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

«PREDICTION OF METABOLISM AND SOLUBILITY OF TABLETED DRUGS ACCORDING TO BIOPHARMACEUTICAL DRUG DISPOSITION CLASSIFICATION SYSTEM»

Created by: student of 5th year,
group 10601Fa
Kaouthar Bouaalam
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

BDDCS	Biopharmaceutical Drug Disposition Classification System
A	hydrogen bond acidity
B	hydrogen bond acidity
S	dipolarity / polarizability
E	excess molar refraction
logP	partition coefficient
logD	distribution coefficient
RotB	number of rotatable bonds
HBA	number of HBond acceptors
HBD	number of HBond donors
MW	molecular weight
PSA	polar surface area
PNN	Probabilistic Neural Network

INTRODUCTION

Actuality. The researched topic is highly relevant and timely in the field of pharmaceutical sciences. It addresses critical challenges in drug development and patient safety by offering a method to predict key pharmacokinetic properties, such as metabolism and solubility, of tablet-form medications. This is particularly important in the current landscape of personalized medicine, where the aim is to tailor treatments to individual patients, enhancing drug efficacy and minimizing adverse effects.

The Biopharmaceutical Drug Disposition Classification System (BDDCS) provides a framework for predicting the absorption, distribution, metabolism, and excretion of drugs based on their solubility and metabolism characteristics. By categorizing drugs into four distinct classes based on these two factors, BDDCS helps pharmaceutical scientists predict how a drug will behave in the body and the potential challenges in its formulation. This predictive power is essential for optimizing drug design and formulation strategies, particularly for drugs that may have issues with bioavailability, solubility, or metabolism.

With the increasing demand for more effective and safe medications, understanding drug metabolism and solubility is crucial for minimizing the risks associated with drug development, especially during the preclinical and clinical trial phases. BDDCS offers a practical tool for identifying drugs that may pose bioavailability challenges early in the development process, allowing for the design of more efficient drug delivery systems and formulations. Additionally, the system helps in assessing the impact of various factors, such as food-drug interactions, transporters, and liver function, on the drug's pharmacokinetics.

As the pharmaceutical industry continues to move toward more sophisticated and efficient methods for drug discovery, the ability to predict and optimize metabolism and solubility through systems like BDDCS becomes increasingly important. This approach not only accelerates the drug development timeline but also ensures that medicines are safer and more effective for patients. Given the

complexities of the human body and the variability in individual responses to drugs, BDDCS provides a systematic way of improving the predictability of drug outcomes and enhancing the overall success rate of drug candidates. Therefore, the importance of this topic lies in its ability to contribute to the efficiency of the pharmaceutical industry, reduce development costs, and improve patient outcomes.

Aim and tasks of research. To develop predictive models for the metabolism and solubility of tableted drugs based on the principles of the Biopharmaceutical Drug Disposition Classification System, using molecular descriptors and chemometric methods.

Tasks of research:

1) to identify informative and necessary sets of molecular descriptors separately for the prediction of metabolism and solubility classes of tableted drugs;

2) to optimize the spread parameter in the Probabilistic Neural Network to achieve the best classification performance;

3) to evaluate the predictive efficiency of the Probabilistic Neural Network in classifying drugs into predefined categories (e.g., high vs. low solubility; extensive vs. poor metabolism) rather than predicting exact numerical values.

Research methods: Kruskal-Wallis test, Probabilistic Neural Network, correlation analysis. The software package MATLAB R2024b and ChemOffice 2020 (trial license) were used in this work.

Novelty and significance of the results. This study presents a novel approach to the classification of tableted drugs according to their metabolism and solubility profiles within the framework of the Biopharmaceutical Drug Disposition Classification System. Unlike traditional models that aim to predict exact numerical values, the proposed methodology focuses on binary classification (e.g., high vs. poor solubility; extensive vs. poor metabolism), which is more practical and relevant for early-stage drug development.

A key novelty of this work lies in the separate selection of informative molecular descriptors for each predicted property, allowing for more targeted and interpretable models. Additionally, the use of a Probabilistic Neural Network with

optimized spread values offers a simple yet effective tool for drug classification based on a limited number of descriptors.

The significance of the results lies in the potential application of these models in drug discovery and development pipelines. The ability to quickly and reliably classify compounds by BDDCS-related properties can assist in prioritizing drug candidates, reducing experimental costs, and improving formulation strategies at early stages.

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: Kaouthar Bouaalam, Yaroslava Pushkarova. Probabilistic neural network models for predicting metabolism of tableted medicines according to biopharmaceutical drug disposition classification system. 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. 164-165.

Structure of work: 42 pages, 2 sections, supplementary information, 29 references.

MAIN PART

1. LITERATURE REVIEW

1.1. Biopharmaceutical Drug Disposition Classification System: main aspects

The Biopharmaceutical Drug Disposition Classification System (BDDCS) is an advanced framework introduced by Wu and Benet in 2005. It was designed to predict drug disposition and categorize drugs based on their solubility and extent of metabolism. BDDCS helps identify how drugs are absorbed, distributed, metabolized, and excreted, providing valuable insights for drug development and regulatory assessment. Prior to BDDCS, drug classification systems primarily focused on solubility and permeability (like the Biopharmaceutics Classification System, BCS). However, BDDCS introduced a more nuanced approach by incorporating the extent of metabolism, offering a more comprehensive understanding of drug disposition. This framework emerged from the recognition that metabolism significantly influences a drug's pharmacokinetics and pharmacodynamics, leading to more accurate predictions of drug behavior in the human body [1-3].

BDDCS classifies drugs into four categories according to their solubility and metabolism extent [4-5]:

1. Class 1: High solubility, extensive metabolism.

Drugs in this class are well absorbed and primarily eliminated via metabolism. They are less affected by efflux and uptake transporters.

2. Class 2: Low solubility, extensive metabolism.

These drugs are poorly soluble but undergo significant metabolism. Efflux transporters like P-glycoprotein may limit their bioavailability.

3. Class 3: High solubility, poor metabolism.

Drugs in this class are highly soluble but undergo minimal metabolic transformation. Excretion is typically renal or biliary without major metabolic change.

4. Class 4: Low solubility, poor metabolism.

These drugs are characterized by poor solubility and limited metabolism. They often exhibit low oral bioavailability.

BDDCS has multiple applications in the field of pharmacokinetics and pharmacodynamics [6, 7]:

− predicting drug disposition and elimination pathways,

− evaluating the impact of transporters on drug absorption and bioavailability,

− assessing food-drug interactions, particularly for Class 2 drugs,

− estimating drug-drug interaction potential and Central Nervous System exposure,

− predicting the risk of drug-induced liver injury,

− facilitating drug repurposing by assessing metabolic and solubility profiles.

BDDCS plays a crucial role in pharmaceutical sciences by providing insights into drug formulation, bioavailability prediction, and personalized medicine. It aids in determining the appropriate route of administration, optimizing dosage forms, and minimizing adverse drug reactions. The system helps in anticipating potential bioavailability issues, guiding formulation scientists to improve solubility or select alternate delivery methods. Additionally, BDDCS assists in clinical decision-making, particularly when selecting medications for patients with compromised liver function or those at risk of drug interactions [8, 9].

Case Studies and Practical Examples:

− the use of BDDCS to predict bioavailability issues in poorly soluble drugs, leading to the development of novel formulation strategies like solid dispersions [10],

− application in regulatory science, where BDDCS helps streamline the approval process by predicting pharmacokinetic profiles based on drug classification [11],

− research utilizing BDDCS to optimize oral drug delivery systems by modifying physicochemical properties [12].

1.2. Importance of Predicting Metabolism and Solubility (Poor/Low or Extensive / High)

Predicting the metabolism and solubility of pharmaceutical compounds is a critical aspect of drug development and personalized medicine. Metabolism primarily affects how drugs are processed in the body, while solubility determines the extent of drug absorption. These properties influence bioavailability, therapeutic efficacy, dosing strategies, and the risk of adverse effects. Accurate prediction of whether a drug exhibits poor or extensive metabolism and solubility is essential for optimizing pharmacokinetic profiles and ensuring patient safety [13-17].

Significance of Metabolism Prediction:

1. Pharmacokinetics and Bioavailability: extensive metabolism can reduce bioavailability, as the drug is rapidly broken down before reaching systemic circulation; poor metabolism can lead to drug accumulation, increasing the risk of toxicity.

2. Dosing Considerations: drugs with extensive metabolism may require higher doses or sustained-release formulations to maintain therapeutic levels; poorly metabolized drugs often need dose adjustments to prevent accumulation, particularly in patients with compromised metabolic function.

