10th Iranian Biennial
Chemometrics Seminar

10TH BIENNIAL CHEMOMETRICS SEMINAR

جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فناوری

دهمین سمینار دوسالانه کمومتریکس انجمن شیمی ایران

The 10th Biennial Iranian Chemometrics Conference

۲۳ و ۲۴ مهرماه ۱۴۰۴ / دانشگاه ارومیه / سالن پژوهشکده آرتمیا

ISC
Islamic World Science Citation Center

محـــــورهای همایش

- روش های پیشرفته در کمومتریکس، معرفی و بکارگیری
- مدل سازی و بهینه سازی فرآیند
- علم داده و تحلیل داده های بزرگ (Big Data Analysis)
- یادگیری ماشین و هوش مصنوعی
- آموزش و استاندار سازی در کمومتریکس بهبود آموزش و تعیین استانداردهای بین المللی در این حوزه

دبیرخانه همایش

آدرس دبیرخانه
دانشگاه ارومیه/دانشکده شیمی

Site address
Chemometrics10.urmia.ac.ir

Email address
chemometrics10@urmia.ac.ir

تلفن: ۰۴۴۳ ۱۱۴ ۴۱۰۷
نمابر: ۰۴۴۳ ۱۱۴ ۴۱۰۲

شهرداری ارومیه · شورای اسلامی شهر ارومیه · شرکت آب و فاضلاب آذربایجان غربی · اداره کل میراث فرهنگی، گردشگری و صنایع دستی آذربایجان غربی · کبوودان شیمی زرین · شرکت مصباح انرژی (سهامی خاص) · پرتو افزار صنعت

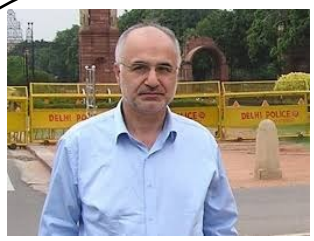
Program book proceeding of the

10th Iranian Chemical Society Biennial Chemometrics Seminar

Oct. 15, 16-2025

Urmia University

In the name of god



Speech by the President of Urmia University

Distinguished professors, esteemed researchers, dear students, and honorable guests,

It is a great honor and pleasure for Urmia University to host this distinguished gathering of scholars and scientists at the 10th Biennial Iranian Chemometrics Seminar. Your kind presence adds spirit and vitality to this academic event and reflects the dynamic nature and strength of Iran's scientific community, which continues to advance knowledge through collaboration and shared experience. As one of the oldest and most reputable universities in the country, Urmia University has always considered it its mission to promote science, nurture innovation, and foster academic and cultural exchange among universities and research centers. Organizing scientific meetings and specialized seminars plays a vital role in fulfilling this mission creating a space where ideas can be shared, experiences discussed, and new paths of cooperation and progress opened.

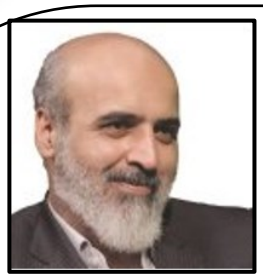
Undoubtedly, this event is the result of the collective effort of dedicated professors, researchers, and students who have worked diligently and passionately to make this seminar possible. I would like to express my sincere appreciation to Dr. Morteza Bahram, the Scientific Secretary, Dr. Aydin Bahrami, the Executive Secretary, and all members of the scientific and organizing committees for their commitment, precision, and teamwork in arranging this meaningful academic gathering.

I hope this seminar will serve as an opportunity to strengthen scientific interaction, inspire new ideas, and contribute to the continued advancement of higher education and research in our country.

Wishing all organizers and participants success, health, and fruitful achievements.

Dr. Mahmoud Reza Zadbari

President of Urmia University



Welcome Address by the Seminar Chairman

Dear Colleagues, Distinguished Professors, Researchers, and Esteemed Guests,

It is a great honor and pleasure to warmly welcome you all to the 10th Biennial Iranian Chemometrics Seminar, hosted by Urmia University. This significant scientific gathering provides an invaluable opportunity for scholars, experts, and students to exchange ideas, present their latest findings, and explore new perspectives in the ever-evolving field of chemometrics.

Chemometrics, from its very beginning, has served as a bridge between chemistry, mathematics, and statistics a science dedicated to extracting meaningful information from complex chemical data. Without the tools of chemometrics, much of this hidden knowledge would remain buried within numbers. It is fascinating to note that many concepts now popular in machine learning and artificial intelligence were, in fact, born within the field of chemometrics decades ago. Techniques such as Principal Component Analysis (PCA), Partial Least Squares (PLS) modeling, Factor Analysis, and Cross-Validation were among the first systematic steps toward recognizing patterns and predicting behaviors in multivariate data. In this sense, chemometrics can rightfully be regarded as a forerunner of machine learning in the chemical sciences. Today, in the age of big data and intelligent algorithms, the boundary between chemometrics and AI has become increasingly blurred. The integration of classical chemometric methods with modern deep learning techniques opens new avenues for understanding, prediction, and even the guidance of chemical phenomena. The future of chemometrics lies in this synergy, combining intelligent models, high-dimensional data, and advanced analytical methods to transform the way we conduct research and innovation.

We are truly grateful to our distinguished invited speakers from Iran and abroad, who generously share their expertise and insights, contributing to the depth and diversity of this seminar. I would also like to extend my sincere appreciation to the Iranian Chemical Society, the Chemometrics Committee, the Islamic World Science Citation Center (ISC), and all supporting institutions for their continued encouragement and collaboration.

Special thanks are due to Dr. Aydin Bahrami, the Executive Secretary of the Seminar, as well as the organizing and scientific committees, students, and staff of the University of Urmia, whose dedication and hard work have made this event possible.

I sincerely hope that this seminar will serve as a fruitful venue for knowledge exchange, inspire new research directions, and strengthen professional relationships within the chemometrics community

Let us seize this opportunity to share ideas, foster collaborations, and contribute together to the advancement of this exciting scientific field.

With warm regards and best wishes for a successful and inspiring seminar,

Dr. Morteza Bahram

Scientific Secretary 10th Biennial Iranian Chemometrics Seminar

University of Urmia, Iran

An Overview of the Biennial Iranian Chemometrics Seminars

No.	City	Year	Chairman	Organizer
1st	Arak	2006	Dr. Gholam Hassan Azimi	Arak University
2nd	Urmia	2009	Prof. Morteza Bahram	Urmia University
3rd	Tabriz	2011	Dr. Abdolhossein Naseri	Tabriz University
4th	Shiraz	2013	Prof. Bahram Hemmateenejad	Shiraz University
5th	Tehran	2015	Prof. Jahanbakhsh Ghasemi	Tehran University
6th	Babolsar	2017	Prof. Mohammad Hossein Fatemi	University of Mazandaran
7th	Shahrood	2019	Prof. Nasser Goudarzi	Shahrood University of Technology
8th	Tehran	2021	Dr. Ahmad Mani-Varnosfaderani	Tarbiat Modares University
9th	Qazvin	2023	Prof. Mohammadreza Khanmohammadi	Imam Khomeini International University
10th	Urmia	2025	Prof. Morteza Bahram	Urmia University

Scientific Committee

No.	Full Name	Academic Rank	University
1	Morteza Bahram	Professor	Urmia University
2	Hamid Abdollahi	Professor	Institute for Advanced Studies in Basic Sciences, Zanjan
3	Mohammad Hossein Fatemi	Professor	University of Mazandaran
4	Abdolhossein Naseri	Associate Professor	University of Tabriz
5	Bahram Hemmati Nejad	Professor	University of Shiraz
6	Hadi Parastar	Associate Professor	Sharif University of Technology
7	Maryam Vosough	Associate Professor	Institute of Chemistry and Chemical Engineering
8	Mansour Arab Cham Jangali	Professor	Shahrood University of Technology
9	Ahmad Mani	Associate Professor	Tarbiat Modares University

10	Seyedeh Maryam Sajjadi	Associate Professor	University of Semnan
11	Ali Niazi	Professor	Islamic Azad University, Central Tehran Branch
12	Taghi Khayamian	Professor	Isfahan University of Technology
13	Karim Asadpour Zeinali	Professor	University of Tabriz
14	Jahanbakhsh Ghasemi	Professor	University of Tehran
15	Mohammad Reza Khan Mohammadi	Professor	Imam Khomeini International University
16	Khalil Farhadi	Professor	Urmia University
17	Mohammad Alizadeh Khaled Abad	Professor	Urmia University
18	Reza Emamali Sabzi	Professor	Urmia University
19	Naser Samadi	Professor	Urmia University
20	Seyed Ali Hosseini	Professor	Urmia University
21	Maryam Khoshkam	Associate Professor	Mohaghegh Ardabili University

Executive Committee

No.	Full Name	Academic Rank	University
1	Aidin Bahrami	Assistant Professor	Urmia University
2	Reza Emamali Sabzi	Professor	Urmia University
3	Naser Samadi	Professor	Urmia University
4	Saber Amiri	Assistant Professor	Urmia University
5	Reza Dadashi	Postdoctorl researcher	Urmia University
6	Samin Mohammadi	Master student	Urmia University

Keynote Speakers

Prof. Hamid Abdollahi
Speech Topic: *Chemometrics Education in the Era of Digital Chemistry*
Speech Topic: *Assessing Matrix Effects with Multivariate Curve Resolution*

Dr. Hadi Parastar
Speech Topic: *From Latent Structure to Neural Networks: Bridging Chemometrics and Deep Learning in Chromatography and Imaging*

Prof. Roma Tauler
Speech Topic: *Recent Developments in Multivariate Curve Resolution*

Prof. Jahan B Ghasemi
Speech Topic: *From punch card to AI in chemistry*

Dr. Somaieh Soltani
Speech Topic: *The Functions of Chemometrics Integrated with the Emerging Paradigm of Machine Learning and Artificial Intelligence in Pharmaceutical Sciences*

10th Iranian Biennial Chemometrics Seminar Timetable

Tuesday, October 14, 2025	
Program Section	**Time**
Registration & Welcome	14:00-20:00
Wednesday, October 15, 2025	
First section	
Program Section	**Time**
Recitation of the Holy Quran and the National Anthem	08:30 – 08:40
Welcome Address by the President of Urmia University, Urmia University Anthem	08:40 – 08:50
Speech by the Scientific Secretary of the Seminar and Dean of the Faculty of Chemistry	08:50 – 09:00
Speech by the Representative of the Iranian Chemical Society and Head of the Chemometrics Committee of Iran (Dr. Mohammad Hossein Fatemi)	09:00 – 09:10
Introductory Video Clip of Urmia University	09:10 – 09:20
Speech by the Seminar Sponsor – Mr. Manouchehri (CEO of Kaboudan Chemistry Company and Former Lecturer of Urmia University)	09:20 – 09:30
Speech: Prof. Hamid Abdollahi (Chemometrics Education in the Era of Digital Chemistry)	09:30 – 10:00
Break and poster presentation	10:00 – 11:00

Wednesday, October 15, 2025		
Second section		
Speech Topic	**Presenter**	**Time**
Chair:Dr. M.Fatemi, Dr. K. Asadpour		
From Latent Structure to Neural Networks: Bridging Chemometrics and Deep Learning in Chromatography and Imaging	*Dr. Hadi Parastar*	11:00 – 11:30
Comparative Study of Nonlinear Hard and Soft Discrimination in Genomic Cancer Classification	*Dr. Somaye Vali Zade*	11:30 – 11:50
A transfer learning workflow based on MCR and CNN to overcome limitations in Vis-NIR HSI data modeling: a case study	*Dr. Fatemeh Sadat Hashemi-Nasab*	11:50 –12:10
Multivariate Calibration Models for ATR-FTIR Spectroscopy Analysis of Olive Oil Adulterated with Corn, Sunflower, and Palm Oils	*Esra modares Askari*	12:10 –12:30
Lunch & Prayer		12:30 – 14:30
Third section		
Speech Topic	**Presenter**	**Time**
Chair:Dr.M.khoshkam, Dr. S.Amiri		
The Functions of Chemometrics Integrated with the Emerging Paradigm of Machine Learning and Artificial Intelligence in Pharmaceutical Sciences	*Dr. Somaieh Soltani*	14:30 –15:00

AI-Driven Design and Optimization of Novel Carbon Allotropes via Crystallographic Space Group Exploration and Machine Learning Potentials	***Dr. Ebrahim Nemato-Kande***	**15:00 –15:20**
Moving toward Smart Sensors for eDiagnostics and eMonitoring	***Dr. Hamed Golmohammadi***	**15:20 –15:40**
Workshop (Hyperspectral and RGB image)	***Dr. Ahmed Doroudi***	**15:40 –16:00**
Break and poster presentation		**16:00 – 17:00**
Fourth section		
Speech Topic	***Presenter***	***Time***
Chair: Dr. H. Parastar, Dr. S.Soltani		
Recent Developments in Multivariate Curve Resolution	***Prof. Roma Tauler***	**17:00 –17:30**
Identification and Measurement of H2S, COS and SO2 in Natural Gas Using a Machine Learning Empowered Colorimetric Sensor Array	***Ahmad Moslehipour***	**17:30 –17:50**
A Deep Learning Workflow Based on Convolutional Neural Networks for GC-IMS Feature Extraction and Classification	***Farbod Bayat-Afshary***	**17:50 –18:10**
Traditional Ashiq Live Music		**18:10 – 18:30**

Tour of Tabib House and Urmia University Central Laboratory	**18:30 – 20:00**
Dinner	**20:00 – 21:00**

Thursday, October 16, 2025

Fifth section

Speech Topic	*Presenter*	*Time*
Chair: Dr. J. Ghasemi, Dr. M. Hormozinejad		
Assessing Matrix Effects with Multivariate Curve Resolution	***Prof. Hamid Abdollahi***	**08:00 –08:30**
Multi-Class Classification of Odor Perception Using Graph Neural Networks and Molecular Graph Representations	***Mahia V. Solout***	**08:30 –08:50**
Biologically Informed Prediction of Drug Synergy: A Transcriptomic Resistance Signature Approach	***Dr. Sajjad Gharaghani***	**08:50 –09:10**
Break and poster presentation		**09:10 – 10:30**

Sixth section

Chair:Dr. H.*Abdollahi, *Dr. M. Alizade Khaledabad		
From punch card to AI in chemistry	***Prof. J.B Ghasemi***	**10:30 –10:50**
Divergent Biological Conclusions: A Comparison of PLS-DA and MCR-ALS in Metabolomics	***Dr. Maryam Khoshkam***	**10:50 –11:10**

Application of support vector machine in predicting potentiometric selectivity (Mg2+/Ca2+) of some amide ligands	***Reza Mahmoudzadeh Laki***	**11:10 –11:30**
Application of multivariate tools in the discrimination of geographical origin of Iranian polyfloral honeys by analysis of physicochemical and biochemical characterization	***Dr. Saber Amiri***	**11:30 –11:50**
Closing Ceremony		**11:50 – 12:40**
Lunch & Prayer		**12:40 – 14:30**
Free Urmia City Tour		**15:00 – ------**

Poster presentation Wednesday, October 15, 2025		
Topic	**Presenter**	**ID No.**
A Comprehensive FTIR Spectral Database for Authentication and Quality Control of Edible and Essential Oils	***Niloofar Rahmani***	**P1**
Machine Learning Enhanced Discrimination of Biothiol and Thiol Ratios Using Gold Nanorod Amalgamation	***Golara Akhondi***	**P2**
Application of Central Composite Design in Modeling and Optimization of a Bio-Based Time–Temperature Indicator Using Fatty Acids	***Maryam Safari***	***P3***
Optimization of Aluminum Anodizing Conditions and Evaluation of Potassium Permanganate-Loaded Anodic Alumina for Ethylene Removal in Active Packaging of Fruits and Vegetables	***Saeid Golmirzaei***	***P4***
Optimizing Anthocyanin Extraction from Prunus domestica Flesh Using Experimental Design	***Zahra Valizadeh***	***P5***
Design of Experiment for Optimized Ultrasound-Assisted Extraction of Natural Pigments	***Zeynab Rasuli***	***P6***
Optimized Extraction of Pectin from Apple Pomace Using Response Surface Methodology for Sustainable Industrial Applications	***Elham Teymorirad***	***P7***
		P8

Machine Learning Classification of Solid-Phase Microextraction (SPME) Types Based on Operational and Analytical Parameters from Experimental Data	***Erfan Hashemi***	
Multivariate image – QSAR of phenyl piperidine derivatives as potent dual NK1R antagonists/serotonin transporter (SERT) inhibitor	***Neda Ladooni***	***P9***
PH-Dependent Sensor Array Based on Carbon Quantum Dots for Detection and Discrimination of Amines	***Tahereh Rasoulizadeh***	***P10***
Solvent-tuned Sensor Arrays Based on colored Carbon Quantum Dots for Discrimination of Multiple Metal Ions	***Tahereh Rasoulizadeh***	***P11***
Experimental design for optimization of the removal of Rhodamine B using a novel GQD incorporated Hydrogel	***Reza Dadashi***	***P12***
Rank Annihilation Factor Analysis for Kinetic Spectrophotometric Determination of MCPA herbicide in Unknown Samples	***Nima Karimnejad***	***P13***
Application of artificial neural network for the simultaneous determination of two rizatriptan and sumatriptan drugs via gold nanoparticle aggregation-based colorimetric sensing	***Masoud Esmailpak***	***P14***
Chemometric modeling of an electrochemical aptasensor for gluten detection using simulated electrochemical sensor data and comparison with experimental results	***Masoumeh Yousef-Zadeh***	***P15***
		P16

ComplexGNN: Hybrid Graph and Descriptor Learning for Interpretable Bandgap Prediction in Inorganic Crystals	***Dorsa Benvan***	
Machine Learning–Driven Discovery of a Potent Lead Compound for GPR17 via Graph Neural Networks and Multistage Structural Validation	***Noushin Agh Babaie***	***P17***
Using Machine Learning in the In-Silico Design of Selective Dopamine Receptor Ligands: Advancements in Targeted Therapies for Neurological and Psychiatric Disorders	***Melika F. Aghdam***	***P18***
Spice Adulteration Detection Using Hyperspectral and Multispectral Imaging	***Mahshid Ziaee jazi***	***P19***
Seed quality assessment using hyperspectral and multispectral imaging techniques	***Yeganeh Nazari***	***P20***
AI-Driven Design of PCSK9 Allosteric Inhibitors for Atherosclerosis via Graph Neural Networks and Structure-Based Validation	***Behnaz Horri***	***P21***
Leveraging Machine Learning and Graph Neural Networks for the Discovery of JAK2 Inhibitors: Structure and Ligand-Based Approaches to Target Inflammatory and Autoimmune Diseases	***Shaghayegh K. Nezhad***	***P22***
Enhancing pH Measurement Accuracy Using Hyperspectral Imaging and commercial pH Strips	***Arshia Rostampour***	***P23***
Comparative investigation of BORGEN norms	***Hamideh Bakhshi***	***P24***

Multi-criteria evaluation of combined preprocessing techniques for first-order spectroscopic calibration	***Mohammad Khademloo***	***P25***
Feasible band boundaries computation in bilinear matrix decomposition using essential data	***Somaye Vali Zade***	***P26***
Evaluation of Classical and Statistical Mechanics-Based Models for Adsorption Data: Insights into Accurate Fitting and Physical Interpretation	***Fatemeh Khabbazi***	***P27***
Machine Learning Bases Optimal Design of Carbon Nanosheets	***Ebrahim Nemati-Kande***	***P28***
Oriented Hyper-Cyclic Geometry with Curvature: A Weighted Directed Graph Model for Molecular Asymmetry and Chirality	***Khalil Shahbazpoor***	***P29***
A Fluorescent Concentration-Based Sensor Array Utilizing Hydrogen Peroxide-Etched Sulfur Quantum Dots (SQDs-H_2O_2) for Simultaneous Detection of Chromium(VI) and Bismuth(III) Ions via Multivariate Calibration	***Sima Mojarrad***	***P30***
Molecular Modeling and Design of Novel Inhibitors for GABAA Receptors	***Farshad Jafari***	***P31***
Artificial neural networks coupled Grey wolf optimizer for modeling of reactive red 198 adsorption onto novel poly(o-toluidine) / strontium hexaferrite /analsim nanocomposite	***H. Alijani***	***P32***

Virtual screening study of some GSK-3 inhibitors, such as Benzamide, Pyrazolopyrimidine, and β-phenylalanine derivatives, as anti-diabetic agents	***M. Piria***	***P33***
The application of machine learning to identify the chemical features that affect the stability of low-dimensional perovskite layers	***Mahmoud Samadpour***	***P34***
Optimization of Cd Adsorption by Chitosan Using Response Surface Methodology (RSM) Based on Box-Behnken Model	***Marziyeh Piri***	***P35***
Explainable Artificial Intelligence (XAI) for Regulatory-Compliant Chemometric Models in Pharmaceutical Analysis	***Nastaran Bahadori***	***P36***
Prediction of blood-brain barrier penetration coefficient of some central nervous system drugs using molecular docking modeling and quantitative structure-activity relationship	***Pezhman Maghouli***	***P37***
FastCCS: A Deep Learning Framework for Accurate Collision Cross-Section Prediction from SMILES with Broad Applications in Cheminformatics and Omics	***Amir Aghajan***	***P38***
Enhanced Rhodamine B Removal by Green-Synthesized Cu/TiO_2 Adsorbent: Evaluation of Adsorption Isotherms	***Seyed Mehdi Sajjadi***	***P39***
Multiple Linear Regression and Radial Basis Function Neural Networks for Identifying Effective Parameters and Ensuring Robustness in LC–MS/MS Simultaneous Analysis	***Rabee Mahdavi***	***P40***

CCD-Based Integrated Extraction–Encapsulation of β-Carotene from Carrot: A Chemometrics Approach humanized	***Zahra Sadri***	***P41***
A CCD-based two-stage DOE framework for the extraction and stabilization of lycopene from tomato pomace using ultrasound-assisted extraction and solid lipid encapsulation	***Fariba Abdi Saray***	***P42***
Explainable Machine Learning–Guided Optimization of µSPE-DLLµE for Phenol Extractiasadion	***Marziyeh Hosseinchi***	***P43***
Deep Learning-Enabled Vis-NIR Hyperspectral Imaging for Fast and Accurate Detection, Classification and Quantification of Dimethoate Residues in Apples	***Parya Omrani***	***P44***

Poster presentation		
Thursday, October 16, 2025		
Topic	***Presenter***	**ID No.**
Focused Molecular Docking Study of Minocycline and Doxycycline with Selected Cyclodextrins: Toward Efficient Host–Guest Complexation	***Melika asadian***	***P45***
Harnessing γ-Cyclodextrin for Enhanced Macrolide Antibiotics Delivery and Environmental Mitigation: A Computational Perspective	***Melika asadian***	***P46***

Prediction of lipid/water distribution coefficient of the organic compounds using the Monte Carlo approach and hybrid optimal descriptors	***Shahram Lotfi***	***P47***
QSPR modeling of chromatography retention indices of plant essential oils using SMILES and graph invariants	***Shahram Lotfi***	***P48***
Monte Carlo algorithm in the prediction of Value of Binding Constants of Organic Ligands to Beta-Cyclodextrin	***Shahram Lotfi***	***P49***
Simultaneous analysis of four cardiovascular drugs in synthetic mixtures and polypill tablets using triple devisor mean centering and derivative spectrophotometric techniques	***Maryam Maghfourian***	***P50***
SPA-PL-MARS QSAR Modeling of Xanthenone Derivatives as HCT-15 Colorectal Cancer Inhibitors: An Integrated Variable Selection and Nonparametric Regression Approach	***Maryam Allahverdi***	***P51***
Classification of Extra Virgin Olive Oil Adulteration Using FTIR Spectroscopy and Machine Learning Algorithms	***Esra modares Askari***	***P52***
Alternative Conditional Expectation Method for Modeling the QSAR (IC50) of Oxazolidinone Derivatives as 17β-Hydroxysteroid Dehydrogenase Type 3 Inhibitors	***Mohaddeseh Habibi***	***P53***
Comparison of Regression Methods (ACE, SVR, PLS, MLR) in QSAR, a Kinesin-5 Inhibitor Case Study	***Mohammadreza Pourmohammad***	***P54***
		P55

QSAR Modeling and Prediction of the Hallucinogenic Potency of Amphetamine Derivatives Using a Docking-Guided QSAR Approach	***Zahra Imanikia***	
Predicting the Viscosity of Magnesium Oxide-Based Nanofluids in Deep Eutectic Solvents Using Quantitative Structure-Property Relationship Modeling	***Mohammad Hossein Hajian***	***P56***
Discrimination of Iranian black teas through volatile compound fingerprints and chemometric modeling	***Adineh Aminianfar***	***P57***
Selective Tankyrase Inhibitors for Wnt-Driven Cancers: QSAR, Docking, Virtual Screening, and ADMET-Based Selection	***Narjes Najafi***	***P58***
Development of a chemometrics classification model in order to identification of diabetic patients for their saliva infrared spectrum	***Sepideh Khodkavandi***	***P59***
Untargeted Metabolomics Based on NMR Spectroscopy and Supervised Kohonen Network for Plasma Biomarker Identification in Early Breast Cancer Detection	***Maryam Kashi***	***P60***
Exploring Independence in Area of Feasible Solutions of Three-Component Systems: Calculations and Significance	***Fatemeh Sadat Hashemi-Nasab***	***P61***
Design the ElectroSensor- Liquid Phase Micro Extraction chip for Morphine detection as a model drug in urine sample	***Maryam Mehrban***	***P62***

A Hybrid NARX Neural Network and Metaheuristic Optimization Approach for Predicting the Galvanostatic Charge-Discharge Dynamics of Supercapacitors	***Hamed Azimi***	***P63***
Application of Group LASSO as dimension reduction method in QSAR Studies using a novel descriptor grouping strategy	***Farzaneh Kia***	***P64***
Application of Group Exponential LASSO and as bi-level feature selection method in QSAR Studies	***Farzaneh Kia***	***P65***
Effect of Descriptor Grouping Strategies on Bi-Level Feature Selection and Predictive Performance in GEL-Based QSAR Modeling	***Farzaneh Kia***	***P66***
Rapid Methanol Detection in Ethanol Solutions via Non-Linear Raman Spectroscopy Modeling	***Somaye Vali Zade***	***P67***
Correlation between Lubricant Viscosity Index and Mooney Viscosity of Polymers: A Predictive Model using Artificial Neural Networks	***Mohammad Hossein Aghamohammadi***	***P68***
Viscosity Enhancement of MWCNT/SiC Hybrid Nanofluids in SAE 20W50 Engine Oil: Experimental Analysis and Machine Learning Modeling	***Mohammad Hossein Aghamohammadi***	***P69***
QSAR and molecular docking studies for a chemometrics approach to the investigation of pyrimidine derivatives as novel inhibitors of anticancer drugs	***Dariush Fathali***	***P70***
Application of preprocessing methodes in image processing for simultaneous determination of three drugs	***Parisa Shakhsari***	***P71***

Application of preprocessing methods in image processing in QSAR and drug design	***Parisa Shakhsari***	***P72***
Quantitative structure activity relationship study of p38α MAP kinase inhibitors	***Maryam Nouri majd***	***P73***
Three-dimensional QSAR study on XIa (FXIa) inhibitors and design of new compounds based on CoMFA, CoMSIA and molecular docking	***Ladan Kazemnezhad***	***P74***
Three-dimensional QSAR study on iazolo-pyrimidine derivatives as WRN inhibitors for the treatment of MSI tumors and design of new compounds based on CoMFA and molecular docking	***Mobina Haji Abbasi***	***P75***
Dispersive liquid-liquid microextraction for the extraction of ibrutinib drug from real samples using UV-Vis spectrophotometry and optimization by experimental design	***Sara Kharat Zabardast***	***P76***
Dispersive liquid-liquid microextraction (DLLME) for extraction and preconcentration of Enzalutamide in real samples using the UV-Vis spectrophotometry	***Sanaz Saber Siyahpoush***	***P77***
Classification of some opioids using principal component analysis	***Morteza Bahram***	***P78***

Chemometrics Education in the Era of Digital Chemistry

Hamid Abdollahi
Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, Zanjan

E-mail address: abd@iasbs.ac.ir

Abstract

In an era defined by unprecedented digital immersion, where machine learning has won Nobel Prizes in physics and chemistry and over 10,000 publications on machine learning in chemistry emerge annually, we stand at the threshold of transformative chemical education. Digital chemistry, a paradigm shift that fuses computational tools, cheminformatics, artificial intelligence (AI), and automation, accelerates material discovery and handles vast, heterogeneous datasets from automated laboratories. This renders traditional wet-lab practices insufficient on their own. In this landscape, chemometrics emerges as a pivotal cornerstone of data-driven chemical analysis, bridging the gap between empirical synthesis and scalable, reproducible ML pipelines. This integration enhances prediction, reproducibility, and innovation amid a backdrop of rising scientific retractions.

However, undergraduate chemistry curricula often lag behind, sidelining chemometrics until advanced stages and depriving students of essential computational thinking and programming skills. This talk advocates for the urgent integration of chemometrics from the introductory level and the redesign of courses to incorporate AI and large language models (LLMs) for smarter prompting. Drawing on expert insights, we emphasize hands-on, low-barrier projects, such as building simple machine learning (ML) models for analyzing simulated data. These projects foster mentorship and dual wet-dry perspectives.

Ultimately, chemometricians must deepen expertise in modern ML techniques to evolve efficient, interpretable methods. By embracing these shifts, we future-proof graduates, empowering them to lead data-centric chemistry and drive interdisciplinary breakthroughs.

From Latent Structures to Neural Networks: Bridging Chemometrics and Deep Learning in Chromatography and Imaging

Hadi Parastar
Department of Chemistry, Sharif University of Technology, Tehran, Iran
E-mail address: h.parastar@sharif.edu

ABSTRACT

Modern analytical instrumentation in chromatography and imaging has revolutionized our ability to address increasingly complex analytical problems. High-dimensional analytical techniques such as gas chromatography-ion mobility spectrometry (GC-IMS) [1] and visible-near infrared hyperspectral imaging (Vis-NIR HSI) [2] are increasingly central to addressing complex challenges in food authentication and safety. However, the richness and scale of these datasets demand advanced data analysis strategies that balance interpretability with predictive power.

In the present contribution, I will explore how classical chemometric tools, particularly principal component analysis (PCA) and multivariate curve resolution (MCR), can be effectively combined with modern deep learning (DL) approaches such as convolutional neural networks (CNNs) and U-Net architectures to deliver robust solutions [3]. Applications include the analysis of volatile organic compounds (VOCs) with GC-IMS for the geographical discrimination and fraud detection of saffron, olive oil, and citrus fruits, as well as the use of Vis-NIR HSI for saffron authentication and the rapid detection of pesticide residues on agricultural products, illustrated with apples as a case study.

These examples highlight the complementary roles of chemometrics and DL: PCA and MCR enable feature extraction, data compression, and interpretation, while CNNs and U-Nets provide powerful nonlinear modeling and image-based segmentation. Together, they establish a synergistic workflow capable of transforming raw, high-dimensional chromatographic and imaging data into actionable insights for food quality, authenticity, and safety.

Keywords: Chemometrics, Deep learning, Gas chromatography-ion mobility spectrometry, Hyperspectral imaging, Convolutional neural network, Food authentication.

References:

[1] H. Parastar, P. Weller, “How Machine Learning and Gas Chromatography-Ion Mobility Spectrometry Form an Optimal Team for Benchtop Volatilomics”, Anal. Chem., 97, (2025), 1468-1481.
[2] C. Yang, D.F. Barbin, N. Watson, M. Povey, X. Zou, “Hyperspectral imaging and deep learning for quality and safety inspection of fruits and vegtables: a review”, J. Agric. Food. Chem., 73, (2025), 10019-10035.
[3] B. Debus, H. Parastar, P. Harrington, D. Kirsanov, “Deep learning in analytical chemistry”, Trends Anal. Chem., 145, (2021), 116459.

Comparative Study of Nonlinear Hard and Soft Discrimination in Genomic Cancer Classification

Somaye Vali Zade[a], Parvin Shahsavand[a] , Hossein Rastegar[b], Hamid Abdollahi[c]*

[a] *Halal Research Center of IRI, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran*

[b] *Cosmetic Products Research Center, Iran Food and Drug Administration, MOHE, Tehran, Iran, Future Studies Group, the Academy of Medical Sciences of The I.R. Iran, Tehran, Iran*

[c] *Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran*

[*] E-mail: smy.valizade@gmail.com

ABSTRACT

Accurately classifying cancer subtypes based on transcriptomic data is central to modern genomic medicine. However, the high dimensionality and noisy nature of gene expression profiles complicates the design of robust classifiers. This study introduces a hybrid framework that integrates nonlinear regression models—Kernel Partial Least Squares (KPLS), Radial Basis Function Neural Networks (RBF-ANN), and Feed-Forward Neural Networks (FF-ANN)—along with hard and soft discrimination approaches, to improve the prediction of cancer subtypes from gene expression data. The nonlinear models project gene expression features into a reduced, class-oriented response space. This space is then subjected to linear discriminant analysis (LDA) for deterministic classification or quadratic discriminant analysis (QDA) for probabilistic classification. We demonstrate the method's utility on two benchmark datasets, Acute Leukemia and Small Round Blue-Cell Tumors (SRBCT), showing that combining nonlinear transformation with soft discrimination improves classification accuracy, particularly in ambiguous or overlapping genomic profiles. This framework is a promising strategy for improving the robustness and interpretability of diagnoses in high-throughput transcriptomic studies.

Keywords : Gene expression classification, Nonlinear transformation, Cancer subtypes, Transcriptomics, Functional genomics, Soft discrimination,

References (Times New Roman 10pt Bold):

[1] Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., ... & Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science, 286(5439), (**1999**), 531–537.

[2] S. V. Zade, E Forooghi, M Ranjbar, B Jannat, F. Hashempour-boltork, H. Abdollahi, A Combined Classification Modeling Strategy for Detection and Identification of Extra Virgin Olive oil Adulteration Using Raman Spectroscopy, Chemometrics and Intelligent Laboratory Systems 240 (**2023**): 104903.

A transfer learning workflow based on MCR and CNN to overcome limitations in Vis-NIR HSI data modeling: a case study

Fatemeh Sadat Hashemi-Nasab[1,2], Hananeh Chogan[1], Hadi Parastar[1*]

1.Department of Chemistry, Sharif University of Technology, P.O. Box 11155-9516, Tehran, Iran
2. Department of Analytical Chemistry, Faculty of Chemistry, Kharazmi University, Tehran 15719-14911, Iran
Email: h.parastar@sharif.edu

Hyperspectral Imaging (HSI) is a powerful tool for capturing detailed spatial and spectral information, providing a wealth of data for complex chemical analysis [1]. This high-dimensional information across multiple wavelengths makes HSI invaluable for analyzing the chemical composition and structure of samples [2]. However, analyzing HSI data presents unique challenges due to its high dimensionality. Machine learning (ML) and deep learning (DL), particularly convolutional neural networks (CNNs), have shown promise in processing HSI data, but their performance often depends on large datasets to achieve reliable accuracy, which can be costly and time-intensive to collect, particularly in chemical experiments with limited samples [3]. Moreover, models and algorithms that preserve the spatial context of HSI data can yield better results, as they retain the inherent structural relationships crucial for accurate classification [4]. To address data scarcity, transfer learning and data augmentation offer feasible solutions that enable DL models to achieve high performance with smaller sample sizes [5]. In this study, a novel strategy for classifying saffron samples based on their geographical origin using Vis-NIR hyperspectral imaging (HSI) data was explored. Initially, multivariate curve resolution (MCR) was applied to extract key features. Subsequently, MCR distribution maps were combined with pure spectra from the 600–750 nm range to construct cubic data for each sample. These cubic datasets were then sliced into 150 2D images per sample, allowing spatial structure to be preserved while the dataset was expanded. To prevent data leakage, it was ensured that slices from the same sample did not appear in both the training and test sets. A dataset comprising 256 saffron samples from four regions was used, and a CNN model was trained, achieving a test accuracy of 96%. This model was saved as a base model for future use. Afterwards, the same CNN architecture was trained using a smaller dataset of 38 new samples from three regions. Without the use of transfer learning, an accuracy of 62% was obtained; however, after the pre-trained model was applied and fine-tuned, the accuracy increased to 92%. The effectiveness and practicality of transfer learning and data augmentation for handling limited HSI datasets were demonstrated. These approaches enable the application of DLmethods, such as CNNs, to real-world challenges in food authenticity and classification. ***Keywords:*** Deep learning, Transfer learning, CNN, Hyperspectral Imaging, Food Authenticity

References

[1] Fatemeh Sadat Hashemi-Nasab and Hadi Parastar, *Food Chem.*, **2022**, 393, 133450.
[2] Bhatti, Uzair Aslam, Meng-Tze Huang. *Expert Syst. Appl.* **2023**, 229, 120496.
[3] Chen, X., Du, H., Liu, Y. *Sci. Rep.,* **2024**, 14, 7209.
[4] Zhang, J., Zhao, J., Lin, H., Tan, Y. & Cheng, J.-X. *J. Phys. Chem. Lett.,* **2020**, 11, 8573–8578.
[5] Bansal, Ms. Aayushi and Sharma, Dr. Rewa and Kathuria, Dr. Mamta, *ACM Comput. Surv.*, **2022**, 54, 29.

10th Iranian Biennial Chemometrics Seminar
10TH BIENNIAL CHEMOMETRICS SEMINAR

Multivariate Calibration Models for ATR-FTIR Spectroscopy Analysis of Olive Oil Adulterated with Corn, Sunflower, and Palm Oils

Esra modares Askari[a], Mohammadreza Khanmohammadi Khorrami[a], Mahsa Mohammadi[a], Miguel de la Guardia[b]*

[a]*Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran*

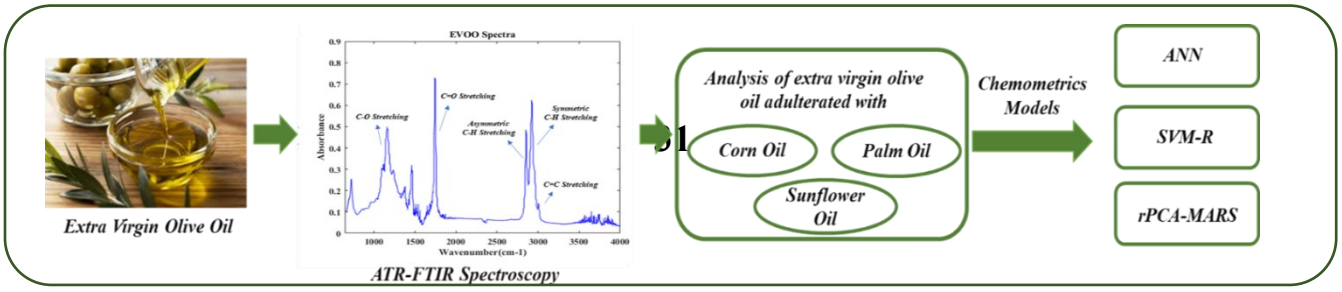
[b]*Departamento de Quı´mica Analı´tica, Research Building, Universitat de Vale`ncia, 50 Dr. Moliner St., 46100 Burjassot, Vale`ncia, Spain*

*Corresponding author: E-mail addresses: m.khanmohammadi@SCI.ikiu.ac.ir (M.Khanmohammadi. Khorrami)

ABESTRACT

Extra virgin olive oil (EVOO), a vital agricultural product in the Mediterranean region, is highly valued for its biological properties as well as its sensuous appeal [1]. For financial benefit, dishonest market participants may adulterate EVOO with less expensive plant or nut oils [2]. Numerous analytical methods have been developed for the identification and measurement of adulteration and authentication of olive oil. Chemometrics and spectroscopy have been utilized successfully in recent years to detect adulteration. The ease with which multi-component investigations may be undertaken is what makes Fourier transform infrared FTIR such an effective quantitative tool [3]. These procedures have proven to be rapid and non-destructive for verifying the validity of culinary oils, notably EVOO. While normal approaches should include the environment, the suggested strategy allows for quick, uncomplicated measurements with no negative environmental impact. Thus, rapid analysis using attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy has been developed for the detection of EVOO adulterated with corn oil, sunflower oil (SFO), and palm oil. Measurements were conducted on EVOO and samples adulterated with varying contents of oils, ranging from 1.0 to 35% by weight. Adulterated mixtures were vigorously stirred to ensure homogeneity before analysis. ATR-FTIR spectra were analyzed using a robust principal component analysis–multivariate adaptive regression splines (rPCA-MARS) model to assess the adulterated samples. Artificial neural networks (ANN) and support vector machines regression (SVM-R) were used to predict olive oil adulteration. The performance of these models, along with linear and cubic rPCA-MARS models, was evaluated using R^2 and RMSE. The linear rPCA-MARS model demonstrated the highest accuracy in predicting adulteration, with an R^2pred value of 0.970 and MSE (pred**)** of 4.950. This model outperformed ANN, SVM-R, and piecewise cubic rPCA-MARS models based on key statistical metrics.

Keywords: Extra virgin olive oil adulteration, ATR-FTIR, EVOO, SVM-R, ANN, rPCA-MARS



References

[1] G. Gurdeniz, F. Tokatli, and B. Ozen, "Differentiation of mixtures of monovarietal olive oils by mid-infrared spectroscopy and chemometrics", Eur. J. Lipid Sci. Technol, 109, (**2007**), 1194–1202.

[2] D. L. García-González, R. Aparicio-Ruiz, and R. Aparicio, "Virgin olive oil—Chemical implications on quality and health", Eur. J. Lipid Sci. Technol., 110, (**2008**), 602–607.

[3] F. R. Van de Voort, "Fourier transform infrared spectroscopy applied to food analysis", Food Res. Int, 25, (**1992**), 397–403.

Functions of Chemometrics within the Emerging Paradigm of Machine Learning and Artificial Intelligence in Pharmaceutical Sciences

Somaieh Soltani

Pharmacy Faculty, Tabriz University of Medical Sciences, Tabriz, Iran Research Center for Computational and Theoretical Molecular Engineering, Khazar University, Baku, Azerbaijan

Since its formal naming in 1971, chemometrics has provided pharmaceutical sciences with a systematic approach to extract information from complex data. Its applications span experimental design and formulation optimization; analytical characterization, including signal resolution, quantitative assays, and solid-state or dissolution profiling; pharmacokinetics and clinical decision support, such as PK/PD modeling and bioequivalence evaluation; manufacturing and quality through process monitoring, continuous manufacturing, and real-time release testing; safety and biomarker discovery; as well as multi-omics integration, mechanism elucidation, target mapping, and more recently, digital twins for predictive workflows. Historically, these efforts relied on structured and labeled datasets, with emphasis on supervised models and exploratory unsupervised methods such as PCA. While nonlinear and hybrid approaches were developed, the intrinsic heterogeneity of drug-related phenomena such as solid-state transitions, dissolution kinetics, multicomponent formulations, and biological variability continues to stretch the limits of classical frameworks.

