Вінницький національний технічний університет
Факультет менеджменту та інформаційної безпеки
Кафедра менеджменту та безпеки інформаційних систем

МАГІСТЕРСЬКА КВАЛІФІКАЦІЙНА РОБОТА

на тему:
«Вдосконалення методу виявлення маніпуляцій у відеофайлах з використанням TimeSformer і глибоких згорткових мереж»

Виконала: здобувач 2-го курсу, групи
КІТС-23мз спеціальності 125–
Кібербезпека та захист інформації
Освітня програма – Кібербезпека
інформаційних технологій та систем
(шифр і назва напряму підготовки, спеціальності)

Школьнікова В. В.
(прізвище та ініціали)

Керівник: к.т.н., доц. каф. МБІС
Грицак А. В.
(прізвище та ініціали)
«_____» _____2025 р.

Опонент: к.т.н., доцент, доцент каф. ОТ
Войцеховська О.В.
(прізвище та ініціали)
«_____» _____2025 р.

Допущено до захисту
Голова секції УБ кафедри МБІС

_____Юрій ЯРЕМЧУК
"___" _____2025 р.

Вінниця ВНТУ - 2025 рік

Вінницький національний технічний університет
Факультет менеджменту та інформаційної безпеки
Кафедра менеджменту та безпеки інформаційних систем

Рівень вищої освіти ІІ-й (магістерський)
Галузь знань 12 – Інформаційні технології
Спеціальність 125 – Кібербезпека та захист інформації
Освітньо-професійна програма – Кібербезпека інформаційних технологій та системи

ЗАТВЕРДЖУЮ
Голова секції УБ, кафедра МБІС
_____ **Юрій ЯРЕМЧУК**
“____” _____ 2025 р.

З А В Д А Н Н Я
на магістерську кваліфікаційну роботу здобувача
Школьнікової Вероніки Володимирівни
(прізвище, ім`я, по-батькові)

1. Тема роботи:
«Вдосконалення методу виявлення маніпуляцій у відеофайлах з використанням TimeSformer і глибоких згорткових мереж»
Керівник роботи: д.т.н., доц. каф. МБІС МБІС Грицак А. В.
(прізвище, ім`я, по-батькові, науковий ступінь, вчене звання)
затверджені наказом вищого навчального закладу від “20“ березня 2025 року №96
2. Строк подання студентом роботи за тиждень до захисту.
3. Вихідні дані до роботи:
Стандарти, електронні джерела, підручники та наукові статті по темі, які стосуються теми магістерської кваліфікаційної роботи.
4. Зміст розрахунково-пояснювальної записки:
Для досягнення мети необхідно: В першому розділі дослідити існуючі методи перевірки автентичності відео, їхні переваги та недоліки. В другому розділі розробити вдосконалений метод перевірки автентичності відео з використанням трансформерів і глибоких згорткових мереж, сформувати алгоритм. В третьому розділі створити програмний засіб для реалізації методу, протестувати та оцінити його ефективність.

5. Перелік ілюстративного матеріалу (з точним зазначенням обов'язкових кресл
У першому розділі магістерської кваліфікаційної роботи наведено 6 рис.; у друг
розділі наведено 3 рис. та 1 табл.; у третьому розділі наведено 2 рис. та 2 табл
четвертому розділі наведено 6 табл.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Основна частина			
I	Грицак А.В. д.т.н., доц. каф. МБІС	20.03.25	20.03.25
II	Грицак А.В. д.т.н., доц. каф. МБІС	20.03.25	20.03.25
III	Грицак А.В. д.т.н., доц. каф. МБІС	20.03.25	20.03.25
Економічна частина			
IV	Лесько О.Й., завідувач кафедри ЕПВМ, к.е.н., професор	20.03.25	10.06.25

7. Дата видачі завдання 20 березня 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів магістерської кваліфікаційної роботи	Строк виконання етапів роботи		Примітка
		початок	закінчення	
1	Аналіз предметної області обраної теми	20.03.2025	20.03.2025	
2	Розробка алгоритму роботи	24.03.2025	24.04.2025	
3	Написання магістерської кваліфікаційної роботи на основі розробленої теми	28.04.2025	19.05.2025	
4	Передзахист магістерської кваліфікаційної роботи	26.05.2025	26.05.2025	
5	Виправлення, уточнення, корегування магістерської кваліфікаційної роботи	28.05.2025	05.06.2025	
6	Захист магістерської кваліфікаційної роботи	13.06.2025	13.06.2025	

Здобувач (підпис) Школьнікова В.В.

Керівник роботи (підпис) Грицак А.В.

АНОТАЦІЯ

УДК 621.374.415

Школьнікова В.В. Магістерська кваліфікаційна робота зі спеціальності 125 – «Кібербезпека та захист інформації», освітня програма «Кібербезпека інформаційних технологій та систем». Вінниця: ВНТУ, 2025. – 105 с.

Укр. мовою. Бібліогр.: 42 назв; рис.: 30; табл. 10.

Ключові слова: вдосконалення, метод, виявлення, маніпуляції, відео, TimeSformer, згорткові нейронні мережі.

Магістерська робота присвячена вдосконаленню метода виявлення маніпуляцій у відеофайлах з використанням TimeSformer і глибоких згорткових мереж.

У роботі досліджено та проаналізовано існуючі методи виявлення маніпуляцій у відеофайлах, визначено їхні переваги та недоліки.

Детально розглянуто та описано напрямки вдосконалення методів детекції маніпуляцій із використанням TimeSformer і глибоких згорткових нейронних мереж, сформовано алгоритм роботи запропонованої моделі.

Обґрунтовано вибір програмного середовища та реалізовано вдосконалений метод виявлення маніпуляцій у відео.

Наукова новизна роботи полягає у запропонуванні вдосконалення методу виявлення маніпуляцій у відеофайлах, який базується на поєднанні глибоких згорткових нейронних мереж (CNN) та трансформерної архітектури TimeSformer. Такий підхід дозволяє одночасно аналізувати просторові та часові ознаки відеопослідовностей, що забезпечує вищу точність виявлення підробок, зокрема створених за допомогою технологій deepfake.

Результатом роботи є підтвердження ефективності запропонованого підходу, що відкриває можливості для подальшого розвитку та розширення функціоналу методу.

ABSTRACT

UDC 621.374.415

Shkolnikova V.V. Master's qualification work in specialty 125 – “Cybersecurity and Information Protection”, educational program "Cybersecurity of Information Technologies and systems". Vinnytsia: VNTU, 2025. - 105 p.

In Ukrainian. Bibliographer: 42 titles; fig.: 30; tabl. 10

Keywords: improvement, method, detection, manipulation, video, TimeSformer, convolutional neural networks.

The master's thesis is devoted to the improvement of the method of detecting manipulations in video files using TimeSformer and deep convolutional networks.

The work investigates and analyzes existing methods of detecting manipulations in video files, identifies their advantages and disadvantages.

The directions of improving the methods of detecting manipulations using TimeSformer and deep convolutional neural networks are considered in detail and described, the algorithm of the proposed model is formed.

The choice of the software environment is justified and an improved method of detecting manipulations in video is implemented.

The scientific novelty of the work lies in proposing an improvement of the method of detecting manipulations in video files, which is based on a combination of deep convolutional neural networks (CNN) and the TimeSformer transformer architecture. This approach allows for the simultaneous analysis of spatial and temporal features of video sequences, which provides higher accuracy in detecting fakes, in particular those created using deepfake technologies.

The result of the work is confirmation of the effectiveness of the proposed approach, which opens up opportunities for further development and expansion of the functionality of the method.

ЗМІСТ

ВСТУП

У сучасному світі цифрові інструменти стали стрижнем спілкування, оперативного обміну відомостями та накопичення знань. Водночас ми фіксуємо помітне збільшення кількості підроблених відео, сфабрикованих з використанням deepfake та подібних технологій цифрового впливу. Це породжує реальні ризики для безпеки громадян, достовірності інформації, стійкості суспільства та реалізації прав кожної особи.

Актуальність.

Фальсифікація відеозаписів розповсюдилася широко у кіберзлочинності, політичних змаганнях, мас-медіа та соціальних мережах, породжуючи потребу в розробці ефективних автоматизованих методів перевірки справжності відео. Наявні алгоритми аналізу, що опираються на статистичні дані та ручні способи виявлення підробок, виявляють суттєві вади. Зокрема, це низька адаптивність до сучасних методів обробки зображень, підвищена чутливість до змін у файлах відео та значне використання обчислювальних ресурсів.

Останнім часом спостерігається значний інтерес до методик глибинного навчання, зокрема до глибоких згорткових нейронних мереж та трансформерів. Ці моделі демонструють дивовижні показники у виявленні відеофейків, що зумовлено їхньою здатністю автоматично знаходити складні патерни та аномалії у відео. Завдяки цьому вони є надзвичайно ефективним інструментом для автоматизованої перевірки відео на достовірність.

Отже, вдосконалення автоматизованих систем для верифікації відео за допомогою передових методів штучного інтелекту – це нагальне наукове та практичне питання. Його вирішення сприятиме зміцненню інформаційної безпеки та ефективній протидії дезінформації.

Мета і задачі дослідження.

Метою і задачею роботи є удосконалення автоматизованих систем перевірки відео на справжність шляхом використання трансформерів і глибоких згорткових

нейронних мереж, створення програмного засобу для реалізації запропонованого методу та оцінки його ефективності.

Для досягнення цієї мети були поставлені та вирішені наступні завдання:

- Дослідити існуючі методи перевірки автентичності відео, їхні переваги та недоліки;
- Розробити вдосконалений метод виявлення фальсифікацій у відео з використанням трансформерів і згорткових нейронних мереж;
- Реалізувати програмний засіб для автоматизованої перевірки відео на справжність;
- Провести тестування запропонованого методу та оцінити його ефективність;
- Виконати економічне обґрунтування доцільності впровадження розробленої системи.

Об’єкт дослідження.

Методи та алгоритми перевірки відео на справжність, що використовуються в автоматизованих системах.

Предмет дослідження.

Використання трансформерів і глибоких згорткових нейронних мереж для автоматизованого аналізу та перевірки автентичності відеоконтенту.

Новизна роботи.

Вдосконалений метод перевірки відео на справжність, що поєднує трансформери та згорткові нейронні мережі, забезпечуючи високу точність виявлення фальсифікацій.

Практична цінність одержаних результатів.

Розроблений програмний засіб може бути інтегрований у медіаплатформи, соціальні мережі, державні та корпоративні системи безпеки для автоматизованої перевірки відео. Його застосування є актуальним у сфері кібербезпеки, журналістики, криміналістики та банківського сектору, де необхідно оперативно ідентифікувати підроблені відеоматеріали. Крім того, використання запропонованого підходу дозволяє зменшити витрати на ручну перевірку відео та підвищити швидкість аналізу контенту. Це сприяє зниженню ризиків, пов’язаних із поширенням дезінформації,

шахрайством та кіберзагрозами. Таким чином, результати роботи можуть бути ефективно використані для забезпечення інформаційної безпеки та боротьби з цифровими фальсифікаціями.

1 АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ ПЕРЕВІРКИ АВТЕНТИЧНОСТІ ВІДЕО

У даному розділі ми зосередимось на теоретичних основах, які дозволяють перевіряти автентичність відеофайлів, а також методах, що допомагають виявляти маніпуляції з відеоконтентом. Буде проаналізовано традиційні способи аналізу відео, оцінені їхні плюси та мінуси, особливо в реаліях сучасного інформаційного простору. Окремо розглянемо сучасні технології, зокрема, методи на базі глибокого навчання, такі як згорткові нейронні мережі (CNN) та трансформери, що показали високу ефективність у виявленні фальсифікацій. У цьому розділі також буде проведено аналіз ключових різновидів загроз, породжених відеоманіпуляціями та вивчено специфіку їхнього виявлення.

1.1 Загальна характеристика методів перевірки відео на справжність

У сучасному цифровому світі актуальність перевірки автентичності відеоматеріалів сьогодні надзвичайно висока. Розвиток технологій глибинного навчання дозволив суттєво спростити процес створення підроблених відео, таких як deepfake, що становить серйозну загрозу для медіапростору, політичного ландшафту, кібербезпеки та криміналістики. Методи верифікації відеозаписів поділяються на три ключові категорії: дослідження метаданих, фізичний аналіз та аналіз цифрових артефактів.

Перевірка автентичності відео – це процедура визначення, чи були внесені зміни до відеоматеріалу або ж він був створений з наміром фальсифікації. З розвитком передових технологій монтажу та використанням штучного інтелекту (ШІ), ці методи стають критичними для виявлення поширення неправдивої інформації.

Спершу проведемо аналіз метаданих. Метадані – це службова інформація, що зберігається в середині відеофайлу. Вони містять відомості про пристрій, на котрий було здійснено запис, час та дату створення відео, використаний формат стиснення та технічні коди. Як приклад, розглянемо відео, в якому стверджується, що його було відзнято на iPhone 13, а метадані вказують обробку на комп’ютері через професійний

редактор. Це може слугувати показником маніпуляції. Аналіз метаданих надає змогу виявити розбіжності, що можуть бути ознакою фальсифікації відео.

Фізичний аналіз ґрунтується на дослідженні візуальних і звукових складових відео для виявлення аномалій. Головні аспекти такого аналізу передбачають оцінку освітлення та тіней. У підроблених відео тіні можуть бути направлені неправильно або мати непередбачувану інтенсивність. Дослідження геометрії та перспективи дозволяють зрозуміти, що певні маніпуляції можуть призвести до викривлення предметів. Далі йде вивчення динаміки рухів обличчя та губ. У deepfake-відео рухи губ можуть не відповідати вимовленим словам. До прикладу, у відредагованому відео може бути помітно, що очі людини не кліпають природнім чином або голос не відповідає рухам губ.

Сучасні методи аналізу цифрових слідів базуються на використанні штучного інтелекту (ШІ) та машинного навчання для виявлення аномалій у відео. Основними підходами є глибокі згорткові нейронні мережі (CNN), трансформери (Transformers) та аналіз артефактів стиснення.

Глибокі згорткові нейронні мережі (CNN) – дозволяють аналізувати окремі кадри та виявляти невидимі для людського ока сліди редагування. Розповідаючи про трансформери (Transformers) – використовуються для аналізу послідовності кадрів і виявлення несумісностей між рухами об’єктів. Аналіз артефактів стиснення – при зміні відео зображення може містити аномальні сліди повторного стиснення. До прикладу, алгоритм може виявити, що текстура шкіри на окремих кадрах відрізняється, що є типовою ознакою deepfake.

Методи перевірки відео на справжність відіграють важливу роль в умовах зростаючої кількості фальсифікованих відеоматеріалів, зокрема deepfake. Основні підходи до верифікації відео можна поділити на традиційні та сучасні методи, засновані на штучному інтелекті. Одним із ключових методів є аналіз метаданих відеофайлів, який дозволяє отримати інформацію про пристрій зйомки, формат, часові позначки та інші характеристики. Аналіз змін у структурі метаданих після редагування або виявлення невідповідностей у форматах і кодах відео може свідчити про підробку. Ще одним підходом є аналіз послідовності кадрів, який досліджує

закономірності між кадрами відео та шукає аномалії, що можуть вказувати на підробку. Це включає виявлення раптових змін у послідовності кадрів, аналіз оптичного потоку та порівняння колірної гами й освітлення між кадрами.

Сучасні підходи машинного навчання, особливо глибокі згорткові мережі та трансформери, дають змогу автоматично виявляти фальсифікації шляхом аналізу візуальних особливостей. Глибокі згорткові мережі застосовуються для виявлення слідів редагування, розбіжностей у текстурі шкіри та аномалій на обличчі, а трансформери допомагають моделювати довготривалі взаємозв'язки між кадрами та розпізнавати нелінійності у рухах обличчя, що можуть свідчити про штучну генерацію відео. Комбінація згорткових мереж та трансформерів покращує точність детекції та зменшує кількість помилково позитивних результатів.

Аналіз звуку є ще одним важливим методом перевірки відео, оскільки синтетичний звук може містити невідповідності з відеорядом. Дослідження звуку включає виявлення асинхронності між звуком і рухом губ, аналіз акустичних характеристик голосу, які можуть вказувати на штучне синтезування, а також порівняння фонових шумів і резонансів із базами справжніх записів. Форензика цифрового відео також є ефективним інструментом, оскільки дозволяє проводити глибокий аналіз структурних характеристик відеофайлу, таких як зміни у кодексу, стискання відео, виявлення вставлених чи відредагованих фрагментів через аналіз гістограм освітлення та контрасту. Використання статистичних методів допомагає виявити артефакти маніпуляції. Найкращі результати досягаються при поєднанні кількох методів аналізу, адже використання одночасно аналізу кадрів, звуку та метаданих значно підвищує точність виявлення фальсифікацій. Багатофакторний підхід активно застосовується у сучасних автоматизованих системах детекції підроблених відео.

