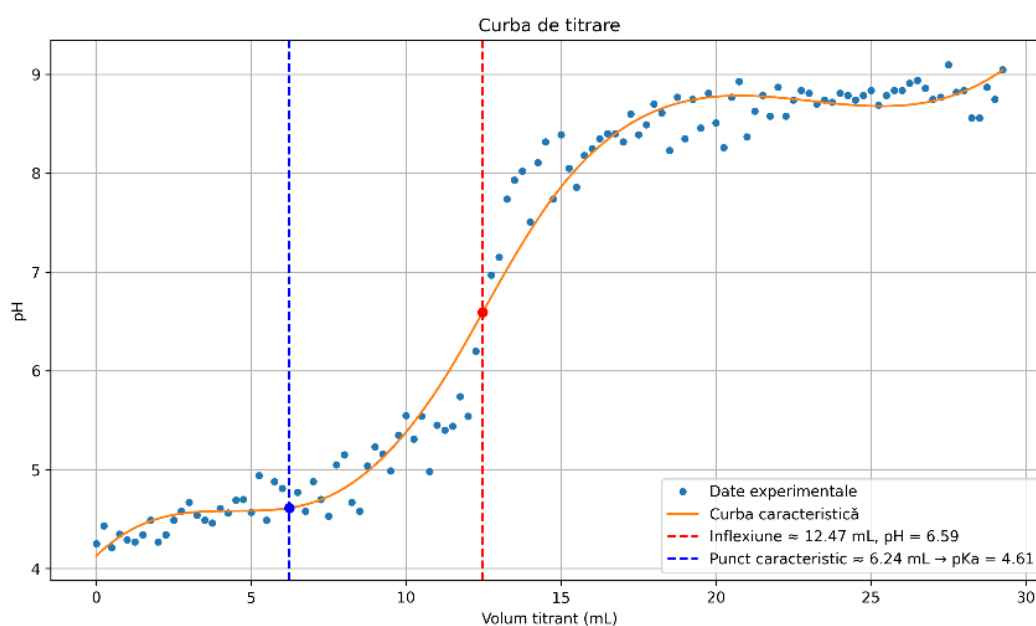


Figura 25. Curba de titrare a acidului acetic cu boratul de sodiu

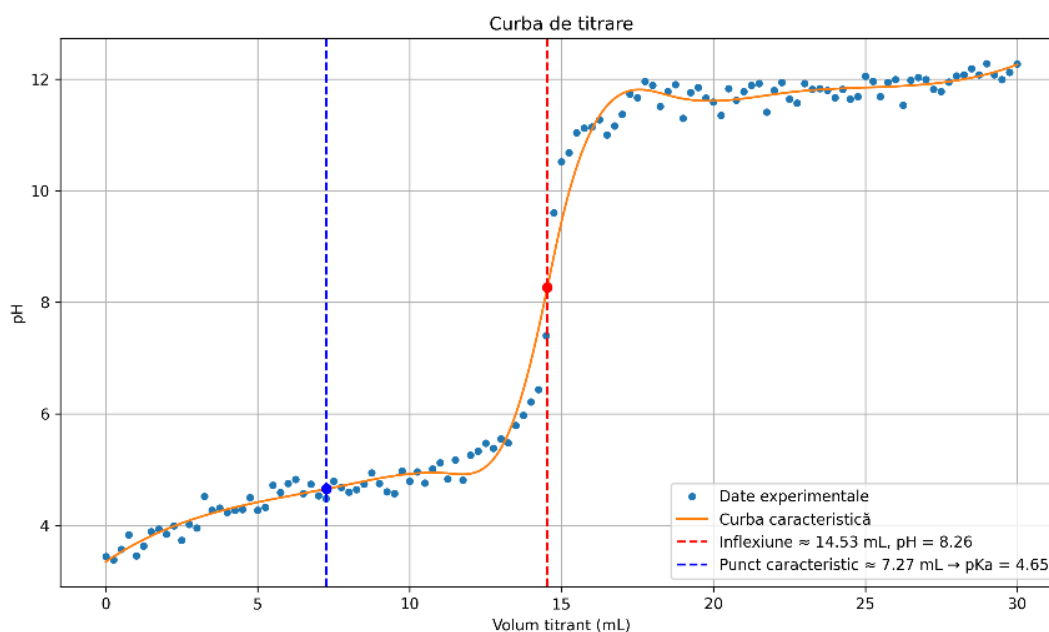


Figura 24. Curba de titrare a acidului acetic cu hidroxidul de sodiu

Curbele experimentale indică un volum de titrant la inflexiune de aproximativ 12,5 mL în cazul boratului și 12,7 mL în cazul NaOH-ului, iar pH-ul la acest punct este de aproximativ 6,6 și 6,8 respectiv. Punctul de jumătate de echivalență, care determină experimental pK_a -ul, corespunde unui volum de aproximativ 6,2 mL și 6,4 mL. Valoarea pK_a dedusă din aceste grafice este de 4,61 pentru borat și 4,76 pentru NaOH. Aceste rezultate sunt extrem de apropiate de valoarea teoretică de 4,75, ceea ce confirmă acuratețea metodei utilizate pentru prelucrarea datelor de titrare.

Acid propanoic

Pentru acidul propanoic, rezultatele sunt ilustrate în Figura 26 pentru titrarea cu borat și Figura 27 pentru titrarea cu NaOH. Acest acid prezintă un lanț metil suplimentar față

