MINISTRY OF HEALTH OF UKRAINE
BOGOMOLETS NATIONAL MEDICAL UNIVERSITY

FACULTY FOR TRAINING OF FOREIGN CITIZENS
Department of Analytical, Physical and Colloid Chemistry

Graduate Master's Thesis

«ANALYSIS OF BIOPHARMACEUTICS CLASSIFICATION SYSTEM BY MEANS OF CHEMOMETRIC METHODS»

Created by: student of 5th year,
group 10601Fa
Sedat Acik
Branch of knowledge: 22 «Healthcare»
Specialty: 226 «Pharmacy, industrial pharmacy»

Supervisor: Candidate of Chemical Sciences, Associate Professor Yaroslava Pushkarova

Reviewer: Candidate of Medical Sciences, Associate Professor Vladyslav Polovynka

KYIV-2025

CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

BCS	Biopharmaceutics Classification System
MW	molecular weight
HBA	number of HBond acceptors
HBD	number of HBond donors
MR	molar refractivity
logP	partition coefficient
logS	solubility
BI	Balaban index
MTI	molecular topological index
RotB	number of rotatable bonds
PSA	polar surface area
WI	Wiener index
PNN	Probabilistic Neural Network

INTRODUCTION

Actuality. The analysis of the Biopharmaceutics Classification System (BCS) using chemometric methods is a highly relevant and promising research topic. The BCS categorizes drugs into four classes based on their solubility and permeability. This approach is used by regulatory authorities such as the Food and Drug Administration and European Medicines Agency to simplify the registration of generic drugs. Considering the growing market for generic drugs, optimizing the registration process by accurately classifying substances using BCS is becoming increasingly important. Chemometric methods, such as QSAR/QSPR, Principal Component Analysis, Partial Least Squares, and others, enable quantitative and qualitative assessment of drug properties based on their chemical structure. These methods allow for predicting the BCS class of new compounds or refining the classification of known drugs without conducting experimental studies. Automating and optimizing the classification process helps reduce costs and time associated with the development of new drugs.

Additionally, the rise of personalized medicine and the need for precise drug formulations tailored to individual patient needs further increases the relevance of BCS classification. By improving the accuracy of BCS predictions, chemometric models can also assist in the development of drug formulations with optimal bioavailability for specific populations, enhancing therapeutic efficacy and safety.

Predicting properties using chemometric models can minimize the number of in vitro and in vivo experiments required, streamlining the drug development pipeline. Due to the increasing volume of data in the pharmaceutical industry, chemometric methods are gaining popularity for their ability to process large datasets. Moreover, the integration of machine learning and artificial intelligence in chemometrics opens up new avenues for refining predictive accuracy and uncovering previously unknown patterns in drug properties. This combination is expected to further accelerate drug discovery, reduce costs, and enhance the precision of bioavailability predictions.

Furthermore, as regulatory bodies are increasingly emphasizing the importance of early-stage in silico predictions in drug approval processes, chemometric methods aligned with BCS provide a valuable tool for regulatory compliance. Combining BCS with chemometric methods creates new opportunities for predicting bioavailability and developing pharmaceuticals, making the topic highly relevant in both research and practical applications. This integration has the potential to transform drug development, ensuring that safe, effective, and accessible medications reach patients faster and more efficiently.

Aim and tasks of research. The aim of this study is to develop an accurate drug classification model according to the Biopharmaceutics Classification System using chemometric methods.

Tasks of research:

− calculate molecular descriptors using software ChemOffice for 122 drug compounds;

− identify sufficient descriptors for accurate drug classification according to the Biopharmaceutics Classification System using the Kruskal-Wallis test;

− to optimize the architecture of a Probabilistic Neural Network for accurate drug classification according to the Biopharmaceutics Classification System;

− predict the BCS class for 23 compounds that have been assigned to multiple BCS classes in different scientific articles.

Research methods: Kruskal-Wallis test, Probabilistic Neural Network.

The software package MATLAB R2024b and ChemOffice 2020 (trial license) were used in this work.

Novelty and significance of the results. The novelty of this study lies in the application of advanced chemometric methods, particularly the use of a Probabilistic Neural Network, for the accurate classification of drugs according to the Biopharmaceutics Classification System. Unlike traditional methods, this approach integrates molecular descriptors that are computationally derived, allowing for a more efficient and reliable drug classification process. The research identifies four key descriptors (number of HBond donors, partition coefficient,

solubility, and polar surface area) that are sufficient for accurate classification, which has not been extensively explored in the context of Biopharmaceutics Classification System.

The significance of the results is twofold:

1. Practical application: the developed model provides an efficient tool for drug developers and pharmaceutical researchers to predict the BCS class of a compound, potentially accelerating the drug development process by streamlining the identification of compounds with optimal bioavailability profiles.

2. Scientific contribution: the study advances the use of chemometric techniques in pharmaceutical sciences, demonstrating their potential to improve predictive models for drug classification. This can have broader applications in other areas of pharmaceutical research, such as Quantitative Structure-Activity Relationship modeling and drug formulation.

Overall, the study contributes to the field by offering a robust, data-driven method for BCS classification, which could aid in the development of more effective drug therapies.

Approbation of research results. The results of this work were presented at Scientific and Practical Conference «Innovations in medicine and pharmacy: contribution of young scientists», February 28, 2025, Kyiv, Bogomolets National Medical University.

Publications. Abstract of report: Sedat Acik, Yaroslava Pushkarova. Analysis of Biopharmaceuticals classification system by means of chemometric methods. Scientific and Practical Conference «Innovations in medicine and pharmacy: contribution of young scientists», February 28, 2025, Kyiv. Ukrainian scientific medical youth journal, 2025, Supplement 1 (153), P. 162.

Structure of work: 41 pages, 2 sections, supplementary information, 44 references.

MAIN PART

1. LITERATURE REVIEW

1.1. Biopharmaceutics Classification System: the main principles

Formulators often prefer oral administration as a method of drug delivery, and it remains the most prevalent approach within the field of drug delivery technologies. Despite its widespread use, this route faces challenges related to absorption and bioavailability within the gastrointestinal tract. When an oral dosage form is administered, the drug is released, dissolving in the surrounding gastrointestinal fluid to form a solution. This dissolution process is constrained by the drug's solubility. Once the drug is dissolved in a liquid medium, it can traverse the cellular membranes lining the gastrointestinal tract. However, this stage of the process is limited by the drug's permeability. Following this, the drug is absorbed into systemic circulation. In essence, the bioavailability and absorption of an orally administered drug are largely influenced by the drug's solubility and permeability characteristics.

The Biopharmaceutics Classification System provides a systematic, scientifically grounded approach to categorizing drug substances based on their aqueous solubility in relation to dose and their intestinal permeability. By also considering the dissolution of the dosage form, the BCS evaluates three crucial factors that regulate the rate and extent of drug absorption from solid oral dosage forms: dissolution, solubilization, and intestinal permeability [1-4].

The Biopharmaceutics Classification System is a scientifically established model used to categorize Active Pharmaceutical Ingredient substances according to their aqueous solubility and intestinal permeability characteristics. The concept of the BCS was introduced by Amidon and his colleagues, who based it on a diagram illustrating the relationship between solubility and permeability. This foundational

idea led to the proposal of waiving in vivo bioequivalence studies for specific oral immediate-release formulations.