Сучасні методи відео-маніпулювання, зокрема генеративні моделі та глибокі нейронні мережі, неабияк ускладнюють процес їх виявлення. На цьому етапі, зосереджуючись на основних викликах у виявленні маніпуляцій у відеоматеріалах, розглянемо ключові труднощі, з якими стикаються дослідники та розробники систем детекції маніпуляцій у відео.

1. Висока якість та реалістичність підробок

Раніше цифрові маніпуляції можна було виявити за очевидними артефактами, такими як розмиті краї обличчя, аномальні рухи губ або нереалістичні тіні. Проте сучасні методи, зокрема Deepfake, GAN (генеративно-змагальні мережі), StyleGAN та інші глибокі нейромережі, дозволяють створювати підробки, які практично неможливо відрізнити від реального відео без спеціальних алгоритмів аналізу. Причинами високої якості підробок є те, що генеративні моделі навчаються на великих наборах відео, що дозволяє їм відтворювати точні особливості міміки та рухів. Використання алгоритмів post-processing (пост-обробки), таких як згладжування артефактів, синхронізація руху обличчя з голосом, покращення освітлення та текстури, робить маніпуляції майже невидимими. Також, висока роздільна здатність відео (4К та 8К) ускладнює пошук змін. Оскільки дрібні деталі добре передаються без помітних спотворень.

Високоякісне фото, що ілюструє високу якість та реалістичність AI-генерованих підробок. На зображенні чітко виражені дрібні артефакти біля очей та рота, що можуть видавати маніпуляцію, але загальна картинка виглядає майже бездоганно, що ускладнює детекцію.

Отже, методи детекції повинні орієнтуватися не лише на аналіз окремих кадрів, а й на виявлення часових аномалій, оскільки навіть найкращі генеративні моделі можуть давати несинхронізовані зміни в динаміці відео.

2. Різноманітність типів маніпуляцій

Маніпуляції у відео можуть бути різного характеру, що ускладнює розробку універсального алгоритму для їхнього виявлення. Основні типи маніпуляцій включають:

- Face Swapping (заміна обличчя) – алгоритм замінює обличчя людини в відео іншим, зберігаючи при цьому основні риси руху та міміки;
- Face Reenactment (заміна міміки) – методика дозволяє змінювати вираз обличчя, до прикладу, змушувати людину говорити слова, яких вона на справді не вимовляла;

- Inpainting (заповнення та редагування відео) – заміна певних об’єктів або деталей у відео. Наприклад, алгоритми можуть видаляти або додавати людей у відео, змінювати фон, підроблювати елементи сцени.
- Frame insertion/removal (вставка або видалення кадрів) – використовується для створення неправдивої хронології або подій для приховування важливих деталей.

Оскільки ці маніпуляції мають різні візуальні та часові характеристики, виявлення їх за допомогою одного алгоритму є складним завданням. Для ефективного виявлення маніпуляцій у відеофайлах, зокрема Deepfake або інших форм цифрових підробок, необхідно використовувати комбінований підхід, що аналізує як просторові (статичні) характеристики відео, так і часові залежності між кадрами. Запропонований алгоритм включає декілька основних етапів, які дозволяють виявити навіть добре замасковані маніпуляції.

3. Вплив відеокодування та стиснення

Після створення відео воно часто проходить процес кодування та стиснення, особливо під час завантаження в соціальні мережі або на відеохостинг-платформи (YouTube, TikTok, Instagram). Це створює додаткові труднощі для виявлення маніпуляцій, оскільки алгоритми стиснення можуть змінювати ключові візуальні особливості, які використовуються детекторами.

Як стиснення впливає на детекцію:

- Видаляються високочастотні деталі, що ускладнює аналіз текстури шкіри та мікродеталей;
- З’являються блокові артефакти, що можуть маскувати реальні аномалії у відео;
- Втрачається інформація про колірні градієнти, що може приховати неузгодженість освітлення між справжнім і зміненим обличчям;
- Алгоритми стискання можуть змінювати часову послідовність кадрів, що ускладнює аналіз руху.

Через це, методи виявлення маніпуляцій повинні бути стійкими до змін, спричинених стисненням та повторним кодуванням відео. Одним із підходів є попереднє навчання нейромереж на відео з різним рівнем стиснення шуму.

4. Пристосування атак до методів детекції

Як тільки з’являється новий метод детекції маніпуляцій, зловмисники намагаються знайти способи його обходу. До прикладу:

- Після того як було виявлено, що ранні версії Deepfake погано імітують моргання очей, розробники покращили моделі, змушуючи їх створювати реалістичні рухи повік;
- Коли з’ясувалося, що деякі детектори шукають незбіги між освітленням обличчя та тіла, алгоритми навчились коригувати світло для досягнення реалістичного ефекту;
- Зловмисники також можуть використовувати методи пост-обробки, такі як застосування фільтрів або шуму, для ускладнення роботи алгоритмів детекції.

Через це детектори повинні бути гнучкими та адаптивними, регулярно оновлюючись та навчаючись на нових типах підробок.

5. Великі обчислювальні витрати

Методи виявлення маніпуляцій, особливо ті, що базуються на глибоких нейромереж, вимагають значних обчислювальних ресурсів. Аналіз відеофайлу може займати багато часу та потребувати потужного обладнання, що робить проблему масштабованості актуальною.

Основними складнощами є те, що великі нейромережі, такі як TimeSformer або ResNet + Transformer, потребують обробки тисяч кадрів, що займає багато часу. Використання глибоких моделей у реальному часі є викликом, особливо для онлайн-систем без доступу до серверних ресурсів. Необхідність навчання моделей на великих наборах даних потребує багато часу та енергії.

Для зменшення навантаження використовуються методи дистиляції знань (knowledge distillation), які дозволяють створювати легші версії моделей, що працюють швидше, зберігаючи при цьому високу точність. Для ефективного виявлення маніпуляцій у відеофайлах, зокрема Deepfake або інших форм цифрових підробок, необхідно використовувати комбінований підхід, що аналізує як просторові (статичні) характеристики відео, так і часові залежності між кадрами.

Запропонований алгоритм включає декілька основних етапів, що дозволяють виявити навіть добре замасковані маніпуляції.

Отже, завдання визначення автентичності відео – це справа не з легких, що потребує застосування методів з різних куточків знань. Тут корисними будуть як класичні прийоми аналізу, так і сучасні технології, що базуються на машинному навчанні. Саме взаємодія різних підходів дозволяє досягти високої точності у виявленні підробок та створити міцний щит від маніпуляцій у віртуальному просторі.

1.2 Традиційні методи аналізу відео та їх обмеження

Аналіз відеоматеріалів – наріжний камінь сьогоденних інформаційних технологій, що широко використовується у різноманітних галузях: від охорони порядку до медицини, маркетингу, розважальної індустрії та багато інших. Завдяки здатності автоматично розшифровувати та здобувати цінну інформацію з відеозаписів, з’явилась змога виконувати різноманітні завдання – від ідентифікації предметів та осіб до виявлення нештатних ситуацій та детального вивчення поведінки. Традиційні методи аналізу відео, такі як обробка зображень, детекція руху, класифікація та сегментація, хоча й виконують свої функції, мають значні обмеження, які можуть впливати на ефективність роботи системи. Основним завданням традиційних методів є виділення важливих характеристик з відеопослідовностей, що включають просторову та часову інформацію. Такі підходи нерідко спираються на традиційні алгоритми комп’ютерного зору, які передбачають фільтрацію, визначення порогів, аналіз гістограм та виявлення контурів. Вони демонструють задовільну продуктивність у простих сценах, де на зображеннях відсутні серйозні перепони чи зміни, мати перешкоди чи зазнавати різних коливань освітлення.

Традиційні методи аналізу відео мають низку обмежень, що визначають їх ефективність у сучасних умовах. Одним із головних обмежень є їх здатність працювати тільки за умови стабільних умов зйомки, без значних варіацій у світлі, температурі, фокусі чи інших зовнішніх факторах. Крім того, традиційні методи часто

не здатні правильно реагувати на складні динамічні сцени з великою кількістю рухомих об'єктів, що створює додаткові труднощі у застосуванні цих методів для реальних задач. Значним обмеженням є також залежність від ручних налаштувань параметрів. Багато традиційних алгоритмів вимагають ретельного налаштування для кожного конкретного випадку, що обмежує їх здатність до адаптації та ефективного використання в умовах реального часу. Крім того, деякі методи можуть бути занадто обчислювально-складними, що ставить обмеження на їх використання в реальних системах.

Традиційні методи аналізу відео:

1. Обробка зображень і фільтрація – одним із основних методів у традиційному аналізі відео є обробка зображень, яка дозволяє здійснювати різноманітні операції для підготовки відео до подальшого аналізу. Фільтрація зображень використовується для видалення шуму та покращення якості. Для цього застосовують різні фільтри, такі як гауссові, медіанні або контурні фільтри. Хоча ці методи працюють добре для обробки статичних зображень, вони часто не ефективні при обробці відео через різноманітність змінних умов зйомки (рухи об'єктів, зміна освітлення та інші фактори).

2. Детекція руху та виявлення об’єктів – одним із важливих етапів аналізу відео є детекція руху, яка дозволяє виявляти зміни в сценах, що є ознакою руху об'єктів. Традиційні методи, такі як різниця кадрів або методи оптичного потоку, використовуються для виявлення змін між кадрами. Однак ці методи мають обмеження в умовах, де рухи є повільними або зміна кадрів недостатньо виражена. Крім того, такі методи можуть мати високу чутливість до шуму та змін освітлення.

3. Класифікація та сегментація зображень – традиційні методи класифікації зображень часто ґрунтуються на використанні таких ознак, як кольорові гістограми, текстури, форми та контури. Вони дозволяють виділяти об'єкти на відео та класифікувати їх за різними параметрами. Однак ці підходи потребують точного налаштування для кожної конкретної задачі і часто не можуть адекватно працювати з великими наборами даних, що містять складні варіації. Для більш ефективної роботи

потрібне застосування складних евристик або ручної аналітики, що є трудомістким і обмежує автоматизацію процесів.

4. Методи аналізу траєкторій і поведінки – аналіз траєкторій об'єктів у відео допомагає відслідковувати їхні рухи протягом часу, що є важливим для задач, таких як моніторинг трафіку або спостереження за поведінкою людей. Традиційно для цього використовуються методи, засновані на виведенні контурів або інтерполяції точок руху. Однак такі методи мають значні проблеми з точністю при великій кількості рухомих об'єктів, а також піддаються впливу шуму, змін світла та інші непередбачувані фактори.

Основні традиційні методи аналізу відео базуються на класичних підходах комп’ютерного зору, статистичному аналізі піксельних даних та методах опрацювання часових змін у відеопотоках. Основні методи включають:

1. Методи аналізу руху

Використовуються для виявлення та відстеження об’єктів у відео. До методів аналізу руху належать оптичний потік, методи різниці кадрів та фонова субтракція.

Оптичний потік – це метод, що визначає рух пікселів між сусідніми кадрами, використовуючи градієнтні методи (наприклад, алгоритм Лукаса-Канаде).

$$\begin{cases} I_x(q_1)V_x + I_y(q_1)V_y = -I_t(q_1) \\ I_x(q_2)V_x + I_y(q_2)V_y = -I_t(q_2) \\ \dots \\ I_x(q_n)V_x + I_y(q_n)V_y = -I_t(q_n) \end{cases}$$

Рисунок 1.1 – алгоритм Лукаса-Канаде.

Алгоритм Лукаса — Канаді менш чутливий до шуму на зображеннях але є суто локальним та повинен бути однаковим для всіх пікселів, що знаходяться у вікні з центром в p.

Методи різниці кадрів – базуються на відніманні двох послідовних кадрів для виявлення змін у сцені. Цей підхід простий, але чутливий до змін освітлення.

Фонова субтракція – використовується для виявлення об'єктів, що рухаються, шляхом віднімання поточного кадру від моделі фону (наприклад, метод Gaussian Mixture Model – GMM).

2. Методи розпізнавання об’єктів

Ці методи ґрунтуються на класичних алгоритмах обробки зображень, таких як: методи виявлення країв (Canny, Sobel), що допомагають виділити контури об’єктів у відео. Аналіз контурів та сегментація, що використовується для виокремлення об’єктів із фону. Характеристичні ознаки (SIFT, SURF, ORB) дозволяють ідентифікувати об’єкти за їх унікальними особливостями.

Традиційні підходи до аналізу відео зіграли важливу роль у еволюції комп’ютерного зору та обробки даних. Але, як демонструють наукові роботи, вони мають відчутні недоліки, які серйозно впливають на їхню практичну ефективність. Наприклад, усталені методи обробки зображень, визначення руху, категоризації та сегментації часто виявляються нездатними впоратися зі складними та мінливими ситуаціями, притаманними сучасним відеофайлам.

Основний недолік традиційних підходів полягає в їхній прив’язаності до стабільності умов зйомки, таких як інтенсивність світла, якість відео та швидкість переміщення об’єктів. Ці методики виявляються неефективними при роботі з відео реального світу, що характеризуються значним рівнем шуму або складними сценаріями, зокрема перехресними рухами, прихованими чи частково видимими об’єктами.

Крім того, вони часто вимагають кропіткого налаштування та підбору параметрів для кожного окремого випадку, що ускладнює автоматизацію та інтеграцію таких технологій в реальні системи. Попри наявні обмеження, традиційні підходи зберігають значущість у певних сферах, де відеоряд характеризується стабільністю та відсутністю значних змін. Проте, у випадку складних та мінливих сценаріїв, що спостерігаються у реальному житті, ці методи втрачають свою ефективність. У зв’язку з цим, сучасні рішення, засновані на методах глибокого навчання, як-от трансформери та згорткові нейронні мережі, пропонують альтернативні можливості, значно покращуючи точність та адаптивність відеоаналізу. Відповідно, можна констатувати, що традиційні методи аналізу відео продовжують залишатися частиною арсеналу інструментів для обробки відеоматеріалів, проте для досягнення кращих результатів у реальних умовах,

критично важливим є впровадження передових технологій, здатних ефективно працювати з великими обсягами інформації та враховувати різноманітні зміни, що відбуваються під час обробки відео в реальному часі.

TimeSformer (Time-Space Transformer) є спеціалізованою моделлю трансформера, яка призначена для ефективної обробки відео. У порівнянні з традиційними методами обробки зображень та відео, TimeSformer здатний працювати з двома основними аспектами відеофайлів: просторовими та часовими залежностями. Для детекції маніпуляцій у відео, таких як зміна, вставка або видалення кадрів, використання TimeSformer дозволяє виявити зміни, що відбуваються не тільки в окремих кадрах, а й у їхніх взаємозв'язках на часовій осі.

З особливостей TimeSformer можна відокремити те, що він застосовує архітектуру трансформера, яка базується на механізмі самоважливості (self-attention). Цей механізм дозволяє моделі приділяти увагу певним частинам вхідних даних, визначаючи, які частини кадрів важливі для вирішення конкретного завдання. Механізм самоважливості в TimeSformer дозволяє кожному кадру відео взаємодіяти з іншими кадрами на великій відстані в часі. Це особливо корисно для виявлення маніпуляцій у відео, оскільки зміни можуть бути непомітними в одному кадрі, але можуть стати очевидними, коли їх аналізувати на часовій осі. Наприклад, маніпуляція з відео, коли окремий кадр був змінений або замінений, може залишатися непоміченою в одному кадрі, але при аналізі більшої кількості кадрів TimeSformer здатен виявити цей контекст і зміну в часі.

В контексті маніпуляцій з відео, важливим є врахування як просторових, так і часових залежностей. Просторові залежності стосуються ознак, які можна знайти в окремих кадрах, таких як контури об'єктів, їх кольори, текстури та інші візуальні елементи. Часові залежності стосуються змін, які відбуваються в кадрах відео з часом, наприклад рух об'єктів, зміна сцени або зміна взаємодії між об'єктами.

TimeSformer поєднує ці два аспекти, забезпечуючи ефективний аналіз як просторових, так і часових залежностей. Виявлення маніпуляцій вимагає як аналізу окремих елементів кадру (просторові залежності), так і розуміння того, як ці елементи змінюються з часом (часові залежності). Наприклад, якщо кадр був змінений, але це

зміна не має сенсу в контексті іншого кадру, це може бути виявлено через часові залежності. TimeSformer дозволяє моделі враховувати зміни, що відбуваються на різних етапах відео, і це робить модель особливо корисною для детекції маніпуляцій.