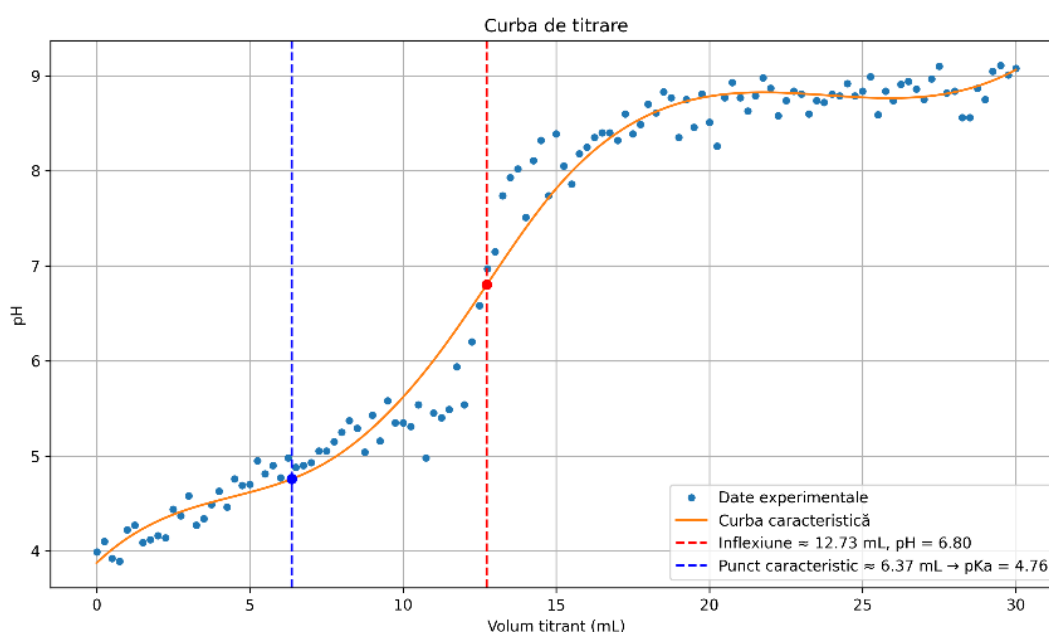


Figura 27. Curba de titrare a acidului propanoic cu borat de sodiu

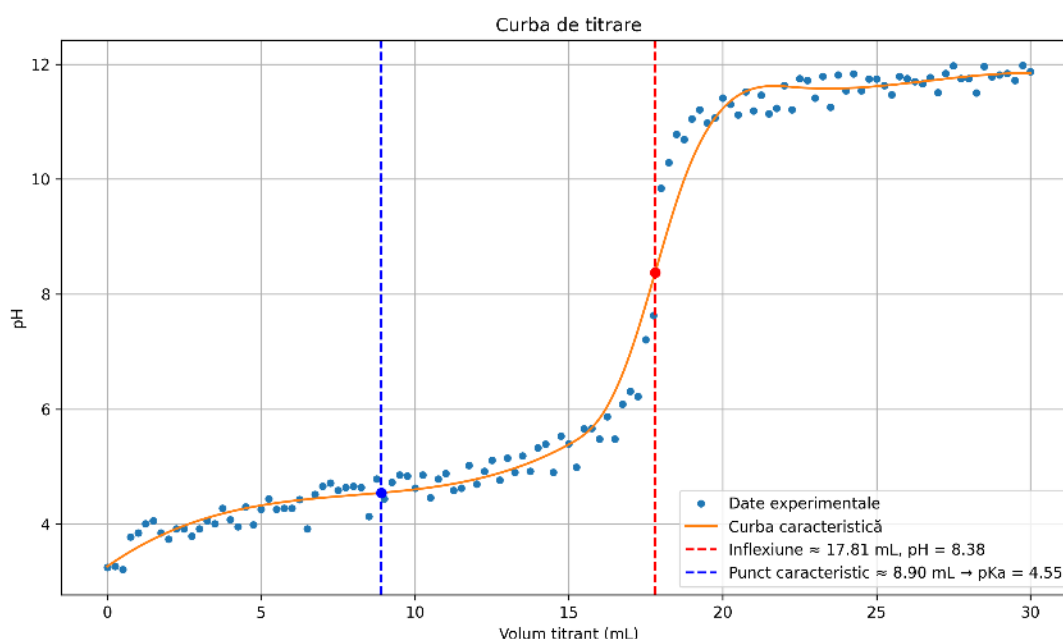


Figura 26. Curba de titrare a acidului propanoic cu hidroxid de sodiu

de acidul acetic, ceea ce accentuează efectul inductiv electronodonator și reduce ușor aciditatea, deci pK_a -ul ar trebui să fie mai mare.

Curba de titrare cu NaOH prezintă un punct de inflexiune la aproximativ 17,8 mL, cu un pH de 8,38, iar cea cu borat are inflexiunea la 14,5 mL, pH-ul fiind de 8,26. Determinarea valorii pK_a , prin metoda punctului caracteristic, a condus la un pK_a de 4,65 pentru titrarea cu borat și 4,55 pentru titrarea cu NaOH. Aceste valori, deși ușor mai mici față de referințele din literatură (unde pK_a -ul acidului propanoic este raportat în jur de 4,87), confirmă tendința previzibilă de variație a acidității în funcție de structura chimică.

Toate aceste observații arată că metoda de determinare experimentală a pK_a -ului utilizată în acest studiu, prin intermediul identificării inflexiunii curbei de titrare și a punctului de jumătate de echivalență, este robustă, reproductibilă și aplicabilă diferitelor tipuri de acizi slabi. Coerența valorilor obținute în cele două tipuri de titrare confirmă utilitatea metodei indiferent de forța bazei folosite, cu o precizie adecvată pentru aplicațiile didactice și experimentale în cercetare chimică.

6.3. CONCORDANȚA PREDICȚIILOR CU DATELE EXPERIMENTALE

În figura de mai jos este sintetizată performanța sistemului de predicție automată bazat pe rețele neuronale grafice, aplicat unui set de acizi carboxilici monoprotici (care pot ceda un singur proton – ion pozitiv de hidrogen), utilizând comparativ date obținute prin titrare clasică și literatură științifică validată. Dintr-o perspectivă inginerescă, accentul cade pe capacitatea de generalizare, stabilitatea numerică și interpretabilitatea arhitecturii algoritmice.