The rise of artificial intelligence (AI) has accelerated this evolution. Transformer-based attention models and self-supervised architectures demonstrate how spectroscopy, microscopy, omics, and process analytical technology (PAT) signals can be integrated into unified, predictive, and auditable decision frameworks. Regulatory initiatives reflected this shift in ICH Q14 lifecycle management of multivariate models into analytical procedure development, while ICH Q13 highlighted the role of advanced modeling frameworks (chemometrics and AI) in continuous manufacturing and real-time release testing. Together, these developments reposition chemometrics not as supportive but as a central scientific and decision-making framework in pharmaceutical sciences.

In this keynote, I will trace the path from the earliest applications of chemometrics to today's AI-enhanced measurement era, where representation learning, active learning, and attention-based integration turn signals from physics, chemistry, and biology into clinically and industrially meaningful predictors. Looking ahead, I will argue for the re-emergence of "self-supervised chemometrics" and its role in the development of foundation models capable of learning directly from heterogeneous, unstructured data and transferring knowledge across instruments, formulations, and regulatory contexts. Realizing this vision will require not only technical advances but also attention to interpretability, robustness, and regulatory acceptance. Ultimately, chemometrics integrated with AI can harmonize pharmaceutical measurement into a multi-disciplinary orchestra whose performance is judged by its contribution to safer, more effective, and more equitable health outcomes.

Key words: Self supervised Chemometrics; Pharmaceutical Sciences; Process Analytical Technology (PAT); Machine Learning; Artificial Intelligence.

AI-Driven Design and Optimization of Novel Carbon Allotropes via Crystallographic Space Group Exploration and Machine Learning Potentials

Ebrahim Nemato-Kande[1]*, *Khalil Shahbazpoor*[2]*,
[1]*Department of Physical Chemistry, Chemistry Faculty, Urmia University, Urmia, Iran*
[2]*Department of Mathematics, Science Faculty, Urmia University, Urmia, Iran*

E-mail address (corresponding author): e.nemati@urmia.ac.ir
kh.shahbazpoor@urmia.ac.ir

ABSTRACT

This study presents a computational framework for the de novo design and optimization of hypothetical carbon allotropes using a high-throughput approach that integrates crystallographic symmetry operations, graph theory, and machine learning interatomic potentials (MLIPs). The methodology involves generating initial structures by populating unit cells with carbon atoms based on predefined hybridization states (sp, sp^2, sp^3) and symmetrically expanding them according to user-specified space groups [1]. A key innovation is the use of a distance-based penalty function during AI-driven relaxation, which enforces a hard minimum interatomic distance (1.0 Å) to ensure geometric plausibility. The framework employs MLIPs, such as CHGNet [2], MACE [3], or M3GNet [4], via the Atomic Simulation Environment (ASE) to perform energy minimization efficiently without the computational cost of density functional theory (DFT). This approach enables the exploration of a vast configurational space, including varied lattice parameters, numbers of orbits, and connectivity patterns, to identify stable carbon polymorphs. The automated pipeline outputs optimized structures in CIF format, facilitating further analysis and validation. This work demonstrates the potential of AI-enhanced computational methods in accelerating the discovery of novel materials with tailored properties, leveraging graph theory for initial structure generation and MLIPs for accurate force field predictions [4].

Keywords: Carbon allotropes, machine learning interatomic potentials, crystallographic space groups, structural optimization, materials discovery,

References:

[1] Ojih, J., Al-Fahdi, M., Yao, Y., Hu, J., & Hu, M. (2024). Journal of Materials Chemistry A, 12, 8502–8515.
[2] Deng, B., Zhong, P., Jun, K., Riebesell, J., Han, K., Bartel, C. J., & Ceder, G. (2023). Nature Machine Intelligence, 5(12), 1031–1041.
[3] Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., & Csanyi, G. (2022). Advances in Neural Information Processing Systems, 35, 2332–2345.
[4] Chen, C., & Ong, S. P. (2022). Nature Computational Science, 2(11), 718–728.

Moving toward Smart Sensors for eDiagnostics and eMonitoring

Hamed Golmohammadi *
Nanosensor Bioplatforms Laboratory, Chemistry and Chemical Engineering Research Center of Iran, 14335-186 Tehran, Iran

golmohammadi@ccerci.ac.ir, ha.golmohammadi@gmail.com

ABSTRACT

The development of biomedical sensors as highly promising devices/tools in early diagnosis and therapy monitoring of many diseases and disorders has recently witnessed unprecedented growth. By learning from defects of current sensors in smart diagnostics of pandemics and despite the recent substantial progress in this field, achieving the ideal sensors capable of addressing the growing expectations and needs on the eve of Industry 5.0 for smart diagnostic/monitoring sensing and moving faster toward decentralized/personalized healthcare diagnostics systems is still in high demand. An ideal smart biomedical sensor must be cost-effective, sensitive, selective, easy-to-use/sample collection/dispose, and also real-time connected for remote monitoring, reporting, programming, and controlling. Indeed, a smart biomedical sensor will be able to precisely and accurately monitor and analyze the concentration of some important (bio)chemicals and physical parameters of the body, alert any changes in their levels, and share the obtained data with users, pre-determined people, doctors, and medical centers. According to the definition of ideal smart biomedical sensors as remotely monitorable, reportable, programmable and controllable (bio)sensors, as well as the definition of smart diagnostics as a diagnosis based on smart data analyzed/resulted from the decentralized real-time connected diagnostic devices/tests' big data, the criterion of real-time connectivity (i.e. the ability of diagnostic sensors to connect to a reader or mobile phone) can be considered as the most important criterion and the main prerequisite toward development of smart biomedical sensors. Indeed, to achieve smart biomedical sensors, while having all EASSURED criteria of WHO, they must be connected to each other, other smart systems, and eventually decision-makers/end-users to be able to generate the smart data -as a prerequisite data for smart decisions-with the help of digital technologies including deep learning (DL), big data analytics (BDA), machine learning (ML), artificial intelligence (AI), edge/fog/cloud computing, and cybersecurity. Due to remotely access, resource sharing, time saving, etc., internet is the most ideal solution for such inter/intra-connectivity through a concept called the Internet of Things (IoT) [1-6].
Herein, an overview is provided to highlight the importance and necessity of an inevitable transition in the era of digital health/Healthcare 4.0 towards smart eDiagnostics and eMonitoring and how to approach it via new digital technologies and IoT gateways (smartphones, readers), etc. Furthermore, alongside highlighting the characteristics/criteria that should be met by the developed sensors towards smart sensing, the challenging issues ahead are delineated along with a comprehensive outlook on this necessary field.

Keywords: Smart sensor, Internet of things, Artificial intelligence, Diagnostics, Digital health, Smartphone

References

[1] T. Naghdi, S. Ardalan, Z. A. Adib, A. R. Sharifi, H. Golmohammadi, Moving toward smart biomedical sensing, Biosensors and Bioelectronics 223 (**2023**) 115009.
[2] A. R. Sharifi, S. Ardalan, R. S. Tabatabaee, S. S. Gorgani, H. Yousefi, K. Omidfar, M. A. Kiani, C. Dincer, T. Naghdi, H. Golmohammadi, Smart Wearable Nanopaper Patch for Continuous Multiplexed Optical Monitoring of Sweat Parameters, Analytical Chemistry 95 (**2023**) 16098-16106.
[3] S. Ardalan, M. Hosseinifard, M. Vosough, H. Golmohammadi, Towards smart personalized perspiration analysis: An IoT-integrated cellulose-based microfluidic wearable patch for smartphone fluorimetric multi-sensing of sweat biomarkers, Biosensors and Bioelectronics 168 (**2020**) 112450.
[4] H. Golmohammadi, F. Parnianchi, A. R. Sharifi, T. Naghdi, R. S. Tabatabaee, M. Peyravian, S. Kashanian, Spicy Recipe for At-Home Diagnostics: Smart Salivary Sensors for Point-of-Care Diagnosis of Jaundice, ACS Sensors 6 (**2024**) 3455–3464.
[5] T. Naghdi, A. R. Sharifi, R. S. Tabatabaee, E. Azizi, M. A. Kiani, H. Golmohammadi, Neonatal Point-of-Care Testing. TrAC Trends Anal. Chem. 175 (**2024**) 17732.
[6] R. S. Tabatabaee, T. Naghdi, M. Peyravian, M. A. Kiani, H. Golmohammadi, An Invisible Dermal Nanotattoo-based Smart Wearable Sensor for eDiagnostics of Jaundice, ACS nano 18 (**2024**), 28012-28025.

Recent Developments in Multivariate Curve Resolution Alternating Least Squares (MCR-ALS)

Roma Tauler, e-mail: Roma.Tauler@idaea.csic.es, Chemometrics for environmental omics: http://www.ch4eo.info/, MCR-ALS web page http://mcrals.wordpress.com/download/mcr-als-toolbox/, Institute of Environmental Assessment and Water Research, IDAEA-CSIC, http://www.idaea.csic.es/, Jordi Girona 18-26, Barcelona 08034, SPAIN

The **Multivariate Curve Resolution Alternating Least Squares (MCR-ALS)** method is a powerful analytical approach for decomposing complex data sets into simpler, interpretable components[1,2]. One of its key strengths lies in resolving systems where multiple components interact, making it particularly valuable in areas such as **chemical analysis**, **spectroscopy**, and other **multi-dimensional** data applications[3].

MCR-ALS achieves this by applying various **constraints** to guide the resolution process and enhance model interpretability. A fundamental assumption is **multilinearity**[4], used to analyze multiway data. This assumes that the system can be described as a linear combination of factors defined in each data way, mode, direction, or dimension. However, in real-world scenarios, strict multilinearity may not always hold. To address this, **relaxing multilinearity**[5,6] allows the model to better accommodate non-linearities, increasing its flexibility and applicability. Incorporating **chemical knowledge** into the modeling process is another important strategy. This can involve the use of **hard modeling**, where physicochemical assumptions are explicitly imposed, or **soft modeling**, which allows for more adaptable constraints without enforcing strict physical laws—thus enabling more nuanced interpretations. Combining both approaches in **hard-soft modeling** often leads to an optimal solution for real-world problems[7,8]. **Sparsity constraints**[9,10] can also be applied to ensure that only a limited number of significant variables dominate the solution, reducing complexity and improving interpretability. Additionally, to derive **quantitative information**, MCR-ALS integrates strategies for **calibration and correlation have been proposed**[11,12]. In cases where significant portions of data are missing or incomplete, adapted **least squares** approaches have been described to maintain the stability and accuracy of the model[13,14].

One of the most critical challenges in MCR-ALS is the estimation of **uncertainties**, particularly those stemming from **rotation ambiguities**[15]—the fact that multiple mathematically valid solutions may exist, each corresponding to a different rotation of the component profiles across the various data modes. Another source of uncertainty is **error propagation**[16], where initial measurement errors can influence the final resolved profiles. **Interval estimation** of these uncertainties in multivariate curve resolution solutions have been proposed based on principles of error propagation in linear least squares[17]. To address these challenges, the **MCR-WALS**[18] (Weighted ALS) and the **MLPCA-MCR-ALS**[19] methods have been developped, introducing uncertainty estimates directly into the MCR calculations to improve robustness. In parallel, methods such as **MCR-(N)BANDS**[20,21] have been specifically designed to handle both **rotation ambiguities** and **error propagation** under conditions of noise and uncertainty.

MCR-ALS continues to have wide-reaching and impactful applications across various fields. In the analysis of large **environmental monitoring**[22,23] data sets, it enables the resolution, apportionment, and assessment of pollution sources. In **GCxGC** and **LCxLC multidimensional chromatographic analysis**[24,25], MCR-ALS enhances the separation power of these techniques. In **hyperspectral imaging**[26], it allows for the extraction of spatially and spectrally resolved chemical information. In **mass spectrometry**, the **ROIMCR**[27,28] method aids in the meaningful analysis of the acquired complex MS data by focusing on the extraction of the regions of interest, which together with the **MSident**[29] method empower non-targeted analysis approaches for new compound discovery and identification. More recently, MCR-ALS is

demonstrating significant effectiveness in the analysis of **ion mobility spectrometry (IMS)** data when coupled with **liquid chromatography-mass spectrometry (LC-IMS-MS)**[30], providing valuable insights into ion behavior and separation of isomeric and isobaric chemical compounds. Overall, the ongoing development of MCR-ALS, with improved constraints, uncertainty modeling, and domain-specific adaptations, has consolidated its position as a versatile and powerful tool in modern analytical science.

REFERENCES

1. Multivariate curve resolution applied to second order data, R. Tauler, **Chemometrics and Intelligent Laboratory Systems**, 1995, 30, 133-146. https://doi.org/10.1016/0169-7439(95)00047-X

2. MCR-ALS GUI 2.0: new features and applications J Jaumot, A de Juan, R Tauler **Chemometrics and Intelligent Laboratory Systems** 2015, 140, 1-12 https://doi.org/10.1016/j.chemolab.2014.10.003

3. Chemometrics applied to unravel multicomponent processes and mixtures. Revisiting latest trends in multivariate resolution, A.de Juan, R. Tauler R., **Analytica Chimica Acta**, 2003, 500, 195-210. DOI:10.1016/S0003-2670(03)00724-4

4. Multivariate curve resolution of multiway data using the multilinearity constraint R.Tauler **J.Chemometrics**, 2020, e3279. https://doi.org/10.1002/cem.3279

5. Flexible Trilinearity Alignment (FTA) and Shift Invariant Transformation (SIT) onstraints in Three-Way Multivariate Curve Resolution Data Analysis. X.Zhang, R.Tauler **Journal of Chemometrics**, 2024; 0:e3581 https://doi.org/10.1002/cem.3581

6. A unifying framework for modelling non-negative bi-linear, tri-linear and "in-between" data in chemometrics. Part I: Theoretical framework and concepts P.A. Schneide, N.Gallagher, J.L.Hinrich, R.Bro, R.Tauler. Chemometrics and Intelligent Laboratory Systems 2025, 265, 105492. https://doi.org/10.1016/j.chemolab.2025.105492

7. Combining hard- and soft-modelling to solve kinetic problems. A.de Juan, M.Maeder, M.Martínez and R.Tauler. **Chemometrics and Intelligent Laboratory Systems**, 2000, 54, 123-141. https://doi.org/10.1016/S0169-7439(00)00112-X

8. Kinetic studies of nitrofurazone photodegradation by hybrid hard- and soft-multivariate curve resolution applied to UV-spectral data. Michele De Luca, Sílvia Mas, Giuseppina Ioele, Gaetano Ragno, Romà Tauler. **International Journal of Pharmaceutics**, 2010, 386, 1-2, 99-107. http://dx.doi.org/10.25258/ijpqa.14.3.27

9. Application of a sparseness constraint in multivariate curve resolution e Alternating least squares. Siewert Hugelier, Sara Piqueras, Carmen Bedia, Anna de Juan, Cyril Ruckebusch. Analytica Chimica Acta 2018, 1000, 100e108. http://dx.doi.org/10.1016/j.aca.2017.08.021

10. Evaluating different sparsity measures for resolving LC/GC-MS data in the context of multivariate curve resolution. Ahmad Mani-Varnosfaderani, Mohammad Javad Masroor, Yadollah Yamini. **Chemometrics and Intelligent Laboratory Systems**, 2020, 200, 104004. https://doi.org/10.1016/j.chemolab.2020.104004

11. Alternative calibration approaches for lC-MS quantitative determination of coeluted compounds in complex

environmental mixtures using multivariate curve resolution
Emma Peré-Trepat, Silvia Lacorte and Romà Tauler. **Analytica Chimica Acta**, 2007, 595, 228-237. https://doi.org/10.1016/j.aca.2007.04.011

12. Application of the area correlation constraint in the MCR-ALS quantitative analysis of complex mixture samples. M. Bayat, M. Marin-Garcia, J.B. Ghasemi, R. Tauler
Analytica Chimica Acta, 2020, 1113, 52-65. https://doi.org/10.1016/j.aca.2020.03.057

13. Dealing with missing data blocks in Multivariate Curve resolution. Towards a general framework based on a single factorization model. A.Gómez-Sánchez, C.Ruckebusch, R.Tauler, A.de Jua. **Trends in Analytical Chemistry** 179 (2024) 117869.
https://doi.org/10.1016/j.trac.2024.117869

14. Multivariate curve resolution of incomplete and partly trilinear multiblock datasets.
A. Queral-Beltran, M.Marín-García, S.Lacorte, R.Tauler. Chemometrics and Intelligent Laboratory Systems, 2024, 247, 105081
https://doi.org/10.1016/j.chemolab.2024.105081

15. Uniqueness and rotation ambiguities in Multivariate Curve Resolution methods
Hamid Andollahi and Romà Tauler. **Chemometrics and Intelligent Laboratory Systems**, 2011, 108, 100–111. DOI:10.1016/j.chemolab.2011.05.009

16. Error propagation along the different regions of multivariate curve resolution feasible solutions. M.Dadashi, H.Abdollahi, R.Tauler. **Chemometrics and Intelligent Laboratory Systems**, 2017, 162, 203-213. https://DOI: 10.1016/j.chemolab.2017.01.009

17. Interval estimation in multivariate curve resolution by exploiting the principles of error propagation in linear least squares. A.Mani-Varnosfaderani, E.S.Park b, R.Tauler. **Chemometrics and Intelligent Laboratory Systems** 206 (2020) 104166
https://doi.org/10.1016/j.chemolab.2020.104166

18. A comparison of positive matrix factorization and the weighted multivariate curve resolution method. Application to environmental data. Ivana Stanimirova, Romà Tauler and Beata Walczak. **Environmental Science and Technology**, 2011, 45, 10102-10110.
https://doi.org/10.1021/es201024m

19. Maximum Likelihood Principal Component Analysis as initial projection step in Multivariate Curve Resolution analysis of noisy data. Mahsa Dadashi, Hamid Abdollahi and Romà Tauler. **Chemometrics and Intelligent Laboratory Systems**, 2012, 118, 33–40; https://doi.org/10.1016/j.chemolab.2012.07.009; Application of Maximum Likelihood Multivariate Curve Resolution to noisy data sets. Mahsa Dadashi, Hamid Abdollahi and Roma Tauler **Journal of Chemometrics**, 2013, 27, 34-41. https://doi.org/10.1002/cem.2489

20. MCR-BANDS: An user friendly MATLAB program for the evaluation of rotational ambiguity in Multivariate Curve Resolution. Joaquim Jaumot and Romà Tauler.
Chemometrics and Intelligent laboratory Systems, 2010, 103, 96–107. https://doi.org/10.1016/j.chemolab.2010.05.020

21. N-BANDS: A new algorithm for estimating the extension of feasible bands in multivariate curve resolution of multicomponent systems in the presence of noise and rotational ambiguity. Alejandro C. Olivieri and Romà Tauler. **J. of Chemometrics** 2021;35:e3317. https://doi.org/10.1002/cem.3317

22. Chemometric analysis for river water quality assessment at the intake of drinking water treatment plants Stefan Platikanov, David Baquero, Susana González, Jordi Martín-Alonso, Miquel Paraira, Jose Luis Cortina, Roma Tauler

Science of The Total Environment 2019, 667, 552-562
https://doi.org/10.1016/j.scitotenv.2019.02.423

23. Temporal air quality (NO2, O3, and PM10) changes in urban and rural
stations in Catalonia during COVID-19 lockdown: an association with human mobility and satellite data. Eva Gorrochategui, Isabel Hernandez · Eva Pérez-Gabucio · Sílvia Lacorte · Romà Tauler. **Environmental Science and Pollution Research**
https://doi.org/10.1007/s11356-021-17137-7

24. Chemometrics comparison of gas chromatography with mass spectrometry and comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry Daphnia magna metabolic profiles exposed to salinity
H.Parastar, E.Garreta-Lara, B.Campos, C.Barata, S.Lacorte, R.T auler
Journal of Separation Science, 2018, 41, 11 (Epub 2018 Feb 27)
https://DOI:10.1002/jssc.201701336

25. Untargeted Comprehensive Two-Dimensional Liquid Chromatography Coupled with High-Resolution Mass Spectrometry Analysis of Rice Metabolome Using Multivariate Curve Resolution
M.Navarro-Reig, J.Jaumot, A.Baglai, G.Vivó-Truyols, P.J.Schoenmakers, R.Tauler
Analytical Chemistry, 2017, 89 (14), pp. 7675-7683.
http://DOI: 10.1021/acs.analchem.7b01648

26. Multimodal multisample spectroscopic imaging analysis of tumor tissues using multivariate curve resolution
C. Bedia, A.Sierra, R.Tauler
Chemometrics and Intelligent Laboratory Systems 2021, 15, 104366
https://doi.org/10.1016/jchemolab.2021.104366

27. ROIMCR: a powerful analysis strategy for LC-MS metabolomic datasets
E. Gorrochategui; J. Jaumot; R. Tauler.
BMC Bioinformatics, 2019, 20, 1,256
https://doi.org/10.1186/s12859-019-2848-8

28. MSroi: A pre-processing tool for mass spectrometry-based studies
Miriam P erez-Cova, Carmen Bedia, Dwight R. Stoll, Roma Tauler, Joaquim Jaumot
Chemometrics and Intelligent Laboratory Systems 2021, 215, 104333
https://doi.org/10.1016/j.chemolab.2021.104333

29. MSident: Straightforward identification of chemical compounds from MSresolved
Spectra. Carlos Perez-Lopez, Antoni Ginebreda, Joaquim Jaumot, Flavia Yoshie Yamamoto, Damia Barcelo, Roma Tauler. **Cheometrics and Intelligent Laboratory Systems, 2024, 105063**
https://doi.org/10.1016/j.chemolab.2024.105063

30. Comprehensive Data Analysis using Regions of Interest Multivariate Curve Resolution for Ion Mobility Mass Spectrometry: analysis of plastic additives.
Ana Torres-Agullo, Roma Tauler, Silvia Lacorte, Yamil Simón-Manso Unpublished results; submitted for publication

Identification and Measurement of H_2S, COS and SO_2 in Natural Gas Using a Machine Learning Empowered Colorimetric Sensor Array

*Ahmad Moslehipour, Bahram Hemmateenejad**

Chemistry Department, Shiraz University, Shiraz, 71454, Iran

E-mail address (corresponding author): hemmatb@shirazu.ac.ir

ABSTRACT

Accurate detection of hydrogen sulfide (H_2S), sulfur dioxide (SO_2), and carbonyl sulfide (COS) in gas sweetening plants is critical for ensuring natural gas quality and operational safety. However, the corrosive nature of these sulfur-containing gases presents significant analytical challenges. H_2S is a toxic, corrosive, and flammable substance. SO_2 and COS are colorless gases with noxious odors. Exposure to lower concentrations of these gases can result in various physiological effects, including eye irritation, sore throat, cough, nausea, shortness of breath, and pulmonary edema [1]. Sensor arrays, defined as a group of sensors deployed in a specific geometry pattern, have demonstrated extensive applications across a range of disciplines. Irrespective of the biosensors, which produce specific signals for target analytes based on a lock-and-key approach, sensor arrays employ a series of cross-reactive sensors that respond to a collection of analytes but with differential selectivity. Colorimetric sensor arrays (CSAs) are a class of sensor arrays using colorful colorimetric reagents (e.g., acid-base, metallochromic, solvatochromic, and redox indicators) that change color when interacting with the analytes [2, 3]. This study introduces a simple, cost-effective colorimetric sensor array capable of discriminating these gases even in complex, corrosive matrices. The array employs three cross-reactive indicators—murexide, chloranilic acid, and hexaammineruthenium(III) chloride—each immobilized in separate vacuum flasks. Upon gas exposure, distinct color changes are captured via smartphone imaging, and RGB values are extracted for analysis. Principal component analysis (PCA) successfully differentiates the gases based on their unique response patterns. Quantitative analysis is achieved by correlating the first PCA scores with gas concentrations, yielding high linearity (R^2 = 0.9945 for H_2S, 0.9447 for COS, and 0.9968 for SO_2). The array demonstrates robustness in multicomponent gas mixtures, with validation against gas chromatography showing excellent agreement (relative standard deviation and relative error <7%). This method offers a rapid, inexpensive, and reliable alternative for real-time monitoring of sulfur-containing gases in industrial settings, particularly where conventional techniques are impractical. Its portability, combined with machine learning-assisted analysis, positions it as a promising tool for on-site gas quality control.

Keywords: Sulfur-Containing Gases, Colorimetric Sensor Arrays, Hydrogen sulfide (H_2S), Sulfur Dioxide (SO_2), Carbonyl Sulfide (COS)

References

[1] Wang, Kehang, Chenghao Bi, Lev Zelenkov, Xiuzhen Liu, Mingzhao Song, Wenxin Wang, Sergey Makarov, and Wenping Yin. "Fluorescent sensing for the detection and quantification of sulfur-containing gases." ACS sensors 9, no. 11 (2024): 5708-5727.

[2] Shojaeifard, Zahra, Ahmad Moslehipour, and Bahram Hemmateenejad. "Optical and electronic nose and tongue." In Nature-Derived Sensors, pp. 173-200. Elsevier, 2025.

[3] Hemmateenejad, Bahram, and Knut Baumann. "Data pre-processing for paper-based colorimetric sensor arrays." Chemometrics and Intelligent Laboratory Systems 254 (2024): 105237.

A Deep Learning Workflow Based on Convolutional Neural Networks for GC-IMS Feature Extraction and Classification

*Farbod Bayat-Afshary, Nima Naderi, Hadi Parastar**
Department of Chemistry, Sharif University of Technology, Tehran, Iran;
E-mail address: h.parastar@sharif.edu

ABSTRACT

The analysis of volatile organic compounds (VOCs) using gas chromatography-ion mobility spectrometry (GC-IMS) has become prominent due to its high sensitivity, rapid analysis and minimal sample preparation requirements. Compared to GC-MS, it offers greater portability and energy efficiency [1]. However, the inherent second-order nature of GC-IMS data introduces significant analytical hurdles in peak alignment, feature extraction, and classification. Classical chemometric/machine learning (ML) approaches often address these challenges by unfolding the second-order data structure into a first-order vector. This process of vectorization, while necessary for many traditional algorithms, compromises the data's native structure and leads to the loss of valuable spatial information between peaks, which can negatively impact model performance [2]. This study introduces a deep learning (DL) workflow based on convolutional neural networks (CNNs) that circumvents this limitation by treating the chromatographic data as image-like structures, thereby preserving the rich, second-order information [3]. While image-based DL has been applied to other chromatographic techniques, its application to GC-IMS data remains underexplored. To demonstrate the efficacy of this approach, two different CNN architectures, simplified and complex, were evaluated using olive oil and citrus GC-IMS datasets. Initially, a complex CNN architecture was tested, but due to the relatively simple nature of GC-IMS data, it quickly overfitted and achieved low accuracy; therefore, a simplified model was adopted. The simplified CNN achieved classification accuracies of 96.4% for olive oil and 98.3% for citrus samples. To address the common challenge of limited sample sizes, a data augmentation strategy was developed specifically tailored to the nature of GC-IMS data. The method simulates plausible instrumental drift, enhancing model performance for datasets with subtle chemical differences without introducing artifacts. A core innovation of this work is a novel strategy to transform the CNN from a "black box" into an interpretable analytical tool. Saliency maps are first used to identify discriminative features on a per-sample basis. Critically, this work then pioneers the use of principal component analysis (PCA) to consolidate these individual maps into a single, representative loading image for each class.

Keywords: Gas chromatography-ion mobility spectrometry, Deep learning, Volatile organic compound, Convolutional neural network, Chemometrics.

References:

[1] H. Parastar, P. Weller, “Towards greener volatilomics: Is GC-IMS the new Swiss army knife of gas phase analysis?”, Trends Anal. Chem., 170, (2024), 117438.
[2] H. Parastar, P. Weller, “How Machine Learning and Gas Chromatography-Ion Mobility Spectrometry Form an Optimal Team for Benchtop Volatilomics”, Anal. Chem., 97, (2025), 1468-1481.
[3] B. Debus, H. Parastar, P. Harrington, D. Kirsanov, “Deep learning in analytical chemistry”, Trends Anal. Chem., 145, (2021), 116459.

Assessing Matrix Effects with Multivariate Curve Resolution

Ali Pahlevan, Somaiyeh Khodadadi Karimvand and Hamid Abdollahi

Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, Zanjan

E-mail address: abd@iasbs.ac.ir

Abstract

Multivariate calibration models in analytical chemistry often suffer from matrix effects due to sample composition and instrumental variations, causing prediction inaccuracies from spectral and concentration mismatches. Traditional strategies like standard addition and local modeling inadequately address both issues simultaneously. This study presents a novel MCR-ALS-based matrix-matching approach for robust matrix effect assessment, decomposing spectral data into concentration and spectral profiles to evaluate spectral similarity via net analyte signal projections and Euclidean distance, and concentration alignment across calibration subsets and unknowns. Validated on simulated data, NIR corn spectra, and NMR alcohol mixtures, the method identifies optimal calibration sets minimizing matrix-induced errors like spectral shifts and interferent overlaps, outperforming conventional techniques in accuracy and robustness. By leveraging MCR's bilinear decomposition to isolate interferent impacts, this framework enhances predictive precision in diverse matrices, offering versatile applicability in analytical challenges such as environmental and pharmaceutical analysis.

Multi-Class Classification of Odor Perception Using Graph Neural Networks and Molecular Graph Representations

Sogol Meknatkhah[a], Mahia Vazifeh Solout[b], Negar Amiri[b], Jahan B. Ghasemi[b*]

[a]*Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

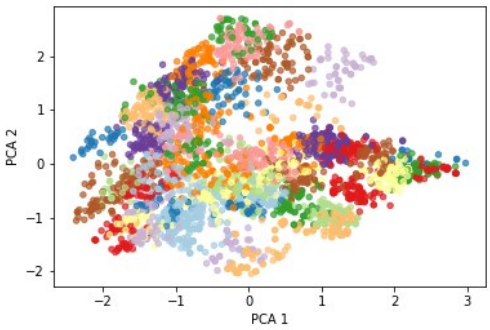
[b]*Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran*

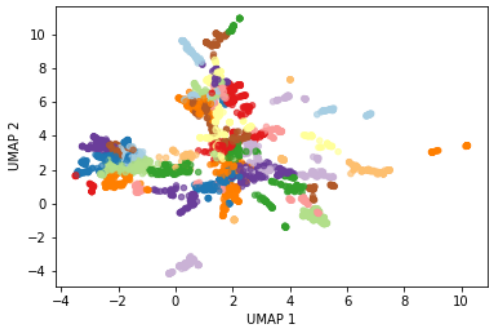
E-mail address (corresponding author): Jahan.ghasemi@ut.ac.ir

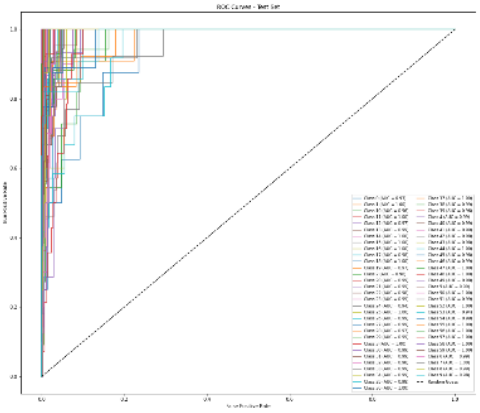
ABSTRACT

Olfaction is the oldest and most fundamental chemical sense, allowing lifeforms to perceive and interpret their surroundings. This sense has long fascinated scientists, inspiring numerous theories aimed at elucidating its underlying mechanisms [1]. Unlike vision or hearing, predicting olfactory perception from molecular structures remains a major challenge. Some molecules with different functional groups can share similar odors, while structurally similar molecules may produce completely different perceptions. Understanding the quantitative structure–odor relationship (QSOR) is therefore critical not only for deciphering the coding principles of olfactory information but also for predicting and designing smells and flavors in applications such as food technology [2]. Recent advances in artificial intelligence have opened new avenues for decoding human olfactory perception from the chemical features of odorant molecules []. However, the accuracy of these predictions strongly depends on the molecular representation used. Among various encoding methods, including SMILES strings and molecular fingerprints, molecular graphs have shown particular promise because they capture atomic- and bond-level features. Graph Neural Networks (GNNs) are specifically designed to extract information from graph-structured data, making them well-suited for QSOR studies. In this work, we present a GNN-based model for predicting the odor of molecules using molecular graphs enriched with atomic and bond-level features. A dataset comprising 5,955 molecules annotated with 160 smell percepts was employed to develop a multi-class classification model. The dataset was randomly split into training, validation, and test sets, and the model achieved an F1 score of 0.59 on the test set, demonstrating promising predictive performance. Data visualization using PCA and UMAP highlights the underlying structure of the odor space, and the AUC-ROC curves confirm the model's effectiveness in distinguishing between different smell classes. Overall, this study demonstrates the potential of GNNs for advancing the quantitative prediction of olfactory perception, providing a foundation for future efforts in molecular design and sensory analysis.

Keywords: Olfactory prediction, Graph neural network, QSOR, Molecular graphs







References:

[1] Hoehn, Ross D., David E. Nichols, Hartmut Neven, and Sabre Kais. "Status of the vibrational theory of olfaction." Frontiers in Physics 6 (2018): 25.

[2] Zhang, Mengji, Yusuke Hiki, Akira Funahashi, and Tetsuya J. Kobayashi. "A deep position-encoding model for predicting olfactory perception from molecular structures and electrostatics." npj Systems Biology and Applications 10, no. 1 (2024): 76.

[3] Achebouche, Rayane, Anne Tromelin, Karine Audouze, and Olivier Taboureau. "Application of artificial intelligence to decode the relationships between smell, olfactory receptors and small molecules." Scientific reports 12, no. 1 (2022): 18817.

Biologically Informed Prediction of Drug Synergy: A Transcriptomic Resistance Signature Approach

*Mozhgan Mozaffarilegha, Sajjad Gharaghani**

Laboratory of Bioinformatics and Drug Design (LBD), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

s.gharaghani@ut.ac.ir

ABSTRACT

Combination therapies are essential for treating complex diseases such as cancer, where they improve efficacy, minimize resistance, and reduce toxicity by leveraging synergistic effects. Despite their therapeutic potential, identifying effective combinations is highly challenging due to the vast number of possible pairings and the substantial costs associated with experimental validation. Computational approaches, particularly machine learning (ML) and deep learning (DL), have advanced drug synergy prediction by integrating heterogeneous datasets and modeling drug–cell line interactions. However, most existing models primarily rely on drug-specific features such as chemical structures, while incorporating limited functional drug information and cellular context. To address these limitations, we propose a novel framework that integrates Drug Resistance Signatures (DRS) as biologically informed representations of drug information. By embedding transcriptomic resistance profiles into predictive models, DRS provide a mechanistically relevant and comprehensive feature space for characterizing drug action. We systematically evaluated the predictive utility of DRS across multiple ML algorithms, including LASSO, Random Forest, AdaBoost, and XGBoost, as well as the DL model SynergyX. Their performance was benchmarked against conventional drug signatures and chemical structure–based descriptors. Validation on independent datasets—ALMANAC, O'Neil, OncologyScreen, and DrugCombDB—confirmed the robustness, accuracy, and generalizability of the DRS-based framework. Results demonstrate that resistance-informed transcriptomic features substantially enhance predictive performance compared with structure-only approaches, while also offering improved interpretability of underlying drug mechanisms. Overall, this study underscores the value of integrating biologically grounded resistance signatures into computational models for drug synergy prediction. The proposed framework provides a powerful and generalizable strategy to accelerate the identification of effective combination therapies, thereby supporting the development of more precise and clinically relevant cancer treatment.

Keywords: synergy, machine learning, deep learning, drug resistance signatures, cancer therapy

From punch card to AI in chemistry

Jahan B. Ghasemi

Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran.

E-mail address: Jahan.ghasemi@gmail.com

Chemometrics has always aimed to extract meaningful chemical knowledge from complex data, evolving continuously alongside advances in computation and instrumentation. This presentation traces the journey of chemometrics from its analog and manual calculation origins through the early days of digital computing to the birth of pioneering algorithms that defined the field. The transition into the digital revolution expanded the scope and precision of chemometric analysis, setting the stage for the current era—where artificial intelligence (AI) and machine learning have become transformative tools in chemistry. However, this shift introduces new challenges, including the need for large, high-quality, and well-labeled datasets, and the critical task of embedding fundamental chemical principles into AI models. Rather than replacing traditional methods, the integration of AI with classical chemometric techniques such as PCA and PLS represents a synergistic evolution. By bridging historical insight with modern innovation, chemometricians can ensure that the discipline remains at the forefront of scientific discovery and chemical data interpretation.

Divergent Biological Conclusions: A Comparison of PLS-DA and MCR-ALS in Metabolomics

Maryam Khoshkam[1,2]*

[1] *Chemistry Department, Faculty of Sciences, University of Mohaghegh Ardabili , 56199-11367, Ardabil, Iran*
[2]*Department of Environmental Science, Faculty of Science, University of Zanjan, 45371-38791, Zanjan, Iran*

E-mail address (corresponding author): khoshkam@uma.ac.ir, khoshkam@znu.ac.ir

ABSTRACT

Metabolomics relies heavily on chemometric tools to extract meaningful biological information from complex spectral data [1]. This study investigates the impact of chemometric method selection [1-5] on biological interpretation by comparing two widely used techniques: Partial Least Squares-Discriminant Analysis (PLS-DA) and Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) [6]. Both methods were applied to a ^{1}H-NMR metabolomics dataset to discriminate between samples from 20 healthy individuals and 81 diseased patients. Although both models demonstrated strong predictive power, they yielded divergent sets of statistically significant metabolites. Consequently, subsequent pathway analysis implicated distinct biological pathways. PLS-DA emphasized perturbations in glycine, serine, and threonine metabolism; pyrimidine metabolism; histidine metabolism; aminoacyl-tRNA biosynthesis; pantothenate and CoA biosynthesis; and beta-alanine metabolism. In contrast, MCR-ALS highlighted alterations in caffeine metabolism; sulfur metabolism; pyruvate metabolism; glycine, serine, and threonine metabolism; and arginine and proline metabolism. These discrepancies are attributed to the algorithms' inherent mechanics: PLS-DA prioritizes variables with high covariance to the class label, while MCR-ALS resolves components based on spectral conformity and bilinearity. This study demonstrates that the choice of chemometric model is a critical factor that can steer biological conclusions. We advocate for method triangulation—applying complementary chemometric approaches—to enhance the validity and depth of pathway analysis and ensure robust biological interpretation in metabolomics.

Keywords: 1HNMR based metabolomics, PLS-DA, MCR-ALS, pathway analysis, biological analysis, COVID 19

References

[1] C.Y. Hamany Djande, C. Pretorius, F. Tugizimana, L.A. Piater, I.A. Dubery, "Metabolomics: A Tool for Cultivar Phenotyping and Investigation of Grain Crops", Agronomy, 10, (**2020**), 831-861.
[2] P. García-Pérez, P.P. Becchi, L. Zhang, G. Rocchetti, L. Lucini, "Metabolomics and chemometrics: The next-generation analytical toolkit for the evaluation of food quality and authenticity", Trends in Food Science & Technology, 147, (**2024**), 104481-104495.
[3] V. Maritha, P.W. Harlina, I. Musfiroh, A.M. Gazzali, M. Muchtaridi, "The Application of Chemometrics in Metabolomic and Lipidomic Analysis Data Presentation for Halal Authentication of Meat Products", Molecules, 27 ,(**2022**), 7571-7584.

[4] N. Sulaiman, M. Sanny, C.B. Yeoh, A.H. Ahmad Tarmizi, N. Nordin, F.K. Ahmad Bustamam, R.A. Abd Razak, A.F. Abdull Razis, A. Khatib, "Application of metabolomics and chemometrics for authentication of vegetable oils and adulteration detection: a review", European Food Research and Technology, (**2025**).
[5] L. Yi, N. Dong, Y. Yun, B. Deng, D. Ren, S. Liu, Y. Liang, "Chemometric methods in data processing of mass spectrometry-based metabolomics: A review", Analytica Chimica Acta, 914, (**2016**) 17-34. "
[6] Y. Pérez, M. Casado, D. Raldúa, E. Prats, B. Piña, R. Tauler, I. Alfonso, F. Puig-Castellví, MCR-ALS analysis of (1)H NMR spectra by segments to study the zebrafish exposure to acrylamide", Analytical and Bioanalytical Chemistry, 412, (2020), 5695-5706.

Application of support vector machine in predicting potentiometric selectivity (Mg^{2+}/Ca^{2+}) of some amide ligands

*Reza Mahmoudzadeh Laki, Eslam Pourbasheer**

Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili, P.O. Box 179, Ardabil, Iran

E-mail address (corresponding author): e.pourbasheer@uma.ac.ir

ABSTRACT

In this study, in order to model the prediction of potentiometric selectivity (Mg^{2+}/Ca^{2+}) of several amide ligands, quantitative structure-activity relationship (QSPR/QSAR) approaches were used along with two stepwise methods and genetic algorithm as descriptor selection techniques [1]. Experimental data and molecular structures were entered into the modeling after calculating and screening the descriptors. The SW method identified five descriptors and GA five descriptors (with two common items) as key parameters [2]. Various models including SW-MLR, GA-MLR, GA-SVM were created, and by identifying an outlier and removing it, GA-MLR and GA-SVM models were re-developed [3]. The models obtained by GA-MLR and GA-SVM (with one outlier removed) had R^2_{train}=0.802 and R^2_{test}=0.757, RMSE=0.606 and 0.705, and Q^2=0.440, respectively, indicating the high predictive power and fit of the models. These values were a significant improvement compared to the reference paper (R^2=0.66 and RMSE=0.53 even after removing 11 outliers). Regression coefficient analysis showed that positive coefficients were associated with increased selectivity and negative coefficients with inhibitory effects of structural or electronic features. In addition to higher accuracy, the findings provide important mechanistic insights into the role of molecular features in selectivity.

Keywords: Amide ligands; Genetic Algorithm; Potentiometric selectivity; QSPR; Support Vector Machine

References:

[1] Ganjali, M. R., Alahdadi, I., Aghazadeh, M., & Pourbasheer, E., All solid state (ASS) polymeric membrane sensor (PMS) for the monitoring of nanomolar nickel concentration. International Journal of Electrochemical Science, **2015**. 10(9), 6913-6923.

[2] Pourbasheer, E. and R.M. Laki, *QSPR prediction of fullerene derivatives solubility in different solvents using the genetic algorithm− Support vector machine.* Journal of Molecular Liquids, **2025**. 15(426), 127307.

[3] Goyal, S., Effective software defect prediction using support vector machines (SVMs). International Journal of System Assurance Engineering and Management, **2022**. 13(2), 681-696.

