КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (20), 2023
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2023.20.2034
УДК 316.776.32:004.738.52

Тищенко Віталій Сергійович
Ассистент кафедри управління інформаційною та кібернетичною безпекою
Державний університет телекомунікацій, Київ, Україна
ORCID ID: 0000-0003-3849-6243
tvs5vetal@email.com

АНАЛІЗ МЕТОДІВ НАВЧАННЯ ТА ІНСТРУМЕНТІВ НЕЙРОМЕРЕЖ ДЛЯ ВИЯВЛЕННЯ ФЕЙКІВ

Анотація. У цій статті проводиться аналіз різних методів навчання та інструментів нейромереж для виявлення фейків. Розглядаються підходи до виявлення фейків на основі текстових, візуальних та змішаних даних, а також використання різних типів нейромереж, таких як рекурентні нейронні мережі, згорткові нейронні мережі, глибока нейронна мережа, генеративні змагальні мережі та. Також розглядаються методи навчання з вчителем та без вчителя, такі як автокодувальні нейромережі та глибокі варіаційні автокодувальні нейромережі.

На основі проаналізованих досліджень, звертається увага на проблеми, пов'язані з обмеженнями в обсязі та якості даних, а також недостатньою ефективністю інструментів для виявлення складних типів фейків. Було проаналізовано програми та інструменти, які засновані на нейронних мережах, та зроблено висновки про їх ефективність та відповідність різним типам даних та задач виявлення фейків.

У результаті дослідження встановлено, що моделі машинного та глибинного навчання, а також методи змагального навчання та спеціальні інструменти для виявлення фейкових медіа є ефективними у виявленні фейків. Однак на ефективність і точність цих методів та інструментів можуть впливати такі фактори, як якість даних, методи, що використовуються для навчання та оцінювання, а також складність фейкових медіа, які виявляються. На основі аналізу методів навчання та характеристик нейромереж, визначено переваги та недоліки у виявленні фейків. Постійні дослідження і розробки в цій галузі мають вирішальне значення для підвищення точності та надійності цих методів та інструментів для виявлення фейків.

Ключові слова: фейкові новини; інструменти виявлення фейків; нейромережі; методи навчання; методи виявлення дезінформації та фейкових новин у мережі Інтернет.

ВСТУП

Останніми роками проблема фейкових новин набуває все більшого поширення в медіапросторі. Поширення неправдивої інформації може завдати шкоди та підірвати довіру до традиційних джерел новин та інформації. З розвитком соціальних мереж і можливістю будь-кого створювати та поширювати контент проблема фейкових новин стала ще складнішою.

Щоб розв'язувати цю проблему, дослідники розробляють і вдосконалюють інструменти для виявлення та боротьби з фейковими новинами. Одним із підходів, який набув популярності, є використання нейронних мереж — типу алгоритму машинного навчання, який може навчитися виявляти закономірності в даних.

Проблема фейків (fake news) стала досить актуальною в сучасному світі, а боротьба з нею стає все складнішою і витонченішою. Оскільки фейків може бути надто багато для ручного перевірки та їх виявлення, нейромережеві інструменти стають все важливішими в цій області [1].

У аналізі виявлення фейків застосовуються різні типи нейромереж, серед яких можна виділити наступні:

1. Рекурентні нейронні мережі (RNN) - використовуються для аналізу тексту та останніх кадрів у відео та зображенн.

2. Згорткові нейронні мережі (CNN) - використовуються для аналізу зображень і відео.

3. Трансформаторні нейронні мережі — ефективні для аналізу тексту, в тому числі для виявлення перекладеного фейкового тексту.

4. Генеративні змагальні мережі (Generative Adversarial Networks, GAN) - архітектура глибоких нейронних мереж, яка може генерувати синтетичні дані та може бути використана для виявлення фальшивих зображень або відео.

5. Глибока нейронна мережа (DNN) - нейронна мережа, яка має понад три приховані шари та може бути використана для виявлення фейків.

Важливо зазначити, що ефективність цих моделей залежить від різних факторів, таких як якість даних, методи навчання та оцінка моделі. Подальші дослідження в цій галузі необхідні для підвищення точності та надійності цих моделей для виявлення фейкових медіа за допомогою нейронних мереж.

Кожен з цих методів має свої переваги та недоліки і використовується залежно від конкретної задачі та особливостей дослідження. Проте зазвичай використовують комбінацію різних методів для отримання найкращого результату. Ключовим фактором успіху при виявленні фейків є якість та обширність даних, що використовуються для навчання і тестування моделей. Також важливо, щоб моделі були постійно оновлювані і підлаштовувались під нові типи фейків, які можуть з'являтися від щодня.

Постановка проблеми. Одним з можливих підходів до розв'язання цієї проблеми є використання нейромереж - машинного навчання, яке дозволяє розв'язувати складні завдання за допомогою аналізу великої кількості даних. У контексті виявлення фейків, нейромережі можуть бути натреновані на даних з різних джерел, щоб розпізнавати ознаки неправдивої інформації. Однак, існують різні інструменти на основі нейромереж, які можуть використовуватись для виявлення фейків, і важливо зробити порівняльний аналіз цих інструментів, щоб визначити їх ефективність та переваги в різних контекстах. Результати даного дослідження можуть бути корисними для фахівців у галузі боротьби з фейками, журналістів, соціальних мереж та інших зацікавлених сторін, які прагнуть забезпечити якість та достовірність інформації, що публікується в мережі Інтернет.

Аналіз останніх досліджень і публікацій. На сьогодні існує багато інструментів і методів для виявлення фейків за допомогою нейромереж. При цьому, кожен з цих методів має свої переваги та недоліки.

Одним з найпоширеніших підходів є застосування глибоких нейронних мереж (deep neural networks). Основним перевагою цього підходу є висока точність виявлення фейків за рахунок використання глибокого навчання та великої кількості даних для навчання моделі [2]. Однак, цей підхід потребує великих обчислювальних ресурсів та тривалого часу на навчання.

Інший підхід, що використовують для виявлення фейків, полягає в застосуванні генеративно-протистоячих мереж (generative adversarial networks, GAN). За допомогою GAN можна відтворювати майже нерозрізнимі від оригіналу зображення та порівнювати їх з оригіналом для виявлення фейків. Однак, цей підхід також вимагає значних обчислювальних ресурсів та відносно великої кількості даних для навчання.

