



Використання методів машинного навчання в розробленні назальних лікарських форм церебропротективної дії

Б. С. Бурлака ^{id}*A-E, І. Ф. Беленічев ^{id}A,F

Запорізький державний медичний університет, Україна

A – концепція та дизайн дослідження; B – збір даних; C – аналіз та інтерпретація даних; D – написання статті; E – редагування статті; F – остаточне затвердження статті

Для ресурсозбереження активних фармацевтичних інгредієнтів і допоміжних речовин на ранніх етапах дослідження під час планування експерименту доцільно використати знання щодо прогнозованих та експериментальних фізико-хімічних властивостей, що перебувають у різних агрегаційних базах даних. Знайдена інформація дає можливість скоротити час на розроблення складу та опрацювання технології. Але різноманіття характеристик активних сполук і допоміжних речовин не завжди наведене в названих сервісах. Останнім часом моделі машинного навчання, що дають можливість отримувати прогнозування з високою ймовірністю, широко застосовують у різних наукових напрямках. Отже, актуальним і перспективним є опрацювання моделей машинного навчання для прогнозу наявності фармацевтичних несумісностей у рецептурі назальних лікарських форм.

Мета роботи – опрацювання моделей машинного навчання для *in silico* прогнозу раціонального складу назальних лікарських форм церебропротективної дії.

Матеріали та методи. Як матеріал використовували датасет, що містив дані щодо сполук (діючих і допоміжних) і ознаки щодо наявності або відсутності взаємодії (фармацевтичної несумісності). Наповнення датасету для навчання (training datasets) здійснювали шляхом контент-аналізу даних бібліотеки PubMed (pubmed.ncbi.nlm.nih.gov) у ручному режимі за ключовими словами («pharmaceutical incompatibilities», «physico-chemical compatibility», «incompatible excipients») за останні 10 років. Обсяг датасету, що одержали, – 1185 рядків. Використали набір методів бінарної класифікації машинного навчання (rusaret.org) із застосуванням мови програмування python 3.8 (python.org) у середовищі управління пакетами miniconda (conda.io). Програмування пайплайну (pipeline) здійснили за допомогою пакета jupyter notebook (jupyter.org). Генерацію ознак сполук MACCS (Molecular ACCess System keys) у навчальному датасеті виконали за допомогою пакета RDKit (rdkit.org). Специфікації спрощеного наведення молекул у рядку введення (SMILES) в автоматичному режимі шукали за допомогою сервісу PubChem (pubchem.ncbi.nlm.nih.gov).

Результати. У результаті дослідження обрали дві перспективні моделі машинного навчання бінарної класифікації, якість роботи яких перевіряли на датасеті для перевірки. Статистичне оцінювання обраних моделей свідчить про високу ймовірність *in silico* прогнозу щодо наявності або відсутності фармацевтичних несумісностей під час розроблення назальних рецептур церебропротективних лікарських форм і розміщення на вебсервері експертної системи ExpSys Nasalia (nasalia.zsmu.zp.ua) в розділі розрахунки.

Висновки. Опрацювали моделі машинного навчання для *in silico* прогнозу раціонального складу назальних лікарських форм церебропротективної дії. Підтвердження якості прогнозу фармацевтичних несумісностей із використанням опрацьованих моделей виконали на датасеті для перевірки. Отримали статистичні показники моделей tree_blender (AUC 0.9521, F1 0.9747, MCC 0.9094), boost_blender (AUC 0.9593, F1 0.9821, MCC 0.9352). Використання моделей машинного навчання у фармацевтичній розробці сприятиме ресурсозбереженню та оптимізації складу рецептури.

Ключові слова: машинне навчання, лікарські форми, церебропротектори.

Актуальні питання фармацевтичної і медичної науки та практики. 2021. Т. 14, № 2(36). С. 232–238

The use of machine learning methods in the development of nasal dosage forms with cerebroprotective action

B. S. Burlaka, I. F. Bielenichev

In order to save resource of active pharmaceutical ingredients and excipients, in the early stages of research, when planning an experiment, it is advisable to use data of the predicted and experimental physicochemical properties stored in different aggregation databases. The information found will reduce the time for composition development and for technology processing. However, the variety of active compounds characteristics and excipients is not always reflected in these services. Recently, machine learning models have been widely used in various scientific fields; they allow to obtain predictions with high reliability. Given the above, it is relevant and promising to develop models of machine learning to predict the presence of pharmaceutical incompatibilities in the formulation of nasal dosage forms.

ARTICLE INFO



<http://pharmed.zsmu.edu.ua/article/view/232053>

UDC 615.45.032.21:616.831].015:004.81
DOI: [10.14739/2409-2932.2021.2.232053](https://doi.org/10.14739/2409-2932.2021.2.232053)

Current issues in pharmacy and medicine: science and practice 2021; 14 (2), 232–238

Key words: machine learning, dosage forms, cerebroprotectors.