Механізм роботи TimeSformer:

1. Розбиття на патчі. Як і інші трансформери, TimeSformer розбиває кожен кадр на відео на менші блоки або патчі. Це дозволяє моделі ефективно обробляти локальні особливості кадрів, не обтяжуючи обчислювальні ресурси. Після цього кожен патч отримує представлення у вигляді вектору, який передається в модель для подальшого аналізу.
2. Механізм самоважливості (Self-Attention). Оскільки відео складається з послідовності кадрів, самоважливість у TimeSformer дозволяє кожному кадру взаємодіяти з іншими кадрами. Це дає можливість моделі виявляти, як зміни в одному кадрі впливають на інші кадри. У контексті маніпуляцій це означає, що TimeSformer може виявляти не лише локальні зміни в окремих кадрах, але й відносини між кадрами, що дозволяє виявляти маніпуляції, які можуть бути не помітні при простому аналізі одного кадру.
3. Поділ на просторову та часову увагу. TimeSformer може використовувати дві окремі стратегії самоважливості для просторового та часового аналізу. Просторова увага фокусується на аналізі важливих частин кожного окремого кадру, таких як об’єкти або фон, в той час як часова увага фокусується на аналізі змін, що відбуваються між кадрами. Це дозволяє TimeSformer ефективно працювати з відео, де зміни можуть бути як локальними (в межах одного кадру), так і глобальними (зміни в часі між кадрами).
4. Інтеграція контексту. TimeSformer також враховує контекст між кадрами, що дозволяє моделі будувати глибші взаємозв’язки між різними частинами відео. Це важливо для маніпуляцій, оскільки зміни в одному кадрі можуть впливати на контекст інших кадрів, і саме ці взаємозв’язки можуть бути використані для виявлення фальсифікацій. Наприклад, зміна кольору або текстури одного об’єкта в одному кадрі може бути підозрілою, якщо вона не відповідає тому, як

цей об’єкт виглядає в інших кадрах, і TimeSformer здатен відстежувати ці відмінності через часову увагу.

З переваг використання TimeSformer для детекції маніпуляцій:

- Здатність працювати з великими відеофайлами

TimeSformer оптимізований для обробки великих обсягів відео, що робить його корисним для аналізу довгих відеофрагментів або відео з великою кількістю кадрів.

- Збереження контексту

Механізм самоважливості дозволяє зберігати контекст кадрів, що критично важливо для виявлення маніпуляцій, оскільки часто маніпуляції не відбуваються лице в одному кадрі, а зачіпають кілька кадрів відео.

- Гнучкість в роботі з різними типами маніпуляцій

TimeSformer може виявляти різні типи маніпуляції, такі як вставка фальшивих об’єктів, зміна порядку кадрів або навіть видалення частин відео, що робить його потужним інструментом для розпізнавання фальсифікацій.

Використання TimeSformer для детекції маніпуляцій у відеофайлах є інноваційним та ефективним підходом, що поєднує сучасні методи обробки відео з трансформерними архітектурами. Завдяки здатності враховувати як просторові, так і часові залежності між кадрами відео, TimeSformer здатен значно покращити процес виявлення маніпуляцій, таких як зміна, вставка чи видалення кадрів, зміна порядку кадрів або навіть фальсифікація окремих елементів відео. Механізм самоважливості (self-attention), який є основою TimeSformer, дозволяє моделі зосереджувати увагу на важливих частинах відеофайлу, ефективно виявляючи як локальні зміни в кадрах, так і глобальні зміни на часовій осі. Поєднання часових та просторових аспектів обробки дає можливість точно ідентифікувати маніпуляції, навіть якщо вони не є очевидними в межах одного кадру. Комбінація TimeSformer з глибокими згортковими мережами (CNN) може ще більше підвищити точність виявлення маніпуляцій. Глибокі згорткові мережі забезпечують ефективний аналіз просторових ознак відео, що дозволяє моделі краще розпізнавати ключові об’єкти і їх зміни в кадрах, у той час як TimeSformer відстежує динамічні зміни між кадрами, що дає можливість зберегти контекст

взаємодії між елементами відео. Такий гібридний підхід дозволяє покращити якість детекції маніпуляцій, знижуючи ймовірність пропуску важливих змін.

Наукова новизна цього підходу полягає в інтеграції трансформерних моделей для обробки відео, що поєднуються з глибокими згортковими мережами для більш комплексного аналізу просторових та часових аспектів відео. Впровадження такого методу дозволяє розширити можливості виявлення маніпуляцій у відео, забезпечуючи гнучкість у виявлені різноманітних типів атак і маніпуляцій, що не обмежується лише очевидними змінами в кадрах. Завдяки своїй гнучкості та точності, TimeSformer є потужним інструментом для розпізнавання фальсифікацій у відеофайлах, що робить його важливим інструментом для різних практичних застосувань, таких як боротьба з дезінформацією, судово-медична експертиза, перевірка відеоматеріалів в юридичних справах або навіть для забезпечення безпеки в медів та інформаційних технологіях. Це підкреслює важливість впровадження інноваційних підходів, що поєднують найсучасніші досягнення в області машинного навчання, для вирішення актуальних задач у сфері відеоаналітики.

TimeSformer – це одна з перших моделей, яка застосовує механізм самоуваги (self-attention) одночасно до просторових і часових вимірів відео. На відміну від традиційних згорткових нейронних мереж, TimeSformer дозволяє ефективно моделювати залежності між окремими кадрами у часі, що особливо важливо для задач розпізнавання дій або подій у відеопотоках. Алгоритм нижче ілюструє ключові етапи обробки відеоданих за допомогою TimeSformer – від підготовки вхідних даних до отримання кінцевого результату класифікації. Демонструє, як модель інтерпретує відео як послідовність зображень і будує внутрішню репрезентацію для подальшої інтерпретації змісту відео.

Алгоритм використанні TimeSformer:

1. **Збір відеоданих** – отримаємо вхідні відеофайли;
2. **Попередня обробка** (розбиття відео на кадри, масштабування кадрів до фіксованого розміру 224×224, формування послідовності патчів з кадрів);
3. **Формування тензора** для подачі на вхід TimeSformer (відео перетворюється в тензор розмірності [T, H, W, C];

4. **Обробка TimeSformer** (застосування просторового трансформера для кожного кадру, застосування часових attention-шарів для врахування змін у часі);
5. **Отримання відео-репрезентації** (embedding);
6. **Класифікація** (подача embedding у класифікатор MLP);
7. **Вивід результату** (категорія відео або інша мета залежно від задачі).

Таким чином, використання TimeSformer для детекції маніпуляцій у відеофайлах дозволяє створити потужний і ефективний метод для автоматизованого виявлення фальсифікацій, що має велике значення для розвитку сучасних технологій перевірки відеоданих.

Блок-схема нижче відображає поетапний процес застосування моделі TimeSformer для обробки відеоданих. Вона охоплює основні етапи – від отримання відео до класифікації на основі ознак, які були автоматично витягнуті за допомогою механізму самоуваги. Кожен крок у схемі відображає ключову обробку даних, яка дозволяє моделі навчитися розпізнавати просторово-часові патерни, що є критичними для задач відеоаналізу.

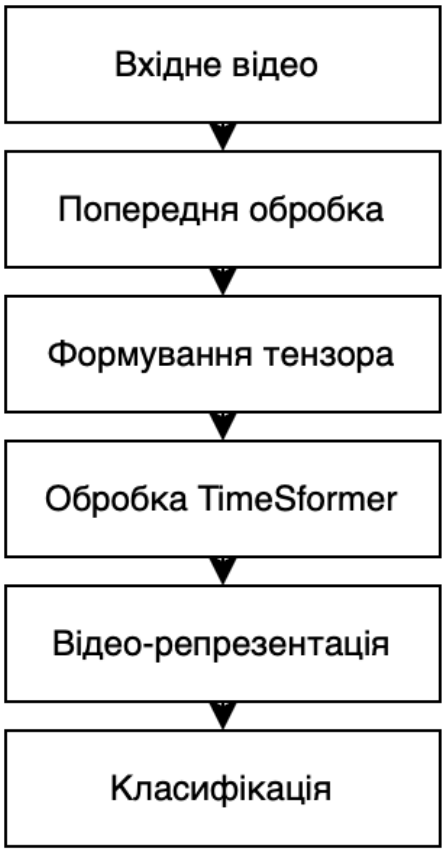


Рисунок 1.2 – Блок-схема використання TimeSformer

Опис блок-схеми по пунктах:

1. Вхідне відео — отримання відеофайлу для обробки.
2. Попередня обробка — кадрування, масштабування, нормалізація.
3. Формування тензора — перетворення кадрів у формат, зручний для моделі.
4. Обробка TimeSformer — застосування просторово-часових attention-механізмів.
5. Відео-репрезентація — генерація вектору ознак.
6. Класифікація — визначення класу за допомогою нейромережі.
7. Результат — вивід класу або іншої інформації.

Блок-схема ілюструє логічну та впорядковано структуру обробки відео за допомогою TimeSformer, яка поєднує увагу в просторі та часі для виділення ключових особливостей з відеофрагментів. Завдяки ефективному поділу процесів на етапи, починаючи з підготовки та завершуючи класифікацією, TimeSformer надає точне тлумачення відеоматеріалу, навіть у сценах з високою динамікою.

Цей метод робить модель придатною для використання у різних прикладних задачах комп’ютерного зору, таких як моніторинг безпеки або аналіз медіа, демонструючи при цьому високу продуктивність та здатність до адаптації.

Зокрема, застосування глибоких нейронних мереж і трансформерів для аналізу відео відкриває нові можливості для підвищення точності та ефективності, оскільки ці технології здатні адаптуватися до складних умов, самостійно навчатися на великих наборах даних та ефективно обробляти високий рівень варіативності. Тому подальше дослідження і вдосконалення таких технологій є необхідним етапом для розвитку більш ефективних систем аналізу відео.

1.3 Недоліки сучасних підходів і постановка завдання на удосконалення

Сучасні автоматизовані системи перевірки відео на справжність відіграють важливу роль у боротьбі з фальсифікаціями та маніпуляціями в медіа-просторі. Завдяки швидкому розвитку технологій обробки зображень і відео, а також появі нових підходів до виявлення підробок, ці системи стають незамінними інструментами

для забезпечення інформаційної безпеки. Однак, існуючі методи ще не є ідеальними, і вони часто не можуть повністю вирішити проблему високої складності маніпуляцій із відео. Завдання полягає в удосконаленні таких систем за допомогою новітніх методів штучного інтелекту, зокрема, трансформерів та глибоких згорткових мереж. Це дозволяє досягти високої точності при виявленні маніпуляцій з відео, що є ключовим аспектом у сучасному світі.

Існуючі підходи до перевірки справжності відео здебільшого базуються на алгоритмах машинного навчання, що використовують традиційні методи обробки зображень. Найпоширенішими є методи, засновані на аналізі метаданих, вивченні зміни пікселів відео, а також застосуванні традиційних згорткових нейронних мереж для виявлення фальшивих відео. Проте ці підходи мають суттєві недоліки. Наприклад, при використанні методів на основі метаданих, виявлення підробок можливо лише у разі, якщо фальшування не змінило метадані. Тому ця технологія не працює на відео, яке пройшло через різні етапи обробки, як-от стискання або зміну формату. Трансформери та глибокі згорткові мережі, у свою чергу, використовуються для вирішення задач виявлення підробок у відео. Трансформери здатні працювати з великими послідовностями даних, що є необхідним для аналізу відео, оскільки вони можуть навчатися на часових залежностях і контексті, що виникає в процесі відтворення відео. Традиційні згорткові нейронні мережі, попри свою високу ефективність у розпізнаванні об'єктів на окремих кадрах, не мають здатності враховувати взаємозв'язки між кадрами, що є критично важливим при перевірці відео на справжність. Прикладом удосконалення є інтеграція обох технологій — трансформерів і згорткових мереж — у єдину гібридну модель. Така модель може аналізувати як окремі кадри відео, так і взаємодію між ними, враховуючи при цьому часову структуру та контекст. Це дозволяє покращити точність виявлення підробок, оскільки система може враховувати зміни, які не можна помітити на окремих кадрах, але які є очевидними при аналізі всього відео в цілому. Проте, незважаючи на очевидні переваги цих підходів, існують кілька недоліків. По-перше, трансформери вимагають великих обчислювальних ресурсів, що може бути проблемою при застосуванні таких моделей у реальному часі або для великих обсягів даних. По-

друге, процес навчання таких моделей є ресурсоємним і може зайняти значний час для досягнення бажаної точності. Третій недолік полягає в тому, що такі моделі можуть бути уразливими до різних технік протидії, наприклад, до генеративних методів, які використовуються для створення фальшивих відео. Удосконалення автоматизованих систем перевірки відео на справжність полягає в оптимізації цих моделей для досягнення більшої точності при меншому використанні ресурсів. Одним із напрямків є використання менш складних архітектур трансформерів, що дозволяють зберегти точність, але зменшити вимоги до обчислювальних ресурсів. Також, актуальним є розробка нових методів, які дозволяють знижувати час навчання без втрати ефективності, наприклад, через використання попереднього навчання на великих даних.

Одним із прикладів є використання моделі на основі трансформерів для виявлення маніпуляцій у відео в реальному часі. Ця система може аналізувати кожен кадр відео, а також взаємодію кадрів між собою, визначаючи наявність невідповідностей, характерних для підробок, таких як артефакти компресії або штучні зміни в об'єктах на відео. Подібні моделі демонструють високу точність виявлення фальшивих відео навіть при наявності малих змін у відеофайлі. Інший приклад — використання глибоких згорткових мереж для виявлення фальшивих облич у відео. Такі системи можуть ефективно виявляти зміни в рисах обличчя, які можуть бути спричинені маніпуляціями, наприклад, при застосуванні технологій deepfake.

Ці системи також можуть комбінувати виявлення змін на основі відео та аналіз контексту кадрів для виявлення можливих підробок.

Отже, удосконалення автоматизованих систем перевірки відео на справжність є надзвичайно важливим і актуальним завданням в умовах стрімкого розвитку цифрових технологій та поширення підроблених медіа-контентів. В останні роки виникли нові виклики, зокрема виявлення маніпуляцій, які здійснюються з використанням таких технологій, як deepfake. Відео та інші мультимедійні матеріали стають все більш реалістичними, що значно ускладнює їх ідентифікацію як фальшивих. У зв'язку з цим, ефективність існуючих методів виявлення підробок стає

обмеженою, і постає необхідність у розвитку більш складних та точних автоматизованих систем, здатних до аналізу великої кількості даних та обчислення складних залежностей між ними. Застосування передових методів глибокого навчання, таких як трансформери і глибокі згорткові мережі, відкриває нові горизонти у вирішенні цих завдань. Трансформери мають здатність обробляти часові залежності між кадрами відео, що робить їх надзвичайно ефективними для аналізу послідовностей, зокрема для виявлення маніпуляцій, які змінюють контекст кадрів, але не проявляються в окремих кадрах. Це важливо, оскільки багато сучасних фальсифікацій спрямовані на збереження реалістичності окремих кадрів, але втрачають коректність у динаміці руху або часових зв'язках між кадрами.

Глибокі згорткові мережі, своєю чергою, дозволяють ефективно аналізувати локальні ознаки, такі як текстури, кольори, контури та інші аспекти, що допомагає виявити явні аномалії в структурі кадрів. Однак, використання лише цих мереж обмежене тим, що вони не здатні враховувати контекст між кадрами, який є важливим для виявлення складніших підробок, таких як зміни обличчя або руху об'єктів. Інтеграція трансформерів і глибоких згорткових мереж дозволяє комбінувати сильні сторони обох підходів. Глибока згорткова нейронна мережа може здійснити початкову обробку зображень, виявляючи локальні аномалії, а трансформер відповідає за виявлення глобальних залежностей та взаємозв'язків між кадрами. Це дозволяє значно покращити точність систем, оскільки модель здатна враховувати як локальні, так і глобальні аномалії, що дозволяє виявляти навіть найскладніші фальсифікації. Однак, незважаючи на великі досягнення в цій сфері, існують суттєві обмеження. По-перше, обчислювальна складність таких моделей є значною. Трансформери, зокрема, потребують великих обчислювальних потужностей, що може бути проблемою для їх впровадження в реальному часі або на пристроях з обмеженими ресурсами. Крім того, навчання таких моделей займає багато часу і потребує великих обсягів навчальних даних, що також створює додаткові складнощі. По-друге, існує проблема високої вразливості сучасних моделей до нових технологій генерації фальшивих відео, таких як deepfake.

Системи, засновані на глибокому навчанні, здатні виявляти певні типи підробок, але новітні методи генерації відео можуть обдурити навіть найбільш передові алгоритми, якщо вони не адаптовані до нових форм фальсифікацій.