The BCS principle put forward the notion of eliminating the need for in vivo bioequivalence studies for certain immediate-release formulations, offering a more efficient approach to assessing bioequivalence. The BCS has gained international recognition among technical industries, academic institutions, and regulatory authorities. The core concept of the BCS is that if two pharmaceutical products exhibit the same concentration profile throughout the gastrointestinal tract, they will produce identical plasma concentration profiles after oral administration. This relationship can be mathematically represented and analyzed [5-7].

The Biopharmaceutics Classification System is based on three fundamental principles: solubility, permeability, and dissolution. These principles are essential for the accurate classification of drugs within the BCS framework.

Solubility: A drug substance is considered highly soluble if its highest dose strength can dissolve in 250 mL or less of water across a pH range of 1 to 7.5 at a temperature of 37 °C.

Permeability: A drug is deemed highly permeable when the extent of its absorption in humans exceeds 90% of the administered dose, as determined by a mass-balance method or by comparison with a reference intravenous dose.

Dissolution: A drug substance is classified as rapidly dissolving if at least 85% of the given amount of the drug dissolves within 30 minutes when tested using USP Apparatus 1 or 2 in a buffer solution volume not exceeding 900 mL.

These principles facilitate the reliable determination of a drug's BCS classification, allowing for streamlined bioequivalence assessments and regulatory decisions [8-12].

According to the Biopharmaceutics Classification System, drugs are categorized into four distinct classes based on their intestinal permeability and solubility characteristics. This classification is grounded in fundamental principles such as solubility, dissolution, and permeability, all of which significantly influence the absorption process. Among these classes, Class I substances exhibit

the highest absorption rates due to their favorable properties. In contrast, Class II drugs have limited solubility, which can affect their absorption. Class III drugs are characterized by limited permeability, while Class IV drugs demonstrate poor absorption due to both low solubility and low permeability [13].

The classification can be summarized as follows:

- Class I: High Solubility – High Permeability
- Class II: Low Solubility – High Permeability
- Class III: High Solubility – Low Permeability
- Class IV: Low Solubility – Low Permeability

1.2. Chemometric methods: the main methods, their tasks and using

Chemometrics is a scientific discipline that leverages mathematical and statistical techniques to create or choose the most effective measurement procedures and experimental designs. Its primary goal is to enhance both the quality and quantity of chemical information derived from chemical data. This interdisciplinary field merges chemistry, mathematics, and computer science to extract valuable insights from chemical datasets. Chemometrics utilizes multivariate analysis, pattern recognition, data mining, and other computational methods to address challenges in chemistry and associated domains.

Chemometrics can be described as the application of statistical techniques to chemical analysis data, aiming to deepen the understanding of chemical systems. One of its key aspects is the extraction of chemical information from collected data, which often originates from analytical instruments.

In practice, chemometrics involves the use of statistical models to analyze data, identify patterns, and make reliable predictions. It is widely applied in areas like spectroscopy, chromatography, and process control, enabling the optimization of experimental conditions and improvement of analytical accuracy while simultaneously reducing both analysis time and associated costs.

In the realm of analytical chemistry, chemometrics proves invaluable as it facilitates the extraction of meaningful information from extensive and complex datasets, thus supporting more informed decision-making. It plays a crucial role in quality control, method development, and the discovery of new chemical knowledge [14-17].

Principal Component Analysis is a widely utilized chemometric technique aimed at data simplification. It transforms the original data into a new set of variables known as principal components, which are orthogonal and uncorrelated with each other.

The primary objective of Principal Component Analysis is dimensionality reduction, which helps decrease the number of variables within a data set while preserving as much variability as possible. This method is especially valuable for researchers when identifying patterns and visualizing data, as it aids in recognizing underlying structures.

Principal Component Analysis is commonly employed for exploratory data analysis to reduce data complexity and reveal hidden patterns within the data [18, 19].

Partial Least Squares Regression is considered one of the most significant methods in chemometrics, used to model the relationship between input and output variables. This technique is particularly advantageous when dealing with datasets containing numerous predictors that are often highly collinear.

The primary purpose of Partial Least Squares Regression is to construct a predictive model that effectively represents the relationships among variables. It is extensively applied in Quantitative Structure-Activity Relationship modeling to estimate activity based on molecular structure. Additionally, Partial Least Squares Regression is frequently employed in spectroscopy for analyzing spectral data [20, 21].

Multivariate Curve Resolution is a method utilized to decompose complex datasets into the spectra and concentrations of pure components. The main objective of Multivariate Curve Resolution is to break down intricate mixtures into

individual components, making it especially valuable for analyzing spectroscopic data. In the field of food science, Multivariate Curve Resolution is commonly applied to identify and examine composite mixtures of food ingredients. This method offers valuable insights into the composition of complex mixtures, proving essential in a wide range of analytical applications [22, 23].

Artificial Neural Networks are computational models inspired by the structure and function of the human brain, designed to aid in modeling, pattern recognition, and classification tasks. Artificial Neural Networks are particularly effective when dealing with non-linear relationships and handling complex datasets. Their primary function is pattern recognition, allowing the detection of intricate patterns within data. Additionally, Artificial Neural Network algorithms can perform classification tasks.

In chemometrics, Artificial Neural Networks are utilized for analyzing complex data, proving useful in various applications such as spectroscopy, process analysis, control, and multivariate calibration. Their ability to model complex relationships and make accurate predictions makes them highly valuable in chemometric analysis.

Artificial Neural Networks are computational frameworks that mimic brain-like neural networks, facilitating learning processes. These networks consist of interconnected artificial neurons that work together to process information in a connectionist manner. Artificial Neural Networks are specifically designed to identify patterns and acquire knowledge from data. They function similarly to the human brain, aiding in information processing by learning from experience, making decisions, and recognizing complex patterns. Their approach to connectionist computing is instrumental in interpreting and analyzing complex information [24-26].

Among the various architectures of Artificial Neural Networks, the Feedforward Neural Network stands out as the simplest type, designed for specific problem types. In this model, data flows in a single direction, moving from input to output without any feedback loops. Due to their straightforward structure,

Feedforward Neural Networks are well-suited for tasks such as image classification and pattern recognition.

Convolutional Neural Networks are specialized neural networks structured to process data with a grid-like topology, such as images. They employ convolutional layers to autonomously learn spatial feature hierarchies, making them particularly effective in image and video recognition, object detection, and segmentation tasks.

Recurrent Neural Networks are designed to handle sequential data and tasks where the current output depends on previous inputs. Their unique architecture includes directed cycles, allowing them to maintain a memory of prior information. This characteristic makes Recurrent Neural Networks especially useful in applications like time-series forecasting and natural language processing, where context from past data is essential [27].

Types of learning process [24-27]:

− supervised learning: in this approach, the data is labeled, meaning that the input-output pairs are known. The primary goal is to learn the relationship between inputs and their corresponding outputs while minimizing the error when outputs overlap.

− unsupervised learning: in this type of learning, the network independently identifies patterns and structures within the data. It is useful for tasks such as clustering, dimensionality reduction, and error detection.