Інший метод використовує аналіз глибини пікселів зображення та використання зображень високої якості для виявлення фейків. Цей метод є ефективним для виявлення

фейкових зображень, створених за допомогою глибинного навчання. Однак, цей метод не є ефективним для виявлення фейкових зображень, створених за допомогою інших технік.

Загалом, можна сказати, що всі інструменти та методи для виявлення фейків за допомогою нейромереж мають свої переваги та недоліки, і вибір конкретного методу залежить від характеристик вхідних датасетів.

У роботі [3] автори розглядають питання виявлення фейків у зображеннях. Використовуються глибокі згорткові нейронні мережі для виконання класифікації зображень на фейкі та ні. Особливість підходу полягає у використанні механізму зміни розміру зображення для підвищення точності класифікації. Результати експерименту показали, що запропонований підхід дозволяє ефективно виявляти фейки в зображеннях.

В одному з досліджень Чжоу та інші [4] досліджували здатність соціальних медіа агрегувати судження великої спільноти користувачів. У своєму подальшому дослідженні вони пояснили підходи машинного навчання з кінцевою метою розробки кращого виявлення чуток. Вони досліджували труднощі поширення чуток, класифікацію чуток і обман для просування таких рамок. Вони також досліджували використання таких корисних стратегій для створення захоплюючих структур, які можуть допомогти людям визначитися з вибором щодо оцінки цілісності даних, зібраних із різних платформ соціальних мереж.

Восоугі та інші [5] розпізнали характерні ознаки чуток, досліджуючи три аспекти інформації, що поширюється в Інтернеті: мовний стиль, характеристики людей, які беруть участь у поширенні інформації, і тонкощі поширення інформації в мережі. Автори проаналізували запропонований ними алгоритм на основі 209 чуток, що представляють 938 806 твітів, зібраних із реальних подій, включаючи вибухи на Бостонському марафоні 2013 року, заворушення у Фергюсоні 2014 року та епідемію лихоманки Ебола 2014 року. Вони висловили ефективність запропонованої ними основи всіма існуючими методами. Основною метою їхнього дослідження було запровадити новий спосіб оцінки подібності стилів між різними текстовими вмістами. Вони реалізували численні моделі машинного навчання та досягли точності 51% для виявлення фейкових новин.

Чен та інші [6] запропонували модель навчання без нагляду, яка поєднує рекурентні нейронні мережі та автоматичні кодери, щоб відрізнити чутки як аномалії від інших надійних мікроблогів на основі поведінки користувачів. Результати експерименту показують, що запропонована ними модель змогла досягти точності 92,49% з показником F1 89,16%.

O'Brien та інші [7] застосували стратегії глибокого навчання для класифікації фейкових новин. У своєму дослідженні вони досягли точності 93,50% за допомогою методу чорної скриньки.

Метою статті є проведення аналізу методів навчання та інструментів на основі нейромереж для виявлення фейків. Стаття спрямована на визначення ефективності та переваг різних інструментів та методів навчання нейромережі в різних контекстах та на основі цього дослідження можна зробити висновки про найбільш ефективний інструмент для виявлення фейків.

ISSN 2663 - 4023

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Виявлення фейкових новин є активним напрямком досліджень в останні роки. Запропоновано різні підходи, зокрема традиційні методи машинного навчання, методи глибокого навчання та методи обробки природної мови (NLP).

Традиційні методи машинного навчання зазвичай використовують алгоритми керованого навчання для вивчення моделей класифікації, які розрізняють правдиві та неправдиві новини. Інженерія ознак є важливим аспектом у таких підходах, коли набір ознак витягується вручну з новинних статей за допомогою різних методів, таких як bag-of-words і tf-idf. Потім ці ознаки вводяться в класифікатор, такий як дерева рішень, випадкові ліси та машини опорних векторів, щоб робити прогнози. Крім того, алгоритми неконтрольованого навчання використовуються для вивчення закономірностей у даних, щоб виявити аномалії, які можуть бути потенційними індикаторами фейкових новин.

Методи глибокого навчання стали популярними для виявлення фейкових новин завдяки їхній здатності автоматично вивчати відповідні ознаки з необроблених даних. Згорткові нейронні мережі (ЗНМ) та рекурентні нейронні мережі (РНМ) є найбільш поширеними архітектурами для виявлення фейкових новин. Моделі на основі CNN використовують згорткові шари для вилучення ознак на різних рівнях, після чого додається повністю зв'язний шар для класифікації. Моделі на основі ШНМ обробляють вхідну послідовність ітеративно, фіксуючи часові залежності між словами. Двонаправлені ШНМ також широко використовуються для виявлення фейкових новин, оскільки вони можуть фіксувати інформацію як з минулого, так і з майбутнього контексту.

Методи, засновані на НЛП, також досліджувалися для виявлення фейкових новин. У цих методах інструменти обробки природної мови, такі як аналіз настроїв, розпізнавання сутностей і моделювання тем, використовуються для виявлення патернів у тексті, які вказують на фейкові новини.

Отже, існуючі підходи до виявлення фейкових новин використовують поєднання традиційних методів машинного навчання, моделей глибокого навчання та методів на основі НЛП. Кожен підхід має свої сильні та слабкі сторони, і вибір підходу залежить від конкретних потреб та обмежень проблеми, що розглядається. Необхідні подальші дослідження для розробки більш точних і надійних методів виявлення фейкових новин, які можуть адаптуватися до мінливої природи фейкових новин [8].

Методика і типи навчання. Існує два основних методи навчання нейромережевих моделей: контрольоване та неконтрольоване навчання.

Навчання під контролем передбачає надання моделі маркованих навчальних даних, де кожна точка даних асоціюється з відповідною міткою або виходом. Під час навчання модель намагається вивчити взаємозв'язок між вхідними даними та відповідним виходом шляхом коригування вагових коефіцієнтів та зсувів. Після завершення навчання модель може робити прогнози або класифікації на основі нових, раніше не бачених даних. У контексті виявлення фейкових новин контрольоване навчання можна використовувати для навчання нейромережевої моделі класифікації новинних статей як справжніх або фейкових на основі маркованих прикладів.