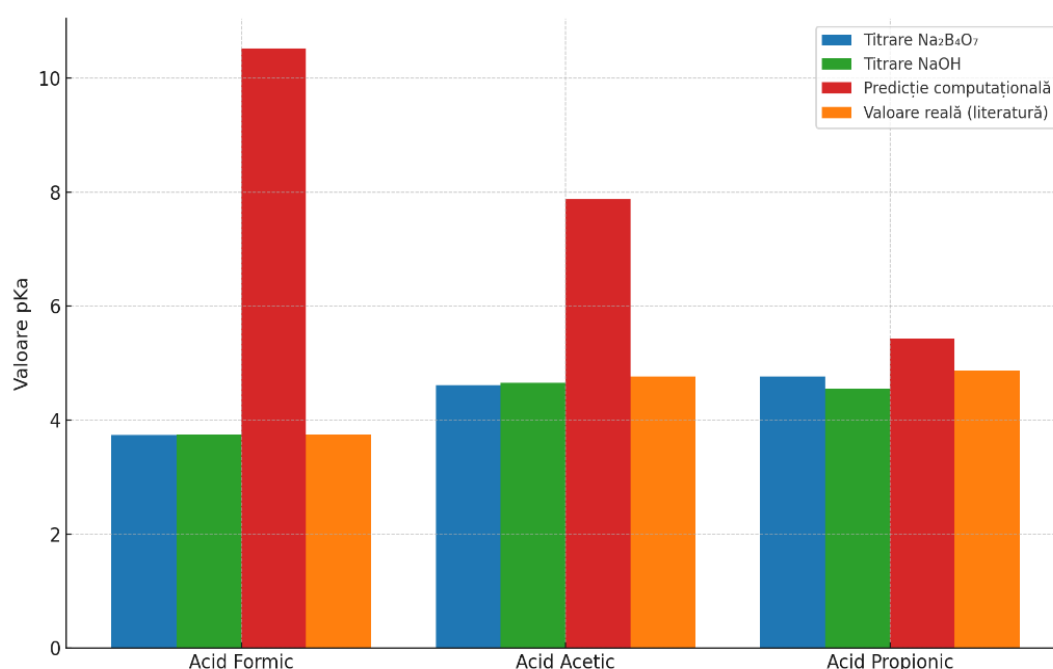


Figura 28. Compararea valorilor pK_a obținute experimental, prezise computațional și de referință pentru cei trei acizi organici

Modelul operează pe reprezentări moleculare sub formă de graf, în care nodurile corespund atomilor, iar muchiile legăturilor chimice, facilitând o propagare controlată a informației prin mecanisme de transmitere de mesaj. Fiecare moleculă este astfel transformată într-un vector numeric global prin agregare, urmată de regresie pentru estimarea scalarei pKa. Analiza diferențelor dintre predicții și valori experimentale evidențiază atât potențialul, cât și limitele arhitecturii implementate.

O eroare redusă pentru moleculele cu lanțuri scurte și structuri regulate denotă o convergență algoritmică favorabilă în acea regiune a spațiului molecular, în timp ce deviațiile semnificative, în special pentru compuși mai mici sau cu distribuții subreprezentate în setul de antrenament, indică o posibilă suprageneralizare sau deficit de eșantionare. Astfel de fenomene sunt caracteristice rețelelor de învățare profundă și necesită strategii suplimentare de echilibrare a datelor și ajustare fină.

Prin urmare, rezultatele obținute confirmă importanța evaluării riguroase prin metrici ingineresti, precum eroarea absolută medie și abaterea standard, alături de implementarea unor metode robuste de validare încrucișată și augmentare a datelor. Aceste direcții sunt esențiale pentru asigurarea fiabilității și scalabilității sistemelor de predicție automată utilizate în cadrul aplicațiilor interdisciplinare din chimia computațională.

7. CONCLUZII, CONTRIBUȚII ȘI DIRECȚII VIITOARE DE DEZVOLTARE

Capitolul de față reunește cele mai importante concluzii rezultate din activitățile desfășurate, evidențiind atât performanțele sistemului propus, cât și gradul de realizare a obiectivelor formulate inițial. Sunt prezentate contribuțiile originale aduse prin această lucrare, precum și direcțiile de cercetare și dezvoltare care pot valorifica rezultatele obținute. Analiza este realizată dintr-o perspectivă interdisciplinară, cu accent pe integrarea metodelor de inteligență artificială și a tehnologiilor de automatizare în contextul cercetării chimice moderne.

7.1. CONCLUZII GENERALE

Această lucrare a demonstrat viabilitatea unei abordări integrate hardware–software pentru estimarea și validarea experimentală a constantei acido-bazice pK_a , combinând algoritmi de învățare automată cu un sistem automatizat de titrare. Modelul de rețea neuronală grafică, antrenat pe un set de date validat științific, a fost capabil să estimeze valorile pK_a ale unor compuși chimici cu o acuratețe moderată. Acesta a fost integrat într-un flux funcțional complet, în care predicția *in silico* este urmată de o confirmare experimentală efectuată automat, cu ajutorul unei platforme Arduino dotate cu senzor de pH și control software realizat în Python.

