

КІБЕРНЕТИКА та КОМП'ЮТЕРНІ ТЕХНОЛОГІЇ

УДК [004.8+528.8+502/504+614.841.2](477) DOI:10.34229/2707-451X.24.3.10

І.Д. ДЕНКОВ, Є.В. НАЗАРЕНКО

ПРОГНОЗУВАННЯ НЕБЕЗПЕКИ ЛІСОВИХ ПОЖЕЖ НА ОСНОВІ ВІДКРИТИХ СУПУТНИКОВИХ ДАНИХ GOOGLE EARTH ENGINE

Вступ. Лісові пожежі мають як позитивний (регулятивний), так і негативний (руйнівний) вплив. Протягом останніх років внаслідок глобального потепління частка руйнівного впливу стає все більшою. Лісові пожежі тривають довше в регіонах, де вони традиційно відбуваються під час пожежонебезпечного сезону, і виникають у місцях, де їх раніше не було. Одним зі способів розв'язання проблеми є визначення ділянок з високим ризиком займання, тобто прогнозування пожеж до їх виявлення для подальшого вжиття необхідних заходів. Традиційні варіанти розв'язання цієї задачі, наземне і авіаційне патрулювання, є трудомісткими, а останнє ще й фінансово затратним. В Україні ця проблема підсилюється внаслідок інтенсивного мінування та можливих обстрілів прикордонних територій, коли традиційний моніторинг лісниками стає небезпечним.

Це пояснює зростаючу роль спостереження за допомогою супутникових систем. Космічний моніторинг має низку переваг: вищу оперативність, більшу площу охоплення земної поверхні тощо. З урахуванням низької вартості використання відкритих супутникових даних, подібні методи матимуть потенціал до застосування і за межами України – у важкодоступних регіонах з високим ризиком руйнівних пожеж і там, де військові конфлікти чи терористична загроза роблять лісові масиви небезпечними для відвідування.

1. Пов'язані роботи

Існують різні підходи до даної проблематики. Наприклад, чинники, що призводять до пожеж, шукали у параметрах сонячної активності [1]. Для регіонів, де в один і той же час протягом літнього сезону виникали пожежі (Греція, Португалія і Каліфорнія), було обрховано кореляцію супутникових даних сонячної активності з наданими метеостанціями параметрами, безпосередньо пов'язаними з пожежною небезпекою: температурою та вологістю повітря.

Розроблено математичну модель прогнозування пожежної небезпеки в лісах за даними супутникових спостережень. Модель було навчено на прикладі лісів України, а роботу перевірено на прикладі лісів України та Польщі.

Ключові слова: пожежна небезпека в лісових масивах, супутникові дані, кореляція, РСА, UMAP, важкодоступні території.

© І.Д. Денков, Є.В. Назаренко, 2024

Є вже приклади власне прогнозування пожеж на основі кліматичних та біофізичних параметрів з застосуванням згорткових нейронних мереж [2] та методів опорних векторів, випадкового лісу, k найближчих сусідів [3]. В першому випадку використовувались суто супутникові дані, навчання проводилось з використанням пожежних масок у Сицилії та Каліфорнії. Модель продемонструвала кращі результати для лісів, ніж для чагарників та орних земель. В другому випадку було використано суто дані геоінформаційних систем для природного парку в Португалії. Вищезазначені методи порівнювали між собою відносно здатності прогнозувати малі (площею менше 50 га) та великі (більше 50 га) пожежі.

Інші дослідники порівнювали між собою багатопланові перцептрони і згорткові нейромережі [4]. Стосовно вхідних даних застосували комбінований підхід: було використано як супутникові дані (надані Глобальною системою асиміляції наземних даних NASA), так і наземні (від Геологічної служби США). На основі супутникової інформації Глобальної бази даних по пожежних викидах побудували бінарні карти пожеж, у подальшому використані в навчанні. Автори продовжували дослідження в цьому напрямку. Зосередившись на використанні згорткових нейромереж, вони зайнялись прогнозуванням потенціальних площ територій, що вигоріли для різних періодів до кінця XXI століття [5]. Згідно результатам, у середньо-довгостроковій та довгостроковій перспективах площа пожеж може зростати орієнтовно в 3–4 рази, а самі пожежі – зміщуватись у більш високі широти.