Application of multivariate tools in the discrimination of geographical origin of Iranian polyfloral honeys by analysis of physicochemical and biochemical characterization

Saber Amiri *, Mohammad Alizadeh Khaledabad, Ali Jahanbakhsh Hesar
Department of Food Science and Technology, Faculty of Agriculture, Urmia University, Urmia, Iran
* Corresponding author; e-mail address: sa.amiri@urmia.ac.ir

Abstract

Honey is one of the foods that, firstly, is very prone to fraud due to the great diversity of honey in terms of flower origin and geographical origin, and secondly, it has a special chemical composition that is sensitive to climatic conditions as well as harvest. Given that the geographical origin can be effective in determining the quality characteristics of honey. As a result of mixing (intentionally or unintentionally) different types of honey and the inability to identify its origin is unacceptable and is considered a form of fraud. So far, physicochemical properties of honey from different geographical origins of the world have been investigated but no information could be found on the identification of geographical origins by physicochemical and biochemical properties of Iranian honeys. This study was aimed to develop a chemometric model for geographical origin classification of the region's honey based on their physicochemical and biochemical parameters. A total of 126 honey samples were collected from three major provinces West Azerbaijan (Urmia), East Azerbaijan (Arasbaran), Ardabil (Sareyn) of Iran. The results of PCA showed that the first four components had Eigenvalue higher than one (8.44, 3.40, 1.32, and 1.15, respectively) The first principal component portioned 44.41%, the second principal component 17.89% and the third principal component 6.99%, the fourth principal component 6.04%, with total variance portion equal to 75.34%, of the total variance portion. The results of cluster varibles divided variles in 5 sets. Cluster 1 had 9 variables including Chroma (C*), Conductivity (mS/cm), b, Total Phenol (mg), FRSA (%), L, Insoluble (%), Fru/Glu, and Diastase (DN). Cluster 2 had 3 variables including Total Sugar (%), Reducing Sugar (%), and HMF (mg/kg). Cluster 3 had 2 variables including Sucrose (%) and a. Cluster 4 had 3 variables including pH, Moisture (%), and Ash (%). Cluster 5 had 2 variables including Hue (h*) and Total Acidity (meq/kg). The most significant variable of each set (Chroma (C*), Total Sugar (%), pH, Hue (h*), and Sucrose (%)) was selected for discrimination analysis. The QDA for the discriminate the geographical origin of honey samples using Chroma (C*), Total Sugar (%), pH, Hue (h*), and Sucrose (%) matrix. All the 126 samples were in the training set with 1 (0.79%) misclassified and -2LL was 9.80. Just one of the EA honey samples was incorrectly classified in WA set. The scatter plot for the Chroma (C*), Total Sugar (%), pH, Hue (h*), and Sucrose (%) matrix is shown in Fig. 2 (A). The first two canonicals explained 100.00% of the total variance with significant effect on discrimination ($p<0.05$). Variance corresponding to canonical 1 was 67.75%, the variance corresponding to canonical 2 was 32.25%.

Keywords: Chemometrics, Geographical origin, Multi-floral blossom honey, Principal component analysis, Quadratic discriminant analysis.

Rank Annihilation Factor Analysis for Kinetic Spectrophotometric Determination of MCPA herbicide in Unknown Samples

*Nima Karimnejad, Reza Dadashi, Morteza Bahram**
Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

*Corresponding author (*E-mail address):* m.bahram@urmia.ac.ir

ABSTRACT

The contamination of soil and water with herbicides is a major concern for both the environment and human health. MCPA (4-(2-methyl-4-chlorophenoxy)acetic acid) is one of the most widely used phenoxy herbicides for controlling broadleaf weeds in cereal crops and pastures. Its extensive application increases the risk of leaching into natural resources, potentially harming non-target species and posing risks to human health [1]**. Therefore, reliable and sensitive methods are needed for monitoring MCPA in unknown and complex samples. Classical analytical methods such as HPLC, GC, and mass spectrometry are accurate but often costly, time-consuming, and require complicated sample preparation. Kinetic spectrophotometry is a simpler and more affordable alternative, but it can be limited by overlapping spectra and matrix interferences, which reduce measurement accuracy** [2]**. To address this challenge, we used Rank Annihilation Factor Analysis (RAFA), a powerful chemometric tool for resolving multivariate data. RAFA eliminates interfering signals by rank reduction, enabling accurate identification and quantification of the analyte even in complex datasets. When combined with kinetic spectrophotometry, RAFA provides improved sensitivity, selectivity, and robustness compared with traditional approaches** [3]**. In this study, kinetic spectral data were obtained using the Marquis reagent, a mixture of formaldehyde and concentrated sulfuric acid, which produces characteristic color reactions with different compounds. This reagent facilitated the generation of distinct kinetic data suitable for RAFA processing. Herbicide samples containing MCPA were analyzed, and the method showed strong performance in resolving overlapping signals and minimizing interference. Overall, the combination of RAFA with kinetic spectrophotometry represents an efficient, cost-effective, and reliable strategy for the determination of MCPA in**

environmental and biological matrices. This approach offers a promising alternative to conventional chromatographic methods and has significant potential for practical applications in environmental monitoring and agricultural safety.

Keywords: Herbicide, Marquis reagent, MCPA, RAFA

References

[1] P.A. Morton, C. Fennell, R. Cassidy, D. Doody, O. Fenton, P. Mellander, P. Jordan, A review of the pesticide MCPA in the land-water environment and emerging research needs, Wiley Interdiscip. Rev. Water. 7 (2020) e1402.

[2] O. Selcuk, Y. Demir, C. Erkmen, S. Yıldırım, B. Uslu, Analytical methods for determination of antiviral drugs in different matrices: recent advances and trends, Crit. Rev. Anal. Chem. 52 (2022) 1662–1693.

[3] A. Afkhami, L. Khalafi, Spectrophotometric investigation of the effect of β-cyclodextrin on the intramolecular cyclization reaction of catecholamines using rank annihilation factor analysis, Anal. Chim. Acta. 599 (2007) 241–248.

A Comprehensive FTIR Spectral Database for Authentication and Quality Control of Edible and Essential Oils

*Niloofar Rahmani, Ahmad Mani-Varnosfaderani**

Chemometrics and Cheminformatics Laboratory, Department of Chemistry, Tarbiat Modares University, Tehran, Iran

E-mail address (corresponding author): a.mani@modares.ac.ir

ABSTRACT

Reliable methods for oil quality control and adulteration detection are critical for food authentication [1, 2]. This study aims to develop a comprehensive Fourier transform infrared (FTIR) spectral database of over 3000 samples from 30 edible and essential oils, which provide a valuable resource for detecting adulteration and quality control in the oil industry. Spectral similarity patterns among oil samples were explored using Spearman correlation and Euclidean distance, which revealed meaningful clustering and compositional differences based on FTIR spectral signatures. By applying advanced machine learning techniques, including support vector machine and random forest (RF), we successfully classified oils with 100% accuracy, distinguishing between different oil types based on their unique molecular fingerprints. FTIR data revealed clear differences between edible and essential oils in the fingerprint region (600–1500 cm^{-1}), reflecting their unique chemical compositions. We also used a similarity-based approach to estimate the proportion of each oil in specific query samples. The Euclidean-based analysis of 100 random oil blends accurately predicted the proportion of each oil in the complex mixtures. To detect refined sunflower oil adulteration, 600 binary blends of high-quality oils, including olive, avocado, coconut, and sesame, were made with adulteration levels ranging from 10% to 50%. Our results showed that the RF model provides highly accurate predictions, with the coefficient of multiple determination (R^2) values of 0.923-0.987 and the root mean square error (RMSE) values of 0.364-0.990. Furthermore, network analysis facilitated the visualization of spectral similarity structures among samples, which offered clear insights into clustering patterns. Overall, the developed database and analytical framework advance capabilities in oil authentication and offer practical value for food quality control.

Keywords: FTIR spectroscopy, Edible oils, Essential oils, Machine learning, Authentication, Adulteration detection.

References:

[1] Sota-Uba, I., Bamidele, M., Moulton, J., Booksh, K., & Lavine, B. K., Authentication of edible oils using Fourier transform infrared spectroscopy and pattern recognition methods, Chemometrics and Intelligent Laboratory Systems, 210, (**2021**), 104251.

[2] Mousa, M. A., Wang, Y., Antora, S. A., Al-Qurashi, A. D., Ibrahim, O. H., He, H.-J., . . . Kamruzzaman, An overview of recent advances and applications of FT-IR spectroscopy for quality, authenticity, and adulteration detection in edible oils. Critical reviews in food science and nutrition, 62, (**2022**), 8009-8027.

Machine Learning Enhanced Discrimination of Biothiol and Thiol Ratios Using Gold Nanorod Amalgamation

*Golara Akhondi, Afsaneh orouji, Mohammad Reza Hormozi-Nezhad**
Sharif University of Technology, department of Chemistry
E-mail address (corresponding author): hormozi@sharif.edu

ABSTRACT

Biothiols, with their thiol groups, play a key role in biological processes and health, and changes in their level can have significant effects on the body. [1] In this study, a gold nanorod-based colorimetric probe is introduced that, with the help of machine learning, enables the simultaneous identification, differentiation, and measurement of important biothiols including cysteine, glutathione, cystine, and glutathione disulfide, as well as thiol-to-disulfide ratios. The use of advanced machine learning algorithms such as linear discriminant analysis and partial least squares regression has enabled the complex spectroscopic patterns resulting from changes in the size and shape of nanorods to be accurately analyzed and interpreted. [2] These computational methods have been able to reliably extract subtle differences between biothiols and their ratios even at low concentrations, significantly increasing the accuracy and sensitivity of detection. The result of this clever combination of gold nanostructures with machine learning is the development of a cost-effective and highly sensitive probe that performs reliably and reliably even in complex biological samples. The statistical results of these methods show linear relationships with low detection limits and high analytical sensitivity for all biothiols and thiol-disulfide ratios, demonstrating the power of machine learning in improving the accuracy of biological diagnostics.

Keywords: machine learning, linear discriminant analysis, gold nanorods, colorimetric probe, biothiols

References

[1] Spearow, J. L.; Copeland, L. Improving Therapeutics for COVID-19 with Glutathione-Boosting Treatments That Improve Immune Responses and Reduce the Severity of Viral Infections. OSF Preprints **2020**, 2020, 10.
[2] Yang, C.; Zhang, H. A Review on Machine Learning−Powered Fluorescent and Colorimetric Sensor Arrays for Bacteria Identification. Microchim. Acta **2023**, 190 (11), 451.

Application of Central Composite Design in Modeling and Optimization of a Bio-Based Time–Temperature Indicator Using Fatty Acids

*Maryam Safari [1], Elham Jannatdoust [1, *], Farshad Kheiri [1], Morteza Bahram [2], Mahammad Sirousazar [1]*

[1] *Faculty of Chemical Engineering and Advanced Materials, Urmia University of Technology, Urmia, Iran.*

[2] *Faculty of Chemistry, Urmia University, Urmia, Iran.*

e.jannatdoust@uut.ac.ir

ABSTRACT

Time–temperature indicators (TTIs) have attracted increasing attention as essential tools in intelligent packaging systems, where maintaining product quality and safety is strongly dependent on temperature control. TTIs provide irreversible color changes that reflect the cumulative thermal history of products, thereby enabling consumers and producers to identify potential spoilage or loss of quality [1,2]. The development of reliable and environmentally friendly TTIs requires careful selection of components as well as robust modeling tools. Among statistical approaches, central composite design (CCD) is widely recognized for its ability to explore factor interactions, optimize experimental conditions, and generate predictive mathematical models with high accuracy.

In the present study, four biodegradable plant-based fatty acids - coconut (23 °C), peanut (3 °C), canola (–10 °C), and castor (–18 °C) - were employed to develop a bio-based TTI [3]. CCD was applied to investigate the main, interaction, and quadratic effects of these fatty acids on the melting point of their mixtures. The results revealed that canola fatty acid had the most significant linear effect, while the strongest interaction occurred between coconut and canola. The most pronounced quadratic effect was related to the squared term of canola. The developed model exhibited a high coefficient of determination ($R^2 > 0.95$), confirming its predictive reliability. Based on this model, the melting point of mixtures could be adjusted within the range of –15 to 20 °C. The optimized formulation provides an efficient and eco-friendly TTI with strong potential for monitoring and preserving temperature-sensitive products in intelligent packaging applications.

Keywords: Time–temperature indicator, Central composite design, Bio-based fatty acids,

References

[1] Pereira Jr, V.A., I.N.Q. de Arruda, and R. Stefani, Active chitosan/PVA films with anthocyanins from Brassica oleraceae (Red Cabbage) as Time–Temperature Indicators for application in intelligent food packaging. Food Hydrocolloids, 2015. 43: p. 180-188.

[2] Gerrans, J., et al., An efficient smart pharmaceutical packaging technology framework to assess the quality of returned medication through non-intrusively recording storage conditions after dispensation. Technologies, 2023. 11(3): p. 75.

[3] Guasch-Ferré, M., et al., Associations of monounsaturated fatty acids from plant and animal sources with total and cause-specific mortality in two US prospective cohort studies. Circulation research, 2019. 124(8): p. 1266-1275.

Optimization of Aluminum Anodizing Conditions and Evaluation of Potassium Permanganate-Loaded Anodic Alumina for Ethylene Removal in Active Packaging of Fruits and Vegetables

*Saeid Golmirzaei [1], Elham Jannatdoust [1, *], Farshad Kheiri [1], Mahammad Sirousazar [1]*

[1] Faculty of Chemical Engineering and Advanced Materials, Urmia University of Technology, Urmia, Iran.

e.jannatdoust@uut.ac.ir

ABSTRACT

Active packaging has emerged as an innovative approach in the food industry, aiming to extend the shelf life and preserve the quality of fresh produce. Ethylene, a key plant hormone, accelerates ripening and spoilage in fruits and vegetables, and its accumulation in storage environments can drastically reduce product longevity and quality. Therefore, developing effective adsorbents to remove ethylene is essential for maintaining postharvest quality. In this study, anodic alumina was synthesized as a nanoscale porous substrate for ethylene adsorption using an electrochemical anodizing method [1,2]. Three electrolytes -sulfuric acid, phosphoric acid, and oxalic acid- were evaluated, and the effects of critical variables including electrolyte type and concentration, anodizing time, and current density on the resulting structure were investigated. Phosphoric acid was found to produce anodic alumina with superior porosity and high specific surface area.

Potassium permanganate was subsequently loaded onto the anodic alumina to serve as an oxidizing agent [3]. Two loading methods, immersion and electrochemical, were compared, and the optimal method was selected based on loading efficiency. Structural and morphological characterization was performed using BET analysis for surface area and pore distribution, as well as SEM for microstructural evaluation. Ethylene adsorption performance was assessed using laboratory-synthesized ethylene, and removal efficiency was quantified through spectrophotometric measurements.

The results demonstrated that the optimized anodic alumina loaded with potassium permanganate exhibited high efficiency in ethylene removal. This study highlights that nanoscale anodic alumina functionalized with potassium permanganate provides a practical, sustainable, and scalable solution for active packaging applications, offering significant potential for the postharvest management of fresh produce.

Keywords: Active packaging, Ethylene adsorption, Anodic alumina, Postharvest quality

References

[1] Domagalski, J.T., E. Xifre-Perez, and L.F. Marsal, Recent Advances in Nanoporous Anodic Alumina: Principles, Engineering, and Applications. Nanomaterials, 2021. 11(2): p. 430.

[2] Álvarez-Hernández, M.H., et al., Potassium Permanganate-Based Ethylene Scavengers for Fresh Horticultural Produce as an Active Packaging. Food Engineering Reviews, 2019. 11(3): p. 159-183.

[3] Ebrahimi, A., et al., Preparation of novel nano–based films impregnated by potassium permanganate as ethylene scavengers: An optimization study. Polymer Testing, 2021. 93: p. 106934.124(8): p. 1266-1275.

Optimizing Anthocyanin Extraction from Prunus domestica Flesh Using Experimental Design

Zahra Valizadeh[1], *Khalil Farhadi*[1*], *Sima Najafi*[1]

[1] *Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran*

khalil.farhadi@yahoo.com

ABSTRACT

Plants are rich sources of bioactive compounds, including alkaloids, flavonoids, terpenoids, phenolics, and natural pigments. Among these, anthocyanins—water-soluble pigments from the flavonoid family—are highly valued in the food, pharmaceutical, and smart packaging industries due to their diverse red-to-blue coloration and strong antioxidant and anti-inflammatory properties.[1] The importance of these compounds lies not only in promoting human health but also in enhancing product shelf life and increasing the added value of agricultural products. Therefore, optimizing anthocyanin extraction from plant sources, particularly fruits, has become a significant focus of recent research. In this study, the ultrasonic-assisted extraction (UAE) of anthocyanins from Prunus domestica flesh was optimized. UAE, as a green and efficient technique, utilizes acoustic cavitation to enhance mass transfer and disrupt cell walls. This significantly improves extraction yield while reducing processing time and solvent use.[2] A central composite design (CCD) in Design-Expert software was employed to evaluate the effects of extraction time, temperature, solvent-to-solid ratio, and solvent concentration on total anthocyanin content (TAC), total phenolic content (TPC), and antioxidant activity (DPPH).The final TAC model, based on 30 experiments including six replicates, demonstrated excellent fit; the high coefficient of determination ($R^2 = 0.9912$) closely matched the predicted ($R^2 = 0.9803$) and adjusted R^2 (0.9863), confirming model accuracy and predictive power. High adequacy precision (63.09) and a low coefficient of variation (CV = 0.919%) further indicated strong reproducibility and experimental stability. Under optimized conditions, the maximum TPC, TAC, and DPPH antioxidant activity were 8.24 mg gallic acid equivalents per g FW, 1.17 mg cyanidin-3-glucoside equivalents per g fresh weight (FW), and 71.42%, respectively. These results demonstrate the high efficiency of UAE in preserving and recovering bioactive compounds from Prunus domestica, highlighting its potential applications in the food, pharmaceutical, and natural colorant industries.

Keywords: Prunus domestica, Anthocyanin, Antioxidant, Experimental Design, Extraction Optimization

References:

1. Patil, J.S., et al., *Green Extraction of Anthocyanins from Prunus domestica Using Natural Deep Eutectic Solvents: Anti-sickling Properties and Phenolic Impact Assessment.* Food Analytical Methods, 2024. **17**(10): p. 1432-1445.

2. Haddadi-Guemghar, H., et al., *Optimisation of microwave-assisted extraction of prune (Prunus domestica) antioxidants by response surface methodology.* International Journal of Food Science and Technology, 2014. **49**(10): p. 2158-2166.

Design of Experiment for Optimized Ultrasound-Assisted Extraction of Natural Pigments

Zeynab Rasuli[1], *Khalil Farhadi*[1*], *Morteza Bahram*[1], *Sima Najafi*[1]
[1]*Department*[1] *of Chemistry Faculty of Science Urmia University Urmia Iran*
khalil.farhadi@yahoo.com

ABSTRACT

Natural pigments such as β-carotene and chlorophyll are widely utilized in nutrition, pharmaceuticals, and cosmetics owing to their antioxidant activity and diverse bioactive functions. In this study, ultrasonic-assisted extraction (UAE) with ethanol was employed to recover these pigments from plant matrices, with process optimization performed using a design of experiments (DOE) approach [1]. Three independent variables extraction temperature, time, and solvent-to-sample ratio were systematically evaluated. All factors significantly influenced pigment yield, with temperature and solvent-to-sample ratio exerting the strongest effects. Prolonged extraction times at elevated temperatures were found to negatively impact yields. DOE optimization identified the optimal extraction conditions as 32 °C, 91 min, and a solvent-to-sample ratio of 1:9, under which maximal recoveries of β-carotene and chlorophyll were achieved [2]. The predictive model exhibited strong performance ($R^2 = 0.95$; RMSE = 0.12 mg g^{-1}), demonstrating excellent predictive reliability. Compared with conventional solvent extraction, UAE under optimized conditions afforded higher pigment yields while reducing solvent consumption and processing time. Experimental data were consistent with model predictions, underscoring the robustness and generalizability of the approach. Pigment purification was performed by flash chromatography on silica gel using a tailored solvent system, followed by HPLC, NMR, and FTIR analyses to confirm purity. These analytical techniques verified pigment integrity, validating both extraction and measurement accuracy. Overall, this DOE-driven optimization framework highlights the potential of UAE with a food-grade ethanol solvent system for scalable industrial applications. Under optimized conditions, yields increased by 35% for β-carotene and 42% for chlorophyll compared with conventional extraction.

Keywords: Natural pigments, β-carotene, Chlorophyll, Ultrasonic-assisted extraction, Design expert

References:

1. Linares, G. and M.L. Rojas, *Ultrasound-assisted extraction of natural pigments from food processing by-products: a review.* Frontiers in Nutrition, **2022. 9:** p. 891462.
2. Li, Y., et al., *Optimization of extraction conditions of carotenoids from Dunaliella parva by response surface methodology.* Molecules, **2022. 27**(4): p. 1436.

Optimized Extraction of Pectin from Apple Pomace Using Response Surface Methodology for Sustainable Industrial Applications

Elham Teymorirad, Mohammad Kazemzadeh, Morteza Bahram, Khalil Farhad*

Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

E-mail address: khalil.farhadi@yahoo.com

ABSTRACT

Pectin, a naturally occurring polysaccharide within plant cell walls, is highly valued for its gelling properties and versatile applications in the food, pharmaceutical, and environmental sectors [1]. This study investigates the extraction and optimization of pectin from apple pomace, an agricultural by-product, to promote sustainable resource utilization and mitigate environmental pollution [2]. Given the substantial reliance on imported pectin in developing countries, coupled with limited resources and high production costs, utilizing local agricultural waste such as apple pomace presents a viable strategy for achieving industrial self-sufficiency and cost reduction [3]. As a biocompatible and safe alternative to synthetic additives, pectin enhances public health by minimizing the need for chemical preservatives in food and pharmaceutical applications [4].

The extraction process was optimized using an acid-based method with hydrochloric acid, employing response surface methodology (RSM) and a Box-Behnken design facilitated by Design-Expert software. Optimal extraction conditions were determined as 0.1 M hydrochloric acid, a temperature range of 85–90°C, an extraction duration of 75–80 minutes, and 70% ethanol, yielding an extraction efficiency of 20% (equivalent to 5 g of pectin from 25 g of dried apple pomace). The utilization of apple pomace not only reduces biological waste and environmental impact but also aligns with sustainable development goals by optimizing natural resource use. Furthermore, localized pectin production offers economic benefits through reduced import dependency, foreign exchange savings, and job creation in agricultural processing industries [5].

Keywords: Pectin, Apple pomace, RSM, Extraction, Agricultural waste

References

1. Ciriminna, R., et al., *Pectin: New science and forthcoming applications of the most valued hydrocolloid.* Food Hydrocolloids, 2022. 127: p. 107483.

2. Cherian, E., et al., *Investigation into pectin extraction and technological implementations in the food industry.* Journal of the Science of Food and Agriculture, 2024. 104(15): p. 9102–9110.

3. Mounya, K.S. and A.R. Chowdary, *Optimization of ultrasound-assisted pectin recovery from cocoa by-products using response surface methodology.* Journal of the Science of Food and Agriculture, 2024. 104(11): p. 6714–6723.

4. Dirpan, A., et al., *Exploring the potential of pectin as a source of biopolymers for active and intelligent packaging: a review.* Polymers, 2024. 16(19): p. 2783.

5. Akram, W., et al., *Designing & optimisation of dual Ca2+ and SO42–ionic cross-linked sericin/pectin microbeads using response surface methodology for colon-specific delivery.* Journal of Biomaterials Science, Polymer Edition, 2025: p. 1–26.

Machine Learning Classification of Solid-Phase Microextraction (SPME) Types Based on Operational and Analytical Parameters from Experimental Data

Erfan Hashemi, Zeinab Sodaei, Majid Hashemianzadeh, Afsaneh Mollahosseini**

1 Molecular Simulation Research Laboratory, Department of Chemistry, Iran University of Science and Technology, Tehran, Iran

2 Research Laboratory of Spectroscopy and Micro and Nano Extraction, Department of Chemistry, Iran University of Science and Technology (IUST), Narmak, Tehran, Iran

E-mail address:
hashemianzadeh@iust.ac.ir
amollahosseini@iust.ac.ir

ABSTRACT

Solid-phase microextraction (SPME) is a sample preparation technique that has garnered significant attention due to its straightforward operation, high separation efficiency, environmental friendliness, and compatibility with various analytical instruments. In the SPME process, the extraction, enrichment, and sample introduction steps are integrated into a single step [1]. Operational parameters significantly affect extraction efficiency and analytical performance in this method. Traditionally, optimizing these parameters relies on extensive experimental trials, which can be time-consuming and resource-intensive [2]. In this study, we present a machine learning-based framework to classify SPME types using key operational and performance features, including coating thickness, equilibrium time, fiber length, and temperature related to analytes Naphthalene and Anthracene. A curated dataset comprising 203 records was compiled from literature sources, encompassing a diverse range of SPME conditions and experimental setups. Several supervised classification algorithms, such as XGBoost, Random Forest, CatBoost, AdaBoost, were trained. Using Random Forest with selected features, the classification accuracy reached 78% and 5-fold cross-validation evaluated. Feature importance analysis revealed that parameters

temperature, equilibrium time, coating thickness, and fiber length had the most significant influence sequentially on classification outcomes, providing insight into the most effective factors for SPME performance. These results demonstrate the potential of machine learning to complement experimental workflows in analytical chemistry techniques performance based on measurable parameters. By identifying the most influential factors, this approach can streamline method development, reduce experimental workload, and enhance reproducibility.

Keywords: "machine learning" "SPME" "solid phase microextraction" "analytical chemistry" "PAHs" "polycyclic aromatic hydrocarbons"

References

[1] X. Feng, X. et al., "Solid phase microextraction for the bioanalysis of emerging organic pollutants", TrAC Trends Anal. Chem., 177, (**2024**), 117786.

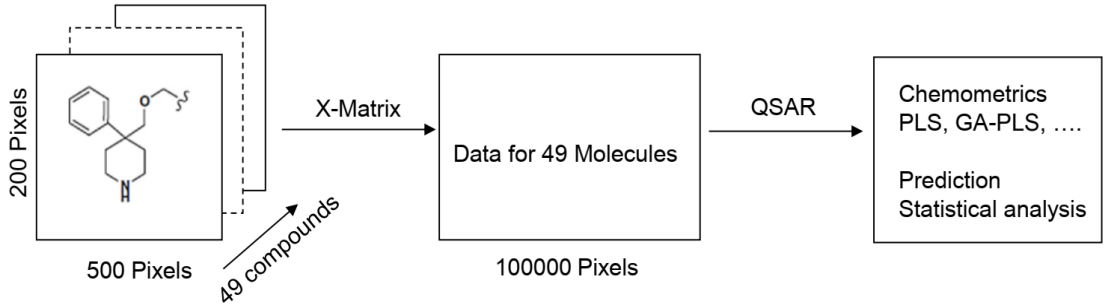
[2] Z. Lan, Z. Zirong, …and the last author, "Length-controlled hydrophobic CF_3-COF as a highly efficient absorbent coating for dual-mode solid-phase microextraction of sixteen polycyclic aromatic hydrocarbons in water samples", Sci. Total Environ., 925, (**2024**), 171726.

Multivariate image – QSAR of phenyl piperidine derivatives as potent dual NK1R antagonists/serotonin transporter (SERT) inhibitor

Neda Ladooni, Atisa Yazdanipour, Ali Niazi*
Department of Chemistry, CT.C., Islamic Azad University, Tehran, Iran
E-mail address: atisayazdanipour@iau.ac.ir

ABSTRACT

Depression is a critical mood disorder that affects millions of patients. Serotonin transporter (SERT) is the target for several antidepressant drugs including the selective serotonin reuptake inhibitors (SSRIs). [1]. In this study, a series of phenyl piperidine derivatives as potent dual NK1R antagonists/serotonin transporter (SERT) inhibitors were applied to quantitative structure–activity relationship (QSAR) analysis by multivariate image analysis.



In this study we investigate the effect of pixel selection by application of genetic algorithms for PLS model. GAs is very useful in the variable selection in modeling and calibration because of the strong effect of the relationship between presence/absence of variables in a calibration model and the prediction ability of the model itself. The subset of pixels, which resulted in the low prediction error, was selected by genetic algorithm. The resulted model showed high prediction ability with RMSEP of 0.1625, 0.0812 and 0.0106 for PCR, PLS and GA-PLS models, respectively. Then, MIA-QSAR study is proposed to model and predict the IC_{50} as Serotonin transporter activity of a series of phenyl piperidine derivatives by GA-PLS modeling method.

Keywords: Multivariate image analysis, QSAR, PLS, Genetic algorithms, Serotonin.

References

[1] S. Zare, M. Fereidoonnezhad, D. Afshar and Z. Ramezani, A comparative QSAR analysis and molecular docking studies of phenyl piperidine derivatives as potent dual NK1R antagonists/serotonintransporter (SERT) inhibitors, Computational Biology and Chemistry, 67 (2017) 22.37.
[2] M. Sarkhosh, N. Khorshidi, A. Niazi and R. Leardi, Application of genetic algorithms for pixel selection in multivariate image analysis for a QSAR study of trypanocidal activity for quinone compounds and design new quinone compounds, Chemometrics and Intelligent Laboratory Systems, 139 (2014) 168-174.
[3] M. Mirshafiei, A. Niazi and A. Yazdanipour, Application of MIA-QSAR in designing new protein P38 MAP kinase compounds using a genetic algorithm, Periodica Polytechnica Chemical Engineering 67 (2023) 141-160.

PH-Dependent Sensor Array Based on Carbon Quantum Dots for Detection and Discrimination of Amines

Tahereh Rasoulizadeh, Masoumeh Hasani*

Department of Analytical Chemistry, Faculty of Chemistry and Petroleum Sciences, Bu-Ali Sina University, Hamedan, Iran

*** E-mail: hasani@basu.ac.ir**

ABSTRACT

The development of analytical methods for amine detection is of great importance, as these compounds are commonly found as contaminants in pharmaceutical, agricultural, and food processing wastewaters [1, 2]. Here, we present a pH-dependent fluorescence array sensor based on carbon dots synthesized from sucrose (SCD) for the detection and discrimination of various amines, including propylamine, ethylamine, butylamine, hexylamine, dicyclohexylamine, diethylenetriamine, tripropylamine, tributylamine, hexamethyleneimine, aniline, and diethylaniline, at different concentrations. The SCD-based sensor displayed distinct fluorescence response patterns at pH 5 and 9, resulting from varying interaction between SCD and the target amines. These pH-dependent responses generate unique fluorescence "fingerprints" that enable amine discrimination using multivariate pattern recognition techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), partial least squares discriminant analysis (PLS-DA), and hierarchical cluster analysis (HCA). By combining responses from both pH conditions, the sensor array classified all amines with 97–100% accuracy in PLS-DA models. Quantitative analysis was achieved by correlating the dominant latent variable (factor 1) from PLS-DA with amine concentration, establishing a linear detection range of 50–300 µM. Detection limits were determined to be 0.041 µM for aniline, 0.091 µM for tripropylamine, 0.024 µM for dicyclohexylamine, and 0.098 µM for hexamethyleneimine. Furthermore, the method demonstrated reliable performance in complex environments, including binary mixtures and real-world samples such as cheese, confirming its practical applicability. This study evaluates the potential of sucrose-derived carbon dot-based fluorescent arrays as effective probe for the detection of a wide range of amine compounds, demonstrating their use in complex and challenging analytical environments.

Keywords: Carbon Quantum Dots, PH-Dependent Sensor array, Fluorescence, PLS-DA, Discrimination of amines

References

[1] C.M. Almeida, J. M. Magalhaes, M. F. Barroso and L. Duraes, Latest advances in sensors for optical detection of relevant amines: insights into lanthanide-based sensors, ***J. Mater. Chem. C***, **10**, (2022), 15263-15276**.**

[2] Y. X. Sun, Z. J. Yan, W. X. Liu, X. M. Chen, M. H. Ding, L. L. Tang and F. Zeng, Rapid and Visual Detection of Volatile Amines Based on Their Gas–Solid Reaction with Tetrachloro-p-Benzoquinone, Molecules, 29, (2024),.

Solvent-tuned Sensor Arrays Based on colored Carbon Quantum Dots for Discrimination of Multiple Metal Ions

Tahereh Rasoulizadeh, Masoumeh Hasani*

Department of Analytical Chemistry, Faculty of Chemistry and Petroleum Sciences, Bu-Ali Sina University, Hamedan, Iran

*** E-mail: hasani@basu.ac.ir**

ABSTRACT

Rapid and accurate detection of multiple heavy metal ions is of great importance for environmental monitoring, food safety and public health protection [1,2]. In this study, we developed two sensor arrays, including a smartphone-assisted colorimetric array and a fluorescence-based array, utilizing solvent-tunable, green-emitting carbon quantum dots (o-gCQDs) for accurate detection of ten heavy metal ions including Mn^{2+}, Cu^{2+}, Cr^{3+}, Co^{2+}, Ni^{2+}, Zn^{2+}, Hg^{2+}, Pb^{2+}, Cd^{2+}, and UO_2^{2+} at various concentrations. The optical responses of o-gCQDs were modulated by varying the solvent composition from pure DMF to DMF/water mixtures, enabling distinct colorimetric and fluorescence signatures upon interaction with different metal ions. In the colorimetric sensor, RGB responses were captured in three solvent environments: pure DMF, a 40:60 water/DMF mixture, and an 80:20 water/DMF mixture. The fluorescence sensor employed emission spectra from o-gCQDs in pure DMF, 80:20 water/DMF, and 90:10 water/DMF mixtures. Each solvent environment functioned as an individual sensing element, enabling the generation of diverse and discriminative optical response patterns across the sensor arrays. Multivariate statistical methods, including principal component analysis (PCA), linear discriminant analysis (LDA), and partial least squares discriminant analysis (PLS-DA), were applied to the collected data. The fluorescence sensor array achieved classification accuracies of 98–100% using PLS-DA, demonstrating high sensitivity and specificity. The colorimetric array also effectively distinguished between metal ions based on their RGB response profiles. This work highlights a robust and versatile sensing strategy based on solvent-regulated nanomaterials, offering an accurate approach for multi-analyte heavy metal ions detection with potential applications in environmental and food safety monitoring.

Keywords: Solvent-tunable Sensor, Fluorescence, Smartphone, Pattern recognition, Discrimination

References

[1] M. F. Rauf, Z. Lin, M. K. Rauf and J. M. Lin, Innovative Microfluidic Technologies for Rapid Heavy Metal Ion Detection, Chemosensors, 13, (2025), 149.

[2] R. Massoud, M.R Hadiani, P. Hamzehlou, and K. Khosravi-Darani, Bioremediation of heavy metals in food industry: Application of Saccharomyces cerevisiae, Electron. J. Biotechnol, 37, (2019), 56–60.

Experimental design for optimization of the removal of Rhodamine B using a novel GQD incorporated Hydrogel

Reza Dadashi[*a], *Morteza Bahram*[a], *Nima Mostafazadeh*[b]

[a]Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

[b]Electrochemistry Research Laboratory, Department of Physical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

[*]Corresponding author (*E-mail address):* r.dadashi@urmia.ac.ir

ABSTRACT

The removal of synthetic dyes such as Rhodamine B (RhB) from wastewater is an enduring environmental challenge due to their persistence, toxicity, and visibility even at low concentrations [1]**. Traditional removal techniques often lack selectivity, efficiency, or cost-effectiveness** [2]**. In this context, we synthesized graphene quantum dot (GQD)-incorporated tragacanth-poly(acrylic acid-co-acrylamide) hydrogel to apply for dye pollution removal. The results showed that the presence of GQDs in the hydrogel framework increases the efficiency of dye pollutant removal. This composite hydrogel combines the high surface area, tunable functional groups, and photochemical stability of GQDs with the porous, water-swellable network of hydrogels to enhance the adsorption of RhB. To optimize the removal performance, we employed a Design of Experiments (DoE) approach, specifically leveraging a Response Surface Methodology (RSM). Independent variables such as pH, contact time, and hydrogel dose were systematically varied to understand their individual and interactive effects on RhB removal efficiency. This statistical framework reduces the number of required experiments, accelerates optimization, and reveals key parameter interactions** [3]**. The properties of the GQD-hydrogel composite were evaluated by different analyses. Preliminary findings indicate high adsorption capacity, rapid kinetics, and strong affinity for RhB, benefiting from the synergistic contributions of GQD's π-π interactions and hydrogel swelling behavior. The optimized conditions for maximum dye removal were obtained by Central Composite Design (CCD).**

Recent studies underscore the potential of GQDs for environmental remediation due to their tunable surface properties and adsorption capabilities [4]**. Moreover, graphene-based**

hydrogel nanocomposites have demonstrated exceptional dye removal efficiency and scalability in water treatment applications. In summary, this study presents a cost-effective, efficient, and scalable strategy for RhB removal using a GQD-embedded hydrogel. The integration of DoE-based optimization ensures practical applicability and environmental relevance. Our findings suggest that such nanocomposite hydrogels hold significant promise for broader pollutant remediation tasks beyond dye removal.

Keywords: GQD, Hydrogel, Rhodamine B, Dye pollution, Experimental Design

References

[1] A.A. Al-Gheethi, Q.M. Azhar, P.S. Kumar, A.A. Yusuf, A.K. Al-Buriahi, R.M.S.R. Mohamed, M.M. Al-Shaibani, Sustainable approaches for removing Rhodamine B dye using agricultural waste adsorbents: A review, Chemosphere. 287 (2022) 132080.

[2] G. Bal, A. Thakur, Distinct approaches of removal of dyes from wastewater: A review, Mater. Today Proc. 50 (2022) 1575–1579.

[3] X. Shi, W. Ruan, J. Hu, M. Fan, R. Cao, X. Wei, Optimizing the removal of rhodamine B in aqueous solutions by reduced graphene oxide-supported nanoscale zerovalent iron (nZVI/rGO) using an artificial neural network-genetic algorithm (ANN-GA), Nanomaterials. 7 (2017) 134.

[4] N. Kumar, R. Gusain, S. Pandey, S.S. Ray, Hydrogel nanocomposite adsorbents and photocatalysts for sustainable water purification, Adv. Mater. Interfaces. 10 (2023) 2201375.

Rank Annihilation Factor Analysis for Kinetic Spectrophotometric Determination of MCPA herbicide in Unknown Samples

*Nima Karimnejad, Reza Dadashi, Morteza Bahram**
Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

*Corresponding author (*E-mail address):* m.bahram@urmia.ac.ir

ABSTRACT

The contamination of soil and water with herbicides is a major concern for both the environment and human health. MCPA (4-(2-methyl-4-chlorophenoxy)acetic acid) is one of the most widely used phenoxy herbicides for controlling broadleaf weeds in cereal crops and pastures. Its extensive application increases the risk of leaching into natural resources, potentially harming non-target species and posing risks to human health [1]**. Therefore, reliable and sensitive methods are needed for monitoring MCPA in unknown and complex samples. Classical analytical methods such as HPLC, GC, and mass spectrometry are accurate but often costly, time-consuming, and require complicated sample preparation. Kinetic spectrophotometry is a simpler and more affordable alternative, but it can be limited by overlapping spectra and matrix interferences, which reduce measurement accuracy** [2]**. To address this challenge, we used Rank Annihilation Factor Analysis (RAFA), a powerful chemometric tool for resolving multivariate data. RAFA eliminates interfering signals by rank reduction, enabling accurate identification and quantification of the analyte even in complex datasets. When combined with kinetic spectrophotometry, RAFA provides improved sensitivity, selectivity, and robustness compared with traditional approaches** [3]**. In this study, kinetic spectral data were obtained using the Marquis reagent, a mixture of formaldehyde and concentrated sulfuric acid, which produces characteristic color reactions with different compounds. This reagent facilitated the generation of distinct kinetic data suitable for RAFA processing. Herbicide samples containing MCPA were analyzed, and the method showed strong performance in resolving overlapping signals and minimizing interference. Overall, the combination of RAFA with kinetic spectrophotometry represents an efficient, cost-effective, and reliable strategy for the determination of MCPA in**

environmental and biological matrices. This approach offers a promising alternative to conventional chromatographic methods and has significant potential for practical applications in environmental monitoring and agricultural safety.

Keywords: Herbicide, Marquis reagent, MCPA, RAFA

References

[1] P.A. Morton, C. Fennell, R. Cassidy, D. Doody, O. Fenton, P. Mellander, P. Jordan, A review of the pesticide MCPA in the land-water environment and emerging research needs, Wiley Interdiscip. Rev. Water. 7 (2020) e1402.

[2] O. Selcuk, Y. Demir, C. Erkmen, S. Yıldırım, B. Uslu, Analytical methods for determination of antiviral drugs in different matrices: recent advances and trends, Crit. Rev. Anal. Chem. 52 (2022) 1662–1693.

[3] A. Afkhami, L. Khalafi, Spectrophotometric investigation of the effect of β-cyclodextrin on the intramolecular cyclization reaction of catecholamines using rank annihilation factor analysis, Anal. Chim. Acta. 599 (2007) 241–248.

Application of artificial neural network for the simultaneous determination of two rizatriptan and sumatriptan drugs via gold nanoparticle aggregation-based colorimetric sensing

Masoud Esmailpak, Morteza Bahram*, Reza Dadashi
Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

*Corresponding author (*E-mail address):* m.bahram@urmia.ac.ir

ABSTRACT

Migraine is a common and debilitating neurological disorder, often treated with triptan-class drugs such as rizatriptan and sumatriptan [1]. Simultaneous monitoring of these drugs in biological samples is crucial for evaluating therapeutic efficacy and preventing drug interactions. This study introduces a novel approach for designing a colorimetric nanosensor based on gold nanoparticles (AuNPs), capable of simultaneous detection of these drugs without the need for surface modification. AuNPs are widely recognized in the development of chemical and biosensors due to their unique optical properties, high stability, and biocompatibility [2]. In this work, the sensing mechanism relies on surface plasmon resonance shifts and nanoparticle aggregation. Interaction of the drugs with citrate-capped AuNPs induces a visible color change from red to blue, which can be detected either by the naked eye or UV–Vis spectroscopy.

To improve accuracy and signal resolution in the presence of both drugs, chemometric methods including principal component analysis (PCA) and artificial neural networks (ANN) were applied. This approach reduces spectral overlap and allows modeling of complex data from drug mixtures [3]. For this purpose, in the MATLAB program and the artificial intelligence defined nntool of the Feed-Forward backprop type, using the principal component obtained from PCA, the neural network was started to be trained and the obtained data was analyzed.

The integration of nanotechnology with artificial intelligence provides a novel framework for pharmaceutical analysis, offering high sensitivity and selectivity, as well as applicability in complex biological matrices such as serum and urine.

The significance of this study lies in the development of a simple, cost-effective, and rapid tool for monitoring migraine therapy. This approach minimizes the need for expensive instrumentation and has potential for the development of portable diagnostic kits. Thus, the combination of gold nanoparticles and chemometric methods not only opens new avenues in pharmaceutical analysis but also provides a valuable platform for broader applications in biosensing and personalized medicine.