− reinforcement learning: the network learns by interacting with an environment to obtain feedback in the form of rewards or punishments. This feedback guides the network towards making better decisions over time.

− backpropagation: this algorithm is widely used for training artificial neural networks. It calculates the gradient of the loss function concerning each weight by applying the chain rule. The weights are then adjusted in the opposite direction of the gradient to reduce the loss.

Conclusions to section 1

The Biopharmaceutics Classification System is a framework used to classify drugs based on their solubility and intestinal permeability. It plays a crucial role in drug development, guiding the formulation and regulatory approval processes.

Chemometric methods are essential tools in modern chemistry and related fields, providing robust techniques for analyzing complex chemical data. The primary aim of chemometrics is to extract useful information from experimental data using mathematical and statistical methods.

2. EXPERIMENTAL PART

2.1. Data set

Studied data set included 122 drug compounds. They form training sub-set for developing the classification model according to the Biopharmaceutics Classification System.

For each drug compound we have calculated 11 physical-chemical and topological descriptors by means of ChemOffice software [28]:

1) molecular weight (MW);

2) number of HBond acceptors (HBA);

3) number of HBond donors (HBD);

4) molar refractivity (MR);

5) partition coefficient (logP);

6) solubility (logS);

7) Balaban index (BI);

8) molecular topological index (MTI);

9) number of rotatable bonds (RotB);

10) polar surface area (PSA);

11) Wiener index (WI).

List of 122 drug compounds, their classification according to the Biopharmaceutics Classification System, values of eleven molecular descriptors are presented in Table 2.1.

Short description of each class of drugs according to the Biopharmaceutics Classification System [29-33]:

− BCS Class I drug substances dissolve quickly in aqueous media, which positively impacts their bioavailability. These substances have high solubility and are easily absorbed from the gastrointestinal tract, allowing them to quickly enter the systemic circulation from the site of administration. The rapid absorption of these substances is crucial for their effectiveness. For drug substances in this class,

simple formulations with immediate release are generally sufficient due to their high solubility and permeability. Examples of this class include capsules and tablets with a low content of excipients.

− BCS Class II drug substances have limited solubility in water but are effectively absorbed in the gastrointestinal tract. This means that although these drugs are not easily dissolved or solubilized in aqueous environments, they are still well absorbed when ingested. The relationship between solubility and absorption significantly impacts the bioavailability of these compounds, which in turn determines their therapeutic effectiveness. For Class II drugs, the primary focus is on improving solubility. Techniques such as solid dispersions, micronization, and nanosizing are utilized, along with solubilizing agents like cyclodextrins and surfactants. Furthermore, lipid-based formulations can also enhance the bioavailability of these drugs. Examples of such formulations include self-emulsifying drug delivery systems and self-microemulsifying drug delivery systems.

− BCS Class III drug substances are characterized by their high solubility in the gastrointestinal tract. However, despite their ability to dissolve easily in aqueous environments, these substances face challenges related to absorption. Although they have excellent solubility, they do not efficiently permeate the intestinal membranes to enter systemic circulation. To improve the bioavailability of Class III drugs, various strategies are employed, particularly the use of permeability enhancers to facilitate better absorption. A common approach involves the use of surfactants as permeation enhancers. These surfactants increase drug permeability by interacting with cell membranes, making it easier for the drug to pass through the gastrointestinal tract.

− BCS Class IV drug substances are characterized by both poor solubility and poor absorption in the gastrointestinal tract. These substances generally have limited dissolution in aqueous environments and are not easily permeable through the gastrointestinal tract membranes. As a result, Class IV drugs have significant limitations in terms of effectiveness. To address these challenges related to

solubility and permeability, various enhancement techniques are employed, such as solid dispersions and nanosizing. These methods increase the surface area of the drug, which helps improve its dissolution in biological media. Other strategies include the use of nanoparticles, liposomes, and targeted drug delivery systems, all of which aim to improve the bioavailability of these drugs.

Table 2.1. List of 122 drug compounds and values of their eleven molecular descriptors

№	Drug compound	BCS Class [34]	Descriptor										
			MW	HBA	HBD	MR	logP	logS	BI	MTI	RotB	PSA	WI
1	Amlodipine	1	408,879	5	2	10,865	1,840	-3,675	981585	12808	10	99,88	1836
2	Bisoprolol	1	325,449	5	2	9,235	2,198	-2,293	861539	12108	12	59,95	1650
3	Donepezil	1	379,500	4	0	11,121	3,600	-4,609	801263	17668	6	38,77	2333
4	Doxazosin	1	451,483	9	1	12,039	2,096	-4,628	1433183	25474	5	111,21	3556
5	Doxepin	1	279,383	2	0	8,841	3,983	-4,310	210631	6973	3	12,47	882
6	Enalapril	1	376,453	4	2	10,155	2,046	-3,044	989295	14260	11	95,94	1982
7	Ephedrine	1	165,236	2	2	5,066	1,390	-1,434	28816	1604	3	32,26	202
8	Ergonovine	1	325,412	4	3	9,454	0,902	-2,696	323993	9446	4	64,60	1263
9	Ethynyl estradiol	1	296,410	2	2	8,753	4,192	-4,393	204957	7264	1	40,46	941
10	Ethosuximide	1	141,170	2	1	3,687	0,080	-0,644	10689	812	1	46,17	108
11	Fluoxetine	1	309,332	5	1	8,087	4,616	-4,707	383228	7877	7	21,26	1148
12	Glucose	1	180,156	6	5	3,761	-3,292	1,990	53802	1252	5	118,22	206
13	Imipramine	1	280,415	2	0	9,006	3,952	-4,339	210631	7074	4	6,48	882
14	Ketorolac	1	255,273	3	1	6,995	2,118	-3,048	136862	5117	3	57,61	693
15	Labetalol	1	328,412	4	4	9,526	2,524	-3,708	636957	11893	8	95,58	1607

16	Levodopa	1	197,190	4	4	4,944	-1,757	0,161	62340	2160	3	103,78	321
17	Lomefloxacin	1	351,354	7	2	8,869	2,840	-4,333	454900	9316	3	72,88	1362
18	Loratadine	1	382,888	2	0	10,835	3,827	-5,250	567281	12972	3	41,90	1772
19	Metoprolol	1	267,369	4	2	7,691	1,643	-1,819	322987	6707	9	50,72	906
20	Metronidazole	1	171,156	4	1	4,063	-0,123	-0,748	27473	1288	3	87,64	193
21	Mirtazapine	1	265,360	3	0	8,154	2,934	-3,735	125403	5476	0	18,84	687
22	Nicotinamide	1	122,127	2	1	3,346	-0,341	-0,479	7074	679	1	55,45	88
23	Norethisterone	1	298,426	2	1	8,780	3,140	-3,597	204957	7264	1	37,30	941
24	Ondansetron	1	293,370	4	0	8,641	2,508	-3,966	217302	7706	2	35,91	997
25	Phenobarbital	1	232,239	3	2	6,139	1,327	-2,589	92531	3322	2	75,27	458
26	Phenylalanine	1	165,192	2	2	4,638	-1,371	0,316	30269	1573	3	63,32	212
27	Prednisolone	1	360,450	5	3	9,840	1,009	-2,525	425311	10240	2	94,83	1425
28	Primaquine	1	259,353	4	2	7,839	2,082	-2,935	181736	5550	6	59,64	726
29	Proguanil	1	253,734	5	3	7,140	1,872	-3,281	173976	4376	4	88,79	608
30	Propranolol	1	259,349	3	2	7,834	3,041	-3,174	198625	6105	6	41,49	792
31	Pyridoxine	1	169,180	4	3	4,328	-0,518	-0,211	26520	1314	2	73,05	186
32	Quinapril	1	438,524	4	2	12,203	3,139	-4,375	1650793	22541	11	95,94	3064
33	Quinidine	1	324,424	4	1	9,562	2,584	-3,400	330149	9833	4	45,06	1286