2. Методи та інструменти

Метою дослідження була побудова математичної моделі визначення ризику виникнення пожежі на основі кліматичних та біофізичних супутникових даних для лісів України і лісів подібної кліматичної зони з подальшим масштабуванням на інші кліматичні зони та інші типи рослинного покриву (степи, чагарники, орні землі тощо). При цьому врахували параметри, кожний з яких сам по собі може не впливати суттєво на пожежну безпеку, але які в сукупності є вагомим її фактором (тип лісового масиву, погодні умови і їх зміна, склад атмосфери, стан рослинності тощо).

Для вирішення задачі дослідження працювали з Google Earth Engine (GEE) – хмарною платформою, що містить колекції попередньо оброблених результатів зондування Землі багатьма штучними супутниками. Ці колекції є динамічними, тобто автоматично поповнюються в ході поточних космічних спостережень. Платформа дозволяє користувачам виконувати аналіз і обробку надвеликих обсягів геопросторових даних. Робота з уже готовими наборами даних спрощує дослідження, оскільки підготовка необроблених супутникових даних є нетривіальним процесом.

Окрім великого обсягу і широкої різноманітності інформації перевагою GEE є відкритий доступ до неї. Таким чином, GEE є потужним інструментом відкритої науки, який сприяє раціональнішому використанню фінансових та інших ресурсів, спрямованих на наукові дослідження. Однією зі складових відкритої науки є можливість незалежної перевірки отриманих результатів експериментів чи спостережень та повторного їх використання [6]. GEE належить до платформ, що надають ресурси для цього, сприяючи прискоренню досліджень, популяризації цікавих результатів і паралельно полегшуючи захист від псевдонауки і плагіату.

Навчальні дані збирались виключно з лісової зони України – Полісся, Карпат та прилеглих областей (на рис. 1 виділені прямокутниками) за 2017–2020 роки. Вікно, яке також можна побачити на рис. 1, послідовно обходило ці регіони, збираючи інформацію за кожний місяць вищезазначених років (окрім зимових). Дані, які стосувалися не лісових ділянок, ігнорувалися. За допомогою набору даних FireCCI51: MODIS Fire_cci Burned Area Pixel Product, Version 5.1 було отримано дані про місця пожеж (координати та дата). Отриману інформацію було оброблено за допомогою методу морфологічної фільтрації, аби залишилися тільки джерела займання. Щоб утворити навчальну вибірку, кожній пожежній ділянці була поставлена у відповідність ділянка, де не було пожежі. Для цього брали ті самі координати, але рівно на рік раніше (якщо і там виявляли пожежу, то ще на рік раніше і т.д.). Врешті-решт було отримано 3278 точок. Пожежні було марковано як 1, решту як –1.