Keywords: Gold Nanoparticle, Chemometric Methods, Sensor, Migraine Drugs

References

[1] O. Singh, S. Sharma, M. Naagar, M.K. Maity, Eletriptan as treatment option for acute migraine, Int. J. Innov. Res. Anal. 2 (2022) 15–24.

[2] L. Nie, F. Liu, P. Ma, X. Xiao, Applications of gold nanoparticles in optical biosensors, J. Biomed. Nanotechnol. 10 (2014) 2700–2721.

[3] G. Ioele, M. De Luca, E. Dinç, F. Oliverio, G. Ragno, Artificial neural network combined with principal component analysis for resolution of complex pharmaceutical formulations, Chem. Pharm. Bull. 59 (2011) 35–40.

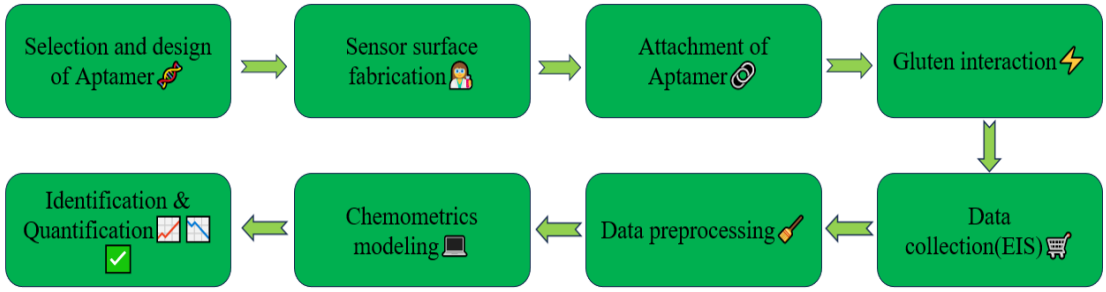
Chemometric modeling of an electrochemical aptasensor for gluten detection using simulated electrochemical sensor data and comparison with experimental results

*Masoumeh Yousef-Zadeh, Reza Emamali Sabzi**
Department of Science, Chemistry Faculty Urmia University, Urmia, Iran
E-mail: rezasabzi@yahoo.com

ABSTRACT

The presence of gluten in food products poses a serious health risk to individuals with celiac disease, necessitating rapid and reliable detection methods [1]. In this study, we designed an electrochemical aptasensor capable of selectively detecting gluten molecules, and applied chemometric models to predict gluten concentration based on the sensor's electrochemical response. A dataset was generated using simulated sensor current values (µA) corresponding to gluten concentrations ranging from 1 fM to 1 µM . Additional sensor features were created to represent variations in sensor response under different measurement conditions. Chemometric techniques of Singular Value Decomposition (SVD), principal component analysis (PCA), were used to examine patterns in the data and build predictive models. The preliminary results analyzed show that SVD and PCA provide high prediction accuracy for estimating gluten concentration within the desired range from sensor data, may promise its efficiency in the real world to provide the technical support for a valid assay in food safety fields [2]. The proposed approach demonstrates the potential of combining electrochemical sensing with multivariate data analysis for the development of portable, low-cost, sensitive and selective gluten detection systems. A summary of the use of chemometrics in this work is shown in Scheme 1.

Keywords: Gluten, Electrochemical Aptasensor, Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Eigen decomposition



Scheme. 1*:* The carthonic fabrication process of the electrochemical aptasensor for gliadin detection with Dimensionality Reduction Techniques.

References:

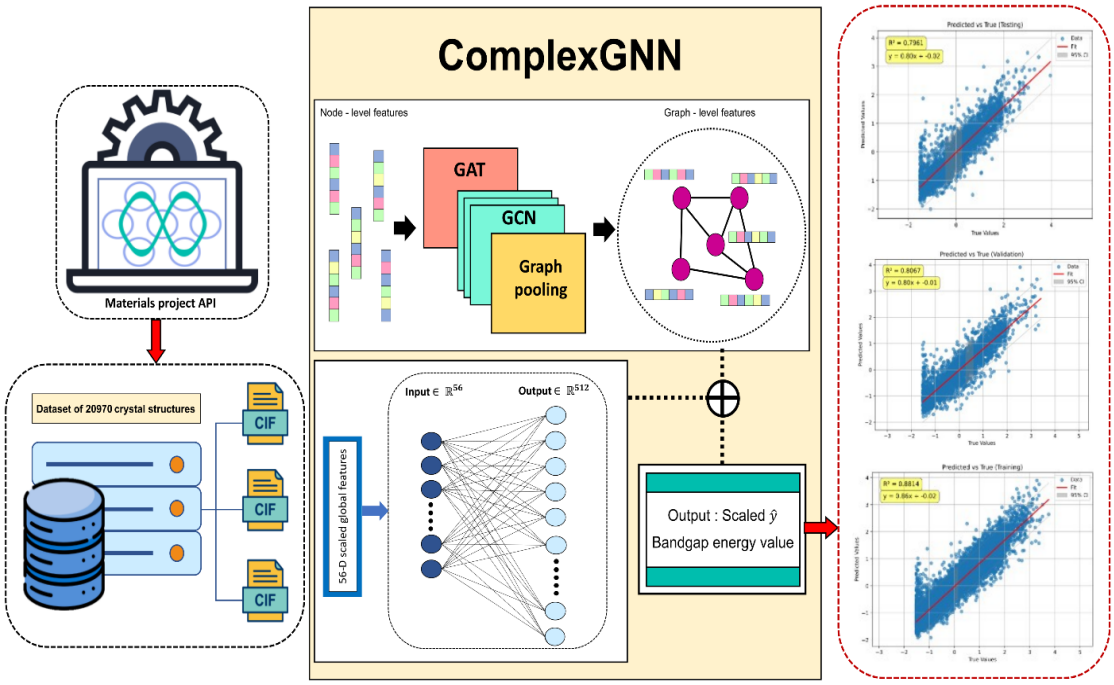
[1] R. Svigelj, N. Dossi, C. Grazioli, et al, Paper-based aptamer-antibody biosensor for gluten detection in a deep eutectic solvent (DES). Anal Bioanal Chem 414, 3341–3348 (2022).
[2] M. Gyimadu, G. Bell, A Comprative Analysis of Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) as Dimensionality Reduction Techniqes. arXiv preprint arXiv:2506.16663(2025).

ComplexGNN: Hybrid Graph and Descriptor Learning for Interpretable Bandgap Prediction in Inorganic Crystals

Dorsa Benvan [a], *Jahan B Ghasemi* [a,*], *Xuanhua Li* [b], *Jiaguo Yu* [c]

[a] *Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran.*
[b] *State Key Laboratory of Solidification Processing, Center for Nano Energy Materials, School of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an, 710072, China.*
[c] *Laboratory of Solar Fuel, Faculty of Materials Science and Chemistry, China University of Geosciences, 68 Jincheng Street, Wuhan 430078, P. R. China.*

*E-mail address (corresponding author): jahan.ghasemi@ut.ac.ir



ABSTRACT

Predicting the electronic bandgap of crystalline materials is considered a critical challenge in materials science and renewable energy, since bandgap values are known to govern solar absorption and photocatalytic efficiency. Accurate estimation of this property is required because the bandgap determines the capacity of semiconductors to capture visible light, drive electron-hole separation, and facilitate redox reactions applied in solar cells, photocatalytic water splitting, and environmental remediation [1-10]. Although accurate predictions are offered by first-principles

methods such as density functional theory (DFT), scalability is limited by high computational cost. DFT calculations have been employed extensively, yet they remain time-consuming and resource-intensive, and their accuracy varies according to the approximations used, which reduces reliability for high-throughput screening [11,12]. In this study, a dual-stream graph neural network (GNN) architecture, named ComplexGNN, is introduced, in which atomic-level graph features are combined with global crystallographic descriptors to predict bandgaps of inorganic crystals. Crystal lattices are represented as graphs, with atoms treated as nodes and interatomic relationships as edges, enabling local features to be learned through graph attention and convolutional layers. At the same time, 56 global descriptors capturing geometric, chemical, and symmetry-related properties are incorporated to provide complementary long-range information, ensuring that both bonding patterns and overall structural attributes are reflected. A dataset of 20970 crystal structures was compiled from the Materials Project database, and both local environments and global descriptors were used for training. Robust performance (average test-set $R^2 \approx 0.80$; MAE ≈ 0.32) was achieved, and strong generalization across randomized splits was confirmed by 50-fold Monte Carlo cross-validation. Applicability domain analysis using Williams plots indicated that over 96% of samples fell within confidence regions, with only a small fraction of unusual materials identified as outliers. Furthermore, unsupervised clustering of the hybrid feature space and principal component analysis (PCA) of handcrafted descriptors revealed clear structure-property relationships, supporting the interpretability of the learned latent space. These analyses demonstrated that not only accurate predictions were made, but also that materials were organized in chemically meaningful ways, with systematic trends in bandgap variation identified. It was therefore highlighted that integration of graph-based learning with physics-aware descriptors enables accurate, interpretable, and scalable materials property prediction, contributing to accelerated semiconductor and photocatalyst discovery.

Keywords: Machine learning, Materials science, Energy materials, Predictive modeling, Structure- property relationships.

References

[1] A. Fujishima, K. Honda, “Electrochemical photolysis of water at a semiconductor electrode”, Nature, 238, (1972), 37–38.

[2] F. Mohamadpour, A. M. Amani, “Photocatalytic systems: reactions, mechanism, and applications”, RSC Advances, 14, (2024), 20609–20645.

[3] Y. He, J. Huang, B. Wang, Y. Qu, “Construction of Z-scheme heterojunction C_3N_4/N-CQDs@$W_{18}O_{49}$ for full-spectrum photocatalytic organic pollutant degradation”, Applied Surface Science, 610, (2023), 155255.

[4] R. Gupta, U. Alam, N. Verma, “Efficient Photofixation of Nitrogen to Ammonia over an S-Scheme-Based $NiSnO_3$-g-C_3N_4 Heterojunction”, Chemical Engineering Journal, 479, (2024), 147644.

[5] A. Garcia-Baldoví, A. M. Asiri, H. Garcia, “Photocatalytic CO_2 reduction to methanol: How can the dilemma be

solved?", Current Opinion in Green and Sustainable Chemistry, 41, (2023), 100831.

[6] I. Ahmad, Y. Zou, J. Yan, Y. Liu, S. Shukrullah, M. Y. Naz, H. Hussain, W. Q. Khan, N. R. Khalid, "Semiconductor photocatalysts: A critical review highlighting the various strategies to boost the photocatalytic performances for diverse applications", Advanced Colloid and Interface Science, 303, (2022), 102830.

[7] L. Zhang, J. Zhang, H. Yu, J. Yu, "Emerging S-scheme photocatalyst", Advanced Materials, 34, (2022), 2107668.

[8] N. Serpone, E. Borgarello, M. Grätzel, "Visible light induced generation of hydrogen from H2S in mixed semiconductor dispersions; improved efficiency through inter-particle electron transfer", Journal of the Chemical Society, Chemical Communications, (1984), 342–344.

[9] D. Salazar-Marín, G. Oza, J. A. D. Real, A. Cervantes-Uribe, H. Pérez-Vidal, M. K. Kesarla, J. G. T. Torres, S. Godavarthi, "Distinguishing between type II and S-scheme heterojunction materials: A comprehensive review", Applied Surface Science Advances, 19, (2024), 100536.

[10] S. Rana, A. Kumar, T. T. Wang, G. Sharma, P. Dhiman, A. García-Penas, "A review of carbon material-based Z-scheme and S-scheme heterojunctions for photocatalytic clean energy generation", New Carbon Materials, 39 (3), (2024), 458–482.

[11] C. H. Lin, J. Rohilla, H. H. Kuo, C. Y. Chen, T. F. Mark Chang, M. Sone, P. P. Ingole, Y. C. Lo, Y. J. Hsu, "Density-Functional Theory Studies on Photocatalysis and Photoelectrocatalysis: Challenges and Opportunities", Solar RRL, 8 (10), (2024), 2300948.

[12] R. Wang, B. Wang, A. S. Abdullahi, H. Fan, "Understanding the prototype catalyst TiO2 surface with the help of density functional theory calculation", WIREs Computational Molecular Science, 14 (1), (2024), e1686.

Machine Learning–Driven Discovery of a Potent Lead Compound for GPR17 via Graph Neural Networks and Multistage Structural Validation

*Noushin Agh Babaie, Morteza Farni, Jahan B. Ghasemi**
Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran.
Jahan.ghasemi@ut.ac.ir

ABSTRACT

G-protein coupled receptor 17 (GPR17) has recently emerged as a critical therapeutic target in neuroinflammation and myelin repair, making its inhibition a promising strategy for multiple sclerosis (MS) and related neurodegenerative diseases [1, 2]. In this work, we present a machine learning–driven framework that integrates graph neural networks (GNNs) with multistage structural validation to accelerate the identification of potent GPR17 inhibitors.

A curated dataset of 317 ligands with experimentally reported IC_{50} values was compiled from BindingDB resource, standardized into pIC_{50} values, and rigorously validated using RDKit. Each ligand was encoded as a molecular graph with atom- and bond-level features, while global descriptors such as molecular weight, TPSA, and logP were added. To effectively capture chemical information, we implemented a GNN architecture based on NNConv layers from PyTorch Geometric [3], which dynamically integrates edge-conditioned convolutions through a filter-generating multilayer perceptron. This design enabled simultaneous learning of node and bond features, significantly improving predictive accuracy.

Hyperparameter optimization using the Optuna framework yielded an optimal configuration of four message-passing layers, hidden dimension 128, batch size 32, and learning rate 1e–3. The model achieved robust predictive performance with R^2 values of 0.868 (training), 0.727 (validation), and 0.790 (test), supported by low RMSE and MAE values. Applicability domain analysis via William's plot confirmed strong reliability, with only four outliers identified among 317 compounds.

The trained model was deployed to external datasets: (i) top antihistamine candidates from molecular docking, and (ii) 592 pharmacophore-screened compounds from the ZINC library. Notably, the model highlighted Barmastine (predicted $pIC_{50} \approx 10$), Sulukast, and Astemizole (BBB-permeable) as strong repurposing candidates. In parallel, pharmacophore screening and docking identified Iralukast (CGP 45715A), a cysteinyl leukotriene receptor antagonist, as the top-ranked hit (predicted $pIC_{50} \approx 12.37$; GoldScore = 67.6). MD simulations and MM-PBSA analyses confirmed stable binding interactions across key residues Lys216, Arg214, and Tyr140, designating these as pharmacological "hot spots."

Collectively, the convergence of docking, pharmacophore modeling, MD validation, and GNN predictions underscores the translational potential of antihistamine- and leukotriene-derived scaffolds as GPR17 inhibitors. Our findings highlight a cost-effective AI-enabled drug repurposing strategy [4], providing high-confidence leads such as Barmastine, Astemizole, Sulukast, and Iralukast for experimental validation and preclinical development in neuroinflammatory disorders.

Keywords: GPR17, graph neural network (GNN), Drug Repurposing, Inhibitor Optimization.

References (Times New Roman 10pt Bold):

[1] Adamu A, Li S, Gao F, Xue G. The role of neuroinflammation in neurodegenerative diseases: Current understanding and future therapeutic targets. Frontiers in aging neuroscience. 2024; 16:1347987.

[2] Lecca D, Raffaele S, Abbracchio MP, Fumagalli M. Regulation and signaling of the GPR17 receptor in oligodendroglial cells. Glia. 2020; 68(10):1957–67.

[3] rGilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE, editors. Neural Message Passing for Quantum Chemistry. International Conference on Machine Learning; 2017.

[4] Pushpakom, Sudeep, et al. "Drug repurposing: progress, challenges and recommendations." Nature reviews Drug discovery 18.1 (2019): 41-58.

Using Machine Learning in the In-Silico Design of Selective Dopamine Receptor Ligands: Advancements in Targeted Therapies for Neurological and Psychiatric Disorders

*Melika F. Aghdam[a], Mahia V. Solout[a], Fatemeh Amini[a], Sogol Meknatkhah[b], Noushin Agha Babaie[a], Morteza Farnia[a], Jahan B. Ghasemi[a]**

[a] *Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran*

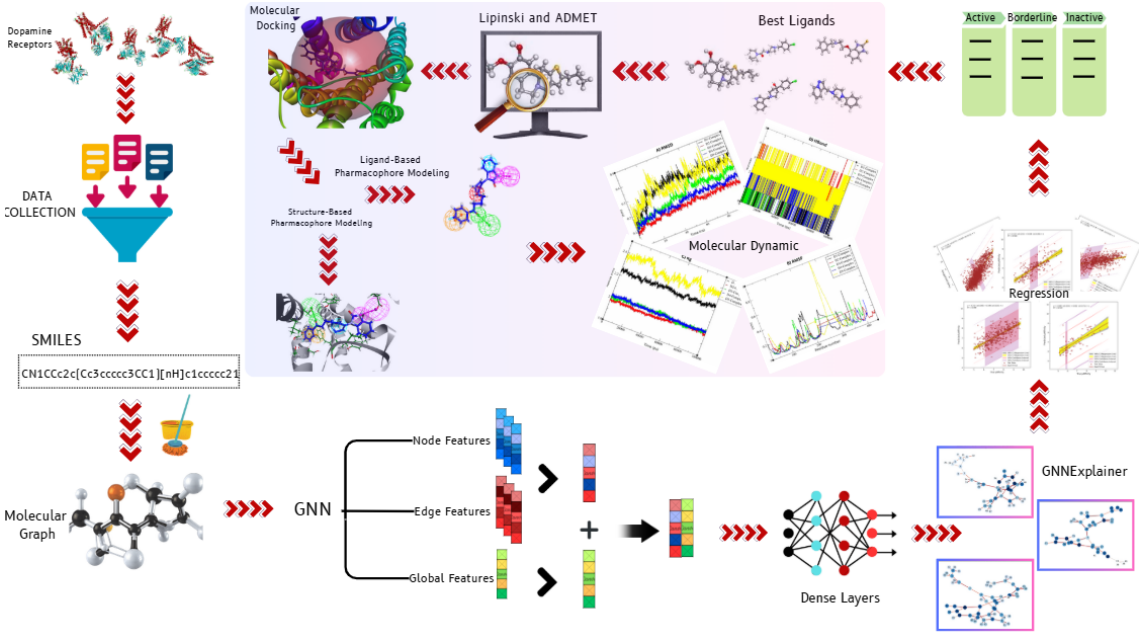
[b] *Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran*

E-mail address (corresponding author): jahan.ghasemi@ut.ac.ir

ABSTRACT

Accurately predicting ligand–receptor binding affinities is crucial for advancing structure-based drug design. Here, we present a novel graph neural network (GNN) model employing edge-conditioned message passing (NNConv), enabling the integration of bond-specific features into molecular graph representations[1]. Using a curated dataset of approximately 24,000 ligands targeting dopamine receptor subtypes D1–D5, our GNN models achieved strong predictive performance with low error metrics (RMSE ~0.6~0.8) and satisfactory determination coefficients (R^2), while residual analysis confirmed the reliability of predictions across activity ranges. Model interpretability was further assessed using the GNNExplainer algorithm[2], which highlighted chemically meaningful substructures driving predictions, such as aromatic rings, halogens, and electronegative atoms (O, N), consistent with known determinants of bioactivity. Complementary molecular docking and pharmacophore analyses revealed distinct binding profiles across bioactivity classes. Active ligands consistently engaged key conserved residues, notably ASP3.32, PHE6.51, and SER5.42/5.43, through directional hydrogen bonds and π–π interactions, resembling the binding patterns of Rotigotine. Inactive ligands exhibited superficial hydrophobic contacts with limited polar anchoring, while borderline ligands showed partial engagement with receptor cavities. ADMET and Lipinski's rule assessments confirmed favorable pharmacokinetic properties for the most potent ligands. Finally, molecular dynamics (MD) simulations of the top-ranked ligands for each dopamine receptor subtype confirmed the dynamic stability and persistent interactions of these complexes at their respective binding sites, further validating the structural predictions. These findings demonstrate that edge-aware GNN models, combined with interpretability and multi-level structural validation, offer a powerful and scalable strategy for decoding ligand–receptor interactions and accelerating neuropharmacological drug discovery.

Keywords: Dopamine Receptors, Machine Learning, Graph Neural Network (GNN), Molecular Docking, Pharmacophore Modeling, Molecular Dynamic, Drug-target Interaction



References:

1. Simonovsky, M. and N. Komodakis. *Dynamic edge-conditioned filters in convolutional neural networks on graphs*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

2. Ying, Z., et al., *Gnnexplainer: Generating explanations for graph neural networks.* Advances in neural information processing systems, 2019. **32**.

Spice Adulteration Detection Using Hyperspectral and Multispectral Imaging

Mahshid Ziaee jazi[a], Mahia Vazifeh Solout[a], José Manuel Amigo[b,c], Jahan B. Ghasemi[a],*
[a] Department of Chemistry, Faculty of Science, University of Tehran, Iran
[b]Department of Analytical Chemistry, University of the Basque Country UPV/EHU, Barrio Sarriena S/N, Leioa 48940, Spain
[c] IKERBASQUE, Basque Society for the Promotion of Science, Plaza Euskadi, 5, Bilbao 48009, Spain
Jahan.ghasemi@gmail.com

ABSTRACT

Food adulteration occurs globally and affects nearly all food commodities. It is not only an economic issue but also poses serious health risks to consumers [1]. The worldwide demand for spices and herbs has been rising, driven by their expanding role in the food service sector—for example, as additives in ready-made meals, the modern consumer's pursuit of novel flavors and ethnic cuisines, and their perceived health benefits [2]. In this context, where spices are highly valued for their culinary and functional properties as well as their high economic value supported by recognized quality standards, fraudulent practices aimed at increasing trade margins have become a serious concern. Spices are particularly vulnerable to food fraud due to their extensive global trade networks, making them commodities that require increased regulation and advanced detection methods. Traditional approaches for detecting food adulterants, including chemical assays and physical examinations, have long been employed but are often labor-intensive, time-consuming, and unsuitable for high-throughput industrial applications. In contrast, modern optical methods such as Near-Infrared (NIR) Spectroscopy and RGB image analysis provide rapid, non-destructive alternatives for food quality assessment [3]. Building upon these advances, hyperspectral imaging (HSI) and multispectral imaging (MSI) offer unique advantages by simultaneously capturing spatial and spectral information, thereby enabling comprehensive characterization of both the chemical and physical attributes of complex food matrices. This study explored the application of HSI and MSI for detecting adulteration in powdered spices relevant to the Iranian market. Imaging data were being analyzed using chemometric and machine learning methods. Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and additional classification algorithms were being applied to extract key features and construct predictive models. Through this approach, the characteristic spectral features of common adulterants were identified, and the comparative performance of HSI and MSI, combined with chemometric methods, was systematically evaluated.

Keywords
Hyperspectral imaging, Multispectral imaging, Spice adulteration, ML-driven detection

References

[1] Lanjewar, Madhusudan G., Pranay P. Morajkar, and Jivan Parab. "Detection of tartrazine colored rice flour adulteration in turmeric from multi-spectral images on smartphone using convolutional neural network deployed on PaaS cloud." Multimedia Tools and Applications 81, no. 12 (2022): 16537-16562.

Seed quality assessment using hyperspectral and multispectral imaging techniques

Yeganeh Nazari[a], Mahia Vazifeh Solout[a], José Manuel Amigo[b,c], Jahan B. Ghasemi[a]*

[a] Department of Chemistry, Faculty of Science, University of Tehran, Iran

[b]Department of Analytical Chemistry, University of the Basque Country UPV/EHU, Barrio Sarriena S/N, Leioa 48940, Spain

[c] IKERBASQUE, Basque Society for the Promotion of Science, Plaza Euskadi, 5, Bilbao 48009, Spain

Jahan.ghasemi@gmail.com

ABSTRACT

Seed quality, defined by germination potential and physicochemical attributes, is a fundamental factor in plant breeding and crop production. High-quality seeds provide a strong start for plant growth and are directly linked to improved yield. Beyond agriculture, seeds are often consumed directly as food, where their quality is closely associated with eating properties such as texture, flavor, and nutritional value. To meet consumer expectations and agricultural demands, seeds must be carefully processed and stored after harvest, and their quality must be monitored reliably [1]. In this context, rapid, accurate, and preferably non-destructive methods for seed quality assessment are highly desirable [2]. Hyperspectral imaging (HSI) has recently emerged as a powerful analytical tool for non-destructive seed quality and safety evaluation. By simultaneously acquiring spectral and spatial information, HSI combines the strengths of spectroscopy and imaging, enabling the chemical characterization of heterogeneous samples together with the spatial distribution of their components. Multispectral imaging (MSI), while more limited in spectral resolution, offers a cost-effective and faster alternative, making it particularly attractive for large-scale applications. This study explored the use of HSI and MSI for seed quality assessment in combination with chemometric and machine learning approaches. Analytical methods, including Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), Support Vector Machine (SVM), and other classification algorithms, were applied to extract key features and develop predictive models. The efficiency of these methods was evaluated for classifying seeds according to quality attributes, detecting defects, and differentiating between seed categories. Ultimately, the results underscore the potential of HSI and MSI as non-invasive and reliable tools for assessing seed quality, offering valuable insights for modern agriculture, food safety, and seed industry practices.

Keywords

Hyperspectral imaging, Multispectral imaging, Seed quality, non-invasive method

References

[1] Feng, Lei, Susu Zhu, Fei Liu, Yong He, Yidan Bao, and Chu Zhang. "Hyperspectral imaging for seed quality and safety inspection: A review." Plant methods 15, no. 1 (2019): 91.

[2] Fathi, Golshid, Seyed Ahmad Mireei, Mehrnoosh Jafari, Morteza Sadeghi, Hassan Karimmojeni, and Majid Nazeri. "Spatial analysis of hyperspectral images for detecting adulteration levels in bon-sorkh (Allium jesdianum L.) seeds: Application of voting classifiers." Smart Agricultural Technology 10 (2025): 100810.

AI-Driven Design of PCSK9 Allosteric Inhibitors for Atherosclerosis via Graph Neural Networks and Structure-Based Validation

Behnaz Horri, Jahan B. Ghasemi, Hasan Behnejad*
Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran
E-mail address (corresponding author): *jahan.ghasemi@ut.ac.ir*

ABSTRACT

Atherosclerosis (AS) is a chronic, progressive vascular disorder and a leading cause of cardiovascular morbidity and mortality worldwide. Its pathogenesis is driven by the accumulation of low-density lipoprotein cholesterol (LDL-C) within arterial walls, promoting lipid-induced inflammation, plaque formation, and ischemic events. Hepatic clearance of LDL is mediated by the low-density lipoprotein receptor (LDLR), which normally recycles after LDL internalization. Proprotein convertase subtilisin/kexin type 9 (PCSK9) disrupts this process by binding LDLR and directing it toward lysosomal degradation, elevating plasma LDL-C and accelerating atherogenesis. Structurally, PCSK9 contains a prodomain, catalytic domain, and C-terminal domain, with an induced-fit allosteric pocket emerging as a promising therapeutic target[1].

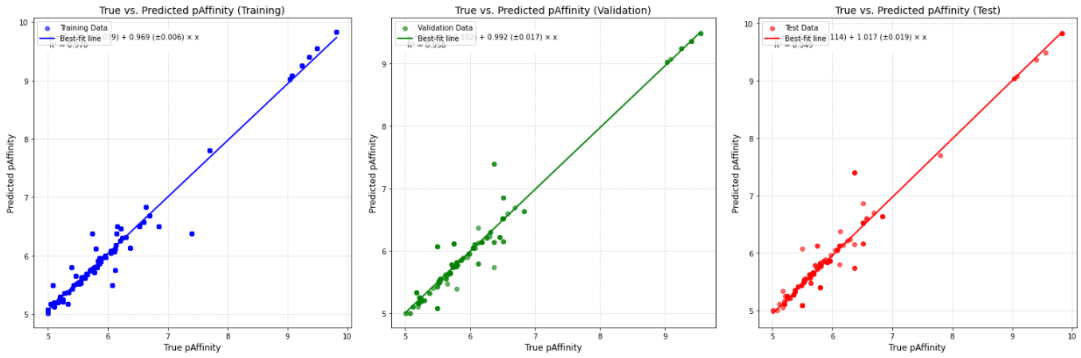
Traditional drug discovery methods, including high-throughput screening, have identified candidate inhibitors but are resource-intensive and limited in efficiency. Integrating machine learning (ML) and artificial intelligence (AI) provides a transformative alternative, enabling rapid bioactivity prediction, compound prioritization, and experimental burden reduction. In particular, graph neural networks (GNNs) represent molecules as graphs, capturing complex structural and topological relationships beyond conventional descriptors[2].

A curated dataset of 1,039 PCSK9 ligands was assembled using a custom workflow integrating multiple databases. For each compound, SMILES strings were retrieved, validated using RDKit, and converted into molecular graph representations suitable for GNN modeling. IC_{50} values were standardized to pIC_{50}, and the dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets. Predictive performance was rigorously evaluated using statistical metrics and Williams plots for applicability domain analysis, ensuring both accuracy and robustness[3].

To validate functional relevance, pharmacophore modeling, molecular docking, and molecular dynamics simulations were performed to characterize and refine ligand–protein interactions at the allosteric pocket of PCSK9 (PDB ID: 6U26). ADMET profiling was concurrently conducted to evaluate drug-likeness and safety properties. These complementary analyses not only optimized binding conformations and interaction networks but also increased confidence in the predicted affinity and selectivity of candidate inhibitors[4].

Collectively, this AI-assisted framework demonstrates the power of GNN-based modeling, integrated with structure-based validation, to accelerate the discovery of orally bioavailable, cost-effective PCSK9 inhibitors. By uniting ML-driven insights with conventional computer-

aided drug design (CADD) and ADMET evaluation, the transition from virtual screening to experimental validation is streamlined, paving the way for next-generation therapeutics against atherosclerosis[5, 6].



Keywords: Atherosclerosis, PCSK9, Allosteric Pocket, Graph neural networks (GNNs), Machine learning (ML), Structure-based drug design (SBDD), Drug discovery,

References

1. Kuzmich, N., et al., *PCSK9 as a Target for Development of a New Generation of Hypolipidemic Drugs.* Molecules, 2022. 27(2).
2. Zhang, Z., et al., *Graph neural network approaches for drug-target interactions.* Curr Opin Struct Biol, 2022. 73: p. 102327.
3. Chen, J.H., et al., *Integration of machine learning and experimental validation reveals new lipid-lowering drug candidates.* Acta Pharmacol Sin, 2025. 46(9): p. 2547-2563.
4. Petrilli, W.L., et al., *From Screening to Targeted Degradation: Strategies for the Discovery and Optimization of Small Molecule Ligands for PCSK9.* Cell Chem Biol, 2020. 27(1): p. 32-40 e3.
5. Iwaloye, O., et al., *Computer-aided drug design in anti-cancer drug discovery: What have we learnt and what is the way forward?* Informatics in Medicine Unlocked, 2023. 41.
6. Zuo, H., et al., *Drug design targeting the AMPK signalling pathway with herbal medicines for atherosclerosis therapy.* LaboratoriumsMedizin, 2017. 41(2): p. 105-112.

Leveraging Machine Learning and Graph Neural Networks for the Discovery of JAK2 Inhibitors: Structure and Ligand-Based Approaches to Target Inflammatory and Autoimmune Diseases

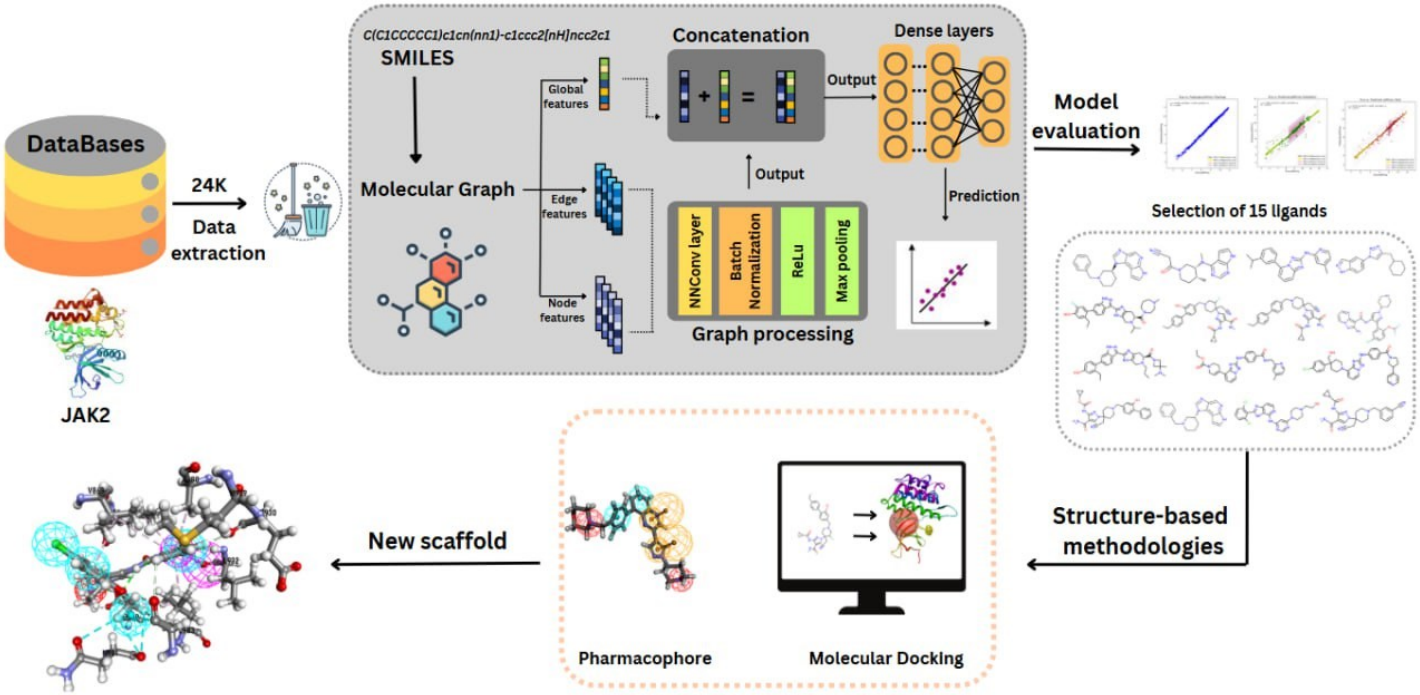
Shaghayegh K. Nezhad[a], Mahia V. Solout[a], Sogol Meknatkhah[b], Fatemeh Amini[a], and Jahan B. Ghasemi[a,]*

*Corresponding Author : Jahan B. Ghasemi, Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran, Tel: +982161112726 Email: jahan.ghasemi@ut.ac.ir

ABSTRACT

The Janus kinase (JAK) family, activated by cytokines through the JAK/STAT signaling pathway, plays a pivotal role in mediating inflammatory and autoimmune responses in diseases such as psoriasis. Consequently, the inhibition of JAK family members represents a promising therapeutic strategy[1, 2]. In recent years, machine learning techniques have garnered significant attention for their speed and efficiency, becoming increasingly integral to modern drug discovery pipelines[3]. This study aimed to identify potent JAK2 inhibitors by integrating machine learning with graph-based modeling approaches. We developed a graph neural network (GNN) architecture that leverages graph convolutions enriched with bond-aware features, enabling differential weighting of bond types during feature aggregation. This design overcomes the limitations of traditional descriptor-based models by capturing fine-grained structural and topological information essential for bioactivity prediction. A curated dataset of 1,869 JAK2 inhibitors with experimentally determined affinity values was compiled and partitioned into training (70%), testing (15%), and validation (15%) sets. The model's predictive performance was rigorously evaluated using Williams plots and statistical metrics. To validate the functional relevance of the predictions, structure-based techniques including molecular docking and pharmacophore modeling were applied to assess ligand binding within the JAK2 ATP-binding site. This integrative approach not only identified high-affinity ligands with strong structural complementarity but also revealed that distinct chemical scaffolds can converge on similar interaction profiles. These findings substantiate scaffold hopping as a viable strategy for diversifying inhibitor chemotypes while preserving target engagement. Overall, our results highlight the potential of combining deep learning with structural biology to accelerate the discovery of selective JAK2 inhibitors for inflammatory and autoimmune diseases.

Keywords (Times New Roman 10pt, Bold): “Psoriasis”, “JAK2 Inhibitors”,” Drug-target interaction”, “Graph Neural Networks (GNNs)”, “Molecular Docking”, “Pharmacophore Modeling”.



References

1. Funk, P.J., et al., *Comparing available JAK inhibitors for treating patients with psoriasis.* Expert Rev Clin Immunol, 2022. **18**(3): p. 281-294.
2. Yamaoka, K., et al., *The Janus kinases (Jaks).* Genome Biology, 2004. **5**(12): p. 253.
3. Deng, J., et al., *Artificial intelligence in drug discovery: applications and techniques.* Brief Bioinform, 2022. **23**(1).

Enhancing pH Measurement Accuracy Using Hyperspectral Imaging and commercial pH Strips

Arshia Rostampour[*1] *Hamid Abdolahi*[1]
[1]*Institute for Advanced Studies in Basic Sciences*
*a.rostampour@iasbs.ac.ir

ABSTRACT

Determining pH is a fundamental process across scientific disciplines, ranging from chemistry and biology to environmental monitoring and industrial applications [1]. One of the most accessible and widely used methods for pH measurement is the commercial pH strip, which relies on visible color changes to indicate acidity or alkalinity. However, despite their convenience and low cost, pH strips suffer from major drawbacks, including limited resolution (typically 1 pH unit) and susceptibility to human error during visual interpretation [2]. To address these issues, previous researchers have explored the use of conventional RGB cameras[3], which improved performance but remain limited by their limited spectral resolution(3 color channels). We believe there is significant untapped potential in this methodology, and therefore, in this work, we propose the integration of hyperspectral imaging (HSI) a higher-information acquisition technique with a commercially available Merck four-pad pH strip, aiming to improve accuracy and minimize human error. For the experimental phase, buffer solutions spanning pH 1.30 to 13.33 were prepared in ~0.1 increments by titration of Bohr–Robinson buffer with NaOH or HCl, resulting in a total of 120 samples. A pH strip was immersed in each solution for 2 seconds, and after approximately 20 seconds a hyperspectral image in the 400–1000 nm range (visible to near-infrared) was acquired using a PartoAfroz hyperspectral camera operating at 2.7 FPS. From each strip, four spectra (one per pad) were obtained by averaging the pixel responses with a custom MATLAB script, and these four spectra were augmented row-wise to construct the calibration matrix X. The dataset was then divided into training and validation sets in a 7:3 ratio using the Kennard–Stone algorithm. Ground-truth pH values were measured with a Metrohm digital pH electrode, providing a resolution of ±0.01. To evaluate the feasibility of the proposed method, two regression models were tested: partial least squares regression(linear model) and radial basis function(nonlinear model). The Durbin–Watson statistic revealed strong autocorrelation in the residuals of the PLS model, indicating nonlinear behavior in the dataset. This issue was resolved by the RBF model, which achieved a root mean square error (RMSE) of ~0.06 for cross-validation samples. These results demonstrate a substantial improvement in the analytical performance of pH strips and confirm the potential of hyperspectral imaging, coupled with nonlinear modeling, to transform inexpensive pH strips into high-precision analytical tools for accurate pH measurement.

Keywords: **“Hyperspectral Imaging” “pH” “visible spectrum”**

References:

[1] D. Voet and J. G. Voet, Biochemistry, 4th ed., Wiley, 2010. ISBN: 978-0470-57095-1

[2] R. W. Sabnis, Handbook of Acid-Base Indicators, CRC Press, 2007. DOI:10.1201/9780849382192.

[3] Xiong XL, Ma YP, Liu H, Huang CZ, Zhou J. Efficient and Accurate pH Determination with pH Test Strips Based on Machine Learning. Anal Chem. 2024 Jul 16;96(28):11498-11507

Comparative investigation of BORGEN norms

Hamideh Bakhshi[a], Hamid Abdollahi[a*], Róbert Rajkó[b]

a Department of Chemistry, Institute for Advanced Studies in Basic Sciences (IASBS), 444 Prof Yousef Sobouti Blvd, Zanjan 45137-66731, Iran
[b] John von Neumann University, H-6000 Kecskemét, Izsáki út 10., Hungary
*To whom correspondence should be addressed. E-mail: abd@iasbs.ac.ir

ABSTRACT

Lawton and Sylvestre first proposed an analytical solution for determining feasible regions in the self-modeling curve resolution (SMCR) method for two-component systems [1]. Borgen and co-workers [2] later extended this framework to three-component systems, while Rajkó and István[3] revitalized Borgen's approach by providing a rigorous theoretical interpretation and a computational geometry-based algorithm for constructing Borgen plots. In these developments, data were normalized using the vector 1-norm, although Borgen and Kowalski [2] had previously introduced a more general standardization scheme, now known as Borgen norms. This study presents a comparative analysis of Borgen norms using simulated two- and three-component datasets. Initially, two-component chromatographic data were simulated, and the feasible regions for each component were determined using the Lawton and Sylvestre approach under different normalization schemes, including normalization to the first eigenvector, 1-norm, and normalization to the pure profile. Graphical representation revealed that the feasible region for a given component varies depending on the normalization method. Quantitatively, the area under the profile was calculated for each component and normalization type, showing significant differences across methods. For instance, in the two-component dataset, the area under the concentration profile of the first component was 1.8148 for normalization to the first eigenvector, 0.2958 for 1-norm, 0.4572 for normalization to the pure profile of component one, 5.1858 for normalization to the pseudo-inverse of the first component, and 11.5851 for the pseudo-inverse of the second component. Similar results were observed for other components and the three-component dataset. In summary, normalization using the Borgen norm corresponding to the first eigenvector aligns all points vertically at x=1, whereas 1-norm normalization places points along x+y=1. These differences indicate that each Borgen norm projects the data in a distinct direction, resulting in feasible regions of varying size for each component.

Keywords: "SMCR method", "rotational ambiguity", "Borgen norm".

