Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет комп'ютерних наук
(повна назва)

Кафедра програмної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів прогнозування для передбачення стихійних явищ
(тема)

Виконав:
студент (ка) 2 курсу, групи ІПЗм-22-2

Корнієнко Д.М.
(прізвище, ініціали)

Спеціальність 121 – Інженерія програмного забезпечення
(код і повна назва спеціальності)
Тип програми освітньо-наукова

Керівник доц. Голян Н.В.
(посада, прізвище, ініціали)

Допускається до захисту
Зав. кафедри _____ З.В.Дудар
(підпис) (прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет комп’ютерних наук
Кафедра програмної інженерії
Рівень вищої освіти другий (магістерський)
Спеціальність 121 – Інженерія програмного забезпечення
Тип програми освітньо-наукова програма
Освітня програма Інженерія програмного забезпечення
(шифр і назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
«___» _____ 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Корнієнко Дмитро Михайлович
(прізвище, ім’я, по батькові)

1. Тема роботи «Дослідження методів прогнозування для передбачення стихійних явищ»

Затверджена наказом по університету від 29.03.2024р. № 250 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 17.06.2024

3. Вихідні дані до роботи опис досліджуваних методів прогнозування для передбачення стихійних явищ, вимоги до розробки методів прогнозування для передбачення стихійних явищ для проведення досліджень за обраною предметною областю, мови програмування Pyton, React, СУБД Mongo DB, середовища розробки Intellij Idea 2023, середовища розробки PyCharm 2024

4. Перелік питань, що потрібно опрацювати в роботі
аналіз та порівняння методів прогнозування для передбачення стихійних явищ, вибір підходящих методів прогнозування для дослідження, написання програмних рішень, проведення експериментів та аналіз отриманих результатів

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз предметної галузі та постановка задачі	23.01 – 14.02.24	*виконано*
2	Аналіз та вибір методів для дослідження	15.02 – 24.02.24	*виконано*
3	Аналіз та моделювання предметної області	17.02 – 28.02.24	*виконано*
4	Планування експериментів	25.02 – 28.02.24	*виконано*
5	Програмна реалізація кожного з обраних методів для дослідження	25.02 – 01.04.24	*виконано*
6	Експериментальні дослідження	02.04 – 20.04.24	*виконано*
7	Аналіз результатів експериментальних досліджень та розробка рекомендацій	20.04 – 23.04.24	*виконано*
8	Написання та оформлення статті та тез доповіді	17.04 – 23.04.24	*виконано*
9	Підготовка пояснювальної записки	01.04 – 03.06.24	*виконано*
10	Підготовка презентації та доповіді	3.06 – 07.06.24	*виконано*
11	Нормоконтроль	03.06	*виконано*
12	Рецензування	03.06	*виконано*
13	Занесення диплома в електронний архів	15.06	*виконано*
14	Попередній захист	15.06	*виконано*
15	Допуск до захисту у зав. кафедри	15.06	*виконано*

Дата видачі завдання 29 лютого 2024р

Студент (ка) _____ Корнієнко Д.М.
(підпис)

Керівник роботи _____ доц. Голян Н.В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ / ABSTRACT

Кваліфікаційна робота магістра містить: 64 сторінка, 4 рисунків, 6 таблиць, 10 джерела.

АРХІТЕКТУРА СПРИЙНЯТТЯ, МЕТОД ПРОГНОЗУВАННЯ, МОНІТОРИНГ КАТАСТРОФ, НЕЙРОНА МЕРЕЖА, ПРИРОДНІ ЯВИЩА, ПРОГНОЗУВАННЯ КАТАСТРОФ.

Об’єктом дослідження є методи передбачення стихійних явищ.

Метою роботи є дослідження методів передбачення стихійних явищ.

У роботі розглядаються методи передбачення стихійних явищ, способи їх реалізації та кінцеві результати для порівняльної характеристики.

У результаті був проведений аналіз предметної галузі, поставлені завдання дослідження, аналіз існуючих методів та алгоритмів, планування та проведення дослідження та виявлено найбільш влучний метод для прогнозування стихійних явищ.

DISASTER FORECASTING, DISASTER MONITORING, FORECASTING METHOD, NATURAL PHENOMENA, NEURON NETWORK, PERCEPTION ARCHITECTURE.

The object of research is methods of predicting natural phenomena.

The purpose of the work is to research the methods of predicting natural phenomena.

The work considers methods of predicting natural phenomena, methods of their implementation, and final results for comparative characterization.

As a result, we analyzed the subject area, set research objectives, analyzed existing methods and algorithms, planned and conducted the study, and identified the most appropriate method for predicting natural phenomena.

Я, Корнієнко Дмитро Михайлович, студента гр. ІПЗм-22-2, здобувач вищої освіти на другому(магістерському) рівні кафедри «Програмна інженерія», заявляю: моя кваліфікаційна робота на тему «Дослідження методів прогнозування для передбачення стихійних явищ», що буде представлена в екзаменаційну комісію для публічного захисту, виконана самостійно, в ній не містяться елементи плагіату і вона може бути опублікована в електронному архіві відкритого доступу ElArKhNURE. Всі запозичення з друкованих таелектронних джерел мають відповідні посилання.

Я ознайомлена з діючим положенням «Про протидію академічному плагіату в ХНУРЕ», згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту та застосування дисциплінарних заходів.

ЗМІСТ

Вступ 7
1 Аналіз предметної області 9
1.1 Аналіз предметної галузі 9
1.2 Аналіз факторів, що впливають на передбачення стихійних явищ 10
1.3 Постановка задачі 112
2 Аналіз існуючіх методів, рішень і алгоритмів 14
2.1 Опис існуючих алгоритмів 14
2.2 Опис існуючих аналогів 14
2.3 Параметричні методи прогнозування 176
2.4 Непараметричні методи 178
2.5 Комбіновані методи прогнозування 210
3 Планування експерементальної частини дослідження 24
3.1 Опис набору даних 24
3.2 Вибір методів дослідження 26
3.3 Метрики оцінювання 30
4 Проведення експерементальної частини дослідження 31
4.1 Аналіз та підготовка набору даних 31
4.2 Опис експерименту 343
4.4 Отримані результати 35
4.5 Аналіз результатів 38
4.6 Рекомендації щодо подальших досліджень 388
Висновки 42
Перелік джерел посилання 43
Перелік джерел посилання за науковими напрямами керівника та науковців кафедри програмної інженерії 44
Додаток А Звіт результатів перевірки на унікальність тексту в базі ХНУРЕ 45
Додаток Б Слайди презентації 46
Додаток В Апробація результатів роботи 54

Додаток Г Експертний висновок результатів перевірки кваліфікаційної роботи на відповідність оформлення вимогам ДСТУ 3008: 2015 64

ВСТУП

Стихійні лиха – це небезпечні природні явища геофізичного, геологічного та атмосферного походження, що характеризуються раптовими руйнуваннями і часто призводять до загибелі багатьох людей, руйнування житлової інфраструктури, навіть подальших катастроф. Вони можуть відбуватися в різний час і з різною інтенсивністю. Торнадо і зливові повені – це короткочасні руйнівні явища, що вражають відносно невеликі території. Інші небезпеки, такі як посухи, виникають повільно, але можуть впливати на цілі континенти або країни протягом місяців або навіть років. Екстремальні погодні явища можуть спричиняти кілька небезпек одночасно або послідовно. Наприклад, тропічні циклони зі сильними вітрами і дощами можуть призводити до повеней та зсувів. У помірних широтах сильні грози можуть супроводжуватися градом, торнадо, сильним вітром і дощем, а також зливовими повенями. Зимові шторми з сильним вітром і снігопадами можуть викликати лавини, що призводить до масового стоку води та повеней під час наступного сезону танення снігу.

Стихійні лиха можуть бути різного масштабу та ступеня впливу. Великі землетруси, такі як ті, що сталися в Японії у 2011 році чи в Індонезії у 2004 році, спричинили катастрофічні цунамі, що завдали значної шкоди та призвели до численних жертв. Виверження вулканів, як наприклад виверження вулкану Ейяф'ятлайокудль у 2010 році, можуть спричиняти тривалі перебої в повітряному русі через викиди вулканічного попелу. Урагани, такі як ураган Катріна у 2005 році, спричинили масові руйнування і затоплення великих територій.

Стихійні лиха можуть мати як прямий, так і непрямий вплив на суспільство. Прямі наслідки включають загибель людей, руйнування будівель, інфраструктури та сільськогосподарських угідь. Непрямі наслідки можуть включати економічні втрати, зниження врожайності, зростання цін на продукти харчування, а також погіршення якості життя населення. Крім того, стихійні лиха можуть спричиняти вторинні катастрофи, такі як епідемії хвороб, що поширюються через забруднену воду та погані санітарні умови.

Існують також техногенні загрози, такі як прориви дамб, хімічні аварії та радіаційні загрози, які часто неможливо передбачити заздалегідь. Наприклад, аварія на Чорнобильській АЕС у 1986 році стала однією з найбільших техногенних катастроф в історії людства, спричинивши значні людські та екологічні втрати. Хімічні аварії, такі як витік хімічних речовин на заводі Union Carbide в Бхопалі, Індія, у 1984 році, призвели до загибелі тисяч людей та довготривалих наслідків для здоров'я населення.