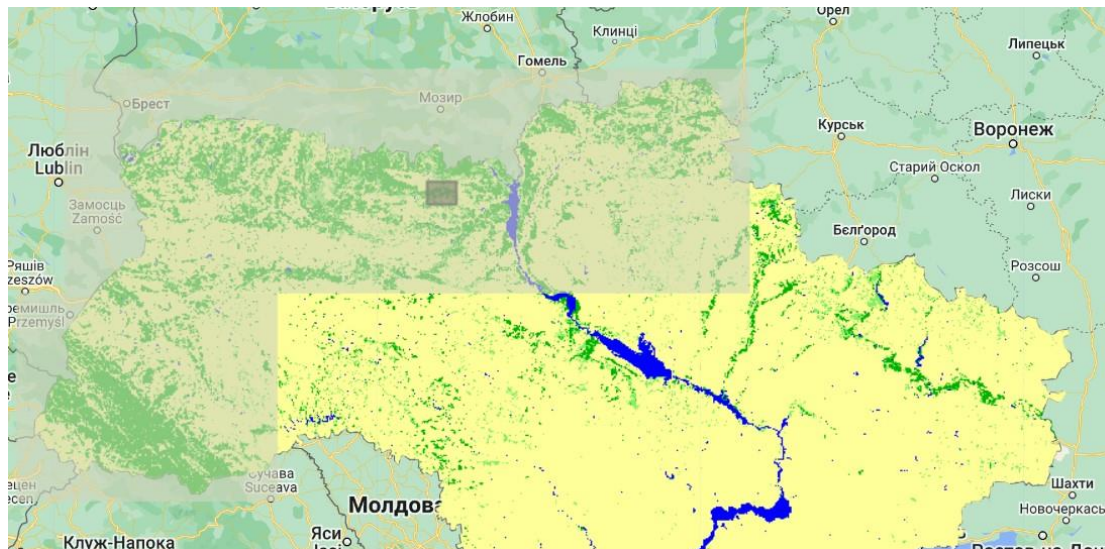


РИС. 1. Регіони, по яких збиралась інформація про пожежі

Кліматичні та біофізичні параметри, які вважалися спроможними вплинути на пожежну небезпеку, збирали з наступних наборів даних:

- MCD18C2.061 Photosynthetically Active Radiation Daily 3-Hour;
- MERRA-2 M2T1NXSLV: Single-Level Diagnostics V5.12.4;
- MOD11A1.061 Terra Land Surface Temperature and Emissivity Daily Global 1km;
- GCOM-C/SGLI L3 Land Surface Temperature (V2);
- GSMaP Operational: Global Satellite Mapping of Precipitation – V6;
- Sentinel-3 OLCI EFR: Ocean and Land Color Instrument Earth Observation Full Resolution;
- MOD17A2H.006: Terra Gross Primary Productivity 8-Day Global 500m;
- PML_V2 0.1.7: Coupled Evapotranspiration and Gross Primary Product (GPP).

Відкинувши результати, де було недостатньо даних, отримали наступні показники:

- загальний обсяг фотосинтетично активної радіації в даний момент часу;
- кількість непарного кисню в атмосферному стовпі;
- обсяг атмосферних опадів;
- обсяг поглинання хлорофілу;
- обсяг поглинання кисню;
- валова первинна продукція;
- дані про випаровування вологи (з ґрунту, з поверхні рослин, перехолення випаровування рослинним пологом).

Показники знімали раз на добу протягом тижня, що передувало пожежі, чи протягом тижня до вихідної дати у не пожежних точках. У разі більшої кількості вимірів за добу бралася середня (наприклад, для кількості кисню у стовпі або для його поглинання) або сумарна (у випадку опадів) величина. Після цього обчислювали їх перші та другі похідні. За похідні бралися різниці між послідовними показниками. У випадку перших похідних, наприклад, показник за другий день мінус показник за перший день (перша різниця), показник за третій мінус показник за другий (друга різниця) тощо. У випадку других похідних – друга різниця мінус перша різниця і так далі. Таким чином, за тиждень виходило 7 вихідних показників, 6 перших похідних, 5 других, тобто для кожного параметру – по 18 атрибутів для 3278 точок. Далі перевірили кореляцію кожного з отриманих атрибутів з фактом пожежі, залишивши врешті-решт тільки ті, де кореляція за модулем була не менша за 0,2.

Для параметрів, що вимірювалися лише раз на тиждень, на кореляцію перевіряли лише один атрибут (власне, вихідне вимірювання).

Врешті-решт залишилось 15 атрибутів. Таким чином, вибірка являла собою 3278 векторів у 15-вимірному просторі. За своєю природою отримані дані були взаємно залежними, отже потребували ортогоналізації. Для повноти експерименту ортогоналізацію виконали двома способами. В першому за основу взяли метод головних компонент PCA, перетворили вихідний базис на ортогональний і привели до нього вектори вибірки, зробивши таким чином атрибути статистично незалежними. В другому варіанті вже обрали нелінійний метод, UMAP, але підхід був такий самий: перехід до нового базису та приведення векторів до нього. При цьому розмірність простору не зменшували: менш значущі атрибути (в PCA, наприклад, це ті, яким відповідають менші власні значення ортогонального розкладу) не відкидали.

В обох випадках отримані атрибути було піддано статистичній обробці з використанням байєсівського підходу. Значення кожного з атрибутів було розбито на 8 рівномірних інтервалів (за допомогою медіан), для кожного з яких було обраховано умовну ймовірність пожежі.

Далі для кожної точки обраховували добутки відношень умовних ймовірностей пожеж до умовних ймовірностей не пожеж. Там, де отриманий результат був більше за 1, вважали, що модель прогнозує пожежу, де менше – відсутність пожежної небезпеки.

