Bioinformatics

http://dx.doi.org/10.7124/bc.000B29
UDC 577.004.9:[543.51+543.52]

T.G. Maiula[1], O.A. Bieda[1, 2], T.V. Pylaieva[1], S.M. Yarmoluk[1]
[1] Institute of molecular biology and genetics of NAS of Ukraine,
Ukraine, 03143, Kyiv, Akademik Zabolotny str., 150
[2] LLC «Scientific service company "Otava"»
Ukraine, 03056, Kyiv, Borschagivska str., 117/129
t.h.maiula@imbg.org.ua

IN SILICO MODELING AND PREDICTION OF ANTIDIABETIC POTENTIAL OF BIOACTIVE COMPOUNDS FROM *GALEGA OFFICINALIS L.* NON-ALKALOID EXTRACT

Aim. *To develop and validate in silico the models for predicting the antidiabetic activity of natural compounds and to test their performance on representative medicinal plant components.* ***Materials and Methods.*** *Two machine learning models based on the XGBoost and LightGBM algorithms were constructed and verified using a set of compounds from traditional Chinese medicine (TCM) formulations with experimentally confirmed hypoglycemic activity. The validated models were subsequently applied to analyze the components of Galega officinalis L. (non-alkaloid fraction), whose composition was determined by gas and liquid chromatography-mass spectrometry (GC-MS and LC-MS).* ***Results.*** *The created models achieved an accuracy of 80 – 81% and were verified by correctly identifying active compounds among those known from TCM formulations to be effective in type 2 diabetes management. This confirms their ability to accurately classify bioactive natural substances. The models were applied to the components of the NA extract of G. officinalis L., a promising plant for subsequent studies on antidiabetic effects.* ***Conclusions.*** *The developed in silico models enable the prediction of antidiabetic activity of naturally derived compounds. Their verification on reference compound sets and application to G. officinalis L. extract demonstrate the potential of this approach for identifying low-toxicity bioactive substances.*

Keywords: *in silico, machine learning, antidiabetic properties, Galega officinalis L.*

Citation: Maiula T.G., Bieda O.A., Pylaieva T.V., Yarmoluk S.M. (2025) *In silico* modeling and prediction of antidiabetic potential of bioactive compounds from *Galega officinalis L.* non-alkaloid extract. *Biopolymers & Cell*, 4(41), 309—317. http://dx.doi.org/10.7124/bc.000B29

Introduction

Diabetes mellitus is a metabolic disorder characterized by chronic hyperglycemia resulting from the interaction of genetic and environmental factors [1, 2]. Its pathophysiology involves either an absolute or relative deficiency of insulin secretion or reduced sensitivity of peripheral tissues to insulin, leading to systemic disturbances in carbohydrate metabolism [3, 4]. Chronic hyperglycemia causes long-term damage to organs, particularly the eyes, kidneys, nervous, and cardiovascular systems [5].

Current treatment strategies include dietary therapy, physical exercise, hypoglycemic agents, and insulin administration [6—12]. However, oral hypoglycemic drugs, insulin secretagogues, and α-glucosidase inhibitors are limited by adverse effects such as gastrointestinal disorders, allergic reactions, metabolic disturbances, and hepatotoxicity [13]. Since their mechanisms are largely single-target, the search for safer, multitarget, and multifunctional therapies has become a medical priority.

TCM, whether applied as single or multicomponent preparations, exerts multitarget and integrative effects that improve metabolic regulation and organ protection, offering distinct advantages in managing chronic diseases, including diabetes [14]. These formulations demonstrate stable efficacy, fewer adverse effects, and suitability for long-term use by acting on multiple biochemical pathways, normalizing metabolism, and protecting organs such as the pancreas, liver, and kidneys.

Table 1. **Example of processed data structure**

molecule_chembl_id	pchembl_value	activity_label
CHEMBL100004	7.16	1.0
CHEMBL100013	5.56	0.0
CHEMBL100067	5.75	0.0
CHEMBL100150	7.25	1.0
CHEMBL100109	8.16	1.0

With the modernization of TCM, research on plant-based formulations for diabetes treatment has gained increasing attention [15—18]. Yet the complexity of multicomponent interactions makes manual analysis impractical. Consequently, data-driven approaches employing large-scale datasets and computational modeling are increasingly used to reveal intercomponent relationships and provide a more precise analytical basis for traditional medicine [19].