References

[1] Lawton, William H., and Edward A. Sylvestre. "Self-modeling curve resolution." Technometrics 13.3 (1971): 617-633.
[2] Borgen, Odd S., and Bruce R. Kowalski. "An extension of the multivariate component-resolution method to three components." Analytica Chimica Acta 174 (1985): 1-26.
[3] Rajkó, Róbert, and Krisztina István. "Analytical solution for determining feasible regions of self-modeling curve resolution (SMCR) method based on computational geometry." Journal of Chemometrics: A Journal of the Chemometrics Society 19.8 (2005): 448-463

Multi-criteria evaluation of combined preprocessing techniques for first-order spectroscopic calibration

Mohammad Khademloo[a], Hamid Abdollahi[a]*
a Department of Chemistry, Institute for Advanced Studies in Basic Sciences (IASBS), 444 Prof Yousef Sobouti Blvd, Zanjan 45137-66731, Iran
E-mail address: abd@iasbs.ac.ir

ABSTRACT:

Accurate calibration in spectroscopic analysis is highly dependent on the quality of input data. However, raw spectral measurements are often influenced by noise, baseline drift, and light scattering, which may mask relevant chemical information and significantly reduce predictive performance. Preprocessing is therefore a critical step to improve data quality before calibration. A variety of methods such as smoothing, baseline correction, and scatter compensation have been proposed for this purpose. Nonetheless, the effectiveness of these methods varies greatly depending on the characteristics of the dataset, making the selection of an optimal preprocessing strategy a challenging task [1]. To address this challenge, the MATLAB toolbox Prospectool has been introduced. It systematically tests a wide range of preprocessing methods on the given data, performs model building, and reports four evaluation criteria (including cross-validation prediction error). Depending on the research objective, one of these criteria can be used to select the best model and the most suitable preprocessing options for the dataset [2]. In our research, we applied Prospectool to real spectroscopic data and, to consider all criteria simultaneously, employed the Sum of Ranking Differences (SRD) method. SRD aggregates multiple metrics by ranking models relative to an ideal reference and quantifying their deviation [3].As a case study on near-infrared (NIR) spectra of rice samples for predicting amylose content, we constructed 409 PLS models with various preprocessing strategies plus one model without preprocessing. The PLS model built on raw spectra yielded a cross-validation error of 5.30, whereas the best configuration—Savitzky–Golay smoothing (window size 43, polynomial order 2), first derivative transformation, and second-order detrending—reduced the error to 3.16. These findings demonstrate that combining Prospectool's automated exploration with SRD-based multi-criteria decision-making enables rigorous, data-driven selection of preprocessing strategies, thereby improving the accuracy and robustness of spectroscopic calibration models in real-world applications.

Keywords: Spectral Preprocessing, First-order Calibration, Partial Least Squares Regression (PLSR), Sum of Ranking Differences (SRD) Method, Multi-Criteria Evaluation

References:

[1] L. Xu, Y.P. Zhou, L.J. Tang, H.L. Wu, J.H. Jiang, G.L. Shen, R.Q. Yu, Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration, Anal. Chim. Acta 616 (2008) 138–143.
[2] J. Ezenarro, D. Schorn-García, O. Busto, R. Boqué, ProSpecTool: A MATLAB toolbox for spectral preprocessing selection, Chemom. Intell. Lab. Syst. 247 (2024) 105096.
[3] K. Héberger, K. Kollár-Hunek, Sum of ranking differences for method discrimination and its validation: Comparison of ranks with random numbers, J. Chemom. 25 (2011) 151–158.

Feasible band boundaries computation in bilinear matrix decomposition using essential data

Somaye Vali Zade[a], Mathias Sawall[b] , Klaus Neymeyr[b,c], Alejandro C. Olivieri[d], Roma Tauler[e], Hamid Abdollahi[f]*

[a] *Halal Research Center of IRI, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran*
[b] *University of Rostock, Institute of Mathematics, Ulmenstraße 69, 18057, Rostock, Germany*
[c] *Leibniz-Institute for Catalysis, Albert-Einstein-Straße 29a, 18059, Rostock, Germany*
[d] *Instituto de Química Rosario (CONICET-UNR), Facultad de Ciencias Bioquímicas y Farmac´ euticas, Universidad Nacional de Rosario, Rosario, Argentina*
[e] *Institute of Environmental Assessment and Water Research, IDAEA-CSIC, Barcelona, Spain*
[f] *Institute for Advanced Studies in Basic Sciences, 45195-1159, Zanjan, Iran*
[*] E-mail: smy.valizade@gmail.com

ABSTRACT

Multivariate curve resolution methods are usually confronted with non-unique pure component factors. This rotational ambiguity can be represented by ranges of feasible profiles, which are equally compatible with the imposed constraints. Sensor-wise N-BANDS is an effective algorithm for the calculation of the bounds of feasible profiles in the presence of noise, but suffers from high computational cost. Sensor-wise N-BANDS has been combined with the concept of essential data points to speed up the computation. The combined algorithm provides full curve resolution independent of the number of chemical species. The effectiveness of the proposed algorithm is demonstrated for simulated chromatographic data and experimental spectroelectrochemical data. With the new proposal, the boundaries of the set of feasible profiles in bilinear matrix decomposition can be estimated in a reasonable time, for any number of components, and in the presence of instrumental noise.

Keywords : Bilinear decomposition, Essential points, Rotational ambiguity, Multi-component systems

References (Times New Roman 10pt Bold):

[1] A.C. Olivieri, Estimating the boundaries of the feasible profiles in the bilinear decomposition of multi-component data matrices, Chemometr. Intell. Lab. Syst. 216 (**2021**) 104387.
[2] M. Ghaffari, N. Omidikia, C. Ruckebusch, Essential spectral pixels for multivariate curve resolution of chemical images, Anal. Chem. 91 (17) (**2019**) 10943–10948.
[3] S.V. Zade, et al., Data point importance: information ranking in multivariate data, J. Chemometr. 37 (1) (**2023**) e3453.

Evaluation of Classical and Statistical Mechanics-Based Models for Adsorption Data: Insights into Accurate Fitting and Physical Interpretation

*Fatemeh Khabbazi, Ebrahim Nemati-Kande**
Department of Physical Chemistry, Chemistry Faculty, Urmia University, Urmia, Iran
E-mail address: e.nemati@urmia.ac.ir

ABSTRACT

In this research, classical models and those derived from statistical mechanics were employed to evaluate adsorption data from the literature. The goal was to examine their performance and effectiveness in fitting data [1]. The study utilized a wide range of materials, including gases and predominantly soluble substances, on common adsorbents like Graphene and nanomaterial surfaces such as boron nitride. The results indicated that while classical models like Langmuir and Temkin successfully fit the data, they offer little insight into the adsorption mechanism [2]. In contrast, models based on statistical mechanics, due to their stronger theoretical foundation, not only fit the data very successfully [with a coefficient of determination, R^2, close to 1] but are also capable of elucidating the physical meaning of their parameters. This provides a valuable perspective on the mechanism and nature of adsorption [3]. For example, in multilayer models, the energy and thus the half-saturation concentration varies for each layer. Surfaces with higher energy have a greater tendency for adsorption and reach saturation at lower concentrations. The first layer, formed directly on the adsorbent surface, possesses higher energy and enjoys greater stability. Subsequent layers, influenced by weaker intermolecular interactions, behave differently. In models where all layers have identical energy, the total adsorption capacity is distributed, with the first layer being dominant. Conversely, in models with different layer energies, each layer is occupied independently [1]. In conclusion, utilizing advanced thermodynamic models based on statistical mechanics requires more experimental data to accurately determine the specific physical meaning of each parameter. Therefore, it is recommended that researchers in adsorption kinetics and equilibrium employ these more advanced and contemporary models, most of which have been developed in the last decade.

Keywords: Surface adsorption, Modeling, nanomaterial, Statistical mechanics, Data fitting

References:

[1] O. Amrhar, L. El Gana, M. Mobarak, "Calculation of adsorption isotherms by statistical physics models: a review", Environmental Chemistry Letters, 19, **(2021)**, 4519–4547.

[2] M. Ben Yahia, H. Yahya Saeed AlZahrani, M. Ben Yahia, "Application of advanced statistical physics modeling for the physicochemical analysis of adsorption isotherms of manganese (II) on porphyrins H_2TPP and H_2TTP", Brazilian Journal of Chemical Engineering, 38, **(2021)**, 373–388.

[3] L. Sellaoui, F. E. Soetaredjo, S. Ismadji, A. Bonilla-Petriciolet, C. Belver, J. Bedia, et al, "Insights on the statistical physics modeling of the adsorption of Cd^{2+} and Pb^{2+} ions on bentonite-chitosan composite in single and binary systems", Chemical Engineering Journal, 354, **(2018)**, 569–576.

Machine Learning Bases Optimal Design of Carbon Nanosheets

Khalil Shahbazpoor[1], *Ebrahim Nemati-Kande*[2]*,
[2]*Department of Mathematics, Science Faculty, Urmia University, Urmia, Iran*
[1]*Department of Physical Chemistry, Chemistry Faculty, Urmia University, Urmia, Iran*

E-mail address (corresponding authors): e.nemati@urmia.ac.ir

ABSTRACT
This work presents an automated computational framework for generating carbon nanosheets with specific hybridization states (sp, sp^2, sp^3) while preserving exact space group symmetry. The methodology combines mathematical group theory with chemical bonding principles to create physically realistic carbon structures. Our algorithm begins by constructing a quotient graph that encodes the desired coordination environment for each carbon atom in the asymmetric unit [1]. We then generate initial atomic positions constrained by hybridization-specific geometry: linear for sp, and trigonal planar for sp^2, configurations. Space group symmetry operations are applied to generate the full periodic structure in a two-dimensional space. A crucial step involves removing atoms that violate minimum distance constraints (1.2 Å), ensuring no unphysically close contacts. The lattice optimization then proceeds while strictly preserving the space group symmetry, minimizing an objective function that penalizes bond stretching, angle bending, and close contacts, preserving the 20 Å vacuum over the z directory. The implementation systematically explores parameter spaces including multiple space groups of 143 to 194 lattice parameters, and hybridization combinations. For each configuration, we generate both optimized and non-optimized structures, providing comprehensive datasets for further quantum mechanical calculations [2]. The automated pipeline produces structures ready for Quantum ESPRESSO calculations, enabling high-throughput screening of carbon materials with tailored properties. The method demonstrates utility for exploring hypothetical carbon phases that maintain specific hybridization patterns while achieving space group symmetry, offering new avenues for carbon materials design with applications in electronics, mechanics, and energy storage [3].

Keywords: Carbon allotropes, Machine Learning, High-throughput screening, Computational design

References:
[1] D. Chauhan, G. Hou, V. Ng, S. Chaudhary, M. Paine, K. Moinuddin, M. Rabiee, M. Cahay, N. Lalley, V. Shanov, and M. J. Schulz, "Carbon Hybrid Materials—Design, Manufacturing, and Applications", *Nanomaterials*, vol. 13, no. 3, (2023), 431.
[2] S. Peng, Y. Wang, R. Dong, Y. Li, and S. Ramakrishna, "Electrospun carbon nanofibers and their hybrid composites as advanced materials for energy conversion and storage", *Nano Energy*, vol. 22, (2016), 361-395.
[3] Z. Zhao, H. Chen, W. Zhang, S. Yi, H. Chen, Z. Su, B. Niu, Y. Zhang, and D. Long, "Defect engineering in carbon materials for electrochemical energy storage and catalytic conversion", *Materials Advances*, vol. 4, (2023), 835-867.

Oriented Hyper-Cyclic Geometry with Curvature: A Weighted Directed Graph Model for Molecular Asymmetry and Chirality

Khalil Shahbazpoor[1]*, *Ebrahim Nemati-Kande*[2]*
[2]*Department of Mathematics, Science Faculty, Urmia University, Urmia, Iran*
[1]*Department of Physical Chemistry, Chemistry Faculty, Urmia University, Urmia, Iran*

E-mail address (corresponding authors): kh.shahbazpoor@urmia.ac.ir
e.nemati@urmia.ac.ir

ABSTRACT
This paper introduces the Oriented Hyper-Cyclic Geometry with Curvature (NESH Structure), a novel geometric framework that models molecules as finite, directed hyper-graphs where vertices represent atoms and directed edges encode covalent bonds with metric weights. Unlike traditional chemical graphs [1], NESH incorporates bond lengths, curvature derived from electron density, and chiral holonomy, enabling the quantification of asymmetric distances and stereochemical properties. The model is applied to benzene and cyclohexane, revealing how directional weighting and curvature influence path lengths and molecular chirality. Key advantages include intrinsic asymmetry, curvature-driven reactivity, and a unified representation of topology and metric data [2]. The NESH Structure offers a robust tool for analyzing cyclic systems, stereoisomer enumeration [3], and chiral resolution, with potential applications in drug design and catalytic screening.

Keywords: Oriented hyper-cyclic geometry, Curvature-driven reactivity, Asymmetric distance, Chiral holonomy, Molecular chirality

References:
[1] Bader, R. F. W. (1990). Atoms in Molecules: A Quantum Theory. Oxford University Press.
[2] Cotton, F. A. (1990). Chemical Applications of Group Theory (3rd ed.). Wiley.
[3] FDA. (2021). Chiral Drugs: Development and Review. FDA Guidance Document.

A Fluorescent Concentration-Based Sensor Array Utilizing Hydrogen Peroxide-Etched Sulfur Quantum Dots (SQDs-H_2O_2) for Simultaneous Detection of Chromium(VI) and Bismuth(III) Ions via Multivariate Calibration

Sima Mojarrad[1], Abdolhossein Naseri[1], Mehdi Moghtader[2]*

[1] Department of Analytical Chemistry, faculty of Chemistry, University of Tabriz, *Tabriz, Iran*

[2] Department of Chemistry, Urmia University, Urmia, Iran

E-mail address (corresponding author): a_naseri@tabrizu.ac.ir

ABSTRACT

Sulfur quantum dots (SQDs) have emerged as a prominent fluorescent nanomaterial since their introduction, distinguished by their excellent water solubility, tunable photoluminescence, and intrinsic biocompatibility [1]. This research capitalizes on these properties by implementing a surface engineering strategy, wherein treatment with hydrogen peroxide induces porosity on the SQD surface [2]. This modification concurrently enhances the quantum yield and reduces the nanoparticle size, resulting in an optically superior material. The foundational discovery was that the fluorescence intensity of these porous SQDs is quantitatively quenched upon the addition of Cr(VI) and Bi^{3+} ions [3]. Based on this phenomenon, a novel sensor array was designed for the simultaneous detection and identification of these heavy metal contaminants. This array generates a unique, fingerprint-like fluorescent response pattern for each analyte, facilitating discrimination even in complex mixtures where individual sensor elements may exhibit low selectivity. For precise quantification, the methodology integrated fluorescence spectroscopy with multivariate calibration, specifically Partial Least Squares (PLS) regression. To ensure model robustness and optimal predictive power, a full factorial experimental design (comprising 6 levels for Bismuth and 6 levels for Chromium) was utilized to construct the calibration set of binary mixtures. The developed method demonstrated exceptional analytical performance, yielding detection limits of 2.46 µM and 8.55 µM for Bi^{3+} and Cr(VI) ions, respectively, a linear dynamic range of 1–100 µM (with R^2 values of 0.9912 and 0.9813), and low prediction errors (RMSEP of 4.10 and 3.96). The practical applicability and accuracy were conclusively validated through successful deployment in real-world matrices, including municipal and spring water samples, achieving excellent recovery values ranging from 90.31–103.25% for Bi^{3+} and 95.53–112.93% for Cr(VI).

Keywords: sulfur quantum dot, Partial Least Squares, Fluorescence, Heavy metal contaminant

References

[1] Pal A, Arshad F, Sk MP. Emergence of sulfur quantum dots: Unfolding their synthesis, properties, and applications. Advances in Colloid and Interface Science. 2020;285:102274.

[2] Wang H, Wang Z, Xiong Y, Kershaw SV, Li T, Wang Y, et al. Hydrogen peroxide assisted synthesis of highly luminescent sulfur quantum dots. Angewandte Chemie. 2019;131(21):7114-8.

[3] Tan Q, An X, Pan S, Liu H, Hu X. Hydrogen peroxide assisted synthesis of sulfur quantum dots for the detection of chromium (VI) and ascorbic acid. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. 2021;247:119122.

Molecular Modeling and Design of Novel Inhibitors for GABAA Receptors

Farshad Jafari, Fatemeh Amini, Noushin Agha Babaie, Morteza Farnia, Jahan B. Ghasemi*
Chemistry Faculty, School of Sciences, University of Tehran, Tehran, Iran
Jahan.ghasemi@gmail.com

ABSTRACT

Gamma-aminobutyric acid (GABA) is the principal inhibitory neurotransmitter in the central nervous system, playing a vital role in maintaining the balance between excitatory and inhibitory signaling. Dysregulation of the GABAergic system is implicated in several neurological and psychiatric disorders, underscoring the importance of identifying novel ligands targeting the GABAA receptor.[1][2]

In this study, a three-dimensional Quantitative Structure–Activity Relationship (3D-QSAR) model was developed for ligands binding to the benzodiazepine site of the GABAA receptor. A dataset of 51 ligands with reported activity was compiled from published studies, and models were constructed using SYBYL software. Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) demonstrated strong predictive power, yielding cross-validated R2 values of 0.755 and 0.652, and non-cross-validated R2 values of 0.966 and 0.903, respectively. The contour maps generated by CoMFA and CoMSIA highlighted critical steric, electrostatic, hydrophobic, and hydrogen-bonding features influencing ligand activity, thereby providing a visual framework for structure–activity interpretation.

A pharmacophore model was also generated to identify the essential steric and electronic features required for activity, providing an additional framework for ligand design

[2]. molecular docking was then conducted using the GABAA receptor crystal structure (PDB ID: 6X3X) to identify important interactions in binding site. Docking revealed the best poses of active ligands with conserved residues in the benzodiazepine pocket.

Using the QSAR results, the most important substitutions were identified and the most important features in increasing ligand activities were extracted through the model pharmacophore, and using the properties of our scaffold hopping, a new framework for GABA inhibition was presented. To further validate ligand–receptor interactions, molecular dynamics (MD) simulations were performed with GROMACS using the CHARMM force field. The MD trajectories confirmed the stability of receptor–ligand complexes and provided additional insights into conformational flexibility and dynamic hydrogen bonding. By integrating 3D-QSAR, docking, and MD simulations. Overall, this study underscores the potential of novel ligand scaffolds with enhanced binding profiles relative to diazepam and demonstrates the utility of computational approaches in accelerating the discovery of therapeutics targeting GABAergic dysfunction. Several of these newly designed ligands exhibited superior predicted binding affinities compared to diazepam, highlighting their promise as lead compounds for further development.

Keywords: 3D-QSAR, Molecular docking, GABAA receptor, Molecular dynamics simulations, Pharmacophore modeling, Computational drug design.

References:

[1] Miller PS, Aricescu AR. Crystal structure of a human GABAA receptor. Nature. 2014 Aug
[2] W. Sieghart, “Structure and pharmacology of γ-aminobutyric acidA receptor subtypes,” Curr. Top. Med. Chem., 1995.

Artificial neural networks coupled Grey wolf optimizer for modeling of reactive red 198 adsorption onto novel poly(o-toluidine) / strontium hexaferrite /analsim nanocomposite

H. Alijani[1], V. Kiarostamim[1], M. Davallom[1], A.M.. Ghaedi[2], and S. Mortazavi Nik[1]*
[1]Department of Chemistry, North Tehran Branch, Islamic Azad University, Tehran, Iran
[2]Department of Chemistry, Gachsaran Branch, Islamic Azad University, Gachsaran, Iran
v_kiarostami@iau.ac.ir

ABSTRACT
Efficient control of operational parameters with application of chemometrics methods has provided powerful tools for modeling and optimization, reducing experimental effort while enhancing accuracy (1). Among these, metaheuristic algorithms such as the grey wolf optimizer (GWO)(2) have gained attention due to their ability to balance exploration and exploitation, rapid convergence, and robustness compared with traditional approaches. Furthermore, in multilayer artificial neural networks (ANN) architectures, the nonlinear relationship between independent variables and the output can be obtained. This modeled relationship is then formulated as an objective function, which metaheuristic algorithms optimize to determine the best conditions for independent factors, thereby yielding highly accurate predictive and optimization models for adsorption and removal efficiency (3). In this study, ANN coupled with grey wolf optimizer (ANN-BA) as a novel metaheuristic strategy in chemistry was used to optimize the adsorption of reactive red 198 (RR198) dye onto a poly(o-toluidine) / strontium hexaferrite /analsim (POT/SrM/An) nanocomposite from aqueous solutions. The POT/SrM/GO nanocomposite was synthesized through an in-situ polymerization method, and its physicochemical and magnetic features were characterized using various analytical techniques. Four operational parameters influencing the adsorption efficiency were optimized in a batch system by employing ANN-GWO and central composite design (CCD). The ANN-GWO model, trained via the Levenberg–Marquardt back-propagation algorithm, demonstrated superior predictive performance, achieving about 97.8% removal in synthetic and wastewater solutions, which was nearly 7% higher than the CCD model. Under these conditions, experimental findings exhibited an excellent fit with the pseudo-second-order kinetic model and the Langmuir adsorption isotherm (q_m = 101 mg/g), with determination coefficients (R^2) of 0.9998 and 0.9142, respectively. Thermodynamic evaluation indicated that the adsorption of RR198 was spontaneous ($\Delta G < 0$) at all tested temperatures, endothermic in nature (ΔH = 20.43 kJ/mol), and associated with a slight entropy increase (ΔS = 0.086 kJ/mol·K).

Keywords: Grey wolf optimizer, CCD, poly(o-toluidine), strontium hexaferrite, Analsim, Reactive red 198

References:

[1] M. Kumari, “A review on metaheuristic algorithms: Recent and future trends. Metaheuristics-Based Materials Optimization”, Elsevier,2025.103-128.
[2] D. Das, A.S. Sadiq, and S. Mirjalili, “,A Brief Review of Grey Wolf Optimizer: Variants and Applications. Optimization Algorithms in Machine Learning: A Meta-heuristics Perspective”, Springer, 2025, 67-83.
[3] S. Ebrahimpoor, V. Kiarostami, M. Khosravi, M. Davallo, A.M. Ghaedi, “ Bees metaheuristic algorithm with the aid of artificial neural networks for optimization of acid red 27 dye adsorption onto novel polypyrrole/SrFe12O19/graphene oxide nanocomposite”, Polym. Bullm,. 76 (2021} 6529–6553.

Virtual screening study of some GSK-3 inhibitors, such as Benzamide, Pyrazolopyrimidine, and β-phenylalanine derivatives, as anti-diabetic agents

M. Piri[a]*, M. H. Fatemi[b]

[a] Chemometrics Laboratory, Department of Chemistry, University of Mazandaran, Babolsar, IRAN. b Chemometrics Laboratory, Department of Chemistry, University of Mazandaran, Babolsar, IRAN.
E-mail address: mohadesepiri@email.com

ABSTRACT

The study aimed to discover novel chemical scaffolds as potential GSK-3 inhibitors via virtual screening, based on 85% similarity with Benzamide, Pyrazolopyrimidine, and β-phenylalanine derivatives from the PubChem database. Additionally, in silico-predicted Absorption, Distribution, Metabolism, and Excretion (ADME) properties were explored for the most promising molecules. The findings illuminate the molecular characteristics of analogs binding to GSK-3 and identify potential inhibitors for drug development. This study's insights contribute significantly to designing and optimizing therapeutic strategies against type 2 diabetes.

Keywords: Virtual screening; Support vector machine; Absorption, Distribution, Metabolism, and Excretion (ADME); type 2 diabetes.

References:

[1] Hermida, M. A., Kumar, J. D., & Leslie, N. R. (2017). GSK3 and its interactions with the PI3K/AKT/mTOR signaling network. Advances in biological regulation, 65, 5-15.

[2] Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., ... & Bryant, S. H. PubChem substance and compound databases. Nucleic acids research, 2016, 44(D1), D1202-D1213.

[3] Huang, S., Cai, N., Pacheco, P. P., Narrandes, S., Wang, Y., & Xu, W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer genomics & proteomics, 2018, 15(1), 41-51.

[4] Hamid, A. A., Abdul-Rasheed, O. F., Mahdi, M. F., & Atia, A. J. Design, synthesis, characterization, and biological evaluation of new diazole-benzamide derivatives as glucokinase activators with antihyperglycemic activity. Egyptian Journal of Chemistry, 2022, 65(8), 451-469.

[5] Dessalew, N., Patel, D. S., & Bharatam, P. V. 3D-QSAR and molecular docking studies on pyrazolopyrimidine derivatives as glycogen synthase kinase-3β inhibitors. Journal of Molecular Graphics and Modelling, 2007, 25(6), 885-895.

[6] Jiang, Y. K. Molecular docking and 3D-QSAR studies on β-phenylalanine derivatives as dipeptidyl peptidase IV inhibitors. Journal of molecular modeling, 2010, 16(7), 1239-1249.

[7] Zhang, M. Q., & Wilkinson, B. Drug discovery beyond the 'rule-of-five'. Current opinion in biotechnology, 2007, 18(6), 478-488.

The application of machine learning to identify the chemical features that affect the stability of low-dimensional perovskite layers

***Mahmoud Samadpour**[*], Alireza Fathi Beiraghvandi*

Physics Department, K. N. Toosi University of Technology, Tehran, Iran

samadpour@kntu.ac.ir

ABSTRACT

Low-dimensional perovskite layers have attracted widespread attention due to their unique optoelectronic properties and high potential for improving the performance of next-generation optoelectronic devices. However, precisely understanding the relationships between the chemical structures of the precursors used in their fabrication and the stability of these layers remains challenging. This research presents a machine learning-based approach to identify and evaluate the key chemical features that have the greatest impact on the stability of low-dimensional perovskite layers. A comprehensive collection of experimental and computational data, including the structural and chemical parameters of organic ammonium salts, was gathered and analyzed using advanced machine learning algorithms. The results from photoluminescence mapping showed the spatial variations of the quasi-Fermi level splitting and the degradation process. SHAP analysis revealed that an optimal combination of features—spatial geometry, the number and type of hydrogen bonds, and electron charge distribution—can significantly enhance the stability of these layers against moisture and heat. In addition to providing a fast and accurate tool for screening new materials, this approach offers valuable insights for the targeted design of stable perovskite structures.

Keywords: perovskite, chemical structure, machine learning, SHAP analysis

Optimization of Cd Adsorption by Chitosan Using Response Surface Methodology (RSM) Based on Box-Behnken Model

Marziyeh Piri[a]*, *Ebrahim Sepehr*[b]

[a] West Azerbaijan Agricultural and Natural Resources Research and Education Center, Soil and Water Research Department, Education and Extension Organization (AREEO), Urmia, Iran, [b]Department of Soil Science, Urmia University, Urmia, Iran

E-mail address (corresponding author): piri.ma@yahoo.com

Abstract

Water pollution with heavy metals is one of the most serious environmental threats in recent decades. One of the effective methods for removing heavy metals from aqueous solutions is adsorption by effective adsorbents. The factors affecting Cd adsorption by chitosan, including pH (3-6), ionic strength (0.01-0.06 M) and concentration (0-200 mg/l), were optimized using response surface methodology based on Box-Behnken model. One-way analysis of variance showed that the quadratic model is the best model for determining the interaction of study variables. The optimal conditions for maximum cadmium adsorption from aqueous solutions were obtained in the range of pH 5.5, ionic strength 0.03 and concentration 197.94 mg/l, respectively. The predicted adsorption values for the mentioned optimal conditions for cadmium adsorption were 52.8 mg/g. The obtained regression model showed that initial concentration and pH were the most effective parameters on increasing the response (adsorption) performance, respectively, and ionic strength had an inverse effect on the rate of Cd adsorption from aqueous solutions by chitosan. Therefore, chitosan can be used as an effective adsorbent for cadmium adsorption from contaminated water sources and the response surface methodology can be effective in optimizing the cadmium adsorption from aqueous solutions by chitosan.

Keywords: Adsorption, Cadmium, Response Surface Method, Chitosan, Box-Behnken Model

Explainable Artificial Intelligence (XAI) for Regulatory-Compliant Chemometric Models in Pharmaceutical Analysis

*Nastaran Bahadori**

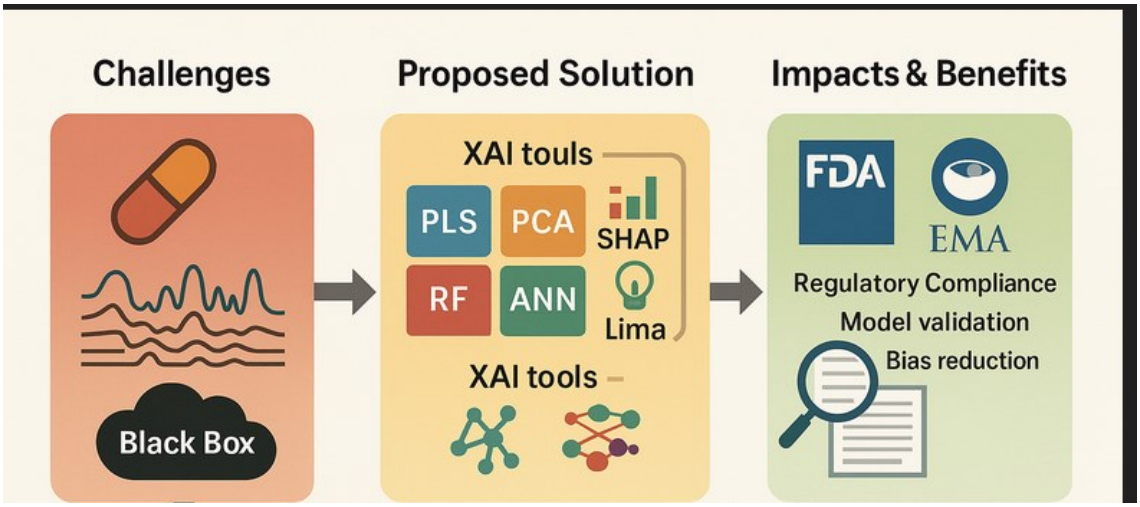
M.Sc. in Analytical Chemistry, Department of Analytical Chemistry, Faculty of Chemistry, University of Guilan, Rasht, Iran

E-mail address (corresponding author): bahadori97.n@gmail.com

ABSTRACT

The digital transformation (DX) in the pharmaceutical industry is still ongoing and the recent crises and technology creates a new and innovative way of doing research and development in the domains of drug development, data management, and regulatory compliance. Artificial intelligence (AI), and more specifically, chemometric models in the pharmaceutical sciences, is shown to be an effective agent for dealing with intricate, complex, and multidimensional data to help advance research in the areas of drug formulation, impurity determinations, and quality control of pharmaceutical products. However, the black-box processes of AI and deep learning, yet popular in applications for linearly separable information representations, still presents challenges with interpretability and regulatory acceptance [4,5,9]. Explainable artificial intelligence (XAI) provides value in assisting with this situation since it has potential tools and techniques that can help explain model rationale as LIME, SHAP, counterfactual explanations, feature importance, and saliency maps that provide transparency and create trust with the end user [4,10]. Also, the regulations from the leading regulatory agencies (FDA, EMA, ICH) require that AI models used in the analysis of pharmaceutical products be explainable, traceable and reproducible; adopting documentation showing validation and transparency in the decision process. Thus, XAI implementing into chemometrics workflow will help in regulatory compliance, reduce model bias, and instill confidence with stakeholders by exposing variable influences and interpretations of model outputs [4,11]. A variety of chemometric models, including Partial Least Squares (PLS), Principal Component Analysis (PCA), Support Vector Machines (SVM), Random Forest (RF), and Artificial Neural Networks (ANN), are increasingly being combined with XAI methods applied to enhance interpretability in functions like spectral analysis, impurity prediction, and assessment of formulation stability. There remains a wide set of challenges that still need to be addressed: (i.) technical barriers (that is, compute power), (ii.) heterogeneous data structures, (iii.) limited descriptive or prescriptive XAI implementation in industry, and (iv.) the developments of standards across jurisdictions [11,16]. Progression will require harmonizing chemometric methods and XAI approaches, improving data infrastructure, strengthening policies and standards to ensure transparency, and executing effective training to the end user and regulators - all of which fit the larger

Pharma 4.0 vision of a completely digital pharmaceutical ecosystem.



Keywords: "Digital transformation", "pharmaceutical analysis," "chemometric models", "XAI", "regulatory compliance", "transparency"

References:

[1] M. Mohan, K. Rajendiran, Y. Rajaram, S. Balaji and A. Cassinadane, "Current trends in pharmaceutical industry: Post-CoVid-19 pandemic effects", Bioinformation, 20(12), (2024), 1784–1788.

[2] D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia and R.K. Tekade, "Artificial Intelligence in Drug Discovery and Development", Drug Discovery Today, 26(1), (2020), 80–93.

[3] H.C.S. Chan, H. Shan, T. Dahoun, H. Vogel and S. Yuan, "Advancing Drug Discovery via Artificial Intelligence", Trends in Pharmacological Sciences, 40(10), (2019), 801.

[4] Q. Ding, R. Yao, Y. Bai, L. Da, Y. Wang, R. Xiang, X. Jiang and F. Zhai, "Explainable Artificial Intelligence in the Field of Drug Research", Drug Design Development and Therapy, 19, (2025), 4501–4516.

[5] S.K. Niazi, "Regulatory Perspectives for AI/ML Implementation in Pharmaceutical GMP Environments", Pharmaceuticals, 18(6), (2025), 901.

[6] P. Singh, A. Thakur and D. Yadav, "AI driven Innovations in Pharmaceuticals: Optimizing Drug Discovery and Industry Operations", RSC Pharmaceutics, 2, (2025), 437–454.

[7] L. Nene, B.T. Flepisi, S.J. Brand, C. Basson and M. Balmith, "Evolution of Drug Development and Regulatory Affairs: The Demonstrated Power of Artificial Intelligence", Clinical Therapeutics, 46(8), (2024), e6-e14.

[8] D. Aboushady, L. Samir, A. Masoud, Y. Elshoura, A. Mohamed, R.S. Hanafi and S.E. Deeb, "Chemometric Approaches for Sustainable Pharmaceutical Analysis Using Liquid Chromatography", Chemistry, 7(1), (2025),

11.

[9] M. Gallegos, V. Vassilev-Galindo, I. Poltavsky, Á. Martín Pendás and A. Tkatchenko, "Explainable chemical artificial intelligence from accurate machine learning of real-space chemical descriptors", Nature Communications, 15(1), (2024), 4345.

[10] D. Minh, H.X. Wang, Y.F. Li and T.N. Nguyen, "Explainable artificial intelligence: a comprehensive review", Artificial Intelligence Review, 55, (2021), 3503–3568.

[11] S. Pasas-Farmer and R. Jain, "From Discovery to Delivery: Governance of AI in the Pharmaceutical Industry", Green Analytical Chemistry, 13, (2025), 100268.

[12] A. Biancolillo and F. Marini, "Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis", Frontiers in Chemistry, 6, (2018), 576.

[13] K.R. Patel, L.M. Prajapati, A.K. Joshi, M.L. Kharodiya and J.R. Patel, "Application of Chemometrics in Simultaneous Spectrophotometric Quantification of Etophylline and Theophylline: The drugs with same chromophore", Iranian Journal of Pharmaceutical Sciences, 9(3), (2013), 17–28.

[14] A. EL-Gindy and G.M. Hadad, "Chemometrics in Pharmaceutical Analysis: An Introduction, Review, and Future Perspectives", Journal of AOAC INTERNATIONAL, 95(3), (2012), 609–623.

[15] G.G. Dumancas, L. Viswanath, A. Lubguban, B. Solivio, R. Malaluan and M. Waller, "Spectroscopic and Chemometric Techniques for Pharmaceutical Applications", ACS in focus, American Chemical Society, (2023).

[16] S. Hong, Y. Zhou, J. Shang, C. Xiao and J. Sun, "Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review", Computers in Biology and Medicine, 122, (2020), 103801.

[17] V. Malheiro, J. Duarte, F. Veiga and F. Mascarenhas-Melo, "Exploiting Pharma 4.0 Technologies in the Non-Biological Complex Drugs Manufacturing: Innovations and Implications", Pharmaceutics, 15(11), (2023), 2545.

Prediction of blood-brain barrier penetration coefficient of some central nervous system drugs using molecular docking modeling and quantitative structure-activity relationship

Pezhman Maghouli, Mohammad Hossein Fatemi*
Laborotory of chemometrics, faculty of Chemistry, University of Mazandaran, Babolsar,Iran
E-mail address: Pmaghouli93@Gmail.com

ABSTRACT
The blood–brain barrier (BBB) represents one of the most intricate and indispensable physiological interfaces, governing the selective transport of molecules into the central nervous system (CNS). While this barrier ensures essential neuroprotection, it also forms a major obstacle in developing CNS-active drugs, as many therapeutic agents fail to achieve sufficient penetration. Quantitative assessment of BBB permeability is commonly expressed by the LogBB index, defined as the logarithmic ratio of a compound's concentration in the brain to that in the blood. This index enables rapid screening, cost reduction, and reliable prediction of CNS drug effects. Experimental methods for measuring LogBB are valid but often restricted by high costs, lengthy procedures, ethical concerns, and occasionally limited accuracy[1,2].
In this study, 58 molecules with reported experimental LogBB values were selected from the literature[3]. To improve prediction accuracy and obtain deeper mechanistic insights, a combined strategy of molecular docking with the L-type amino acid transporter 1 (LAT1) transporter and QSAR modeling was applied. The choice of LAT1 as a transporter protein is particularly important, as it plays a critical role in CNS drug delivery and strengthens the framework of this study[4]. Docking was used to examine drug–protein interactions and estimate binding energies. QSAR established quantitative relationships between structural features and BBB permeability. The integration of these two approaches allowed both three-dimensional and structural characteristics to be considered, leading to more accurate and generalizable statistical models.
The results showed that all three modeling techniques (ANN, MLR, SVM) had strong predictive performance. However, ANN achieved a clear statistical advantage compared with MLR and SVM. For the prediction dataset, the R^2, MSE, and RMSE values were 0.872, 0.060, and 0.245 for MLR; 0.947, 0.048, and 0.220 for SVM; and 0.972, 0.026, and 0.161 for ANN. In the test set, these values were 0.861, 0.088, and 0.297 for MLR; 0.929, 0.089, and 0.298 for SVM; and 0.961, 0.051, and 0.226 for ANN. Comparison with reference data confirmed that the proposed framework, particularly in its ANN configuration, significantly improved the predictive accuracy of BBB permeability.
In conclusion, the integrated docking–QSAR framework is not only valuable for CNS drug design but also provides a powerful tool to study the potential penetration and effects of pollutants and microplastics on the BBB. This highlights its broader significance as a cross-disciplinary approach, bridging pharmaceutical research with environmental toxicology.

Keywords: Protein, molecular docking, central nervous system, QSAR modeling, blood-brain barrier, permeability

References:

[1] Bradbury, M. (1985). The blood-brain barrier. Transport across the cerebral endothelium. Circulation research, 57(2), 213–222.
[2] Nicolazzo, J. A., Charman, S. A., & Charman, W. N. (2006). Methods to assess drug permeability across the blood-brain barrier. Journal of pharmacy and pharmacology, 58(3), 281–293.
[3] Karelson, M., Dobchev, D., Tamm, T., Tulp, I., Jänes, J., Tämm, K., Lomaka, A., Savchenko, D., & Karelson, G. (2008). Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors. Arkivoc, 16, 38–60.
[4] Lee, Y., Wiriyasermkul, P., Jin, C., Quan, L., Ohgaki, R., Okuda, S., Kusakizako, T., Nishizawa, T., Oda, K., Ishitani, R., Yokoyama, T., Nakane, T., Shirouzu, M., Endou, H., Nagamori, S., Kanai, Y., & Nureki, O. (2019). Cryo-EM structure of the human L-type amino acid transporter 1 in complex with glycoprotein CD98hc. Nature Structural & Molecular Biology, 26(6), 510–517.

FastCCS: A Deep Learning Framework for Accurate Collision Cross-Section Prediction from SMILES with Broad Applications in Cheminformatics and Omics

Amir Aghajan[1], *Saeed Masoum*[*,1], *Amir Hossein Alinoori*[2]

[1] *Department of Analytical Chemistry, Faculty of Chemistry, University of Kashan, Kashan, Iran.*

[2] *Spectroscopy and Microelectronics Department, Institute of Materials and Energy, Isfahan, Iran*

E-mail address: masoum@kashanu.ac.ir

ABSTRACT

Ion mobility spectrometry (IMS) is an excellent technique for studying complex samples [1]. One of IMS's most vital capabilities is enabling the observation and comparison of dynamic changes in analyte structure [2]. This has proven challenging for most biological samples previously. Collision cross-section (CCS) values from IMS serve as crucial, structure-dependent descriptors in analytical and structural chemistry [3,4]. Using CCS values as an additional molecular descriptor improves chemical identification processes and reduces false positives in non-targeted metabolomics [5,6]. The expansive CCS databases enable precise annotation and identification of diverse molecular classes [7]. Due to its ability to enhance reporting standards and increase confidence levels in metabolomics studies, employing liquid chromatography ion mobility mass spectrometry (LC-IM-MS) produces three distinct datasets: retention time, mass-to-charge ratio (m/z), and CCS offers unique orthogonal information that can be utilized to acquire an improved understanding of the analyzed compounds compared to conventional methods [8]. As IMS is increasingly used in biological separations and proteomics research, a high-throughput, rapid computational method is required to determine CCS. Herein, we introduce FastCCS, a deep learning framework that enables accurate CCS prediction directly from Simplified Molecular Input Line Entry System (SMILES) strings and ion adduct types. Trained on a curated dataset of over 82,000 molecular structures and 51 adduct ion states, the largest chemically diverse CCS dataset to date. FastCCS achieves a median relative error of 1.7% ($R^2 = 0.99$), outperforming state-of-the-art CCS prediction tools. Key to its performance is a novel descriptor set integrating topological, geometric, and physicochemical features to capture molecular complexity. FastCCS is freely accessible via an open-access web interface (https://fastccs.com), providing a scalable cheminformatics platform that significantly enhances compound identification in complex biological and chemical matrices. The web interface, which uses deep learning to increase analysis accuracy and speed, has tremendous potential in proteomics, metabolomics, and other life sciences, as well as bioinformatics and customized medicine.

Keywords: Deep Learning, Artificial Intelligence, Ion Mobility, Chemometrics, Spectroscopy, Omics, Collision Cross Section, Mass spectrometry, Prediction, Metabolomics.