*E-mail: burlakabogdan@gmail.com

Received: 31.03.2021 // Revised: 16.04.2021 // Accepted: 21.04.2021

The aim of the study is to develop models of machine learning for *in silico* forecast of the rational composition of nasal dosage forms with cerebroprotective action.

Materials and methods. A dataset, containing data on compounds (active and auxiliary) and characteristics on the presence or absence of interaction (pharmaceutical incompatibility), was used as material. Training datasets were filled by content analysis of PubMed library data (pubmed.ncbi.nlm.nih.gov) manually, by keywords “pharmaceutical incompatibilities”, “physico-chemical compatibility”, “incompatible excipients”) for the last 10 years. The resulting dataset comprises 1185 lines. The methods employed were a set of methods for binary classification of machine learning (pycaret.org) using the programming language Python 3.8 (python.org) in the package management environment Miniconda (conda.io). Pipeline programming was performed using Jupyter notebook package (jupyter.org). The generation of MACCS (Molecular ACCess System keys) in the training dataset was performed using RDKit package (rdkit.org). Specifications of the simplified representation of molecules in the input line (SMILES), in automatic mode, were searched using PubChem service (pubchem.ncbi.nlm.nih.gov).

Results. The obtained data allowed to choose two perspective models of machine learning of binary classification, whose quality was checked on a dataset for verification. Statistical evaluations of the selected models indicate a high probability of *in silico* prognosis for the presence or absence of pharmaceutical incompatibilities in the development of nasal formulations of cerebroprotective dosage forms. They are posted on the web server of the expert system ExpSys Nasalia (nasalia.zsmu.zp.ua) in the calculations section.

Conclusions. As a result of our research, we have developed machine learning models for *in silico* prediction of the rational composition of nasal dosage forms with cerebroprotective action. Confirmation of the quality of the pharmaceutical incompatibilities prediction, using the developed models, is checked on a dataset for check. The statistical indicators of the tree_blender (AUC 0.9521, F1 0.9747, MCC 0.9094) and boost_blender (AUC 0.9593, F1 0.9821, MCC 0.9352) models were obtained. The use of machine learning models in pharmaceutical development will contribute to resource conservation and optimization of the composition of the formulation.

Key words: machine learning, dosage forms, cerebroprotectors.

Current issues in pharmacy and medicine: science and practice 2021; 14 (2), 232–238

Использование методов машинного обучения в разработке назальных лекарственных форм церебропротективного действия

Б. С. Бурлака, И. Ф. Беленичев

Для ресурсосбережения активных фармацевтических ингредиентов и вспомогательных веществ на ранних этапах исследования при планировании эксперимента целесообразно использовать данные по прогнозируемым и экспериментальным физико-химическим свойствам, которые находятся в разных агрегационных базах данных. Найденная информация позволит сократить время на разработку состава и обработки технологии. Однако многообразие характеристик активных соединений и вспомогательных веществ не всегда представлено в указанных сервисах. В последнее время модели машинного обучения широко применяют в различных научных направлениях и позволяют прогнозировать результат с высокой достоверностью. Таким образом, актуальна и перспективна разработка моделей машинного обучения для прогноза наличия фармацевтических несовместимостей в рецептуре назальных лекарственных форм.

Цель работы – разработка моделей машинного обучения для *in silico* прогноза рационального состава назальных лекарственных форм церебропротективного действия.

Материалы и методы. В качестве материала использовали датасет, содержащий данные соединений (действующих и вспомогательных) и признаки о наличии или отсутствии взаимодействия (фармацевтической несовместимости). Наполнение датасета для обучения (training datasets) осуществляли путем контент-анализа данных библиотеки PubMed (pubmed.ncbi.nlm.nih.gov) в ручном режиме по ключевым словам («pharmaceutical incompatibilities», «physico-chemical compatibility», «incompatible excipients») за последние 10 лет. Объём полученного датасета – 1185 строк. Использовали набор методов бинарной классификации машинного обучения (pycaret.org) с применением языка программирования python 3.8 (python.org) в среде управления пакетами miniconda (conda.io). Программирование пайплайна (pipeline) осуществляли с помощью пакета jupyter notebook (jupyter.org). Генерацию признаков соединений MACCS (Molecular ACCess System keys) в датасете для обучения проводили с помощью пакета RDKit (rdkit.org). Спецификации упрощенного представления молекул в строке ввода (SMILES) в автоматическом режиме искали с помощью сервиса PubChem (pubchem.ncbi.nlm.nih.gov).