Метою цього дослідження є вивчення способів прогнозування природних загроз. Прогнозування стихійних лих є складним завданням через велику кількість змінних і їх взаємодій, але воно є критично важливим для зменшення наслідків катастроф. Сучасні методи прогнозування включають використання супутникових даних, метеорологічних моделей, історичних даних та штучного інтелекту. Результати цього дослідження будуть корисними для дослідників та науковців, які вивчають метеорологічні явища.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз предметної галузі

Стихійні лиха – це природні явища, що виникають внаслідок природних процесів і мають значний вплив на навколишнє середовище та життя людей. Вони можуть бути викликані широким спектром явищ, таких як урагани, землетруси, виверження вулканів, повені, грози та інші катастрофічні події. Ці явища є невід'ємною частиною природного середовища і можуть значно змінювати географічні території та екосистеми. Вони можуть мати серйозні наслідки для людей, включаючи людські жертви, руйнування інфраструктури та втрату майна. Крім того, природні катастрофи можуть спричиняти екологічні проблеми, такі як забруднення та втрата біорізноманіття. [2]

Попри свою руйнівну силу, природні явища також відіграють важливу роль у формуванні ландшафтів і створенні природних ресурсів, які є важливими для збереження екосистем і біорізноманіття. Вони сприяють природному перетворенню і оновленню середовища, створюючи нові екологічні ніші і можливості для різноманітних організмів.

Оскільки природні явища є непередбачуваними та можуть завдати значної шкоди, важливо розробляти ефективні методи прогнозування і захисту від їхніх наслідків. У цьому контексті значну роль відіграють сучасні технології, зокрема нейронні мережі. Нейронні мережі є потужним інструментом для аналізу і прогнозування природних явищ, оскільки вони здатні обробляти великі обсяги даних і виявляти складні взаємозв'язки між різними змінними. [4, 6]

Наприклад, згорткові нейронні мережі (CNN) ефективні для розпізнавання шаблонів і закономірностей у великих масивах даних, таких як зображення природних явищ. Це дозволяє використовувати їх для виявлення ознак наближення стихійних лих, таких як урагани або повені, за супутниковими знімками. Рекурентні нейронні мережі (RNN), у свою чергу, можуть бути використані для моделювання та прогнозування часових рядів, таких як зміни температури або рівня води під час стихійних лих. Це дозволяє створювати точні прогнози і своєчасно попереджати про можливі небезпеки.

Глибокі нейронні мережі (DNN) можуть бути застосовані для аналізу великих обсягів різноманітних даних, включаючи метеорологічні показники, сейсмічні дані, дані про зміни клімату та інші фактори, що впливають на виникнення природних катастроф. Завдяки цьому можна виявляти складні взаємозв'язки і залежності, які допомагають краще розуміти природу стихійних лих і мінімізувати їхній вплив на людей і навколишнє середовище.

Кожна нейронна мережа має свої сильні та слабкі сторони, і вибір конкретної моделі для аналізу природних явищ залежить від характеру даних, наявних ресурсів та конкретних цілей. Однак загалом, застосування нейронних мереж у вивченні та прогнозуванні природних явищ має великий потенціал і може значно покращити здатність суспільства протидіяти стихійним лихам, зменшуючи їхній вплив на людське життя та екосистеми.

1.2 Аналіз факторів, що впливають на передбачення стихійних явищ

Передбачення стихійних явищ є складним завданням, що вимагає врахування різноманітних факторів. Метеорологічні дані, такі як температура повітря і води, атмосферний тиск, опади та вітри, є ключовими для прогнозування погодних аномалій, включаючи урагани, повені та посухи. Геологічні фактори, такі як сейсмічна активність, рух тектонічних плит і вулканічна активність, допомагають передбачати землетруси та виверження вулканів. Гідрологічні показники, такі як рівень води в річках, сніговий покрив і вологість грунту, впливають на прогнозування повеней і посух. Сучасні технології, зокрема супутникові спостереження, комп'ютерні моделі та радарні системи, значно підвищили точність прогнозів стихійних явищ. Антропогенні фактори, такі як зміна клімату, вирубка лісів, урбанізація та інтенсивне використання водних ресурсів, також впливають на виникнення та інтенсивність цих подій. Зміна клімату збільшує частоту та інтенсивність багатьох стихійних явищ, а вирубка лісів і урбанізація змінюють природні ландшафти, що може посилити повені та зсуви. Інтенсивне використання водних ресурсів може призвести до посух, а забруднення та надмірне споживання води впливають на її доступність і якість. Інтеграція різних типів даних і

міждисциплінарний підхід є ключовими для успішного прогнозування та пом'якшення наслідків стихійних явищ. Сучасні технології та наукові дослідження значно покращили точність прогнозів, однак залишаються значні виклики, пов'язані з непередбачуваністю природи та постійними змінами довкілля.

1.3 Постановка задачі

Мета даної роботи полягає у всебічному дослідженні та аналізі методів передбачення стихійних явищ з використанням нейронних мереж. Особлива увага буде приділена методам, таким як згорткові нейронні мережі (CNN), рекурентні нейронні мережі (RNN) та глибокі нейронні мережі (DNN). Важливо визначити їхню ефективність та застосовність у передбаченні стихійних явищ на основі наявних даних.

Ключові завдання дослідження включають:

- ретроспективний аналіз ефективності різних архітектур нейронних мереж у передбаченні стихійних явищ на основі наявних даних. Це завдання передбачає детальний розгляд історичних даних та оцінку точності різних моделей нейронних мереж у передбаченні стихійних явищ. Буде досліджено, наскільки добре ці моделі справлялися з прогнозуванням різних типів стихійних лих у минулому;
- порівняння різних методів щодо їхньої здатності моделювати складні залежності в динаміці стихійних явищ. Це завдання спрямоване на виявлення того, які моделі нейронних мереж краще справляються з моделюванням складних, нелінійних залежностей, характерних для динаміки стихійних явищ. Будуть проаналізовані переваги і недоліки кожного підходу;
- розробка та навчання моделей для передбачення стихійних явищ на основі обраних методів. Цей етап включає створення та тренування нових моделей нейронних мереж, які можуть бути ефективними для передбачення стихійних явищ. Будуть використані сучасні алгоритми

машинного навчання та великі набори даних для підвищення точності прогнозів;

– аналіз отриманих результатів та формулювання висновків щодо ефективності та застосовності вибраних нейронних мереж для даного завдання. Після розробки та тренування моделей буде проведено детальний аналіз отриманих результатів. Це дозволить оцінити, наскільки ефективні обрані нейронні мережі у передбаченні стихійних явищ, а також виявити їхні сильні та слабкі сторони.

Ця робота спрямована на виявлення оптимального підходу до передбачення стихійних явищ та встановлення їхньої застосовності у контексті реальних умов та завдань прогнозування. Підсумковий аналіз результатів дослідження дозволить зрозуміти, які методи можуть бути найбільш корисними для передбачення та управління стихійними явищами.

Розробка точних методів передбачення стихійних явищ має велике значення для зниження ризиків і мінімізації шкоди, яку вони можуть завдати. Ефективні моделі прогнозування можуть допомогти урядам і організаціям швидше реагувати на потенційні загрози, що, у свою чергу, може врятувати життя, зберегти майно та захистити навколишнє середовище. Використання нейронних мереж у цьому контексті є перспективним напрямом, який може значно покращити точність і надійність прогнозів стихійних явищ.

Після завершення даного дослідження буде рекомендовано подальший розвиток і вдосконалення методів прогнозування стихійних явищ з використанням нейронних мереж. Це може включати розширення наборів даних, використання нових архітектур нейронних мереж, а також інтеграцію результатів у системи раннього попередження та управління надзвичайними ситуаціями.

Це дослідження може стати важливим внеском у науку про передбачення стихійних явищ та допоможе покращити підготовленість до природних катастроф у майбутньому.

2 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ, РІШЕНЬ І АЛГОРИТМІВ

2.1 Опис існуючих алгоритмів

Дослідження методів прогнозування стихійних явищ за допомогою нейронних мереж включає різні підходи і методи, що дозволяють аналізувати, моделювати та прогнозувати природні явища з використанням алгоритмів штучного інтелекту. Ось деякі з них:

- глибинне навчання. Використання багатошарових нейронних мереж для аналізу великих обсягів даних про стихійні явища. Глибинні нейронні мережі здатні виявляти складні залежності та патерни в даних, що дозволяє покращити точність прогнозів;
- рекурентні нейронні мережі (RNN). Застосування рекурентних нейронних мереж для обробки послідовних даних, таких як часова серія метеорологічних показників. RNN дозволяють враховувати залежності між попередніми та поточними значеннями, що є важливим для прогнозування динаміки стихійних явищ;
- конволюційні нейронні мережі (CNN). Використання CNN для аналізу просторових даних, таких як супутникові зображення та карти опадів. Конволюційні нейронні мережі ефективно виділяють важливі особливості в зображеннях, що дозволяє точніше визначати області ризику [5];
- автокодувальники. Застосування автокодувальників для виявлення аномалій у великих масивах даних. Цей підхід може бути корисним для виявлення нетипових змін в метеорологічних та геофізичних даних, які можуть передувати стихійним явищам;
- змішане навчання. Комбінування різних типів нейронних мереж, таких як CNN та RNN, для отримання більш точних прогнозів. Змішане навчання дозволяє одночасно аналізувати як просторові, так і часові залежності в даних [7, 8];
- розподілені обчислення: Використання розподілених обчислювальних систем для тренування великих нейронних мереж на масивних наборах

даних. Це дозволяє прискорити процес навчання і підвищити точність моделей;

– підкріплювальне навчання: Використання підкріплювального навчання для оптимізації моделей прогнозування стихійних явищ. Цей підхід дозволяє адаптувати моделі до змінних умов навколишнього середовища і покращувати їхню точність з часом.