З іншого боку, неконтрольоване навчання не покладається на марковані дані для навчання. Натомість модель відповідає за виявлення закономірностей і взаємозв'язків у даних самостійно. Кластеризація та зменшення розмірності є поширеними методами неконтрольованого навчання, які використовуються для виявлення фейкових новин. При кластеризації модель намагається згрупувати схожі точки даних разом, тоді як при

ISSN 2663 - 4023

зменшенні розмірності модель намагається зменшити кількість вхідних ознак, зберігаючи при цьому дисперсію даних. Ці методи можуть допомогти виявити кластери схожих новинних статей або тем, які потім можна проаналізувати для виявлення потенційного контенту фейкових новин.

Загалом, як контрольовані, так і неконтрольовані методи навчання мають свої сильні та слабкі сторони, і вибір методу залежить від конкретної проблеми та наявних ресурсів. У той час як контрольоване навчання вимагає маркованих даних і обмежується якістю маркування, неконтрольоване навчання може дати уявлення про глибинну структуру даних, але може бути складним для інтерпретації та впровадження.

Отже, вибір методу навчання залежить від конкретної проблеми та наявних ресурсів. Як контрольовані, так і неконтрольовані методи навчання виявилися перспективними в контексті виявлення фейкових новин, причому контрольоване навчання особливо корисне для завдань класифікації, а неконтрольоване навчання - для виявлення закономірностей і взаємозв'язків у даних.

Існує багато інструментів для виявлення фейкових новин, багато з яких використовують методи машинного навчання та нейронних мереж. В дослідженні представлено декілька прикладів:

1. Factmata: Factmata поєднує машинне навчання, обробку природної мови та краудсорсингову експертизу для аналізу новинних статей на предмет достовірності

Factmata - це система надання послуг, яка використовує штучний інтелект для класифікації контенту в Інтернеті на основі його достовірності. Принцип роботи Factmata полягає в тому, щоб використовувати машинне навчання та аналіз даних для оцінки достовірності та точності вмісту, що надходить до системи. Для досягнення цієї мети, Factmata використовує технології, такі як згорткові нейронні мережі та навчання з підкріпленням, що дозволяють залучити широкий спектр даних для класифікації та управління контентом в реальному часі. Результати класифікації можуть використовуватися для підтримки різноманітних цілей, включаючи боротьбу зі спамом, підробленням інформації, дезінформацією та фейками, що входять до системи [9].

2. OpenAI GPT: Одна з мовних моделей, розроблених OpenAI, GPT здатна генерувати дуже переконливі фейкові новини. OpenAI GPT (Generative Pre-trained Transformer) - це модель глибокого навчання, зокрема згорткової нейронної мережі, яка використовується для генерації тексту, перекладу, аналізу тексту і багатьох інших завдань в області обробки природних мов. Основний принцип роботи OpenAI GPT полягає в застосуванні техніки попереднього навчання на великих наборах текстових даних, що дозволяє моделі здатніше розуміти мову та генерувати більш якісний текст.

OpenAI GPT реалізований за допомогою трансформерної архітектури, яка забезпечує здатність моделі до оброблення довгих послідовностей та зводить до мінімуму проблему витікання градієнту. OpenAI GPT також використовує механізм самоуваги для визначення відповідних зв'язків між різними словами в тексті [10].

В результаті, OpenAI GPT може генерувати якісний текст на основі великих об'ємів даних, а також здійснювати інші операції з обробки природних мов, такі як переклад, аналіз тощо.

Однак її також можна використовувати для навчання моделей для виявлення фейкових новин.

3. Fakebox.org: Цей інструмент надає користувачам простий у використанні інтерфейс для подання статей, які, на їхню думку, є фейковими. Система використовує машинне навчання для аналізу контенту і виставляє оцінку, яка вказує на ймовірність того, що стаття є фейковою [11].

4. Fake News Detector: Розширення для Chrome, яке використовує обробку природної мови та машинне навчання для виявлення фейкових новин [12].

Загалом ці інструменти демонструють потенціал машинного навчання та нейронних мереж у виявленні фейкових новин. Однак важливо зазначити, що ці інструменти не є досконалими і можуть потребувати додаткового людського нагляду та перевірки.

Для виявлення фейкових новин можна використовувати згорткову нейронну мережу (CNN). Процес використання CNN для виявлення фейкових новин полягає в тому, що спочатку текст статті зображується у вигляді числового вектору, після чого застосовується згортка (англ. convolution), яка дозволяє визначити ключові ознаки тексту. Далі, отримані ознаки передаються в повнозв'язні шари, які здійснюють остаточний вивід статті як фейкової або правдивої.

Наприклад, у дослідженні [13] згорткову нейронну мережу було використано для виявлення фейкових новин на основі аналізу текстової інформації. У цьому дослідженні було використано базу даних з фейковими та правдивими новинами, після чого було навчено згорткову нейронну мережу на основі цих даних. Згодом, було проведено тестування на нових даних, та отримано досить високі результати точності виявлення фейкових новин.

У дослідженні [14], CNN навчалися виявляти фейкові новини, аналізуючи текстову інформацію новинних статей. Набір даних, використаний у цьому дослідженні, містив як фейкові, так і справжні новини. Архітектура CNN, розроблена для цього дослідження, складалася з декількох згорткових шарів, за кожним з яких слідував шар об'єднання, а потім серія повністю з'єднаних шарів для класифікації.

CNN навчали на маркованому наборі даних, а потім оцінювали на окремому тестовому наборі новинних статей, досягнувши високої точності в класифікації фейкових новинних статей. Загалом це дослідження показало, що CNN може бути ефективним інструментом для виявлення фейкових новин шляхом аналізу тексту новинних статей.

Аналіз рекурентних нейронних мереж (RNN) дуже важливий для виявлення фейків та аналізу послідовностей даних, таких як текстові повідомлення та відео. Якщо говорити про архітектуру RNN, то вона складається з нейронів, що мають зв'язки з попередніми станами, що дозволяє зберігати інформацію від попередніх шарів нейронної мережі.

Успіх RNN зумовлений їх здатністю досліджувати взаємозв'язки між елементами в послідовностях даних, що не завжди можуть бути виявлені з використанням звичайних методів машинного навчання.