Результаты. В результате исследования выбрали две перспективные модели машинного обучения бинарной классификации, качество работы которых проверяли на датасете для проверки. Статистические оценки избранных моделей свидетельствуют о высокой вероятности *in silico* прогноза о наличии или отсутствии фармацевтических несовместимостей при разработке назальных рецептур церебропротективных лекарственных форм и размещены на веб-сервере экспертной системы ExpSys Nasalia (nasalia.zsmu.zp.ua) в разделе расчёты.

Выводы. В результате исследований разработаны модели машинного обучения для *in silico* прогноза рационального состава назальных лекарственных форм церебропротективного действия. Подтверждение качества прогноза фармацевтических несовместимостей с использованием обработанных моделей проведено на датасете для проверки. Получены статистические показатели моделей tree_blender (AUC 0,9521, F1 0,9747, MCC 0,9094), boost_blender (AUC 0,9593, F1 0,9821, MCC 0,9352). Использование моделей машинного обучения в фармацевтической разработке будет способствовать ресурсосохранению и оптимизации состава рецептуры.

Ключевые слова: машинное обучение, лекарственные формы, церебропротекторы.

Актуальные вопросы фармацевтической и медицинской науки и практики. 2021. Т. 14, № 2(36). С. 232–238

Одна з проблем у фармацевтичному розробленні нових лікарських форм – обґрунтування складу рецептури для наступного опрацювання технології екстемпорального або промислового виготовлення. Після вибору активних речовин, що забезпечуватимуть очікуваний терапевтичний ефект, і виду лікарської форми потрібно теоретично та практично обґрунтувати вміст допоміжних речовин у лікарській формі. Серед вимог, котрі ставлять до допоміжних речовин, – безпечність, біосумісність, надання лікарській формі належних властивостей (консистенція, форма, фізико-хімічні властивості), індиферентність [1,2].

Нині є чимало активних фармацевтичних інгредієнтів, а також допоміжних речовин органічної та неорганічної природи, з різними фізико-хімічними властивостями. Під час розроблення нових лікарських форм у рецептурі виникають різноманітні поєднання діючих і допоміжних речовин, які можуть характеризуватися фармацевтичною несумісністю (*Incompatibilita pharmaceutica*). Це явище зумовлене взаємодією між інгредієнтами рецептури, що призводить до суттєвої зміни фізико-хімічних властивостей готової лікарської форми [3].

Для експериментального вивчення можливої взаємодії інгредієнтів рецептури використовують кілька підходів. Створюють бінарну суміш активного фармацевтичного інгредієнта та допоміжної речовини, яку тестують без додавання води і з нею. Як інструментальні методи використовують ізотермічне стрес-тестування (*isothermal stress testing – IST*), високоефективну рідинну хроматографію (*high performance liquid chromatography – HPLC*), термічні методи – диференційна сканувальна калориметрія (*differential scanning calorimetry – DSC*) і термогравіметрія (*thermal gravimetry – TG, DTA*) [4–6].

Для ресурсозбереження активних фармацевтичних інгредієнтів і допоміжних речовин на ранніх етапах дослідження доцільно під час планування експерименту використати знання щодо прогнозованих, експериментальних фізико-хімічних властивостей, що містяться в різних агрегаційних базах даних, як-от Pubchem (pubchem.ncbi.nlm.nih.gov) або Drugbank (drugbank.com). Знайдена інформація дасть змогу скоротити час на розроблення складу та опрацювання технології. Але різноманіття характеристик активних сполук і допоміжних речовин не завжди наведене в цих сервісах.

Останнім часом моделі машинного навчання широко застосовують у різних наукових напрямках, вони дають змогу отримувати прогнозування з високою ймовірністю [7–10]. Отже, актуальним і перспективним є опрацювання моделей машинного навчання для прогнозу наявності фармацевтичної несумісності в рецептурі назальних лікарських форм.

Мета роботи

Опрацювання моделей машинного навчання для *in silico* прогнозу раціонального складу назальних лікарських форм церебропротективної дії.

Матеріали і методи дослідження

Для опрацювання моделей машинного навчання потрібно використати певний обсяг інформації, що міститиме дані щодо сполук (діючих і допоміжних) та ознаки щодо наявності або відсутності взаємодії (фармацевтичної несумісності). Тому для наповнення датасету для навчання (*training datasets*) використовували метод контент-аналізу даних бібліотеки PubMed (pubmed.ncbi.nlm.nih.gov) у ручному режимі за ключовими словами («*pharmaceutical incompatibilities*», «*physicochemical compatibility*», «*incompatible excipients*») за останні 10 років. Аналізуючи посилання на фахову літературу, в датасет вносили композицію з двох сполук (традиційні або хімічні назви) та класифікаційні позначки: 1 або 0 (композиція: 1 – сумісна, 2 – несумісна). Обсяг датасету, що одержали, – 1185 рядків.