In this study, the machine learning models based on the XGBoost and LightGBM algorithms were used to predict the antidiabetic activity of natural compounds. These *in silico* models effectively classify complex pharmacological datasets, capture nonlinear relationships between molecular descriptors and activity, and resist overfitting [20, 21]. The models were validated using compounds from TCM formulations with confirmed hypoglycemic activity and subsequently applied to the components of the NA extract of *G. officinalis L.*, a promising candidate in diabetes-related research [22—26].

Therefore, the aim of this study was to develop and validate *in silico* models for predicting the antidiabetic activity of natural compounds and to test them on medicinal plant components.

Materials and Methods

To construct *in silico* machine learning models for predicting the antidiabetic activity of natural compounds, the data were obtained from the ChEMBL database, an open-access repository of bioactive small molecules with experimentally confirmed IC_{50} and EC_{50} values for diverse biological targets.

The dataset was downloaded in CSV format using the keyword "antidiabetic," yielding over 150,000 entries. Data preprocessing was conducted in Google Colab using *pandas*, *numpy*, *rdkit*, *joblib*, *xgboost*, and *lightgbm* libraries.

Entries lacking numerical pChEMBL values were excluded, and the remaining data were grouped by the unique molecular identifier (molecule_chembl_id), averaging pChEMBL values for each compound (Table 1).

For activity classification, the following rule was applied:

- active compounds (1): pChEMBL ≥ 7.0;
- inactive compounds (0): pChEMBL < 6.0;
- intermediate values (6.0 ≤ pChEMBL < 7.0) were excluded to ensure a clearer separation between the active and inactive classes.

For each compound identified by its molecule_chembl_id, the SMILES notation was retrieved via the ChEMBL Web Resource Client interface. Subsequently, molecular descriptors and structural fingerprints were calculated using the RDKit library, following the approach described in previous studies [20].

A total of 2,222 features were obtained for each molecule and used as input parameters for the models. The dataset comprised 26,628 active and 22,236 inactive compounds. The class distribution was sufficiently balanced, allowing model training without the need for additional resampling techniques such as oversampling or undersampling.

Machine learning models were constructed using two gradient boosting algorithms, XGBoost and LightGBM. The training process was carried out on 80% of the dataset, with the remaining 20% reserved for testing.

To enhance the robustness of the results, a five-fold cross-validation procedure was implemented. Model performance was evaluated using standard metrics, including Accuracy, Precision, Recall, and F1-score [20].

Results and Discussion

At the initial stage of the study, the aforementioned models were tested, and both demonstrated comparable accuracy values. Table 2 presents the classification performance obtained on the test set using these models.

The obtained results indicate a high predictive capability of both models.

Both models demonstrated >80% classification accuracy, a strong performance for bioinformatics datasets involving large descriptor sets. High precision values (77—79%) indicate that the models

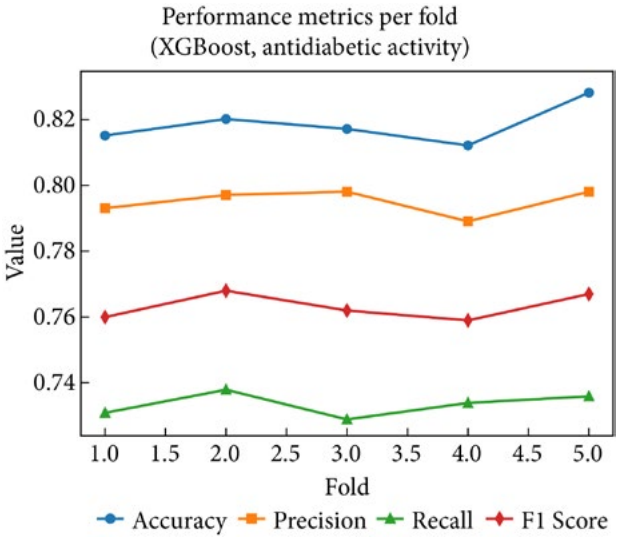

Fig. 1. Performance of the model based on the XGBoost algorithm during each cross-validation stage