3. Drug-Drug Interactions: predicting metabolism helps identify potential interactions, especially with drugs metabolized by the same enzymes; extensive metabolism increases the likelihood of competitive inhibition when co-administered with enzyme inhibitors.

4. Personalized Medicine: metabolic profiling allows for tailoring therapy based on individual metabolic capacity, minimizing side effects and optimizing efficacy.

Significance of Solubility Prediction:

1. Oral Bioavailability: poor solubility limits drug dissolution and absorption, leading to subtherapeutic plasma concentrations; enhancing solubility through formulation strategies improves bioavailability.

2. Route of Administration: drugs with poor solubility may require non-oral routes to achieve adequate systemic exposure.

3. Therapeutic Efficacy: adequate solubility is crucial for achieving the desired pharmacological effect, particularly in oral formulations.

4. Formulation Development: solubility prediction guides the choice of excipients and formulation techniques, directly impacting drug performance.

1.3. Introduction to Neural Networks

Neural networks are a subset of machine learning algorithms designed to simulate the way the human brain processes information. They are the foundation for deep learning and have revolutionized fields like artificial intelligence, natural language processing, computer vision, and autonomous systems. Neural networks model complex relationships between input and output data, making them highly effective in solving problems involving large, complex datasets [18, 19].

A neural network consists of interconnected layers of units called neurons. These neurons are organized into layers:

1. Input Layer: this layer receives the raw input data; each neuron in this layer represents one feature or attribute of the input data.

2. Hidden Layers: these layers perform computations on the input data. There can be multiple hidden layers, which allow neural networks to capture complex patterns. Each neuron in a hidden layer applies a weighted sum of inputs followed by an activation function.

3. Output Layer: this layer produces the final output, whether it's a prediction, classification, or any other form of result. The number of neurons in the output layer depends on the type of task.

Each neuron in a neural network processes the input it receives by applying a mathematical operation (usually a weighted sum) followed by an activation function. The activation function introduces non-linearity to the model, allowing it to capture complex patterns.

Training a neural network involves adjusting its weights based on the errors in its predictions. This process is done using optimization techniques, the most common being backpropagation and gradient descent [20-22].

1. Forward Propagation: in this step, the input data is passed through the network, layer by layer, until it produces an output.

2. Backpropagation: this algorithm computes the gradient of the loss function with respect to each weight in the network and propagates this error backward through the layers.

3. Gradient Descent: the weights are updated by moving them in the direction that reduces the loss.

Neural networks have found applications in numerous fields: Image Recognition, Natural Language Processing, Autonomous Systems, Drug Discovery, Healthcare.

Conclusions to section 1

BDDCS provides a robust framework for understanding drug disposition by integrating solubility and metabolism data. Its predictive power makes it a crucial tool in drug development, facilitating more accurate predictions of ADME characteristics, potential drug interactions, and formulation strategies. Its ongoing evolution ensures that it remains a cornerstone in modern pharmaceutical research and drug development.

Accurate prediction of metabolism and solubility is fundamental to drug development, influencing dosing strategies, therapeutic outcomes, and safety

profiles. By understanding whether a compound exhibits poor or extensive metabolism and solubility, researchers can optimize formulations and personalize treatment approaches, ultimately improving patient outcomes.

In the future, neural networks are expected to become more efficient and interpretable. With advancements in transfer learning, neural architecture search, and unsupervised learning, neural networks will continue to push the boundaries of artificial intelligence and machine learning.

2. EXPERIMENTAL PART

2.1. Set of tableted drugs and descriptors

The researched set involves 220 tableted drugs, characterized by 11 descriptors:

1) hydrogen bond acidity (A),

2) hydrogen bond acidity (B),

3) dipolarity / polarizability (S),

4) excess molar refraction (E),

5) partition coefficient (logP),

6) distribution coefficient (logD),

7) number of rotatable bonds (RotB),

8) number of HBond acceptors (HBA),

9) number of HBond donors (HBD),

10) molecular weight (MW),

11) polar surface area (PSA).

The descriptors № 1-9 was taken from [15]. The descriptors № 10, 11 have been calculated by means of ChemOffice software [23]. The Kruskal-Wallis test and Probabilistic Neural Network were realized using the software package Matlab R2024b [24].

Under «Prediction of Metabolism and Solubility of Tableted DRUGS According to Biopharmaceutical Drug Disposition Classification System» we understand prediction of class according to BDDCS as follows:

1) extensive or poor metabolism (class 1 or class 2),

2) high or low solubility (class 1 or class 2).

Set of tableted drugs and descriptors are presented in Table 2.1.

Table 2.1. Set of tableted drugs and descriptors

№	Drug compound	BDDCS class: 1 – extensive metabolism, 2 – poor metabolism	BDDCS class: 1 – high solubility, 2 – low solubility	Descriptor										
				A	B	S	E	logP	logD	RotB	HBA	HBD	MW	PSA
1	Acetaminophen	1	1	0,95	0,80	1,63	3,24	0,49	0,34	2	2	2	151,17	55,41
2	Acetohexamide	1	1	0,59	1,46	2,79	4,32	2,25	2,44	6	6	2	324,40	103,22
3	Aliskiren	1	1	1,20	3,06	3,65	2,53	3,51	0,52	20	7	4	551,77	156,62
4	Alosetron	1	1	0,35	1,38	2,64	3,24	1,74	0,76	2	5	1	294,36	44,23
5	Alprazolam	1	1	0,00	0,84	1,95	1,30	2,56	2,50	1	4	0	308,77	33,27
6	Alprenolol	1	1	0,29	1,36	1,12	0,50	-0,86	1,22	8	2	2	249,36	46,24
7	Ambrisentan	1	1	0,57	1,52	2,32	2,91	3,33	5,46	7	5	1	376,46	70,11
8	Ambroxol	1	1	0,73	1,12	1,89	0,73	2,66	1,70	3	3	3	364,08	64,50
9	Aminophenazone	1	1	0,00	1,79	1,88	0,81	1,04	0,75	2	4	0	231,30	26,79
10	Amlodipine	1	1	0,36	2,19	2,26	2,65	3,43	2,89	10	7	2	408,89	105,47
11	Amoxapine	1	1	0,16	1,43	1,68	0,99	3,41	1,73	1	3	1	313,79	34,47
12	Anastrozole	1	1	0,00	1,00	2,38	2,06	1,29	0,97	4	5	0	293,37	61,36
13	Asenapine	1	1	0,00	0,91	1,58	1,53	4,58	2,55	0	1	0	285,78	10,59
14	Azathioprine	1	1	0,35	1,56	2,86	1,02	0,51	0,65	3	6	1	277,27	99,66
15	Bambuterol	1	1	0,38	2,25	2,20	9,73	0,56	-1,03	10	6	2	367,45	94,24
16	Benazepril	1	1	0,71	2,08	2,75	1,45	1,82	1,14	10	7	2	424,50	101,61
17	Benidipine	1	1	0,13	2,13	2,99	1,64	7,41	4,82	9	6	1	535,65	112,85
18	Benserazide	1	1	2,01	2,63	2,78	1,18	-2,90	-2,11	6	5	7	257,25	169,84
19	Benznidazole	1	1	0,35	0,57	0,98	1,87	0,90	0,91	6	4	1	260,25	87,81
20	Bepridil	1	1	0,00	1,32	1,81	2,52	6,20	4,10	10	3	0	366,55	10,84
21	Beraprost	1	1	1,20	1,51	2,03	2,16	2,04	2,87	8	4	3	398,50	95,05
22	Betamethasone	1	1	0,80	1,97	2,95	1,98	1,79	1,87	2	5	3	392,47	104,22
23	Betaxolol	1	1	0,29	1,53	1,31	1,10	2,32	1,02	11	3	2	307,44	55,03
24	Biperiden	1	1	0,31	1,17	1,32	1,19	4,94	1,76	5	2	1	311,47	23,73
25	Bopindolol	1	1	0,46	1,48	2,14	0,56	4,98	3,75	9	4	2	380,49	64,71
26	Bromazepam	1	1	0,47	1,27	1,93	1,35	1,70	2,06	1	4	1	316,16	52,40
27	Bromperidol	1	1	0,31	1,45	2,16	0,24	4,00	2,37	6	3	1	420,33	42,01
28	Bupropion	1	1	0,13	0,94	1,32	3,44	3,21	3,31	4	2	1	239,75	32,84
29	Butorphanol	1	1	0,73	1,32	1,42	2,30	3,73	3,17	2	2	2	327,47	46,61