References:

[1] F. Lanucara, S.W. Holman, C.J. Gray, C.E. Eyers, "The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics", Nature Chemistry, 6, (**2014**), 281–294.
[2] E.S. Baker, W. Uritboonthai, A. Aisporna, C. Hoang, H.M. Heyman, L. Connell, D. Olivier-Jimenez, M. Giera, G. Siuzdak, "METLIN-CCS Lipid Database: An authentic standards resource for lipid classification and identification", Nature Metabolism, 6, (**2024**), 981–982.
[3] J.N. Dodds, E.S. Baker, "Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead", Journal of the American Society for Mass Spectrometry, 30, (**2019**), 2185–2195.
[4] J.A. Picache, B.S. Rose, A. Balinski, K.L. Leaptrot, S.D. Sherrod, J.C. May, J.A. McLean, "Collision cross section compendium to annotate and predict multi-omic compound identities", Chemical Science, 10, (**2019**), 983–993.
[5] I. Blaženović, T. Shen, S.S. Mehta, T. Kind, J. Ji, M. Piparo, F. Cacciola, L. Mondello, O. Fiehn, "Increasing Compound Identification Rates in Untargeted Lipidomics Research with Liquid Chromatography Drift Time-Ion Mobility Mass Spectrometry", Analytical Chemistry, 90, (**2018**), 10758–10764.
[6] E.S. Baker, C. Hoang, W. Uritboonthai, H.M. Heyman, B. Pratt, M. MacCoss, B. MacLean, R. Plumb, A. Aisporna, G. Siuzdak, "METLIN-CCS: an ion mobility spectrometry collision cross section database", Nature Methods 2023 20:12, 20, (**2023**), 1836–1837.
[7] K.L. Leaptrot, J.C. May, J.N. Dodds, J.A. McLean, "Ion mobility conformational lipid atlas for high confidence lipidomics", Nature Communications 2019 10:1, 10, (**2019**), 1–9.
[8] K.M. Hines, J. Herron, L. Xu, "Assessment of altered lipid homeostasis by HILIC-ion mobility-mass spectrometry-based lipidomics", Journal of Lipid Research, 58, (**2017**), 809–819.

Enhanced Rhodamine B Removal by Green-Synthesized Cu/TiO_2 Adsorbent: Evaluation of Adsorption Isotherms

Seyed Mehdi Sajjadi [1,*]
[1]*Department of Chemical Engineering, University of Bonab, Bonab, Iran.*
E-mail address (corresponding author): seyedmehdi_sajjadi@yahoo.com,
Mehdi_sajjadi@ubonab.ac.ir

ABSTRACT

The rapid growth of industrial sectors such as textiles, pharmaceuticals, and leather tanning has significantly increased the discharge of synthetic dyes into aquatic environments. Among these pollutants, Rhodamine B (RhB) is of particular concern because it is toxic, highly stable, and resistant to conventional biological, chemical, and physical treatments. Incomplete degradation and the generation of secondary contaminants are major drawbacks of traditional methods, which highlights the urgent need for sustainable and more efficient alternatives. Nanomaterials have recently gained wide attention in wastewater treatment due to their high surface area, tuneable surface chemistry, and strong adsorption capacity. In this context, green synthesis of nanoparticles offers additional advantages by avoiding hazardous chemicals and employing plant-derived bioactive compounds as natural reducing and stabilizing agents [1-3].

In the present study, Cu/TiO_2 nanoparticles were synthesized via a green route and evaluated as adsorbents for RhB removal from aqueous solution. The NPs were characterized using XRD, FESEM, BET and EDAX analysis, revealing uniform dispersion, an average particle size of approximately 20.5 nm, and a high surface area of 92 m^2/g. Additionally, 3D surface analysis indicated a moderate level of surface roughness (RMS~7.5 nm), which supports the presence of sufficient accessible sites for interaction with dye molecules. Batch adsorption experiments were carried out at initial RhB concentrations ranging from 1 to 10 mg/L with a contact time of 150 min. Equilibrium data were analyzed using three classical isotherm models, Langmuir, Freundlich, and Temkin, to assess adsorption behaviour through chemometric modelling and statistical fitting. The results showed that the Freundlich isotherm exhibited the strongest correlation with the experimental data (R^2 = 0.9995), indicating heterogeneous surface adsorption and multilayer coverage. The Langmuir model also provided a good fit (R^2 = 0.9946), suggesting partial monolayer formation, while the Temkin model showed weaker applicability (R^2 = 0.9273). The maximum adsorption capacity (q_{max}) was 19.12 mg/g, demonstrating the high uptake potential of the synthesized adsorbent. Furthermore, analysis of the Freundlich constant revealed that (n^{-1}) was 0.7713, which lies within the range of $0 < n^{-1} < 1$, confirming that the adsorption of RhB was normal and favorable.

This study illustrates the integration of green nanotechnology with chemometric analysis, emphasizing that adsorption isotherm modelling not only provides mechanistic insights into surface interactions but also enhances predictive capability for pollutant removal. Such data-driven evaluations are crucial for designing efficient, cost-effective, and environmentally sustainable wastewater treatment strategies.

Keywords: Cu/TiO_2 nanoparticles, Rhodamine B adsorption, Green synthesis, Adsorption isotherms

References:

[1] Sharma, J., et al., "Toxic effects of Rhodamine B on antioxidant system and photosynthesis of Hydrilla verticillata", Journal of Hazardous Materials Letters, 3, **(2022)**, p. 100069.
[2] Dutta, S., et al., "Contamination of textile dyes in aquatic environment: Adverse impacts on aquatic ecosystem and human health, and its management using bioremediation", Journal of Environmental Management, 353, **(2024)**, p. 120103.
[3] Sajjadi, S.M. and G. Hossinzadeh, "Textile dyes removing from the wastewater by green synthesized Cu-doped ZnO photocatalysts under the simulated sunlight illumination", Ceramics International, 50, **(2024)**, p. 36271-36285.

Multiple Linear Regression and Radial Basis Function Neural Networks for Identifying Effective Parameters and Ensuring Robustness in LC–MS/MS Simultaneous Analysis

Rabee Mahdavi[a], *Zahra Talebpour* [a,b]*, *Somaye Vali Zade*[c]

[a] *Department of Analytical Chemistry, Faculty of Chemistry, Alzahra University, Vanak, Tehran, Iran*

[b] *Analytical and Bioanalytical Research Centre, Alzahra University, Vanak, Tehran, Iran*
[c] *Halal Research Center of IRI, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran*
* E-mail: ztalebpour@alzahra.ac.ir

ABSTRACT
Robust analytical methods are essential for reliable performance in the presence of complex matrices and variable experimental conditions, a challenge that is particularly critical in the LC-MS/MS technique. The Analytical Quality by Design (AQbD) framework, combined with computational tools, was used to define the Method Operable Design Region (MODR) and ensure method robustness. In this study, two complementary approaches were employed: Multiple Linear Regression (MLR), providing interpretable polynomial models, and a Radial Basis Function (RBF) neural network, capturing complex nonlinear interactions [1, 2]. Both were applied to develop an LC–MS/MS method for the simultaneous analysis of several antiviral drugs. A Box–Behnken design with four factors (flow rate, formic acid content, cone voltage, and acetonitrile percentage) at three levels was employed to assess their main effects and their binary interactions on retention time, peak area, and resolution, and to determine the MODR. Chromatographic separation was performed on a Nucleodur C18 column (150 × 4.6 mm, 5 µm) with a gradient of water containing formic acid and acetonitrile. Detection was carried out using a Waters Quattro Micromass mass spectrometer with a positive ESI source operating in MRM mode. The predictive performance of both computational approaches was evaluated using R^2 and the extent of the MODR. MLR identified flow rate and cone voltage as the most influential factors, providing a simple and interpretable model. The RBF neural network, with higher R^2 values (0.91–0.99) for some responses, captured complex non-linear interactions and predicted a wider, more flexible operable region, better accommodating experimental variation while maintaining robustness. In both approaches, the lack of fit was not significant, confirming model adequacy.

Keywords: “AQbD”, “RBF”, “MLR”, “LC–MS/MS”, “Robustness”

References:
[1] Jiang, Q., et al. "An efficient multilayer RBF neural network and its application to regression problems." *Neural Computing and Applications,* 34.6, (**2022**), 4133-4150.
[2] Passarin, P.B.S., et al. "AQbD approach for the development of an analytical procedure for the separation of cephalosporin drugs and their degradation products." *Chemometrics and Intelligent Laboratory Systems,* 261 (**2025**), 105389.

CCD-Based Integrated Extraction–Encapsulation of β-Carotene from Carrot: A Chemometrics Approach humanized

*Zahra Sadri, Zahra Khorsandi, Nazila Khorrammaslack, Sima Najafi, Fariba Abdisaray, Morteza Bahram, Khalil Farhadi**
Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran.
E-mail address (corresponding author): khalil.farhadi@yahoo.com

ABSTRACT

To apply Central Composite Design (CCD) for efficiently exploring the multi-factor space governing the integrated extraction–encapsulation workflow of β-carotene from carrot and to achieve robust optimization of extraction and microencapsulation parameters within an intelligent laboratory framework [1]. A six-factor CCD examined extraction time, extraction temperature, solid-to-liquid ratio, encapsulation temperature, encapsulation wall-material percentage, and surfactant amount. The design encompassed 50 experimental runs, including center points, to enable curvature estimation. The β-Carotene yield after ultrasound-assisted extraction and encapsulation performance (e.g., encapsulation efficiency and stability) served as primary responses. Quadratic polynomial models were fitted and evaluated by ANOVA; model adequacy was assessed via R^2 (0.9211), adjusted R^2 (0.9630), predicted R^2 (0.9350), and lack-of-fit (not significant). Desirability-based multi-response optimization identified optimal settings. The CCD revealed significant main effects and first- and higher-order interactions among factors, enabling accurate prediction of extraction yield and encapsulation stability [2]. Optimized conditions achieved higher β-carotene recovery and improved stability relative to baseline runs; confirmatory experiments validated model predictions. The CCD-driven integrated extraction–encapsulation workflow, embedded within chemometric and intelligent laboratory systems, provides a scalable, data-driven framework for developing carrot-derived β-carotene products with superior antioxidant activity and nutritional quality for food and pharmaceutical applications [3].

Keywords: Central Composite Design, Encapsulation, β-carotene, Carrot, Surfactant, ANOVA, Chemometrics

References:

[1] M. Umair, S. Jabbar, …and L. Zhao, "Ultrasound-Assisted Extraction of Carotenoids from Carrot Pomace and Their Optimization through Response Surface Methodology", Molecules, 26, (**2021**), 6763.
[2] T. P. Vo, H. K. Linh Tran, …and D. Q. Nguyen, "Extraction and Emulsification of Carotenoids from Carrot Pomaces Using Oleic Acid", ACS Omega, 8, (**2023**), 39523-39534.
[3] E. Malagutti, S.G.B. Pinho, …and M. Martelli Tosi, "Optimizing Encapsulation of Active Compounds of Carrot By-Product in TPP-Chitosomes", Foods, 16, (**2024**), 2604.

A CCD-based two-stage DOE framework for the extraction and stabilization of lycopene from tomato pomace using ultrasound-assisted extraction and solid lipid encapsulation

*Fariba Abdi Saray, Katayoun Yekani, Sima Najafi, Nazila Khorram Maslack, Morteza Bahram, Khalil Farhadi**

Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

E-mail address (corresponding author): khalil.farhadi@yahoo.com

ABSTRACT

Lycopene is a powerful lipophilic antioxidant, but its stability, bioavailability, and solubility limit its practical applications[1]. This work presents a CCD-based, two-stage design of experiments framework within a chemometrics and intelligent laboratory systems (CLS) paradigm to optimize both extraction and stabilization of lycopene from tomato pomace. In stage one, ultrasound-assisted extraction (UAE)[2] with food-grade sunflower oil yields a lycopene-rich oleoresin. In stage two, the oleoresin is encapsulated into solid lipid microparticles (SLMs) via hot homogenization and ultrasonication, followed by a freeze-drying step to enhance stability. A central composite design simultaneously optimizes extraction (temperature, extraction time, solvent-to-solid ratio) and encapsulation (core–shell ratio, coating material, and drying conditions), with responses including lycopene yield, encapsulation efficiency, loading capacity, and lycopene content. The identity of the encapsulated lycopene was confirmed by HPLC-UV analyses. Predictive models show robust fits ($R^2 \approx .966$) and cross-validation metrics around 3% for RMSE and bias. Under optimized conditions, lycopene yield and encapsulation efficiency improved relative to conventional processes, with encapsulation efficiency reaching 99%, loading capacity around 30%, and lycopene content near 24 g per 100 g. The framework is readily scalable to pilot and industrial scales, offering a sustainable pathway to convert agri-food waste into stabilized lycopene powder for nutraceutical and food fortification applications. By integrating DOE, chemometrics, and automated workflows, this approach enables data-driven decision making and real-time optimization from upstream extraction to downstream stabilization.

Keywords: lycopene- encapsulation- solid lipid microparticles – freeze drying- ultrasound assisted extraction- sunflower oil

References:

[1] A. Kyriakoudi , K. Klimantakis , P. Kalaitzis , C. Biliaderis, I. Mourtzinos, “Enrichment of sunflower oil with tomato carotenoids and its encapsulation in Ca-alginate beads: Preparation, characterization and chemical stability upon in vitro digestion”, Food Hydrocolloids, 151, (2024), 19855- 19866

[2] J. Li, R. Campardelli,, G. Firpo, J. Zhang, P. Peregor, “Oil-in-water nanoemulsions loaded with lycopene extracts encapsulated by spray drying: Formulation, characterization and optimization”, Chemical Engineering, 70, (2024), 73- 81.

Explainable Machine Learning–Guided Optimization of µSPE-DLLµE for Phenol Extraction

Marziyeh Hosseinchi[1], *Khalil Farhadi*[*1], *Morteza Bahram*[1], *Sima Najafi Ghamat*[2]
[1]*Department of Chemistry Faculty of Science Urmia University Urmia Iran*
[2]*Department of Chemistry Faculty of Science of Alzahra University*
E-mails: khalil.farhadi@yahoo.com

ABSTRACT

An analytical strategy, µSPE-DLLµE (micro-solid phase extraction combined with dispersive liquid-liquid microextraction), was developed for the extraction and pre-concentration of phenol from aqueous matrices [1]. To capture the complex, non-linear interactions in µSPE-DLLµE, a suite of regression and machine learning models was evaluated, including Multiple Linear Regression (MLR), Ridge Regression, Random Forest (RF), Gradient Boosting (GB), XGBoost, and a Multi-Layer Perceptron (MLP) neural network [2]. Models were trained using 5-fold cross-validation and evaluated on an independent test set with R^2, RMSE, and MAE as performance metrics. The MLP model emerged as the most accurate predictor, achieving $R^2 = 0.967$ on 12 external samples and $R^2 = 0.983$ with RMSE = 0.014 on validation data. SHapley Additive exPlanations (SHAP) identified pH and adsorbent amount as the most influential parameters, consistent with the known chemistry of phenol adsorption. Partial dependence plots revealed clear nonlinear interactions, with adsorption efficiency plateauing beyond ~10 mg of adsorbent. Experimentally, the optimized µSPE-DLLµE achieved an enrichment factor of 435, recoveries of 92.4–98.7%, a detection limit of 0.5 µg L^{-1}, and a linear range of 1.7–120 µg L^{-1}, with relative standard deviations below 4.1%, confirming robustness and reproducibility.

This integrated, explainable ML–extraction framework offers a reliable and scalable platform for phenol monitoring in both environmental and food-safety applications.

Keywords: SHAP; MLP; µSPE-DLLµE, Phenol

References:

[1] M.A. Farajzadeh, M. Abbaspour, M.R.A. Mogaddam, H. Ghorbanpour, Determination of some synthetic phenolic antioxidants and bisphenol A in honey using dispersive liquid–liquid microextraction followed by gas chromatography-flame ionization detection, Food Analytical Methods. 8 (2015) 2035–2043.

[2] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature. 559 (2018) 547–555.

Deep Learning-Enabled Vis-NIR Hyperspectral Imaging for Fast and Accurate Detection, Classification and Quantification of Dimethoate Residues in Apples

*Parya Omrani, Hadi Parastar**
E-mail address: h.parastar@sharif.edu

ABSTRACT

Today, one of the critical challenges facing humanity is the extensive use of pesticides to increase crop yields, resulting in human exposure to these chemicals. The residual levels of pesticides on or within agricultural products pose a serious health risk for consumers, as their presence cannot be detected through smell or by the naked eye. To ensure high yields, a substantial number of pesticides is used in their production, making rapid detection of these chemicals crucial. Hyperspectral imaging (HSI) is one method that allows for the rapid detection of pesticides [1]. This technique is a fast, non-destructive method that enables simultaneous spatial and spectral data collection, which has led to its rapid adoption in analytical chemistry. HSI offers advantages for qualitative and quantitative analyses, including accessibility, low cost, reduced analysis time, no need for sample preparation and non-destructive testing. This analytical tool provides complex data which needs advanced data analysis tools to extract qualitative and quantitative information. However, traditional chemometric and machine learning (ML) approaches often require extensive preprocessing and struggle with high-dimensional data, limiting their ability to classify or quantify residues accurately. This study introduces a deep learning (DL) framework utilizing convolutional neural networks (CNNs) for the detection, classification, and quantification of dimethoate residues in apples via Vis-NIR HSI [2]. The methodology involved acquiring HSI data from apple samples and treated with varying dimethoate concentrations. The model development step involved extracting regions of interest (ROI) from the global HSI, followed by training a CNN model to automatically extract hierarchical spatial-spectral features from raw hypercubes, bypassing manual feature engineering. For classification, the CNN discriminated between different residue concentrations in the range of 20–500 ppm, achieving an accuracy exceeding 97%. Saliency maps highlighted critical regions corresponding to significant spectral features, providing interpretable insights into residue distribution and feature importance visualization. Additionally, a CNN-based regression model was applied to quantify residues in a lower concentration range of 20–80 ppm, yielding an R^2 value of 96%, 94% and 92% for the training, cross-validation and test sets, respectively. These results underscore the superiority of DL-enhanced HSI over classical methods of PLS and MCR, offering rapid, non-invasive analysis with high sensitivity and specificity. By enabling precise residue monitoring, this approach supports food safety regulations, reduces consumer exposure risks, and promotes sustainable agriculture.

Keywords: Hyperspectral imaging, Pesticides, Deep learning, Convolutional neural network, Chemometrics.

References:

[1] Zhang, Zi-Yu, et al. "How to Identify Pesticide Targets?" J. Agric. Food Chem., 73.3 (2025): 1790-1800.
[2] B. Debus, H. Parastar, P. Harrington, D. Kirsanov, “Deep learning in analytical chemistry”, *Trends Anal. Chem.,* 145, (2021), 116459.

Focused Molecular Docking Study of Minocycline and Doxycycline with Selected Cyclodextrins: Toward Efficient Host–Guest Complexation

*Melika asadian, Roya Kiani-anbouhi**
Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran.
kiani@sci.ikiu.ac.ir

Abstract
The uncontrolled release of antibiotics into aquatic environments has highlighted the urgent need for strategies that reduce their persistence and potential risks. While many studies have broadly compared the interaction of multiple antibiotics with different cyclodextrins (CDs), a focused approach can provide deeper mechanistic insights. In this work, we specifically examined the binding of doxycycline and minocycline with β- and γ-cyclodextrins, which were previously identified as the most effective hosts in tetracycline–CD systems. Molecular docking simulations revealed that γ-CD formed the most stable inclusion complexes, particularly with doxycycline, whereas β-CD showed moderate but consistent interactions with both antibiotics. The rationale for selecting these two antibiotics lies in their high clinical usage and structural variations, which influence host–guest affinity. This selective comparison not only confirms the superior binding of γ-CD but also emphasizes the importance of tailoring host molecules to specific antibiotics. Such a narrowed yet detailed analysis can guide the design of targeted removal methods in water treatment and potential pharmaceutical formulations.

Harnessing γ-Cyclodextrin for Enhanced Macrolide Antibiotics Delivery and Environmental Mitigation: A Computational Perspective

*Melika Asadian, Roya Kiani-anbouhi**

Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran.
kiani@sci.ikiu.ac.ir

Abstract
Macrolide antibiotics, such as azithromycin and erythromycin, are cornerstone agents in treating bacterial infections but often struggle with poor aqueous solubility and environmental persistence. γ-Cyclodextrin (γ-CD) offers a viable strategy to overcome these challenges, thanks to its large hydrophobic cavity, which enhances drug solubility, stability, and bioavailability. In this work, molecular docking was used to explore the encapsulation potential of γ-CD for selected macrolides. By focusing on compounds with the strongest and weakest binding affinities, we identified key structural features that influence γ-CD binding efficacy—most notably, hydrogen bonding networks and hydrophobic fit. Computational methods complemented by experimental frameworks have recently advanced the design of CD-based delivery systems. Furthermore, similar studies suggest γ-CD is particularly effective at accommodating bulky molecules due to its ample cavity size. The findings provide a computational foundation to guide the rational design of γ-CD-based formulations for improved drug delivery and environmental remediation.
Keywords: γ-Cyclodextrin, Computational Perspective, Molecular Docking, Macrolide Antibiotic.

Prediction of lipid/water distribution coefficient of the organic compounds using the Monte Carlo approach and hybrid optimal descriptors

Shahram Lotfi*
Department of Chemistry, Payame Noor University (PNU), 19395-4697 Tehran, Iran
Corresponding author e-mail: Sh.lotfi@pnu.ac.ir; Sh.lotfei@gmail.com

ABSTRACT

The lipid content in organisms mainly consists of two categories: storage lipids and membrane lipids. Storage lipids, primarily made up of triacylglycerides, are the key constituents of adipose (fat) tissue. Membrane lipids, on the other hand, are found within biological membranes and are largely composed of phospholipids. These two lipid types differ significantly in how they accumulate organic substances, due to variations in their molecular structures and the nature of their interactions with contaminants. In environmental toxicology, two important parameters are used to evaluate chemical behavior and pollutant distribution between organisms and their surroundings: (ii) The storage lipid–water partition coefficient (log $K_{lip/w}$), (ii) The phospholipid–water partition coefficient (log K_{pw}) [1, 2].

However, experimental measurements of lipid/water distribution coefficient of the organic compounds are complex, time-consuming, laborious and costly. So, development of computational methods (e.g., Quantitative Structure- Property Relationship (QSPR)) to predict the $K_{lip/w}$ or K_{pw} would be beneficial. In this study, QSPR modeling was performed for 305 organic compounds to predict the lipid/water partition coefficient (log $K_{lip/w}$) using the CORAL software based on the Monte Carlo algorithm. The molecular structures of the selected compounds were drawn using BIOVIA software. These structures were converted into SMILES format and defined as input files for the CORAL software.

The experimental values for a data set including 305 compounds with measured $K_{lip/w}$ were taken from the previous literatures [1]. These experimental values were converted into the logarithm scale and were used as a dependent variable for constructing current QSPR models.

The CORAL software creates QSPR model in form of one-variable correlation between the hybrid-based descriptor and log $K_{lip/w}$ according to the following mathematical function:

$$\log K_{lip/w} = C_0 + C_1 \times {}^{Hybrid}\mathrm{DCW}(T^*, N^*)$$

Therefore, here, the hybrid- based descriptor applied in developed QSAR model is obtained using the following equation:

$$^{\mathrm{Hybrid}}\mathrm{DCW}(T^*, N^*) = DCW_{SMILES}(T^*, N^*) + DCW_{HSG}(T^*, N^*) \quad (2)$$

The QSPR models for the log $K_{lip/w}$ achieved for the four random splits are the following:

Split 1 $\log K_{lip/w} = -3.385\ (\pm 0.031) + 0.148\ (\pm 0.001) * \mathrm{DCW}(7,35)$ (13)

Split 2 $\log K_{lip/w} = -3.110\ (\pm 0.019) + 0.134\ (\pm 0.001) * \mathrm{DCW}(7,35)$ (14)

Split 3 $\log K_{lip/w} = -2.842\ (\pm 0.031) + 0.118\ (\pm 0.001) * \mathrm{DCW}(7,35)$ (15)

Split 4 $\log K_{lip/w} = -2.470\ (\pm 0.0241) + 0.096\ (\pm 0.001) * \mathrm{DCW}(7,35)$ (16)

All developed models are robust and have good predictability. The statistical quality obtained for split 6 is better than other models, so it is chosen as an outstanding model ($R^2_{Validation} = 0.9645$, $\bar{r}^2_m = 0.8672$, $MAE = 0.2475$, $RMSE = 0.4474$, $CCC = 0.9620$, $IIC = 0.6493$, $CII = 0.9588$).

Keywords: QSPR, $K_{lip/w}$, Monte Carlo method, CORAL software, Hybrid descriptor

References

[1] M. Li, J. Li, Y. Lu, C. Han, X. Wei, G. Ma, H. Yu, Developing the QSPR model for predicting the storage lipid/water distribution coefficient of organic compounds, Frontiers of Environmental Science & Engineering 15(2) (**2021**) 24.
[2] M.I. Khawar, A. Mahmood, D. Nabi, Exploring the role of octanol-water partition coefficient and Henry's law constant in predicting the lipid-water partition coefficients of organic chemicals, Scientific Reports 12(1) (**2022**) 14936.

QSPR modeling of chromatography retention indices of plant essential oils using SMILES and graph invariants

Shahram Lotfi*
Department of Chemistry, Payame Noor University (PNU), 19395-4697 Tehran, Iran
Corresponding author e-mail: Sh.lotfi@pnu.ac.ir; Sh.lotfei@gmail.com

ABSTRACT

Essential oils of plant have been extensively used in many fields, particularly in the pharmaceutical, food, household, and cosmetic industries. The analysis of volatility and polarity of essential oil components is generally performed by capillary GC with polarity stationary phase. In chromatography, the identification of compounds is done by comparing obtained peaks with a reference standard. However, there are cases where reference standards are either unavailable or the chemical information of essential oil componentsis not documented in reference libraries. To overcome these limitations, theoretical methodologies are applied to estimate the retention indices (RIs) of new essential oil components. The retention behavior of essential oil components determined by GC and GC-MS is costly and time-consuming. Quantitative structure–retention relationship (QSRR) could provide an effective and convenient way to predict the retention data of the essential oil components [1].

This study aims to develop robust QSRR models to predict the RI of 505 frequently reported components of plant essential oils using CORAL's built-in Monte Carlo algorithm. The experimental RIs of components of plant essential oils was taken from the article published by Li-Tang Qin [2]. A hybrid optimal descriptor, combining information from SMILES and hydrogen-suppressed graph (HSG) representations, was employed in model construction. This hybrid descriptor is calculated as the sum of correlation weights (CWs) derived from both SMILES notation and the molecular graph, as shown in Equation (1).

$$^{\text{Hybrid}}\text{DCW}(\text{T}^*,\text{N}^*) = {}^{SMILES}DCW(\text{T}^*,\text{N}^*) + {}^{HSG}DCW(\text{T}^*,\text{N}^*) \quad (1)$$

$$^{\text{SMILES}}\text{DCW}\,(\text{T}^*,\text{N}^*) = \sum \text{CW}(\text{S}_\text{K}) + \sum \text{CW}(\text{SS}_\text{K}) + \sum \text{CW}(\text{SSS}_\text{K}) + \sum \text{CW}(\text{NOSP}) + \sum \text{CW}(\text{HALO}) + \sum \text{CW}(\text{HARD}) + \sum CW(FLS) + \sum CW(APP) \quad (2)$$

$$^{HSG}DCW\ (\text{T}^*,\ \text{N}^*) = \sum \text{CW}(EC0_k) + \sum \text{CW}(EC1_k) + \sum \text{CW}(EC2_k) + \sum \text{CW}(EC3_k) + + \sum \text{CW}(pt2_k) + \sum CW(pt3_k) + \sum CW(pt4_k) + \sum CW(VS2_k) + \sum CW(VS3_k) + \sum \text{CW}(nn_k) + CW(C5) + CW(C6) \quad (3)$$

Then the calculated CWs are employed to design a predictive model of the RI according to the following equation;

$$LogRI = C_0 + C_1 \times DCW(T^*,N^*) \quad (4)$$

The QSRR models obtained by the Monte Carlo optimization for the five splits formulated with TF2 of endpoint is the represented by the following relationship:

Split 1 pIC50 = 4.4002 (0.0047) + 0.0396 (0.0001) * DCW(3,35) (5)
Split 2 logRI = 3.2924 (0.0058) + 0.0431 (0.0001) * DCW(3,35) (6)
Split 3 logRI = 3.5013 (0.0049) +0.0323 (0.0001) * DCW(3,35) (7)
Split 4 logRI = 2.7641 (0.0112) + 0.0517 (0.0001) * DCW(3,35) (8)
Split 5 logRI = 3.8564 (0.0038) +0.0465 (0.0001) * DCW(3,35) (9)

The top-performing model, obtained from split 5, exhibited superior validation metrics (R^2 = 0.9420, IIC = 0.7976, CII = 0.9221, CCC = 0.9245, Q^2 = 0.8432, RMSE=0.6523, MAE=0.5321). Furthermore, key molecular fragments contributing to increases and decreases in retention time

were identified from the best model.

Keywords: QSRR, Monte Carlo method, CORAL software, Retention indices, Plant essential oils

References

[1] S. Ahmadi, S. Lotfi, H. Hamzehali, P. Kumar, A simple and reliable QSPR model for prediction of chromatography retention indices of volatile organic compounds in peppers, RSC advances 14(5) (2024) 3186-3201.

[2] L.T. Qin, S.S. Liu, F. Chen, Q.S. Wu, Development of validated quantitative structure–retention relationship models for retention indices of plant essential oils, Journal of separation science 36(9-10) (2013) 1553-1560.

Monte Carlo algorithm in the prediction of Value of Binding Constants of Organic Ligands to Beta-Cyclodextrin

Shahram Lotfi*
Department of Chemistry, Payame Noor University (PNU), 19395-4697 Tehran, Iran
Corresponding author e-mail: Sh.lotfi@pnu.ac.ir; Sh.lotfei@gmail.com

ABSTRACT

Binding of organic chemicals to macromolecules is of high relevance in environmental science and related fields, including pharmacy, agriculture, the food industry and explosives. Cyclodextrins (CDs) have been one of the most extensively studied complexation agents in various fields, especially as pharmaceutical excipients. The binding constants could be determined via some experimental methods, such as nuclear magnetic resonance and phase solubility technique. But the experimental methods usually take much time and effort owing to the low aqueous solubility of guest molecules. Cyclodextrin binding constant modeling deserves special attention due to its practical importance. Quantitative structure-activity relationship (QSPR) methodology has been extensively used for evaluating the formation abilities of molecular complexes, and characterizing their properties [1].

In this study, the dataset of 229 compounds used for QSPR modelling was extracted from previous literatures [2]. This dataset comprises binding constants (ln K) of 1:1 β-CD complexes with different organic compounds. In the current QSAR modelling, a hybrid optimal descriptor was employed and it was calculated with correlation weights of molecular attributes extracted from SMILES together with correlation weights of molecular attributes extracted from the hydrogen suppressed graph (HSG). The CORAL software creates QSPR model in form of one-variable correlation between the hybrid-based descriptor and lnK according to the following mathematical function:

$$\ln K = C_0 + C_1 \times {}^{Hybrid}\mathrm{DCW}(T^*, N^*) \quad (1)$$

Here, the hybrid- based descriptor applied in developed QSPR model is obtained using the following equation:

$${}^{\mathrm{Hybrid}}\mathrm{DCW}(T^*, N^*) = DCW_{SMILES}(T^*, N^*) + DCW_{HSG}(T^*, N^*) \quad (2)$$

Were the first part of the hybrid descriptor computed as the following relationship:

$${}^{\mathrm{SMILES}}\mathrm{DCW}(T^*, N^*) = \sum \mathrm{CW}(S_K) + \sum \mathrm{CW}(SS_K) + \sum \mathrm{CW}(\mathrm{NOSP}) + \sum \mathrm{CW}(\mathrm{HALO}) + \sum \mathrm{CW}(\mathrm{BOND}) + \sum \mathrm{CW}(\mathrm{HARD}) + \sum CW(FLS) + \sum CW(APP) \quad (3)$$

$${}^{HSG}DCW(T^*, N^*) = \sum \mathrm{CW}(EC0_k) + \sum \mathrm{CW}(EC1_k) + \sum \mathrm{CW}(pt2_k) + \sum \mathrm{CW}(nn_k) \quad (4)$$

QSPR models were developed for four randomly generated splits of the available dataset into active training, passive training, calibration, and validation subsets as detailed below:

Split 1 lnK = 1.8276 (0.0439704) + 0.0295379 (0.0004063) * DCW(7,15) (10)
Split 2 lnK = 2.1452 (0.0325354) + 0.0303548 (0.0002633) * DCW(7,15) (11)
Split 3 lnK = 2.1891 (0.0353161) +0.0373232 (0.0003673) * DCW(7,15) (12)
Split 4 lnK = 1.9241 (0.0291274) + 0.0354516 (0.0003041) * DCW(7,15) (13)

Results demonstrated that the models were robust, simple, and predictive, with validation metrics including R^2, Q^2, IIC, and CII ranging from 0.8923 to 0.9484, 0.8155 to 0.8832, 0.7542 to 0.7865, and 0.7963 to 0.8134, respectively.

Keywords: QSPR, Monte Carlo algorithm, CORAL software, Binding Constants, β-Cyclodextrin

References

[1] P. Di, J. Chen, L. Liu, W. Li, Y. Tang, G. Liu, In silico prediction of binding capacity and interaction forces of organic compounds with α-and β-cyclodextrins, Journal of Molecular Liquids 302 (2020) 112585.

[2] P. Cysewski, M. Przybyłek, Predicting value of binding constants of organic ligands to beta-cyclodextrin: Application of MARSplines and descriptors encoded in SMILES string, Symmetry 11(7) (2019) 922.

Simultaneous analysis of four cardiovascular drugs in synthetic mixtures and polypill tablets using triple devisor mean centering and derivative spectrophotometric techniques

Maryam Maghfourian[1], *Vahid Kiarostami*[1]*, *Farzad Kobarfard*[2], *Mehran Davallo*[1]
[1]*Department of Chemistry, North Tehran Branch, Islamic Azad University, Tehran, Iran*
[2] *Department of Medicinal Chemistry, School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
v_kiarostami@iau.ac.ir

ABSTRACT
Cardiovascular diseases, along with cancer and infectious diseases, are among the leading causes of death worldwide. Hypertension is the principal modifiable driver of cardiovascular disease worldwide, and fixed-dose polypills combining drugs have emerged as an evidence-based strategy to simplify combination therapy and improve long-term adherence. Accurate simultaneous determination of each active ingredient in commercial polypill tablets is therefore critical for dose uniformity, regulatory compliance and patient safety, motivating the continued development of robust analytical approaches (1). UV–Vis spectrophotometry is a low-cost, rapid technique well suited to the quantification of a single absorbing pharmaceutical ingredient when its absorption band is free from interference by co-formulated components. However, in multi-component systems that contain several chromophores, severe spectral overlap can preclude direct application of simple UV–Vis measurements. Chemometrics techniques have proven effective in overcoming the limitations of conventional UV–Vis spectrophotometry for overlapping spectra. In particular, the ratio derivative spectrometry and mean centering of ratio spectra is widely used to enhance resolution of overlapping absorbance bands (2, 3). Accordingly, in the present study we applied mean-centering with a triple-divisor strategy and derivative spectrophotometry as chemometrics techniques for the first time for the simultaneous quantitative analysis of a quaternary mixture of enalapril, hydrochlorothiazide, aspirin and atorvastatin in laboratory prepared mixtures and in a commercial Polypill E tablet (Alborz Daru Co.). Compared with derivative UV–Vis spectroscopy, the triple-divisor mean-centering strategy yielded more satisfactory recoveries in both synthetic mixtures and commercial polypill formulations (80–105%), and exhibited superior analytical performance. Furthermore, the application of triple-divisor mean-centering of ratio spectra ensured that the figures of merit for each analyte across multiple wavelengths consistently remained within the acceptable range.

Keywords: Mean centering, Triple devisor, derivative spectroscopy, Hydrochlorothiazide, Aspirin, Enalapril, Atorvastatin.

References:

[1] M. Stolarczyk, A. Apola, A. Maślanka, A. Kwiecień and W. Opoka, “Quantitative and qualitative analysis of fixed-dose-combination products (FDC, polypill) applied in polytherapy of hypertensive disease”, Acta Pol. Pharm.75 (**2018**), 1083-1094.
[2] N.W. Alia, N.S.Abdelwahab, M. Abdelkawy, and A.A. Emam, “A comparative study of ICH validated novel spectrophotometric techniques for resolving completely overlapping spectra of quaternary mixtures”, Spectrochim Acta A Mol Biomol Spectrosc, 154 (2016), 114-122.
[3] L.H. Nurani, C.A. Edityaningrum, I. Irnawati, A.R. Putri, A. Windarsih, A. Guntarti, and A. Rohman, "Chemometrics-assisted UV-vis spectrophotometry for quality control of pharmaceuticals: A review”, Indones. J. Chem., 23 (2023), 542-567.

SPA-PL-MARS QSAR Modeling of Xanthenone Derivatives as HCT-15 Colorectal Cancer Inhibitors: An Integrated Variable Selection and Nonparametric Regression Approach

Maryam Allahverdi[a], Mohammadreza Khanmohammadi Khorrami[a], Mohaddeseh Habibi[a], Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran*

*m.khanmohammadi@sci.ikiu.ac.ir

ABSTRACT

In this study, we used a dataset of 46 C1-O-substituted-3-(3-butylamino-2-hydroxy-propoxy)-xanthen-9-one derivatives for QSAR modeling [1]. PaDEL-Descriptor software generated a total of 1,444 molecular descriptors, which were then refined through a two-step preprocessing pipeline[2]. The first step involved removing near-constant descriptors (those with more than 80% constant values), followed by the exclusion of highly collinear descriptors (with a Pearson correlation greater than 0.95 or VIF greater than 5). This process resulted in 392 optimized descriptors. To ensure representative data partitioning, the SPXY method was used, which takes both predictor and response variable spaces into account [3]. For QSAR modeling, we compared two approaches: Multiple Linear Regression (MLR) and Piecewise-Linear Multivariate Adaptive Regression Splines (PL-MARS). While MLR faced challenges due to intercorrelated topological descriptors, the PL-MARS model proved advantageous, offering better predictive power. By integrating the Successive Projections Algorithm (SPA) for optimal variable selection and PL-MARS for nonlinear modeling, we achieved high predictive accuracy. The PL-MARS model resulted in an R^2 value of 0.970 and an adjusted R^2 of 0.966, with low error metrics (SSE = 0.036 and MSE = 0.004), showing its robustness and reliability. Additionally, the Applicability Domain (AD) of the model was evaluated using a Williams plot. This diagnostic tool helped identify response outliers and influential compounds in the training set. Specifically, three influential compounds and one response outlier were identified, suggesting that these compounds, due to their structural uniqueness, should be excluded when designing new therapeutic agents using this model. No such issues were detected in the test set. Overall, the SPA-PL-MARS approach proved to be a powerful and interpretable framework for designing and optimizing xanthenone-based anticancer agents. The findings highlight the potential of this approach to develop highly predictive QSAR models and contribute to advancing cancer therapeutics.

Keywords QSAR, SPA, MLR, PL-MARS, XY-jointed algorithm, Cancer Inhibitors

References

[1] Park S, Hong E, Kwak SY, Jun K-Y, Lee E-S, Kwon Y, et al. Synthesis and biological evaluation of C1-O-substituted-3-(3-butylamino-2-hydroxy-propoxy)-xanthen-9-one as topoisomerase IIα catalytic inhibitors. European Journal of Medicinal Chemistry. 2016;123:211-25.

[2] Yap CW. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of computational chemistry. 2011;32(7):1466-74.

[3] He Z, Li M, Ma Z. Design of a reference value-based sample-selection method and evaluation of its prediction capability. Chemometrics and Intelligent Laboratory Systems. 2015;148:72-6.

Classification of Extra Virgin Olive Oil Adulteration Using FTIR Spectroscopy and Machine Learning Algorithms

Esra modares Askari[a], Mohammadreza Khanmohammadi Khorrami[a], Mahsa Mohammadi[a], Miguel de la Guardia[b]*

[a]*Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran*

[b]*Departamento de Quı´mica Analı tica, Research Building, Universitat de Vale`ncia, 50 Dr. Moliner St., 46100 Burjassot, Vale`ncia, Spain*

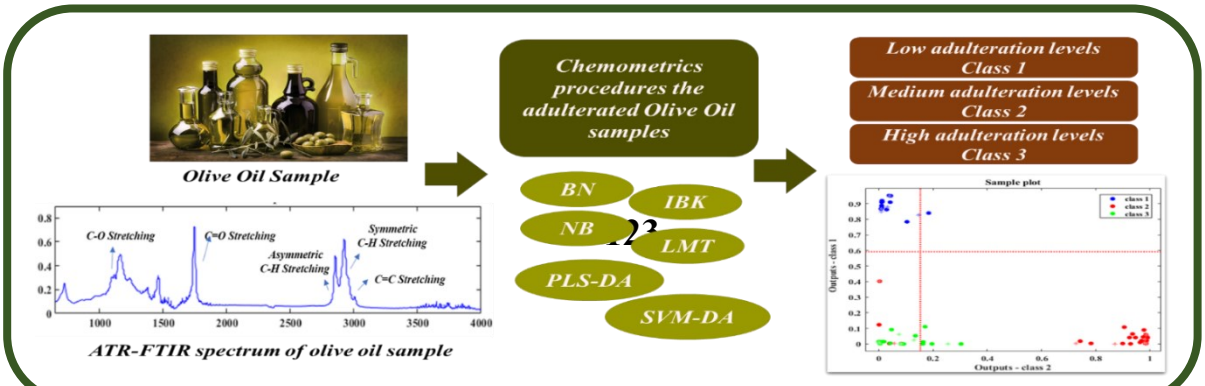
*Corresponding author: E-mail addresses: m.khanmohammadi@SCI.ikiu.ac.ir (M.Khanmohammadi. Khorrami)

ABSTRACT

The authentication of Extra Virgin Olive Oil (EVOO) is critical to ensuring quality standards and consumer trust [1]. Given the global rise in food adulteration, especially in olive oil, identifying fraud and implementing stricter regulations has become a necessity in the field of food safety [2]. The results of studies have shown that combining chemical and statistical methods significantly increases the accuracy of adulteration detection [3]. Chemometric models, which integrate machine learning and spectroscopy, offer a viable solution. In this study, chemometric methods for classification were applied to Fourier transform infrared (FTIR) spectral data from adulterated EVOO samples. Samples were classified into three purity classes: low (44.75–59% EVOO), medium (63.75–73.25% EVOO), and high (78–97% EVOO). Thus, a total of 75 samples were prepared by blending EVOO with escalating percentages (1%–35% w/w) of the three lower-cost oils. Adulterated mixtures were vigorously stirred to ensure homogeneity before analysis. All of the 75 samples were split into a training set (53 samples, 70%) for model training and a prediction set (22 samples, 30%) for validation. The curse of dimensionality poses challenges in high-dimensional spectral datasets. Robust PCA (rPCA) addresses the challenge of dimensionality in high-dimensional spectral datasets by decomposing raw data into two matrices: a low-rank matrix representing the underlying structure and a sparse matrix capturing outliers or noise. This study evaluates seven machine learning algorithms for detecting EVOO adulteration with lower-cost vegetable oils using attenuated total reflectance Fourier transform infrared ATR-FTIR spectroscopic data. We compare traditional chemometric methods with advanced machine learning models, finding that the Instance-Based K-nearest neighbor (IBK) algorithm outperformed other classifiers, achieving 98% accuracy and a mean absolute error (MAE) of 0.008. Notably, IBK surpassed both Partial Least Squares Discriminant Analysis (PLS-DA) and Support Vector Machine Discriminant Analysis (SVM-DA), establishing a robust framework for rapid, accurate EVOO quality control.