Проте, у RNN є деякі недоліки, такі як проблема довгої залежності, яка може призвести до затухання градієнту. З цієї причини, деякі нові архітектури, такі як LSTM та GRU, були розроблені, щоб боротися з цією проблемою.

У загальному розумінні, RNN та їхні модифікації є потужними інструментами для аналізу послідовностей даних та виявлення фейків, але для кращих результатів їх

У загальному, RNN та їхні модифікації є потужними інструментами для аналізу послідовностей даних та виявлення фейків, але для кращих результатів їх ефективне використання потребує знань з глибинного навчання та оптимізації гіперпараметрів.

Також важливо враховувати якість та обсяг тренувальних даних для навчання класифікатора, так як на це значно впливає загальна ефективність системи.

Крім того, залежно від типу фейкової інформації, можуть бути кращі альтернативні архітектури нейронних мереж, які можуть краще пристосуватись до конкретного типу

даних. Наприклад, конволюційні нейронні мережі (CNN) можуть бути ефективними для обробки зображень та відео, а Transformer мережі можуть бути кращим рішенням для аналізу мови та тексту, типову модель нейромережі проілюстровано на рисунку 1.

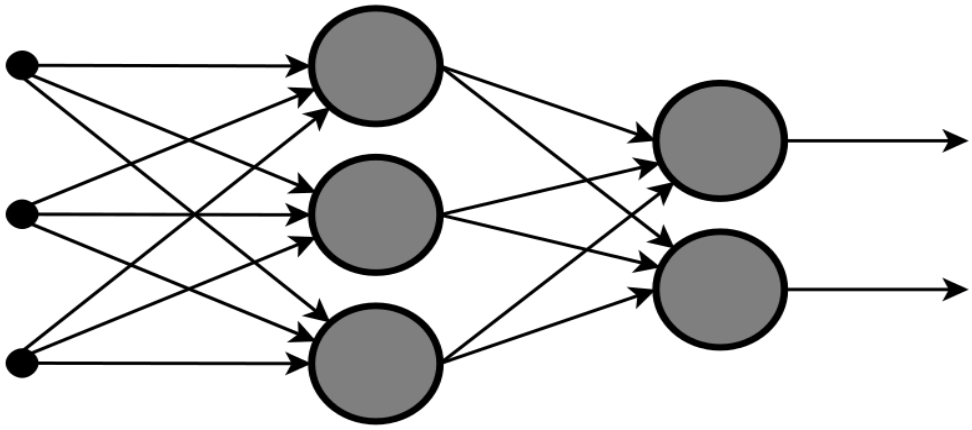

Рис.1. Модель архітектури рекурентних нейронних мереж

Узагальнюючи, RNN та їхні модифікації є потужними інструментами для розпізнавання та виявлення фейкової інформації, якщо вони використовуються правильно та з врахуванням особливостей вихідних даних.

Проте, для досягнення кращих результатів необхідно глибоке розуміння глибинного навчання та оптимізації гіперпараметрів. Крім того, для досягнення кращої ефективності необхідно мати належні та об'ємні навчальні дані. В залежності від типу фейкової інформації, можуть бути кращі альтернативні архітектури нейронних мереж, які можуть краще пристосуватись до конкретного типу даних.

Отже, використання RNN та їхніх модифікацій є важливим інструментом для виявлення та розпізнавання фейкової інформації, але для досягнення найкращих результатів необхідно враховувати їх обмеження та вибір альтернативних архітектур, крім того, для досягнення кращої ефективності, можуть використовуватися альтернативні навчальні методи та архітектури, такі як CNN і Transformer мережі.

Трансформерна нейронна мережа - це тип нейромережевої архітектури, який привернув увагу в останні роки завдяки своїм успіхам у вирішенні завдань обробки природної мови, таких як машинний переклад, моделювання мови та класифікація текстів. Трансформерна архітектура була представлена в статті 2017 року за авторством Васвані та ін., і являє собою відхід від традиційних архітектур рекурентних нейронних мереж (RNN) [15].

Однією з ключових особливостей архітектури Transformer є механізм самоуваги, який дозволяє моделі приділяти увагу різним позиціям у вхідній послідовності під час прогнозування. Цей механізм самоуваги дозволяє моделі обробляти довгострокові залежності більш ефективно, ніж традиційні RNN, і призвів до найкращих результатів у низці тестів з обробки мови.

Ще однією важливою особливістю архітектури Transformer є використання позиційних вбудовувань, які дозволяють моделі кодувати позицію кожної лексеми у вхідній послідовності. Це дозволяє моделі вивчати взаємозв'язки між різними токенами в послідовності і робити більш точні прогнози як показано на рисунку 2 [16].

ISSN 2663 - 4023

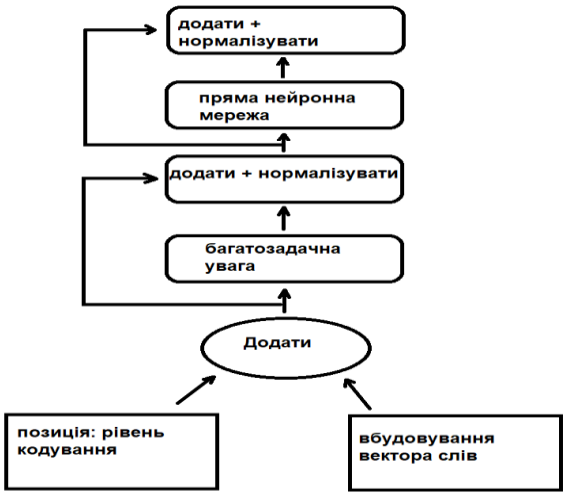

Рис.2. Модель Трасформерної нейронної мережі

Нейронна мережа Transformer виявилася потужним інструментом для обробки природної мови завдяки своїй здатності фіксувати складні взаємозв'язки між різними лексемами та ефективно обробляти довгострокові залежності.