Використали набір методів бінарної класифікації машинного навчання (rusaret.org), застосовуючи мову програмування python 3.8 (python.org) у середовищі управління пакетами *miniconda* (conda.io). Програмування пайплайну (*pipeline*) здійснювали за допомогою пакета *jupyter notebook* (jupyter.org). Генерацію ознак сполук MACCS (*Molecular ACCESS System keys*) у навчальному датасеті виконали за допомогою пакета *RDKit* (rdkit.org). Специфікації спрощеного наведення молекул у рядку введення (*SMILES*) в автоматичному режимі шукали за допомогою сервісу *PubChem* (pubchem.ncbi.nlm.nih.gov).

Результати

Процес підготовки (*pipeline*) даних для дослідження програмували в такій послідовності: завантаження в робочу пам'ять дослідного датасету з назвами сполук і класифікаційними ознаками (1 або 0) (*рис. 1*), трансформація датасету для отримання *SMILES* сполук (*рис. 2*), генерація ознак молекул (*feature engineering*) шляхом послідовного оброблення *SMILES* (*рис. 3*), нормалізація датасету – перевірка на наявність пропущених даних та їх заповнення.

Обговорення

Після підготовки даних отримали датасет, що містив MACCS ключі для дослідних пар сполук (167 ознак на кожну сполуку), а також класифікаційну позначку 1 або 0. Для машинного навчання дослідний датасет випадковим алгоритмом розподіляли на датасет для навчання (769 сполук) і датасет для перевірки (192 сполуки). Потім виконували попередній аналіз моделей, що одержали, без оптимізації на датасеті для навчання, та обирали 3 перспективні моделі класифікації за критерієм AUC (*Area Under the Curve*). Враховуючи, що дослідний датасет містить бінарні групи 2×167 ознак, усередині групи може виникати колінеарність (явище, коли одна змінна в наборі лінійно корелює з іншою характеристикою), додатково налаштовували класифікатор для мінімізації цього явища.

```

from typing import Union, List, Dict
from tqdm.notebook import trange, tqdm
import requests
import json
import numpy as np
import pandas as pd
from rdkit.Chem import AllChem, Descriptors, MolFromSmiles, Crippen, rdMolDescriptors, rdchem, QED, rdMolOps, MACCSkeys
import pycaret.classification as pycaret_class
from pycaret.utils import check_metric

dataset_in = pd.read_csv(filepath_or_buffer='incopartible_exipients.csv', sep=',')
dataset_in.set_index('num', inplace=True)

```

Рис. 1. Імпортування необхідних бібліотек і дослідного датасету в робочу пам'ять.

```

get_feature_from_name(compound:str)->Union[Dict[str,str]]:
cid=''
smiles=''
title=''
error=''
url = 'https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/{}/property/Title,CanonicalSMILES/json'.format(compound)
try:
    response = requests.get(url)
    response.raise_for_status()
    raw=json.loads(response.text)
    try:
        cid=raw['PropertyTable']['Properties'][0]['CID']
    except Exception as e:
        cid=''
        error = e
    try:
        smiles = raw['PropertyTable']['Properties'][0]['CanonicalSMILES']
    except Exception as e:
        smiles = ''
        error = e
    try:
        title = raw['PropertyTable']['Properties'][0]['Title']
    except Exception as e:
        title = ''
        error = e
    except Exception as e:
        error = 'Pubchem error - {}'.format(e)
    return {'cid':cid, 'smiles':smiles, 'title':title, 'error':''}

```

Рис. 2. Функція отримання SMILES із традиційної або хімічної назви сполуки.

```

def get_fingerprint_maccs(smiles:str, prefix:str, num:int)->pd.DataFrame:
    if len(smiles)>0:
        try:
            mol_maccskeys = [int(x) for x in MACCSkeys.GenMACCSKeys(MolFromSmiles(smiles)).ToBitString()]
            res_df = pd.DataFrame([mol_maccskeys,mol_maccskeys]).iloc[[0]]
            old_column=[el for el in range(0,167)]
            new_column=['{}_{}'.format(prefix, el) for el in old_column]
            res_df.rename(columns=dict(zip(old_column, new_column)), inplace=True)
            res_df['num']=num
            res_df.set_index('num', inplace=True)
            return res_df
        except Exception as e:
            columns_new=['{}_{}'.format(prefix, el) for el in range(0,167)]
            res_df = pd.DataFrame(columns=columns_new)
            res_df['num']=1
            res_df.set_index('num', inplace=True)
            return res_df
    else:
        columns_new=['{}_{}'.format(prefix, el) for el in range(0,167)]
        res_df = pd.DataFrame(columns=columns_new)
        res_df['num']=1
        res_df.set_index('num', inplace=True)
        return res_df