Таким чином, отримана модель складається з:

- перетворення координат векторів (у PCA та UMAP – у кожного своє);
- параметрів дискретизації (меж інтервалів для кожного з 15 атрибутів);
- сукупності усіх умовних ймовірностей.

3. Обговорення результатів

Результати роботи моделі для вихідної вибірки подано у вигляді матриць невідповідностей (рис. 2) і ROC-кривих (рис. 3).

PCA	Пожежа	Не пожежа		UMAP	Пожежа	Не пожежа
Пожежа	94%	23%		Пожежа	82%	24%
Не пожежа	6%	77%		Не пожежа	18%	76%

РИС. 2. Матриці невідповідностей для моделі прогнозування пожежної небезпеки у лісах України

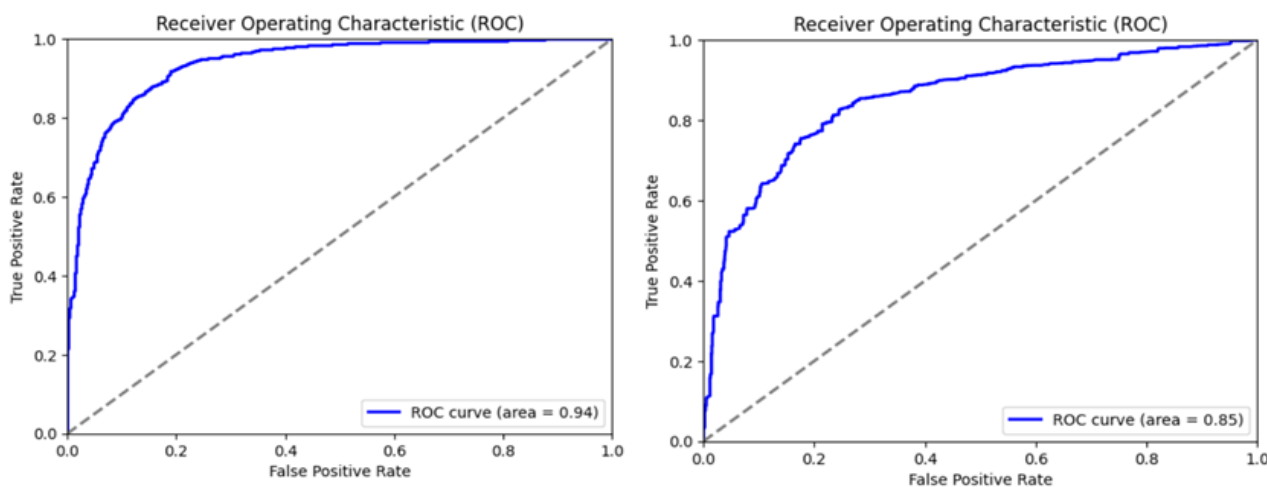


РИС. 3. ROC-криві для моделі прогнозування в Україні (PCA ліворуч, UMAP праворуч)

Таким чином, навіть без застосування машинного навчання модель вже демонструє вагомий результат.

Далі модель було перевірено на тестових даних. Тестову вибірку обирали, виходячи з того, що це має бути той самий період часу (2017–2020), але інший лісовий регіон тієї самої кліматичної зони. Таким чином, вибір зупинили на лісах Польщі. Дані по пожежах і не пожежах, а також інформацію про кліматичні та біофізичні параметри було зібрано вищеповисаним способом. На їх основі одразу розраховали 15 атрибутів, які було відібрано раніше. Після цього отримані вектори у 15-вимірному просторі (у випадку Польщі їх отримали 314) пропустили через нашу модель.

Модель продемонструвала значно гірші результати на тестовій вибірці. Зокрема модель з використанням PCA показала дуже погані результати щодо прогнозування на не пожежних ділянках (менше 50 %), тому для Польщі наведемо матрицю відповідності (рис. 4) та ROC-криву (рис. 5) тільки для UMAP.