Згорткові нейронні мережі (CNN) - це популярний тип нейронних мереж, які використовуються в аналізі зображень і відео. Вони працюють шляхом навчання фільтрів, які виділяють ознаки з вхідних зображень або відеокадрів. Потім фільтри застосовуються до ділянок вхідних даних, що перекриваються, створюючи карту ознак, яка виділяє цікаві для нас області на зображенні. Потім карти ознак проходять через додаткові шари, які виконують об'єднання, нормалізацію і нелінійні перетворення, в результаті чого створюється вектор ознак, який можна використовувати для класифікації або інших завдань.

CNN продемонстрували відмінну продуктивність у широкому спектрі завдань аналізу зображень, включаючи розпізнавання, сегментацію і виявлення об'єктів. Вони також успішно використовуються в обчислювальній біології, розпізнаванні мови та обробці природної мови.

Однією з ключових інновацій, що уможливила успіх CNN, є використання зворотного поширення для навчання мережі. Зворотне поширення - це алгоритм оптимізації на основі градієнта, який дозволяє мережі навчатися ваговим коефіцієнтам фільтрів шляхом багаторазових ітерацій на наборі даних з маркованими прикладами.

CNN є потужним інструментом для аналізу зображень і відкрили нові можливості для комп'ютерного зору та машинного навчання.

Загалом, CNN є потужним інструментом для аналізу зображень і відкрили нові можливості для комп'ютерного зору та машинного навчання.

Однією з ключових інновацій, що уможливила успіх CNN, є використання зворотного поширення для навчання мережі. Зворотне поширення - це алгоритм оптимізації на основі градієнта, який дозволяє мережі навчатися ваговим коефіцієнтам фільтрів шляхом багаторазових ітерацій на наборі даних з маркованими прикладами як продемонстровано на рисунку 3.

ISSN 2663 - 4023

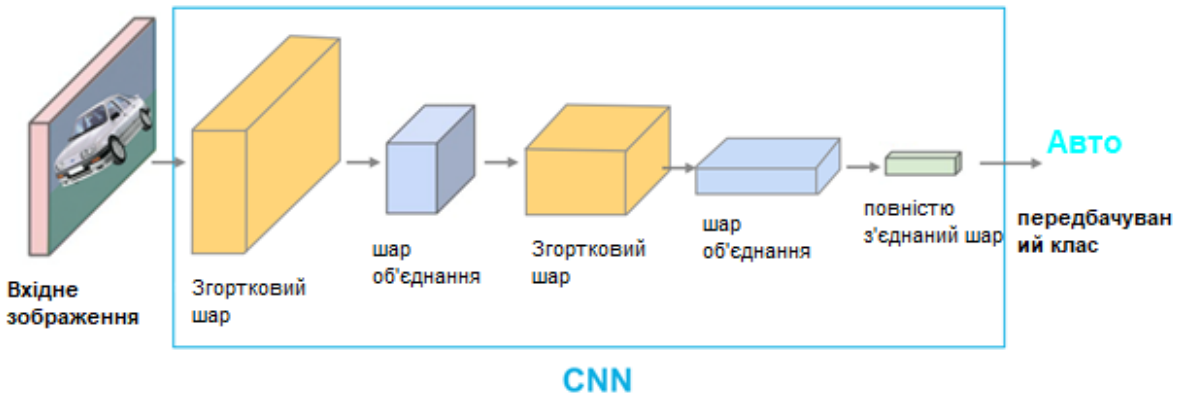

Рис.3. Модель архітектури згорткової нейромережі

Вцілому, CNN є потужним інструментом для аналізу зображень і відкрили нові можливості для комп'ютерного зору та машинного навчання.

Генеративні змагальні мережі (GAN) - це тип нейронних мереж, які складаються з двох моделей: моделі генератора, яка створює фейкові дані, і моделі дискримінатора, яка намагається відрізнити справжні дані від фейкових. GAN показали великі перспективи у створенні реалістичних зображень, але вони також досліджуються на предмет їхнього потенціалу у виявленні фальшивих новин і зображень.

Дослідження показали, що нейронні мережі на основі GAN можуть ідентифікувати фейкові зображення з високою точністю, і зараз докладаються зусилля для подальшого вдосконалення цих інструментів. Крім того, GAN досліджуються як потенційний інструмент для створення фейкових новин, щоб перевірити надійність систем виявлення.

Використання нейронних мереж на основі GAN для виявлення фейкових новин і зображень є перспективним напрямом, але для вдосконалення цих інструментів і підвищення їхньої точності необхідні подальші дослідження і розробки [17].

Архітектура генеративної змагальної мережі (GAN) складається з двох основних компонентів: генератора і дискримінатора.

Генератор приймає на вхід випадковий вектор шуму і генерує синтетичну вибірку даних. Мета генератора — створити вибірки, які є достатньо правдоподібними, щоб обдурити дискримінатор.

Дискримінатор приймає на вхід вибірку даних (реальну або синтетичну) і видає ймовірність того, що ця вибірка є реальною. Мета дискримінатора - правильно визначити, чи є дана вибірка реальною або синтетичною.

В процесі навчання генератор і дискримінатор навчаються в змагальній манері. Генератор намагається генерувати кращі зразки, щоб обдурити дискримінатор, в той час, як дискримінатор намагається краще відрізняти справжні зразки від синтетичних.

Загалом, архітектура GAN розроблена таким чином, щоб вивчати базовий розподіл даних і створювати реалістичні синтетичні зразки, які важко відрізнити від реальних зразків. Вона може бути модифікована і розширена різними способами, включаючи використання різних нейромережевих архітектур і функцій втрат, для покращення її продуктивності на конкретних завданнях, приклад моделі архітектури даної мережі зображено на рисунку 4.