Keywords

Adulteration, Classification, EVOO, BN, IBK, LMT, MLP, NB, PLS-DA, SVM-DA



References

[1] International Olive Council, "Trade standard applying to olive oils and olive-pomace oils (COI/T.15/NC No 3/Rev. 16)", **2021**.

[2] P. Zunin, R. Boggia, S. Lanteri, et al., "Authentication of extra virgin olive oils by analysis of their minor components", Food Chemistry, 188, (**2015**), 17–23.

[3] G. Downey, P. McIntyre, and A. N. Davies, "Detecting and quantifying sunflower oil adulteration in extra virgin olive oils from the Eastern Mediterranean by visible and near-infrared spectroscopy", Journal of Agricultural and Food Chemistry, 51, (**2003**), 5730–5735.

Alternative Conditional Expectation Method for Modeling the QSAR (IC50) of Oxazolidinone Derivatives as 17β-Hydroxysteroid Dehydrogenase Type 3 Inhibitors

Mohaddeseh Habibi[a], *Mohammadreza Khanmohammadi Khorrami**[a],
Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran
*m.khanmohammadi@sci.ikiu.ac.ir

ABSTRACT

This study investigates the quantitative structure-activity relationship (QSAR) of oxazolidinone and thiazolidinedione scaffolds as inhibitors of 17-β-hydroxysteroid dehydrogenase type 3 (*17-β-HSD3*) to accurately predict their pIC_{50} values.[1] Dimensionality reduction was performed using robust Principal Component Analysis (rPCA), and the duplex method partitioned the dataset into calibration and test sets. Predictive models were developed using ridge regression (RR), principal component regression (PCR), and alternating conditional expectations (ACE).[2, 3] While RR and PCR effectively addressed multicollinearity through linear modeling, they were limited in capturing nonlinear relationships. On the test set, RR and PCR yielded adjusted R^2 values of 0.479 and 0.482 and RMSEs of 0.366 and 0.312, respectively. In contrast, the ACE method optimized transformations of both response and predictor variables, successfully modeling nonlinear dependencies. ACE significantly outperformed the linear methods, achieving an adjusted R^2 of 0.972 and an RMSE of 0.099. These results demonstrate the superiority of nonlinear regression techniques for enhancing QSAR model accuracy. Using QSAR modeling, we conducted a systematic analysis of oxazolidinone and thiazolidinedione derivatives to predict pIC_{50} values. Among the various techniques tested, the nonlinear ACE method—especially the innovative Spline-ACE regression—demonstrated a substantial improvement over traditional linear approaches (such as Polynomial-ACE, RR, and PCR), offering significantly better predictive accuracy by effectively addressing complex nonlinear relationships that are crucial for creating reliable QSAR models.

Keywords: "Ridge regression" "principal component regression" "Alternative Conditional Expectation" "17-β-HSD3 inhibitors"

References:

1. **Koichiro Harada, H.K., Jun Abe, Mari Haneta, Arnel Conception, Shinichi Inoue, Satoshi Okada, Kazuhiko Nishioka, *Discovery of potent and orally bioavailable 17b-hydroxysteroid dehydrogenase type 3 inhibitors.* Bioorganic & Medicinal Chemistry, 2012. 20.**
2. **Breiman, L., & Friedman, J. H., *Estimating Optimal Transformations for Multiple Regression and Correlation.* . Journal of the American Statistical Association, 1985. 80.**
3. **Dara, S., et al., *Machine Learning in Drug Discovery: A Review.* Artificial Intelligence Review, 2022. 55(3): p. 1947-1999.**

Comparison of Regression Methods (ACE, SVR, PLS, MLR) in QSAR, a Kinesin-5 Inhibitor Case Study

*Mohammadreza Pourmohammad, Mohammadreza Khanmohammadi Khorrami**
Department of Chemistry, Faculty of Science, Imam Khomeini International University, Qazvin, Iran
E-mail address: m.khanmohammadi@ikiu.ac.ir

Abstract
In recent years, the use of Computational based drug design, especially QSAR methods has become increasingly widespread. Using a suitable regression method is a crucial part of building a QSAR model to showcase the relationship between the activity and the molecular structure and various regression methods can be employed to discover this relationship. By developing a reliable regression model, it is possible to predict the activity of new molecules or, with the help of effective molecular descriptors, design and synthesize entirely new molecules with improved activity [1]. The objective of this study if to compare the performance of four regression methods including three well-known regression methods: Multiple Linear Regression (MLR), Partial Least Squares regression (PLS), and Support Vector Regression (SVR). Additionally, Alternating Conditional Expectations (ACE) regression, a novel regression method in chemistry and QSAR studies, was evaluated as the fourth regression approach [2]. This case study was done on a dataset of 52 molecules as kinesin-5 protein inhibitor and potential anticancer agents [3]. After optimizing the molecular structures using the AMBER force field, a total of 1819 molecular descriptors of 1,2 and 3-Dimension were calculated using PaDEL-Descriptor and MOE software packages. Additionally, Robust Principal Component Analysis (RPCA) was done in data preprocessing to reduce dimensionality and prepare 4 new independent variables for regression. Data splitting was performed using the SPXY algorithm to ensure an even distribution of data point in the training and test sets [4]. The R^2 values for the ACE model were 0.990 and 0.989 for the training and test sets, respectively. The ACE algorithm achieves this high Correlation by transforming variables. In comparison, these values were 0.506 and 0.678 for the MLR, 0.536 and 0.624 for the PLS, and 0.857 and 0.691 for the SVR, indicating the superior performance of the ACE method. Other statistical criteria, such as SSE and RMSE, were also reported and the reliability of these criteria was assessed using Sum of Ranking Differences (SRD) analysis [5], Suggesting Akaike and Bayesian Information Criterion (AIC, BIC) as reliable criteria for model evaluation. Finally, these criteria were used for model comparison and the results were consistent with those from R^2 and R_a^2 values, confirming the superiority of the ACE method over the other regression methods employed here.

Keywords: QSAR, RPCA, ACE, SVR, PLS, SRD

References:
[1] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, "QSAR modeling: where have you been? Where are you going to?", Journal of medicinal chemistry, 57(12), (**2014**), 4977-5010.

[2] L. Breiman, J.H. Friedman, "Estimating optimal transformations for multiple regression and correlation.", Journal of the American statistical Association, 80(391), (**1985**), 580-598.
[3] K. Schiemann, D. Finsinger, F. Zenke, C. Amendt, T. Knöchel, D. Bruge, H.P. Buchstaller, U. Emde, W. Stähle, S. Anzali, "The discovery and optimization of hexahydro-2H-pyrano [3, 2-c] quinolines (HHPQs) as potent and selective inhibitors of the mitotic kinesin-5.", Bioorganic & medicinal chemistry letters, 20(5), (**2010**), 1491-1495.
[4] R.K.H. Galvao, M.C.U Araujo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, "A method for calibration and validation subset partitioning.", Talanta, 67(4), (**2005**), 736-740.
[5] K. Héberger, "Sum of ranking differences compares methods or models fairly.", TrAC Trends in Analytical Chemistry, 29(1), (**2010**), 101-109.

QSAR Modeling and Prediction of the Hallucinogenic Potency of Amphetamine Derivatives Using a Docking-Guided QSAR Approach

*Zahra Imanikia *, Mohammad Hossein Fatemi*
Chemometrics Laboratory, Faculty of Chemistry, University of Mazandaran,Babolsar,Iran
E-mail address: Matinimanikia75@Gmail.com

ABSTRACT

In the present study, the hallucinogenic potency of amphetamine derivatives was modeled and predicted using a quantitative structure–activity relationship (QSAR) approach guided by molecular docking. The dataset comprised 17 amphetamine-related compounds in which; their hallucianogenic potencies were expressed relative to mescaline (mescaline units, MU) on a logarithmic scale [1]. Since the most chemical compounds are distributed in human body and transported in the bloodstream as complexes with human serum albumin (HSA), therefore molecular docking studies were performed to investigate and evaluate the binding of the interested amphetamine derivatives to HSA. After docking the most stable bound conformer for each compound was frizzed for molecular descriptor calculation [2].

These structures were used to compute more than 1800 molecular descriptors per compound with PaDEL and Dragon software. Descriptor screening was then carried out to remove descriptors with low variance and high intercorrelation. Important descriptors were selected using stepwise linear selection combined with a breakpoint method. The results indicated that descriptors of XLogP3, MATS6p, Mor09u, topological polar surface area (TPSA), and HTm are highly relevant to the hallucinogenic potency of the studied compounds. These descriptors were employed as independent variables, with hallucinogenic potency as the dependent variable, to develop a linear QSAR model.

The resulting model showed excellent statistical performance: for the training set, SE = 0.06529 and R^2 = 0.992; for the prediction set, SE = 0.12558 and R^2= 0.876. Also the obtained model is evaluated by Cross-Validation and chance-correlation method. These results demonstrate the capability of a docking-guided QSAR strategy to model and predict the hallucinogenic potency of amphetamine derivatives in applicability domain of model.

Keywords: Molecular docking, Amphetamine, Human serum albumin, QSAR modeling, Log mescaline unit , hallucinogenic

References:

[1] Abdou, M. M. (2001). A molecular approach to the study of structure-activity correlation for some amphetamines. *Journal of psychoactive drugs*, *33*(3), 295-300.

[2] Mohammadnia, F., Fatemi, M. H., & Taghizadeh, S. M. (2020). Study on the interaction of anti-inflammatory drugs with human serum albumin using molecular docking, quantitative structure–activity relationship, and fluorescence spectroscopy. Luminescence, 35(2), 266-273.

Predicting the Viscosity of Magnesium Oxide-Based Nanofluids in Deep Eutectic Solvents Using Quantitative Structure-Property Relationship Modeling

Mohammad Hossein Hajian, Mohammad Hossein Fatemi*
Laboratory of Chemometrics, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran
E-mail address: mohammadhosseinhajian8@gmail.com

ABSTRACT

The primary objective of this research is to model the viscosity of nanofluids using a Quantitative Structure-Property Relationship (QSPR) approach. In the current era, global sustainability initiatives are intensely focused on enhancing energy efficiency across various sectors, including advanced cooling technologies such as those used in vehicle radiators and air conditioning systems. Among these technologies, nanofluids, as a novel achievement of nanotechnology, have attracted significant attention. Nanofluids are colloidal suspensions in which nanometer-sized particles (including nanoparticles, nanotubes, and nanorods) with dimensions of 1 to 100 nm are dispersed in a base fluid such as water, ethylene glycol, ionic liquids, or deep eutectic solvents[1-4]. The dispersion of nanoparticles within the base fluid results in a significant improvement in thermophysical properties. Possessing distinctive features such as high viscosity and thermal conductivity, nanofluids have the potential to replace conventional fluids in heat transfer systems. To predict the (in logarithmic scale) viscosity of 144 magnesium oxide-based nanofluids in deep eutectic solvents prepared experimentally in the laboratory, the QSPR method was employed using the Monte Carlo algorithm within the CORAL software. In this regard, four separate random splits of the dataset were defined, and the quasi-SMILES notation was used to represent the features of the nanofluids. Optimal descriptors based on quasi-SMILES were calculated, and the influence of key variables such as temperature, nanoparticle mass fraction, and base fluid type on viscosity was investigated. A rigorous evaluation of the developed model confirmed its satisfactory statistical quality (best statistical parameters for viscosity: $R^2 = 0.9620$, $Q^2 = 0.9469$, RMSE = 0.1298).In the subsequent phase of this project, QSPR models based on the linear Multiple Linear Regression (MLR) and non-linear Support Vector Machine (SVM) and Artificial Neural Network (ANN) methods were developed to predict the logarithm of viscosity for the same compounds using STATISTICA software. Four independent variables, including temperature, nanoparticle mass fraction, base fluid type, and the optimal descriptor (DCW), were considered, with the logarithm of viscosity as the dependent variable. A comparison of the statistical parameters for the training and test sets demonstrated that the ANN model, with a coefficient of determination ($R^2 = 0.9954$) and a root mean square error (RMSE = 0.05855) for the test set, outperformed the SVM and MLR models, exhibiting superior predictive accuracy and precision.

Keywords: Nanofluids, thermophysical properties, nano-QSPR, CORAL software, deep eutectic solvent
References

[1] Asadi A, Aberoumand S, Moradikazerouni A, Pourfattah F, Żyła G, Estellé P, et al. Recent advances in preparation methods and thermophysical properties of oil-based nanofluids: A state-of-the-art review. Powder technology. 2019;352:209-26.
[2] Mahian O, Kolsi L, Amani M, Estellé P, Ahmadi G, Kleinstreuer C, et al. Recent advances in modeling and simulation of nanofluid flows-Part I: Fundamentals and theory. Physics reports. 2019;790:1-48.
[3] Kasaeian A, Daneshazarian R, Mahian O, Kolsi L, Chamkha AJ, Wongwises S, et al. Nanofluid flow and heat transfer in porous media: a review of the latest developments. International Journal of Heat and Mass Transfer. 2017;107:778-91.
[4] Amani M, Amani P, Mahian O, Estellé P. Multi-objective optimization of thermophysical properties of eco-friendly organic nanofluids. Journal of Cleaner Production. 2017;166:350-9.

Discrimination of Iranian black teas through volatile compound fingerprints and chemometric modeling

Adineh Aminianfar [1,2], *Mohammad Hossein Fatemi* [1]*, *Fatemeh Azimi* [1]

[1]*Department of Analytical Chemistry, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran*

[2] *Gilan Province Water and Wastewater Quality Monitoring and Supervision Center, National Water and Wastewater Engineering Company (NWWEC), Ministry of Energy, Iran*

E-mail address (corresponding author): mhfatemi@umz.ac.ir

ABSTRACT

Tea, as the second most widely consumed beverage in the world after water, has attracted considerable attention due to its unique aroma and complex chemical composition [1, 2]. Despite widespread consumption in Iran, the chemical profile of Iranian tea has remained largely unexplored. Volatile compounds play a decisive role in the qualitative and aromatic characteristics of tea, and identifying their patterns, particularly in relation to factors such as geographical origin and cultivation altitude, is important for tea quality control, authentication, and flavor characterization [3, 4]. In this study, chromatographic fingerprints of black teas from northern Iran were obtained using HS-SPME-GC-MS. The resulting data, after normalization and autoscaling, were subjected to exploratory methods such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) to investigate hidden structures and clustering. Subsequently, to evaluate the discriminative capacity of samples, supervised classification models including partial least squares–discriminant analysis (PLS-DA), linear discriminant analysis (LDA), and support vector machines (SVM) were employed. Moreover, the variable importance in projection (VIP) index was used for variable selection, so that selected compounds could reveal regional and altitudinal variations. Results showed that compounds, including methyl dodecanoate, heptadecane, nonadecane, 1-heptadecanamine, and 1-iododotriacontane discriminated altitude groups, while compounds such as 5-methylundecane, morpholine-4-octadecyl, caffeine, decane, and nonadecane differentiated samples based on geographical regions. These compounds, as potential chemical markers, possess a strong capacity for application in quality control, product authentication, and indices linked to aroma and flavor. In altitude-based classification, the SVM algorithm achieved stable performance with 100% test accuracy and only one training error, outperforming LDA, which, despite its perfect accuracy in training (100%), exhibited a substantial drop in test accuracy (66.7%). Conversely, in region-based classification, the LDA model demonstrated the best performance (100% test accuracy), showing that the chemical data at the regional level aligned closely with this method, whereas SVM performed slightly lower (83.3% test accuracy). Overall, the findings indicate that volatile compound patterns in tea are influenced by both geographical region and cultivation altitude, and these differences can be meaningfully predicted using supervised models. Furthermore, a limited set of compounds with the highest VIP scores was sufficient to capture the main discriminations, highlighting their key role in tea quality control and authenticity assessment.

Keywords: Chemometrics, Iranian black tea, Headspace Solid-Phase Microextraction–Gas Chromatography–Mass Spectrometry, Supervised and Unsupervised Analysis

References:

[1] X. Qin, J. Zhou, C. He, L. Qiu, D. Zhang, Z. Yu, Y. Wang, D. Ni, Y. Chen, "Non-targeted metabolomics characterization of flavor formation of Lichuan black tea processed from different cultivars in Enshi", Food Chem. X. 19 (2023) 100809.

[2] Q. Wang, D. Qin, G. Huang, X. Jiang, K. Fang, Q. Wang, E. Ni, B. Li, C. Pan, H. Li, others, "Identification and characterization of the key volatile flavor compounds in black teas from distinct regions worldwide", J. Food Sci. 87 (2022) 3433–3446.

[3] S. Fang, J. Ning, W.-J. Huang, G. Zhang, W.-W. Deng, Z. Zhang, "Identification of geographical origin of Keemun black tea based on its volatile composition coupled with multivariate statistical analyses", J. Sci. Food Agric. 99 (2019) 4344–4352.

[4] M. Wang, J. Li, X. Liu, C. Liu, J. Qian, J. Yang, X. Zhou, Y. Jia, J. Tang, L. Zeng, "Characterization of key odorants in Lingtou Dancong oolong tea and their differences induced by environmental conditions from different altitudes", Metabolites. 12 (2022) 1063.

Selective Tankyrase Inhibitors for Wnt-Driven Cancers: QSAR, Docking, Virtual Screening, and ADMET-Based Selection

Narjes Najafi, Mohammad Hossein Fatemi*
Laboratory of Chemometrics, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran
E-mail address: N.Najafi1987@gmail.com

ABSTRACT

The canonical Wnt (Wingless/Int-1) pathway plays an important role in embryonic development, adult tissue homeostasis, and cancer. In many Wnt-addicted tumors, recurrent mutations in adenomatous polyposis coli (APC), axin-1/2, or catenin beta-1 (CTNNB1) lock this pathway in an active state, driving β-catenin accumulation and oncogenic transcription [1-2]. Tankyrases 1 and 2 (TNKS1/2; hereafter TNKS), poly(ADP-ribose) polymerase (PARP)-family enzymes that poly(ADP-ribosyl)ate the Axin scaffold to promote its ubiquitin-mediated degradation, are therefore attractive targets. Inhibiting TNKS stabilizes Axin, restores β-catenin turnover, and offers a mechanism-based opportunity where approved Wnt-pathway therapies remain lacking.

We pursued an integrated ligand and structure based workflow to identify TNKS inhibitors. Datasets comprising quinazolinone series and related scaffolds were harmonized for consistency; descriptors were scaled and collinearity reduced via variance and inter-correlation filters. Two models—multiple linear regression (MLR) and support vector machine (SVM) [3-4]—were trained on stratified splits within a defined applicability domain (Williams plot). The MLR model achieved cross-validated coefficient of determination (Q^2) = 0.763 (standard error of prediction, SPRESS = 0.228), and the SVM achieved Q^2 = 0.653 (SPRESS = 0.276). Both showed strong fit and external predictivity (training-set correlation, R_{train} = 0.918; test-set correlation, R_{test} = 0.914 for MLR; R_{train} = 0.928; R_{test} = 0.927 for SVM). Y-scrambling indicated the models were not driven by chance correlations.

Docking served as a complementary analysis that confirmed ligand placement within the tankyrase ADP-ribosyltransferase (ART) catalytic pocket, specifically the nicotinamide (NI) subsite. The interaction map indicated stable hydrogen bonds with SER1186A, GLY1231A, ASP1134A, and ILE1228A; in addition, hydrophobic contacts and, in some poses, alignment of an aromatic ring consistent with π-stacking were observed. This interaction pattern supports credible occupancy of the active site and provides a plausible structural basis for inhibition, reinforcing the quantitative findings.

Validated Quantitative Structure–Activity Relationship (QSAR) [3] models were subsequently applied to an independent, lead-like library pre-filtered for toxicity and developability. The prospective screen prioritized several novel candidates with favorable predicted absorption, distribution, metabolism, excretion, and toxicity (ADMET)[5], expanding the accessible chemical space for tankyrase inhibition.

Collectively, these results support an efficient, mechanism-informed path to safe and effective TNKS inhibitors. By coupling externally predictive QSAR with structurally rational docking and prospective screening, this study advances credible candidates for biochemical validation in Wnt-driven cancers.

Keywords: “Tankyrase inhibitors”, “Wnt/β-catenin signaling”, “TNKS1/2”, “QSAR”, “Molecular docking”, “ADMET” .

References:

[1] J.W. Johannes, L. Almeida, B. Barlaam, P.A. Boriack-Sjodin, R. Casella, R.A. Croft, A.P. Dishington, L. Gingipalli, C. Gu, J.L. Hawkins, J.L. Holmes, T. Howard, J. Huang, S. Ioannidis, S. Kazmirski, M.L. Lamb, T.M. McGuire, J.E. Moore, D. Ogg, A. Patel, K.G. Pike, T. Pontz, G.R. Robb, N. Su, H. Wang, X. Wu, H-J. Zhang, Y. Zhang, X. Zheng, and T. Wang, “Pyrimidinone Nicotinamide Mimetics as Selective Tankyrase and Wnt Pathway Inhibitors Suitable for in Vivo Pharmacology”, ACS Medicinal Chemistry Letters, vol. 6, (**2015**), 254–259.
[2] H-P. Buchstaller, U. Anlauf, D. Dorsch, D. Kuhn, M. Lehmann, B. Leuthner, D. Musil, D. Radtki, C. Ritzert, F. Rohdich, R. Schneider, and C. Esdar, “Discovery and Optimization of 2-Arylquinazolin-4-ones into a Potent and Selective Tankyrase Inhibitor Modulating Wnt Pathway Activity”, Journal of Medicinal Chemistry, vol. 62, (**2019**), 9921–9940.
[3] A. Talevi, “Computer-Aided Drug Discovery and Design: Recent Advances and Future Prospects”, Methods in Molecular Biology, vol. 2714, (**2023**), 1–20.
[4] M. Shah, M. Patel, M. Shah, M. Patel, and M. Prajapati, “Computational transformation in drug discovery: A comprehensive study on molecular docking and quantitative structure activity relationship (QSAR)”, Intelligent Pharmacy, vol. 2, (**2024**), 589–595.
[5] G. Klopman, L. R. Stefan, and R. D. Saiakhov, “ADME evaluation. 2. A computer model for the prediction of intestinal absorption in humans”, European Journal of Pharmaceutical Sciences, vol. 17, (**2002**), 253–263.

Development of a chemometrics classification model in order to identification of diabetic patients for their saliva infrared spectrum

Sepideh Khodkavandi, Mohammad Hossein Fatemi, Mohammad Javad Chaichi*
Laboratory of Chemometrics and Spectroscopy, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran
E-mail address: sepideh.kh77@gmail.com

ABSTRACT

Diabetes is one of the most prevalent metabolic disorders worldwide, with an increasing incidence, and its early diagnosis is of critical importance [1]. Non-invasive and rapid diagnostic approaches, such as saliva-based analysis, have emerged as promising alternatives to blood samples [2,3]. In this study, Fourier-transform infrared spectroscopy (FTIR) was employed to identify and classify diabetic and healthy individuals using saliva samples. Infrared spectra were collected from 33 individuals (11 healthy and 22 diabetic) in the range of 4000–650 cm^{-1}, yielding approximately 3400 data points per sample. The data were preprocessed by normalization and noise reduction, and subsequently analyzed in SPSS and Minitab. Principal Component Analysis (PCA) was applied for dimensionality reduction, and Hierarchical Cluster Analysis (HCA) was employed for classification [4].

The results showed that PCA with the first three components achieved a classification accuracy of 78.78%. HCA applied to the raw data achieved 90% accuracy. HCA on the three PCA components resulted in 42% accuracy, while clustering on selected significant wavelengths after PCA yielded 63%. These findings indicate that combining preprocessing, PCA, and HCA improves classification performance, although dimensionality reduction can reduce clustering accuracy in certain cases.

Overall, the results highlight the potential of saliva-based FTIR spectroscopy as a rapid and non-invasive method for diabetes diagnosis and monitoring, and suggest its applicability in the development of automated clinical screening tools.

Keywords: Diabetes, FT-IR spectroscopy, Saliva, PCA, HCA

References

[1] Magliano, D. J., Boyko, E. J., & Atlas, I. D. (2021). What is diabetes?. In *IDF DIABETES ATLAS [Internet]. 10th edition*. International Diabetes Federation.

[2] Avantes BV. (2021, November 24). Non-Invasive Diabetes Diagnosis and Monitoring. AZoM. Retrieved on January 26, 2024 from https://www.azom.com/article.aspx?ArticleID=14393.

[3] Nogueira, M. S. (2022). FTIR spectroscopy as a point of care diagnostic tool for diabetes and periodontitis: A saliva analysis approach. Photodiagnosis and Photodynamic Therapy.

[4] Caixeta, D. C., Carneiro, M. G., Rodrigues, R., Alves, D. C. T., Goulart, L. R., Cunha, T. M., Espindola, F. S., Vitorino, R., & Sabino-Silva, R. (2023). Salivary ATR-FTIR Spectroscopy Coupled with Support Vector Machine Classification for Screening of Type 2 Diabetes Mellitus. Diagnostics, **13**(8), 1396.

Untargeted Metabolomics Based on NMR Spectroscopy and Supervised Kohonen Network for Plasma Biomarker Identification in Early Breast Cancer Detection

Maryam Kashi [1], *Atieh Akbari* [2], *Mohammad Esmaeil Akbari* [2], *Hadi Parastar*[1*]
[1]*Department of Chemistry, Sharif University of Technology, Tehran, Iran*
[2] *Cancer Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran*
E-mail address: h.parastar@sharif.edu

ABSTRACT

Breast cancer (BC) is the most common malignancy among women worldwide, with a significant mortality rate [1]. This study highlights the importance of metabolomics in the early detection and monitoring of BC by analyzing the metabolic composition of blood plasma. This study used proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy combined with supervised Kohonen network (SKN) [2] machine learning algorithms to evaluate plasma samples collected from 72 BC patients and 75 healthy individuals (30–50 years old). For this purpose, 8 mL of blood (fasting) was collected from each participant using sodium citrate tubes and then centrifuged at 4000 rpm for 10 min. The metabolome content was then extracted using a liquid-liquid extraction (LLE) method using methanol as the extraction solvent. Since the number of metabolites is very important in metabolomics, to optimally extract the metabolome from the plasma biological matrix and also to ensure accuracy, metabolites were extracted under the optimal conditions obtained by the response surface methodology (RSM). After sample preparation, the spectral data collected by NMR spectroscopy underwent preprocessing steps, including phase correction, baseline adjustment, before modeling. The classification model showed impressive performance, achieving 93% sensitivity and specificity and 96% accuracy for correctly identifying the control groups. For breast cancer patients, the model achieved 96% sensitivity, 96% specificity, and 93% overall accuracy. Through topological map analysis, the study identified vital biomarkers such as betaine, methionine, choline [3], histidine, and tyrosine, which distinguished patients from healthy subjects with 98% accuracy. The integration of advanced analytical tools such as NMR spectroscopy and SKN machine learning offers a promising strategy for improving BC diagnosis and patient management. These advances could facilitate early detection and personalized treatment approaches, ultimately improving survival outcomes for patients.

Keywords: Chemometrics, Breast cancer; Supervised Kohonen Network; Machine Learning; Response Surface Methodology; Untargeted Metabolomics.

References:

[1] J. Kim et al., "Global patterns and trends in breast cancer incidence and mortality across 185 countries," Nat. Med., 31 **(2025)** 1-9.
[2] W. Melssen, R. Wehrens, and L. Buydens, "Supervised Kohonen networks for classification problems," Chemom. Intell. Lab. Syst., 83 **(2006)** 99–113.
[3] M. Kashi and H. Parastar, "Self-organizing maps for exploration and classification of nuclear magnetic resonance spectra for untargeted metabolomics of breast cancer", J. Pharm. Biomed. Anal., 249 **(2024)** 116377.

Exploring Independence in Area of Feasible Solutions of Three-Component Systems: Calculations and Significance

Fatemeh Sadat Hashemi-Nasab[1,2], Somaye Vali Zade[3], Hamid Abdollahi[4], Hadi Parastar[1,*]

[1]Department of Chemistry, Sharif University of Technology, P.O. Box 11155-9516, Tehran, Iran

[2] Department of Analytical Chemistry, Faculty of Chemistry, Kharazmi University, Tehran 15719-14911, Iran

[3]Halal Research Center of IRI, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran

[4]Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran

* Corresponding authors email address:h.parastar@sharif.edu, h.parastar@gmail.com

Soft modeling uses matrix decomposition to uncover meaningful chemical patterns in complex datasets while minimizing residuals [1]. While PCA is common, it lacks chemical interpretability, limiting its usefulness in chemometrics [2]. ICA, in contrast, extracts statistically independent components, making it effective for resolving overlapping signals in chemical data such as chromatography, spectroscopy, and hyperspectral imaging [2]. This research investigates the issue of rotational ambiguity in independence, focusing on three-component systems. Rotational ambiguity gives rise to an Area of Feasible Solutions (AFS), a low-dimensional representation of all chemically plausible solutions. The relationship between statistical independence qantified using mutual information (MI) and the AFS is examined to assess the reliability of ICA-based solutions. A new perspective on the AFS is introduced to enhance understanding in this context. To evaluate practical implications, both simulated chromatographic datasets (with varying levels of complexity and noise) and an experimental dataset are analyzed. Results show that MI values vary with data characteristics and tend to be higher when spectral patterns are more independent, supporting the concept of data duality. Clear trends in MI variation were observed across different directions in the AFS, and noise analysis confirmed the stability of these findings. Additionally, mean-field independent component analysis (MF-ICA) consistently produced chemically meaningful responses in ternary systems, aligning with AFS boundaries, while mutual information-based least dependent component analysis (MILCA) and joint approximate diagonalization of eigenmatrices (JADE) did not.

Keywords: Mutual Information (MI), Blind Source Separation (BSS), Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), Independent Component Analysis (ICA), Area of Feasible Solutions (AFS), Chemometrics.

References

[1] M.F. Delaney, Chemometrics, Anal. Chem., 56 (1984) 261-277.
[2] A. Golshan, H. Abdollahi, M. Maeder, Anal. Chem., 83 (2011) 836-841.

Design the Electro-Sensor- Liquid Phase Micro Extraction chip for Morphine detection as a model analyte in urine sample

Maryam Mehrban, Tayyebeh Madrakian, Abbas Afkhami, Hossein Khalilian*

Faculty of Chemistry Bu-Ali Sina University, Hamedan, Iran
madrakian@gmail.com

ABSTRACT

Owing to the progress made in the miniaturization of analytical techniques, chip-based methodologies that can amalgamate various laboratory operations onto a diminutive chip are especially suited to investigation [1]. This study presents the design of a millifluidic chip that integrates liquid phase microextraction (LPME) with analyte detection through an electrochemical sensor. Upon incorporating the sample solution that includes the the model analyte, morphine, into the donor part of the milifluidic chip, passive diffusion transpires as it traverses the supported liquid membrane, influenced by the pH gradient. The innovative sensor utilizes a platinum electrode (Pt) that is modified with MDA–Trz–COF, a synthesized covalent organic framework (COF), referred to as COF/ Pt. Following the extraction duration in the acceptor part, a differential pulse voltammetry (DPV) determination of morphine is conducted. The parameters influencing the electro-sensor LPME chip were refined through central composite design methodologies. In ideal circumstances, the maximum current responses observed during the oxidation of morphine on COF /Pt versus various concentrations of morphine within the linear range of 5-700 pmol L^{-1}, a limit of detection of 1.5 pmol L^{-1}, and a limit of quantification of 5 pmol L^{-1}, respectively. The efficacy of the proposed method for the preconcentration and quantification of morphine in urine samples was thoroughly examined.

Keywords: “Liquid Phase Micro Extraction” “ElectroSensor” “Morphine” “central composite design” “Millifluidic chip” “Covalent organic framework”

References:

[1] N, Ullah, M, Tuzen, “A new trend and future perspectives of the miniaturization of conventional extraction methods for elemental analysis in different real samples: a review”, Critical Reviews in Analytical Chemistry, 54.6, **(2024)**, 1729-1747.

A Hybrid NARX Neural Network and Metaheuristic Optimization Approach for Predicting the Galvanostatic Charge-Discharge Dynamics of Supercapacitors

Hamed Azimi[1], Ebrahim Ghorbani Kalhor[1], Mohammad Behbahani[2], Seyed Reza Nabavi[3], Mohammad Taghi Vardini[1]*

[1]Department of Chemistry, Tabriz Branch, Islamic Azad University, Tabriz, Iran
[2]Department of Chemistry, Faculty of Science, Shahid Chamran University of Ahvaz, Ahvaz, Iran
[3]Department of Applied Chemistry, Faculty of Chemistry, University of Mazandaran, Babolsar, Iran
E-mail address (corresponding author): ekalhor@iaut.ac.ir

ABSTRACT

This research presents a novel artificial intelligence-based approach for the complete modeling and prediction of the dynamic behavior of supercapacitors. The main innovation of this study is the transition from predicting static parameters—which has been the primary focus of previous studies on capacity prediction [1], Remaining Useful Life (RUL) [2], and State of Health (SOH) [3]—to simulating the entire Galvanostatic Charge-Discharge (GCD) curve, effectively creating a powerful virtual simulator. For this purpose, a hybrid model based on a Nonlinear Autoregressive with Exogenous Inputs (NARX) recurrent neural network was designed. Experimental data were digitally extracted from the GCD curves of a modified graphene electrode (CDs/rGO) at three different current densities (0.5, 1, and 2 A/g) from a reputable scientific paper [4], and were subsequently pre-processed using linear interpolation to create a uniform time series. After data preparation, the optimal architecture and key hyperparameters of the network were systematically determined using a Bayesian Optimization approach. Subsequently, the NARX network was trained by optimizing its parameters (weights and biases) using six different metaheuristic algorithms: FA, PSO, IWO, ACO, GA, and COVIDOA. A comprehensive evaluation of the models' performance, utilizing six key statistical metrics (including R^2, RMSE, MSE, R, WI, and NSE), demonstrated the superior performance of the Firefly Algorithm (FA), Particle Swarm Optimization (PSO), and Invasive Weed Optimization (IWO). These superior models achieved near-ideal performance metrics (e.g., $R^2 > 0.999$ and RMSE ≈ 0.001) and were able to reconstruct the GCD curves in excellent agreement with the actual data. This research demonstrates that the developed hybrid model is an efficient tool for the accurate simulation of supercapacitor dynamics, enabling the calculation of key performance parameters such as energy density, power density, coulombic efficiency, and equivalent series resistance (ESR) without the need for costly experiments.

Keywords: Supercapacitor, NARX Neural Network, Bayesian Optimization, Metaheuristic Algorithms, Time Series Forecasting, Galvanostatic Charge-Discharge

References:

[1] Tawfik WZ, Mohammad SN, Rahouma KH, Salama GM, Tammam E. "Machine learning models for capacitance prediction of porous carbon-based supercapacitor electrodes" Physica Scripta 2024;99(2):026001.

[2] Qi G, Ma N, Wang K. “Predicting the remaining useful life of supercapacitors under different operating conditions” Energies 2024;17(11):2585.
[3] Haris M, Hasan MN, Qin S. “State of health prediction of supercapacitors using multi-trend learning of NARX neural network” Materials Today Sustainability 2022;20:100201.
[4] Dang Y-Q, Ren S-Z, Liu G, Cai J, Zhang Y, Qiu J. “Electrochemical and capacitive properties of carbon dots/reduced graphene oxide supercapacitors” Nanomaterials 2016;6(11):212.

Application of Group LASSO as dimension reduction method in QSAR Studies using a novel descriptor grouping strategy

[a]Farzaneh Kia*, [a]Mansour Arab Chamjangali, [b]Mozhgan Beglari
[a]Department of chemistry, Sharood University of Technology, Shahrood, Iran
[b]Department of Chemical Industry, Technical and Vocational University(TVU), Tehran, Iran
E-mail address: Farzaneh.kia18@yahoo.com

ABSTRACT
In Quantitative Structure–Activity Relationship (QSAR) studies, penalized regression methods such as the Least Absolute Shrinkage and Selection Operator (LASSO) [1], Smoothly Clipped Absolute Deviation (SCAD) [2], and Adaptive LASSO (ALASSO) [3] are widely used for feature selection to improve predictive models while, before applying these methods, dimensionality reduction is essential for managing high-dimensional datasets and enhancing model stability and interpretability [4–6]. For the first time in QSAR studies, we employed Group Lasso (GLASSO) [7] as a preprocessing step before LASSO, ALASSO, and SCAD. To this end, a novel descriptor grouping strategy was applied to organize 1643 molecular descriptors into meaningful groups, preserving both chemical information and their statistical relevance to biological response. Descriptors were first categorized based on properties such as topological, chemical, geometrical, and electronic features. Within each category, descriptors were further grouped according to their absolute Pearson correlation coefficient with the response, creating 311 structured groups. This targeted grouping allowed GLASSO to efficiently reduce dimensionality while retaining biologically relevant information for subsequent penalized regression modeling. Using this approach, the structured groups obtained through GLASSO for a dataset comprising 73 anti-HIV-1 compounds were subsequently employed as input for LASSO, SCAD, and ALASSO to perform further feature selection. The results were evaluated both with and without the GLASSO-based preprocessing step. For the test set, the predictive performance of each method was summarized in terms of coefficient of determination (R^2), mean squared error (MSE), and the number of significant variables (n). Specifically, for LASSO with GLASSO preprocessing the results were $R^2_{test} = 0.92$, $MSE_{test} = 0.05$, n = 15, whereas without GLASSO the corresponding values were $R^2_{test} = 0.93$, $MSE_{test} = 0.05$, n = 30. For SCAD, the GLASSO-preprocessed model yielded $R^2_{test} = 0.89$, $MSE_{test} = 0.07$, n = 8, compared to $R^2_{test} = 0.88$, $MSE_{test} = 0.08$, n = 16 without GLASSO. Similarly, for Adaptive LASSO (ALASSO), the GLASSO-based approach achieved $R^2_{test} = 0.93$, $MSE_{test} = 0.05$, n = 8, while the direct model without GLASSO produced $R^2_{test} = 0.84$, $MSE_{test} = 0.11$, n = 11. These results clearly demonstrate the effectiveness of GLASSO as an efficient dimension reduction technique when combined with our novel descriptor grouping strategy.

Keywords: QSAR, Group Lasso, Feature Selection, Dimension Reduction, Penalized Regression Methods

References:
[1] Tibshirani R. *Regression shrinkage and selection via the lasso*. Journal of the Royal Statistical Society Series B: Statistical Methodology. 58. (1996). 267-288.
[2] Fan J. and Li R. *Variable selection via nonconcave penalized likelihood and its oracle properties*. Journal of the American Statistical Association. 96. (2001). 1348-1360.
[3] Zou H. *The adaptive lasso and its oracle properties*. Journal of the American Statistical Association. 101. (2006). 1418-1429.
[4] Mozafari Z., Arab Chamjangi M., Arashi M., and Goudarzi N. *Performance of smoothly clipped absolute deviation as a variable selection method in the artificial neural network-based QSAR studies*. Journal of Chemometrics. 35. (2021). 1-20.
[5] Mozafari Z., Arab Chamjangi M., Arashi M., and Goudarzi N. *Combination of least absolute shrinkage and selection operator with Bayesian Regularization artificial neural network (LASSO-BR-ANN) for QSAR studies using functional group and molecular docking mixed descriptors*. Chemometrics and Intelligent Laboratory Systems. 200. (2020). 1-13.
[6] Mozafari Z., Arab Chamjangi M., Arashi M., and Goudarzi N. *Suggestion of active 3-chymotrypsin like protease (3CLPro) inhibitors as potential anti-SARS-CoV-2 agents using predictive QSAR model based on the combination of ALASSO with an ANN model*. SAR and QSAR in Environmental Research. 32. (2021). 863-888.
[7] Yuan M. and Lin Y. *Model selection and estimation in regression with grouped variables*. Journal of the Royal Statistical Society Series B: Statistical Methodology. 68. (2006). 49-67.

Application of Group Exponential LASSO and as bi-level feature selection method in QSAR Studies

[a]Farzaneh Kia*, [a]Mansour Arab Chamjangali,[b]Mozhgan Beglari
[a]Department of chemistry, Sharood University of Technology, Shahrood, Iran
[b]Department of Chemical Industry, Technical and Vocational University(TVU), Tehran, Iran
E-mail address: Farzaneh.kia18@yahoo.com

ABSTRACT
Quantitative Structure–Activity Relationship (QSAR) modeling often involves high-dimensional datasets with thousands of molecular descriptors, making efficient feature selection a crucial step in building robust predictive models [1]. Penalized regression methods have been widely applied in this context, with LASSO [2] being one of the most common approaches. However, LASSO treats each descriptor independently and may retain redundant variables, particularly when descriptors are highly correlated [3].For the first time in QSAR studies, we employed group penalized regression methods as an efficient feature selection strategy. For this purpose, we employed a threshold-based correlation network [4] to organize descriptors into related groups for structured penalized regression, resulting in 191 groups. Group penalized regression methods were then applied to select variables in a structured and stable manner. Specifically, Group LASSO (GLASSO) [5] was used, which selects or discards predefined groups of variables based on a combined penalty. Since it does not allow within-group selection, Group Exponential LASSO (GEL) [4] was also employed to enable simultaneous selection at both group and within-group levels, thereby overcoming the limitations of classical LASSO. Subsequently, the descriptors selected by GEL were used as input features to train an artificial neural network (ANN), enabling further evaluation and enhancement of predictive performance.A dataset of 335 molecular descriptors (after the preprocessing step) related to anti-HIV-1 activity was analyzed. The descriptors selected by GEL resulted in a simple and interpretable linear model with acceptable predictive accuracy (R^2_{test} = 0.87, MSE_{test} = 0.11). To further improve predictive performance, the six significant descriptors identified by GEL were subsequently employed as input features to train an artificial neural network (ANN). The GEL–ANN pipeline with a 6-9-1 architecture achieved the highest overall accuracy, yielding strong correlation coefficients (R^2_{val} = 0.93, R^2_{tes} = 0.81) and low prediction errors (MSE_{val} = 0.28, MSE_{test} = 0.15). These findings suggest that while GEL ensures sparsity and interpretability, coupling it with ANN effectively captures nonlinear relationships in high-dimensional QSAR datasets, resulting in stable, parsimonious, and highly accurate predictive models.

Keywords: QSAR modeling, Feature selection, Group LASSO, GEL, Artificial neural network, Drug discovery.