UMAP	Пожежа	Не пожежа
Пожежа	69%	45%
Не пожежа	31%	55%

РИС. 4. Матриця невідповідностей для моделі прогнозування пожежної небезпеки в лісах Польщі

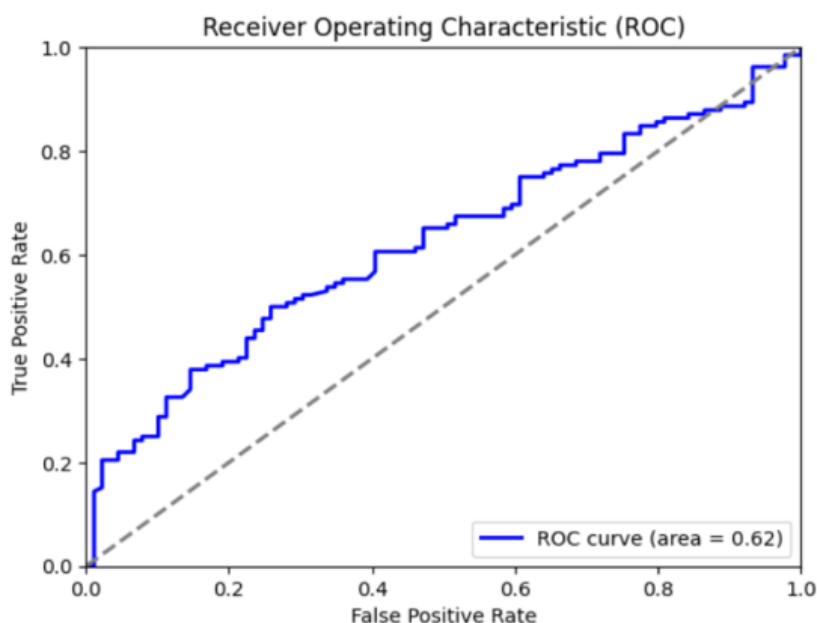


РИС. 5. ROC-крива для моделі прогнозування у випадку Польщі

Додатково для результатів роботи алгоритму UMAP зробили візуалізацію (рис. 6). В обох випадках видно, що множини не є роздільними. Це в значній мірі пояснює складність роботи з даними і, відповідно, складність отримання коректних результатів.

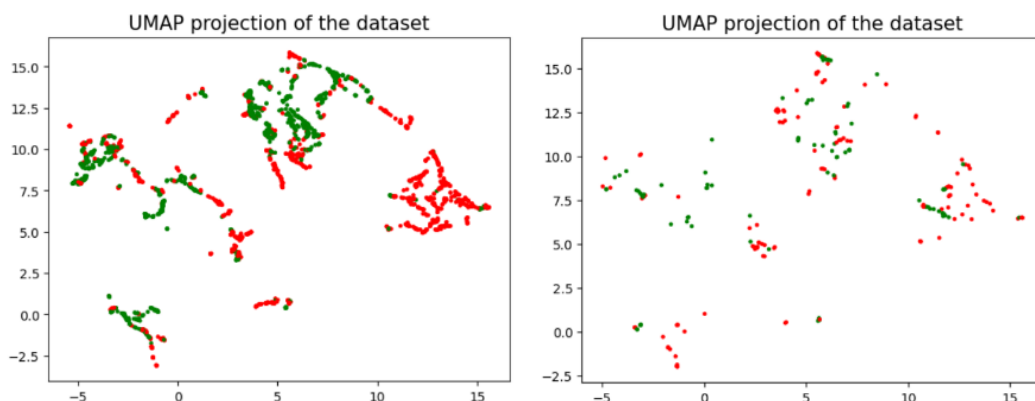


РИС. 6. Візуалізація роботи алгоритму UMAP у 2-вимірному просторі для України (ліворуч) і Польщі (праворуч), червоним позначено пожежі, зеленим – не пожежі

Також побудували кусково-лінійні діаграми розподілу для атрибутів, що показали найкращу кореляцію з показником «пожежа/не пожежа» (+1/-1). У випадку України найкращими кореляціями були 0,44, 0,18 і 0,17, а для Польщі – 0,27 і 0,23. Результати – відповідно на рис. 7 і рис. 8. З графіків видно, що поодинокі атрибути не демонструють достатньої кореляції і не можуть бути застосовані для прогнозування, але в сукупності, як було зазначено вище, спроможні бути інструментом для прогнозування з певною точністю.