30	Capecitabine	1	1	0,60	2,40	2,41	1,87	0,84	0,97	8	9	3	359,36	124,33
31	Carbidopa	1	1	1,69	1,77	1,79	1,03	-0,45	-2,68	4	4	5	226,23	132,39
32	Cerivastatin	1	1	1,20	1,80	2,25	1,45	3,56	1,21	11	6	3	459,56	104,98
33	Chlorambucil	1	1	0,57	0,80	1,60	1,62	3,63	1,42	9	3	1	304,22	41,70
34	Chlorpheniramine	1	1	0,00	1,02	1,49	2,33	3,15	1,58	5	2	0	274,80	11,42
35	Chlorpromazine	1	1	0,00	0,99	1,45	3,63	5,30	3,32	4	2	0	318,87	1,75
36	Cilazapril	1	1	0,71	2,51	2,70	0,73	0,50	-2,09	9	9	3	417,51	106,36
37	Clemastine	1	1	0,00	0,97	1,55	1,98	5,45	3,11	6	2	0	343,90	9,97
38	Clomipramine	1	1	0,00	0,89	1,66	2,09	5,92	3,59	4	2	0	314,86	1,75
39	Clonazepam	1	1	0,47	1,09	2,25	1,22	2,38	2,34	2	3	1	315,72	86,30
40	Clorazepate	1	1	1,04	1,34	2,14	1,19	2,51	1,80	2	6	4	314,73	82,98
41	Cortisone	1	1	0,41	1,90	1,63	1,39	1,30	1,44	2	5	2	360,45	99,93
42	Cyclizine	1	1	0,00	1,21	1,55	1,06	3,80	2,12	3	2	0	266,39	2,35
43	Cyclobenzaprine	1	1	0,00	0,83	1,41	1,74	5,10	3,29	3	1	0	275,40	1,18
44	Cyclophosphamide	1	1	0,14	1,18	2,20	1,84	0,80	0,23	5	4	1	261,09	44,31
45	Cyproheptadine	1	1	0,00	0,83	1,85	2,67	6,41	5,18	0	1	0	287,41	1,18
46	Darifenacin	1	1	0,49	1,58	2,18	1,99	3,62	2,70	7	3	1	426,56	56,52
47	Debrisoquine	1	1	0,34	1,16	1,37	1,54	0,90	-1,86	1	3	2	175,24	52,16
48	Desipramine	1	1	0,13	0,90	1,58	2,08	4,47	1,44	4	2	1	266,39	15,14
49	Dexamethasone	1	1	0,80	1,97	2,95	1,61	1,79	1,87	2	5	3	392,47	104,22
50	Diazepam	1	1	0,00	1,04	1,72	1,04	2,96	2,96	1	3	0	284,75	28,76
51	Dolasetron	1	1	0,31	1,52	1,76	2,45	2,34	2,70	3	5	1	324,38	60,48
52	Doxazosin	1	1	0,23	2,60	4,45	1,89	3,53	0,61	5	6	1	451,49	104,90
53	Enalapril	1	1	0,71	1,92	2,61	1,72	0,67	-0,32	11	7	2	376,46	101,91
54	Escitalopram	1	1	0,00	1,08	1,87	2,15	3,13	0,48	5	3	0	324,40	28,67
55	Estradiol	1	1	0,81	0,95	2,30	2,66	3,78	4,13	0	1	2	272,39	45,43
56	Eszopiclone	1	1	0,00	2,43	3,20	2,66	1,25	0,74	4	9	0	388,82	79,29
57	Ethinylestradiol	1	1	0,90	1,02	3,79	2,04	3,86	4,52	0	1	2	296,41	45,43
58	Famciclovir	1	1	0,23	1,64	1,76	2,07	0,09	-0,67	9	9	1	321,34	113,10
59	Flunitrazepam	1	1	0,00	1,15	2,15	1,34	1,78	1,25	2	3	0	313,29	72,91
60	Fluvoxamine	1	1	0,23	1,14	0,95	1,41	3,32	1,25	10	2	1	318,34	58,37
61	Frovatriptan	1	1	0,93	1,33	2,10	0,96	0,72	-1,79	2	4	3	243,31	74,79
62	Galantamine	1	1	0,31	1,45	2,02	2,48	1,03	1,18	1	2	1	287,36	41,94
63	Granisetron	1	1	0,26	1,56	2,38	2,07	1,72	-1,27	3	5	1	312,42	47,74

64	Guanabenz	1	1	0,48	1,20	1,02	1,66	2,98	1,84	3	4	3	231,09	78,44
65	Hexobarbital	1	1	0,24	1,33	1,50	2,17	1,63	1,58	1	5	1	236,27	70,56
66	Hydrocodone	1	1	0,00	1,42	2,12	1,69	1,13	0,75	1	2	0	299,37	37,66
67	Hydromorphone	1	1	0,27	1,32	1,79	1,30	0,72	-0,04	0	2	1	285,35	51,42
68	Hydroxyzine	1	1	0,23	1,80	2,41	3,12	4,00	1,99	8	4	1	374,91	33,70
69	Imidapril	1	1	0,71	2,22	2,85	2,05	1,53	-1,97	11	9	2	405,45	121,36
70	Imipramine	1	1	0,00	0,95	1,59	2,05	5,04	2,84	4	2	0	280,42	1,75
71	Isoniazid	1	1	0,47	1,39	1,85	1,73	-0,67	-0,89	2	4	2	137,14	74,33
72	Ivabradine	1	1	0,00	2,35	3,25	2,04	3,97	2,40	10	3	0	468,60	57,04
73	Ivermectin	1	1	0,68	4,23	3,21	2,67	5,39	6,92	15	28	6	875,12	173,88
74	Labetalol	1	1	1,00	1,72	2,15	0,99	2,50	0,75	8	4	4	328,41	106,25
75	Letrozole	1	1	0,00	0,97	2,92	2,21	1,24	1,91	3	5	0	285,31	61,36
76	Melphalan	1	2	0,78	1,37	1,90	1,01	-0,21	-0,71	8	4	2	305,21	69,67
77	Omeprazole	1	2	0,35	2,05	3,18	3,07	2,57	2,13	5	4	1	345,42	71,04
78	Prochlorperazine	1	2	0,00	1,47	2,11	2,23	4,38	3,97	4	3	0	373,95	2,92
79	Thioridazine	1	2	0,00	1,13	1,93	2,05	6,00	4,03	4	2	0	370,58	1,75
80	Triamcinolone acetonide	1	2	0,56	2,14	3,13	2,49	2,21	2,50	2	6	2	434,51	99,25
81	Aminoglutethimide	1	2	0,56	1,34	1,79	2,45	0,77	1,41	2	4	2	232,28	78,79
82	Aripiprazole	1	2	0,41	1,75	2,53	1,85	5,31	5,55	7	4	1	448,40	43,70
83	Armodafinil	1	2	0,49	1,47	3,20	1,63	0,94	1,17	5	3	1	273,36	64,62
84	Astemizole	1	2	0,13	1,64	2,70	2,48	6,09	4,20	8	4	1	458,58	35,37
85	Bevantolol	1	2	0,29	1,82	2,14	2,26	3,00	2,19	10	2	2	381,91	33,17
86	Bicalutamide	1	2	0,71	1,63	3,05	1,79	2,71	4,94	7	6	2	430,38	110,56
87	Bosentan	1	2	0,60	2,48	3,50	2,41	4,17	1,15	11	8	2	551,63	142,96
88	Buspirone	1	2	0,00	2,16	2,18	2,26	2,19	3,39	6	7	0	385,51	60,25
89	Carbamazepine	1	2	0,39	0,92	2,06	1,84	2,38	2,67	1	3	1	236,28	46,81
90	Carvedilol	1	2	0,62	2,09	3,00	2,08	4,04	3,47	10	3	3	406,49	78,42
91	Cefditoren Pivoxil	1	2	0,50	3,45	4,52	1,70	2,71	-1,66	13	11	2	620,73	176,19
92	Cefpodoxime Proxetil	1	2	0,50	3,50	1,68	1,30	0,80	1,92	14	12	2	557,61	183,53
93	Chlorzoxazone	1	2	0,45	0,50	1,32	1,68	2,51	2,19	0	1	1	169,57	42,53
94	Cilostazol	1	2	0,41	1,63	2,37	0,91	3,53	3,05	7	6	1	369,47	81,66
95	Cinacalcet	1	2	0,13	0,63	1,37	0,50	6,35	4,05	7	1	1	357,42	14,57
96	Cisapride	1	2	0,50	2,17	3,15	1,54	3,81	2,69	10	5	2	465,96	88,69
97	Citalopram	1	2	0,00	1,08	2,25	2,19	3,13	0,48	5	3	0	324,40	28,67