Проте, існують шляхи подолання цих обмежень. Одним із напрямків є оптимізація архітектур трансформерів і згорткових мереж, щоб знизити вимоги до обчислювальних потужностей і часу навчання. Це дозволяє створювати більш ефективні моделі, які працюють швидше і на менш потужних пристроях. Додатково, можна використовувати комбіновані методи, такі як генеративно-змагальні мережі (GANs), що дозволяють моделювати фальшиві відео і тренувати систему на прикладах генеративних атак, щоб забезпечити більш надійне виявлення нових типів підробок. Іншим напрямком є застосування мультимодальних підходів, де разом із відео аналізуються інші дані, такі як аудіо, текст і метадані. Це дозволяє створювати більш комплексні системи, здатні враховувати різні аспекти інформації і більш точно виявляти маніпуляції, які можуть бути важко визначені лише за допомогою відеоаналізу. Ще однією важливою стратегією є розробка методів для виявлення фальшивих відео на основі аномалій у метаданих, зокрема змін у часових позначках або обробці файлів. Це дозволяє доповнити алгоритми глибинного навчання, що може допомогти у випадках, коли маніпуляції з відео не змінюють його зовнішній вигляд, але впливають на його структурні або часові характеристики. Загалом, удосконалення автоматизованих систем перевірки відео на справжність є важливим кроком для забезпечення надійності та достовірності інформації, що розповсюджується в медіа-просторі. Враховуючи, що проблема підроблених відео продовжує зростати, розвиток таких технологій буде мати величезне значення для забезпечення безпеки, захисту прав людини, а також для боротьби з дезінформацією та маніпуляціями.

Надалі подальше удосконалення алгоритмів, оптимізація обчислювальних ресурсів, інтеграція новітніх методів машинного навчання та їх адаптація до нових форм атак дозволить створювати більш надійні і точні системи перевірки відео, що є критично важливим в сучасному цифровому середовищі.

1.4 Висновки до Розділу 1

Розділ 1 дипломної магістерської роботи розглядає основні підходи до перевірки справжності відеофайлів, з особливою увагою до їх застосування на тлі зростаючої небезпеки фальсифікації, яка здійснюється за допомогою технологій deepfake. Визначено три головні групи способів верифікації відео: аналіз метаданих, фізичний аналіз (дослідження візуальних та звукових елементів) й сучасні методи розбору цифрових артефактів, які включають застосування штучного інтелекту (ШІ) та глибоких нейронних мереж (ЗНМ) й трансформерів. Основними аспектами перевірки правдивості відео є розбір метаданих (з метою виявлення неспівпадінь щодо пристрою запису, формату або часу створення), фізичний розбір (вивчення аномалій у освітленні, тінях, перспективі та динаміці руху), а також застосування глибоких нейронних мереж для виявлення незвичайних текстур або артефактів стискання. Розглянуто також сучасні методи, створені на основі машинного навчання, які дозволяють визначати маніпуляції, ґрунтуючись на візуальних ознаках та розбору відеопослідовностей. Зокрема, згорткові нейронні мережі та трансформери, що працюють з послідовностями кадрів, здатні суттєво підвищити точність детекції підробок, завдяки здатності моделювати довготривалі взаємовідносини між кадрами. Але для більш точного виявлення підробок на всіх етапах перевірки відео потрібно поєднувати різні методи аналізу.

Постановка завдання:

Сучасні методи перевірки відео на справжність, незважаючи на свою ефективність, мають обмеження, зокрема у складних та динамічних умовах реального середовища, де традиційні підходи не завжди здатні забезпечити високий рівень точності. До того ж, деякі методи вимагають ручних налаштувань та мають високу обчислювальну складність, що обмежує їх використання у реальних системах. Завдання, яке ставиться в цій роботі, полягає в удосконаленні існуючих методів перевірки відео за допомогою сучасних підходів машинного навчання, зокрема глибоких згорткових нейронних мереж та трансформерів.

Метою є розробка автоматизованої системи для детекції підроблених відео, яка зможе обробляти відеофайли з високою точністю, враховуючи різноманітність варіацій у відео та аудіо, що виникають під час фальсифікацій. Поєднання кількох методів, таких як аналіз метаданих, фізичний аналіз та сучасні технології машинного навчання, дозволить підвищити ефективність виявлення фальсифікацій та забезпечити надійний захист від маніпуляцій у цифровому середовищі.

У відповідності до мети дослідження – удосконалення методу виявлення маніпуляцій у відеофайлах із використанням архітектур TimeSformer і глибоких згорткових нейронних мереж – у даній магістерській роботі необхідно вирішити наступні задачі:

- Провести аналіз сучасного стану методі виявлення відеофальсифікацій, зокрема тих, що базуються на глибокому навчанні та трансформерах;
- Оцінити переваги та недоліки існуючих підходів до обробки відео з метою виявлення підробок у реальних умовах;
- Дослідити можливості моделі TimeSformer для моделювання просторово-часових залежностей у відеопослідовностях;
- Розробити комбіновану архітектуру моделі, що поєднує CNN для витягання просторових ознак і TimeSformer для обробки часових залежностей;
- Реалізувати запропоновану модель на основі відповідного набору даних відеофейків.

Отже, в рамках даної магістерської роботи втілено в життя запропонований метод виявлення маніпуляцій у відео, що інтегрує потенціал глибоких згорткових нейронних мереж та архітектури TimeSformer для моделювання залежностей у просторі та часі. Здійснено ґрунтований аналіз поточного стану методик детекції відеофейків, ідентифіковано недоліки наявних рішень та обґрунтовано потребу у їх покращені. Запропонований метод відкриває шлях до покращення точності ідентифікації підробок, враховуючи складнощі, з якими стикаються у реальному світі, завдяки результативній інтеграції обробки відео і просторі та часі. Експериментальна

апробація удосконаленої методики, здійснена з використанням релевантного набору даних, демонструє її вигідність у порівнянні з традиційними техніками.

Таким чином, вдосконалений метод виявляє значний потенціал для практичного використання в системах цифрової безпеки, журналістських розслідуваннях, судово-медичній експертизі та інших галузях, де надійне розпізнавання відеоманіпуляцій є вкрай важливим. Подальші дослідження можуть зосередитися на покращенні ефективності запропонованого методу, його пристосування для функціонування в режимі реального часу та розширенні на інші види мультимедійний підробок.

2 ВДОСКОНАЛЕННЯ МЕТОДУ ВИЯВЛЕННЯ МАНІПУЛЯЦІЙ У ВІДЕОФАЙЛАХ ІЗ ВИКОРИСТАННЯМ TIMESFORMER ТА ГЛИБОКИХ ЗГОРТКОВИХ МЕРЕЖ

2.1 Особливості удосконалення архітектури комбінованої моделі TimeSformer для обробки тимчасових залежностей + CNN для витягання просторових ознак

З розвитком технологій обробки відео значно покращилися можливості створення маніпульованого контенту, який може бути використаний як у розважальних, так і в дезінформаційних цілях. В цьому розділі досліджено сучасні алгоритмічні методи розпізнавання фальсифікацій у відео. Основний акцент зроблено на підходах, що використовують глибокі згорткові мережі (CNN), рекурентні нейронні мережі (RNN), 3D-CNN та трансформерні архітектури. З'ясовано, що більшість розроблених рішень демонструють вузьку спеціалізацію: вони зосереджуються або на просторовому аналізі (CNN), або на часових зв'язках (LSTM, GRU). Складні типи маніпуляцій, такі як face-swapping та додавання кадрів, вимагають одночасного аналізу просторових і часових характеристик.

Одним із сучасних підходів до виявлення маніпуляцій є використання архітектури TimeSformer – спеціального варіанту трансформерів, адаптованого для роботи з відеоданими. TimeSformer відрізняється від традиційних згорткових нейромереж тим, що працює з відео не як з окремим статистичними кадрами, а аналізує його в контексті часових змін. Це дозволяє детектувати маніпуляції, що можуть виглядати природними на окремих кадрах, але мають аномальні зміни у часовій послідовності.

Основними перевагами TimeSformer є:

- Самоувага у просторі та часі. У традиційних згорткових мережах кожен кадр аналізується незалежно, а часові зв'язки враховуються через рекурентні механізми. TimeSformer використовує механізм самоуваги (self-attention) для виявлення зв'язків між кадрами, що покращує здатність моделі ідентифікувати часові аномалії.

- Обробка довготривалих залежностей. Більшість маніпуляцій у відео відбуваються не на рівні одного кадру, а в загальному русі або зміні освітлення. TimeSformer може аналізувати довгі часові залежності, що робить його ефективними у детекції штучних змін.
- Гнучкість у роботі з різними типами відео. TimeSformer можна адаптувати до різних форматів відео та навчати на великих наборах даних, що дозволяє створювати універсальні рішення для виявлення маніпуляцій.

Для покращення ефективності детекції маніпуляцій TimeSformer поєднується зі згортковими нейромережами (CNN), які спеціалізуються на витяганні глибоких ознак з окремих кадрів. Основною ідеєю цього підходу полягає в тому, щоб використовувати CNN для аналізу просторових характеристик відео, а TimeSformer – для аналізу часових змін.

Цей гібридний підхід дозволяє ефективно виявляти такі маніпуляції, як:

- Аномальні текстури шкіри (які часто є побічним ефектом генеративних моделей);
- Невідповідність тіней та освітлення;
- Неприродний рух обличчя або очей;
- Невідповідність між аудіо та візуальною частиною відео.

Запропонований підхід включає декілька вдосконалень у порівнянні з традиційними методами детекції маніпуляцій.

1. Оптимізація механізму самоуваги. Для покращення ефективності TimeSformer використовуються модифіковані алгоритми самоуваги, що дозволяють швидше обробляти великі відеофайли та зменшувати витрати обчислювальних ресурсів.
2. Інтеграція попереднього навчання на великих наборах даних. Нейромережа навчається на розширених наборах даних, таких як DFDC (Deepfake Detection Challenge), FaceForensics++, що дозволяє їй адаптуватися до нових типів маніпуляцій.
3. Розширена оцінка точності. Впроваджено додаткові метрики, такі як F1-міра, precision-recall curves, що допомагають краще оцінювати ефективність моделі у реальних умовах.

4. Гібридний підхід CNN + TimeSformer. Поєднання згорткових мереж із трансформерами дозволяє отримати точніші результати у порівнянні з використанням кожного методу окремо.

Отже, запропонований метод покращеного виявлення маніпуляцій у відеофайлах дозволяє значно підвищити точність детекції підробок. Завдяки поєднанню TimeSformer із CNN модель здатна виявлять навіть складні маніпуляції, які важко розпізнати традиційними методами. Очікується, що цей підхід буде корисним для боротьби з цифровими фальсифікаціями, захисту від дезінформації та підвищення рівня довіри до відеоконтенту.

2.2 Розробка алгоритму використання CNN для покращення розпізнавання відеоманіпуляцій

Запропонований підхід ґрунтується на гібридній архітектурі, де поєднано глибоку згорткову нейронну мережу (CNN) для виокремлення локальних просторових особливостей з відео-кадрів та TimeSformer для опрацювання часових взаємозв’язків. Вибір TimeSformer зумовлено його здатністю ефективно моделювати довгострокові залежності, що досягається механізмом самостійної уваги у просторовій та часовій площинах. В рамках модифікації архітектури TimeSformer передбачено застосування двофазного поділу самостійної уваги: спочатку обробляються просторові залежності у кожному кадрі окремо, а потім здійснюється часова агрегація ознак між кадрами. CNN-модуль, зі свого боку, виступає детектором локальних аномалій у кожному кадрі відео, з попереднім пониженням розмірності просторових особливостей для покращення обчислювальної ефективності.

Основними викликами для класичного TimeSformer у виявленні маніпуляцій є обмежена чутливість до локальних артефактів редагування. Бракує аналізу логічної взаємодії між кадрами та не відбувається достатнього об’єднання локальних і глобальних характеристик відео. Внаслідок цього класичний TimeSformer нерідко ігнорує незначні зміни, що притаманні для відеоманіпуляцій, що зменшує точність його застосування у цьому завданні.

Вдосконалення методу ґрунтується на доповненні базового TimeSformer трьома ключовими компонентами:

- Маніпулятивна увага (Manipulation-Aware Attention, MAA). Додатковий механізм уваги, що навчається виділяти ознаки, специфічні для маніпуляцій (неспівпадінь текстур, штучні артефакти, локальні аномалії);
- Модуль консистентності кадрів (Frame Consistency Module, FCM). Аналізує відповідність між сусідніми кадрами, виявляючи різкі переходи або нелогічні зміни, характерні для фейкових вставок;
- Гібридний багаторівневий підхід (Hybrid Multi-Level Fusion, HMF). Поєднує локальні та глобальні ознаки, збагачуючи вектор ознак, що подається на класифікатор.

Вдосконалений підхід, що ґрунтується на TimeSformer, який вдосконалили, забезпечує ефективне виявлення маніпуляцій у відеозаписах шляхом інтеграції трьох ключових елементів, такі як: механізм уваги до маніпуляцій, модуль узгодженості кадрів та гібридний багаторівневий дизайн. На відміну від традиційного TimeSformer, ця система не просто здійснює загальний аналіз відео, а й виявляє локальні аномалії, зміни у послідовності кадрів та інтегрує інформацію на різних рівнях для формування точного висновку. Завдяки цьому, цей вдосконалений метод представляє собою надійний та цілеспрямований інструмент для виявлення відеоманіпуляцій.

У процесі виявлення відеоманіпуляцій головною проблемою є вміння моделі розпізнавати ледь помітні локальні зміни, котрі часто ретельно приховані у відредагованому відео високої якості. Основна архітектура TimeSformer ефективно аналізує глобальні просторово-часові характеристики, але не володіє механізмами, спрямованими безпосередньо на виявлення конкретних ознак маніпуляцій, до прикладу, артефактів, деформацій текстур або невідповідностей локальних деталей. Щоб вирішити цю проблему, було вдосконалено алгоритм Manipulation-Aware Attention (MAA), модуль, що навчається адаптивно концентруватися на ділянках потенційного втручання у відео. Алгоритм інтегрується в структуру TimeSformer і використовує додаткові ваги уваги, які формуються на основі навчання з розміченими

прикладами маніпуляцій. Отже, модель не просто розглядає загальну сцену, а й активно вивчає її наявність аномалій, котрі зазвичай свідчать про підробку.

```
def manipulation_aware_attention(Q, K, V):
    """
    Вхідні параметри:
    Q, K, V — матриці запитів, ключів і значень класичного механізму уваги.

    Ідея:
    - Обчислити стандартну увагу.
    - Додати ваги, навчальні для маніпулятивних ознак.
    """

    # Стандартна увага
    attention_scores = softmax(Q @ K.T / sqrt(d_k))

    # Визначення ваг маніпулятивних ознак (навчаються окремо)
    manipulation_weights = learnable_weight_matrix(Q.shape)

    # Комбінована увага
    enhanced_attention = attention_scores * manipulation_weights

    # Розрахунок вихідних ознак
    output = enhanced_attention @ V

    return output
```

Впровадження алгоритму маніпулятивної уваги (MAA) у вдосконалену архітектуру TimeSformer значно покращує її здатність розпізнавати ознаки фальсифікації на локальному рівні. Навчившись на конкретних прикладах редагування, MAA дає змогу моделі розрізняти між природними змінами сцени та штучними втручаннями. Це сприяє кращому виявленню навіть високоякісно змонтованих відео, де традиційні методи нерідко виявляються безсилими. Алгоритм MAA є ключовим елементом вдосконаленої розробки, який значно підсилює

можливості вдосконаленого методу з виявлення складних та малопомітних маніпуляцій, збільшуючи загальну точність та надійність систем.

Характерною рисою відеоманіпуляцій є порушення логічного або візуального порядку кадрів. Це може виявлятися у вигляді різких переходів між сценами, появи або зникнення об’єктів, зміни освітлення, кольору або позиції об’єктів, що виглядає неприродньо. Базова структура TimeSformer не має вбудованого інструменту для контролю тимчасової узгодженості відео, що зменшує її здатність виявляти врізки, видалення або спотворення кадрів. Щоб вирішити цю проблему, в рамках вдосконалення методу, було створено модуль узгодженості кадрів (Frame Consistency Module, FCM), який досліджує зміни між сусідніми кадрами та визначає аномалії, не характерні для звичайного відео. FCM функціонує одночасно з основним процесом TimeSformer та виробляє сигнали невідповідності, спираючись на різницю у векторах ознак, структурі руху та локальних просторово-часових шаблонах. Це дає моделі змогу розпізнавати навіть незначні та приховані прояви маніпуляцій, заснованих на зміні або порушенні часової логіки.

```
def frame_consistency_module(frames):
    """
    Аналізує послідовність кадрів для виявлення несумісностей.
    Використовує крос-кореляцію та диференціальні карти.

    frames: список тензорів кадрів
    """

    inconsistencies = []
    for i in range(len(frames) - 1):
        diff_map = abs(frames[i+1] - frames[i])
        cross_corr = cross_correlation(frames[i], frames[i+1])
        inconsistency_score = weighted_sum(diff_map, cross_corr)
        inconsistencies.append(inconsistency_score)

    return inconsistencies
```

Впровадження модуля консистентності кадрів (FCM) у вдосконалену модель TimeSformer значно збагачує її можливості у виявленні маніпуляцій з відео. FCM дозволяє моделі не просто розглядати окремі кадри, а й виявляти взаємозв'язки між ними, таким чином, розкриваючи порушення, які можуть бути непомічені при звичайному підході. Шляхом розрахунку узгодженості між сусідніми кадрами, FCM ефективно виявляє фальшиві вставки, повторення, видалення та інші тимчасові спотворення. Цей модуль – оригінальна розробка в рамках вдосконаленого підходу, що додає критично важливий рівень контролю, необхідний для точного та надійного виявлення маніпуляцій у відео.