34	Ramipril	1	416,518	4	2	11,369	2,796	-3,815	1210619	18586	11	95,94	2546
35	Riboflavin	1	376,369	8	5	9,505	-1,874	-1,091	658031	11526	5	155,05	1698
36	Salbutamol	1	239,315	4	4	6,763	1,443	-1,881	160268	4090	5	72,72	560
37	Salicylic acid	1	138,122	2	2	3,494	2,208	-2,133	11308	822	1	57,53	114
38	Sertraline	1	306,230	1	1	8,693	5,114	-5,534	167768	5628	2	12,03	770
39	Sildenafil	1	474,580	8	1	12,585	1,544	-4,491	1451240	22672	7	106,91	3086
40	Sotalol	1	272,363	4	3	7,234	1,079	-2,399	226065	5254	6	78,43	706
41	Terbinafin	1	291,438	1	0	9,771	5,917	-5,550	425236	10280	6	3,24	1273
42	Theophylline	1	180,167	4	1	4,527	-0,746	-0,783	25429	1509	0	65,01	211
43	Timolol	1	316,420	7	2	8,197	1,309	-2,271	324259	7477	7	78,68	1063
44	Tramadol	1	263,381	3	1	7,824	2,379	-2,578	167956	5141	4	32,70	670
45	Venlafaxine	1	277,408	3	1	8,288	2,741	-2,860	219265	6061	5	32,70	792
46	Zidovudine	1	267,245	6	2	6,356	-0,884	-1,192	175676	4676	3	127,63	701
47	Zolpidem	1	307,397	3	0	9,293	2,263	-3,652	323265	8804	4	35,91	1137
48	Azithromycin	2	748,996	13	5	19,727	1,985	-4,083	13032353	65841	7	180,08	9356
49	Celecoxib	2	381,373	7	1	9,145	3,832	-5,812	595269	11379	4	75,76	1654
50	Cisapride	2	465,950	7	2	12,232	3,271	-5,132	1915042	24586	10	86,05	3560
51	Clopidogrel	2	321,819	2	0	8,663	2,510	-3,598	207525	6212	4	29,54	867

52	Danazol	2	337,463	3	1	9,795	4,188	-4,664	311044	9947	1	41,82	1299
53	Diflunisal	2	250,201	4	2	6,037	4,471	-4,734	134867	3924	2	57,53	597
54	Dipyridamole	2	504,636	12	4	13,613	1,449	-3,733	2027221	25976	12	143,32	3652
55	Fenoprofen	2	242,274	2	1	6,933	3,148	-3,145	146840	4864	4	46,53	650
56	Flurbiprofen	2	244,265	2	1	6,796	3,761	-3,871	141234	4581	3	37,30	626
57	Glipizide	2	445,538	6	3	11,752	3,693	-5,980	1735550	25212	10	129,09	3430
58	Ibuprofen	2	206,285	1	1	6,124	3,646	-3,119	89861	3076	4	37,30	404
59	Indomethacin	2	357,790	3	1	9,505	3,350	-4,590	474843	9936	5	66,84	1424
60	Irbesartan	2	428,540	6	1	12,314	4,730	-6,568	1188964	23576	7	81,78	3125
61	Itraconazole	2	705,641	10	0	18,815	6,527	-9,752	8072800	85682	11	98,04	12096
62	Ketoconazole	2	531,434	7	0	13,892	4,063	-6,421	2228313	32881	8	66,84	4692
63	Lansoprazole	2	369,362	8	1	8,789	1,493	-4,141	545169	10846	6	63,05	1631
64	Lorazepam	2	321,157	3	2	8,297	3,479	-4,796	196116	5637	1	61,69	819
65	Lovastatin	2	404,547	3	1	11,256	4,440	-4,789	999139	16462	7	72,83	2246
66	Mefenamic acid	2	241,290	2	2	7,149	3,994	-4,347	136170	4586	3	49,33	602
67	Montelukast	2	586,187	3	2	17,462	7,861	-9,013	3867382	46793	12	69,89	6351
68	Nalidixic acid	2	232,239	4	1	6,212	2,173	-2,882	94940	3397	2	69,97	470
69	Naproxen	2	230,263	2	1	6,574	2,829	-3,008	106914	3894	3	46,53	530

70	Nevirapine	2	266,304	4	1	7,555	2,468	-3,696	123447	5188	1	57,06	676
71	Nitrofurantoin	2	238,159	5	1	5,496	-0,095	-1,682	117288	3681	3	122,81	580
72	Ofloxacin	2	361,373	7	1	9,293	2,210	-3,848	442670	10338	2	73,32	1484
73	Oxaprozin	2	293,322	3	1	8,294	3,613	-4,230	277217	7923	5	58,89	1063
74	Phenazopyridine	2	213,244	5	2	6,742	2,687	-3,827	87196	3711	2	89,12	485
75	Phenytoin	2	252,273	2	2	7,223	2,252	-3,487	121875	4679	2	58,20	617
76	Raloxifene	2	473,587	5	2	13,605	5,162	-6,877	1633844	28431	7	70,00	3843
77	Rifampicin	2	822,953	14	6	22,008	2,414	-7,100	17103588	98085	5	220,15	13890
78	Risperidone	2	410,493	6	0	11,277	2,426	-4,609	941544	20311	4	57,50	2793
79	Rofecoxib	2	314,355	3	0	8,225	3,190	-4,713	277278	7984	3	60,44	1061
80	Simvastatin	2	418,574	3	1	11,720	4,649	-4,963	1159529	17942	7	72,83	2440
81	Sulfamethoxazole	2	253,276	5	2	6,276	0,690	-2,829	108079	3926	3	93,78	535
82	Sulindac	2	356,411	3	1	9,878	3,504	-5,585	506021	10917	4	54,37	1517
83	Tamoxifen	2	371,524	2	0	12,070	7,076	-7,047	887252	16848	8	12,47	2141
84	Tolmetin	2	257,289	3	1	7,173	2,622	-3,268	184391	5410	4	57,61	736
85	Acebutolol	3	336,432	5	3	9,369	1,678	-2,589	891316	11420	11	87,66	1568
86	Alendronic acid	3	249,096	4	6	4,748	-2,802	1,075	101097	1756	5	161,31	281
87	Allopurinol	3	136,114	4	2	3,386	-0,485	-0,796	7659	777	0	65,85	105