ISSN 2663 - 4023

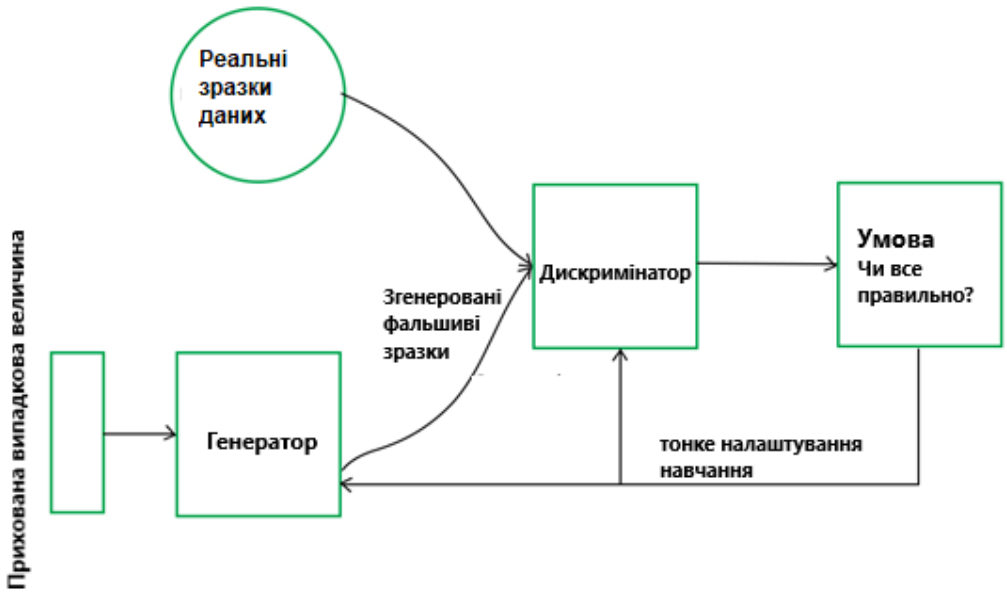

Рис.4. Модель архітектури Генеративної змагальної мережі

Слід зазначити, що ті самі принципи архітектури GAN можна застосувати до багатьох інших областей, окрім зображень. Наприклад, архітектуру GAN досліджували для генерування синтетичного тексту, мови, музики тощо.

DNN розшифровується як глибока нейронна мережа, яка є різновидом штучної нейронної мережі (ШНМ), що зазвичай має кілька шарів між вхідним і вихідним шарами.

Глибокі нейронні мережі є потужним інструментом для вирішення широкого спектра завдань машинного навчання, включаючи комп'ютерний зір, розпізнавання мови, обробку природної мови тощо. Вони здатні вивчати дуже складні та абстрактні представлення даних, і здатні досягти найсучаснішої продуктивності в широкому діапазоні завдань.

Розробка та реалізація глибоких нейронних мереж зазвичай містить ряд методів, включаючи зворотне поширення, регуляризацію, нормалізацію та методи оптимізації, такі як стохастичний градієнтний спуск. Існує також широкий спектр бібліотек і фреймворків з відкритим вихідним кодом для побудови, навчання і розгортання глибоких нейронних мереж. В цілому, глибокі нейронні мережі - це потужна область досліджень машинного навчання, що швидко розвивається, з широким спектром потенційних застосувань у багатьох різних сферах [18].

DNN (Deep Neural Network) — це тип штучної нейронної мережі, що складається з кількох шарів нейронів між вхідним та вихідним шаром. Кожен шар приймає значення від попереднього та обраховує вагову суму з активацією.

Архітектура DNN зазвичай складається з кількох повнозв’язних шарів з активацією, таких як ReLU або sigmoid. Кожен шар нейронів приймає значення від попереднього та обчислює вагову суму з активацією, передаючи результат наступному шару. Останній шар мережі відповідає за вихідні значення, наприклад, мітки класів в задачі класифікації, або значення регресії [19].

ISSN 2663 - 4023

Для навчання DNN зазвичай використовується алгоритм зворотного поширення помилок (backpropagation), який дозволяє оновлювати ваги мережі на основі різниці між очікуваним та фактичним виходом мережі, як зображено на рисунку 5.

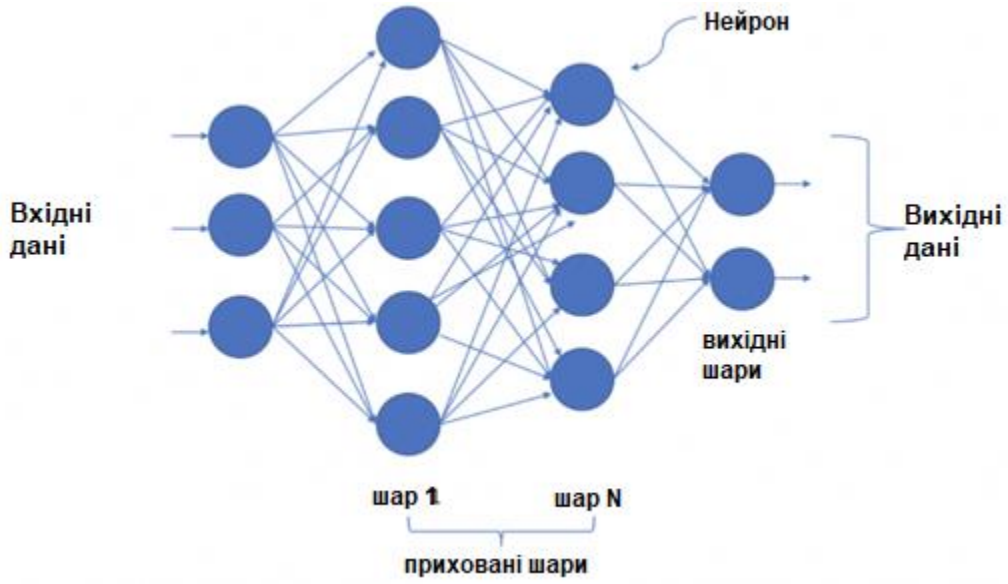

Рис.5. Модель архітектури глибокої нейронної мережі

Аналіз доступної літератури демонструє, що нейронні мережі, зокрема згорткові нейронні мережі (CNN), показали добрі результати у виявленні фейкових новин. Проте розвиток таких алгоритмів потребує використання даних з мітками, що часто є обмеженим або важко доступним. Тому необхідно провести подальше дослідження для розробки нових підходів для навчання нейронних мереж з використанням нагляду без учителів або слабкого нагляду. Окрім того, необхідно більш тісно співпрацювати між дослідниками та соціальними медіа-платформами з метою розробки ефективних інструментів виявлення фейкових новин, які можна було б впровадити в широкому масштабі.

Для виявлення фальшивих медіа, таких як зображення та відео, використовуються різні методи навчання та мережеві інструменти. Одним із популярних підходів є використання моделей машинного і глибокого навчання, таких як згорткові нейронні мережі (CNN) і рекурентні нейронні мережі (RNN), для вилучення ознак і виявлення закономірностей у справжніх і фальшивих медіа. Методи змагального навчання, такі як генеративні змагальні мережі (GAN), також можуть бути використані для виявлення фейків.