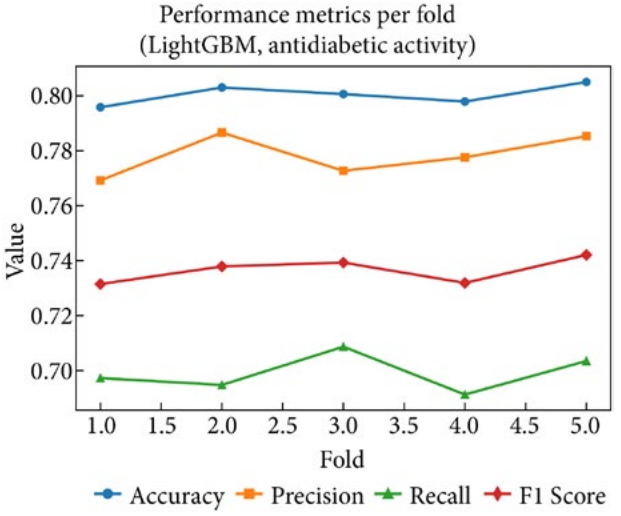

Fig. 2. Performance of the model based on the LightGBM algorithm during each cross-validation stage

Table 2. **Average model metrics after five-fold cross-validation**

Metric	XGBoost, %	LightGBM, %
precision	79.51	77.83
recall	73.36	69.91
f1-score	76.31	73.65
accuracy	81.82	80.05

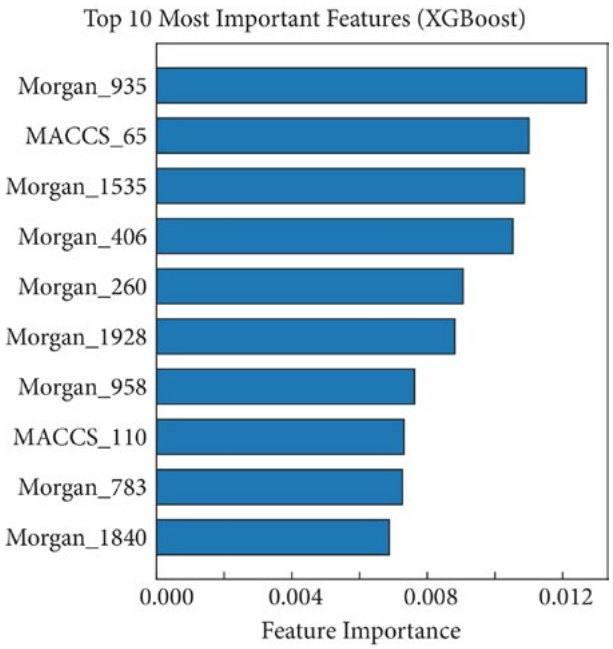

Fig. 3. Histogram of the most influential descriptors for the model based on the XGBoost algorithm

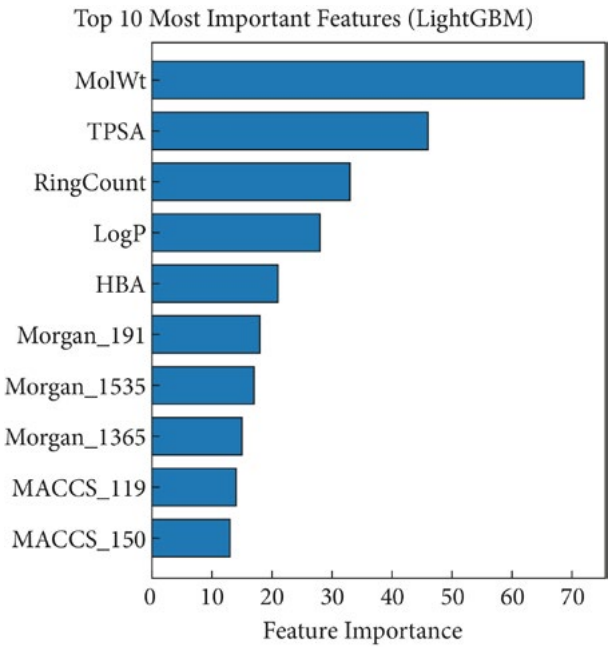

Fig. 4. Histogram of the most influential descriptors for the model based on the LightGBM algorithm

can accurately identify active compounds while maintaining a low rate of false positives, whereas recall values (70—73%) reflect their ability to effectively capture the majority of truly active samples. The F1-score values above 73% confirm a well-balanced trade-off between precision and recall, suggesting that the models are neither overfitted nor biased toward a particular class.

The average performance metrics presented in Table 2 were calculated based on five-fold cross-validation. Minor fluctuations in individual fold results were observed, but the changes remained consistent and smooth, indicating stable model training. The dynamics of these metrics are shown in Figures 1 and 2.

Thus, both models demonstrated sufficient stability and generalization capability, allowing their use for in silico prediction of the antidiabetic activity of natural compounds.