```

Рис. 3. Функція генерації ключів MACCS для дослідних сполук.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
et	Extra Trees Classifier	0.9071	0.9684	0.9409	0.9275	0.9333	0.7794	0.7834	0.267
catboost	CatBoost Classifier	0.9201	0.9677	0.965	0.9252	0.9438	0.8054	0.8128	4.737
xgboost	Extreme Gradient Boosting	0.909	0.9659	0.9462	0.9261	0.9349	0.783	0.7895	0.206
rf	Random Forest Classifier	0.9034	0.9646	0.9382	0.9247	0.9303	0.772	0.7775	0.297
lightgbm	Light Gradient Boosting Machine	0.9107	0.9631	0.9408	0.9326	0.9358	0.7883	0.7922	0.101
gbc	Gradient Boosting Classifier	0.9052	0.9591	0.9462	0.9204	0.9324	0.7737	0.7785	0.112
knn	K Neighbors Classifier	0.8736	0.9265	0.9461	0.8814	0.912	0.6881	0.6967	0.088
ada	Ada Boost Classifier	0.8587	0.9246	0.9245	0.8824	0.9004	0.6538	0.6671	0.064
lr	Logistic Regression	0.8587	0.9223	0.9111	0.8894	0.8987	0.6636	0.671	0.347
qda	Quadratic Discriminant Analysis	0.8793	0.8976	0.8488	0.9734	0.9054	0.7406	0.757	0.019
lda	Linear Discriminant Analysis	0.859	0.8615	0.9378	0.8697	0.9019	0.6521	0.6606	0.021
nb	Naive Bayes	0.5537	0.8464	0.3883	0.9203	0.5408	0.2315	0.3175	0.012
dt	Decision Tree Classifier	0.8698	0.8433	0.9138	0.9007	0.9066	0.6916	0.6948	0.012
svm	SVM - Linear Kernel	0.8475	0	0.9085	0.8787	0.8919	0.6319	0.6392	0.014
ridge	Ridge Classifier	0.872	0	0.9432	0.881	0.91	0.6874	0.6976	0.012

Рис. 4. Попередній аналіз моделей машинного навчання бінарної класифікації.

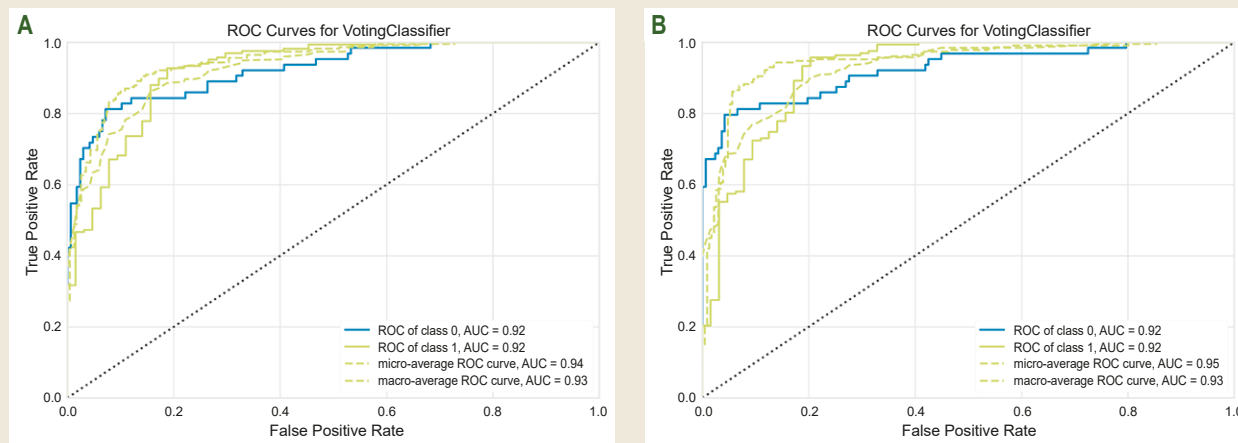


Рис. 5. Криві навчання моделей машинного навчання blending tree, blending boosting. **A:** змішана модель blending tree (tuned_class_rf, tuned_class_et); **B:** змішана модель blending boosting (tuned_class_catboost, tuned_class_lightgbm, tuned_class_xgboost).

За результатами, що наведені на рис. 4, найбільш перспективними щодо обраного критерію AUC для наступного дослідження є моделі: класифікатор додаткових дерев (Extra Trees Classifier, ET) AUC 0,9684, категоріальний бустинг (CatBoost Classifier, CatBoost) AUC 0,9677, екстремальний градієнтний бустинг (Extreme Gradient Boosting, EGB) AUC 0,9659, випадковий ліс (Random Forest Classifier, RF) AUC 0,9646, Light Gradient Boosting Machine (LIGHTGBM) AUC 0,9631. Надалі на етапі з обраними моделями використали різні техніки машин-

ного навчання Ensembling (Stacking, Boosting, Blending) для детальнішої оптимізації. У підсумку для кожного виду моделі отримали ще додатково по три моделі (табл. 1).