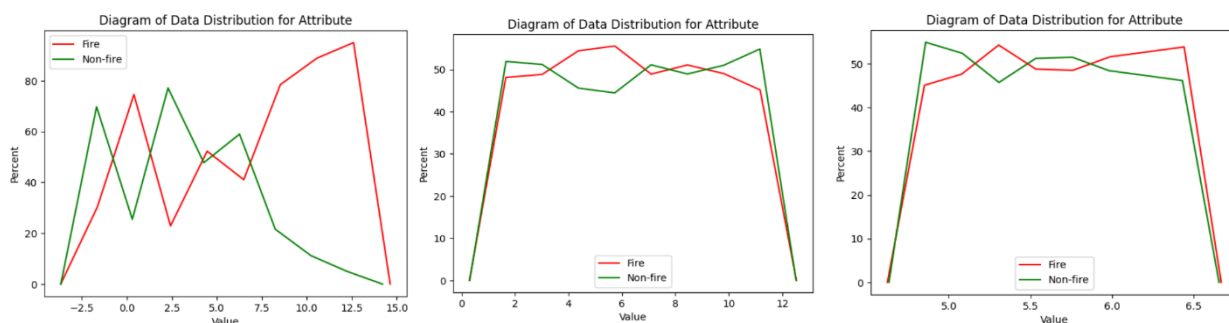


РИС. 7. Діаграми кореляції між окремими параметрами та пожежною небезпекою для України

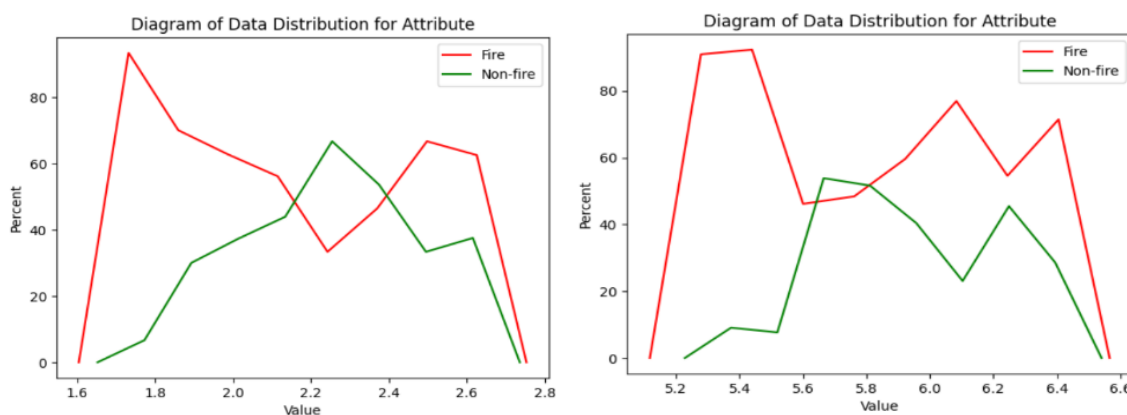


РИС. 8. Діаграми кореляції між окремими параметрами та пожежною небезпекою для Польщі

Погіршення результатів у випадку Польщі може бути зумовлене як недосконалістю моделі (зокрема її перенавчанням), так і причинами, з нею не пов'язаними. До останніх належить малий обсяг тестової вибірки і фактори, не пов'язані безпосередньо з природними процесами: пильніше відстеження лісниками потенційної пожежної небезпеки у господарствах і вжиття заходів заздалегідь, більш суворе законодавство стосовно розведення багатьох тощо. Стосовно можливих недоліків самої моделі слід зазначити, що потрібна подальша перевірка її в інших умовах (наприклад, той самий діапазон дат, але інший регіон або той самий регіон, але в інші роки). Подальші покращення моделі можливі шляхом модифікації алгоритму пошуку точок пожежі, введенням додаткових кліматичних чи біофізичних параметрів чи навпаки відкиданням менш значимих атрибутів після застосування PCA або UMAP, а також використанням алгоритмів машинного навчання.