Similar to the previous study [20], a feature importance analysis was performed, revealing that the algorithms apply distinct strategies for decision formation. The model based on the XGBoost algorithm relies primarily on structural fingerprints that reflect local molecular fragments and topological characteristics, whereas LightGBM implements a more hybrid approach, combining the informativeness of global molecular descriptors with selected structural keys (Fig. 3, 4).

This distribution of feature importance is consistent with the notion that XGBoost is more effective at detecting patterns in large binary feature sets, while LightGBM more efficiently utilizes generalized physicochemical parameters for prediction.

The next stage of the study verified the developed models and assessed their ability to predict the antidiabetic activity of natural compounds. A test set of 118 compounds was collected through data mining from open-access databases and publications on TCM formulations with experimentally confirmed hypoglycemic effects. After removing duplicates and converting structures to SMILES format, the dataset included compounds with validated activity or reported affinity toward diabetes-related targets such as DPP-4, PTP1B, α-glucosidase, and GLUT4 [27].

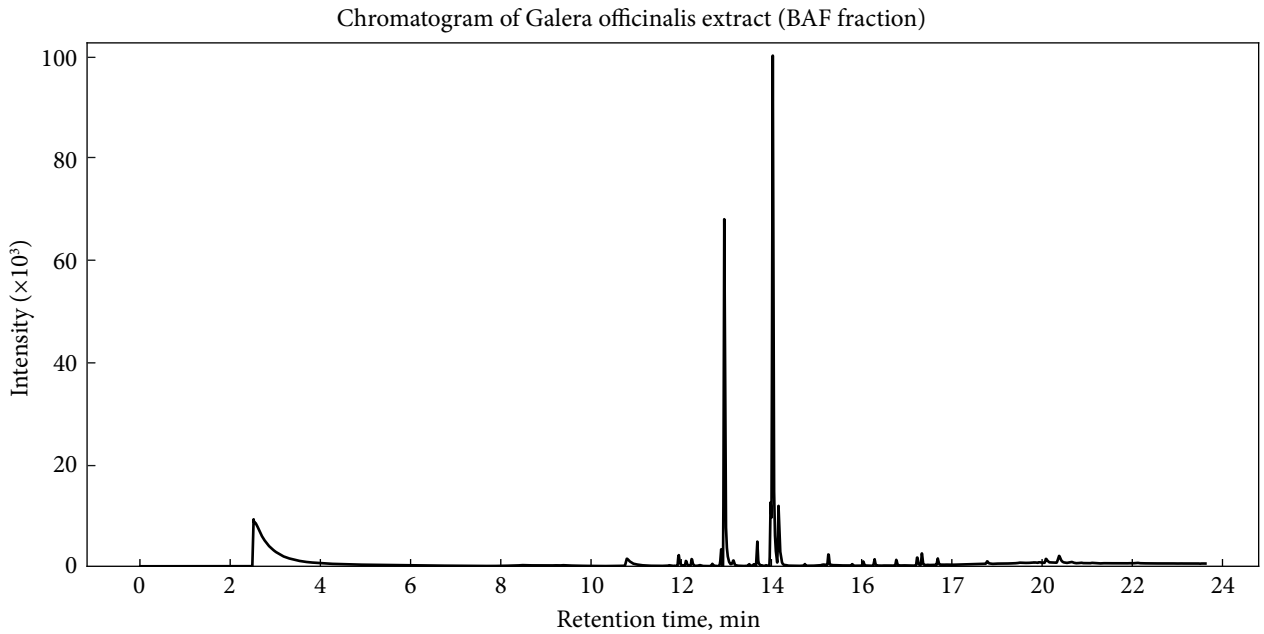

Fig. 5. Chromatogram of the NA extract of *G. officinalis L.* obtained using the EXPEC 2000 gas chromatograph coupled with the EXPEC 3700 mass spectrometric detector

Molecular descriptors and structural fingerprints were calculated in RDKit following the same procedure as for the training dataset. For each compound, 2,222 parameters were generated, covering both global descriptors and structural keys.

The test set was aligned with the training dataset to ensure model compatibility.

The complete table of predicted activities can be accessed as an Excel file in the public GitHub repository at: https://github.com/mTarik12/antidiabetic_activity

Based on the verification results, the model based on the XGBoost algorithm classified 44 compounds as active, while the LightGBM-based model identified 72. This distribution reflects the intrinsic characteristics of the algorithms: both correctly identified compounds with experimentally confirmed antidiabetic activity according to their prediction accuracy. These findings confirm the applicability of the models for further *in silico* prediction of plant-derived compounds that remain underexplored but promising for diabetes treatment.