98	Clotrimazole	1	2	0,00	0,78	2,37	1,18	5,25	5,42	4	2	0	344,85	10,81
99	Clozapine	1	2	0,20	1,65	1,66	2,17	3,71	3,45	1	4	1	326,83	25,63
100	Dapsone	1	2	0,45	1,35	2,84	1,70	0,89	0,94	2	4	2	248,31	92,10
101	Dasatinib	1	2	0,76	2,50	3,47	1,51	2,88	3,30	8	9	3	488,02	101,85
102	Desloratadine	1	2	0,13	0,99	1,55	2,04	3,83	4,78	0	2	1	310,83	24,82
103	Dicoumarol	1	2	0,63	1,57	2,48	1,01	3,66	3,55	2	4	2	336,30	101,12
104	Diloxanide furoate	1	2	0,09	1,16	2,34	1,30	3,09	1,42	6	3	0	328,15	55,94
105	Disulfiram	1	2	0,00	1,16	1,62	2,04	3,88	3,88	9	2	0	296,54	2,35
106	Domperidone	1	2	0,72	1,83	3,13	2,30	4,27	2,99	5	7	2	425,92	66,80
107	Donepezil	1	2	0,00	1,50	2,17	0,91	4,60	3,39	6	2	0	379,50	37,66
108	Drospirenone	1	2	0,00	1,24	3,29	1,91	2,84	3,16	0	3	0	366,50	45,34
109	Ebastine	1	2	0,00	1,41	2,37	2,01	6,94	6,77	10	3	0	469,67	28,24
110	Entacapone	1	2	0,58	1,38	2,85	2,17	1,76	2,38	6	3	2	305,29	128,04
111	Eplerenone	1	2	0,00	1,75	3,73	1,16	0,29	1,05	2	6	0	414,50	81,19
112	Estazolam	1	2	0,00	0,84	2,01	2,92	2,29	3,25	1	4	0	294,75	33,27
113	Etodolac	1	2	0,88	0,90	2,12	1,56	3,43	3,52	4	4	2	287,36	63,59
114	Etoricoxib	1	2	0,00	1,41	2,77	2,19	2,35	2,26	3	4	0	358,85	57,25
115	Etravirine	1	2	0,47	1,42	3,44	3,67	5,22	4,19	4	6	2	435,29	108,94
116	Exemestane	1	2	0,00	1,14	2,60	1,39	3,28	3,11	0	2	0	296,41	36,54
117	Ezetimibe	1	2	0,81	1,77	2,61	2,72	3,96	3,49	6	3	2	409,44	64,57
118	Febuxostat	1	2	0,57	1,11	2,25	2,44	4,40	3,87	5	4	1	316,38	78,88
119	Felodipine	1	2	0,13	1,42	1,83	3,74	5,30	4,83	6	5	1	384,26	68,70
120	Flufenamic acid	1	2	0,72	0,59	1,36	1,11	5,53	5,62	4	3	2	281,24	54,80
121	Flunarizine	1	2	0,00	1,37	2,06	2,48	6,34	4,67	6	2	0	404,51	2,35
122	Gemfibrozil	1	2	0,57	0,71	1,07	2,11	3,94	4,39	6	2	1	250,34	49,94
123	Gliclazide	1	2	0,59	1,66	2,54	2,54	1,09	1,19	5	6	2	323,42	89,39
124	Glimepiride	1	2	0,75	2,15	3,50	1,76	3,96	2,93	11	9	3	490,63	137,24
125	Glipizide	1	2	0,85	2,19	1,00	2,14	2,57	2,00	10	9	3	445,54	138,28
126	Griseofulvin	1	2	0,00	1,58	1,87	2,12	1,91	3,53	3	3	0	352,77	72,65
127	Haloperidol	1	2	0,31	1,45	2,08	1,70	3,85	2,20	6	3	1	375,87	42,01
128	Iloperidone	1	2	0,00	1,73	2,85	2,00	4,27	2,89	8	2	0	426,49	61,00
129	Indobufen	1	2	0,57	1,15	1,66	1,23	3,27	3,30	4	4	1	295,34	59,98
130	Indomethacin	1	2	0,57	1,24	2,49	2,39	4,18	3,10	5	4	1	357,80	68,78
131	Indoramin	1	2	0,57	1,49	2,66	2,29	2,84	1,34	6	4	2	347,46	47,99

132	Irbesartan	1	2	0,63	1,69	2,71	1,05	6,04	3,04	7	7	1	428,54	82,47
133	Ketanserin	1	2	0,26	1,82	2,01	2,22	3,00	2,77	5	6	1	395,44	70,53
134	Ketoconazole	1	2	0,00	2,22	3,76	2,18	3,64	3,50	8	7	0	531,44	57,83
135	Pentazocine	1	2	0,50	1,04	1,38	1,36	4,67	3,08	2	1	1	285,43	24,05
136	Albuterol	2	1	1,19	1,82	1,26	1,91	0,06	-1,69	5	3	4	239,32	82,56
137	Almotriptan	2	1	0,31	1,65	2,16	2,04	1,79	-0,07	6	5	1	335,47	53,85
138	Amiloride	2	1	1,01	2,16	2,12	1,42	0,11	2,89	3	8	5	229,63	159,65
139	Aminocaproic acid	2	1	0,78	0,91	0,95	1,81	-2,24	-1,59	4	3	2	131,18	68,80
140	Amoxicillin	2	1	1,55	2,90	3,59	1,13	-1,87	-2,02	5	7	4	365,41	143,96
141	Atenolol	2	1	0,78	1,85	1,97	2,07	-0,11	-1,57	8	4	3	266,34	92,48
142	Atropine	2	1	0,31	1,31	3,15	0,82	1,30	-0,86	5	4	1	289,38	50,80
143	Baclofen	2	1	0,78	1,02	1,47	2,20	-0,62	-0,94	4	3	2	213,67	68,80
144	Biotin	2	1	0,95	1,22	1,86	1,07	-0,08	0,09	5	5	3	244,31	88,25
145	Bumetanide	2	1	1,16	1,70	1,92	3,51	3,37	0,47	8	6	3	364,42	130,02
146	Cadralazine	2	1	0,57	1,84	2,14	1,12	0,93	0,33	9	8	3	283,33	106,01
147	Captopril	2	1	0,57	1,13	1,77	2,15	0,89	0,27	4	4	1	217,29	60,28
148	Carbenicillin	2	1	1,42	2,41	3,14	2,70	1,64	0,99	6	8	3	378,41	133,95
149	Cefamandole	2	1	1,02	3,18	2,01	1,02	0,11	1,51	8	11	3	462,51	155,70
150	Cefuroxime	2	1	1,29	2,90	3,62	1,48	0,23	0,46	9	9	3	424,39	179,19
151	Celiprolol	2	1	0,61	2,35	2,34	2,47	1,86	0,31	12	6	3	379,50	98,23
152	Cetirizine	2	1	0,57	1,76	2,24	1,39	2,08	-0,50	8	5	1	388,90	51,97
153	Chloroquine	2	1	0,13	1,29	1,63	1,89	5,06	1,93	8	3	1	319,88	25,69
154	Cilazaprilat	2	1	1,28	2,50	2,74	1,66	1,50	-3,65	7	8	3	389,46	120,13
155	Clarithromycin	2	1	0,80	4,49	2,97	1,15	2,37	2,42	8	14	4	747,97	189,50
156	Dalfampridine	2	1	0,23	0,71	1,21	2,01	0,32	-1,43	0	2	1	94,12	37,91
157	Demeclocycline	2	1	2,27	3,57	3,93	2,27	-0,59	-3,97	2	9	6	464,86	197,07
158	Desvenlafaxine	2	1	0,81	1,22	1,34	1,73	2,68	0,45	4	2	2	263,38	34,13
159	Didanosine	2	1	0,31	1,77	1,85	1,62	-1,62	-1,36	2	7	2	236,23	84,65
160	Digitoxin	2	1	1,27	4,02	4,20	1,17	2,85	2,44	7	13	5	764,96	192,62
161	Digoxin	2	1	1,58	4,32	1,82	0,98	1,42	0,85	7	14	6	780,96	215,17
162	Doxycycline	2	1	2,10	3,47	3,88	1,56	-0,51	-3,64	2	10	7	444,45	197,07
163	Enalaprilat	2	1	1,28	1,91	2,08	1,35	0,88	-0,32	9	7	3	348,40	115,68
164	Entecavir	2	1	1,05	2,29	2,18	2,32	-2,58	-0,96	2	9	5	277,29	126,09
165	Ethambutol	2	1	0,78	1,72	0,98	1,73	0,12	-2,13	9	4	4	204,31	74,26

