GDAŃSK UNIVERSITY OF TECHNOLOGY

Subject card

Subject name and code	QSAR, QSPR MODELING, PG_00070143		
Field of study	InfoBioChem		
Date of commencement of studies	February 2026	Academic year of realisation of subject	2025/2026
Education level	second-cycle studies	Subject group	Obligatory subject group in the field of study Subject group related to scientific research in the field of study
Mode of study	Full-time studies	Mode of delivery	at the university
Year of study	1	Language of instruction	Polish Polish language
Semester of study	1	ECTS credits	3.0
Learning profile	general academic profile	Assessment form	assessment
Conducting unit	Department of Physical Chemistry -> Faculty of Chemistry -> Faculties of Gdańsk University of Technology		
Name and surname of lecturer (lecturers)	Subject supervisor	prof. dr hab. inż. Adam Kloskowski	
	Teachers	dr hab. Agnieszka Gajewicz-Skrętna	

Lesson types

Lesson type	Lecture	Tutorial	Laboratory	Project	Seminar	SUM
Number of study hours	0.0	0.0	0.0	60.0	0.0	60

E-learning hours included: 0.0

eNauczanie source address: https://enauczanie.pg.edu.pl/2025/user/index.php?id=5441

Learning activity and number of study hours

Learning activity	Participation in didactic classes included in study plan	Participation in consultation hours	Self-study	SUM
Number of study hours	60	2.0	13.0	75

Subject objectives

This course introduces students to the theoretical foundations and practical applications of qualitative and quantitative structure-activity relationship (QSAR) modeling, as well as qualitative and quantitative structure-property relationship (QSPR) modeling. Students will gain theoretical knowledge and practical skills related to describing chemical structures, calculating molecular descriptors, and using modern computational tools to develop, validate, and interpret predictive models (Q SAR/Q SPR models). This course prepares students to independently apply computational methods to analyze the relationship between the structure of chemical compounds and their biological activity or physicochemical properties.

Learning outcomes

Course outcome	Subject outcome	Method of verification
[K7_W01] knows and understands the methods, techniques, and tools used to solve bioinformatics tasks, including molecular modeling.	The student demonstrates knowledge of the theoretical foundations of [Q]SAR/[Q]SPR modeling methods. The student also possesses the knowledge of machine learning required for the development and validation of classification and regression models applied in materials chemistry, computational toxicology, and drug design.	[SW1] Assessment of factual knowledge [SW2] Assessment of knowledge contained in presentation [SW3] Assessment of knowledge contained in written work and projects
[K7_U05] is able to select computer modeling methods and apply them to solve problems related to the operation and regulation of complex systems.	The student can independently develop and evaluate a simple QSAR/SPR model, as well as prepare a presentation that summarizes the key results. The student can also critically interpret the modeling outcomes and relate them to the current state of knowledge in the field.	[SU2] Assessment of ability to analyse information [SU1] Assessment of task fulfilment [SU3] Assessment of ability to use knowledge gained from the subject [SU4] Assessment of ability to use methods and tools [SU5] Assessment of ability to present the results of task
[K7_K01] is ready to determine the priorities of activities in the implementation of his/her own tasks or those assigned by others	Students understand the need for learning and inspire others to do the same, cooperate in groups by taking on different roles, exhibit creativity in determining priorities necessary for the realization of different tasks, and understand the social aspects of the practical use of knowledge and abilities, as well as the responsibility connected to them.	[SK1] Assessment of group work skills [SK2] Assessment of progress of work [SK3] Assessment of ability to organize work [SK4] Assessment of communication skills, including language correctness [SK5] Assessment of ability to solve problems that arise in practice
[K7_U01] is able to use databases in the field of exact and natural sciences - appropriate to the educational program InfoBioChem	The student can use specific chemical and biological databases to find the data needed to develop, analyze, and validate QSAR/QSPR models.	[SU1] Assessment of task fulfilment [SU2] Assessment of ability to analyse information [SU3] Assessment of ability to use knowledge gained from the subject [SU4] Assessment of ability to use methods and tools [SU5] Assessment of ability to present the results of task