The stricter classification threshold of XGBoost, which relies mainly on structural fingerprints capturing detailed molecular fragments, resulted in fewer hits. In contrast, LightGBM, which combines global physicochemical descriptors with structural keys, detected a broader range of potentially active substances.

The developed models were then applied to identify active components within the NA extract of *G. officinalis L.* Considering the plant's therapeutic potential and the low toxicity of its non-alkaloid compounds, extract samples were obtained from the authors of previous studies [22, 23, 25, 26].

According to earlier reports, α- and β-amyrin are key bioactive triterpenoids in *G. officinalis L.* with anti-inflammatory and antioxidant properties, showing beneficial effects in streptozotocin-induced diabetes models [28]. The hypoglycemic activity of the extract may also be related to the presence of phytol [22].

A preliminary analytical stage was conducted to determine the composition of the NA extract using gas and liquid chromatography-mass spectrometry

Table 3. **List of identified components of the NA extract of *G. officinalis L.* with corresponding retention times**

Compound name	Retention time, m	Peak area, count units
Triethyl citrate	10.79797	75000.695310
Neophytadiene	11.94355	27309.851560
3,7,11,15-Tetramethyl-2-hexadecen-1-ol	12.10805	12762.605470
Phytol	12.23545	16065.736330
Tetradecanoic acid, ethyl ester	12.76293	3590.013916
Hexadecanoic acid, ethyl ester	12.95597	772913.250000
9,12-Octadecadienoic acid, ethyl ester	13.98708	115664.234400
9,12,15-Octadecatrienoic acid, ethyl ester, (Z,Z,Z)-	14.02841	1091594.000000
Octadecanoic acid, ethyl ester	14.15953	126435.875000
Squalene epoxide	17.23298	18099.679690
Squalene	17.33651	21951.875000
beta-Amyrin	20.08881	12668.777340
alpha-Amyrin	20.38025	30043.490230
Cholestane, 1-vinyl-1-hydroxy- (BEST HIT)	20.65656	6368.985840

Table 4. **List of active components of the NA extract of *G. officinalis L.* identified by the model based on the LightGBM algorithm**

Compound name	CAS RN	Retention time, m
Neophytadiene	504-96-1	11.94355
3,7,11,15-Tetramethyl-2-hexadecen-1-ol	7541-49-3	12.10805
Phytol	150-86-7	12.23545

Table 5. **List of alkaloid compounds identified in the alkaloid fraction of *G. officinalis L.* extract**

Compound name	Searching m/z	Retention time in the NA fraction, min	Retention time in the alkaloid fraction, min
Galegine	128	2.7—128 (SIM)	1.9–128 (SIM)
Hydroxygalegine	144	1.6 : 10.1	1.6
Metformin	130	11.4 : 14.1	1.4

(Fig. 5, Table 3). The NA fraction was analyzed on an EXPEC 2000 gas chromatograph with an EXPEC 3700 single quadrupole mass spectrometer. GC-MS was performed on a BP-5MS column (30 m × 0.25 mm i.d.) with helium as the carrier gas (1.5 mL/min). The temperature was held at 50 °C for 5 min, ramped to 300 °C at 15 °C/min, and maintained for 5 min. Mass chromatograms were recorded within an m/z range of 35—600 at a scan rate of 1000 (m/z units)/min.

The BEST HIT mark, assigned to the last identified compound with a retention time of 20.656 min, indicates its correspondence to one of the main peaks on the chromatogram and suggests it as a marker compound of this fraction due to its physicochemical similarity to the observed signal.

To assess the potential antidiabetic activity of major components, the identified compounds were converted into SMILES format, and molecular descriptors and fingerprints were calculated for prediction using the verified models. This enabled estimation of which compounds from the experimentally identified composition may exhibit antidiabetic potential.

According to the prediction results, the model based on the LightGBM algorithm identified three potentially active components in the NA extract of *G. officinalis L.* (Table 4). In contrast, the stricter XGBoost-based model, which relies mainly on

structural fingerprints, did not detect active compounds among the analyzed samples. This supports earlier observations that, for phytochemicals, global physicochemical descriptors play a more decisive role in biological activity, which is better captured by the LightGBM algorithm.