166	Etidronic Acid	2	1	1,56	2,54	1,55	1,46	-2,54	-3,76	2	7	5	206,03	150,80
167	Famotidine	2	1	1,21	2,78	2,24	1,60	-1,17	-0,95	9	9	6	337,45	179,34
168	Gabapentin	2	1	0,78	0,93	0,99	2,28	-0,66	-1,31	3	3	2	171,24	68,80
169	Hydrochlorothiazide	2	1	1,01	1,76	2,77	2,81	-0,37	-0,07	1	7	3	297,74	131,87
170	Hydroxyurea	2	1	0,91	0,98	1,40	2,31	-1,80	-1,80	1	2	3	76,06	86,58
171	Lamivudine	2	1	0,44	2,02	1,92	1,76	-1,46	-0,71	2	6	2	229,26	88,11
172	Levetiracetam	2	1	0,49	1,32	1,87	2,60	-0,34	-0,67	3	4	1	170,21	65,69
173	Levocetirizine	2	1	0,57	1,76	2,24	2,37	2,08	-0,50	8	5	1	388,90	51,97
174	Lisinopril	2	1	1,49	2,47	2,98	1,56	-1,69	-1,35	13	8	4	405,50	143,65
175	Memantine	2	1	0,21	0,66	0,58	1,76	3,03	0,29	0	1	1	179,31	27,97
176	Metformin	2	1	0,55	1,68	0,58	3,39	-1,63	-4,31	2	5	3	129,17	89,74
177	Methazolamide	2	1	0,44	2,01	2,50	1,78	0,09	0,13	2	7	1	236,27	107,10
178	Methyldopa	2	1	1,56	1,45	1,73	1,47	-2,26	-2,38	3	3	4	211,22	114,54
179	Miglitol	2	1	1,18	2,30	1,57	2,36	-1,26	-1,41	3	6	5	207,23	113,97
180	Milnacipran	2	1	0,21	1,33	2,32	1,81	1,91	-1,43	6	3	1	246,36	47,42
181	Nadolol	2	1	0,83	1,90	1,56	2,97	0,38	-0,38	6	4	4	309,41	91,35
182	Naratriptan	2	1	0,68	1,62	2,14	3,77	1,70	-1,76	5	5	2	335,47	67,24
183	Phenylpropanolamine	2	1	0,46	1,22	1,09	3,49	0,58	-0,20	2	2	2	151,21	50,53
184	Piperazine	2	1	0,29	0,89	0,63	2,66	-1,48	-3,49	0	2	2	86,14	29,15
185	Piracetam	2	1	0,49	1,28	1,88	0,63	-1,18	-1,55	2	4	2	142,16	65,69
186	Pirenzepine	2	1	0,42	2,42	3,13	2,38	-0,35	-0,22	3	7	1	351,41	63,98
187	Pramipexole	2	1	0,36	0,97	2,69	0,78	1,17	-0,52	3	3	2	211,33	52,49
188	Pravastatin	2	1	1,51	1,89	2,11	2,64	2,05	1,44	11	7	4	424,54	135,57
189	Pregabalin	2	1	0,78	0,97	0,93	2,65	-0,92	-1,38	5	3	2	159,23	68,80
190	Ranitidine	2	1	0,27	1,97	1,74	0,81	0,67	0,28	10	3	2	314,41	84,19
191	Risedronate	2	1	1,56	2,91	2,15	2,09	-2,62	-5,59	4	8	4	283,12	161,04
192	Rolitetracycline	2	1	1,86	4,09	4,02	1,23	0,47	-2,71	5	10	6	527,58	184,85
193	Sitagliptin	2	1	0,21	1,61	2,52	1,55	0,69	1,12	6	6	1	407,32	71,38
194	Sotalol	2	1	0,74	1,74	1,98	1,35	0,23	-1,37	6	5	3	272,37	88,93
195	Sulpiride	2	1	0,72	2,15	1,78	1,21	1,11	-1,02	7	6	2	341,43	108,62
196	Acetazolamide	2	2	0,85	1,50	2,55	1,45	-0,98	-0,26	3	7	2	222,25	121,43
197	Acyclovir	2	2	0,65	2,18	1,95	2,16	-2,42	-1,76	4	8	3	225,21	112,32
198	Amisulpride	2	2	0,50	2,18	3,16	2,11	1,80	0,12	8	6	2	369,49	107,55
199	Atovaquone	2	2	0,31	1,21	2,54	2,48	6,35	6,17	2	3	1	366,85	59,10

200	Cefpodoxime	2	2	1,07	2,95	3,67	1,82	-0,41	-0,87	8	9	3	427,46	155,91
201	Cefprozil	2	2	1,55	2,89	3,22	1,48	-1,87	-2,55	6	7	4	389,43	143,96
202	Chlorothiazide	2	2	0,64	1,66	2,74	1,85	-1,00	-0,18	1	7	2	295,72	126,91
203	Chlorthalidone	2	2	1,01	1,98	3,05	2,26	0,45	-0,74	2	6	3	338,77	120,90
204	Cinoxacin	2	2	0,57	1,55	2,05	2,67	1,74	-0,71	2	5	1	262,22	91,03
205	Cloxacillin	2	2	0,84	2,32	3,27	0,74	2,52	2,52	5	6	2	435,89	116,46
206	Enoxacin	2	2	0,73	1,96	2,45	1,26	-1,60	-1,09	3	7	2	320,33	85,36
207	Erythromycin stearate	2	2	0,91	4,49	3,30	0,58	1,61	11,16	25	15	4	1018,40	203,27
208	Felbamate	2	2	0,89	1,19	2,12	1,10	0,50	1,19	7	6	2	238,25	110,07
209	Fleroxacin	2	2	0,57	1,81	2,37	2,51	-0,33	-0,52	4	6	1	369,35	61,72
210	Furosemide	2	2	1,25	1,50	2,37	2,36	1,90	3,00	5	6	3	330,75	130,02
211	Niclosamide	2	2	0,77	0,78	2,67	2,50	4,34	5,40	4	2	2	327,13	99,56
212	Rifaximin	2	2	1,99	3,89	4,88	1,75	7,24	2,10	3	11	5	785,90	205,94
213	Roxithromycin	2	2	1,05	5,12	2,90	2,64	2,29	2,99	13	15	5	837,07	224,19
214	Sulfadiazine	2	2	0,59	1,40	2,58	2,91	0,10	-0,41	3	6	2	250,28	99,95
215	Sulfamethizole	2	2	0,59	1,26	2,71	1,01	0,42	0,07	3	6	2	270,33	102,86
216	Sulfisoxazole	2	2	0,59	1,31	2,44	3,06	0,22	0,38	3	4	2	267,31	102,81
217	Valsartan	2	2	1,21	1,82	3,32	5,08	4,86	4,73	11	8	2	435,53	113,69
218	Vitamin B2	2	2	1,33	2,69	2,71	1,69	-0,73	-2,02	5	10	5	376,37	161,81
219	Paliperidone	2	2	0,31	2,01	2,48	2,15	1,12	0,98	4	5	1	426,50	76,16
220	Iopanoic acid	2	2	0,85	0,74	2,00	2,27	4,70	4,17	4	3	2	570,94	68,50

2.2. Used methods and algorithms

The Kruskal–Wallis test is a non-parametric statistical method used to compare three or more independent groups to determine whether they originate from the same distribution. Unlike one-way ANOVA, it does not assume normality of data, making it suitable for non-normally distributed or ordinal data. The test ranks all values across groups and evaluates whether the distributions of ranks differ significantly between the groups. The null hypothesis (H_0) states that all groups have the same median, while the alternative hypothesis (H_1) suggests that at least one group differs. If the p-value is below a chosen significance level (commonly 0.05), H_0 is rejected, indicating a statistically significant difference among the groups [25-27].

The Probabilistic Neural Network is a type of feedforward neural network used primarily for classification tasks. It is based on Bayesian theory and uses kernel-based probability estimation to determine the likelihood that an input belongs to a particular class. The network architecture consists of four layers: input, pattern (where each neuron represents a training sample), summation (which computes class-wise probabilities), and output (which selects the class with the highest probability). Probabilistic Neural Network offer several advantages, including fast training, high classification accuracy, and robustness to noisy data. However, they can be memory-intensive and slower during prediction, especially when the training dataset is large, as the model stores all training samples explicitly [28, 29].

Together, the Kruskal–Wallis test and Probabilistic Neural Network can be useful tools in chemometric or biomedical data analysis, particularly when assessing group differences or building predictive models based on biological or chemical descriptors.

2.3. Prediction of BDDCS class according to metabolism (extensive or poor metabolism)

The Kruskal-Wallis test was calculated for 11 molecular descriptors. The results are listed in Table 2.2. The table value of χ^2 at a significance level of 5% with 1 degree of freedom is 3,80.