Subject contents

Course content – project

The course content covers the following areas:

- The role of data quality in [Q]SAR/[Q]SPR modeling: methods for preliminary data inspection and reliable assessment of data quality issues, such as incomplete data, outliers, noisy data, and small sample size.
- The calculation of molecular descriptors using available tools such as *Mordred, PaDEL, RDKit, OCHEM* , etc.
- The preparation of input data for [Q]SAR/[Q]SPR modeling: selection of independent variables using methods such as stepwise regression (backward elimination or forward selection), genetic algorithms, and Elastic Net; techniques for splitting datasets into training and test sets; data standardization.
- The validation of [Q]SAR/[Q]SPR models: OECD recommendations for [Q]SAR/[Q]SPR validation; internal and external validation; descriptive statistics related to model's goodness-of-fit, robustness, and predictive ability; applicability domain assessment.
- The development of qualitative and/or quantitative [Q]SAR/[Q]SPR models using available computational platforms, such as *Orange Data Mining*: hyperparameter tuning, internal and external model validation, visualization and interpretation of modeling results.
- Overview of platforms such as *EPI Suite, OECD QSAR Toolbox* that provide ready-to-use [Q]SAR/[Q]SPR models.
- Final project: development and presentation of [Q]SAR/[Q]SPR modeling results for a selected classification and/or regression problem.

Prerequisites and co-requisites

Assessment methods and criteria

Subject passing criteria	Passing threshold	Percentage of the final grade
Laboratory classes – project / problem-based task	50.0%	100.0%

Recommended reading

Basic literature

- Roy K, Kar S, Das RN. A Primer on QSAR/QSPR Modeling: Fundamental Concepts (SpringerBriefs in Molecular Science). 2015, Springer. ISBN:9783319172804.
- Todeschini R, Consonni V. Molecular Descriptors for Chemoinformatics. 2009, WileyVCH Verlag GmbH & Co. KGaA. ISBN:9783527318520.

	Supplementary literature	• Dearden JC, Cronin MTD, Kaiser KLE. How not to develop a quantitative structureactivity or structureproperty relationship (QSAR/QSPR). *SAR and QSAR in Environmental Research*, 2009; 20(34), 241266. https://doi.org/10.1080/10629360902949567. • Gramatica P. Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, 2020:5(3). https://doi.org/10.4018/IJQSPR.20200701.oa1. • Gramatica P. Origin of the OECD Principles for QSAR Validation and Their Role in Changing the QSAR Paradigm Worldwide: An Historical Overview. *Journal of Chemometrics*, 2025; 39(3): e70014. https://doi.org/10.1002/cem.70014. • Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A. QSAR modeling: where have you been? Where are you going to? *Journal of Medicinal Chemistry*. 2014; 57(12):4977-5010. https://doi.org/10.1021/jm4004285.
	eResources addresses	
Example issues/ example questions/ tasks being completed	Example topics / questions:: • What is the difference between a QSAR model and a QSPR model? • What are the main types of molecular descriptors? • When are classification models more appropriate than regression models? • What are outliers, and how do they affect the quality of a QSAR/QSPR model? • How can the statistical significance of a QSAR/QSPR model be assessed? • What does overfitting mean in QSAR/QSPR modeling, and how can it be avoided? • What is the essence of the ToplissCostello rule? • What is the difference between internal and external validation of a QSAR/QSPR model? • Under what conditions can QSAR/QSPR model predictions be considered reliable? • What are the most common sources of error in QSAR/QSPR modeling? • Evaluate the precision of a classification model based on a confusion matrix. • Assess the statistical significance of a QSAR/QSPR model. • Compare the predictive abilities of two QSAR models developed using different machine learning methods. • Interpret the applicability domain of a QSAR model. • Employ a classification [Q]SAR/[Q]SPR model developed with the k-nearest neighbors method to classify five novel compounds. • Evaluate a QSAR/QSPR model with respect to OECD recommendations.	
Practical activites within the subject	Not applicable	

Document generated electronically. Does not require a seal or signature.