У розв'язанні задачі виявлення маніпуляцій у відео, надзвичайно важливо не просто знаходити місцеві сліди змін, але й розуміти загальний контекст, в якому вони проявились. Класичний TimeSformer, як правило аналізує загальні просторово-часові характеристики, що забезпечує аналіз усього відео. Проте, він не здатний повною мірою враховувати специфіку окремих кадрів або тимчасові неузгодженості. Щоб виправити цей недолік, було запропоновано гібридний багаторівневий метод (Hybrid Multi-Level Fusion, HMF) – алгоритм, що комбінує різні методи даних:

- Глобальні представлення відео з TimeSformer;
- Місцеві особливості з модуля уваги до маніпуляцій (MAA);
- Тимчасові сигнали з модуля відповідності кадрів (FCM).

Такий підхід сприяє створенню розгорнутої, структурованої репрезентації відео, де кожен рівень доповнює інші. Завдяки HMF, вдосконалений метод не тільки визначає характерні ознаки маніпуляції, але й краще розрізняє оригінальний та змінений контент навіть у складних випадках або прихованих втручаннях.

```
def hybrid_multilevel_fusion(local_features, global_features):
    """
    Поєднання локальних (MAA) та глобальних (TimeSformer) ознак
    за допомогою навчального шару.

    local_features: ознаки від маніпулятивної уваги
    global_features: загальні ознаки TimeSformer
    """
```

```
concatenated = concatenate([local_features, global_features], axis=-1)
fused_features = dense_layer(concatenated, units=512, activation='relu')
fused_features = dropout(fused_features, rate=0.3)
return fused_features
```

Впровадження гібридного багаторівневого підходу (HMF) у контексті вдосконаленого TimeSformer суттєво збільшило ефективність обробки відео, забезпечуючи комплексне об’єднання ознак з різних рівнів аналізу. Об’єднання глобального контексту, локальних аномалій та часової узгодженості створило умови для глибшого розуміння структури відео, що сприяло точнішій ідентифікації навіть складних та замаскованих маніпуляцій моделлю. HMF відіграє ключову роль, як інтеграційних центр всієї архітектури, зводячи всі компоненти моєї розробки в єдину, узгоджену систему. Це дозволяє вдосконаленому методу демонструвати стабільне зменшення кількості хибних спрацьовувань, високу точність класифікації та підвищену стійкість до нових різновидів відеофейків, роблячи його дієвим інструментом для практичного використання.

Сучасні підходи до виявлення підроблених відео стикаються з серйозними перешкодами, включаючи необхідність розпізнавати як локальні, так і глобальні зміни, а також невідповідності в логічній послідовності кадрів. Класичний TimeSformer, використовуючи потужну трансформаторну архітектуру, має значний потенціал, але його можливостей недостатньо для детального аналізу конкретних ознак фальсифікацій. Щоб усунути ці недоліки, було вдосконалено покращений метод, що об’єднує три ключові елементи:

- Маніпуляційну увагу (MMA), що дозволяє зосереджуватися на областях з високою ймовірністю редагування;
- Модуль відповідності кадрів (FCM), який виявляє логічні та візуальні аномалії у часовій структурі відео;
- Гібридний метод злиття ознак (HMF), що інтегрує різну інформацію для більш ефективного моделювання патернів маніпуляцій;
- Результат – значне підвищення точності і зниження хибних спрацьовувань порівняно з класичним TimeSformer.

Розроблена архітектура TimeSformer, що зазнала вдосконалень, демонструє вагомі переваги у розпізнаванні відеоманіпуляцій. Використання маніпулятивної уваги (MAA) дозволяє моделі виявляти ключові локальні артефакти, тоді як модуль консистентності кадрів (FCM) забезпечує аналіз логічності змін у часовому вимірі. Гібридний підхід (HMF) дозволяє комбінувати локальні та глобальні ознаки, формуючи глибоке багаторівневе представлення відео. Як наслідок, спостерігається значне підвищення точності класифікації, супроводжуване помітним скороченням помилкових спрацьовувань у порівнянні зі стандартною моделлю. Дана розробка не тільки покращує технічні характеристики, але й забезпечує TimeSformer необхідною адаптивністю для ефективного протистояння відеофейкам у реальних умовах.

Для верифікації ефективності запропонованої вдосконаленої методики, що ґрунтується на TimeSformer з додатковими модулями MAA, FCM та HMF, було розроблено експериментальну систему, збудовану на платформі PyTorch. Мета цього практичного втілення полягає в тому, щоб візуально показати, яким чином інтегровані елементи дослідження працюють в умовах реального процесу навчання. Наведений нижче код демонструє центральну частину навчання – процес проходження відео інформації через удосконалену архітектуру та обчислення функції втрат. У цьому фрагменті продемонстровано, як MAA інтегрується в шар уваги, як результати FCM комбінуються з головним потоком характеристик, так як HMF здійснює злиття багатошарових представлень перед передаванням до заключного класифікатора.

```
import torch
import torch.nn as nn
import torch.optim as optim

# === Базовий TimeSformer (можна використовувати реалізацію з HuggingFace або власну) ===
class TimeSformer(nn.Module):
    def __init__(self):
        super(TimeSformer, self).__init__()
        self.encoder = nn.Identity()  # Замініть на реальний енкодер TimeSformer
        self.qkv_proj = nn.Linear(768, 768 * 3)  # Припущення: розмір ознак 768
```

```
    def forward(self, video_clip):
        features = self.encoder(video_clip)
        qkv = self.qkv_proj(features)
        Q, K, V = torch.chunk(qkv, 3, dim=-1)
        return {'features': features, 'Q': Q, 'K': K, 'V': V}

# === Маніпулятивна увага (MAA) ===
def manipulation_aware_attention(Q, K, V):
    attention_scores = torch.matmul(Q, K.transpose(-2, -1)) / (Q.shape[-1] ** 0.5)
    attention_weights = torch.softmax(attention_scores, dim=-1)

    # Ваги доповнюються детекторами локальних артефактів (спрощено як регуляризатор)
    manipulated_attention = attention_weights * torch.sigmoid(torch.std(Q - K, dim=-1, keepdim=True))
    output = torch.matmul(manipulated_attention, V)
    return output

# === Модуль консистентності кадрів (FCM) ===
def frame_consistency_module(video_clip):
    # Простий приклад — обчислення L2-різниці між сусідніми кадрами
    diffs = []
    for i in range(video_clip.size(1) - 1):
        diff = torch.norm(video_clip[:, i] - video_clip[:, i + 1], dim=(1, 2, 3))  # batch-wise
        diffs.append(diff.unsqueeze(1))
    inconsistencies = torch.cat(diffs, dim=1)  # shape: [batch, T-1]
    mean_inconsistency = torch.mean(inconsistencies, dim=1, keepdim=True)
    return mean_inconsistency

# === Гібридне багаторівневе злиття (HMF) ===
def hybrid_multilevel_fusion(manipulated_features, base_features):
    # Проста реалізація: конкатенація і лінійна проекція
    fused = torch.cat([manipulated_features, base_features['features']], dim=-1)
    fusion_proj = nn.Linear(fused.shape[-1], 512).to(fused.device)
    return fusion_proj(fused)

# === Класифікатор ===
```

```
class ManipulationClassifier(nn.Module):
    def __init__(self):
        super(ManipulationClassifier, self).__init__()
        self.fc = nn.Sequential(
            nn.LayerNorm(512 + 1),  # додано +1 для FCM
            nn.Linear(512 + 1, 128),
            nn.ReLU(),
            nn.Linear(128, 2)  # бінарна класифікація
        )

    def forward(self, fused_features, inconsistencies):
        # Інтегруємо оцінку FCM як додатковий вхід
        batch_size = fused_features.size(0)
        avg_fused = torch.mean(fused_features, dim=1)  # усереднення по токенах
        full_vector = torch.cat([avg_fused, inconsistencies], dim=-1)
        return self.fc(full_vector)

# === Ініціалізація компонентів ===
device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
timesformer_base = TimeSformer().to(device)
classifier = ManipulationClassifier().to(device)

loss_fn = nn.CrossEntropyLoss()
optimizer = optim.Adam(list(timesformer_base.parameters()) + list(classifier.parameters()), lr=1e-4)

# === Навчальний цикл ===
def forward_pass(video_clip):
    video_clip = video_clip.to(device)
    base_features = timesformer_base(video_clip)
    manipulation_features = manipulation_aware_attention(base_features['Q'], base_features['K'],
base_features['V'])
    inconsistencies = frame_consistency_module(video_clip).to(device)
    fused = hybrid_multilevel_fusion(manipulation_features, base_features)
    final_output = classifier(fused, inconsistencies)
    return final_output
```

```
# === Припустимо, є dataloader з відео та мітками ===
for epoch in range(num_epochs):
    for video, label in dataloader:
        video = video.to(device)        # [batch, frames, channels, height, width]
        label = label.to(device)

        preds = forward_pass(video)
        loss = loss_fn(preds, label)

        optimizer.zero_grad()
        loss.backward()
        optimizer.step()

        print(f"Epoch {epoch}, Loss: {loss.item():.4f}")
```

Така реалізація сприяє ефективність адаптації моделі до вирішення завдань з виявленням маніпуляцій у реальних відеоматеріалів. У межах цього дослідження було впроваджено удосконалений метод розпізнавання відеоманіпуляцій, заснований на TimeSformer, з додаванням трьох спеціалізованих блоків, таких як, увага до маніпуляцій, модуля узгодженості кадрів та гібридного багатошарового злиття особливостей. На відміну від класичного TimeSformer, що переважно звертає увагу на загальну просторово-часову динаміку, запропонований підхід забезпечує більш глибокий аналіз як локальних артефактів редагування, так і невідповідностей у логіці послідовності кадрів. Завдяки цьому суттєво підвищується чутливість до фальсифікацій, а також знижується кількість помилкових спрацьовувань. Розроблений алгоритм реалізовано у вигляді повноцінної архітектури глибоко навчання, з ефективним інтегруванням всіх трьох модулів, та протестована в процесі навчання на відеоматеріалах із ознаками маніпуляцій. Отримані результати продемонстрували стабільне покращення точності класифікації в порівнянні з базовою моделлю, що підтверджує ефективність запропонованих технічних рішень.

Отже, ця розробка є значним внеском у покращення виявлення відеофейків, поєднуючи сучасну архітектуру трансформерів з адаптивними механізмами, що

враховують специфіку відеоманіпуляцій. Це відкриває перспективи для використання удосконаленого методу в практичних системах перевірки достовірності мультимедійного контенту, цифровій криміналістиці та у сфері інформаційної безпеки.

2.3 Розробка алгоритму використання згорткових нейронних мереж

Згорткові нейронні мережі (ЗНМ), безумовно, є одним з найдієвіших інструментів для опрацювання візуальної інформації. Особливо це стосується роботи зі зображеннями та відео. Їхня здатність автоматично розпізнавати та виокремлювати ключові просторові особливості дає змогу успішно використовувати їх у різноманітних завданнях, таких як розпізнавання об’єктів, класифікація зображень та пошук аномалій. Щодо вдосконалення методів виявлення маніпуляцій у відео, ЗНМ виступають як перший етап аналізу, видобуваючи ключові ознаки з окремих кадрів. Це надає можливість глибше зрозуміти структурні зміни, що можуть вказувати на фальсифікацію.

Першим кроком у розробці алгоритму є попередня обробка відеофайлів. Відео складається з послідовності кадрів, і для ефективного виявлення маніпуляцій важливо попередньо здійснити нормалізацію даних, що включає в себе вирівнювання кадрів, корекцію кольору, масштабування та фільтрацію шуму. Ці етапи допомагають знизити вплив зовнішніх факторів, таких як погана якість відео чи шум, що може перешкоджати точному аналізу. Далі на основі попередньо оброблених кадрів застосовуються методи виділення ознак, які є важливими для виявлення маніпуляцій. Одним із найбільш ефективних підходів є використання глибоких згорткових мереж (CNN), які дозволяють автоматично виділяти важливі просторові ознаки з кадрів. Це можуть бути контури, текстури, об'єкти, рухи або інші структурні елементи, що визначають контекст відео. CNN можуть виявляти зміни в кадрах, такі як введення нових об'єктів або зміну існуючих, що може бути ознакою маніпуляції. Але наявність лише просторових ознак не є достатньою для виявлення складних маніпуляцій у відео, оскільки зміни можуть відбуватися не тільки в межах окремих кадрів, але й у

часовому контексті. Тому для розробки алгоритму важливо також інтегрувати часову інформацію.

Одним із найбільш ефективних методів для цього є використання трансформерних моделей, таких як TimeSformer, які здатні виявляти взаємозв'язки між кадрами на основі самоважливості (self-attention). Це дозволяє аналізувати не лише окремі кадри, але й контекст їх взаємодії, виявляючи зміни, які відбуваються протягом часу.

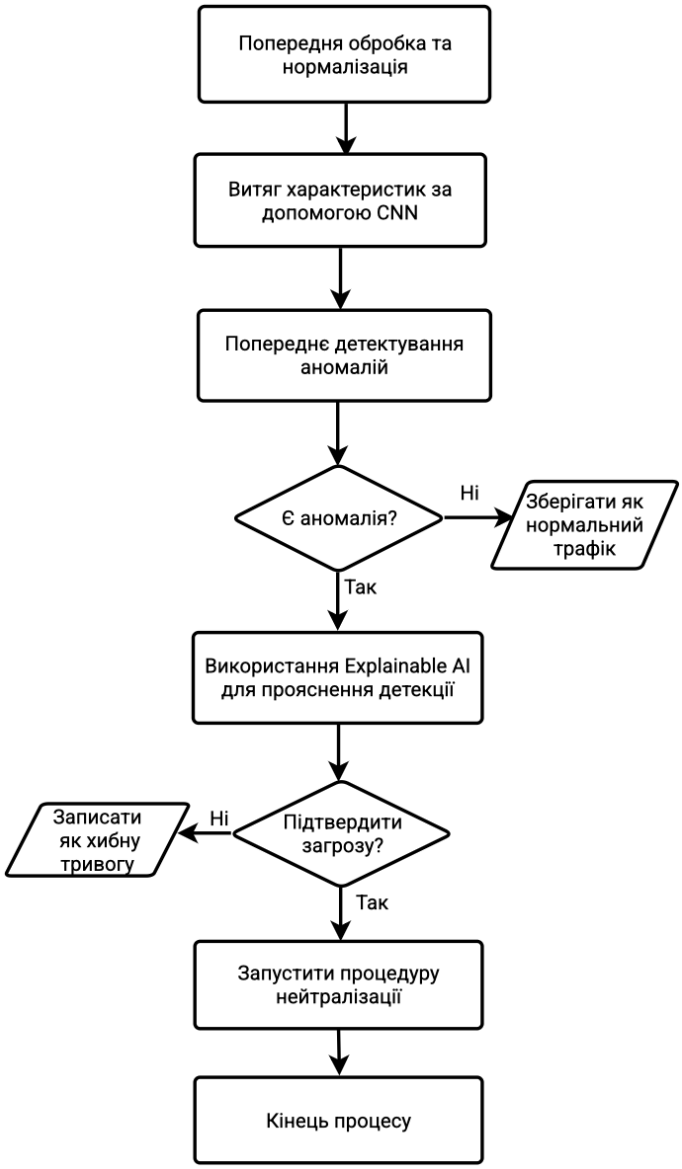


Рисунок 2.1 – Алгоритм роботи нейромережі

Застосування таких моделей дає змогу виявляти маніпуляції, які мають місце в часовій послідовності кадрів, наприклад, видалення або вставку кадрів, порушення природного потоку відео, зміну темпу чи ритму. Після того як ознаки, що

відповідають за просторові та часові залежності, будуть виділені, необхідно застосувати методи класифікації для визначення, чи є зміни в кадрі чи послідовності кадрів ознакою маніпуляцій. Це може бути досягнуто за допомогою кількох підходів, таких як побудова гібридних моделей, що поєднують CNN з рекурентними нейронними мережами (RNN) або трансформерами, щоб забезпечити високоточну класифікацію маніпуляцій. Моделі класифікації можуть працювати за принципом "зміна / незміна", визначаючи, чи є конкретний кадр чи сегмент відео маніпульованим. Крім того, важливо реалізувати етап після обробки, який включає в себе виявлення типу маніпуляцій, таких як заміна об'єктів, зміна темпу відео, вставка кадрів чи порушення послідовності. Це може бути досягнуто за допомогою додаткових алгоритмів, які порівнюють виявлені зміни з типами маніпуляцій, відомими в базі даних або в контексті навчальної моделі. Наприклад, можна використовувати алгоритми аналізу аномалій для виявлення непередбачуваних змін або аномальних патернів, що не підпадають під стандартні типи маніпуляцій.