Ці методи часто комбінуються з іншими підходами, такими як гідрометеорологічне моделювання та геоінформаційні технології, для досягнення найкращих результатів у прогнозуванні та управлінні ризиками, пов'язаними зі стихійними катастрофами. Використання нейронних мереж дозволяє значно підвищити точність і ефективність прогнозів, що є важливим для захисту людських життів та майна від природних загроз.

2.2 Опис існуючих аналогів

Перший аналог найцікавішої програми - Natural Disaster Monitor (див. рис. 2.1).

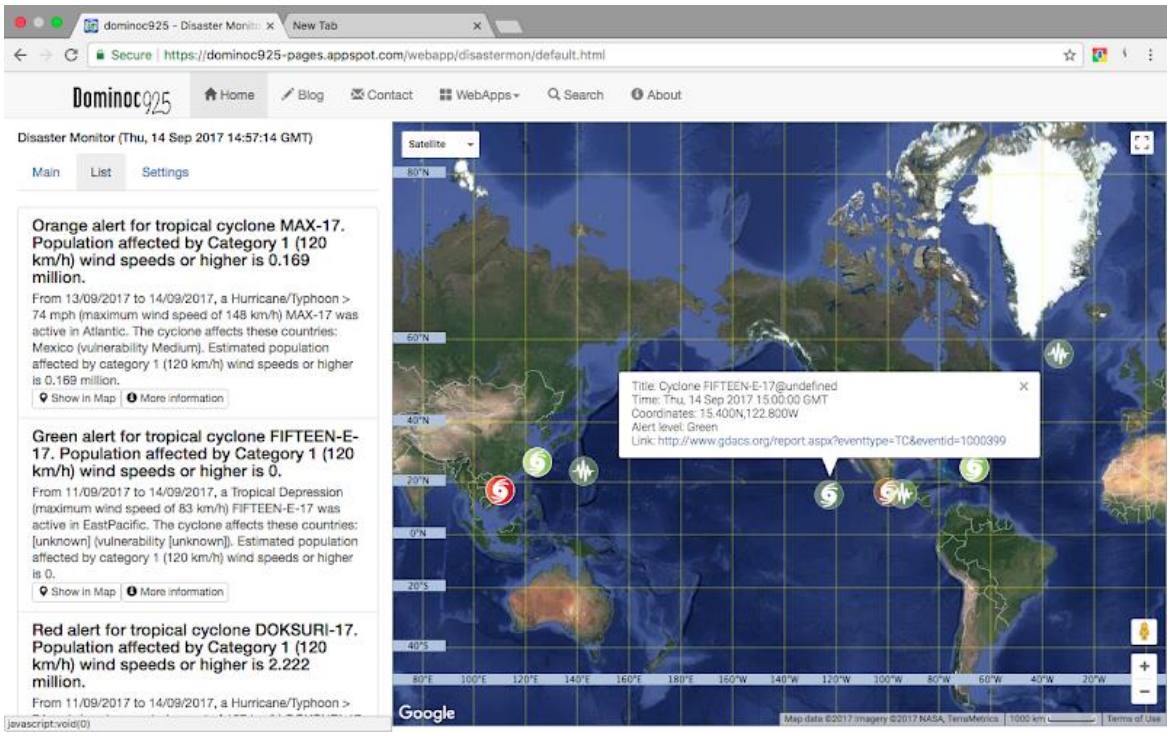

Рисунок 2.1 – Natural Disaster Monitor

Він показує користувачеві в реальному часі, де і коли сталися надзвичайні події, наприклад землетруси, тропічні циклони, повені, вулкани, посухи, лісові пожежі. Але програма не показує ймовірність, а тільки збирає дані з сайту і показує. Він просто бере дані з сайту https://www.gdacs.org, створеного Організацією Об'єднаних Націй та Європейською Комісією на основі даних міжнародних організацій GDACS з використанням інструментів і алгоритмів JRC для обміну інформацією в реальному часі. Цікава ідея використовувати цей сайт, який допоможе нам прогнозувати природні явища. Другий аналог - Disaster Alert (див. рис. 2.2).

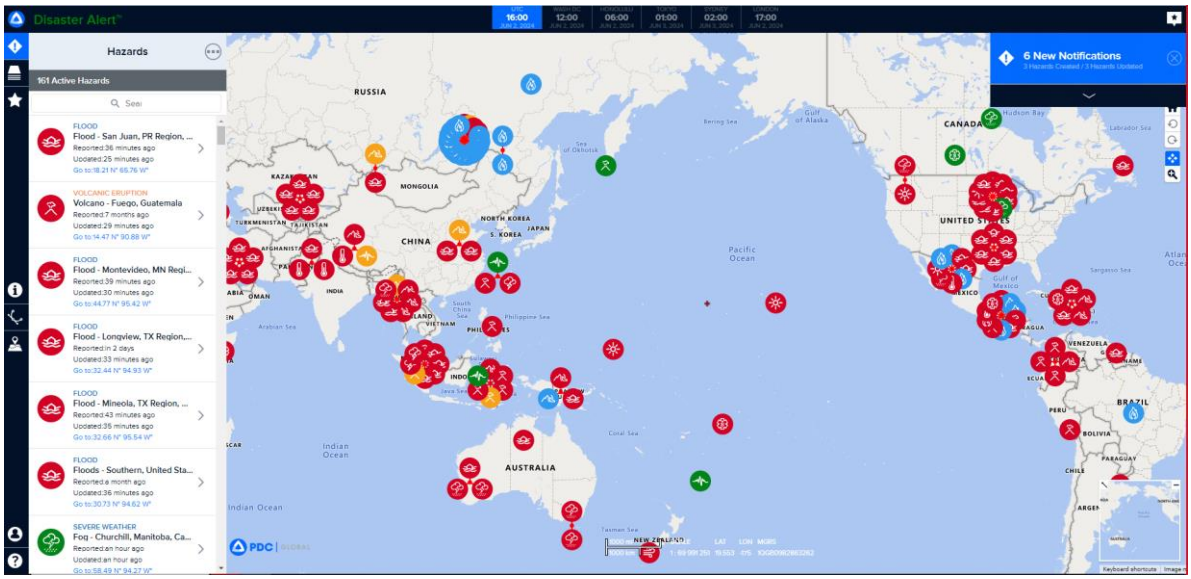

Рисунок 2.2 – Disaster Alert

Як попередній аналог, він також показує користувачеві природні явища. Але на відміну від попереднього аналога, він показує більше інформації про природні явища. Він заснований на PDC, який бере дані зі світу та надсилає користувачеві, коли і де відбуваються природні явища. Третій аналог Tropical Storm Risk (TSR) (див. рис. 2.3).

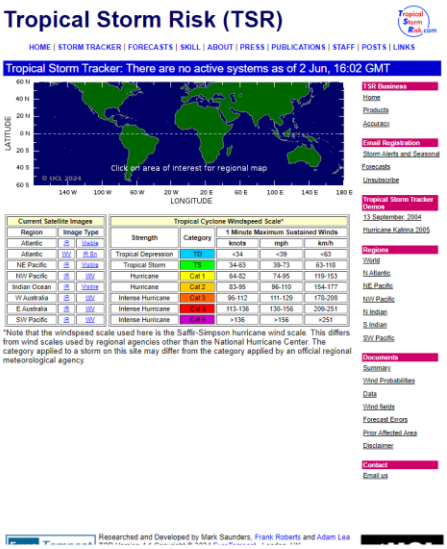

Рисунок 2.3 – Tropical Storm Risk (TSR)

TSR пропонує провідний ресурс для прогнозування та картографування активності тропічних штормів у всьому світі. Він надає прогнози та інформацію, щоб допомогти зрозуміти ризики та прийняти рішення щодо тропічних штормів. Це допомагає користувачеві побачити детальну карту та прогнозувати наслідки тропічних штормів у всьому світі. Чотири аналоги – Hurricane Hound. Він забезпечує відстеження ураганів на прокручуваній карті та категорії ураганів, 36-годинний прогноз, місцезнаходження, відстань від поточного місцезнаходження, максимальний вітер, пориви та час. Це допомагає користувачеві знати, куди йде ураган і коли він буде. Всі розглянуті аналоги мають свої шанси і мінуси. Але вони не передбачають природні явища, вони лише беруть інформацію з міжнародних незалежних відкритих джерел про природні явища та порівнюють місцезнаходження користувача з можливою катастрофою.

2.3 Параметричні методи прогнозування

Параметричні методи прогнозування використовуються для моделювання та аналізу даних шляхом припущення певної форми розподілу даних. Ці методи ґрунтуються на припущенні, що дані можуть бути описані за допомогою обмеженої кількості параметрів. Параметричні методи можна поділити на лінійні та нелінійні, кожен з яких має свої особливості та застосування.

Лінійні методи:

– метод часових рядів. Цей метод використовується для аналізу та прогнозування даних, які залежать від часу. Він передбачає моделювання залежностей між поточними та попередніми значеннями ряду. Основні техніки включають авторегресійні моделі (AR), моделі інтегрованого ковзного середнього (ARIMA) та сезонні ARIMA (SARIMA);
– алгоритм історичного середнього. Цей метод базується на використанні середніх значень попередніх даних для прогнозування майбутніх значень. Він є простим та ефективним, але може бути недостатньо точним при наявності значних коливань у даних;
– алгоритм згладжування. Цей метод використовує техніки згладжування, такі як експоненціальне згладжування, для зменшення впливу випадкових коливань на прогноз. Алгоритми згладжування можуть бути простими (просте експоненціальне згладжування) або складними (подвійне та потрійне експоненціальне згладжування);
– алгоритм фільтрації. Використовує фільтри, такі як Калманівський фільтр, для зменшення шуму в даних та покращення точності прогнозу. Фільтри допомагають відстежувати стан системи, роблячи прогнозування більш точним навіть при наявності шуму та невизначеностей.