Аналізували моделі, які отримали, за комплексом статистичних показників: Accuracy, AUC, F1, MCC (табл. 1). Виявили, що техніки оптимізації машинного навчання покращили статистичні показники досліджуваних моделей. Надалі групували моделі за їхньою природою (tree, boosting) і використовували техніку змішування моделей

Таблиця 1. Статистичні показники оптимізованих моделей машинного навчання (mean \pm sd)

Модель	Аccuracy	AUC	F1	MCC
RF (tuned)	0,9051 \pm 0,0271	0,9657 \pm 0,0157	0,9304 \pm 0,0195	0,7878 \pm 0,0666
RF (bagged)	0,9015 \pm 0,0264	0,9608 \pm 0,0203	0,9305 \pm 0,018	0,7689 \pm 0,0633
RF (calibrated)	0,9034 \pm 0,0351	0,9666 \pm 0,0151	0,9302 \pm 0,0257	0,7777 \pm 0,0811
CatBoost (tuned)	0,9108 \pm 0,0359	0,9655 \pm 0,0227	0,9368 \pm 0,0251	0,7921 \pm 0,0867
CatBoost (bagged)	0,9015 \pm 0,03	0,9643 \pm 0,0214	0,9306 \pm 0,0205	0,7684 \pm 0,0731
CatBoost (calibrated)	0,9145 \pm 0,0313	0,967 \pm 0,0185	0,9395 \pm 0,0214	0,8004 \pm 0,0764
LIGHTGBM (tuned)	0,9108 \pm 0,0331	0,9681 \pm 0,0183	0,9362 \pm 0,0232	0,7916 \pm 0,0804
LIGHTGBM (bagged)	0,9071 \pm 0,0342	0,9645 \pm 0,0211	0,9335 \pm 0,0241	0,7849 \pm 0,0833
LIGHTGBM (calibrated)	0,9145 \pm 0,0278	0,9681 \pm 0,0165	0,9391 \pm 0,0194	0,8018 \pm 0,0664
EGB (tuned)	0,8959 \pm 0,0292	0,9699 \pm 0,0164	0,9289 \pm 0,0186	0,7545 \pm 0,0705
EGB (bagged)	0,9071 \pm 0,0389	0,964 \pm 0,0198	0,9336 \pm 0,0279	0,7843 \pm 0,0916
EGB (calibrated)	0,9126 \pm 0,03	0,9671 \pm 0,0182	0,9381 \pm 0,0207	0,796 \pm 0,0726
ET (tuned)	0,894 \pm 0,0313	0,9645 \pm 0,02	0,9215 \pm 0,0232	0,7645 \pm 0,0737
ET (bagged)	0,9034 \pm 0,0341	0,9636 \pm 0,0191	0,9307 \pm 0,0246	0,7762 \pm 0,0799
ET (calibrated)	0,9052 \pm 0,0419	0,9653 \pm 0,0179	0,9316 \pm 0,0307	0,7828 \pm 0,097

```

unseen_predictions_boost = pycaret_class.predict_model(finalized_blender_specific_boost_tuned_models,
    data=data_unseen)
blender_boost_models_auc = check_metric(unseen_predictions_boost['Compartmentible'],
    unseen_predictions_boost['Label'],
    metric='AUC')
blender_boost_models_f1 = check_metric(unseen_predictions_boost['Compartmentible'],
    unseen_predictions_boost['Label'],
    metric='F1')
blender_boost_models_mcc = check_metric(unseen_predictions_boost['Compartmentible'],
    unseen_predictions_boost['Label'],
    metric='MCC')
unseen_predictions_tree = pycaret_class.predict_model(finalized_blender_specific_tuned_tree_custom_models,
    data=data_unseen)
blender_tree_models_auc = check_metric(unseen_predictions_tree['Compartmentible'],
    unseen_predictions_tree['Label'],
    metric='AUC')
blender_tree_models_f1 = check_metric(unseen_predictions_tree['Compartmentible'],
    unseen_predictions_tree['Label'],
    metric='F1')
blender_tree_models_mcc = check_metric(unseen_predictions_tree['Compartmentible'],
    unseen_predictions_tree['Label'],
    metric='MCC')

print("Датасет для перевірки(модель tree_blender): AUC {}, F1 {}, MCC {}".format(blender_tree_models_auc,
    blender_tree_models_f1,
    blender_tree_models_mcc))
print("Датасет для перевірки(модель boost_blender): AUC {}, F1 {}, MCC {}".format(blender_boost_models_auc,
    blender_boost_models_f1,
    blender_boost_models_mcc))

```

Датасет для перевірки(модель tree_blender): AUC 0.9521, F1 0.9747, MCC 0.9094
Датасет для перевірки(модель boost_blender): AUC 0.9593, F1 0.9821, MCC 0.9352

Рис. 6. *In silico* прогноз на датасеті для перевірки.