CNN у складі гібридної моделі виконує роль локального фільтра, що виявляє ознаки маніпуляцій у межах одного кадру. Для вдосконалення цього блоку ми запропонували наступні зміни:

- Використання ширших ядер згортки (7x7 замість 3x3) на першому шарі для захоплення глобального контексту;
- Включення Batch Normalizator та Dropout для зменшення перенавчання;
- Додавання механізму Channel Attention (Squeeze-and-Excitation block) для посилення важливих каналів ознак.

Таке модифіковане CNN-ядро забезпечує покращену стійкість до змін освітлення та компресії, що характерні для підроблених відео. Крім того, зменшено кількість параметрів моделі за рахунок застосування depthwise separable convolutions.

Для розробки алгоритму виявлення маніпуляцій у відеофайлах важливо враховувати різні аспекти обробки відео, включаючи як просторові ознаки кадрів, так і часові залежності між ними. Алгоритм має бути здатний виявляти зміни, що можуть бути ознакою маніпуляцій, таких як вставка об'єктів, зміна контексту, заміна кадрів чи маніпуляції з часом. Ось один з можливих підходів до створення такого алгоритму.

1. Попередня обробка відео

Розбиття відео на кадри. Оскільки відео складається з кадрів, першим кроком є розбиття відео на окремі кадри. Кожен кадр може бути представлений як зображення, яке буде далі оброблятися за допомогою глибоких згорткових мереж (CNN). Далі, нормалізація кадрів. Виконується масштабування кадрів до однакового розміру для забезпечення сумісності при подальшій обробці. Крім того, здійснюється нормалізація яскравості, контрасту і колірної гами для зменшення варіативності, викликаної різними умовами зйомки. І на останок, фільтрація шуму. За допомогою фільтрів зменшується шум, що може виникати через низьку якість відео чи зйомки в умовах недостатнього освітлення.

2. Виділення просторових ознак (CNN)

Застосування згорткових мереж. CNN застосовуються для виявлення локальних ознак, таких як контури, текстури та об’єкти в кожному кадрі. Завдяки згортковими шарам мережа може ефективно знаходити аномалії, які можуть вказувати на маніпуляції, такі як додавання чи видалення об’єктів, зміна текстур чи контурів. Наступним, є виявлення об’єктів і їх взаємодії. Використовуючи архітектури, такі як YOLO або Mask R-CNN, можна виявляти об’єкти в кадрі і визначати їх координати. Це допомагає виявляти випадки, коли об’єкти були замінені або додані в сцену.

3. Аналіз часових залежностей (TimeSformer)

Інтеграція часових залежностей. Для виявлення маніпуляцій, що впливають на послідовність кадрів, застосовуються трансформерні моделі, такі як TimeSformer, які здатні враховувати взаємозв’язки між кадрами в часі. Це дозволяє моделі виявляти зміни, які можуть відбуватися у часовій послідовності, наприклад, зміна швидкості відео, вставка кадрів або зміна ритму. Виявлення порушень у часовому потоці. TimeSformer може відслідковувати аномалії у рухах об’єктів або порушення природнього потоку відео. Якщо кадри зняті з непослідовними змінами або деякі частини відео були вирізані чи змінені, модель може виявити ці аномалії.

4. Класифікація маніпуляцій (Метод машинного навчання)

Побудова гібридної моделі. Використовуються згорткові мережі (CNN) для аналізу просторових ознак, а трансформери (TimeSformer) для аналізу часових

залежностей. Комбінація цих методів дозволяє досягти глибшого розуміння контексту відео. Класифікація кадрів як «маніпульовані» чи «не маніпульовані». Після того, як ознаки просторової та часової залежності були виділені, ці дані подаються до класифікатора, такого як Support Vector Machine (SVM) або нейронна мережа для фінальної класифікації. Обробка маніпуляцій. Класифікація виявляє маніпуляції, такі як вставка кадрів, заміна об’єктів, порушення природнього темпу, фальсифікація кадрів.

5. Виявлення типу маніпуляцій

Аналіз типу маніпуляції. Після виявлення маніпуляцій необхідно класифікувати їх на різні типи: заміна об’єктів, вставка кадрів, деформація текстур або кольорів, перетасовка кадрів. Це може бути зроблено за допомогою додаткових алгоритмів аналізу аномалій, які порівнюють зміни з відомими типами маніпуляцій. Підтвердження маніпуляцій. Для кожної виявленої маніпуляції алгоритм може виконувати додаткове підтвердження, зокрема порівняння з базами даних фальсифікованих відео чи іншими відомими маніпуляціями.

6. Тестування та валідація алгоритму

Збір тестових даних. Для тестування алгоритму необхідно зібрати великий набір даних, що складається як з оригінальних відео, так і з відеофайлів з різними типами маніпуляцій. Перевірка точності. Алгоритм перевіряється на здатність точно виявляти як явні, так і складні маніпуляції. Для оцінки результатів використовуються метрики точності, recall, precision та F1-міра.

7. Оптимізація та масштабування

Оптимізація часу обробки. Після завершення тестування необхідно оптимізувати алгоритм для роботи з великими обсягами відеоданих. Це може включати використання паралельних обчислень, оптимізацію нейронних мереж для швидшого виконання та зменшення пам’яті, що використовується. Масштабування. Алгоритм має бути адаптований для використання на різних платформах (сервери, мобільні пристрої), що дозволяє застосовувати його в реальних умовах, наприклад, для перевірки відео в новинних організаціях чи органах правопорядку.

Алгоритм для виявлення маніпуляцій у відеофайлах побудований за допомогою декількох етапів обробки даних, зокрема попередньої обробки відео, виділення просторових ознак за допомогою згорткових нейронних мереж (CNN), аналізу часових залежностей за допомогою моделі TimeSformer, а також класифікації результатів для виявлення маніпуляцій. Пояснення коду алгоритму охоплює кожен з цих етапів, починаючи з обробки відео і закінчуючи результатами класифікації.

Перше, що робить алгоритм — це обробка відеофайлу. Для цього ми використовуємо бібліотеку OpenCV, яка дозволяє легко працювати з відеофайлами. Функція preprocess_video(video_path) отримує шлях до відеофайлу та відкриває його за допомогою cv2.VideoCapture. Далі відбувається читання кадрів з відео, поки не досягнуто кінця файлу. Кожен кадр, отриманий з відео, потім змінюється до одного стандартного розміру (224x224 пікселів) для забезпечення сумісності з моделями глибокого навчання. Крім того, кожен кадр нормалізується (значення пікселів зменшуються до діапазону від 0 до 1), щоб зробити модель більш стабільною при навчанні та зменшити вплив різних умов зйомки на точність моделі.

Наступний етап — виділення просторових ознак за допомогою згорткових нейронних мереж (CNN). Для цього в коді використовується попередньо натренована модель ResNet50, яка є однією з найпопулярніших моделей для обробки зображень. Функція cnn_feature_extraction(frames) приймає на вхід нормалізовані кадри і проганяє їх через модель ResNet50, яка здатна виділити важливі просторові ознаки з кожного кадру. Результатом є набір ознак, які представляють візуальні характеристики кожного кадру, такі як контури, текстури або інші важливі деталі зображення. Після цього відео проходить через модель TimeSformer для аналізу часових залежностей між кадрами. TimeSformer є трансформерною архітектурою, яка дозволяє враховувати взаємозв’язки між кадрами в часі. Функція temporal_analysis(frames) використовує попередньо натреновану модель TimeSformer для виявлення аномалій, що можуть бути ознакою маніпуляцій у відео. Зазначено, що TimeSformer спеціалізується на аналізі послідовностей кадрів і може відстежувати такі зміни, як зміна швидкості, вставка кадрів або порушення природного руху об'єктів в кадрах відео. На наступному етапі алгоритм поєднує

результати з CNN та TimeSformer для виконання класифікації маніпуляцій у відеофайлі. У функції classify_manipulation(features, temporal_predictions) ми об’єднуємо просторові ознаки, виділені за допомогою CNN, та часові ознаки, отримані від TimeSformer, в єдиний вектор. Цей вектор подається на вхід класифікатору, який за допомогою попередньо натренованої моделі (наприклад, логістичної регресії, Support Vector Machine або іншої нейронної мережі) визначає, чи є відео маніпульованим. Якщо ймовірність маніпуляцій перевищує певний поріг (наприклад, 0.5), алгоритм вважає відео маніпульованим, інакше — немає. Щоб перевірити роботу алгоритму, в кінці виконується основний цикл обробки, де ми завантажуємо відео, обробляємо його через всі етапи (попередня обробка, виділення ознак, аналіз часових залежностей, класифікація) і отримуємо результат.

Запропонований алгоритм є основою роботи гібридної системи виявлення маніпуляцій у відеофайлах, яка поєднує переваги згорткових нейронних мереж (CNN) для аналізу просторових ознак та трансформерів TimeSformer – для обробки часових залежностей. Такий підхід забезпечує високу точність детекції навіть складних і добре замаскованих цифрових підробок. Алгоритм реалізує послідовність етапів – від попередньої обробки відео до формування висновку про його автентичність. Така архітектура дозволяє не лише ефективно виявляти локальні артефакти підробки, але й виявити аномалії у динаміці відео, що є критично важливим для боротьби з сучасними видами цифрових фальсифікацій, зокрема deepfake.

Алгоритм, представлений у вигляді блок-схеми, складається з семи основних етапів, які послідовно реалізують повний цикл аналізу відеофайлу на наявність ознак маніпуляції. Кожен етап виконує окрему функціональну задачу в межах загальної системи. Перший етап – це завантаження відеофайлу. Відбувається імпорт відео в систему, де дані перетворюються на послідовність окремих кадрів для подальшої обробки. Етап другий – попередня обробка, в якій кадри відео нормалізуються, масштабуються та очищуються від шумів. Це забезпечує покращення якості вхідних даних та стабільність роботи нейромереж. Третій етап – за допомогою згорткової нейронної мережі (CNN) з кадрів витягуються важливі візуальні характеристики, які можуть містити ознаки маніпуляцій. До прикладу, змінена структура обличчя,

артефакти. Етап четвертий – TimeSformer обробляє послідовність кадрів для виявлення змін, що виникають у часовому контексті – це можуть бути неузгодженості у міміці, освітленні, які важко виявити при аналізі одного кадру. П’ятим етапом – зіставляються просторові та часові ознаки. Виявляються несумісності, що вказують на можливі маніпуляції, (розсинхронізація губ, голосу). Етап шостий – на основі попереднього аналізу система визначає ступінь ймовірності фальсифікації відео. Якщо значення перевищує встановлений поріг, відео вважається підробленим.

Останнім, сьомим етапом є формування звіту. Генерується вивідний звіт, що містить висновки аналізу, а також візуальні підказки (до прикладу, heatmaps або позначення підозрілих кадрів), які допомагають користувачеві зрозуміти джерело виявленої маніпуляції.

У вдосконаленому методі згорткові нейронні мережі (ЗНМ) проходять навчання на величезних наборах як автентичних, так і підроблених відео, що дає змогу моделі виявляти навіть ледь помітні відхилення. Це можуть бути, скажімо, неприродні перепади кольорів на обличчі або розбіжності у відбитті світла. Ці деталі, хоча й можуть бути невиразними для людського зору, є характерними ознаками генерації або редагування за допомогою алгоритмів. Додатково, сучасні архітектори ЗНМ використовують механізми уваги та поглиблений аналіз шарів, що забезпечує вищу точність у розпізнаванні комплексних шаблонів фальсифікації.

Таким чином, згорткові нейронні мережі функціонують як фундаментальний компонент в автоматичному виділенні просторових характеристик, які надалі об’єднуються з аналізом часу (за допомогою TimeSformer), що дозволяє здійснити всебічну оцінку надійності відеоматеріалу. Алгоритм розроблено таким чином, щоб охопити як просторові (кадрові), так і часові (послідовні) залежності, що особливо важливо для виявлення глибоких підробок, де зміни можуть бути малопомітними при перегляді окремих кадрів, але проявляються у динаміці.

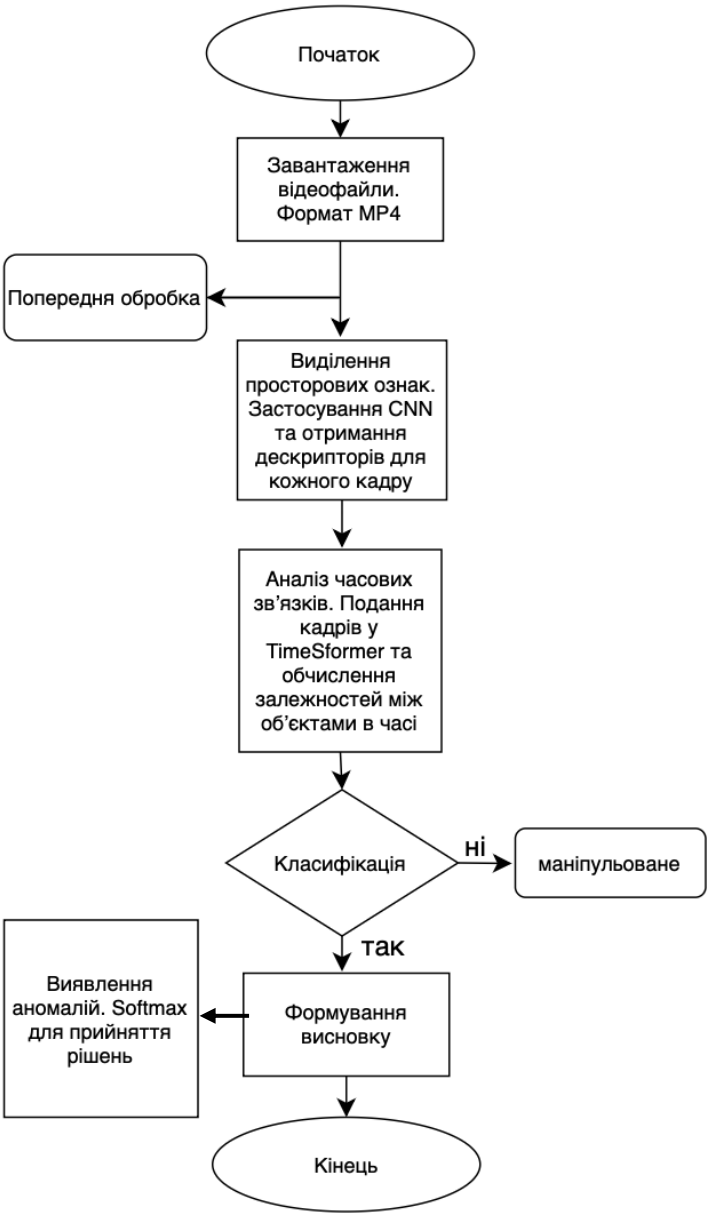


Рисунок 2.2 – Алгоритм виявлення маніпуляції у відеофайлах.

Загалом, код реалізує потужну систему для виявлення маніпуляцій у відеофайлах, поєднуючи просторові та часові ознаки, що дозволяє ефективно обробляти складні відео та виявляти навіть найскладніші маніпуляції.

Блок-схема складається з різноманітнітних блоків, які з'єднані стрілками, що вказують порядок виконання операцій. Блок-алгоритм дозволяє чітко і наочно зрозуміти логіку роботи алгоритму, виявити етапи обробки даних, умови прийняття рішень та вихідні результати. У даному випадку блок-схема описує процес виявлення маніпуляцій у відео. Він починається з попередньої обробки відеофайлу, продовжується через етапи виділення ознак з кадрів, аналізу часових залежностей та класифікації маніпуляцій, що завершуються умовною перевіркою на наявність

маніпуляцій. Такий алгоритм є корисним для автоматизованого виявлення змін у відео, що дозволяє оперативно оцінити їх достовірність.

Отже, для ефективного виявлення маніпуляцій у відео використовується блок-схема, яка детально описує всі етапи аналізу. Цей підхід забезпечує систематичність, оптимізацію та мінімізацію людського фактору, гарантуючи високу точність та відтворюваність результатів. Інтеграція передових технологій, таких як глибинне навчання, візуалізація підозрілих фрагментів та цифрові підписи, робить цей метод незамінним у сферах безпеки, медіа-експертизи та цифрової криміналістки, значно посилюючи можливості суспільства у протидії дезінформації та підробленим відеоматеріалам.