Висновки. Запропоновано підхід до прогнозування лісових пожеж за супутниковими даними. Отримані результати вказують на те, що модель потребує подальшого вдосконалення, але є дієвою і потенційно може бути поширеною на інші регіони. Таким чином, роботу над моделлю буде продовжено, планується покращення якості прогнозів, розширення географії і створення web-додатку.

Авторські внески. Денков І.Д.: дослідження, методологія, концептуалізація, формальний аналіз, узагальнення, ресурси, програмне забезпечення, написання – оригінальна чернетка. Назаренко Є.В.: дослідження, методологія, ресурси, написання – рецензування та редагування.

Фінансування. Автори не отримували фінансування для проведення досліджень та написання статті.

Наявність даних. Дані, які підтверджують висновки цього дослідження, доступні за посиланням https://github.com/Ivan-Denkov/My_Coding/tree/main/FirePrediction_Ukraine.

Список літератури

1. Vyklyuk Y., Radovanović M.M., Stanojević G., etc. Connection of Solar Activities and Forest Fires in 2018: Events in the USA (California), Portugal and Greece. *Sustainability*. 2020. **12** (24). 10261. <https://doi.org/10.3390/su122410261>
2. Santopaolo A., Saif S.S., Pietrabissa A., Giuseppi A. Forest Fire Risk Prediction from Satellite Data with Convolutional Neural Networks. *29th Mediterranean Conference on Control and Automation (MED)*. 22–25 June 2021. <https://doi.org/10.1109/MED51440.2021.9480226>
3. Mohammed Z., Hanae S., Setti L. Comparative study on machine learning algorithms for early fire forest detection system using geodata. *International Journal of Electrical and Computer Engineering (IJECE)*. 2020. **10** (5). P. 5507–5513. <https://doi.org/10.11591/ijece.v10i5.pp5507-5513>
4. Zhang G., Wang M., Liu K. Deep neural networks for global wildfire susceptibility modeling. *Ecological Indicators*. 2021. **127**. 107735. <https://doi.org/10.1016/j.ecolind.2021.107735>
5. Zhang G., Wang M., Yang B., Liu K. Current and Future Patterns of Global Wildfire Based on Deep Neural Networks. *Earth's Future*. 2024. **12** (2). e2023EF004088. <https://doi.org/10.1029/2023EF004088>
6. Тульчинський В.Г., Лавренюк С.І., Коломієць О.В. та ін. Розвиток інтероперабельної інфраструктури геопорталу для обміну даними у двох напрямках. Етап 3. *Розробити прототип підсистеми хмарних геоінформаційних обчислень для геопорталу: технічний звіт*. Київ: Інститут кібернетики ім. В.М. Глушкова НАН України, 2023. 67 с. <https://nrat.ukrintei.ua/searchdoc/0223U005825/> (звернення: 02.07.2024).

Одержано 22.07.2024

Денков Іван Дем'янович,
аспірант Інституту кібернетики імені В.М. Глушкова НАН України, Київ,
<https://orcid.org/0009-0007-7989-2169>
havoc85@gmail.com

Назаренко Євген Володимирович,
молодший науковий співробітник відділу автоматизації програмування
Інституту кібернетики імені В.М. Глушкова НАН України, Київ.
<https://orcid.org/0009-0009-8906-373X>
eugn@ukr.net

УДК [004.8+528.8+502/504+614.841.2](477)

І.Д. Денков *, С.В. Назаренко

Прогнозування небезпеки лісових пожеж на основі відкритих супутникових даних Google Earth Engine

*Інститут кібернетики імені В.М. Глушкова НАН України, Київ** Листування: havoc85@gmail.com

Вступ. Лісові пожежі завдають суттєвої шкоди як природному фонду, так і народному господарству. Останніми роками їх шкідливий вплив виріс внаслідок глобального потепління, а в Україні також і внаслідок збройного конфлікту. Таким чином традиційні методи патрулювання (наземне та авіаційне) є не тільки витратними, а й небезпечними через мінування і можливі обстріли прикордонних територій. Отже зростає роль спостереження за допомогою супутникових систем. Космічний моніторинг є оперативнішим, охоплює більшу площу земної поверхні. Ще одна важлива його перевага – це відкритий доступ до інформації.