Această sinergie software-hardware a evidențiat potențialul ingineriei calculatoarelor aplicate în chimia computațională și analitică. Sistemul creat permite estimarea rapidă a pK_a fără a necesita intervenție umană extinsă, reducând timpul și efortul asociate titrărilor clasice. În plus, prin automatizarea procesului experimental, se asigură un nivel crescut de reproductibilitate și se elimină sursele de eroare subiectivă.

În privința performanței, modelul a oferit predicții acceptabile pentru aproximativ o treime dintre compușii testați, înregistrând o eroare sub 0,5 unități pK_a – suficient de precisă pentru aplicații practice în farmacologie sau chimie aplicată. Deși această rată de succes sugerează o precizie limitată în stadiul actual, ea rămâne competitivă comparativ cu rezultatele publicate în literatura de specialitate, unde modele comerciale sau avansate obțin erori medii absolute în intervalul 0,6–0,8 unități. Totuși, variabilitatea ridicată a erorilor între diferite molecule sugerează o nevoie clară de îmbunătățire a capacității de generalizare a modelului.

Aplicabilitatea practică a sistemului propus este remarcabilă. Acesta poate funcționa ca un instrument de laborator capabil să genereze ipoteze (prin predicția valorilor pK_a) și să le testeze automat, reducând semnificativ durata ciclului de cercetare. Relevanța unei astfel de soluții este amplificată de contextul actual, în care se impune caracterizarea rapidă a unui număr mare de substanțe chimice noi. Într-un asemenea cadru, metodele computaționale asistate de inteligență artificială devin nu doar utile, ci esențiale.

Prin urmare, proiectul a reușit să atingă obiectivele propuse, oferind un prototip funcțional de sistem integrat pentru determinarea constantei pK_a . Acesta constituie un exemplu concret al aplicării metodelor de inteligență artificială și automatizării în cercetarea chimică modernă, și poate reprezenta o bază solidă pentru dezvoltări viitoare, inclusiv extinderea spre alte proprietăți moleculare de interes.

7.2. CONTRIBUȚII

O primă contribuție majoră a fost proiectarea și implementarea unui model de rețea neuronală grafică dedicat predicției constantei acido-bazice pK_a . Aceasta a presupus conceperea arhitecturii modelului, alegerea reprezentării moleculare sub formă de graf, cu noduri corespunzând atomilor și muchii legăturilor chimice, precum și selecția și preprocesarea setului de date, format din compuși cu valori experimentale cunoscute ale constantei pK_a .

Au fost utilizate biblioteci specializate de învățare automată, precum PyTorch Geometric, pentru a implementa acest model, care a fost ulterior antrenat, validat și optimizat iterativ prin ajustarea hiperparametrilor. În plus, a fost realizată o analiză detaliată a performanțelor obținute pe subseturi diferite de date, identificându-se tiparele moleculare care favorizează predicții precise, respectiv cele care generează erori semnificative. Această analiză a stat la baza îmbunătățirilor aduse ulterior modelului.

Cea de-a doua contribuție importantă a constat în dezvoltarea unei aplicații software în limbajul Python, menită să asigure controlul și monitorizarea experimentului de titrare. Aplicația a fost concepută pentru a comunica în timp real cu o platformă Arduino, trimițând comenzi pentru dozarea automată a titrantului și înregistrând simultan valorile de pH măsurate de senzor. Au fost implementate funcționalități pentru calibrarea senzorului de pH, filtrarea zgomotului din datele achiziționate, precum și rutine de analiză numerică pentru identificarea punctului de echivalență și calculul valorii pK_a pe baza curbei de titrare. Interfața aplicației a fost gândită minimalist, fie la nivel de consolă, fie cu elemente grafice simple, pentru a permite utilizatorului controlul integral al experimentului, vizualizarea rezultatelor și compararea valorilor experimentale cu predicțiile modelului. Totodată, au fost programate algoritmi de control ce reglează dozarea incrementală a soluției titrante în funcție de modificările detectate în pH, stabilind o legătură directă și dinamică între predicțiile digitale și comportamentul fizic observat.