88	Ascorbic acid	3	176,124	5	4	3,635	-0,626	-0,294	26787	1156	2	107,22	188
89	Atenolol	3	266,341	4	3	7,478	0,460	-1,497	317560	6458	8	84,58	890
90	Biperiden	3	311,469	2	1	9,610	4,034	-4,300	270410	9078	5	23,47	1142
91	Cefaclor	3	367,804	4	3	9,223	0,406	-2,826	427499	9309	5	112,73	1383
92	Cefazolin	3	454,498	9	2	10,791	-0,844	-2,986	904654	16087	8	151,75	2468
93	Chloramphenicol	3	323,126	4	3	7,313	0,669	-2,133	348207	5269	7	121,37	880
94	Cimetidine	3	252,340	6	3	6,903	1,140	-2,483	190251	4813	8	84,60	664
95	Codeine	3	299,370	4	1	8,347	1,087	-2,144	155703	6202	1	41,93	824
96	Colchicine	3	399,443	6	1	10,859	1,510	-3,370	864522	13802	6	83,09	1944
97	Didanosine	3	236,231	6	2	5,834	-0,103	-1,451	80271	3497	2	86,52	502
98	Ergocalciferol	3	396,659	1	1	12,987	7,154	-6,555	1078625	19222	5	20,23	2428
99	Ergotamine	3	581,673	7	3	16,105	2,053	-5,556	2932081	45947	5	114,45	6191
100	Fexofenadine	3	501,667	4	3	14,891	5,597	-6,603	3058708	39158	10	81,00	5214
101	Folinic acid	3	473,446	9	7	11,744	-1,582	-2,962	2501834	27473	11	215,55	4127
102	Gabapentin	3	171,240	2	2	4,732	1,327	-1,562	27510	1432	3	63,32	193
103	Hydralazine	3	160,180	4	2	4,692	0,671	-1,787	18782	1417	1	62,77	182
104	Levetiracetam	3	170,212	2	1	4,519	-0,693	-0,378	27106	1398	3	63,40	190
105	Levothyroxine	3	776,874	4	3	12,681	3,965	-5,842	569130	8837	5	92,78	1438

106	Lisinopril	3	405,495	5	4	10,988	1,141	-2,838	1355521	16912	13	132,96	2362
107	Losartan	3	422,917	7	2	11,664	4,846	-6,716	1045260	19332	8	84,94	2665
108	Metformin	3	129,167	5	4	3,521	0,607	-1,388	13687	716	3	88,99	96
109	Nadolol	3	309,406	5	4	8,594	1,201	-2,049	390186	8518	6	81,95	1168
110	Penicillamine	3	149,208	2	3	3,860	-1,862	0,942	12267	591	2	63,32	86
111	Propylthiouracil	3	170,230	1	1	4,940	0,477	-1,556	18909	1161	2	32,34	158
112	Pyridostigmine	3	181,214	1	0	4,979	0,306	-1,254	43898	2003	3	32,55	262
113	Ranitidine	3	314,404	5	2	8,651	1,301	-2,613	535958	8479	10	88,34	1227
114	Reserpine	3	608,688	9	1	16,109	3,332	-5,645	4206442	49046	10	114,02	6905
115	Terazosin	3	387,440	8	1	10,303	0,960	-3,188	754065	15647	5	101,98	2195
116	Topiramate	3	339,359	7	1	7,371	-0,602	-1,348	247579	6440	3	115,54	948
117	Valsartan	3	435,528	6	2	12,202	4,590	-6,087	1682497	22840	11	106,72	3122
118	Zalcitabine	3	211,221	4	2	5,296	-1,394	-0,380	58431	2583	2	88,15	369
119	Acetazolamide	4	222,237	5	2	4,649	-0,997	-1,300	42977	1722	3	113,98	257
120	Azathioprine	4	277,262	7	1	6,784	0,720	-2,698	133848	4562	3	116,52	676
121	Cefixime	4	453,444	8	4	10,844	-0,413	-3,100	1216151	16510	9	183,98	2560
122	Oxcarbazepine	4	252,273	2	1	7,223	1,495	-3,017	117397	4490	1	63,40	593

2.2. Assessment of informativeness of molecular descriptors

The Kruskal-Wallis test is used to compare groups or classes of samples. If the calculated value of the test is greater than the critical value, the tested parameter significantly changes depending on the solvent classes; otherwise, there are no statistically significant inter-class differences for the tested parameter [35].

In this study, the Kruskal-Wallis test was used to determine the molecular descriptors whose values have the greatest influence on classifying solvents into different classes. The Kruskal-Wallis test was calculated for 122 solvents characterized by 11 molecular descriptors, using the software package Matlab R2024b [36], and the results are presented in Table 2.2. The critical value of χ^2 at a significance level of 5% is 7,81 (with 3 degrees of freedom).

Table 2.2. Results of the Kruskal-Wallis test calculation for 11 molecular descriptors

	Descriptor										
	MW	HBA	HBD	MR	logP	logS	BI	MTI	RotB	PSA	WI
χ^2	5,91	4,41	**11,35**	5,75	**29,64**	**28,46**	5,08	6,26	2,11	**13,66**	5,88

It has been established that the classification of drug compounds according to the Biopharmaceutics Classification System is most significantly influenced by the following molecular descriptors:

1) number of HBond donors (HBD);

2) partition coefficient (logP);

3) solubility (logS);

4) polar surface area (PSA).

For other descriptors calculated value of χ^2 is lower than critical value. So molecular weight, number of HBond acceptors, molar refractivity, Balaban index, molecular topological index, number of rotatable bonds and Wiener index are not

informativeness for classification of drug compounds according to the Biopharmaceutics Classification System.

2.3. Optimizing the architecture of a Probabilistic Neural Network

A Probabilistic Neural Network (PNN) is a type of neural network based on Bayesian theory and kernel functions primarily used for classification tasks. The architecture of a PNN consists of several layers. The input layer takes feature vectors as input and each neuron in this layer corresponds to a feature. The pattern layer calculates the probability density function for each class using a Gaussian kernel function where each neuron represents a training sample. The summation layer sums the outputs from the pattern layer for each class to estimate the likelihood. The output layer uses the maximum likelihood principle to classify the input into one of the target classes. PNNs are characterized by fast training since they involve simple calculations without iterative optimization. They are capable of forming nonlinear decision boundaries due to the Gaussian function in the pattern layer and they provide probabilistic output indicating the probability of belonging to each class. Although PNNs are known for accuracy and quick training they can be computationally intensive during prediction due to the large number of neurons in the pattern layer.

In a Probabilistic Neural Network, the spread parameter (also known as the smoothing parameter or sigma) plays a crucial role in determining the shape and width of the Gaussian function used in the pattern layer. The spread parameter controls the smoothness of the probability density function estimates. A small spread value results in a narrow Gaussian curve, making the network sensitive to small variations and prone to overfitting. A large spread value results in a wider Gaussian curve, leading to smoother decision boundaries but potentially causing underfitting. Selecting an appropriate spread value is essential for achieving a good balance between model generalization and accuracy [37, 38].