(blending models), що дає змогу поєднувати різні алгоритми машинного навчання з використанням більшості голосів або середніх прогнозованих імовірностей для прогнозування кінцевого результату.

Після змішування одержали дві перспективні моделі машинного навчання бінарної класифікації (рис. 5), якість роботи яких перевіряли на датасеті для перевірки (192 сполуки). Датасет для перевірки не використову-

вали в навчанні цих моделей. Отримали статистичні оцінки моделей (рис. 6): tree_blender – AUC 0,9521, F1 0,9747, MCC 0,9094, boost_blender – AUC 0,9593, F1 0,9821, MCC 0,9352. Вони свідчать про високу ймовірність *in silico* прогнозу щодо наявності або відсутності фармацевтичних несумісностей під час розроблення назальних рецептур церебропротективних лікарських форм.

Опрацьовані моделі сприяють оптимізації рецептури в розробленні нових назальних засобів церебропротективної дії та розміщенні на вебсервері експертної системи ExpSys Nasalia (nasalia.zsmu.zp.ua) в розділі розрахунки.

Висновки

1. У результаті дослідження опрацювали моделі машинного навчання для *in silico* прогнозу раціонального складу назальних лікарських форм церебропротективної дії.

2. Підтвердження якості прогнозу фармацевтичних несумісностей із використанням опрацьованих моделей виконали на датасеті для перевірки. Отримали статистичні показники моделей *tree_blender* (AUC 0.9521, F1 0.9747, MCC 0.9094), *boost_blender* (AUC 0.9593, F1 0.9821, MCC 0.9352), що свідчать про високу ймовірність прогнозу.

3. Використання моделей машинного навчання в фармацевтичній розробці сприятиме ресурсозбереженню та оптимізації складу рецептури.

Перспективи подальших досліджень. Для оптимізації якості *in silico* прогнозу фармацевтичних несумісностей у рецептурі назальних лікарських форм церебропротективної дії доцільно наповнювати датасет для навчання новими експериментальними даними та здійснювати оптимізацію моделей.

Конфлікт інтересів: відсутній.

Conflicts of interest: authors have no conflict of interest to declare.

Відомості про авторів:

Бурлака Б. С., канд. фарм. наук, доцент каф. технології ліків, Запорізький державний медичний університет, Україна.

ORCID ID: [0000-0003-4539-7331](https://orcid.org/0000-0003-4539-7331)

Беленічев І. Ф., д-р біол. наук, професор, зав. каф. фармакології та медичної рецептури з курсом нормальної фізіології, Запорізький державний медичний університет, Україна.

ORCID ID: [0000-0003-1273-5314](https://orcid.org/0000-0003-1273-5314)

Information about authors:

Burlaka B. S., PhD, Associate Professor of the Department of Medicinal Preparations Technology, Zaporizhzhia State Medical University, Ukraine. Bielenichev I. F., Dr.hab., Professor, Head of the Department of Pharmacology and Medical Formulation with Course of Normal Physiology, Zaporizhzhia State Medical University, Ukraine.

Сведения об авторах:

Бурлака Б. С., канд. фарм. наук, доцент каф. технологии лекарств, Запорожский государственный медицинский университет, Украина. Беленічев І. Ф., д-р біол. наук, професор, зав. каф. фармакології та медичинської рецептури з курсом нормальної фізіології, Запорожський державний державний медичний університет, Україна.

Список літератури

- [1] Abrantes C. G., Duarte D., Reis C. P. An Overview of Pharmaceutical Excipients: Safe or Not Safe?. *Journal of pharmaceutical sciences*. 2016. Vol. 105, Iss. 7. P. 2019-2026. <https://doi.org/10.1016/j.xphs.2016.03.019>
- [2] Excipient Stability in Oral Solid Dosage Forms: A Review / M. A. Darji, R. M. Lalge, S. P. Marathe et al. *AAPS PharmSciTech*. 2018. Vol. 19, Iss. 1. P. 12-26. <https://doi.org/10.1208/s12249-017-0864-4>
- [3] Narang A. S., Desai D., Badawy S. Impact of excipient interactions on solid dosage form stability. *Pharmaceutical research*. 2012. Vol. 29, Iss. 10. P. 2660-2683. <https://doi.org/10.1007/s11095-012-0782-9>
- [4] Bharate S. S., Bharate S. B., Bajaj A. N. Interactions and Incompatibilities of Pharmaceutical Excipients with Active Pharmaceutical Ingredients:

A Comprehensive Review. *Journal of Excipients and Food Chemicals*. 2010. Vol. 1, Iss. 3. P. 3-26.