Нелінійні методи:

– вейвлет аналіз. Цей метод використовується для аналізу та обробки сигналів, дозволяючи розбити їх на різні частотні компоненти. Вейвлети можуть бути ефективними для виявлення локальних особливостей та аномалій у даних;
– теорія катастроф. Моделює раптові зміни в поведінці системи, які можуть призвести до суттєвих змін або катастрофічних подій. Цей метод особливо корисний для аналізу систем, які можуть зазнати різких і непередбачуваних змін;
– теорія хаосу: Досліджує складні системи, в яких незначні зміни початкових умов можуть призвести до значних відмінностей у поведінці.

Цей метод використовується для моделювання динамічних систем, які демонструють хаотичну поведінку, що робить прогнозування особливо складним.

Переваги параметричних методів:

– ефективність: Параметричні методи можуть бути дуже ефективними при правильному визначенні моделі та параметрів;
– простота: Лінійні параметричні методи, такі як метод часових рядів та історичне середнє, є відносно простими для реалізації та інтерпретації;
– математична основа: Ці методи мають сильну математичну основу, що дозволяє формально аналізувати їх властивості та точність.

Недоліки параметричних методів:

– обмеженість: Параметричні методи можуть бути недостатньо гнучкими для моделювання складних і нелінійних залежностей у даних;
– припущення: Вони ґрунтуються на припущеннях про розподіл даних, які можуть не відповідати реальності;
– чутливість: Ці методи можуть бути чутливими до вибору початкових параметрів та форми моделі, що може впливати на точність прогнозу.

Параметричні методи залишаються важливим інструментом у прогнозуванні завдяки їхній простоті та ефективності. Однак, для складніших задач можуть бути необхідні комбіновані або непараметричні методи, які забезпечують більшу гнучкість і точність.

2.3 Непараметричні методи

Непараметричні методи прогнозування використовуються для моделювання даних без припущення конкретної форми розподілу. Вони забезпечують гнучкість у виявленні складних залежностей між змінними, що робить їх корисними для аналізу даних, які не підходять під параметричні моделі. Основні непараметричні методи включають машину опорних векторів, непараметричну регресію, нейронні мережі та нечітку логіку.

Основні непараметричні методи:

– машина опорних векторів використовується для класифікації та регресії. Цей метод знаходить оптимальні гіперплощини для розділення даних на різні класи або прогнозування значень. SVM особливо ефективна при роботі з високовимірними даними та здатна моделювати нелінійні залежності завдяки використанню функцій ядра, таких як поліноміальні або радіальні базисні функції (RBF);
– непараметрична регресія - цей метод включає техніки, такі як локальна регресія (LOESS) та регресія методом найближчих сусідів. Непараметрична регресія не припускає конкретної функціональної форми для залежності між змінними, що дозволяє моделювати складні та нелінійні зв'язки. Локальна регресія використовує вагові функції для побудови моделей у кожній точці даних, тоді як метод найближчих сусідів базується на середніх значеннях найближчих точок;
– нейронні мережі складаються з шарів нейронів, які імітують роботу людського мозку. Вони можуть виявляти складні нелінійні залежності у даних завдяки своїй багатошаровій структурі. Основні типи нейронних мереж включають багатошарові перцептрони (MLP), згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN). Нейронні мережі широко використовуються у задачах класифікації, регресії, розпізнавання образів та прогнозування часових рядів [1, 3];
– нечітка логіка використовує нечіткі множини та правила для моделювання невизначеностей у даних. Вона дозволяє створювати системи, що можуть оперувати нечіткими поняттями, такими як "високий", "низький" або "помірний". Нечітка логіка ефективна для моделювання складних систем, де традиційні методи не можуть дати точних результатів через високу ступінь невизначеності.

Переваги непараметричних методів:

– гнучкість. Непараметричні методи можуть моделювати складні та нелінійні залежності без потреби у припущеннях про форму розподілу даних;

– менше обмежень. Ці методи не вимагають попередніх знань про структуру даних, що робить їх більш універсальними;
– ефективність при великій кількості даних. Непараметричні методи можуть добре працювати з великими наборами даних, де параметричні методи можуть зазнавати труднощів.

Недоліки непараметричних методів:

– високі обчислювальні витрати. Непараметричні методи можуть вимагати значних обчислювальних ресурсів, особливо при роботі з великими наборами даних;
– перенавчання. Через свою гнучкість ці методи можуть бути схильні до перенавчання, особливо при недостатній кількості даних;
– складність інтерпретації. Результати непараметричних методів можуть бути складнішими для інтерпретації порівняно з параметричними методами.

Непараметричні методи є потужним інструментом для аналізу даних, що дозволяє виявляти складні залежності та моделювати невизначеності. Завдяки своїй гнучкості та універсальності, вони широко використовуються в різних галузях, включаючи машинне навчання, статистику та обробку сигналів. Незважаючи на певні обчислювальні витрати та складність інтерпретації, їхні переваги значно переважають недоліки, що робить їх цінним інструментом у сучасному аналізі даних. Непараметричні методи продовжують розвиватися, пропонуючи нові можливості для більш точного та ефективного моделювання складних систем у різних галузях знань.

2.4 Комбіновані методи прогнозування

Комбіновані методи прогнозування об'єднують кілька підходів для досягнення більшої точності та надійності в прогнозуванні. Ці методи використовуються для моделювання складних систем, де один метод може бути недостатньо ефективним або точним. Комбіновані методи поєднують переваги

різних підходів, зменшуючи їхні недоліки, що дозволяє досягти більш точних прогнозів.

Основні типи комбінованих методів:

– модель комбінації кластеризації - метод використовує техніки кластеризації для групування подібних даних перед застосуванням прогнозуючих моделей. Кластеризація дозволяє виділити групи даних з подібними характеристиками, що спрощує подальше прогнозування. Після кластеризації до кожного кластеру може бути застосована відповідна модель прогнозування, що враховує специфіку даних у цьому кластері;
– модель комбінації декомпозиції - метод розбиває дані на кілька компонентів, таких як тренд, сезонність і випадковість, та прогнозує кожен компонент окремо. Декомпозиція дозволяє зосередитися на окремих аспектах даних, що поліпшує точність прогнозів. Після прогнозування окремих компонентів результати об'єднуються для отримання загального прогнозу;
– технологія об’єднання результатів прогнозування - підхід поєднує результати різних прогнозуючих моделей для отримання остаточного прогнозу. Використовуються методи, такі як ансамблеве навчання, яке об'єднує декілька моделей для поліпшення точності та зменшення ризику перенавчання. Основні техніки включають методи середнього значення (bagging), бустинг (boosting) та стекінг (stacking).

Переваги комбінованих методів:

– підвищена точність. Об'єднання кількох методів дозволяє компенсувати недоліки окремих моделей та покращити загальну точність прогнозування;
– гнучкість. Комбіновані методи можуть бути адаптовані до різних типів даних та завдань, що робить їх універсальними;
– стійкість. Поєднання декількох моделей зменшує ризик перенавчання та підвищує стійкість до шуму в даних.

Недоліки комбінованих методів:

– складність. Комбіновані методи можуть бути складними для реалізації та налаштування, вимагаючи більше обчислювальних ресурсів та часу;
– інтерпретованість. Поєднання декількох моделей може ускладнити інтерпретацію результатів, оскільки важко зрозуміти, які аспекти даних впливають на остаточний прогноз;
– обчислювальні витрати. Використання кількох моделей одночасно може призвести до значних обчислювальних витрат, особливо при роботі з великими наборами даних.

Комбіновані методи прогнозування є потужним інструментом для досягнення високої точності та надійності в моделюванні складних систем. Використання таких методів дозволяє отримати більш точні прогнози, що є особливо важливим у контексті складних і динамічних систем, таких як стихійні явища. Попри складність реалізації та обчислювальні витрати, переваги комбінованих методів значно перевершують їхні недоліки, що робить їх цінним інструментом у сучасному прогнозуванні.

3 ПЛАНУВАННЯ ЕКСПЕРИМЕНТАЛЬНОЇ ЧАСТИНИ ДОСЛІДЖЕННЯ

3.1 Опис набору даних

З наявних даних можна зробити висновок, що набори даних з інформацією про стихійні явища використовуються для моніторингу та аналізу природних катастроф у різних регіонах. Ці набори даних зазвичай збираються за допомогою різних засобів, таких як супутники, метеорологічні станції, сейсмічні датчики та мобільні пристрої. Вони використовуються урядовими організаціями, дослідниками та іншими зацікавленими сторонами для прогнозування стихійних явищ і мінімізації їх впливу.

Однією з головних проблем у використанні наборів даних про стихійні явища є обмежена доступність відкритих наборів даних. Хоча існує багато джерел даних про природні катастрофи, включаючи приватні компанії та державні установи, доступ до цих даних часто обмежений через конфіденційність або право власності.

У багатьох випадках набори даних про стихійні явища можуть збиратися приватними компаніями, які спеціалізуються на моніторингу та аналізі природних катастроф. Ці компанії можуть неохоче надавати свої дані безкоштовно, оскільки вони покладаються на ці дані як на ключовий компонент своєї бізнес-моделі. Аналогічно, державні органи можуть мати побоювання щодо оприлюднення конфіденційних даних, які можуть поставити під загрозу приватне життя людей або розкрити стратегічну інформацію про інфраструктуру.