Set of 122 drug compunds, which are characterized by four descriptors (number of HBond donors; partition coefficient; solubility; polar surface area) was used for determination of effective architecture of Probabilistic Neural Network for accurate drug classification according to the Biopharmaceutics Classification System. The PNN was built using the software package Matlab R2024b [36].

Part of incorrect classified drug compounds at different spread values is listed in Table 2.3. The proportion of incorrectly classified drug compounds was calculated as the number of incorrectly classified drug molecules divided by the total number of drug compounds.

Table 2.3. Results of Probabilistic Neural Network training

Spread value	Part of incorrect classified drug compounds, %
0,1	0,0
0,2	0,0
0,3	0,0
0,4	0,0
0,5	0,0
0,6	0,0
0,7	0,8
0,8	1,6
0,9	1,6
1,0	1,6

So, the correct training of probabilistic neural network was archived at values of spread parameter from 0,1 to 0,6 (there are no classification errors). One drug compound (Norethisterone) was incorrect classified at value of spread parameter 0,7. Two drug compounds (Norethisterone and Ketorolac) were incorrect classified at values of spread parameter from 0,8 to 1,0.

To check the correctness of the proposed model, 122 drug compounds was divided on two sub sets: training (104 drug compounds, 85%) and testing (18 drug

compounds, 15%). Testing sub set includes «new» drug compounds for neural network. The set of molecular descriptors was unchanged: number of HBond donors; partition coefficient; solubility; polar surface area. The list of testing sub set and results of prediction of their classification according to the Biopharmaceutics Classification System are presented in Table 2.4.

Correct classes prediction for 18 drug compounds of testing sub set was observed at value of spread parameter 0,1. One drug compound (Zidovudine) was incorrect classified at values of spread parameter from 0,2 to 1,0. That is why value of spread parameter 0,1 is recommended as effective value for drug classification according to the Biopharmaceutics Classification System.

Table 2.4. Results of classes prediction according to the Biopharmaceutics Classification System for testing sub set

Drug compound	Correct class	Spread value 0,1	Spread value 0,2−1,0
		Predicted class	Predicted class
Tramadol	1	1	1
Zolpidem	1	1	1
Bisoprolol	1	1	1
Venlafaxine	1	1	1
Zidovudine	**1**	**1**	**2**
Doxepin	1	1	1
Enalapril	1	1	1
Rofecoxib	2	2	2
Simvastatin	2	2	2
Celecoxib	2	2	2
Lorazepam	2	2	2
Naproxen	2	2	2
Cimetidine	3	3	3
Risperidone	2	2	2

Acebutolol	3	3	3
Atenolol	3	3	3
Ergotamine	3	3	3
Nevirapine	2	2	2

2.4. Evaluation of predictive ability of the proposed model

There are 23 compounds which had been assigned to multiple BCS classes in different scientific articles. The fact that some compounds are assigned to multiple BCS classes in different scientific articles can be explained by several factors:

1. Variability in experimental conditions: solubility and permeability measurements can vary depending on the experimental setup, pH conditions, solvents used, and other factors; different studies might use slightly different methodologies, leading to discrepancies in BCS classification.

2. Differences in data sources: some studies might use in vitro data, while others rely on in vivo data or computational predictions; these differences can result in varying classification outcomes.

3. Polymorphism and drug formulations: a single compound can exist in different polymorphic forms or be formulated differently, which may alter its solubility or permeability, thereby influencing its BCS classification.

4. Biological variability: differences in species, individual variability, or pathological conditions can affect the absorption and solubility profiles, leading to inconsistent classification across studies.

5. Revised data and updated methods: as analytical techniques improve and more accurate data become available, older classifications may be revised, leading to inconsistencies when comparing newer and older studies.

6. Borderline cases: some compounds may have solubility and permeability values that are close to the threshold between two BCS classes, making them

susceptible to being categorized differently depending on minor variations in data or interpretation.

The list of compounds sorted into two BCS classes [34, 39-44] is presented in Table 2.5. Also Table 2.5 includes results of evaluation of predictive ability of the proposed model (PNN with value of spread parameter 0,1) and values of four descriptors (number of HBond donors; partition coefficient; solubility; polar surface area), which are the most informativeness according to the Kruskal-Wallis test.

Table 2.5. Results of evaluation of predictive ability of the proposed model

Drug compound	Descriptor				Possible classes	Predicted class
	HBD	LogP	logS	PSA		
Amitriptyline	0	4,932	-5,102	3,24	1/2	1
Chlorpromazine	0	4,845	-5,397	6,48	1/2	1
Digoxin	6	2,185	-5,232	203,06	1/2	1
Valproic acid	1	2,666	-1,882	37,30	1/2	2
Warfarin	1	2,985	-3,934	63,60	1/2	2
Acetylsalicylic acid	1	1,443	-1,724	63,60	1/3	3
Captopril	2	0,581	-1,157	57,61	1/3	1
Chlorpheniramine	0	3,234	-3,781	15,60	1/3	1
Dexamethasone	3	1,143	-2,682	94,83	1/3	1
Fluconazole	1	0,290	-2,127	76,15	1/3	1
Isoniazid	2	-0,819	-0,375	67,48	1/3	3
Lamivudine	2	-1,451	-0,520	88,15	1/3	3
Methyldopa	4	-1,632	0,030	103,78	1/3	1
Paracetamol	2	0,441	-1,058	49,33	1/3	1
Pravastatin	4	2,133	-3,145	124,29	1/3	1
Promethazine	0	4,371	-4,777	6,48	1/3	1
Pyrazinamide	1	0,436	-0,982	67,81	1/3	3
Quinine	1	2,584	-3,400	45,06	1/3	1

Amiloride	4	1,085	-3,161	158,23	1/4	1
Acetaminophen	2	0,441	-1,058	49,33	1/4	1
Spironolactone	0	3,390	-4,746	60,44	2/4	2
Furosemide	3	1,220	-3,571	118,72	3/4	3
Hydrochlorothiazide	3	-0,429	-2,838	118,36	3/4	4

Let’s consider results more detailed:

− Amitriptyline: according to various scientific articles, this compound can belong to BCS Class I or II, the PNN classified it as Class I;

− Chlorpromazine: according to various scientific articles, this compound can belong to BCS Class I or II, the PNN classified it as Class I;

− Digoxin: according to various scientific articles, this compound can belong to BCS Class I or II, the PNN classified it as Class I;

− Valproic acid: according to various scientific articles, this compound can belong to BCS Class I or II, the PNN classified it as Class II;

− Warfarin: according to various scientific articles, this compound can belong to BCS Class I or II, the PNN classified it as Class II;

− Acetylsalicylic acid: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class III;

− Captopril: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Chlorpheniramine: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Dexamethasone: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Fluconazole: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Isoniazid: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class III;

− Lamivudine: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class III;

− Methyldopa: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Paracetamol: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Pravastatin: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Promethazine: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Pyrazinamide: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class III;

− Quinine: according to various scientific articles, this compound can belong to BCS Class I or III, the PNN classified it as Class I;

− Amiloride: according to various scientific articles, this compound can belong to BCS Class I or IV, the PNN classified it as Class I;

− Acetaminophen: according to various scientific articles, this compound can belong to BCS Class I or IV, the PNN classified it as Class I;

− Spironolactone: according to various scientific articles, this compound can belong to BCS Class II or IV, the PNN classified it as Class II;

− Furosemide: according to various scientific articles, this compound can belong to BCS Class III or IV, the PNN classified it as Class III;

− Hydrochlorothiazide: according to various scientific articles, this compound can belong to BCS Class III or IV, the PNN classified it as Class IV.