Table 2.2. Results of the Kruskal-Wallis test calculation

	Descriptor										
	A	B	S	E	logP	logD	RotB	HBA	HBD	MW	PSA
χ^2	**55,98**	**16,65**	0,25	0,10	**66,70**	**64,26**	0,010	**25,41**	**51,13**	**4,96**	**39,62**

The results show that descriptors dipolarity / polarizability, excess molar refraction and number of rotatable bonds are not influenced on classification / prediction of BDDCS class according to metabolism (extensive or poor metabolism).

To assess multicollinearity between descriptors, we calculated the correlation coefficients between each pair of parameters. The calculated values of the correlation coefficients are presented in Table 2.3.

So parameters hydrogen bond acidity and number of HBond donors are multicollinear, hydrogen bond acidity and number of HBond acceptors are multicollinear. And we left one parameter from each pair, which is characterized by higher value of χ^2 (highest influence on prediction of BDDCS class according to metabolism) − hydrogen bond acidity and number of HBond acceptors.

So, for classification / prediction of metabolism (extensive or poor) we have recommended six molecular descriptors: hydrogen bond acidity, partition coefficient, distribution coefficient, number of HBond acceptors, molecular weight and polar surface area.

Table 2.3. Calculated values of correlation coefficients between each pair of descriptors

Descriptor	A	B	S	E	logP	logD	RotB	HBA	HBD	MW	PSA
A	1,0000	0,5369	0,2897	-0,0623	-0,4328	-0,4127	0,1592	0,4622	**0,8908**	0,2006	0,7670
B	0,5369	1,0000	0,5781	-0,0503	-0,2118	-0,1595	0,4977	**0,8449**	0,6459	0,7366	0,8155
S	0,2897	0,5781	1,0000	-0,0028	0,0741	0,0756	0,2455	0,4975	0,2300	0,6018	0,5053
E	-0,0623	-0,0503	-0,0028	1,0000	0,0720	0,0343	0,0117	-0,0097	-0,0784	-0,0417	-0,0567
logP	-0,4328	-0,2118	0,0741	0,0720	1,0000	0,7855	0,1623	-0,2157	-0,4547	0,3719	-0,4194
logD	-0,4127	-0,1595	0,0756	0,0343	0,7855	1,0000	0,1770	-0,0874	-0,3949	0,4057	-0,2919
RotB	0,1592	0,4977	0,2455	0,0117	0,1623	0,1770	1,0000	0,4658	0,2345	0,6097	0,3756
HBA	0,4622	**0,8449**	0,4975	-0,0097	-0,2157	-0,0874	0,4658	1,0000	0,5879	0,6708	0,8003
HBD	**0,8908**	0,6459	0,2300	-0,0784	-0,4547	-0,3949	0,2345	0,5879	1,0000	0,2613	0,8056
MW	0,2006	0,7366	0,6018	-0,0417	0,3719	0,4057	0,6097	0,6708	0,2613	1,0000	0,4997
PSA	0,7670	0,8155	0,5053	-0,0567	-0,4194	-0,2919	0,3756	0,8003	0,8056	0,4997	1,0000

To evaluate the relevance of selected descriptors we divided 220 drugs into a training sub-set (85%) and a testing sub-set (15%) and implemented a Probabilistic Neural Network as the prediction tool. For the implementation of the Probabilistic Neural Network, we used different spread values ranging from 0,1 to 1,0. Results are presented on Table 2.4. The mistakes are highlighted in bold.

The lowest part of incorrect classified drugs was observed at spread value ranging from 0,8 to 1,0. At these spread values only Azathioprine is wrong classified.

We also checked the effect of the parameter «molecular weight» on the classification / prediction of BDDCS class according to metabolism (extensive or poor metabolism), because its χ^2 much less than for other descriptors (Table 2.2). Removing descriptor «molecular weight» from the parameter set led to a significant increase in the proportion of misclassified drugs, so this descriptor is informative and necessary.

Corresponding results are presented on Table 2.5. Proportion of misclassified drugs was increased from 3,03% till 15,15% (from one to five incorrect classified drugs).

Table 2.4. Results of prediction of metabolism (extensive or poor) using Probabilistic Neural Network

Drug	Correct BDDCS class: 1 – extensive metabolism, 2 – poor metabolism	Predicted class using Probabilistic Neural Network at different spread value									
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Acetohexamide	1	1	1	1	1	1	1	1	1	1	1
Aliskiren	1	1	1	1	1	1	1	1	1	1	1
Alosetron	1	1	1	1	1	1	1	1	1	1	1
Alprazolam	1	1	1	1	1	1	1	1	1	1	1
Ambrisentan	1	1	1	1	1	1	1	1	1	1	1
Aminophenazone	1	1	1	1	1	1	1	1	1	1	1
Amlodipine	1	1	1	1	1	1	1	1	1	1	1
Amoxapine	1	1	1	1	1	1	1	1	1	1	1
Anastrozole	1	1	1	1	1	1	1	1	1	1	1
Asenapine	1	1	1	1	1	1	1	1	1	1	1
Azathioprine	1	1	1	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Bambuterol	1	1	1	1	1	1	1	1	1	1	1
Melphalan	1	1	1	1	1	1	1	1	1	1	1
Omeprazole	1	1	1	1	1	1	1	1	1	1	1
Prochlorperazine	1	1	1	1	1	1	1	1	1	1	1
Thioridazine	1	1	1	1	1	1	1	1	1	1	1
Triamcinolone acetonide	1	1	1	1	1	1	1	1	1	1	1
Aminoglutethimide	1	1	1	1	1	1	1	1	1	1	1
Aripiprazole	1	1	1	1	1	1	1	1	1	1	1
Armodafinil	1	1	1	1	1	1	1	1	1	1	1
Astemizole	1	1	1	1	1	1	1	1	1	1	1

Bevantolol	1	1	1	1	1	1	1	1	1	1	1
Bicalutamide	1	1	1	1	1	1	1	1	1	1	1
Bosentan	1	1	1	1	1	1	1	1	1	1	1
Albuterol	2	**1**	2	2	2	2	2	2	2	2	2
Amiloride	2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	2	2	2
Amoxicillin	2	**1**	**1**	**1**	**1**	2	2	2	2	2	2
Atenolol	2	1	2	2	2	2	2	2	2	2	2
Baclofen	2	**1**	**1**	**1**	2	2	2	2	2	2	2
Cadralazine	2	**1**	**1**	**1**	**1**	2	2	2	2	2	2
Captopril	2	**1**	**1**	**1**	2	2	2	2	2	2	2
Acyclovir	2	**1**	**1**	**1**	2	2	2	2	2	2	2
Cefpodoxime	2	**1**	**1**	**1**	**1**	**1**	**1**	2	2	2	2

Table 2.5. Results of prediction of metabolism (extensive or poor) using Probabilistic Neural Network without parameter «molecular weight»

Drug	Correct BDDCS class: 1 – extensive metabolism, 2 – poor metabolism	Predicted class using Probabilistic Neural Network at different spread value									
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Acetohexamide	1	1	1	1	1	1	1	1	1	1	1
Aliskiren	1	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Alosetron	1	1	1	1	1	1	1	1	1	1	1
Alprazolam	1	1	1	1	1	1	1	1	1	1	1
Ambrisentan	1	1	1	1	1	1	1	1	1	1	1
Aminophenazone	1	1	1	1	1	1	1	1	1	1	1
Amlodipine	1	1	1	1	1	1	1	1	1	1	1
Amoxapine	1	1	1	1	1	1	1	1	1	1	1
Anastrozole	1	1	1	1	1	1	1	1	1	1	1
Asenapine	1	1	1	1	1	1	1	1	1	1	1
Azathioprine	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Bambuterol	1	1	1	1	1	1	1	1	1	1	1
Melphalan	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Omeprazole	1	1	1	1	1	1	1	1	1	1	1
Prochlorperazine	1	1	1	1	1	1	1	1	1	1	1
Thioridazine	1	1	1	1	1	1	1	1	1	1	1
Triamcinolone acetonide	1	1	1	1	1	1	1	1	1	1	1
Aminoglutethimide	1	1	1	1	1	1	1	1	1	1	1
Aripiprazole	1	1	1	1	1	1	1	1	1	1	1
Armodafinil	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**

Astemizole	1	1	1	1	1	1	1	1	1	1	1
Bevantolol	1	1	1	1	1	1	1	1	1	1	1
Bicalutamide	1	1	1	1	1	1	1	1	1	1	1
Bosentan	1	1	1	1	1	1	1	1	1	1	1
Albuterol	2	**1**	2	2	2	2	2	2	2	2	2
Amiloride	2	**1**	2	2	2	2	2	2	2	2	2
Amoxicillin	2	2	2	2	2	2	2	2	2	2	2
Atenolol	2	2	2	2	2	2	2	2	2	2	2
Baclofen	2	2	2	2	2	2	2	2	2	2	2
Cadralazine	2	2	2	2	2	2	2	2	2	2	2
Captopril	2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Acyclovir	2	**1**	2	2	2	2	2	2	2	2	2
Cefpodoxime	2	2	2	2	2	2	2	2	2	2	2

2.4. Prediction of BDDCS class according to solubility (high or low solubility)

The Kruskal-Wallis test was calculated for 11 molecular descriptors. The results are listed in Table 2.6. The table value of χ^2 at a significance level of 5% with 1 degree of freedom is 3,80.