According to the previous studies, these components may contribute to the hypoglycemic effect of the extracts [22]. Searches in chemical databases (including CAS registry data) confirmed that these compounds exhibit antidiabetic, antibacterial, and anti-inflammatory properties. For α-amyrin and β-amyrin, anti-inflammatory activity predominates; hence, their hypoglycemic potential may not be recognized by the developed models, which were primarily trained on compounds with direct antidiabetic mechanisms. This underscores the complexity of identifying synergistic effects in plant extracts using machine learning, as their bioactivity often results from multifactorial molecular interactions.

To ensure that the observed activities were unrelated to alkaloids naturally occurring in this plant, LC-MS analysis was performed on the NA extract of *G. officinalis L.* using an Agilent 1260 Infinity II chromatograph coupled with an LC/MSD iQ mass spectrometer. Two extract types were analyzed: the non-alkaloid fraction (chloroform solvent) and the alkaloid fraction (water solvent).

Both samples were filtered through 0.2 μm syringe filters, diluted (1:10), and analyzed under LC-MS conditions. According to literature data [29—32], the main alkaloids of G. officinalis L. are guanidine derivatives — primarily galegine and hydroxygalegine, with smaller amounts of paragalegine and galeganine. Retention times of $[M+H]^+$ ions were used for identification, and summarized results are presented in Table 5.

As shown in the table, the retention times of peaks reconstructed from $[M+H]^+$ ions did not coincide between the alkaloid and non-alkaloid fractions, indicating the absence of typical guanidine alkaloids in the NA fraction. For instance, the peak corresponding to galegine in the alkaloid fraction was not observed in the chloroform extract, while a peak at a different retention time likely represented another compound with distinct chromatographic behavior. These results confirm that the NA extract of *G. officinalis L.* lacks the major alkaloids characteristic of this species, consistent with its preliminary characterization and supporting its use for independent assessment of biological activity of non-alkaloid components.

Conclusions

The developed and validated machine learning models effectively predicted the antidiabetic activity of natural compounds. Their application to the components of *G. officinalis L.* enabled an *in silico* analysis of the NA extract and identification of substances with potential hypoglycemic activity. Both XGBoost- and LightGBM-based models showed strong classification accuracy and consistent performance during cross-validation.

Chromatographic analysis confirmed the absence of major guanidine alkaloids (galegine, hydroxygalegine) in the NA fraction, while the plant itself is known to exhibit antidiabetic effects. This finding suggests that the observed activity originates from non-alkaloid constituents, several of which were predicted by our models and independently supported by chemical knowledge bases as biologically active.

These results emphasize the importance of integrated computational and analytical approaches for studying complex phytochemical systems, where biological effects often arise from synergistic interactions of multiple compounds. The developed workflow combining machine learning with chromatography-based validation provides a powerful tool for rapid identification of bioactive components, reducing time and cost in early-stage screening.

Future work will include experimental validation of the predicted compounds in *in vitro* and *in vivo* models and expansion of the developed algorithms to other classes of natural bioactive molecules. This approach can form the basis for systematic discovery of safe and multitarget phytocomplexes with pronounced antidiabetic potential.

Acknowledgments. The authors express their sincere gratitude to Prof. Sybirna N.O. and Dr. Hachkova H.Ya. for kindly providing the *Galega officinalis* extracts used in this study.

Funding. This research was supported by the Institute of Molecular Biology and Genetics of the National Academy of Sciences of Ukraine under the Simons Foundation Grant for Ukrainian Institutions No. SFI-PD-Ukraine-00017453.