Мета роботи – побудова математичної моделі визначення пожежної небезпеки на основі кліматичних та біофізичних супутникових даних для лісів України, а також подібної кліматичної зони з можливістю подальшого масштабування на інші кліматичні регіони і типи рослинного покриву. Для дотримання принципів відкритої науки обрано Google Earth Engine (GEE), хмарну платформу, що надає відкритий доступ до динамічних колекцій попередньо оброблених результатів дистанційного зондування Землі.

Результати. За допомогою засобів бібліотеки Python для роботи з даними GEE зібрано кліматичні та біофізичні дані для лісів України за 2017–2020 роки. Далі отримані дані було оброблено двома методами: лінійним (PCA) і нелінійним (UMAP) з метою отримання статистично незалежних атрибутів. Обидві отримані вибірки піддали статистичній обробці байєсівським методом. Врешті-решт для кожної точки на карті, для якої збиралася інформація, обрахували показник, який прогнозував пожежну небезпеку, якщо отримане значення більше за 1 і її відсутність у протилежному випадку.

Отримана модель продемонструвала ефективну роботу на навчальних даних. На тестовій вибірці (дані по лісах Польщі за той самий період) результати виявилися гіршими, зокрема модель з застосуванням PCA не прогнозувала відсутність пожежної небезпеки, а модель з застосуванням UMAP в цілому продемонструвала нижчі показники. Це може бути зумовлене як недосконалістю моделі, так і малим обсягом тестової вибірки або факторами, не пов'язаними з природними процесами (зокрема людським).

Висновки. Запропоновано підхід до прогнозування лісових пожеж за супутниковими даними. Отримані результати вказують, що модель вже на цьому етапі є дієвою, хоча ще не було застосовано методи машинного навчання. Проте вона потребує подальшого вдосконалення, отже роботу над моделлю буде продовжено. Разом з покращенням якості прогнозів увагу буде приділено розширенню географії і створенню web-додатку.

Ключові слова: пожежна небезпека в лісових масивах, супутникові дані, кореляція, PCA, UMAP, важкодоступні території.

UDC [004.8+528.8+502/504+614.841.2](477)

Ivan Denkov *, Yevhen Nazarenko

Forest Fire Hazard Forecasting Based on Google Earth Engine Open Satellite Data

*V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv** Correspondence: havoc85@gmail.com

Introduction. Forest fires cause significant damage to both the natural fund and the national economy. In recent years, their harmful influence has increased due to global warming, and in Ukraine also due to the armed conflict. Thus, traditional methods of patrolling (ground and air) are not only costly, but also dangerous due to mines and possible shelling of border areas. Therefore, the role of surveillance using satellite systems is increasing. Space monitoring is more efficient and covers a larger area of the Earth's surface. Another important advantage is open access to information.

The purpose of the paper is to build a mathematical model for determining fire danger based on climatic and biophysical satellite data for the forests of Ukraine, as well as a similar climatic zone with the possibility of further scaling to other climatic regions and types of vegetation cover. To adhere to the principles of open science, Google Earth Engine (GEE), a cloud-based platform that provides open access to dynamic collections of pre-processed Earth remote sensing results, was chosen.

Results. Climatic and biophysical data for the forests of Ukraine for the years 2017-2020 were collected using the tools of the Python library for working with GEE data. Further, the obtained data were processed by two methods: linear (PCA) and non-linear (UMAP) in order to obtain statistically independent attributes. Both obtained datasets were subjected to statistical processing using the Bayesian method. Finally, for each point on the map for which information was collected, an indicator was calculated that predicted fire danger if the obtained value was greater than 1 and its absence in the opposite case.

The resulting model showed its efficiency on training data. On the test dataset (data on Polish forests for the same period), the results turned out to be worse, in particular, the model using PCA did not predict absence of fire danger, and the model using UMAP generally showed lower performance. This can be due to both the imperfection of the model and the small size of the test dataset or factors unrelated to natural processes (in particular the human factor).

Conclusions. An approach to forest fires forecasting based on satellite data is proposed. The obtained results indicate that the model is already effective at this stage, although machine learning methods have not yet been applied. However, it needs further improvement, so work on the model will be continued. Along with improving the quality of forecasts, attention will be paid to the geographic expansion and the creation of a web application.

Keywords: fire danger in forests, satellite data, correlation, PCA, UMAP, hard-to-reach areas.