Table 2.6. Results of the Kruskal-Wallis test calculation

	Descriptor										
	A	B	S	E	logP	logD	RotB	HBA	HBD	MW	PSA
χ^2	1,19	0,00	**18,69**	1,71	**18,71**	**38,81**	0,74	0,21	**5,44**	**17,57**	0,01

For classification / prediction of solubility (high or low) we have recommended five molecular descriptors: dipolarity / polarizability, partition coefficient, distribution coefficient, number of HBond donors, molecular weight. There are no multicollinear parameters (Table 2.3).

To evaluate the relevance of selected descriptors we divided 220 drugs into a training sub-set (85%) and a testing sub-set (15%) and implemented a Probabilistic Neural Network as the prediction tool. For the implementation of the Probabilistic Neural Network, we used different spread values ranging from 0,1 to 1,0. Results are presented on Table 2.7. The mistakes are highlighted in bold.

The lowest part of incorrect classified drugs was observed at spread value ranging from 0,7 to 1,0. At these spread values four drugs are wrong classified: Bambuterol, Aminoglutethimide, Aripiprazole, Atenolol. So, accuracy of prediction is equal to 88%.

Table 2.7. Results of prediction of solubility (high or low) using Probabilistic Neural Network

Drug	Correct BDDCS class: 1 – high solubility, 2 – low solubility	Predicted class using Probabilistic Neural Network at different spread value									
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Acetohexamide	1	1	1	1	1	1	1	1	1	1	1
Aliskiren	1	1	1	1	1	1	1	1	1	1	1
Alosetron	1	1	1	1	1	1	1	1	1	1	1
Alprazolam	1	1	1	1	1	1	1	1	1	1	1
Ambrisentan	1	1	1	1	1	1	1	1	1	1	1
Aminophenazone	1	1	1	1	1	1	1	1	1	1	1
Amlodipine	1	1	1	1	1	1	1	1	1	1	1
Amoxapine	1	1	1	1	1	1	1	1	1	1	1
Anastrozole	1	1	1	1	1	1	1	1	1	1	1
Asenapine	2	1	2	2	2	2	2	2	2	2	2
Azathioprine	2	2	2	2	2	2	2	2	2	2	2
Bambuterol	2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Melphalan	2	2	2	2	2	2	2	2	2	2	2
Omeprazole	2	2	2	2	2	2	2	2	2	2	2
Prochlorperazine	2	2	2	2	2	2	2	2	2	2	2
Thioridazine	2	2	2	2	2	2	2	2	2	2	2
Triamcinolone acetonide	2	**1**	2	2	2	2	2	2	2	2	2
Aminoglutethimide	2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Aripiprazole	1	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Armodafinil	1	1	1	1	1	1	1	1	1	1	1
Astemizole	1	1	1	1	1	1	1	1	1	1	1
Bevantolol	1	1	1	1	1	1	1	1	1	1	1

Bicalutamide	1	1	1	1	1	1	1	1	1	1	1
Bosentan	1	1	1	1	1	1	1	1	1	1	1
Albuterol	1	1	1	1	1	1	1	1	1	1	1
Amiloride	1	1	1	1	1	1	1	1	1	1	1
Amoxicillin	1	1	1	1	1	1	1	1	1	1	1
Atenolol	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Baclofen	2	**1**	**1**	**1**	**1**	**1**	**1**	2	2	2	2
Cadralazine	2	**1**	2	2	2	2	2	2	2	2	2
Captopril	2	2	2	2	2	2	2	2	2	2	2
Acyclovir	2	2	2	2	2	2	2	2	2	2	2
Cefpodoxime	2	**1**	**1**	**1**	**1**	**1**	2	2	2	2	2

We also checked the effect of the parameter «number of HBond donors» on the classification / prediction of BDDCS class according to solubility (high or low solubility), because its χ^2 much less than for other descriptors (Table 2.6). Removing descriptor «number of HBond donors» from the parameter set led to a significant increase in the proportion of misclassified drugs, so this descriptor is informative and necessary.

Corresponding results are presented on Table 2.8. Proportion of misclassified drugs was increased from 12,12% till 21,21% (from four to seven incorrect classified drugs).

Conclusions to section 2

Six molecular descriptors are recommended for accuracy classification / prediction of metabolism (extensive or poor): hydrogen bond acidity, partition coefficient, distribution coefficient, number of HBond acceptors, molecular weight and polar surface area.

Five molecular descriptors are recommended for accuracy classification / prediction of solubility (high or low): dipolarity / polarizability, partition coefficient, distribution coefficient, number of HBond donors, molecular weight.

Results demonstrate the effectiveness of the proposed models in predicting the metabolism and solubility categories of tableted drugs, offering a valuable tool for early-stage drug development. The high accuracy of the models emphasizes the importance of selecting the right molecular descriptors and optimizing model parameters to achieve reliable predictions.

Table 2.8. Results of prediction of solubility (high or low) using Probabilistic Neural Network without descriptor «number of HBond donors»

Drug	Correct BDDCS class: 1 – high solubility, 2 – low solubility	Predicted class using Probabilistic Neural Network at different spread value									
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
Acetohexamide	1	1	1	1	1	1	1	1	1	1	1
Aliskiren	1	1	1	1	1	1	1	1	1	1	1
Alosetron	1	1	1	1	1	1	1	1	1	1	1
Alprazolam	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Ambrisentan	1	1	1	1	1	1	1	1	1	1	1
Aminophenazone	1	1	1	1	1	1	1	1	1	1	1
Amlodipine	1	1	1	1	1	1	1	1	1	1	1
Amoxapine	1	1	1	1	1	1	1	1	1	1	1
Anastrozole	1	1	1	1	1	1	1	1	1	1	1
Asenapine	2	**1**	2	2	2	2	2	2	2	2	2
Azathioprine	2	2	2	2	2	2	2	2	2	2	2
Bambuterol	2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Melphalan	2	2	2	2	2	2	2	2	2	2	2
Omeprazole	2	2	2	2	2	2	2	2	2	2	2
Prochlorperazine	2	2	2	2	2	2	2	2	2	2	2
Thioridazine	2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Triamcinolone acetonide	2	**1**	2	2	2	2	2	2	2	2	2
Aminoglutethimide	2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Aripiprazole	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Armodafinil	1	1	1	1	1	1	1	1	1	1	1
Astemizole	1	1	1	1	1	1	1	1	1	1	1

Bevantolol	1	1	1	1	1	1	1	1	1	1	1
Bicalutamide	1	1	1	1	1	1	1	1	1	1	1
Bosentan	1	1	1	1	1	1	1	1	1	1	1
Albuterol	1	1	1	1	1	1	1	1	1	1	1
Amiloride	1	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Amoxicillin	1	1	1	1	1	1	1	1	1	1	1
Atenolol	1	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Baclofen	2	**1**	**1**	**1**	**1**	**1**	**1**	2	2	2	2
Cadralazine	2	**1**	2	2	2	2	2	2	2	2	2
Captopril	2	2	2	2	2	2	2	2	2	2	2
Acyclovir	2	2	2	2	2	2	2	2	2	2	2
Cefpodoxime	2	1	1	1	1	1	2	2	2	2	2

CONCLUSIONS

1. Predictive models were developed for the classification of tableted drugs based on their metabolism and solubility profiles according to the Biopharmaceutical Drug Disposition Classification System.

2. For metabolism prediction (extensive vs. poor), six molecular descriptors were identified: hydrogen bond acidity, partition coefficient, distribution coefficient, number of hydrogen bond acceptors, molecular weight, and polar surface area, achieving an accuracy of 97%.

3. For solubility classification (high vs. low), five molecular descriptors were used: dipolarity / polarizability, partition coefficient, distribution coefficient, number of hydrogen bond donors, and molecular weight, resulting in an accuracy of 88%.