Оскільки для побудови моделі було вирішено шукати необхідну інформацію на ресурсах метеорологічних служб. Відомим і надійним джерелом є NEXRAD Weather Radar Data (див. рис. 3.1), надані Національним центром даних про клімат (NOAA).

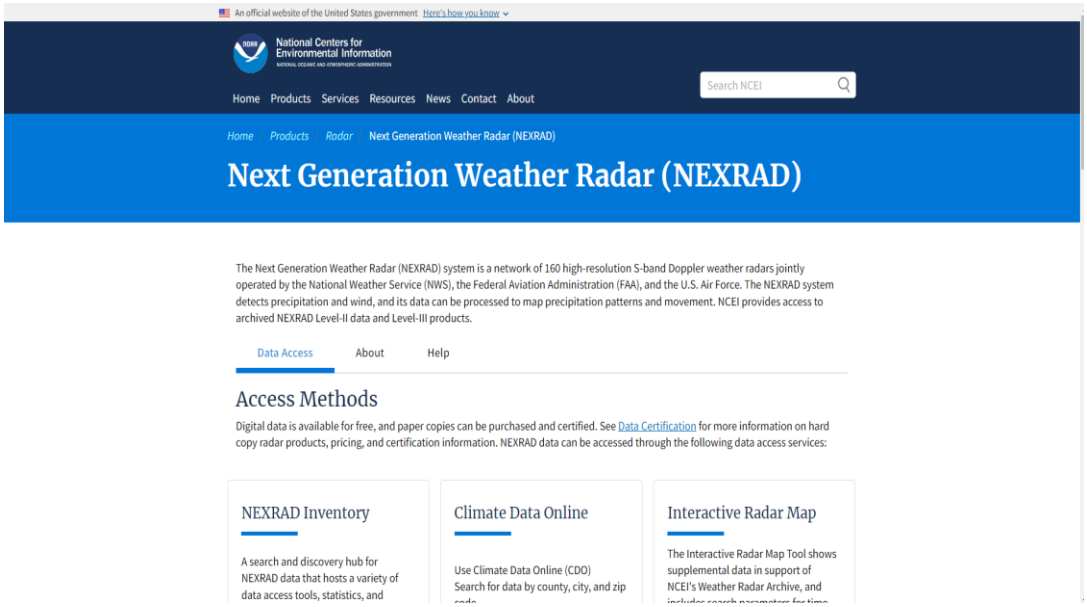

Рисунок 3.1 – NEXRAD Weather Radar Data

NEXRAD Weather Radar Data – це сервіс, який надає доступ до радіолокаційних даних про погодні умови. Він використовується для моніторингу та аналізу погодних явищ, таких як урагани, торнадо та грози. Дані зібрані з мережі радіолокаційних станцій NEXRAD.

Однією з переваг даних, наданих NEXRAD, є те, що вони забезпечують доступ до великого і постійно оновлюваного набору даних, отриманих з фактичних спостережень. Це забезпечує рівень деталізації та точності, якого важко досягти за допомогою інших джерел даних, таких як поодинокі датчики або моделі. Дані розбиті на набори, кожен з яких містить інформацію про погодні умови та радіолокаційні зображення за кожну годину кожного дня конкретного місяця. Для включення даних до моделі вони мають містити інформацію про щонайменше 5 унікальних спостережень на досліджуваній ділянці у відповідний момент часу. Хоча дані, що надаються NEXRAD, мають певні обмеження, наприклад, вони відображають лише частину загальної картини природних явищ в певному регіоні.

Ще однією перевагою даних NEXRAD є те, що вони надаються у стандартизованому форматі, що дозволяє легко порівнювати та аналізувати дані з різних джерел. Це може бути особливо цінним для дослідників і урядових органів,

яким потрібно проаналізувати дані з кількох регіонів або країн. Набір даних включає наступні поля:

- рік – рік спостереження;
- місяць – номер місяця спостереження;
- день – день спостереження;
- година – година спостереження за місцевим часом;
- utc_timestamp – дата та час спостереження у форматі UTC;
- latitude – широта місця спостереження;
- longitude – довгота місця спостереження;
- radar_reflectivity – відображення радара в дБЗ (децибелах Z);
- wind_speed – швидкість вітру в метрах на секунду;
- wind_speed – швидкість вітру в метрах на секунду;
- precipitation_rate – інтенсивність опадів у мм/год.

Ці дані дозволяють аналізувати та прогнозувати стихійні явища, використовуючи сучасні методи машинного навчання, такі як нейронні мережі, включаючи CNN, RNN та DNN. Завдяки детальному та стандартизованому формату даних дослідники можуть отримувати точніші прогнози та приймати обґрунтовані рішення щодо попередження та мінімізації впливу стихійних явищ.

3.2 Вибір методів дослідження

Хоча існує багато методів для прогнозування стихійних явищ, кожен з них має свої переваги та недоліки. Вибір методів на основі нейронних мереж, таких як CNN, RNN і DNN, є доцільним з кількох причин.

По-перше, згорткові нейронні мережі (CNN) є надзвичайно ефективними для обробки даних у вигляді зображень та інших багатовимірних масивів. Вони здатні автоматично виявляти важливі характеристики в даних, що робить їх ідеальними для аналізу супутникових зображень та інших візуальних даних, пов'язаних зі стихійними явищами, такими як урагани, повені та лісові пожежі. Здатність CNN виявляти та аналізувати просторові структури в зображеннях дозволяє їм

витягувати корисну інформацію, яка може бути критично важливою для точного прогнозування цих явищ. Приклад нейроної мережі CNN (див. рис. 3.2).

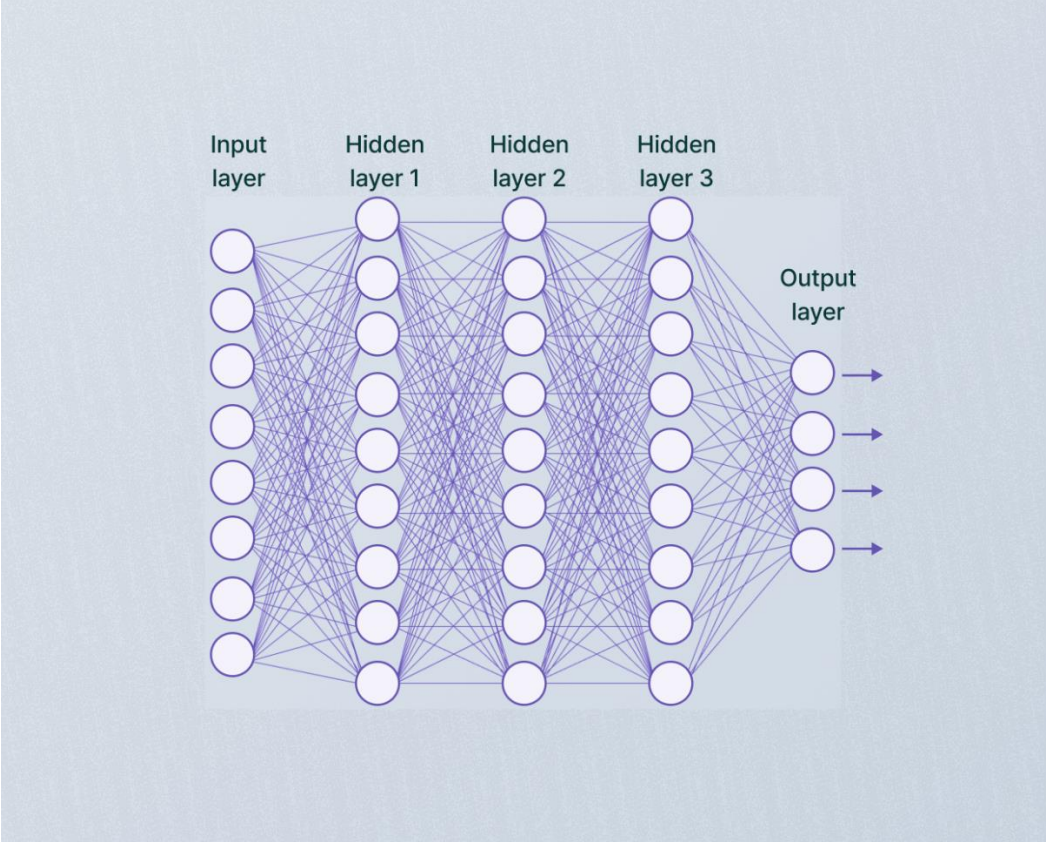

Рисунок 3.2 – Приклад нейроної мережі CNN

По-друге, рекурентні нейронні мережі (RNN) є особливо корисними для аналізу часових рядів і послідовних даних. Вони здатні запам'ятовувати попередні стани і використовувати цю інформацію для прогнозування майбутніх значень. Це важливо для передбачення стихійних явищ, таких як землетруси, де поточні умови можуть бути сильно залежні від попередніх подій. RNN забезпечують можливість враховувати тимчасову динаміку даних, що є ключовим фактором для прогнозування багатьох природних катастроф. Приклад нейроної мережі RNN (див. рис. 3.3).