Conclusions to section 2

It was established, that:

− four descriptors (number of HBond donors; partition coefficient; solubility; polar surface area) are sufficient for accurate drug classification according to the Biopharmaceutics Classification System;

− architecture of Probabilistic Neural Network with a spread value 0,1 is effective for accurate drug classification according to the Biopharmaceutics Classification System.

CONCLUSIONS

1. It was established that four descriptors (number of HBond donors, partition coefficient, solubility, and polar surface area) are sufficient for accurate drug classification according to the Biopharmaceutics Classification System.

2. The architecture of the Probabilistic Neural Network with a spread value of 0.1 proved to be an effective tool for accurate drug classification according to the Biopharmaceutics Classification System.

3. The analysis of scientific literature revealed compounds that had been reported as belonging to multiple BCS classes in different scientific articles. The developed model assigned each of these compounds to a single BCS class.

4. The obtained results demonstrate the significant potential of chemometric methods for developing predictive models, that can optimize the drug development process.

REFERENCES

1. Arrunátegui, L. B., Silva-Barcellos, N. M., Bellavinha, K. R., Ev, L. D. S., & Souza, J. D. (2015). Biopharmaceutics classification system: importance and inclusion in biowaiver guidance. *Brazilian Journal of Pharmaceutical Sciences*, *51*, 143-154.

2. Mehta, M. U., Uppoor, R. S., Conner, D. P., Seo, P., Vaidyanathan, J., Volpe, D. A., ... & Yu, L. X. (2017). Impact of the US FDA "Biopharmaceutics Classification System"(BCS) guidance on global drug development. *Molecular pharmaceutics*, *14*(12), 4334-4338.

3. Reddy, B. B. K., & Karunakar, A. (2011). Biopharmaceutics classification system: a regulatory approach. *Dissolution Technologies*, *18*(1), 31-37.

4. Cook, J. A., Davit, B. M., & Polli, J. E. (2010). Impact of biopharmaceutics classification system-based biowaivers. *Molecular pharmaceutics*, *7*(5), 1539-1544.

5. Amidon, G. L., Lennernäs, H., Shah, V. P., & Crison, J. R. (1995). A theoretical basis for a biopharmaceutic drug classification: the correlation of in vitro drug product dissolution and in vivo bioavailability. *Pharmaceutical research*, *12*, 413-420.

6. MÖLLER, H. (2002). The biopharmaceutical classification system (BCS) and its usage. *Drugs Made in Germany*, *45*(3), 63-65.

7. Dressman, J., Butler, J., Hempenstall, J., & Reppas, C. (2001). The BCS: where do we go from here?. *Pharmaceutical Technology*, *25*(7), 68-68.

8. Niazi, S. K. (2016). − Waiver of In Vivo Bioequivalence Study. In *Handbook of Pharmaceutical Manufacturing Formulations* (pp. 23-35). CRC Press.

9. Lindenberg, M., Kopp, S., & Dressman, J. B. (2004). Classification of orally administered drugs on the World Health Organization Model list of Essential Medicines according to the biopharmaceutics classification system. *European Journal of Pharmaceutics and Biopharmaceutics*, *58*(2), 265-278.

10. Polli, J. E., Lawrence, X. Y., Cook, J. A., Amidon, G. L., Borchardt, R. T., Burnside, B. A., ... & Zhang, G. (2004). Summary workshop report: biopharmaceutics classification system − implementation challenges and extension opportunities. *Journal of pharmaceutical sciences*, *93*(6), 1375-1381.

11. Barends, D., Dressman, J., Hubbard, J., Junginger, H., Patnaik, R., Polli, J., & Stavchansky, S. (2005). Multisource (generic) pharmaceutical products: guidelines on registration requirements to establish interchangeability draft revision. *WHO, Geneva, Switzerland*.

12. Sullivan, J. O., Blake, K., Berntgen, M., Salmonson, T., Welink, J., & Pharmacokinetics Working Party. (2018). Overview of the European Medicines Agency's Development of Product-Specific Bioequivalence Guidelines. *Clinical Pharmacology & Therapeutics*, *104*(3), 539-545.

13. Gothoskar, A. V. (2005). Biopharmaceutical classification system of drugs. *Pharmaceut Rev. Availablen from: URL: http://www. pharmainfo. net/reviews/biopharmaceutical-classification-drugs*.

14. Leardi, R. (2006). Chemometrics in data analysis. *Chromatographic Analysis of the Environment*, 221-241.

15. Pushkarova, Y., Zaitseva, G., & Kaliuzhenko, A. (2023, September). Classification of Residual Solvents by Risk Assessment Using Chemometric Methods. In *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)* (pp. 562-565). IEEE.

16. Brown, S., Tauler, R., & Walczak, B. (Eds.). (2020). *Comprehensive chemometrics: chemical and biochemical data analysis*. Elsevier.

17. Lavine, B., & Workman, J. (2008). Chemometrics. *Analytical chemistry*, *80*(12), 4519-4531.

18. Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433-459.

19. Greenacre, M., Groenen, P. J., Hastie, T., d'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), 100.

20. Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley interdisciplinary reviews: computational statistics*, *2*(1), 97-106.

21. Abdi, H., & Williams, L. J. (2013). Partial least squares methods: partial least squares correlation and partial least square regression. *Computational Toxicology: Volume II*, 549-579.

22. de Juan, A., & Tauler, R. (2021). Multivariate Curve Resolution: 50 years addressing the mixture analysis problem–A review. *Analytica Chimica Acta*, *1145*, 59-78.

23. De Juan, A., Jaumot, J., & Tauler, R. (2014). Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Analytical Methods*, *6*(14), 4964-4976.

24. Pushkarova, Y., & Tymchenko, I. (2024). Predicton of oral drug bioavailability based on chemical structure. *Сучасна медицина, фармація та психологічне здоров'я*, 104.

25. Pushkarova, Y., & Kholodniuk, P. (2023). A New Procedure for Unsupervised Clustering Based on Combination of Artificial Neural Networks. *European Journal of Artificial Intelligence and Machine Learning*, *2*(4), 1-3.

26. Pushkarova, Y., Panchenko, V., & Kholin, Y. (2021, July). Application an Artificial Neural Network for Prediction of Substances Solubility. In *IEEE EUROCON 2021-19th International Conference on Smart Technologies* (pp. 82-87). IEEE.

27. Rosa, J. P., Guerra, D. J., Horta, N. C., Martins, R. M., Lourenço, N. C., Rosa, J. P., ... & Lourenço, N. C. (2020). Overview of artificial neural networks. *Using artificial neural networks for analog integrated circuit design automation*, 21-44.

28. https://chemistrydocs.com/perkinelmer-chemoffice-2020-version-20-0/

29. Hubatsch, I., Ragnarsson, E. G., & Artursson, P. (2007). Determination of drug permeability and prediction of drug absorption in Caco-2 monolayers. *Nature protocols*, *2*(9), 2111-2119.