În paralel, a fost realizată proiectarea și asamblarea sistemului hardware de titrare automatizată, reprezentând o contribuție esențială în validarea experimentală a predicțiilor. Configurația hardware a inclus o placă Arduino Uno, un sistem de control pentru o pompă peristaltică sau un servomotor conectat la o biuretă, un modul de măsurare a pH-ului cu circuit de condiționare, precum și un agitator magnetic controlat electronic, menit să asigure omogenizarea soluției. Întregul montaj a fost realizat pe o platformă compactă, adaptată mediului de laborator, luând în considerare siguranța electrică și robustețea sistemului de dozare.

A fost realizată și calibrarea sistemului, testând funcționalitatea cu soluții standard de acizi cunoscuți și verificând dacă punctele de echivalență sunt detectate corect, atât prin metoda gran, cât și prin derivarea numerică. Comunicarea între partea hardware și aplicația software a fost complet gestionată de autor, rezultând într-un sistem integrat și funcțional, comparabil din punct de vedere al performanței cu echipamente comerciale, dar realizat la un cost semnificativ mai redus, de peste 50 de ori mai mic.

Prin această serie de contribuții, s-a dezvoltat un ecosistem experimental complet, în care predicțiile modelului construit folosind învățarea automată sunt validate printr-un sistem fizic de titrare, închizând astfel bucla dintre simulare și experiment. Lucrarea demonstrează nu doar capacitatea de integrare a unor componente hardware și software complexe, ci și potențialul unor soluții mai puțin costisitoare în cercetarea chimică asistată de inteligență artificială.

7.3. DIRECȚII DE DEZVOLTARE

Pe baza rezultatelor obținute, se conturează o serie de direcții promițătoare de dezvoltare, atât în ceea ce privește componenta de predicție asistată de învățare automată, cât și din perspectiva automatizării experimentale. Aceste direcții vizează îmbunătățirea preciziei, extinderea funcționalităților și adaptarea sistemului la o paletă mai largă de aplicații din domeniul chimiei computaționale și analitice.

Un prim obiectiv major este creșterea performanței modelului de învățare automată. Acuratețea estimării valorilor pK_a poate fi îmbunătățită prin extinderea setului de date de antrenament, incluzând compuși cu o mai mare diversitate structurală și funcțională, precum și prin integrarea de date suplimentare din literatura științifică sau obținute direct cu ajutorul sistemului dezvoltat. De asemenea, pot fi testate arhitecturi neuronale mai sofisticate, cum ar fi rețelele neuronale cu mecanisme avansate de propagare a mesajelor și atenție contextuală sau modele construite prin tehnici de transfer learning, în care preantrenarea pe date sintetice este urmată de ajustarea fină pe date reale.

Totodată, integrarea informației tridimensionale în reprezentarea moleculară, cum ar fi coordonatele atomice sau descriptorii electrostatici, poate contribui la o mai bună captare a relațiilor spațiale care influențează aciditatea compușilor. Întrucât proprietățile precum pK_a sunt adesea dependente de stabilitatea conformerilor și de interacțiunile intramoleculare, luarea în considerare a acestor factori poate îmbunătăți semnificativ capacitatea modelului de a distinge între structuri similare.

Din perspectiva hardware-ului, se preconizează dezvoltarea unui sistem mai autonom și mai inteligent. O posibilă direcție constă în implementarea unui mecanism de feedback adaptiv, care să regleze dinamic debitul de titrant în funcție de variația pH-ului, prevenind astfel depășirea punctului de echivalență și optimizând consumul de reactivi. În plus, integrarea unor senzori suplimentari, precum senzori de temperatură sau debitmetre, ar putea crește acuratețea și reproductibilitatea măsurărilor. O extindere naturală a

platformei constă în conectarea sa la infrastructuri de tip Internet of Things, ceea ce ar permite stocarea datelor în cloud, analiza de la distanță și controlul experimentelor prin interfețe web. Această transformare ar facilita supravegherea simultană a mai multor sisteme, întreținerea predictivă a componentelor și, în perspectivă, implementarea unor laboratoare complet autonome. De asemenea, adaptarea sistemului într-un format portabil, alimentat din surse autonome, ar permite utilizarea sa în aplicații educaționale, de teren sau în medii cu resurse limitate.