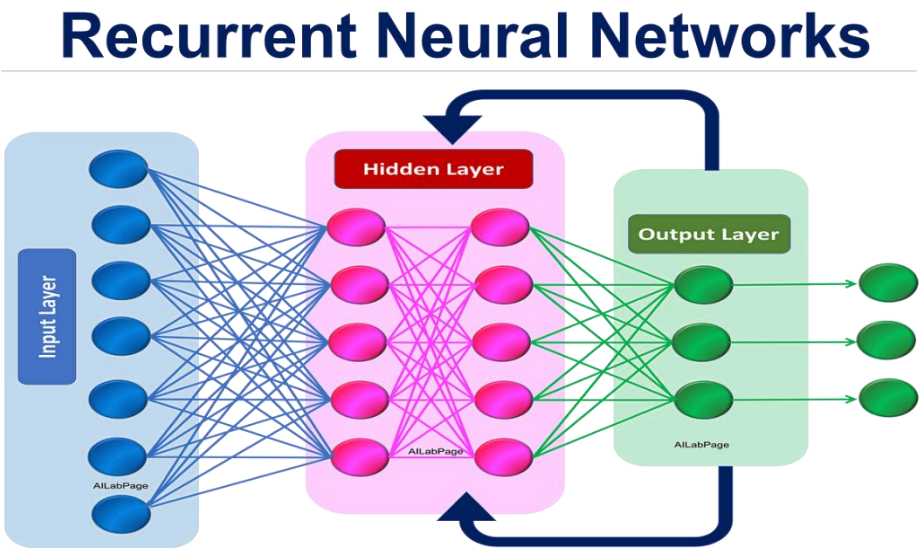

Рисунок 3.3 – Приклад нейроної мережі RNN

По-третє, глибокі нейронні мережі (DNN) здатні моделювати складні нелінійні взаємодії між численними факторами, що впливають на стихійні явища. Вони можуть бути використані для інтеграції різних типів даних, таких як метеорологічні дані, сейсмічні дані та інші релевантні показники, що дозволяє створювати більш точні та надійні прогнози. DNN забезпечують високу гнучкість у моделюванні складних процесів, що є важливим для розуміння та передбачення різних аспектів стихійних явищ. Приклад нейроної мережі RNN (див. рис. 3.3).

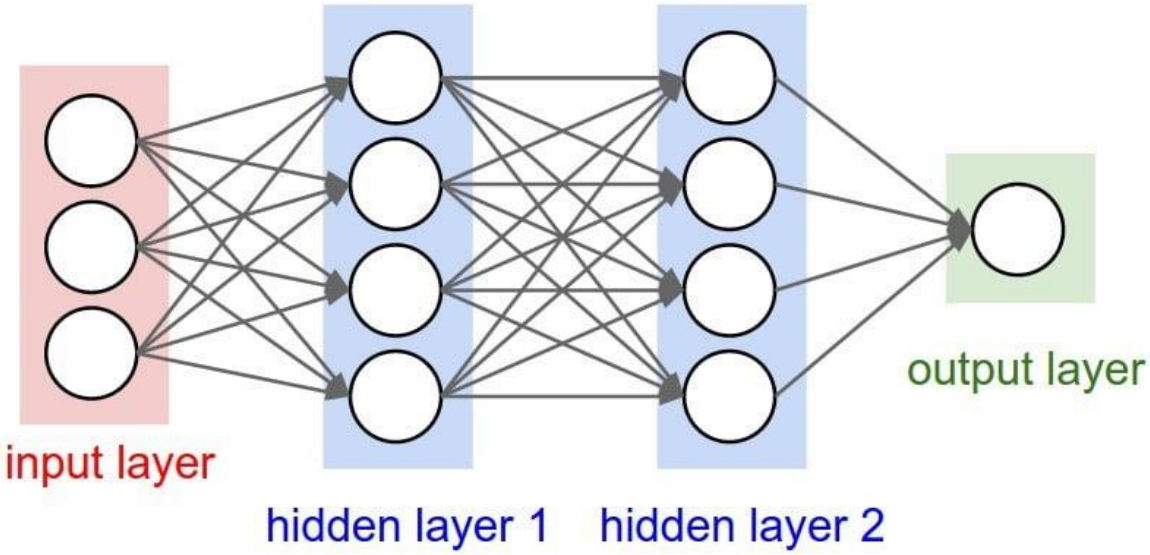

Рисунок 3.3 – Приклад нейроної мережі RNN

Виходячи з обраного набору даних, були обрані наступні методи машинного навчання для їх подальшого аналізу: CNN, RNN та DNN.

CNN добре працює для обробки просторових даних, що дозволяє використовувати їх для аналізу супутникових знімків і інших візуальних даних. Їхня здатність виявляти локальні патерни робить їх надзвичайно корисними для виявлення ранніх ознак стихійних явищ на основі зображень.

RNN підходять для аналізу часових рядів і послідовних даних, що робить їх ефективними для прогнозування явищ, що розвиваються з часом. RNN можуть використовувати контекст попередніх даних для поліпшення точності прогнозів майбутніх подій.

DNN дозволяє інтегрувати і обробляти різноманітні типи даних, враховуючи складні нелінійні взаємодії між ними. Це робить їх потужним інструментом для створення комплексних моделей, які можуть враховувати широкий спектр впливових факторів.

Кожен з цих методів має свої переваги та недоліки, які можуть впливати на його ефективність при прогнозуванні стихійних явищ. Наприклад, CNN можуть бути менш ефективними для аналізу часових рядів, але чудово підходять для обробки візуальних даних. RNN можуть краще прогнозувати послідовні події, але можуть бути складними в навчанні через проблему зникаючих градієнтів. DNN можуть забезпечити високу точність прогнозування, але вимагають великих обчислювальних ресурсів.

Загалом, вибір методу для прогнозування стихійних явищ залежить від конкретних умов і вимог задачі. Часто використовують комбінації з декількох методів, які доповнюють один одного, для отримання кращих результатів. Тому у багатьох дослідженнях застосовують гібридні моделі, що поєднують переваги CNN, RNN і DNN для досягнення найточніших прогнозів.

Таким чином, вибір нейронних мереж, таких як CNN, RNN і DNN, є виправданим підходом до прогнозування стихійних явищ завдяки їх здатності обробляти різні типи даних та моделювати складні взаємодії. Інтеграція цих

методів може забезпечити більш точні і надійні прогнози, що є критично важливим для попередження та управління наслідками природних катастроф.

3.3 Метрики оцінювання

Точність прогнозу буде оцінена за допомогою різних метрик, таких як середня абсолютна відсоткова помилка (MAPE), середня абсолютна помилка (MAE), середньоквадратична помилка (RMSE), коефіцієнт детермінації (R^2).

Середня абсолютна відсоткова помилка (MAPE) — це метрика, яка використовується для оцінки точності прогнозу в задачах регресії. Вона обчислюється як середнє значення абсолютних відсоткових похибок між фактичними та прогнозованими значеннями за формулою 3.1:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (3.1),$$

де N – кількість точок даних;

y_i – спостережене значення;

$\hat{y}_i$ – прогнозоване значення.

MAPE є зручною метрикою, оскільки виражає помилку у відсотках, що робить її інтуїтивно зрозумілою.

Середня абсолютна помилка (MAE) — це метрика, яка використовується для оцінки точності моделей у задачах регресії. Вона вимірює середню величину абсолютних помилок між фактичними і прогнозованими значеннями за формулою 3.2:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i} \quad (3.2),$$

де N – кількість точок даних;

y_i – спостережене значення;

$\hat{y}_i$ – прогнозоване значення.

MAE є зручною метрикою, оскільки її значення легко інтерпретувати як середню величину похибки в тих же одиницях, що й вихідні дані.

Середньоквадратична помилка (RMSE) — це метрика, яка використовується для оцінки точності моделей у задачах регресії. Вона вимірює середнє значення квадратів помилок між фактичними і прогнозованими значеннями за формулою 3.3:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (3.3),$$

де N – кількість точок даних;

y_i – спостережене значення;

$\hat{y}_i$ – прогнозоване значення.

RMSE є важливою метрикою, оскільки вона надає більшу вагу більшим помилкам через квадратичне обчислення, що дозволяє краще ідентифікувати моделі з великими похибками.

Коефіцієнт детермінації (R^2) — це метрика, яка використовується для оцінки якості моделей у задачах регресії. Вона показує, яка частка варіації залежної змінної пояснюється незалежними змінними моделі. Значення R^2 варіюється від 0 до 1, де 1 означає, що модель повністю пояснює варіацію залежної змінної, а 0 означає, що модель не пояснює варіацію взагалі, розраховується за формулою 3.4:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (3.4),$$

де y_i – спостережене значення;

$\hat{y}_i$ – прогнозоване значення.

4 ПРОВЕДЕННЯ ЕКСПЕРИМЕНТАЛЬНОЇ ЧАСТИНИ ДОСЛІДЖЕННЯ

4.1 Аналіз та підготовка набору даних

Першим кроком є збір даних. Це дані, завдяки яким можна передбачити стихійні явища, і їх можна зібрати за допомогою спеціальних пристроїв або станцій, які збирають метеорологічні дані. Для проведення аналізу та тестування моделей нейронних мереж було використано сервіс NEXRAD, який надає дані у відкритому доступі.

Для проведення експерименту було завантажено набір даних Weather Radar Data з Бостону за січень та лютий 2023 року. Ці дані включають різноманітну інформацію про погодні умови, таку як опади, швидкість та напрямок вітру, температура, вологість, а також дані радіолокаційного спостереження. Далі проводиться аналіз та підготовка цього набору даних.

Також було проаналізовано вплив набору даних на оцінку прогнозування стихійних явищ. Експериментальні результати показують, що прогностичний ефект багатомасштабної моделі набагато кращий, ніж у одномасштабного прогнозування. Це повністю відображає важливість використання великого та різноманітного набору даних. Додавання більшої кількості інформації має значну дослідницьку цінність. Ці ресурси можуть забезпечити подальше розуміння та перспективи використання методів науки про дані та машинного навчання для прогнозування та аналізу природних явищ і тенденцій.