30. Hancock, B. C., & Zografi, G. (1997). Characteristics and significance of the amorphous state in pharmaceutical systems. *Journal of pharmaceutical sciences*, *86*(1), 1-12.

31. Leuner, C., & Dressman, J. (2000). Improving drug solubility for oral delivery using solid dispersions. *European journal of Pharmaceutics and Biopharmaceutics*, *50*(1), 47-60.

32. Sugano, K., Okazaki, A., Sugimoto, S., Tavornvipas, S., Omura, A., & Mano, T. (2007). Solubility and dissolution profile assessment in drug discovery. *Drug metabolism and pharmacokinetics*, *22*(4), 225-254.

33. Miller, J. M., Beig, A., Carr, R. A., Webster, G. K., & Dahan, A. (2012). The solubility–permeability interplay when using cosolvents for solubilization: revising the way we use solubility-enabling formulations. *Molecular pharmaceutics*, *9*(3), 581-590.

34. Bergström, C. A., Andersson, S. B., Fagerberg, J. H., Ragnarsson, G., & Lindahl, A. (2014). Is the full potential of the biopharmaceutics classification system reached?. *European Journal of Pharmaceutical Sciences*, *57*, 224-231.

35. Ostertagová E., Ostertag O., Kováč J. Methodology and Application of the Kruskal-Wallis Test. *Applied Mechanics and Materials*. 2014. Vol. 611. P. 115–120.

36. https://www.mathworks.com/products.html

37. Mohebali, B., Tahmassebi, A., Meyer-Baese, A., & Gandomi, A. H. (2020). Probabilistic neural networks: a brief overview of theory, implementation, and application. *Handbook of probabilistic models*, 347-367.

38. Mao, K. Z., Tan, K. C., & Ser, W. (2000). Probabilistic neural-network structure determination for pattern classification. *IEEE Transactions on neural networks*, *11*(4), 1009-1016.

39. Lindenberg, M., Kopp, S., & Dressman, J. B. (2004). Classification of orally administered drugs on the World Health Organization Model list of Essential Medicines according to the biopharmaceutics classification system. *European Journal of Pharmaceutics and Biopharmaceutics*, *58*(2), 265-278.

40. Wu, C. Y., & Benet, L. Z. (2005). Predicting drug disposition via application of BCS: transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharmaceutical research*, *22*, 11-23.

41. Yazdanian, M., Briggs, K., Jankovsky, C., & Hawi, A. (2004). The "high solubility" definition of the current FDA guidance on biopharmaceutical classification system may be too strict for acidic drugs. *Pharmaceutical research*, *21*, 293-299.

42. Ramirez, E., Laosa, O., Guerra, P., Duque, B., Mosquera, B., Borobia, A. M., ... & Frias, J. (2010). Acceptability and characteristics of 124 human bioequivalence studies with active substances classified according to the Biopharmaceutic Classification System. *British journal of clinical pharmacology*, *70*(5), 694-702.

43. Butler, J. M., & Dressman, J. B. (2010). The developability classification system: application of biopharmaceutics concepts to formulation development. *Journal of pharmaceutical sciences*, *99*(12), 4940-4954.

44. Kleberg, K., Jacobsen, J., & Müllertz, A. (2010). Characterising the behaviour of poorly water soluble drugs in the intestine: application of biorelevant media for solubility, dissolution and transport studies. *Journal of Pharmacy and Pharmacology*, *62*(11), 1656-1668.

SUPPLEMENTARY INFORMATION

СЕРТИФІКАТ

№007

ЦИМ СЕРТИФІКАТОМ ЗАСВІДЧУЄТЬСЯ, ЩО

SEDAT ACIK

ВЗЯВ (ЛА) УЧАСТЬ У НАУКОВО-ПРАКТИЧНІЙ КОНФЕРЕНЦІЇ “ІННОВАЦІЇ В МЕДИЦИНІ ТА ФАРМАЦІЇ: ВНЕСОК МОЛОДИХ ВЧЕНИХ”, ТРИВАЛІСТЮ 6 ГОДИН

28 ЛЮТОГО 2025 РОКУ

Кучин Ю.Л.
Ректор ЗВО
НМУ імені О.О. Богомольця

Земсков С.В.
Проректор з наукової
роботи та інновацій
НМУ імені О.О. Богомольця

Костюк І.А.
Голова Товариства молодих
вчених і спеціалістів
НМУ імені О.О. Богомольця

SUMMARY

Sedat Acik

ANALYSIS OF BIOPHARMACEUTICS CLASSIFICATION SYSTEM BY MEANS OF CHEMOMETRIC METHODS

Department of Analytical, Physical and Colloid Chemistry

Scientific supervisor: associate professor Yaroslava Pushkarova

Keywords: drug classification, molecular descriptors, pharmaceutical sciences, chemoinformatics.

Introduction. The analysis of the Biopharmaceutics Classification System using chemometric methods is a highly relevant and promising research topic. The Biopharmaceutics Classification System categorizes drugs into four classes based on their solubility and permeability. Predicting properties using chemometric models can minimize the number of in vitro and in vivo experiments required, streamlining the drug development pipeline. Due to the increasing volume of data in the pharmaceutical industry, chemometric methods are gaining popularity for their ability to process large datasets. Integrating machine learning and artificial intelligence enhances the prediction of biopharmaceutical parameters, making these methods even more powerful. Combining Biopharmaceutics Classification System with chemometric methods creates new opportunities for predicting bioavailability and developing pharmaceuticals, making the topic highly relevant in both research and practical applications. Aim of research*:* the aim of this study is to develop an accurate drug classification model according to the Biopharmaceutics Classification System using chemometric methods.

Materials and methods. Data set – total 145 drug compounds, 11 physical-chemical, and topological descriptors. Methods of investigation: Kruskal-Wallis test, Probabilistic Neural Network.

Results. The Biopharmaceutics Classification System is a framework used to classify drugs based on their solubility and intestinal permeability. It plays a

crucial role in drug development, guiding the formulation and regulatory approval processes.

Chemometric methods are essential tools in modern chemistry and related fields, providing robust techniques for analyzing complex chemical data. The primary aim of chemometrics is to extract useful information from experimental data using mathematical and statistical methods.

It was shown, that four descriptors (number of HBond donors; partition coefficient; solubility; polar surface area) are sufficient for accurate drug classification according to the Biopharmaceutics Classification System; architecture of Probabilistic Neural Network with a spread value 0,1 is effective for accurate drug classification according to the Biopharmaceutics Classification System.

Conclusions. It was established that four descriptors (number of HBond donors, partition coefficient, solubility, and polar surface area) are sufficient for accurate drug classification according to the Biopharmaceutics Classification System.

The architecture of the Probabilistic Neural Network with a spread value of 0.1 proved to be an effective tool for accurate drug classification according to the Biopharmaceutics Classification System.

The analysis of scientific literature revealed compounds that had been reported as belonging to multiple BCS classes in different scientific articles. The developed model assigned each of these compounds to a single BCS class.

The obtained results demonstrate the significant potential of chemometric methods for developing predictive models, that can optimize the drug development process.