O direcție strategică de extindere o reprezintă aplicarea metodologiei la predicția și determinarea altor parametri chimici de interes. Abordarea propusă este generalizabilă și poate fi utilizată pentru estimarea unor proprietăți precum solubilitatea în apă, constanta bazică pK_b , viteza de reacție sau constante de legare. Pentru fiecare dintre acești parametri, se pot dezvolta modele de predicție bazate pe învățare profundă și se pot construi prototipuri experimentale automatizate corespunzătoare. De exemplu, pentru solubilitate, un robot de dizolvare combinat cu un sistem de detecție spectroscopică. Astfel, platforma actuală poate deveni nucleul unui ecosistem de laborator modular, în care modele de inteligență artificială colaborează direct cu instrumente experimentale inteligente pentru a accelera ciclul de cercetare și dezvoltare.

În concluzie, această lucrare a demonstrat fezabilitatea integrării unui model de învățare automată cu un sistem hardware pentru determinarea pK_a și a deschis multiple direcții de extindere viitoare. De la perfecționarea modelelor de predicție și creșterea autonomiei sistemului hardware, până la extinderea către alți parametri chimici esențiali, aceste perspective conturează o traiectorie clară de evoluție. Lucrarea contribuie astfel la dezvoltarea domeniului ingineriei calculatoarelor aplicate în științele naturale și deschide drumul către platforme de laborator inteligente, capabile să sprijine descoperirea științifică prin automatizare, flexibilitate și învățare continuă.

8. BIBLIOGRAFIE

- [1] Wu, Q., Zhang, Y., Xu, H., & Chen, Y. (2022). Machine learning for pKa prediction: Methods, datasets, and applications. *Drug Discovery Today*, 27(9), 2425–2436.
- [2] Pan, X., Lee, J. Y., & Kim, S. (2021). MolGpKa: A graph-convolutional neural network model for predicting acid dissociation constants. *Journal of Chemical Information and Modeling*, 61(12), 5962–5970.
- [3] Miao, H., Wang, W., He, H., & Song, Y. (2024). GR-pKa: A graph retention mechanism for accurate prediction of pKa values. *Briefings in Bioinformatics*, 25(1), bbad452.
- [4] Han, J., & Priefer, R. (2023). A comprehensive review of experimental techniques for determining acid dissociation constants. *International Journal of Pharmaceutics*, 637, 122761.
- [5] Del Castillo-Santaella, T., Ortega, I., & García, J. (2023). Low-cost automated titration system using Arduino and open-source syringe pump. *HardwareX*, 13, e00335.
- [6] Han, J., & Priefer, R. (2023). A comprehensive review of experimental techniques for determining acid dissociation constants. *International Journal of Pharmaceutics*, 637, 122761.
- [7] Yang, S., & Skolnik, S. (2023). A hybrid potentiometric and UV-metric method for high-throughput pKa determination. *Journal of Pharmaceutical and Biomedical Analysis*, 233, 115658.
- [8] Sanli, S., Sanli, N., Alsancak, G., & Alsancak, G. (2019). Determination of acidity constants of biologically active compounds by spectrophotometric and potentiometric methods. *Indian Journal of Pharmaceutical Education and Research*, 53(3), 478–485.
- [9] Saurina, J., Núñez, O., & Puignou, L. (2000). Gradient pH flow injection analysis with multivariate curve resolution–alternating least squares (MCR-ALS) for the determination of acid dissociation constants. *Analytica Chimica Acta*, 410(1–2), 69–78.
- [10] Wu, Jialu & Kang, Yu & Pan, Peichen & Hou, Tingjun. (2022). Machine learning methods for pKa prediction of small molecules: Advances and challenges. *Drug Discovery Today*. 27. 103372.
- [11] Skoog, D. A.; Holler, F. J.; Crouch, S. R. *Principles of Instrumental Analysis*, 6th ed.; Thomson Brooks/Cole: Belmont, CA, 2007.
- [12] Famularo, N.; Kholod, Y.; Kosenkov, D. Integrating Chemistry Laboratory Instrumentation into the Industrial Internet: Building, Programming, and Experimenting with an Automatic Titrator. *Journal of Chemical Education* 93(1), 175–181 (2016).
- [13] Harris, D. C.; Lucy, C. A. *Quantitative Chemical Analysis*, 9th ed.; W.H. Freeman and Company: New York, NY, 2016.
- [14] Savitzky, A.; Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36(8), 1627–1639 (1964).
- [15] Zheng, Jonathan W. and Lafontant-Joseph, Olivier, IUPAC Digitized pKa Dataset, v2.3. Copyright © 2025 International Union of Pure and Applied Chemistry (IUPAC), 2025.