Для дослідження були обрані дані з штату Массачусетс, США.

У нього входять такі значення як wind_speed, precipitation_rate, soil_moisture, humidity, тощо.

У ці роботі одним з основних інструментів програмування є мова Python з кількох причин. Вона має інтуїтивно зрозумілий синтаксис. Також вона має великий вибір бібліотек та фреймворків для моделей машинного навчання. Також одним з плюсів полягає в тому, що це сучасна мова програмування, яка може бути легко інтегрована з іншими мовами, якщо це необхідно. Для моделювання алгоритму прогнозування було використано Python 3.10 та Jupyter Notebook.

Для коректної роботи необхідно попередньо обробити досліджувані дані, зокрема, якимось чином відновити відсутні значення. Це може бути пов'язано зі помилками обладнання яке збірало дані.

Також варто звертати увагу на те, що є залежність появи стихійного явища в залежності від сезону року. Кожен сезон має свої специфічні погодні умови, які можуть призводити до різних типів стихійних явищ. Наприклад, весна часто супроводжується повенями через танення снігу і весняні дощі, а також можливістю виникнення торнадо через зіткнення теплих і холодних повітряних мас. Літо характеризується сезоном ураганів, грозами та блискавками, які можуть спричиняти лісові пожежі. Восени також триває сезон ураганів, а осінні шторми можуть спричиняти сильні вітри та опади. Зима приносить снігові бурі та льодові шторми, що можуть впливати на транспорт і інфраструктуру. Ці сезонні фактори обов'язково враховуються при аналізі та прогнозуванні стихійних явищ, оскільки вони допомагають краще зрозуміти, які ризики є найбільш імовірними у певний час року.

Зібрані дані часто не є досконалими і не можуть бути використані одразу. Вони можуть містити прогалини або непотрібну інформацію, яка перевантажить алгоритм, що в результаті призведе до неточних прогнозів. Прогалини в зібраних телекомунікаційних даних можуть виникати з різних причин, таких як системні проблеми, втрата пакетів, перешкоди та інші технічні несправності. Крім того, дані можуть містити помилки вимірювання датчиків, викиди, а також інші аномалії, які впливають на їхню якість. Викиди можуть виникати через аномальні значення, спричинені несправністю обладнання або зовнішніми факторами. Прогалини можуть бути результатом тимчасової втрати зв'язку з датчиками або відсутністю подій, які фіксуються. Всі ці проблеми ускладнюють процес аналізу та прогнозування, тому важливо правильно обробляти та очищати дані перед їх використанням в алгоритмах машинного навчання. Це включає в себе виявлення та заповнення пропущених значень, фільтрацію викидів і корекцію помилок, що дозволяє підвищити точність і надійність прогнозів.

4.2 Опис експерименту

Для експерименту були взяті дані з штату Масачусетс. Для будування прогнозу використовувалися історичні дані з 1 січня по 31 лютого 2023 року. Усі дані були розділені на навчальну та тестову вибірки.

Були реалізовані моделі для 3 алгоритмів CNN, RNN та DNN. Налаштування параметрів усіх моделей експерименту наведено в таблиці 4.1 (див. табл. 4.1).

Таблиця 4.1 – Опис параметрів алгоритмів (Таблиця виконана самостійно)

Назва алгоритму	Використані параметри
CNN	kernel_size=3x3, pooling_size=2x2, activation='relu, softmax', optimizer='adam', loss='sparse_categorical_crossentropy', epochs=10
RNN	units = 50, activation='relu', optimizer='adam', loss= mean_squared_error', epochs=10
DNN	units = 64, activation='relu', optimizer='adam', loss= mean_squared_error', epochs=10

Якість прогнозування можна оцінити за допомогою показників MAE, MAPE, RMSE та R^2.

Машинне навчання є ресурсоємним процесом, який вимагає великої кількості оперативної пам'яті для ефективної роботи. Для прискорення доступу до оперативної пам'яті необхідно використовувати процесор, що підтримує чотири канали пам'яті замість стандартних двох, які зазвичай використовуються в звичайних кастомних рішеннях. Це значно підвищує швидкість обробки даних та оптимізує роботу системи.

Крім того, важливо враховувати кількість ядер процесора та обсяг пам'яті відеокарти при виконанні завдань машинного навчання. Процесор з більшою

кількістю ядер здатен паралельно обробляти більшу кількість завдань, що прискорює час тренування моделі. Обсяг пам'яті відеокарти також має велике значення, оскільки більші моделі потребують більше графічної пам'яті для обробки складних обчислень.

Час тренування моделі сильно залежить від обчислювальної потужності системи. Для тестування часу тренування була використана система з процесором Intel Core i7 та 16 ГБ оперативної пам'яті. Результати показали, що навіть з такою конфігурацією час тренування може бути значним, тому для більш ефективної роботи машинного навчання рекомендується використовувати ще потужніші системи з більшою кількістю оперативної пам'яті та більш продуктивними процесорами і відеокартами.

Враховуючи всі ці фактори, можна значно підвищити ефективність та швидкість тренування моделей машинного навчання, що дозволяє отримувати більш точні та швидкі результати.

4.4 Отримані результати

В результаті експерименту було проведено порівняння трьох типів нейронних мереж: CNN (Convolutional Neural Network), RNN (Recurrent Neural Network) та DNN (Deep Neural Network). Моделі було порівняно за їхньою здатністю передбачати різні типи стихійних лих таких як урагани, лісові пожежі, тсунамі, наводненя та інші. Для оцінки їхньої продуктивності використовувалися чотири основні метрики помилки: MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error) та R^2 (коефіцієнт детермінації). За допомогую цих метрик можливо виявити нейрону мережу яка більше всього підходить для передбачення стихійних лих із перелічених вище.

У таблиці 4.2 наведено результат оцінювання метрик помилки для прогнозування можливості появи урагану в штаті Масачусетс з 1 січня по 31 лютого 2023 року (див. табл. 4.2). Ці результати базуються на використанні різних моделей прогнозування, які були навчені на історичних даних та даних про поточні умови.

Таблиця 4.2 – Метрики помилки для прогнозування можливості появи урагану (Таблиця виконана самостійно)

Метрика	CNN	RNN	DNN
R^2	0.85	0.83	0.80
MAE	2.3	2.5	2.8
MAPE	5.2%	5.5%	6.0%
RMSE	3.1	3.3	3.5
Точність	88%	85%	82%

Вміст таблиці 4.3 наведено результат оцінювання метрик помилки для прогнозування можливості появи лісової пожежі в штаті Масачусетс з 1 січня по 31 лютого 2023 року (див. табл. 4.3).

Таблиця 4.3 – Метрики помилки для прогнозування можливості появи лісової пожежі (Таблиця виконана самостійно)

Метрика	CNN	RNN	DNN
R^2	0.88	0.85	0.82
MAE	2.0	2.2	2.5
MAPE	4.8%	5.1%	5.5%
RMSE	2.9	3.1	3.4
Точність	90%	87%	84%

Вміст таблиці 4.4 показує результати оцінювання метрик помилки для прогнозування можливості появи цунамі в штаті Массачусетс у період з 1 січня по 31 лютого 2023 року (див. табл. 4.4). Ці результати базуються на використанні різних моделей прогнозування, які були навчені на історичних даних та даних про поточні умови.

Таблиця 4.4 – Метрики помилки для прогнозування можливості появи тсунамі (Таблиця виконана самостійно)

Метрика	CNN	RNN	DNN
R^2	0.86	0.81	0.81
MAE	2.1	2.3	2.6
MAPE	4.9%	5.2%	5.6%
RMSE	2.8	3.0	3.3
Точність	89%	86%	83%

Вміст таблиці 4.5 наведено результат оцінювання метрик помилки для прогнозування можливості появи наводнення в штаті Масачусетс з 1 січня по 31 лютого 2023 року (див. табл. 4.5).

Таблиця 4.5 – Метрики помилки для прогнозування можливості появи наводнень (Таблиця виконана самостійно)

Метрика	CNN	RNN	DNN
R^2	0.87	0.84	0.79
MAE	2.1	2.1	2.5
MAPE	5.3%	6.1%	6.6%
RMSE	2.3	2.5	2.8
Точність	90%	85%	80%

Вміст таблиці 4.6 показує результат прогнозування вірогідності появи стихійних лих в штаті Масачусетс на основі данних які були введені (див. табл. 4.6). На основі проаналізовних данних був виконаний аналіз за допомгою нейроних мереж. Отриманий аналіз був порівняний з реальними даними про стихійні лиха, що відбулися у штаті Масачусетс протягом досліджуваного періоду.

Таблиця 4.6 – Результат прогнозування можливості появи стихійних лих (Таблиця виконана самостійно)

	Ймовірність	CNN	RNN	DNN
Ураган	40%	38%	41%	35%
Лісова пожежа	50%	45%	43%	40%
Тсунамі	10%	10%	10%	8%
Наводненя	50%	49%	48%	45%

Час прогнозування склав 4.94 с.

4.5 Аналіз результатів

В результаті експерименту було отримано результати, всі нейроні мережі показали схожі результати. А ось за часом помітна вагома різниця найшвидшим виявився CNN, а найдовшим RNN.

Метод згорткових нейронних мереж (CNN) зазвичай працює швидше за глибокі нейронні мережі (DNN) у випадках, коли набір даних не дуже великий. Глибокі нейронні мережі можуть мати сотні шарів, що потребує багато часу та обчислювальних ресурсів. Кожен шар глибокої нейронної мережі виконує обробку даних на основі деякої підмножини ознак, що може призвести до перенавчання, якщо кількість ознак дуже велика.