4. The difference in accuracy can be attributed to the more complex nature of solubility, which is influenced by additional factors not fully captured by the selected descriptors.

REFERENCES

1. Talevi, A., & Bellera, C. L. (2022). Biopharmaceutics drug disposition classification system. In *The ADME Encyclopedia: A Comprehensive Guide on Biopharmacy and Pharmacokinetics* (pp. 185-189). Cham: Springer International Publishing.

2. Charalabidis, A., Sfouni, M., Bergström, C., & Macheras, P. (2019). The biopharmaceutics classification system (BCS) and the biopharmaceutics drug disposition classification system (BDDCS): beyond guidelines. *International journal of pharmaceutics*, *566*, 264-281.

3. Ji, L. I., Larregieu, C. A., & Benet, L. Z. (2016). Classification of natural products as sources of drugs according to the biopharmaceutics drug disposition classification system (BDDCS). *Chinese journal of natural medicines*, *14*(12), 888-897.

4. Idkaidek, N. M. (2014). Interplay of biopharmaceutics, biopharmaceutics drug disposition and salivary excretion classification systems. *Saudi Pharmaceutical Journal*, *22*(1), 79-81.

5. Benet, L. Z. (2013). The role of BCS (biopharmaceutics classification system) and BDDCS (biopharmaceutics drug disposition classification system) in drug development. *Journal of pharmaceutical sciences*, *102*(1), 34-42.

6. Bowman, C. M., & Benet, L. Z. (2016). Hepatic clearance predictions from in vitro–in vivo extrapolation and the biopharmaceutics drug disposition classification system. *Drug Metabolism and Disposition*, *44*(11), 1731-1735.

7. Cristofoletti, R., Chiann, C., Dressman, J. B., & Storpirtis, S. (2013). A comparative analysis of biopharmaceutics classification system and biopharmaceutics drug disposition classification system: a cross-sectional survey with 500 bioequivalence studies. *Journal of pharmaceutical sciences*, *102*(9), 3136-3144.

8. Reyes, M., & Benet, L. Z. (2011). Effects of uremic toxins on transport and metabolism of different biopharmaceutics drug disposition classification system xenobiotics. *Journal of pharmaceutical sciences*, *100*(9), 3831-3842.

9. Kikuchi, R., de Morais, S. M., & Kalvass, J. C. (2013). In vitro P-glycoprotein efflux ratio can predict the in vivo brain penetration regardless of biopharmaceutics drug disposition classification system class. *Drug Metabolism and Disposition*, *41*(12), 2012-2017.

10. Jia, X., Chen, J., Cheng, H., Pan, X., Ke, Y., Fu, T., ... & Li, J. (2023). Use of surfactant-based amorphous solid dispersions for BDDCS class II drugs to enhance oral bioavailability: a case report of resveratrol. *International Journal of Pharmaceutics*, *641*, 123059.

11. Hosey, C. M., Chan, R., & Benet, L. Z. (2016). BDDCS predictions, self-correcting aspects of BDDCS assignments, BDDCS assignment corrections, and classification for more than 175 additional drugs. *The AAPS journal*, *18*, 251-260.

12. Alqahtani, M. S., Kazi, M., Alsenaidy, M. A., & Ahmad, M. Z. (2021). Advances in oral drug delivery. *Frontiers in pharmacology*, *12*, 618411.

13. Tran, T. T. V., Tayara, H., & Chong, K. T. (2023). Artificial intelligence in drug metabolism and excretion prediction: recent advances, challenges, and future perspectives. *Pharmaceutics*, *15*(4), 1260.

14. Bergström, C. A., & Larsson, P. (2018). Computational prediction of drug solubility in water-based systems: Qualitative and quantitative approaches used in the current drug discovery and development setting. *International journal of pharmaceutics*, *540*(1-2), 185-193.

15. Ashrafi, A., Teimouri, K., Aghazadeh, F., & Shayanfar, A. (2024). Neural Network Models for Predicting Solubility and Metabolism Class of Drugs in the Biopharmaceutics Drug Disposition Classification System (BDDCS). *European Journal of Drug Metabolism and Pharmacokinetics*, *49*(1), 1-6.

16. Litsa, E. E., Das, P., & Kavraki, L. E. (2021). Machine learning models in the prediction of drug metabolism: Challenges and future perspectives. *Expert opinion on drug metabolism & toxicology*, *17*(11), 1245-1247.

17. Wang, D., Liu, W., Shen, Z., Jiang, L., Wang, J., Li, S., & Li, H. (2020). Deep learning based drug metabolites prediction. *Frontiers in Pharmacology*, *10*, 1586.

18. PUSHKAROVA, Y., & TYMCHENKO, I. (2024). PREDICTION OF ORAL DRUG BIOAVAILABILITY BASED ON CHEMICAL STRUCTURE. *СУЧАСНА МЕДИЦИНА, ФАРМАЦІЯ ТА ПСИХОЛОГІЧНЕ ЗДОРОВ'Я*, 104.

19. 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.

20. Kaliuzhenko, A., & Pushkarova, Y. (2023). *Application of artificial neural networks for solving pharmaceutical issues. Grail Sci. 2023;(24): 766-9.*

21. Pushkarova, Y., Zaitseva, G., & Al Saker, M. (2022, September). Prediction of Toxicity of Phenols Using Artificial Neural Networks. In *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)* (pp. 493-496). IEEE.

22. 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.

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

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

25. Ostertagova, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. *Applied mechanics and materials*, *611*, 115-120.

26. Hecke, T. V. (2012). Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems*, *15*(2-3), 241-247.

27. MacFarland, T. W., Yates, J. M., MacFarland, T. W., & Yates, J. M. (2016). Kruskal–Wallis H-test for oneway analysis of variance (ANOVA) by ranks. *Introduction to nonparametric statistics for the biological sciences using R*, 177-211.

28. 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.

29. Zeinali, Y., & Story, B. A. (2017). Competitive probabilistic neural network. *Integrated Computer-Aided Engineering*, *24*(2), 105-118.

SUPPLEMENTARY INFORMATION

СЕРТИФІКАТ

№003

ЦИМ СЕРТИФІКАТОМ ЗАСВІДЧУЄТЬСЯ, ЩО

KAOUTHAR BAUAALAM

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

Kaouthar Bouaalam

PREDICTION OF METABOLISM AND SOLUBILITY OF TABLETED DRUGS ACCORDING TO BIOPHARMACEUTICAL DRUG DISPOSITION CLASSIFICATION SYSTEM

Department of Analytical, Physical and Colloid Chemistry

Scientific supervisor: associate professor Yaroslava Pushkarova

Keywords: solubility prediction, metabolism prediction, drug classification, chemometric methods, drug design.

Introduction. The researched topic is highly relevant and timely in the field of pharmaceutical sciences. It addresses critical challenges in drug development and patient safety by offering a method to predict key pharmacokinetic properties, such as metabolism and solubility, of tablet-form medications. This is particularly important in the current landscape of personalized medicine, where the aim is to tailor treatments to individual patients, enhancing drug efficacy and minimizing adverse effects.

The Biopharmaceutical Drug Disposition Classification System provides a framework for predicting the absorption, distribution, metabolism, and excretion of drugs based on their solubility and metabolism characteristics. By categorizing drugs into four distinct classes based on these two factors, BDDCS helps pharmaceutical scientists predict how a drug will behave in the body and the potential challenges in its formulation. This predictive power is essential for optimizing drug design and formulation strategies, particularly for drugs that may have issues with bioavailability, solubility, or metabolism.

Materials and methods. Data set – total 220 drugs, 11 descriptors. Methods of research: Kruskal-Wallis test, Probabilistic Neural Network, correlation analysis.

Results. Predictive models were developed for the classification of tableted drugs based on their metabolism and solubility profiles according to the

Biopharmaceutical Drug Disposition Classification System. Results demonstrate the effectiveness of the proposed models in predicting the metabolism and solubility categories of tableted drugs, offering a valuable tool for early-stage drug development. The high accuracy of the models emphasizes the importance of selecting the right molecular descriptors and optimizing model parameters to achieve reliable predictions.

Conclusions. For metabolism prediction (extensive vs. poor), six molecular descriptors were identified: hydrogen bond acidity, partition coefficient, distribution coefficient, number of hydrogen bond acceptors, molecular weight, and polar surface area, achieving an accuracy of 97%.

For solubility classification (high vs. low), five molecular descriptors were used: dipolarity / polarizability, partition coefficient, distribution coefficient, number of hydrogen bond donors, and molecular weight, resulting in an accuracy of 88%.

The difference in accuracy can be attributed to the more complex nature of solubility, which is influenced by additional factors not fully captured by the selected descriptors.