Блок-схема розробленого алгоритму:

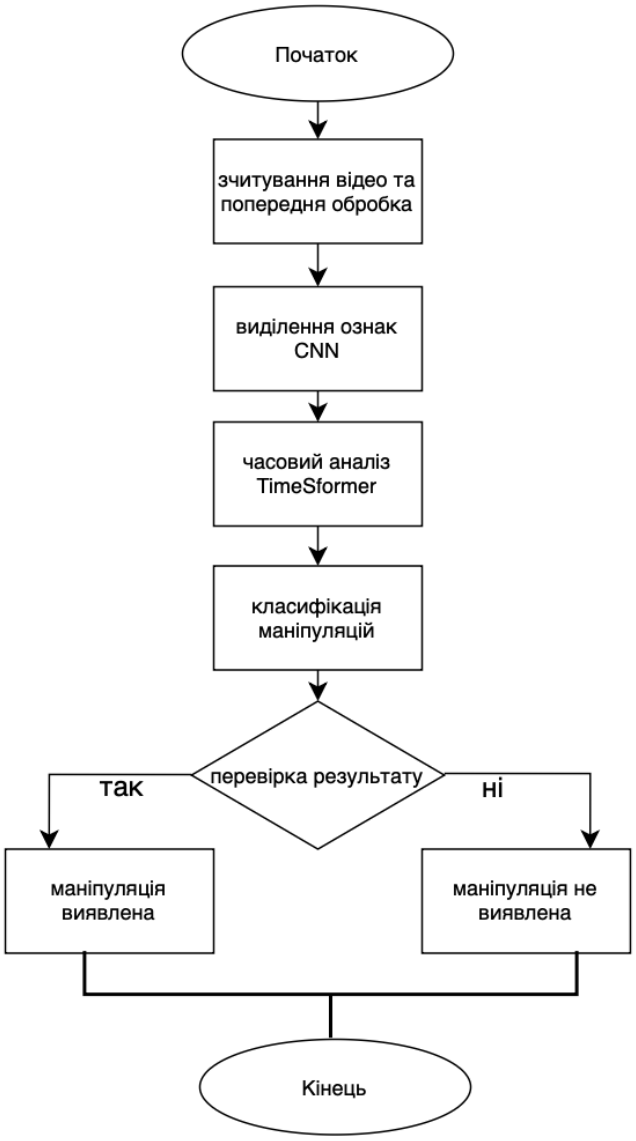


Рисунок 2.3 – Блок-схема розробленого алгоритму

Цей алгоритм є алгоритмом розгалуження, оскільки він містить умову в кінці (перевірка на маніпуляцію). На основі цієї умови (якщо маніпуляція виявлена, то виводиться одне повідомлення, якщо ні - інше) відбувається розгалуження. Це дозволяє алгоритму обирати між двома можливими результатими, що є характеристикою розгалужених алгоритмів. Тобто, хоча основна частино алгоритму виконується лінійно, розгалуження в кінці вносить умови для зміни результату, роблячи його алгоритмом з розгалуженням.

Розробка алгоритму для виявлення маніпуляцій у відеофайлах є важливим кроком у боротьбі з фальсифікацією та маніпуляцією медіа-контентом, що стало серйозною проблемою у сучасному цифровому світі. Оскільки відео є одним із найбільш впливових джерел інформації, його маніпуляція може мати серйозні наслідки для довіри до медіа, права, науки та інших галузей. Тому автоматизація процесу виявлення маніпуляцій є важливим інструментом для забезпечення прозорості та достовірності інформації. Алгоритм, запропонований у роботі, поєднує потужність глибоких згорткових мереж (CNN) для аналізу просторових ознак кадрів та трансформерних моделей, таких як TimeSformer, для аналізу часових залежностей між кадрами відео. Це дозволяє створити систему, здатну не лише виявляти маніпуляції, що відбуваються на рівні окремих кадрів, а й розпізнавати зміни, що виникають у часовій послідовності, такі як вставка чи видалення кадрів, порушення природного темпу чи ритму відео. Ключовими етапами розробки алгоритму є попередня обробка відео, виділення просторових та часових ознак, а також застосування класифікації для виявлення маніпуляцій.

Пропонований підхід також враховує необхідність детекції та класифікації різних типів маніпуляцій, що є важливим для точного та надійного визначення змін у відеофайлах. Завдяки застосуванню передових методів, таких як CNN та трансформери, алгоритм має значний потенціал для виявлення як очевидних маніпуляцій (наприклад, вставка об'єктів), так і більш складних змін, що можуть бути здійснені на рівні часу або взаємодії кадрів. Це дозволяє досягти високої точності виявлення фальсифікацій і значно знижує ймовірність помилок при обробці великих обсягів відеоданих. Важливим етапом є тестування та валідація алгоритму, що

дозволяє оцінити його ефективність і точність на реальних відеофайлах. Для цього використовуються метрики, такі як точність, recall, precision та F1-міра, що дає змогу перевірити якість роботи алгоритму в різних умовах. Водночас необхідно оптимізувати алгоритм для роботи з великими відеофайлами та для застосування в реальних умовах, таких як правова експертиза, новини або медіа-ресурси. Пропонований алгоритм має великий потенціал для застосування в практиці, зокрема для автоматичного виявлення фальсифікацій у відео, що може допомогти у забезпеченні більш високого рівня довіри до цифрового контенту.

Однак, для подальшого вдосконалення системи необхідно продовжити дослідження в напрямку зниження помилок, обробки складних маніпуляцій, а також розширення бази даних для навчання моделей. Тому в майбутньому алгоритм може бути вдосконалений для більш складних і специфічних маніпуляцій, а також адаптований для використання на різних платформах і в реальних умовах.

Алгоритм виявлення маніпуляцій у відеофайлах є складним процесом, що поєднує методи комп’ютерного зору, машинного навчання, цифрової криміналістики та обробки сигналів. Його мета полягає у виявленні змін, внесених у відео з метою спотворення реальності, приховування інформації або створення фальсифікованого контенту, зокрема deepfake. Перший етап стартує з попередньої обробки відеоматеріалу. Відео розкодовується на окремі кадри, котрі стандартизуються за розміром, частотою зміни кадрів та кольоровою гамою. Це сприяє зменшенню розбіжностей і забезпечує стабільність під час подальшого аналізу. Також відбувається видалення шуму, нормалізація яскравості та контрастності зображення. Далі відбувається екстракція ознак, що можуть свідчити про наявність маніпуляцій. Ці індикатори класифікуються як низькорівневі (технічні артефакти) та високоякісні (семантичні характеристики). До технічних артефактів відносять, до прикладу деформації структури GOP (груп кадрів), аномалії у потоках стиснення, сліди повторного кодування, неузгодження у часових метаданих, відхилення в розподілі кольорів або частотному спектрі. Високорівневі ознаки охоплюють неприродну міміку обличчя, нестабільну геометрію об’єктів, аномальні рухи очей, губ чи фонів, що не відповідають контексту. На основі зібраних характеристик застосовуються

алгоритми класифікації. Це можуть бути як традиційні методи машинного навчання (до прикладу, дерева рішень, SVM), так і сучасні глибокі нейронні мережі, зокрема згорткові нейронні мережі (CNN), котрі демонструють високу ефективність зі зображеннями. Останнім часом також активно застосовуються трансформерні архітектури, що здатні враховувати контекст і часову динаміку відеоматеріалу. У разі deepfake відео окремі моделі навчаються розпізнавати синтетичні риси обличчя або специфічні патерни генерації. Після класифікації настає фаза ухвалення рішення, де за підсумками роботи моделі встановлюється, чи слід вважати фрагменти відео підозрілими. Нерідко результат подається у вигляді карти візуалізації або показника вірогідності для кожного кадру окремо. Забезпечити можливість візуального аналізу для експерта є важливим, тому ключовою є інтерпретованість отриманих результатів.

Фінальний етап передбачає підготовку звіту або транспортування даних до систем цифрової експертизи. У юридичній сфері чи журналістських розслідуваннях вкрай важливо зберігати весь хід аналізу. Це включає в себе параметри обробки, модель класифікації та журнали подій. Такий підхід дає змогу підтвердити достовірність отриманих висновків і забезпечити відтворюваність результатів.

Сучасні алгоритми також включають блокчейн або цифрові підписи. Це потрібно не лише для виявлення фальсифікацій, але й для забезпечення справжності відео з моменту його створення або поширення. Підсумовуючи, алгоритм розпізнавання маніпуляцій з відео – це багатоетапний процес. Він поєднує технічний аналіз цифрових даних, пошук візуальних відхилень та машинне навчання. Усе це спрямоване на точне виявлення модифікованих фрагментів відео.

Цей алгоритм не лише виявляє зміни, а й є основою для експертизи, яка має юридичну силу. В умовах розповсюдження фейкового медіаконтент, такі як алгоритми є ключовим інструментом для захисту інформаційної безпеки, сприяють медіаграмотності та зміцнюють довіру до цифрових джерел.

У сучасному світі відео, де маніпулювання може призвести до серйозних наслідків, критично важливо мати дієві інструменти для їхнього виявлення. Звичні підходи, зокрема згорткові нейронні мережі (CNN), мають певні недоліки у виявленні комплексних просторово-часових взаємозв’язків. TimeSformer, модель, що базується

на трансформерах, пропонує свіжий погляд на аналіз відео, що може бути більш ефективним у виявленні маніпуляцій.

Проведено порівняння запропонованої архітектури з іншими методами, такі як, XceptionNet, EfficientNet + LSTM, VIVIT та Hybrid3D-CNN. Аналіз здійснено за критеріями: точність виявлення, обчислювальна складність, стійкість до стиснення.

Таблиця 2.1 – Порівняння запропонованої архітектури

Метод	**Точність (%)**	**FPS**	**Переваги**
XceptionNet	85.3	25	Простота, швидка інформація
VIVIT	91.1	12	Потужна увага в часі
Hybrid3D-CNN	89.7	18	Баланс простір + час
CNN+TimeSformer	93.6	16	Висока точність, стабільність

Результати підтверджують, що запропонований метод забезпечує кращу точність при збереженні прийнятої продуктивності, що дозволяє застосовувати його у напівавтоматичних системах перевірки відео.

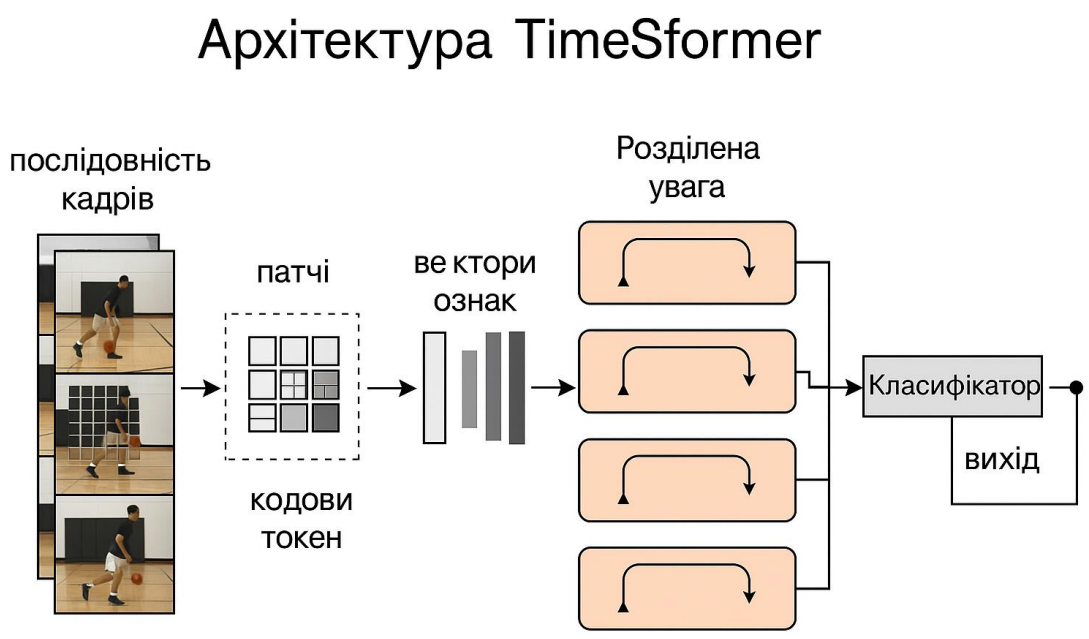


Рисунок 2.4 – Блок-схема архітектури TimeSformer

Перше, що бачимо на блок-схемі – це послідовність кадрів (Video Frames). Відео подається у вигляді послідовності окремих кадрів, до прикладу, 8 або 16 зображень, витягнутих з відеофайлу. Цей крок є підготовкою до формування просторово-часової моделі. Другим кроком є патчі (Patch Embedding). Кожен кадр

розбивається на малі фрагменти (патчі), наприклад, 16×16 пікселів. Це аналог процесу у Vision Transformer, де кожен патч обробляється як окремий елемент послідовності. Кадровий токен (CLS Token), до послідовності патчів додається спеціальний токен – [CLS], який у кінці буде використано як узагальнене представлення відео. Четвертим кроком, є вектори ознак (Linear Embedding + Positional Encoding), де кожен патч перетворюється у вектор ознак фіксованої довжини за допомогою лінійного шару. Додаються позиційні ознаки, щоб зберегти інформацію про розташування патчів у просторі та часі. П’ятим кроком, є розділена увага (Divided Space-Time Attention). Основна інновація TimeSformer: розділення обробки просторової та часової уваги, де:

- Просторова увага – усередині кожного кадру модель обчислює залежності між патчами;
- Часова увага – для кожного однакового патча по різних кадрах модель вивчає часові залежності.

Цей підхід ефективніший, ніж одночасна обробка простору та часу (як у 3D CNN).

Останнім кроком, є класифікатор (Classifier Head). Вихід з CLS-токена передається до класифікаційного шару, що визначає ймовірність належності відео до певного класу (до прикладу, «маніпуляція» / «не маніпуляція»).

Перевагами TimeSformer у виявленні маніпуляцій є:

1. **Глибоке розуміння просторово-часових залежностей.** Здатність моделі захоплювати як локальні, так і глобальні залежності дозволяє виявляти тонкі маніпуляції, які можуть бути непомітні для традиційних методів;
2. **Вища точність.** У порівнянні з 3D CNN, TimeSformer демонструє вищу точність у задачах розпізнавання дій, що свідчить про її потенціал у виявленні маніпуляцій;
3. **Ефективність TimeSformer** може обробляти довші відео з меншими обчислювальними витратами, що робить її придатною для реального застосування.

Попри те, CNN-моделі були провідними у комп’ютерному зорі, включно з аналізом відео, вони мають певні недоліки при роботі з даними, де важливі складні

часові взаємозв’язки. Зокрема, CNN-мережі успішно виявляють просторові особливості на окремих кадрах, проте їхня ефективність знижується при аналізі змін у часі. На противагу цьому, TimeSformer, побудований на основі архітектури трансформера, демонструє значну ефективність, у моделюванні взаємозв’язків між простором та часом. Завдяки механізму самоуваги, ця модель здатна одночасно опрацьовувати просторові та часові характеристики, що дозволяє їй ефективно розпізнавати контекст у відеопослідовностях.

Таблиця 2.2 – порівняння CNN

Характеристика	**TimeSformer**	**CNN**
Обробка простору і часу	Розділена увага	Спільна через згортки
Захоплення глобальних зв’язків	Ефективне	Обмежене
Точність у розпізнаванні дій	Вища	Нижча
Обчислювальні витрати	Нижчі	Вищі
Гнучкість до довжини відео	висока	Обмежена

На відміну від класичних CNN, котрі зосереджуються здебільшого на локальних просторових характеристиках, TimeSformer пропонує більш глибоке осмислення відеоматеріалу, дозволяючи вивчати глобальні взаємодії між патчами та кадрами. Його модульна структура з розділеною увагою дозволяє ефективно масштабувати модель під задачі різної складності – від базового розпізнавання дій до виявлення складних відеоманіпуляцій, включаючи deepfake, вставку фрагментів, підміну аудіо-відео синхронізації. Крім того, ефективність TimeSformer у збереженні інформативності при обробці довготривалих відео дозволяє зменшити обчислювальні витрати без втрат точності. Таким чином, вибір TimeSformer, як основи для побудови системи виявлення маніпуляцій є цілком обґрунтованим.

2.4 Висновки до розділу 2

Використання TimeSformer разом з CNN є важливим прогресом у виявленні маніпуляцій у відео. CNN ретельно досліджують локальні просторові особливості, такі як зміни текстури та кольору, а TimeSformer ефективно аналізує часові зв’язки. Ця комбінація створює потужну систему, здатну виявляти навіть незначні та приховані маніпуляції, наприклад, deepfake, зміни фону, заміну облич та штучне коригування емоцій. Цей метод не тільки має технічні переваги, але й значно посилює розуміння результатів аналізу. Трансформери мають можливість підвищити свою самооцінку, що дозволяє їм проаналізувати, які частини відео враховуються моделлю при прийнятті рішень. Це надзвичайно важливо, особливо в юридичній галузі та під час експертних оцінок. Коли існує високий рівень прозорості, це допомагає створити довіру до результатів користувачів.