REFERENCES

1. *Zhang Y, Cui Y, Li Y, et al., and Miao D.* Network analysis of depressive and anxiety symptoms in older Chinese adults with diabetes mellitus. *Front Psychiatry.* 2024; **15**:1328857.
2. Öztürk *E, Arslan AKK, Yerer MB, Bishayee A.* Resveratrol and diabetes: A critical review of clinical studies. *Biomed Pharmacother.* 2017; **95**:230—4.
3. *Porte D Jr, Baskin DG, Schwartz MW.* Insulin signaling in the central nervous system: a critical role in metabolic homeostasis and disease from C. elegans to humans. *Diabetes.* 2005; **54**(5):1264—76.
4. *Abdul-Ghani MA, Jenkinson CP, Richardson DK, et al., and DeFronzo RA.* Insulin secretion and action in subjects with impaired fasting glucose and impaired glucose tolerance: results from the Veterans Administration Genetic Epidemiology Study. *Diabetes.* 2006; **55**(5):1430—5.
5. *Schuster DP, Duvuuri V.* Diabetes mellitus. *Clin Podiatr Med Surg.* 2002; **19**(1):79—107.
6. *Esposito K, Maiorino MI, Di Palo C, et al., and Campanian Postprandial Hyperglycemia Study Group.* Adherence to a Mediterranean diet and glycaemic control in Type 2 diabetes mellitus. *Diabet Med.* 2009; **26**(9):900—7.
7. *Simpson HC, Simpson RW, Lousley S, et al., and Mann JI.* A high carbohydrate leguminous fibre diet improves all aspects of diabetic control. *Lancet.* 1981; **1**(8210):1—5.
8. *Balducci S, Sacchetti M, Haxhi J, et al., and Pugliese G.* Physical exercise as therapy for type 2 diabetes mellitus. *Diabetes Metab Res Rev.* 2014; **30**(Suppl 1):13—23.
9. *Prasathkumar M, Becky R, Anisha S, et al., and Sadhasivam S.* Evaluation of hypoglycemic therapeutics and nutritional supplementation for type 2 diabetes mellitus management: An insight on molecular approaches. *Biotechnol Lett.* 2022; **44**(2):203—38.
10. *Khan MK, Baig RU.* Pharmacotherapy of diabetes mellitus type 2: a review on various hypoglycemics. *J Clin Nurs Res.* 2022; **6**(5):57—67.
11. *Sağlam FS, Bektas H.* Risk perceptions of patients with type 2 diabetes mellitus regarding insulin therapy and diabetes complications: A cross-sectional study. *J Clin Nurs.* 2023; **32**(15—16):5010—27.
12. *Nkonge KM, Nkonge DK, Nkonge TN.* Insulin Therapy for the Management of Diabetes Mellitus: A Narrative Review of Innovative Treatment Strategies. *Diabetes Ther.* 2023; **14**(11):1801—31.
13. *Akinwunmi OA.* In vitro antioxidant, α-amylase and α-glucosidase activities of methanol extracts from three Momordica species. *Int J Phytomedicine.* 2020; **11**(1):8—14.
14. *Luo H, Li Y, Song H, et al., and Zhang Y.* Role of EZH2-mediated epigenetic modification on vascular smooth muscle in cardiovascular diseases: A mini-review. *Front Pharmacol.* 2024; **15**:1416992.
15. *Chen J, Dun WL.* Review on active ingredients of traditional Chinese medicine in the treatment of type 2 diabetes. *Jiangsu Sci Technol Inf.* 2016; **26**:35—8.
16. *Gao J, Jiang Z, Adams E, Van Schepdael A.* A fast and efficient method for screening and evaluation of hypoglycemic ingredients of Traditional Chinese Medicine acting on PTP1B by capillary electrophoresis. *J Pharm Biomed Anal.* 2024; **244**:116125.
17. *Zhou X, Guo Y, Yang K, et al., and Wang J.* The signaling pathways of traditional Chinese medicine in promoting diabetic wound healing. *J Ethnopharmacol.* 2022; **282**:114662.
18. *Li X, Geng-Ji JJ, Quan YY, et al., and Xie X.* Role of potential bioactive metabolites from traditional Chinese medicine for type 2 diabetes mellitus: An overview. *Front Pharmacol.* 2022; **13**:1023713.
19. *Zhang X, Yang Y, Zhou X.* Application of data mining in treatment of diabetes with traditional Chinese medicine. *Chin Arch Tradit Chin Med.* 2024; **42**(10):1—5.
20. *Maiula TG, Yarmoliuk SM, Konvaliuk II, Kunakh VA.* In silico prediction of neuroprotective properties of natural compounds using Scutellaria baicalensis as an example. *Biopolym Cell.* 2025; **41**(2):139—49.