На основі аналізу методів навчання та характеристик нейромереж, можна виділити наступні переваги та недоліки:

Переваги:

- Нейронні мережі можуть обробляти великі обсяги даних, що робить їх добре придатними для аналізу великих обсягів тексту, зображень і відео.
- Нейронні мережі можуть автоматично вивчати складні патерни та особливості, що робить їх ефективними для виявлення тонких невідповідностей та аномалій у фейкових медіа.

- Нейронні мережі можна адаптувати і тонко налаштовувати під конкретні завдання і домени, що забезпечує високу точність і специфічність у виявленні фейкових медіа.

Недоліки:

- Нейронні мережі потребують великих обсягів високоякісних даних для навчання, які не завжди доступні або можуть бути дорогими для придбання.
- Нейронні мережі можуть бути вразливими до ворожих атак, коли зловмисники маніпулюють вхідними даними, щоб змусити мережу видавати неправильні результати.
- Нейронні мережі можуть бути дорогими в обчислювальному плані, вимагаючи значних обчислювальних потужностей і часу для навчання та запуску.

Підсумовуючи, ефективність нейронних мереж у виявленні фейкових медіа залежить від різних факторів, таких як якість і кількість даних, конкретний тип медіа, що аналізуються, а також архітектура нейронної мережі та методи навчання, що використовуються. Крім того, слід пам'ятати, що не існує єдиного рішення, яке може остаточно ідентифікувати всі фальшиві медіа, і для точної та надійної роботи може знадобитися використання декількох підходів і методів.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У статті проаналізовано методи навчання та інструментів нейромереж для виявлення фейків на основі таких критеріїв, як ефективність, аналіз точності, швидкість виявлення, здатність виявляти фото, відео, аудіо контент та архітектура нейромереж і її вплив на виявлення фейків. В останні роки проблема фейкових новин стала однією з найбільш актуальних у суспільстві. У зв’язку з цим, розробка надійних методів виявлення фейків, зокрема з використанням нейромереж, стає дедалі більш важливою задачею. Однак, існує безліч різних методів навчання та інструментів нейромереж для виявлення фейків, тому важливо дослідити їх ефективність та можливості.

Аналіз свідчить, що моделі машинного та глибинного навчання, а також методи змагального навчання та спеціальні інструменти для виявлення фейкових медіа є ефективними у виявленні фейків. Однак на ефективність і точність цих методів та інструментів можуть впливати такі фактори, як якість даних, методи, що використовуються для навчання та оцінювання, а також складність фейкових медіа, які виявляються. Постійні дослідження і розробки в цій галузі мають вирішальне значення для підвищення точності та надійності цих методів та інструментів для виявлення фейків.

Тому необхідно провести подальше дослідження для розробки нових підходів для навчання нейронних мереж з використанням нагляду без учителів або слабкого нагляду. Окрім того, необхідно більш тісно співпрацювати між дослідниками та соціальними медіа-платформами з метою розробки ефективних інструментів виявлення фейкових новин, які можна було б впровадити в широкому масштабі.

Надалі планується продовжити дослідження проблем поширення фейкових новин у напрямку аналізу рішень для фільтрації та виявлення фейкових новин, розробки алгоритму на основі нейромережі та перевірки новин на справжність за його допомогою.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1 Vosoughi, S., Roy, D., Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

2 Kulsoom, F., Narejo, S., Mehmood, Z., Chaudhry, H. N., Butt, A., Bashir, A. K. (2022). A review of machine learning-based human activity recognition for diverse applications. Neural Computing and Applications, 34(21), 18289–18324. https://doi.org/10.1007/s00521-022-07665-9

3 Singh, B., Sharma, D. K. (2021). Predicting image credibility in fake news over social media using multi-modal approach. Neural Computing and Applications, 34(24), 21503–21517. https://doi.org/10.1007/s00521-021-06086-4

4 Zhou, X., Zafarani, R. (2020). A Survey of Fake News. ACM Computing Surveys, 53(5), 1–40. https://doi.org/10.1145/3395046

5 Vosoughi, S., Mohsenvand, M. N., Roy, D. (2017). Rumor Gauge. ACM Transactions on Knowledge Discovery From Data, 11(4), 1–36. https://doi.org/10.1145/3070644

6 Chen, W., Zhang, Y., Yeo, C. K., Lau, C. T., Lee, B. S. (2018). Unsupervised rumor detection based on users' behaviors using neural networks. Pattern Recognition Letters, 105, 226–233. https://doi.org/10.1016/j.patrec.2017.10.014

7 O'Brien, N., Latessa, S., Evangelopoulos, G., Boix, X. (2018). The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors.

8 Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. Machine Learning With Applications, 4, 100032. https://doi.org/10.1016/j.mlwa.2021.100032

9 About Us.. Factmata. https://factmata.com/about-us/

10 Introducing ChatGPT.. Introducing ChatGPT. https://openai.com/blog/chatgpt

11 Inc., V. Fakebox · Docs · Machine Box · Machine learning in a box. Fakebox · Docs · Machine Box · Machine Learning in a Box. https://machinebox.io/

12 Falcone, J., & bio, S. F. (2023). Looking for Great Deals? Use CNET Shopping to Save Time and Money. CNET. https://www.cnet.com/tech/services-and-software/use-cnet-shopping-to-seek-out-the-best-deals/

13 Khanam, Z., Alwasel, B. N., Sirafi, H., Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. IOP Conference Series: Materials Science and Engineering, 1099(1), 012040. https://doi.org/10.1088/1757-899x/1099/1/012040

14 Shad, H. S., Rizvee, M. M., Roza, N. T., Hoq, S. M. A., Monirujjaman Khan, M., Singh, A., Zaguia, A., Bourouis, S. (2021). Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network. Computational Intelligence and Neuroscience, 2021, 1–18. https://doi.org/10.1155/2021/3111676

15 Transformer neural networks are shaking up AI | TechTarget. Enterprise AI. https://www.techtarget.com/searchenterpriseai/feature/Transformer-neural-networks-are-shaking-up-AI