- [16] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). "Neural Message Passing for Quantum Chemistry." ICML.
- [17] Wu, Z. et al. (2018). "MoleculeNet: A Benchmark for Molecular Machine Learning." Chemical Science.
- [18] Kearnes, S., McCloskey, K., Berndl, M., Pande, V., & Riley, P. (2016). "Molecular graph convolutions: moving beyond fingerprints." Journal of Computer-Aided Molecular Design.
- [19] Streamlit Developers, Streamlit documentation, <https://docs.streamlit.io/>, accesat la 11.06.2025.
- [20] Arduino, Arduino IDE, <https://www.arduino.cc/en/software>, accesat în iunie 2025.
- [21] Hunter, J. D., Matplotlib: A 2D graphics environment, Computing in Science & Engineering, 2007.
- [22] McKinney, W., Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 2010.
- [23] Pedregosa, F. et al., Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 2011.
- [24] Virtanen, P. et al., SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 2020.
- [25] Python Software Foundation, Tkinter, <https://docs.python.org/3/library/tkinter.html>, accesat la 12.06.2025.
- [26] PyFPDF Developers, FPDF for Python, <https://pyfpdf.github.io/fpdf2/>, accesat la 12.06.2025.

**DECLARAȚIE DE AUTENTICITATE A
LUCRĂRII DE FINALIZARE A STUDIILOR***

Subsemnatul CRISAN CRISTIAN

legitimat cu Ci seria TZ nr. 576 817

CNP 5020203350095

autorul lucrării PREDICȚIA CONSTANTEI PKA FOLOSIND O REȚEA NEURONALĂ
ȘI DETERMINAREA AUTOMATIZATĂ A ACESTEIA

elaborată în vederea susținerii examenului de finalizare a studiilor de
LICENȚĂ organizat de către Facultatea

DE AUTOMATICĂ ȘI CALCULATOARE din cadrul Universității

Politehnica Timișoara, sesiunea Iunie 2025 a anului universitar
2024 - 2025, coordonator PROF. DR. HABIL. ÎNG. MIHAI-VICTOR MICEA, luând în

considerare art. 34 din *Regulamentul privind organizarea și desfășurarea examenelor de licență/diplomă și disertație*, aprobat prin HS nr. 109/14.05.2020 și cunoscând faptul că în cazul constatării ulterioare a unor declarații false, voi suporta sancțiunea administrativă prevăzută de art. 146 din Legea nr. 1/2011 – legea educației naționale și anume anularea diplomei de studii, declar pe proprie răspundere, că:

- această lucrare este rezultatul propriei activități intelectuale;
- lucrarea nu conține texte, date sau elemente de grafică din alte lucrări sau din alte surse fără ca acestea să nu fie citate, inclusiv situația în care sursa o reprezintă o altă lucrare/alte lucrări ale subsemnatului;
- sursele bibliografice au fost folosite cu respectarea legislației române și a convențiilor internaționale privind drepturile de autor;
- această lucrare nu a mai fost prezentată în fața unei alte comisii de examen/prezentată public/publicată de licență/diplomă/disertație;
- În elaborarea lucrării ~~am utilizat~~ instrumente specifice inteligenței artificiale (IA) și anume _____ (denumirea) _____ (sursa), pe care le-am citat în conținutul lucrării/nu am utilizat instrumente specifice inteligenței artificiale (IA)¹.

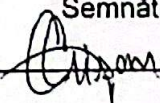
Declar că sunt de acord ca lucrarea să fie verificată prin orice modalitate legală pentru confirmarea originalității, consimțind inclusiv la introducerea conținutului său într-o bază de date în acest scop.

Timișoara,

Data

27.06.2025

Semnătura



¹•Declarația se completează de student, se semnează olograf de acesta și se inserează în lucrarea de finalizare a studiilor, la sfârșitul lucrării, ca parte integrantă.

Se va păstra una dintre variante: 1 - s-a utilizat IA și se menționează sursa 2 – nu s-a utilizat IA