- [5] Бурлака Б. С., Беленічев І. Ф., Гладішев В. В. Термогравіметричні дослідження інтраназальної форми ноопепту. *Фармацевтичний журнал*. 2019. Т. 74, № 6. С. 54-61.
- [6] Дериватографічне вивчення мазі з піроктон оламіном для терапії та профілактики себореїного дерматиту / В. А. Солодовник, В. В. Гладішев, Б. С. Бурлака, І. О. Пухальська. *Актуальні питання фармацевтичної і медичної науки та практики*. 2020. Т. 13, № 2. С. 249-253. <https://doi.org/10.14739/2409-2932.2020.2.207184>
- [7] Бурлака Б. С. Використання сучасного програмного забезпечення в систематизації літературних даних по інтраназальним лікарським засобам. *Фармацевтичний часопис*. 2015. № 1. С. 29-31.
- [8] Experimental and theoretical approaches to the creation of computer program for virtual screening of scavengers no in a range of azageterocycles / V. P. Ryzhenko, I. F. Belenichev, O. A. Ryzhov, S. V. Levich. *Медична інформатика та інженерія*. 2018. № 3. P. 54-57. <https://doi.org/10.11603/mie.1996-1960.2018.3.9474>
- [9] Deo R. C. Machine Learning in Medicine. *Circulation*. 2015. Vol. 132, Iss. 20. P. 1920-1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>
- [10] Yang K. K., Wu Z., Arnold F. H. Machine-learning-guided directed evolution for protein engineering. *Nature methods*. 2019. Vol. 16, Iss. 8. P. 687-694. <https://doi.org/10.1038/s41592-019-0496-6>

References

- [1] Abrantes, C. G., Duarte, D., & Reis, C. P. (2016). An Overview of Pharmaceutical Excipients: Safe or Not Safe?. *Journal of pharmaceutical sciences*, 105(7), 2019-2026. <https://doi.org/10.1016/j.xphs.2016.03.019>
- [2] Darji, M. A., Lalge, R. M., Marathe, S. P., Mulay, T. D., Fatima, T., Alshammari, A., Lee, H. K., Repka, M. A., & Narasimha Murthy, S. (2018). Excipient Stability in Oral Solid Dosage Forms: A Review. *AAPS PharmSciTech*, 19(1), 12-26. <https://doi.org/10.1208/s12249-017-0864-4>
- [3] Narang, A. S., Desai, D., & Badawy, S. (2012). Impact of excipient interactions on solid dosage form stability. *Pharmaceutical research*, 29(10), 2660-2683. <https://doi.org/10.1007/s11095-012-0782-9>
- [4] Bharate, S. S., Bharate, S. B., & Bajaj, A. N. (2010). Interactions and Incompatibilities of Pharmaceutical Excipients with Active Pharmaceutical Ingredients: A Comprehensive Review. *Journal of Excipients and Food Chemicals*, 1(3), 3-26.
- [5] Burlaka, B. S., Belenichev, I. F., & Gladyshev, V. V. (2019). Termogravimetrični doslidzhennia intranazalnoi formy noopeptu [Thermogravimetric investigation of a new intranasal gel with noopept]. *Farmatsevtichnyi zhurnal*, 74(6), 54-61. [in Ukrainian].
- [6] Solodovnyk, V. A., Hladyshev, V. V., Burlaka, B. S., & Pukhalska, I. O. (2020). Deryvatohrafichne vyvchennia mazi z pirokton olaminom dlia terapii ta profilaktyky seboreinoho dermatytu [Derivatographic study of the ointment with piroctone olamine for therapy and prevention of seborrheic dermatitis]. *Current issues in pharmacy and medicine: science and practice*, 13(2), 249-253. [in Ukrainian]. <https://doi.org/10.14739/2409-2932.2020.2.207184>
- [7] Burlaka, B. S. (2015). Vykorystannia suchasnoho programnoho zabezpechennia v systematyzatsii literaturnykh danykh po intranazalnym likarskym zasobam [Use of modern software for systematization of the literature data for intranasal drugs]. *Farmatsevtichnyi chasopys*, (1), 29-31. [in Ukrainian].
- [8] Ryzhenko, V. P., Belenichev, I. F., Ryzhov, O. A., & Levich, S. V. (2018). Experimental and theoretical approaches to the creation of computer program for virtual screening of scavengers no in a range of azageterocycles. *Medical Informatics and Engineering*, (3), 54-57. <https://doi.org/10.11603/mie.1996-1960.2018.3.9474>
- [9] Deo R. C. (2015). Machine Learning in Medicine. *Circulation*, 132(20), 1920-1930. <https://doi.org/10.1161/CIRCULATIONAHA.115.001593>
- [10] Yang, K. K., Wu, Z., & Arnold, F. H. (2019). Machine-learning-guided directed evolution for protein engineering. *Nature methods*, 16(8), 687-694. <https://doi.org/10.1038/s41592-019-0496-6>