References:
[1] Nantasenamat C., Moomro T., Worachartcheewan A., Mandi P., Isarankura-Na-Ayudhya C., and Prachayasittikul V. *Predictive QSAR modeling of aldehyde reductase inhibitors using Monte Carlo feature selection*. European Journal of Medicinal Chemistry. 76. (2014). 352-359.
[2] Tibshirani R. *Regression shrinkage and selection via the lasso*. Journal of the Royal Statistical Society Series B: Statistical Methodology. 58. (1996). 267-288.
[3] Martens R. and Steland A. *Concentration and Model Selection Consistency of the Group Lasso for α-Mixing Errors*. Stat. 14. (2025). 1-8.
[4] Eguiluz V., Chialvo D., Cecchi G., Baliki M., and Apkarian A. *Scale-free brain functional networks*. Physical Review Letters. 94. (2005). 1-4.
[5] Yuan M. and Lin Y. *Model selection and estimation in regression with grouped variables*. Journal of the Royal Statistical Society Series B: Statistical Methodology. 68. (2006). 49-67.
[6] Breheny P. *The group exponential lasso for bi-level variable selection*. Biometrics. 71. (2015). 731-740.

Effect of Descriptor Grouping Strategies on Bi-Level Feature Selection and Predictive Performance in GEL-Based QSAR Modeling

[a]Farzaneh Kia*, [a]Mansour Arab Chamjangali, [b]Mozhgan Beglari
[a]Department of Chemistry, Shahrood University of Technology, Shahrood, Iran
[b]Department of Chemical Industry, Technical and Vocational University(TVU), Tehran, Iran
E-mail address: Farzaneh.kia18@yahoo.com

ABSTRACT

Variables in real-world datasets, especially in chemical and biological studies, frequently exhibit an inherent group structure [1]. Ignoring this inherent group structure may lead to the exclusion of features that appear weak individually but collectively play a crucial role in predictive modeling [2]. In this study, with a focus on the Group Exponential LASSO (GEL) [3], as an advanced bi-level feature selection method that performs simultaneous selection both *within groups* and *between groups*, the impact of different descriptor grouping strategies, such as property-based grouping, correlation-based clustering, and a hybrid approach, was systematically investigated on the performance of GEL in QSAR modeling. To this end, these distinct descriptor grouping strategies were employed in this QSAR study. In the property-based grouping, descriptors were classified into 22 predefined groups according to Dragon software's default categorization, reflecting their topological, chemical, geometrical, and electronic properties. In the correlation-based grouping, a pairwise correlation matrix among descriptors was computed, and a threshold (here, 0.8) was applied. Entries equal to or exceeding the threshold were assigned a value of 1, while others were assigned a value of 0. Connected descriptors, identified through this binary matrix, were subsequently clustered into separate groups, resulting in 191 groups of descriptors. Finally, in the hybrid-based grouping, descriptors were first classified into the 22 Dragon-defined classes. Within each class, the absolute Pearson correlation coefficient was calculated for each descriptor and the response. Twenty correlation thresholds, ranging from 0 to 1 with a step size of 0.05, were applied to define subgroups within each class. These subgroups were then consolidated to form the final structured groups, resulting in a total of 311 groups for subsequent analysis on a dataset of 1643 molecular descriptors related to anti-HIV-1 activity. In the property-based grouping, GEL selected five variables, yielding an R^2_{test} of 0.88 and an MSE_{test} of 0.09, indicating that some inter-descriptor relationships may be overlooked. In the correlation-based approach, 13 variables were selected ($R^2_{test} = 0.85$, $MSE_{test} = 0.11$), capturing more variables but potentially underrepresenting chemical information. The hybrid-based approach, integrating chemical properties and correlation with the response, selected five variables and achieved improved predictive performance (R^2_{test} of 0.89, MSE_{test} of 0.08). These results highlight the advantage of combining chemical and correlation-based grouping for optimal variable selection, balancing dimensionality reduction, predictive accuracy, and interpretability in QSAR modeling with GEL.

Keywords: QSAR modeling, feature selection, Group Exponential LASSO, predictive modeling, dimensionality reduction

References:

[1] D. Fourches, E. Muratov, and A. Tropsha, "Trust, but verify II: a practical guide to chemogenomics data curation.".Journal of chemical information and modeling 56. (**2016**). 1243-1252.
[2] C. Kuzudisli, B. Bakir-Gungor, N. Bulut, B. Qaqish, and M. Yousef, "Review of feature selection approaches based on grouping of features". PeerJ 11. (**2023**). 1-37.
[3] P. Breheny, "The group exponential lasso for bi-level variable selection.". Biometrics 71. (**2015**). 731-740.

Rapid Methanol Detection in Ethanol Solutions via Non-Linear Raman Spectroscopy Modeling

Somaye Vali Zade[a*], *Hossein Rastegar*[b], *Hamid Abdollahi*[c]

[a] *Halal Research Center of IRI, Food and Drug Administration, Ministry of Health and Medical Education, Tehran, Iran*

[b] *Cosmetic Products Research Center, Iran Food and Drug Administration, MOHE, Tehran, Iran, Future Studies Group, the Academy of Medical Sciences of The I.R. Iran, Tehran, Iran*

[c] *Faculty of Chemistry, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran*

[*] E-mail: smy.valizade@gmail.com

ABSTRACT

Methanol contamination in ethanol-based products poses a significant health risk due to its toxicity at low concentrations. This study developed a rapid, non-destructive method for detecting and quantifying methanol in alcoholic beverages using Raman spectroscopy combined with chemometric models, including Partial Least Squares (PLS), Interval PLS (iPLS), Kernel PLS, and Radial Basis Function Artificial Neural Networks (RBF-ANN). The Kernel PLS model, applied to optimized spectral intervals, achieved the lowest relative prediction error (REP%) of 6.84% on a test set. Validation on real samples spiked with 3–20% methanol confirmed the method's ability to quantify methanol with absolute relative errors below 10% in most cases. Compared to conventional techniques like gas chromatography (GC) and high-performance liquid chromatography (HPLC), this approach requires minimal sample preparation and shorter analysis time, making it suitable for quality control in the food, pharmaceutical, and cosmetic industries.

Keywords : Methanol, Non-Linear Multivariate Calibration. Kernel PLS, RBF-ANN

References (Times New Roman 10pt Bold):

[1] M.J.C. Pontes, S.R.B. Santos, M.C.U. Araújo, L.F. Almeida, R.A.C. Lima, E.N. Gaião, U.T.C.P. Souto, Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry, Food Res. Int. 39 (**2006**) 182-189,

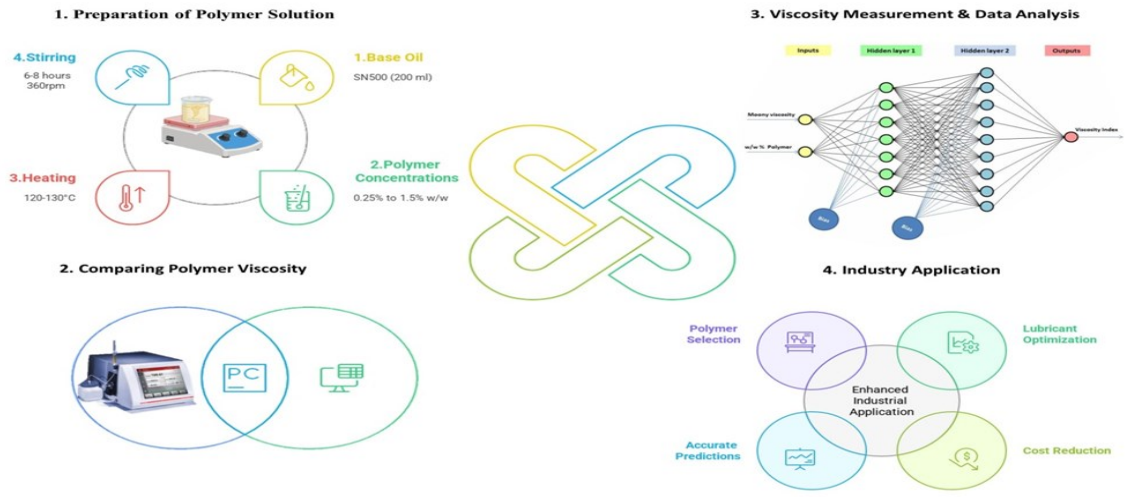
[2] M. Gallignani, S. Garrigues, M. de La Guardia, Direct determination of ethanol in all types of alcoholic beverages by near-infrared derivative spectrometry, Analyst, 118 (**1993**) 1167-1173.

Correlation between Lubricant Viscosity Index and Mooney Viscosity of Polymers: A Predictive Model using Artificial Neural Networks

Mohammad Hossein Aghamohammadi, Ali Niazi, Atisa Yazdanipour, Alireza Feizbaksh
Department of Chemistry, CT.C., Islamic Azad University, Tehran, Iran
E-mail address: ali.niazi@gmail.com; aliniazi@iau.ac.ir

ABSTRACT

This study investigates the correlation between the viscosity index (VI) of lubricating oils and the Mooney viscosity of four viscosity index improver (VII) polymers, with a focus on developing an accurate predictive model using artificial neural networks (ANN). Polymer solutions were prepared at concentrations ranging from 0.25% to 1.5% w/w, and their viscosities were measured at 40°C and 100°C. A 2-7-9-1 ANN model was developed, achieving high accuracy (R^2 = 0.992, RMSE = 1.71). Sensitivity analysis revealed that polymer concentration contributes 58% to VI prediction, while Mooney viscosity accounts for 42%. Among the tested polymers, KEP2060 showed superior performance due to its high molecular weight and unique structure[1]. A simplified linear regression model (VI = 81.72 + 19.41C + 0.33MV) was also proposed for industrial applications where computational resources are limited. These findings provide valuable insights for optimizing lubricant formulations and enhancing product reliability in the automotive, aerospace, and petrochemical industries. The integration of PLS for preliminary screening and ANN for refined predictions enables efficient development processes, reducing reliance on time-consuming experimental trials.



Keywords: Viscosity Index (VI), Viscosity Index Improver Polymers (VII), Mooney Viscosity, Artificial Neural Network (ANN), Partial Least Squares (PLS), Lubricant Optimization.

References:

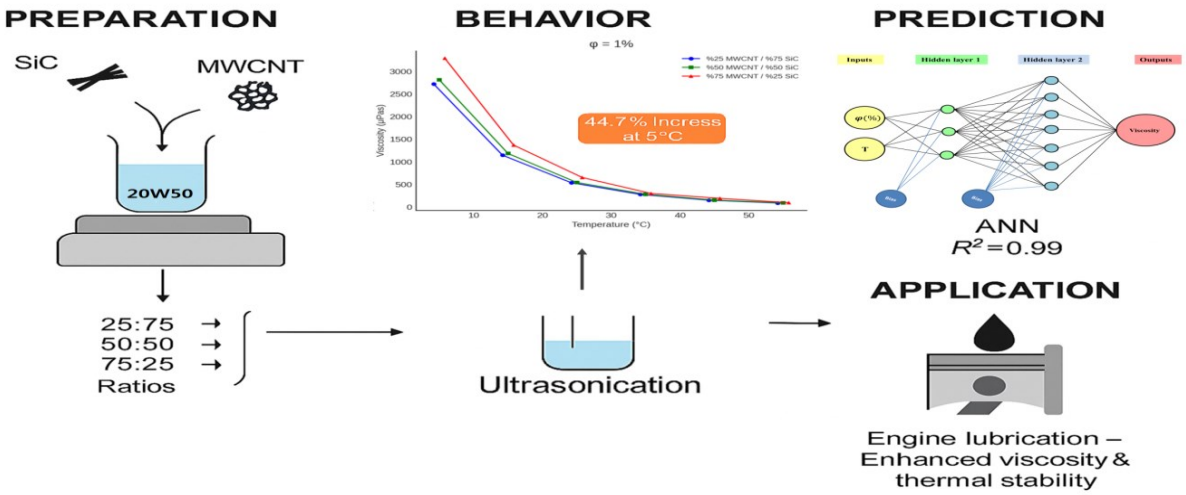
[1] Bhattad A. Review on viscosity measurement: devices, methods and models. Journal of Thermal Analysis and Calorimetry.2023;148(14):6527-6543. https://doi.org/10.1007/s10973-023-12214-0

Viscosity Enhancement of MWCNT/SiC Hybrid Nanofluids in SAE 20W50 Engine Oil: Experimental Analysis and Machine Learning Modeling

Mohammad Hossein Aghamohammadi, Ali Niazi, Atisa Yazdanipour, Alireza Feizbaksh
Department of Chemistry, CT.C., Islamic Azad University, Tehran, Iran
E-mail address: ali.niazi@gmail.com; aliniazi@iau.ac.ir

ABSTRACT

This study investigates the viscosity enhancement of SAE 20W50 engine oil using hybrid SiC/MWCNT nanoparticles at temperatures ranging from 5 to 55°C and concentrations of 0.1–1.0 wt%. The 75:25 MWCNT/SiC ratio at 1.0 wt% resulted in a 44.7% increase in viscosity at 5°C due to the MWCNT network formation and stable dispersion of SiC. An ANN model ($R^2 = 0.996$) performed better than PLS regression in accurately predicting viscosity. Principal Component Analysis highlighted temperature's dominant influence. These nanolubricants, through viscosity enhancement, have the potential to improve friction and wear properties in automotive engines, as well as ensure stable flow in hydraulic systems[1]. Although thermal conductivity was not measured due to resource limitations, the results suggest a significant potential for enhancing the tribological properties of the proposed hybrid nanolubricants. Future work will focus on direct thermal conductivity measurements, optimizing nanoparticle ratios, and investigating extended operating conditions.



Keywords: Hybrid nanofluid; Viscosity prediction; MWCNT/SiC; Artificial neural network(ANN); Tribological properties; Machine Learning .

References:

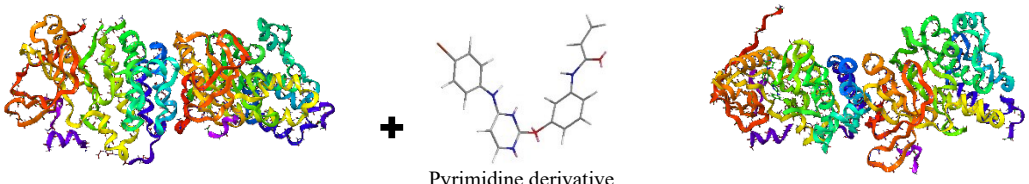
[1] Rahman MM, Islam M, Roy R, Younis H, AlNahyan M, Younes H. Carbon nanomaterial-based lubricants: review of recent developments. Lubricants. 2022;10(11):281. https://doi.org/10.3390/lubricants10110281

QSAR and molecular docking studies for a chemometrics approach to the investigation of pyrimidine derivatives as novel inhibitors of anticancer drugs

Dariush Fathali[1], *Ali Niazi*[*1], *Amirhossein Mohsen Sarrafi*[1], *Atisa Yazdanipour*[1]
Department of Chemistry, CT.C., Islamic Azad University, Tehran, Iran
E-mail address: aliniazi@iau.ac.ir

ABSTRACT

Quantitative structure–activity relationship (QSAR) analysis was conducted on a series of 41 pyrimidine derivatives [1] that were identified as novel inhibitors of the double-mutant EGFR (L858R/T790M) tyrosine kinase (TKIs). The aim was to overcome resistance to first-generation EGFR inhibitors and the non-selective toxicities of second-generation agents in patients with non-small-cell lung cancer (NSCLC)—the analysis employed chemometric methods. The IC_{50} values of the pyrimidine derivatives served as the response variable and are indicative of effective therapeutic dosing. QSAR models [2,3] were built using partial least squares (PLS) and least absolute shrinkage and selection operator (LASSO) regression on selected molecular descriptors to predict activity. All selected descriptor standardscale were performed before modeling for normalization of molecular descriptors. Cross-validation results showed that the PLS model has higher precision and lower error than the LASSO model, with Q^2 (R^2_CV) = 0.810 and RMSE_CV = 0.01065 for PLS, versus Q^2 (R^2_CV) = 0.706 and RMSE_CV = 0.01323 for LASSO. Finally, molecular docking was performed to evaluate the plausibility of the newly designed molecule, focusing on the lowest-energy poses and error criteria.



Keywords: Drug discovery, QSAR, Molecular Docking, Chemometrics, anti-cancer

References:

[1] R. Pawara, I. Ahmad, D. Nayak, S. Belamkar, S. Surana, Ch. Nath Kundu, Ch. Patil, H. Patel, "Design and synthesis of the novel, selective WZ4002 analogue as EGFR-L858R/T790M tyrosine kinase inhibitors for targeted drug therapy in non-small-cell lung cancer (NSCLC)", Journal of Molecular Structure, 1254, (2022), 132313.

[2] M. Mirshafiei, A. Niazi, A. Yazdanipour, "Application of MIA-QSAR in Designing New Protein P38 MAP Kinase Compounds Using a Genetic Algorithm", Periodica Polytechnica Chemical Engineering, 67 (1), (2023), 141-160.

[3] Z. Doroudi, A. Niazi, "Application of Genetic Algorithms for Pixel Selection in MIA-QSAR Studies on Anti-HIV HEPT Analogues for New Design Derivatives", Iranian Journal of Pharmaceutical Research, 18 (3), (2019), 1239-1252.

Application of preprocessing methodes in image processing for simultaneous determination of three drugs

Parisa Shakhsari, Ali Niazi*, Homayon Ahmad Panahi, Amir Hossein Mohsen Sarrafi, Atisa Yazdanipour

Department of Chemistry, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

E-mail: aliniazi@iau.ac.ir

ABSTRACT

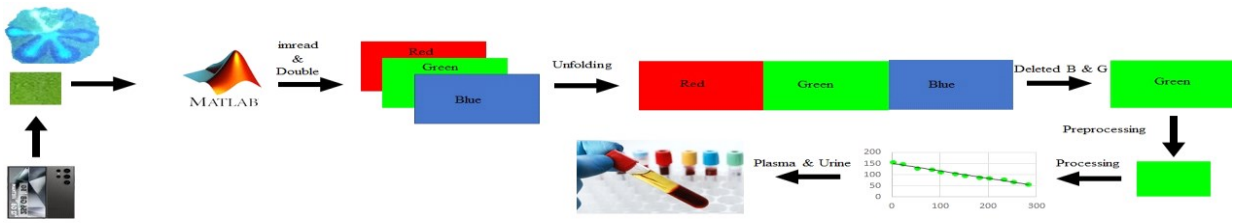
Simultaneous quantification of three pharmaceutical compounds in human plasma and urine has long represented a substantial challenge in the field of bioanalytical chemistry. While techniques such as spectroscopy, capillary electrophoresis, and electrochemical methods offer high precision and sensitivity, their adoption in routine testing is hindered by several drawbacks: the requirement for specialized and costly instruments, complex and time-intensive sample preparation protocols, laborious operational procedures, reliance on highly trained personnel, and use of potentially hazardous chemical reagents. In contrast, fluorimetry approaches provide a more practical alternative for everyday drug analysis: they deliver high sensitivity and selectivity, incur lower costs, enable faster throughput, operate with simpler instrumentation, and offer excellent reproducibility. [1]

In this study, we employed an innovative microfluidic paper fabrication technique in combination with simultaneous mobile phone imaging to detect and quantify three pharmaceutical compounds in real samples. Image preprocessing and processing were conducted using MATLAB and ImageJ software also Artificial Intelligence (AI) to analyze the results. Preprocessing methods in image processing are crucial for simultaneous determination of three drugs, especially when using image-based analytical techniques. These methods enhance the quality of images, reduce noise, and prepare the data for subsequent analysis, leading to more accurate and reliable determination of drug amounts. With preprocessing, image data is optimized for subsequent analytical methods, such as chemometrics analysis or deep learning models, allowing for simultaneous and accurate determination of three drugs based on their distinct visual features.

Data preprocessing constitutes a critical stage in the knowledge discovery process. Although it is less well-recognized than subsequent stages such as data mining, it frequently demands a greater share of time and effort in the overall data analysis workflow. Raw datasets typically exhibit various deficiencies, including inconsistencies, missing values, noise, and/or redundancy. Consequently, if subsequent learning algorithms are supplied with low-quality data, their performance will inevitably degrade. Appropriate preprocessing can therefore substantially enhance the quality and reliability of subsequent pattern discovery and automated decision-making.

Data preparation, as a fundamental component of preprocessing, aims to transform raw input into high-quality data that is optimally aligned with the requirements of the subsequent exploration phase. This preparation step is regarded as an indispensable element of the analytical pipeline. [2]

Keywords: Preprocessing, Image processing, Chemometric, Fluorimetry, Microfluidic-paper, Simultaneous detection



Reference:
[1] S. Azarian, M. Shaghaghi, G. Dehghan and N. Sheibani, "A rapid, simple and ultrasensitive spectrofluorimetric method for the direct detection of metformin in real samples based on a nanoquenching approach", Luminescence, vol. 36, (2021), pp. 658-667.
[2] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez and F. Herrera, "Big data preprocessing: methods and prospects", Big Data Analytics, vol. 1, (2016), p.9

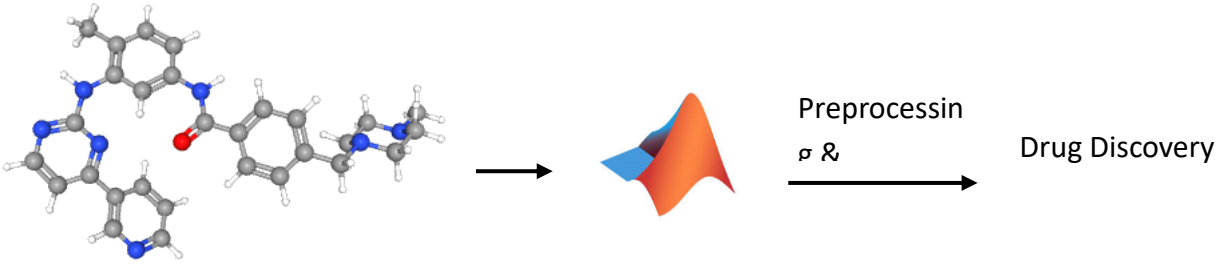
Application of preprocessing methods in image processing in QSAR and drug design

Parisa Shakhsari, Ali Niazi*, Homayon Ahmad Panahi, Amir Hossein Mohsen Sarrafi, Atisa Yazdanipour
Department of Chemistry, Central Tehran Branch, Islamic Azad University, Tehran, Iran.
E-mail: aliniazi@iau.ac.ir

ABSTRACT

Imatinib, the first line therapy for CML treatment, binds the inactive conformation of the kinase domain. Nevertheless, resistance to this drug is frequently observed mainly because of mutations in the kinase domain of Bcr–Abl. For this reason, new Abl inhibitors able to bind both the active and inactive conformations, or that preferentially bind the active form of the protein, were discovered. [1] A major step in constructing the QSAR models is finding one or more molecular descriptors
that represent variation in the structural property of the molecules numerically. Awide variety of descriptors have been
reported on QSAR analysis. [2]
In this study, using a variety of preprocessing and image processing methods for molecules drawn in Chemdraw software and examining them by artificial intelligence in Matlab, the best method for designing a new drug was selected. According to the obtained data and data analysis, the best method is GA-PLS, which RMSEP=4.7491 was obtained for each descriptor.

Keywords Preprocessing, Image processing, Chemometric, QSAR, Drug discovery, Anti-cancer



References:

[1] F. Falchi, F. Manetti, F. Carraro, A. Naldini, G. Maga, E. Crespan, S. Schenone, O. Bruno, C. Brullo, and M. Botta, "3D QSAR Models Built on Structure-Based Alignments of Abl Tyrosine Kinase Inhibitors", *ChemMedChem*, vol. 4, (2009), 976–987.

[2] A. Niazi, S. Jameh-Bozorghi, and D. Nori-Shargh, "Quantitative structure–property relationship study of the solubility of thiazolidine-4-carboxylic acid derivatives using ab initio and genetic algorithm–partial least squares", *Chinese Chemical Letters*, vol. 18, (2007), 621–624.

Quantitative structure activity relationship study of p38α MAP kinase inhibitors

*Maryam Nouri majd, Eslam Pourbasheer**
Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili, Ardabil, Iran
E-mail address: e.pourbasheer@uma.ac.ir

ABSTRACT

The protein kinase p38α - a member of the MAP kinase family - plays a pivotal role in the response to environmental stresses and in regulating the production of proinflammatory cytokines (TNF-α, IL-1β, IL-6, IL-8). This kinase is involved in the pathogenesis of diseases such as cancer, rheumatoid arthritis, and neurological disorders and is considered a promising therapeutic target[1]. The development of new p38α inhibitors is usually associated with expensive and time-consuming laboratory methods. QSAR modeling is considered an efficient strategy for predicting the activity of compounds and optimizing drugs[2]. In this study, the inhibitory activity of 48 pyridine [3, 4] derivatives against p38α was analyzed using SW-MLR (stepwise multiple linear regression) and GA-MLR (genetic algorithm multiple linear regression) methods. 3219 molecular descriptors were calculated by Dragon software and after filtering (removing fixed and correlated descriptors with $r > 0.90$), 490 descriptors were used for modeling. The dataset was divided into two parts: training (38 compounds) and testing (10 compounds) sets, using hierarchical clustering. The SW-MLR model[5] with 6 descriptors showed good quality for training data (R^2_{train}=0.9002, Q^2_{LOO}=0.8119) but performed poorly in predicting test data (R^2_{test}=0.1176, $RMSE_{test}$=1.2173) and identified 5 compounds as outliers. The GA-MLR model[6] with 6 optimized descriptors showed superior performance in both sets (R^2_{train}=0.8353, R^2_{test}=0.6452, $RMSE_{test}$=0.6006) and had only 1 outlier. Analysis of GA-MLR descriptors showed that the IDDE descriptor (information distance index) has the greatest impact on activity. Other descriptors included radial distribution functions (RDF045m, RDF120m) and topological indices (GGI7, H2u, RPCG). Model validation with Y-scrambling, VIF analysis, and Williams plot methods confirmed that the GA-MLR model is robust and generalizable (Q^2_{LGO}=0.7502). Finally; Genetic algorithm is the preferred approach for descriptor selection in QSAR modeling of p38α inhibitors. Atomic distribution (IDDE), atomic mass (RDF045m, RDF120m), and molecular topological features are key determinants of activity. This model is capable of predicting the activity of novel pyridine compounds.

Keywords : QSAR, Genetic algorithm, p38α MAP, Pyridin derivative

References :

[1] E. Pourbasheer, S. Ahmadpour, R. Zare-Dorabei, M. Nekoei, Quantitative structure activity relationship study of p38α MAP kinase inhibitors, Arabian Journal of Chemistry, 10 (2017) 33-40.
[2] E. Pourbasheer, R. Aalizadeh, M. Ganjali, P. Norouzi, J. Shadmanesh, C. Methenitis, QSAR study of Nav1.7 antagonists by multiple linear regression method based on genetic algorithm (GA–MLR), Medicinal Chemistry Research, 23 (2014) 2264-2276.
[3] A. Kaieda, M. Takahashi, H. Fukuda, R. Okamoto, S. Morimoto, M. Gotoh, T. Miyazaki, Y. Hori, S. Unno, T. Kawamoto, T. Tanaka, S. Itono, T. Takagi, H. Sugimoto, K. Okada, G. Snell, R. Bertsch, J. Nguyen, B.-C. Sang, S. Miwatashi, Structure-Based Design, Synthesis, and Biological Evaluation of Imidazo[4,5-b]pyridin-2-one-Based p38 MAP Kinase Inhibitors: Part 1, ChemMedChem, 14 (2019) 1022-1030.
[4] A. Kaieda, M. Takahashi, H. Fukuda, R. Okamoto, S. Morimoto, M. Gotoh, T. Miyazaki, Y. Hori, S. Unno, T.

Kawamoto, T. Tanaka, S. Itono, T. Takagi, H. Sugimoto, K. Okada, W. Lane, B.-C. Sang, K. Saikatendu, S. Matsunaga, S. Miwatashi, Structure-Based Design, Synthesis, and Biological Evaluation of Imidazo[4,5-b]Pyridin-2-one-Based p38 MAP Kinase Inhibitors: Part 2, ChemMedChem, 14 (2019) 2093-2101.
[5] S. Alimohammadi, A. Hamidi, P. Pargolghasemi, N. Nourani, M.S. mainezhad-Namin, QSAR study of antiproliferative drug against A549 by GA-MLR and SW-MLR methods, Chemical Review and Letters, 2 (2019) 193-198.
[6] F. Khosravi, R. Kiani-anbouhi, E. Pourbasheer, QSAR Study on DYRK1A Inhibitors for Regenerative Therapy in Diabetes, Advanced Journal of Chemistry, Section A, 7 (2024) 522-539.

Three-dimensional QSAR study on XIa (FXIa) inhibitors and design of new compounds based on CoMFA, CoMSIA and molecular docking

*Ladan Kazemnezhad , Eslam Pourbasheer**
e.pourbasheer@uma.ac.ir

ABSTRACT

In recent years, significant research efforts have been directed toward developing targeted therapeutics for specific diseases. Factor XIa (FXIa) has emerged as a promising target for novel anticoagulant drugs, as its inhibition reduces thrombosis risk with minimal impact on bleeding propensity. In this study, 3D-QSAR modeling was employed to design new FXIa inhibitors as anticoagulants. The dataset was partitioned into training and test sets, with hierarchical clustering method. CoMFA, CoMFA-Focus, and CoMSIA models were developed to predict compound activity. Statistical evaluation identified the CoMSIA model, with a cross-validated correlation coefficient (q^2) of 0.515 and a non-cross-validated correlation coefficient (r^2) of 0.804, as the optimal predictive model. Contour maps generated from this model highlighted key molecular regions critical for enhancing inhibitory activity. Based on these insights, 11 novel compounds were designed. Molecular docking and ADMET analyses were conducted to further evaluate the model and the designed compounds. ADMET analysis confirmed the drug-like properties of the designed compounds, while molecular docking validated these findings by elucidating key ligand-receptor interaction [1].

Keywords: 3D-QSAR, Anticoagulant, CoMFA , CoMSIA , Factor Xia (FXIa)

References:

[1] Yajing Zhang , Rong Huang , Xueqin Hu , Nan Zheng , Lei Geng , Zequn Yin ,Yajun Duan , Qin Wang , Chenzhong Liao , Xiaoxiao Yang, Zhouling Xie, Discovery of potent, highly selective, and orally bioavailable factor XIainhibitors for anticoagulant therapy European Journal of Medicinal Chemistry 289 (2025) 117436.

[2] Reza Mahmoudzadeh Laki and Eslam Pourbasheer,3D-QSAR Modeling on 2-Pyrimidine Carbohydrazides as Utrophin Modulators for the Treatment of Duchenne Muscular Dystrophy by Combining CoMFA, CoMSIA, and Molecular Docking Studies, ACS Omega 2024, 9, 24707−24720

[3] Mohammad Sarafraz, Eslam Pourbasheer, Reza Mahmoudzadeh Laki, Mohammad Fadaeian, 3D-QSAR study on the DYRK1A inhibitors and design of new compounds by CoMFA and CoMSIA methods, Chem Rev Lett 7 (2024) 926-941 (926-941)

Three-dimensional QSAR study on iazolo-pyrimidine derivatives as WRN inhibitors for the treatment of MSI tumors and design of new compounds based on CoMFA and molecular docking

*Mobina Haji Abbasi, Eslam Pourbasheer**
Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili, P.O. Box 179, Ardabil, Iran
corresponding author E-mail address: e.pourbasheer@ac.ir

ABSTRACT

In recent years, considerable attention has been directed toward the development of targeted therapeutics for cancer treatment. Previous studies have reported the design and synthesis of triazolo-pyrimidine derivatives as inhibitors of Werner syndrome RecQ helicase (WRN), a critical target in microsatellite instability (MSI) tumors. Inhibition of WRN has been shown to selectively disrupt tumor cell proliferation, highlighting its therapeutic potential. To accurately predict the biological activity of these compounds, three-dimensional quantitative structure–activity relationship (3D-QSAR) models were constructed using CoMFA, CoMFA-Focus, and CoMSIA approaches. Statistical analyses indicated that the CoMFA model exhibited the best predictive performance, with a cross-validated correlation coefficient (Q^2) of 0.508, a training set correlation coefficient (R^2_{train}) of 0.859, and a test set correlation coefficient (R^2_{test}) of 0.661. Contour maps generated from this model identified critical structural features influencing inhibitory potency, providing valuable insights for rational drug design. Guided by these results, a series of new compounds were proposed and further assessed through molecular docking and ADMET analyses. Docking studies confirmed that the designed derivatives were capable of forming key interactions within the WRN binding site, consistent with the predicted quantitative relationships. Moreover, ADMET predictions demonstrated that the compounds possess favorable pharmacokinetic profiles and drug-like characteristics, supporting their potential for oral administration. Overall, the integration of 3D-QSAR modeling, molecular docking, and ADMET evaluation provides a comprehensive framework for the rational design of WRN inhibitors. These findings suggest that triazolo-pyrimidine derivatives represent promising candidates for the targeted therapy of MSI tumors. This study not only highlights the potential of WRN inhibition in cancer treatment but also offers structural insights that may guide future optimization and preclinical development of novel anticancer agents.

Keywords: 3D-QSAR, CoMFA, CoMSIA, MSI tumors, WRN inhibitors, Triazolo-pyrimidine derivatives

References :

[1] Sui, Qibang et al. “Design, synthesis, and structure-activity relationship studies of triazolo-pyrimidine derivatives as WRN inhibitors for the treatment of MSI tumors.” European journal of medicinal chemistry vol. 282 (**2025**): 117039. doi:10.1016/j.ejmech.2024.117039.

[2] Mahmoudzadeh Laki, Reza, and Eslam Pourbasheer. “3D-QSAR Modeling on 2-Pyrimidine Carbohydrazides as Utrophin Modulators for the Treatment of Duchenne Muscular Dystrophy by Combining CoMFA, CoMSIA, and Molecular Docking Studies.” ACS omega vol. 9,23 24707-24720. (**2024**), doi:10.1021/acsomega.4c01225.

[3] Sarfaraz Khalifehlou, M. ,, Pourbasheer, E. ,, Mahmoudzadeh Laki, R. , and Fadaeian, M. . “3D-QSAR study on the DYRK1A inhibitors and design of new compounds by CoMFA and CoMSIA methods”, *Chemical Review and Letters*, vol. 7, 6, (**2024**), 926-941. doi: 10.22034/crl.2024.474579.1410.

Dispersive liquid-liquid microextraction for the extraction of ibrutinib drug from real samples using UV-Vis spectrophotometry and optimization by experimental design

*Sara Kharat Zabardast, Eslam Pourbasheer**
Department of Chemistry,Faculty of Science,University of Mohaghegh Ardabili, Ardabil,Iran
E-mail address: e.pourbasheer@uma.ac.ir

ABSTRACT

Ibrutinib is a kinase inhibitor that works by irreversibly inhibiting Bruton's Tyrosine Kinase (BTK) [1][2]. Ibrutinib is a key treatment for select lymphoid malignancies, including chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL). It achieves significant therapeutic outcomes in blood cancers by blocking the molecular signaling pathways that promote cancer cell growth [3][4]. Given the drug's widespread clinical application and the critical importance of accurate concentration monitoring, the development of a sensitive, precise, and efficient analytical method for its measurement in real samples is absolutely essential. This study presents a simple and rapid method for the preconcentration and determination of Ibrutinib in real samples using dispersive liquid-liquid microextraction (DLLME) [5], which is effectively coupled with ultraviolet-visible (UV-Vis) spectrophotometry. DLLME was chosen for its distinct advantages, including its simplicity, high speed, low consumption of toxic solvents, and overall cost-effectiveness. The influential parameters in DLLME were optimized using both the one-at-a-time method and experimental design with the CCD method. Under these optimized conditions, the method's calibration curve exhibited very good linear behavior. The regression coefficient (R^2) was also highly favorable, which confirms the method's high accuracy. Furthermore, the limit of detection (LOD) and limit of quantification (LOQ) for Ibrutinib yielded very good results, demonstrating the proposed method's excellent sensitivity and low RSD. This advanced and innovative technique was successfully applied for the determination of Ibrutinib in real samples, and the results showed high accuracy and precision. These findings, by relying on the design of experiments, firmly establish the proposed method as a reliable and highly efficient tool for the analysis of trace amounts of Ibrutinib in complex matrices.

Keywords: Dispersive Liquid-Liquid Microextraction(DLLME), Ibrutinib, experimental design, UV-VIS Spectrophotometry,

References:

[1]Schwarzbich M-A, Witzens-Harig M.Ibrutinib. In: Martens UM, editor. Small Molecules in Oncology.Berlin,Heidelberg: Springer Berlin Heidelberg;2014, 259-67.

[2]Nocco S, Andriano T, Bose A, Chilov M, Godwin K, Dranitsaris G, et al. Ibrutinib-associated dermatologic toxicities: A systematic review and meta-analysis.CRITICAL REVIEWS IN ONCOLOGYHEMATOLOGY.2022;174..

[3]K. I. Alzanda, H. R. Younis, B. Qasim, & J. N. Jeber, "Development and validation of a dispersive liquid–liquid microextraction method for metoclopramide analysis in pharmaceuticals and biological samples," Russian Journal of General Chemistry, vol. 94, 2024, 2517–2528.

[4]E. Pourbasheer & S. Sadafi, & M. R. Ganjali, "Dispersive liquid–liquid microextraction for preconcentration and determination of phenytoin in real samples using response surface methodology-high performance liquid chromatography," RSC Advances, vol. 4, 2014, 62190–62196.

[5]Pourbasheer,E.,& Sadafi, S.,& Ganjali, M. R..Dispersive liquid–liquid microextraction for preconcentration and determination of phenytoin in real samples using response surface methodology-high performance liquid chromatography. RSC Advances.2014,4, 62190–62196.

Dispersive liquid-liquid microextraction (DLLME) for extraction and preconcentration of Enzalutamide in real samples using the UV-Vis spectrophotometry

*Sanaz Saber Siyahpoush, Eslam Pourbasheer**
Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili, Ardabil, Iran
E-mail address: e.pourbasheer@uma.ac.ir

ABSTRACT

Prostate cancer (PCa) is the most prevalent cancer and the second leading cause of cancer death among men. Androgen deprivation therapy is typically the first-line treatment[1]. A new chemical entity Enzalutamide (EZA) is second generation androgen receptor inhibitor. It is androgen-receptor antagonist that blocks androgens from binding to the androgen receptor and prevents nuclear translocation and co-activator recruitment of the ligand-receptor complex. Therapeutically, it is indicated for patients with metastatic castration resistant prostate cancer following Docetaxel based chemotherapy[2]. A dispersive liquid-liquid microextraction (DLLME) method was developed for the determination of Enzalutamide (EZA) in real samples by using UV-Vis spectrophotometry. Dispersive liquid–liquid microextraction (DLLME) has emerged as a favoured green pre-treatment technique for drug extraction because it utilizes reduced volumes of organic solvents (µL range). DLLME requires a dispersant and an extractant solvent[3]. Rapid injection of a mixture of dispersant and extraction solvent into an aqueous sample creates a cloudy mixture, which increases the contact surface area between the phases and reduces the time needed for the species to reach equilibrium. After centrifugation, the settled organic phase is analyzed at the maximum wavelength. The optimization was carried out by using experimental design methods. The Central Composite Design method is a simple approach that uses fewer experiments while maintaining high accuracy. The dispersive solvent volume, extraction solvent volume, and sample volume were optimized and under these conditions , the calibration curve for the real sample was linear. The regression coefficient (R^2) was also highly favorable, confirms the high accuracy and linearity of the method. Furthermore, the limit of detection (LOD) and limit of quantification (LOQ) for Enzalutamide were very suitable values, demonstrating the excellent sensitivity of the proposed method. This method was successfully applied for the determination of Enzalutamide.

Keywords Dispersive Liquid-liquid microextraction , UV-Vis spectrophotometry, Enzalutamide, Experimental design

References

[1] Hsiao-Chi Wang , Allen C. Gao , Roger Xia , Chun-Te Wu , Ssu-Wei Hsu , Ching-Hsien Chen, Tsung-Chieh Shih , Inhibition of Galectin-1 and Androgen Receptor Axis Enhances Enzalutamide Treatment in Enzalutamide Resistant Prostate Cancer ,Journal of Pharmaceutical and Biomedical Analysis , 236 (2025).

[2] Khan Z. G. , Patil S.S , Deshmukh P.K , Patil P.O, Validated RP-HPLC method for determination of enzalutamide in bulk drug and pharmaceutical dosage form , (2016).

[3] Pol Clivill´e-Cabr´, Luana M. Rosendo , Francesc Borrull , Carme Aguilar , Marta Calull , A comparative study of SPE- and DLLME-based methods for the determination of opioids and benzodiazepines in urine samples using LC–MS/MS , Microchemical Journal , 208 (2025).

Classification of some opioids using principal component analysis

Morteza Bahram[*], *Reza Dadashi*
Department of Analytical Chemistry, Faculty of Chemistry, Urmia University, Urmia, Iran

[*]Corresponding author (*E-mail address*): m.bahram@urmia.ac.ir

ABSTRACT

Heroin (diacetylmorphine) is a powerful opioid derived from morphine, widely abused for its euphoric and analgesic effects. It is typically administered intravenously, leading to a rapid onset of action and a high potential for addiction. Crack cocaine, a freebase form of cocaine obtained by processing cocaine hydrochloride with a strong base, is usually smoked, delivering intense psychoactive effects almost immediately. Both substances are illicit, highly addictive, and pose significant public health risks [1]**. Their chemical similarities with other adulterants and cutting agents make accurate identification in complex samples a challenging task for forensic laboratories.**

The identification and quantification of illicit drugs like heroin and crack cocaine in unknown samples are critical for forensic analysis and law enforcement. Classification techniques in chemometrics are pivotal for the identification and differentiation of substances in complex datasets [2]**. These methods, including Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Support Vector Machines (SVM), are extensively utilized to analyze spectral data from various analytical instruments** [3]**. By transforming and modeling data, these techniques facilitate the classification of samples based on their chemical fingerprints, enabling accurate identification even in the presence of adulterants or complex matrices.**

In this study, a chemometric model was developed for the classification of heroin and crack cocaine and the detection of suspicious samples. Initially, the recorded GC data were preprocessed and encoded to mitigate peak-shift variations. The model was constructed using 25 pure heroin samples and 20 pure crack samples, and the corresponding outlier data were calculated. The resulting model was subsequently applied to identify and classify unknown or suspicious samples, demonstrating high accuracy and reliability in

distinguishing between the two drug types.

Keywords: Heroin, Crack,Chemometrics, opioid, Classification

References

[1] S.A. Moallem, K. Balali-Mood, M. Balali-Mood, M. Balali-Mood, Opioids and opiates, in: Handb. Drug Interact. A Clin. Forensic Guid., Springer, 2011: pp. 159–191.

[2] M. Cocchi, A. Biancolillo, F. Marini, Chemometric methods for classification and feature selection, in: Compr. Anal. Chem., Elsevier, 2018: pp. 265–299.

[3] L.C. Lee, C.-Y. Liong, A.A. Jemain, Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps, Analyst. 143 (2018) 3526–3539.