16 Transformer Neural Network. (2020). DeepAI. https://deepai.org/machine-learning-glossary-and-terms/transformer-neural-network

17 Rocca, J. (2021). Understanding Generative Adversarial Networks (GANs). Medium. https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29

18 Deep Neural Networks. Deep Neural Networks. https://www.tutorialspoint.com/python_deep_learning/python_deep_learning_deep_neural_networks.htm

19 Patwari, K., Hafiz, S. M., Wang, H., Homayoun, H., Shafiq, Z., Chuah, C. N. (2022). DNN Model Architecture Fingerprinting Attack on CPU-GPU Edge Devices. 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P). https://doi.org/10.1109/eurosp53844.2022.00029

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 4 (20), 2023
ISSN 2663 - 4023

Vitalii S. Tyshchenko
Assistant of Information Security and Cyber Security Department
State University of Telecommunications, Kyiv, Ukraine
ORCID ID: 0000-0003-3849-6243
tvs5vetal@gmail.com

ANALYSIS OF TRAINING METHODS AND NEURAL NETWORK TOOLS FOR FAKE NEWS DETECTION

Abstract. This article analyses various training methods and neural network tools for fake news detection. Approaches to fake news detection based on textual, visual and mixed data are considered, as well as the use of different types of neural networks, such as recurrent neural networks, convolutional neural networks, deep neural networks, generative adversarial networks and others. Also considered are supervised and unsupervised learning methods such as autoencoding neural networks and deep variational autoencoding neural networks.
Based on the analysed studies, attention is drawn to the problems associated with limitations in the volume and quality of data, as well as the lack of efficiency of tools for detecting complex types of fakes. The author analyses neural network-based applications and tools and draws conclusions about their effectiveness and suitability for different types of data and fake detection tasks.
The study found that machine and deep learning models, as well as adversarial learning methods and special tools for detecting fake media, are effective in detecting fakes. However, the effectiveness and accuracy of these methods and tools can be affected by factors such as data quality, methods used for training and evaluation, and the complexity of the fake media being detected. Based on the analysis of training methods and neural network characteristics, the advantages and disadvantages of fake news detection are identified. Ongoing research and development in this area is crucial to improve the accuracy and reliability of these methods and tools for fake news detection.

Keywords: fake news; fake news detection tools; neural networks; learning methods; methods for detecting disinformation and fake news on the Internet.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1 Vosoughi, S., Roy, D., Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146–1151. https://doi.org/10.1126/science.aap9559
2 Kulsoom, F., Narejo, S., Mehmood, Z., Chaudhry, H. N., Butt, A., Bashir, A. K. (2022). A review of machine learning-based human activity recognition for diverse applications. Neural Computing and Applications, 34(21), 18289–18324. https://doi.org/10.1007/s00521-022-07665-9
3 Singh, B., Sharma, D. K. (2021). Predicting image credibility in fake news over social media using multi-modal approach. Neural Computing and Applications, 34(24), 21503–21517. https://doi.org/10.1007/s00521-021-06086-4
4 Zhou, X., Zafarani, R. (2020). A Survey of Fake News. ACM Computing Surveys, 53(5), 1–40. https://doi.org/10.1145/3395046
5 Vosoughi, S., Mohsenvand, M. N., Roy, D. (2017). Rumor Gauge. ACM Transactions on Knowledge Discovery From Data, 11(4), 1–36. https://doi.org/10.1145/3070644
6 Chen, W., Zhang, Y., Yeo, C. K., Lau, C. T., Lee, B. S. (2018). Unsupervised rumor detection based on users' behaviors using neural networks. Pattern Recognition Letters, 105, 226–233. https://doi.org/10.1016/j.patrec.2017.10.014
7 O'Brien, N., Latessa, S., Evangelopoulos, G., Boix, X. (2018). The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors.
8 Khan, J. Y., Khondaker, M. T. I., Afroz, S., Uddin, G., Iqbal, A. (2021). A benchmark study of machine learning models for online fake news detection. Machine Learning With Applications, 4, 100032. https://doi.org/10.1016/j.mlwa.2021.100032
9 About Us.. Factmata. https://factmata.com/about-us/
10 Introducing ChatGPT.. Introducing ChatGPT. https://openai.com/blog/chatgpt
11 Inc., V. Fakebox · Docs · Machine Box · Machine learning in a box. Fakebox · Docs · Machine Box · Machine Learning in a Box. https://machinebox.io/

12 Falcone, J., & bio, S. F. (2023). Looking for Great Deals? Use CNET Shopping to Save Time and Money. CNET. https://www.cnet.com/tech/services-and-software/use-cnet-shopping-to-seek-out-the-best-deals/

13 Khanam, Z., Alwasel, B. N., Sirafi, H., Rashid, M. (2021). Fake News Detection Using Machine Learning Approaches. IOP Conference Series: Materials Science and Engineering, 1099(1), 012040. https://doi.org/10.1088/1757-899x/1099/1/012040

14 Shad, H. S., Rizvee, M. M., Roza, N. T., Hoq, S. M. A., Monirujjaman Khan, M., Singh, A., Zaguia, A., Bourouis, S. (2021). Comparative Analysis of Deepfake Image Detection Method Using Convolutional Neural Network. Computational Intelligence and Neuroscience, 2021, 1–18. https://doi.org/10.1155/2021/3111676

15 Transformer neural networks are shaking up AI | TechTarget. Enterprise AI. https://www.techtarget.com/searchenterpriseai/feature/Transformer-neural-networks-are-shaking-up-AI

16 Transformer Neural Network. (2020). DeepAI. https://deepai.org/machine-learning-glossary-and-terms/transformer-neural-network

17 Rocca, J. (2021). Understanding Generative Adversarial Networks (GANs). Medium. https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29

18 Deep Neural Networks. Deep Neural Networks. https://www.tutorialspoint.com/python_deep_learning/python_deep_learning_deep_neural_networks.htm

19 Patwari, K., Hafiz, S. M., Wang, H., Homayoun, H., Shafiq, Z., Chuah, C. N. (2022). DNN Model Architecture Fingerprinting Attack on CPU-GPU Edge Devices. 2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P). https://doi.org/10.1109/eurosp53844.2022.00029