На невеликих об'ємах даних глибокі нейронні мережі (DNN) можуть прогнозувати швидше, оскільки вони зазвичай мають меншу складність та обчислювальну складність, ніж CNN. Однак, зі збільшенням об'єму даних, різниця в часі виконання між цими моделями може стати менш помітною, або навіть помінятися місцями. Отже, вибір між методом залежить від конкретної задачі, розміру даних та характеристик ознак.

На основі результатів експерименту можна зробити висновок, що згорткові нейронні мережі (CNN) є найбільш точною та ефективною моделлю для

передбачення різних екстремальних погодних явищ, таких як урагани, лісові пожежі та інші стихійні лиха. Ці моделі демонструють високу точність у прогнозуванні завдяки здатності ефективно обробляти та аналізувати просторові дані.

Рекурентні нейронні мережі (RNN) показують середні результати в задачах прогнозування погодних явищ. Вони можуть бути гарною альтернативою для певних типів даних, особливо коли важливі тимчасові послідовності. RNN добре справляються з обробкою часових рядів, але іноді можуть поступатися CNN у точності та ефективності.

Глибокі нейронні мережі (DNN), хоча і здатні вирішувати складні задачі, показують найнижчу точність у прогнозуванні екстремальних погодних явищ у порівнянні з CNN та RNN. Ці моделі потребують додаткової оптимізації, щоб досягти кращих результатів у задачах прогнозування стихійних лих. Оптимізація може включати підбір гіперпараметрів, додавання регуляризації та використання більш потужних обчислювальних ресурсів для тренування моделей на великих наборах даних.

Отже, підсумовуючи, CNN виділяються як найбільш потужний інструмент для прогнозування екстремальних погодних явищ завдяки їхній здатності ефективно обробляти складні просторові структури даних. RNN можуть слугувати ефективною альтернативою для задач, де важливі тимчасові залежності. DNN, з іншого боку, потребують ретельної оптимізації для покращення своїх результатів у сфері прогнозування стихійних лих. Вибір моделі залежить від конкретних характеристик даних та вимог до точності прогнозу.

4.6 Рекомендації щодо подальших досліджень

У подальших дослідженнях можна запропонувати кілька напрямків, які можуть значно покращити результати передбачення стихійних явищ. Перш за все, інтеграція карти місцевості у модель прогнозування може значно підвищити точність передбачення стихійних явищ. Детальна інформація про топографію, типи ґрунтів, розташування водних ресурсів та інші географічні фактори може надати

моделі додатковий контекст, що сприятиме точнішому прогнозуванню. Наприклад, дані про висоту місцевості можуть бути корисними для передбачення ризиків затоплення, тоді як інформація про типи рослинності може допомогти краще оцінити ймовірність лісових пожеж. Інтеграція таких даних може дозволити створити більш точні та детальні моделі, які враховують всі важливі фактори ризику для конкретних регіонів.

Варто також дослідити ефективність інших типів нейронних мереж, таких як трансформери, графові нейронні мережі або моделі зворотного розповсюдження. Це може дозволити виявити більш ефективні моделі для прогнозування стихійних явищ або вдосконалити існуючі моделі. Наприклад, графові нейронні мережі можуть бути особливо корисними для обробки даних, що мають складні взаємозв'язки, такі як мережі річок або транспортні системи. Аналіз ефективності різних архітектур нейронних мереж може призвести до значного покращення точності та надійності прогнозів. Використання трансформерів, які добре себе зарекомендували в обробці природної мови, може надати нові підходи до аналізу часових рядів та просторових даних.

Додавання більшої кількості стихійних явищ до моделі може зробити систему більш детальною і корисною для практичних застосувань. Розширення набору даних для включення таких подій, як землетруси, цунамі, бурі та інші екстремальні погодні явища, дозволить створити більш універсальну систему. Це також сприятиме кращому розумінню взаємозв'язків між різними типами стихійних явищ і допоможе розробити комплексні стратегії для їхнього прогнозування та управління ризиками. Таке розширення моделі може також дозволити краще оцінювати кумулятивний вплив різних явищ, які можуть виникати одночасно або послідовно, посилюючи негативні наслідки один одного.

Дослідження додаткових аудіо характеристик може допомогти виявити нові індикатори, пов'язані зі стихійними явищами. Наприклад, аналіз звукових сигналів, що супроводжують урагани або землетруси, може надати додаткові дані для більш точного передбачення їхнього настання або інтенсивності. Це також може бути корисним для розробки систем раннього попередження, які використовують

звукові сигнали для ідентифікації потенційних загроз. Використання аудіо даних може надати нові можливості для виявлення ранніх ознак стихійних явищ, які не завжди можуть бути помічені за допомогою традиційних методів моніторингу.

На основі отриманих моделей варто розробити повноцінну програмну систему, яка буде здатна працювати в реальному часі. Така система може інтегрувати всі вищезазначені аспекти і надавати точні та своєчасні прогнози для широкого спектра стихійних явищ. Реалізація реального часу дозволить системі оперативно реагувати на змінювані умови і забезпечувати ефективніші заходи з попередження та реагування на стихійні лиха. Ця система може бути корисною для урядових організацій, служб екстреного реагування та інших зацікавлених сторін, що займаються управлінням ризиками стихійних явищ. Така інтегрована система може забезпечити безперервний моніторинг та своєчасне виявлення потенційних загроз, що є критично важливим для мінімізації втрат і зменшення шкоди.

Таким чином, ці напрямки досліджень можуть значно покращити точність і надійність прогнозування стихійних явищ, сприяючи зменшенню їхнього негативного впливу на суспільство. Реалізація таких підходів потребує мультидисциплінарного підходу, залучення експертів з різних галузей та використання передових технологій. Це дозволить створити більш ефективні системи для передбачення, моніторингу та управління стихійними явищами, що є надзвичайно важливим для забезпечення безпеки та стабільності у різних регіонах світу.

ВИСНОВКИ

У цій роботі було досліджено доцільність використання різних моделей для передбачення стихійних явищь. Були порівняні нейроні мережі CNN, RNN і DNN. Найкращим з них виявився CNN.

У результаті виконання роботи було описано аналіз предметної галузі, опис основних факторів, що впливають на умови передбачення стихійних явищь, поставлене завдання дослідження та описані його етапи, наведено аналіз існуючих методів, що широко засовуються для передбачення стихійних явищь, проведене планування експериментального дослідження, що включає вибір набору даних, вибір методів, а саме регресійних моделей, та метрик для оцінювання, проведене дослідження і описані його результати.

Загалом, робота дає уявлення про потенціал нейроних мереж для прогнозування погодних явищ і може стати основою для майбутніх досліджень і моделювання в цій галузі.

У майбутньому ми плануємо провести дослідження та порівняти ефективність інших нейроних мереж, які можна застосувати для погодних явищ. Ми можемо також збільшити кількість стихійних явищ які будуть передбачатися.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. How Does Back-Propagation Work in Neural Networks? URL: https://towardsdatascience.com/how-does-back-propagation-work-in-neural-networks-with-worked-example-bc59dfb97f48 (дата звернення: 25.04.2023).

2. Romm J. Climate Change: What Everyone Needs to Know. Oxford University Press, Incorporated, 2022, с. 100-200. (дата звернення: 25.04.2023)

3. Aggarwal C. C. Neural Networks and Deep Learning: A Textbook. Springer, 2018. С. 300-497. (дата звернення: 25.04.2023)

4. Ready to Rumble. Predicting the Unpredictable. 2010. С. 1–11. URL: https://doi.org/10.1515/9781400883547-002 (дата звернення: 02.06.2024).

5. Burkov A. The Hundred-Page Machine Learning Book. Andriy Burkov, 2019. 160 с.

6. Taylor D. C. Science of Natural Disasters: When Nature and Humans Collide. Nomad Press, 2020. 128 с.

7. Back-Propagation. Deep Learning Neural Networks. 2016. С. 23–32. URL: https://doi.org/10.1142/9789813146464_0003 (дата звернення: 02.06.2024).

8. Chollet F. Deep Learning with Python. Manning Publications Co. LLC, 2017. 384 с.

9. Maksym Bekuzarov, Oleksandr Samantsov, Oksana Mazurova,Mariia Shirokopetleva. Neural Network Architecture Editor With Code Generation. Problem of Infocommunications. Science and Technology (PIC S&T′2020), Kharkiv, Ukraine.- 6- 9 October 2020.

10. Maksym Shopynskyi, Nataliia Golian, Iryna Afanasieva. Long short-term memory model appliance for generating music compositions // 2020 IEEE International ScientificPractical Conference Problems of Infocommunications, Science and Technology (PICS&T), 6-9 Oct. 2020, Kharkiv, Ukraine.

ПЕРЕЛІК ДЖЕРЕЛЬ ПОСИЛАННЯ ЗА НАУКОВИМИ НАПРЯМАМИ КЕРІВНИКА ТА НАУКОВЦІВ КАФЕДРИ ПРОГРАМНОЇ ІНЖЕНЕРІЇ

9. Maksym Bekuzarov, Oleksandr Samantsov, Oksana Mazurova,Mariia Shirokopetleva. Neural Network Architecture Editor With Code Generation. Problem of Infocommunications. Science and Technology (PIC S&T′2020), Kharkiv, Ukraine.- 6- 9 October 2020.

10. Maksym Shopynskyi, Nataliia Golian, Iryna Afanasieva. Long short-term memory model appliance for generating music compositions // 2020 IEEE International ScientificPractical Conference Problems of Infocommunications, Science and Technology (PICS&T), 6-9 Oct. 2020, Kharkiv, Ukraine.