21. *Maiula TG, Yarmoliuk SM.* In silico prediction of neuroprotective properties of natural compounds using Scutellaria baicalensis as an example. Proc. IMBG Conf. Kyiv; 2025. p. 7.
22. *Lupak MI, Khokhla MR, Hachkova HY, et al., and Sybirna NO.* Application of biogenic surface-active substances for stabilization of a phytopreparation based on the alkaloid-free fraction of Galega officinalis L. extract. *Studia Biologica.* 2015; **9**(1):25—36.
23. *Sibirna NO, Lupak MI, Hachkova HY, Lutsyk MD.* The effect of Galega officinalis L. extract on the content of glycated hemoglobin and blood glucose in experimental diabetes. *Visnyk Lviv Univ, Ser Biol.* 2014; 66:174—82.
24. *Korulkin D, Muravlyova L, Abilov Z, Auhatayeva A.* Medicinal plants Galega officinalis L. and yacon leaves (Smallanthus sonchifolius): comparative analysis of composition and hypoglycemic activity. *Ukrainica Bioorganica Acta.* 2019; **17**(1):12—8.
25. *Hachkova HY, Lupak MI, Sybirna NO.* Galega officinalis extract regulates carbohydrate metabolism in experimental diabetes. *Biol Syst: Theory Innovation.* 2020; **11**(2):45—52.
26. *Sibirna NO, Hachkova HY, Khokhla MR.* Method for obtaining an alkaloid-free extract from Galega officinalis L. with antidiabetic activity. Patent of Ukraine UA 96839 U. Filed July 15, 2014; published February 25, 2015.
27. *Mysinger MM, Carchia M, Irwin JJ, Shoichet BK.* Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J Med Chem.* 2012; **55**(14):6582—94.
28. *Santos FA, Frota JT, Arruda BR, et al., and Rao VS.* Antihyperglycemic and hypolipidemic effects of α, β-amyrin, a triterpenoid mixture from Protium heptaphyllum in mice. *Lipids Health Dis.* 2012; **11**:98.
29. *Naik SR, Kokil GR.* Development and discovery avenues in bioactive natural products for glycemic novel therapeutics. In: Bioactive Natural Products: Chemistry and Biology. Elsevier; 2013. p. 283—302.
30. *Bednarska K, Kuś P, Fecka I.* Investigation of the Phytochemical Composition, Antioxidant Activity, and Methylglyoxal Trapping Effect of Galega officinalis L. Herb In Vitro. *Molecules.* 2020; **25**(24):5810.
31. *Shymanska OV, Vergun OM, Rakhmetov DB, Brindza J.* Antiradical activity of plant extracts of Galega officinalis L. and G. orientalis Lam. *Plant Introduction.* 2018; **78**(2):28—33.
32. Bioecological features and pharmaceutical use of Galega officinalis L. (literature review). *Am J Pediatr Med Health Sci.* 2024; **2**(2):292—6.

Received 20.10.2025

Т.Г. Маюла[1], О.А. Бєда[1, 2], Т.В. Пилаєва[1], С.М. Ярмолюк[1]
[1] Інститут молекулярної біології і генетики НАН України,
Україна, 03143, м. Київ, вул. Академіка Заболотного, 150
[2] ТОВ «Науково-сервісна фірма "Otava"»
Україна, 03056, м. Київ, вул. Борщагівська, 117/129
t.h.maiula@imbg.org.ua

IN SILICO МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ АНТИДІАБЕТИЧНОГО ПОТЕНЦІАЛУ БІОАКТИВНИХ СПОЛУК БЕЗАЛКАЛОЇДНОГО ЕКСТРАКТУ *GALEGA OFFICINALIS L.*

Мета. Розробити та верифікувати in silico моделі прогнозування антидіабетичної активності природних сполук та перевірити їх на прикладі компонентів лікарських рослин. Матеріали і методи. Побудовано дві моделі машинного навчання на основі алгоритмів XGBoost і LightGBM. Верифікацію проведено на наборі сполук, що входять до складу рецептур традиційної китайської медицини (ТКМ) з доведеною гіпоглікемічною дією. Верифіковані моделі застосовано до компонентів *Galega officinalis L.* (безалкалоїдна фракція), склад якої визначено за допомогою газової та рідинної мас-спектрометрії. ***Результати.*** Побудовані моделі досягли точності 80—81% та успішно пройшли верифікацію, виявивши активні сполуки серед таких, які зазначені у рецептурах ТКМ як ефективні у боротьбі з діабетом 2 типу. Це підтверджує їх здатність коректно класифікувати біоактивні природні речовини. Моделі застосовано до компонентів безалкалоїдного екстракту *G. officinalis L.*, перспективної рослини для подальших досліджень антидіабетичної дії. ***Висновки.*** Розроблені *in silico* моделі дозволяють прогнозувати антидіабетичну активність сполук природного походження. Їх успішна верифікація на наборах сполук та застосування до екстракту *G. officinalis L.* підтвердили потенціал підходу для пошуку малотоксичних біоактивних речовин.

Ключові слова: *in silico*, машинне навчання, антидіабетичні властивості, *G. officinalis L.*