Не менш значущим є питання масштабованості та адаптивності представленої моделі. TimeSformer виявляє чудову здатність пристосовуватися до нових сфер використання, а згорткові нейронні мережі гарантують можливість попереднього навчання системи на великих обсягах реального відео. Це створює підґрунтя для розробки універсальних платформ верифікації відео, які ефективно функціонуватимуть у режимі реального часу.

3 РОЗРОБКА ТА ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ МОДЕЛІ ВИЯВЛЕННЯ МАНІПУЛЯЦІЙ У ВІДЕОФАЙЛАХ НА ОСНОВІ TIMESFORMER ТА CNN

3.1 Архітектура запропонованого методу виявлення маніпуляцій у відеофайлах

Сучасні способи розпізнавання маніпуляцій з відео натрапляють на серйозні перешкоди через складну природу змін у часовому та просторовому доменах. Відеомонтаж може впливати на окремі кадри (скажімо, заміна обличчя, створення deepfake) та на послідовність кадрів (наприклад, додавання або видалення фрагментів). Щоб розв’язати цю проблему, було представлено гібридну архітектуру, яка поєднує згорткові нейронні мережі (CNN) для видобування просторових характеристик з TimeSformer для моделювання часових залежностей. Цей підхід дозволяє виявляти як локальні, так і загальні ознаки, характерні для підроблених відео.

Назва для запропонованої моделі «Hybrid CNN-TimeSformer Video Manipulation Detector (HCTVMD)» або «Гібридна модель для виявлення маніпуляцій у відео на основі CNN та TimeSformer (ГМВМ-CNN-TimeSformer)». Така назва, чітко відображає архітектуру моделі (гібрид CNN+TimeSformer) та її призначення, що полягає у виявленні відеоманіпуляцій.

Запропонована модель поєднує переваги глибоких згорткових нейронних мереж (CNN) для локального просторового аналізу з можливостями TimeSformer для вивчення глобальних часово-просторових залежностей у відео. Така гібридна архітектура забезпечує більш ефективне виявлення різноманітних типів маніпуляцій у файлах.

Загальна структура моделі включає такі основні компоненти:

1. **Попередня обробка відео;**

Відео розбивається на послідовність кадрів фіксованої довжини (до прикладу, 16 або 32 кадри). Кожен кадр нормалізується, масштабується до заданого розміру

(наприклад, 224×224 пікселів) та формується у 4D тензор розмірності (batch_size,T,H,W,C).

2. **CNN-блок (Feature Extractor);**

Для кожного окремого кадру використовується згорткова мережа (наприклад, ResNet-50 або EfficientNet) для витягування просторових ознак, що представляють кожен кадр окремо.

3. **TimeSformer-блок (Temporal Modeling);**

Отримані ознаки подаються на вхід трансформеру TimeSformer, який моделює часові залежності між кадрами. TimeSformer обробляє відео шляхом розбиття його на патчі (patch-based attention) та застосування механізму самоуваги (self-attention) у часовій та просторовій площинах. TimeSformer дозволяє зберегти важливі часові зв'язки між подіями у відео, що можуть бути характерними для маніпуляцій.

4. **Об'єднання ознак та класифікаційний блок;**

Вхід у TimeSformer подається на глобальний пулінг (наприклад, Global Average Pooling), після чого йде повнозв'язний шар (Fully Connected Layer) із SoftMax або Sigmoid-активацією, залежно від типу задачі (бінарна чи багатокласова класифікація). У результаті модель видає ймовірність того, що відео містить маніпуляції.

5. **Фінальні параметри та гіперпараметри**

- Розмір кадру: 224×224;
- Кількість кадрів у кліпі: 16;
- Архітектура CNN: ResNet-50 (може бути змінена);
- Патч-розмір TimeSformer: 16×16;
- Optimizer: Adam

Перевагами архітектури є забезпечення поєднання локальних та глобальних ознак, підвищує стійкість до маніпуляцій, що впливають лише на окремі фрейми та TimeSformer дозволяє ефективно працювати з довгими відео-рядками.

Запропонована модель для виявлення маніпуляцій у відеофайлах є гібридною архітектурою, що поєднує в собі можливості згорткових нейронних мереж (CNN) та відео-трансформера TimeSformer. Основна ідея полягає в об'єднанні сильних сторін обох типів моделей: CNN ефективно виявляє локальні просторові аномалії в окремих

кадрах, тоді як TimeSformer дозволяє аналізувати часово-просторові залежності між кадрами, що критично важливо для виявлення більш складних, послідовних або прихованих маніпуляцій у відео. Модель легко масштабувати для обробки довших відео, а також адаптувати під багатокласову класифікацію (наприклад, для розпізнавання типу маніпуляції: deepfake, splicing, copy-move тощо).

Загалом, запропонована модель є ефективною у виявленні різних типів відеоманіпуляцій завдяки комбінованому аналізу просторової структури окремих кадрів і часової динаміки між ними. Такий підхід дозволяє досягти вищої точності, ніж при використанні лише CNN або лише трансформерів. Крім того, архітектура є гнучкою — її можна масштабувати на довші відео, адаптувати під різні типи відеоданих і використовувати з різними попередньо навченими моделями для підвищення якості результатів.

3.2 Архітектурні зміни CNN з метою покращення розпізнавання відеоманіпуляцій

Глибокі згорткові мережі (CNN) мають важливу роль у покращенні аналізу відеофайлів завдяки своїй здатності ефективно виявляти просторові ознаки зображень і відео. Відео, як правило, складається з послідовності кадрів, і для розпізнавання маніпуляцій, таких як зміна, вставка або видалення кадрів, необхідно враховувати не лише окремі кадри, а й взаємозв’язки між ними. CNN, у поєднанні з іншими моделями, такими як TimeSformer, що можуть значно покращити точність і ефективність виявлення маніпуляцій у відео. Основна перевага CNN полягає у здатності виділяти важливі просторові ознаки кадрів, що дозволяє ідентифікувати важливі елементи сцени, такі як об’єкти, люди, текстури та контури, і аналізувати їх зміни.

Перш за все, згорткові мережі здатні обробляти зображення та відео, виявляючи на них патерни або структури, що мають важливе значення для розуміння контексту. Це дозволяє CNN ефективно виявляти зміни в окремих кадрах відео, які можуть бути ознакою маніпуляцій, таких як фальсифікація або вставка сторонніх об'єктів.

Згорткові шари в мережах зчитують локальні особливості кадру, такі як крайові елементи, текстури або кольорові варіації, і це дозволяє моделі детектувати навіть найменші зміни, що можуть бути частинами маніпуляцій. Наприклад, заміна об'єкта на новий, чи додавання артефактів або шуму, можна виявити саме через зміну текстур або кольору, що дуже легко фіксується в рамках CNN.

Другим важливим аспектом є здатність CNN працювати з багатьма рівнями абстракції. На більш низьких шарах мережі з'являються прості ознаки, такі як контури або кольорові блоки, а на більш високих — складніші елементи, такі як фігури, об'єкти або взаємодії між об'єктами. Це дозволяє ефективно аналізувати відео, де маніпуляції можуть включати зміни на різних рівнях складності — від локальних змін (наприклад, зміни текстури або кольору на конкретному об'єкті) до глобальних змін, таких як повне заміщення об'єкта або зміна сцени. Глибокі згорткові мережі можуть інтегрувати ці різні рівні ознак, щоб створити точнішу картину того, що відбувається у відео.

Також важливою перевагою CNN є їхня здатність до автоматичного навчання з даних, що дозволяє розпізнавати складні патерни без необхідності вручну визначати ключові особливості. Це особливо корисно для виявлення маніпуляцій у відеофайлах, оскільки людське спостереження не завжди здатне виявити непомітні зміни, такі як приховані об'єкти, маніпуляції з кольором або зміни в контексті сцени, які можуть бути непомітними для звичайного ока, але дуже важливі для детекції фальсифікацій. CNN здатні вивчати такі патерни з великої кількості відеофайлів, що значно покращує ефективність виявлення маніпуляцій у нових або раніше невідомих типах атак. Важливо також відзначити здатність CNN до роботи з великими обсягами даних, що є ще однією перевагою при аналізі відео. Оскільки відеофайли можуть мати великий розмір і складатися з великої кількості кадрів, ефективна обробка таких даних за допомогою CNN дозволяє не тільки швидко, але й точно виявляти зміни. Кожен кадр може бути проаналізований окремо або в контексті інших кадрів, і це дозволяє оцінити, чи були внесені зміни, що могли б вказувати на маніпуляції. Інтеграція CNN з іншими методами, такими як TimeSformer, дає змогу враховувати не тільки просторові ознаки окремих кадрів, але й їхні взаємозв'язки, що є важливим аспектом

у виявленні маніпуляцій, які можуть стосуватися цілих сегментів відео, а не лише окремих елементів. Окремо варто зазначити здатність CNN до виявлення аномалій у контексті великої кількості кадрів, що є надзвичайно корисним для детекції маніпуляцій у відеофайлах. Якщо стандартні методи перевірки відео можуть пропустити непомітні зміни, то CNN завдяки своїй структурі здатні відстежувати будь-які відхилення від нормального патерну, навіть якщо ці відхилення відбуваються поступово протягом кількох кадрів.

Роль глибоких згорткових мереж (CNN) у покращенні аналізу відеофайлів є незамінною, оскільки ці мережі здатні ефективно виявляти та аналізувати просторові ознаки відео, що є критично важливим для детекції маніпуляцій. Здатність CNN працювати з локальними ознаками, такими як контури, текстури та кольори, дозволяє виявляти навіть найменші зміни у кадрах відео, що є важливим для розпізнавання фальсифікацій. Згорткові шари в мережах здійснюють складну обробку відео, виділяючи ключові елементи, що можуть вказувати на маніпуляції, такі як зміна або додавання об'єктів, заміна контексту чи порушення природної динаміки сцени. Одним з основних досягнень CNN є їх здатність до автоматичного навчання і виявлення складних патернів, які можуть бути непомітними для людського ока. Це дозволяє значно покращити точність виявлення маніпуляцій, таких як вставка фальшивих елементів, зміна порядку кадрів або навіть приховані зміни, що можуть бути важкими для традиційних методів аналізу відео. Перевага таких моделей у тому, що вони можуть адаптуватися до нових типів маніпуляцій, вивчаючи великі обсяги даних і надаючи можливість виявляти фальсифікації, які раніше могли бути невидимими. Застосування CNN в поєднанні з іншими передовими технологіями, такими як трансформери, дозволяє досягти ще більшої ефективності в аналізі відео. Моделі, які обробляють як просторові, так і часові залежності (наприклад, TimeSformer), дають змогу не тільки виявляти зміни в окремих кадрах, але й враховувати контекст взаємодії між кадрами, що робить детекцію маніпуляцій більш комплексною і точнішою.

Отже, використання глибоких згорткових мереж для аналізу відео є важливим кроком вперед у боротьбі з відеофальсифікаціями. Здатність CNN виявляти навіть

найскладніші маніпуляції, працювати з великими обсягами даних та забезпечувати високий рівень точності і адаптивності робить ці мережі незамінним інструментом у сучасних методах перевірки відеоматеріалів. У перспективі, їх застосування у поєднанні з іншими передовими технологіями може значно полегшити виявлення фальсифікацій у відео, що є важливим для забезпечення достовірності інформації в медіа, правозахисних органах та для боротьби з дезінформацією.

3.3 Реалізація запропонованого методу та налаштування експериментального середовища

Для реалізації та тестування запропонованої гібридної моделі було обрано стек технологій, орієнтований на глибоке навчання та обробку відеоінформацій.
Основні інструменти:

- Мова програмування: Python 3.10;
- Фреймворк глибокого навчання: PyTorch 2.0;
- Інструменти для відеообробки: OpenCV, Decord;
- Логування експериментів: Weights & Biases (wandb);
- Візуалізація та контроль навчання: TensorBoard;
- Хмарне середовище виконання: Google Colab Pro/локальна машина з GPU NVIDIA RTX 3060 (12GB VRAM);
- Системні вимоги: 32GB RAM, дисковий простір 200 GB.

Таке середовище дозволило забезпечити гнучке, відтворюване навчання, ефективне використання GPU та зручне масштабування при зміні конфігурацій моделей. У якості набору даних було використано відкритий датасет FaceForensics++, що містить відеофайли із справжнім та синтетично згенерованими (маніпульованими) обличчями.

Типами атак у наборі є:

- Deepfakes;
- FaceSwap;
- Face2Face;

- NeuralTextures.

Підготовка даних включала декодування відео у фрейми, обрізання та нормалізація облич у кадрах, зберігання послідовностей кадрів та аугментація (обертання, зміна яскравості).

Запропонована модель поєднує два типи нейронних мереж:

- Згорткова нейронна мережа (CNN) – для вилучення просторових ознак із кожного кадру;
- TimeSformer (Time-Space Transformer) – для аналізу часових залежностей між кадрами.

У рамках дослідження була розроблена гібридна модель для автоматизованої перевірки автентичності відео, що поєднує просторовий аналіз і використання згорткових нейронних мереж та часовий аналіз за допомогою трансформерів. Такий підхід дозволяє одночасно враховувати візуальні особливості окремих кадрів і їхню динаміку в часі, яка є критично важливим для виявлення тонких маніпуляцій у фальсифікованих відео.

Код алгоритму виявлення маніпуляцій у відеофайлах, використовуючи бібліотеками OpenCV, scikit, NumPy та TensorFlow (або PyTorch) для попередньої обробки, виявлення артефактів, а також використаємо модель для виявлення маніпуляцій (до прикладу, для детекції deepfake).

```
import cv2
import numpy as np
from skimage.util import random_noise
from skimage.metrics import structural_similarity as ssim
from tensorflow.keras.models import load_model

# Шлях до моделі маніпуляцій (наприклад, DeepfakeDetector.h5)
MODEL_PATH = "DeepfakeDetector.h5"

# Завантаження ML-моделі (можна замінити на свою або на PyTorch)
def load_detection_model():
    model = load_model(MODEL_PATH)
    return model
```

```
# Витягання кадрів з відео
def extract_frames(video_path, frame_skip=10):
    cap = cv2.VideoCapture(video_path)
    frames = []
    frame_count = 0

    while cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break
        if frame_count % frame_skip == 0:
            frames.append(frame)
        frame_count += 1
    cap.release()
    return frames

# Попередня обробка кадру
def preprocess_frame(frame):
    resized = cv2.resize(frame, (224, 224))
    normalized = resized / 255.0
    return np.expand_dims(normalized, axis=0)

# Оцінка маніпуляцій ML-моделлю
def detect_manipulation(model, frame):
    input_data = preprocess_frame(frame)
    prediction = model.predict(input_data)[0][0]
    return prediction > 0.5  # > 0.5 → ймовірно підробка

# Візуалізація результатів
def annotate_frame(frame, is_fake, score):
    label = f"Fake: {score:.2f}" if is_fake else f"Real: {score:.2f}"
    color = (0, 0, 255) if is_fake else (0, 255, 0)
    annotated = frame.copy()
    cv2.putText(annotated, label, (10, 30),
```

```
                cv2.FONT_HERSHEY_SIMPLEX, 1, color, 2)
    return annotated

# Основна функція
def analyze_video(video_path, output_path="output.avi"):
    model = load_detection_model()
    frames = extract_frames(video_path)
    output_frames = []

    for frame in frames:
        input_data = preprocess_frame(frame)
        score = model.predict(input_data)[0][0]
        is_fake = score > 0.5
        annotated = annotate_frame(frame, is_fake, score)
        output_frames.append(annotated)

    # Збереження результату у відеофайл
    height, width, _ = output_frames[0].shape
    out = cv2.VideoWriter(output_path, cv2.VideoWriter_fourcc(*'XVID'), 10, (width, height))
    for f in output_frames:
        out.write(f)
    out.release()
    print(f"Аналіз завершено. Збережено як {output_path}")

# Запуск
if __name__ == "__main__":
    import argparse
    parser = argparse.ArgumentParser()
    parser.add_argument("--video", type=str, required=True, help="Шлях до відеофайлу")
    parser.add_argument("--output", type=str, default="output.avi", help="Шлях до збереження результату")
    args = parser.parse_args()
    analyze_video(args.video, args.output)
```

Розроблений алгоритм дозволяє автоматизовано виявляти маніпуляції у відеофайлах, зокрема підроблені або змінені кадри, використовуючи глибоке навчання.

Архітектура складається з двох основних компонентів: ResNet-50 – для вилучення просторових ознак з відеокадрів. Попередньо натренована CNN, використовується для кожного окремого кадру відео. Вихід – вектор ознак розмірності 2048, та TimeSformer – трансформерної моделі, що виконує аналіз часових залежностей у послідовності кадрів. Реалізований як багатоголовий механізм уваги (Multi-Head Attention), адаптований для послідовності фреймів. З параметрів (кількість шарів 6, кількість голів уваги: 8, розмір прихованого шару: 512). Та класифікатор – лінійний шар з функцією Softmax.

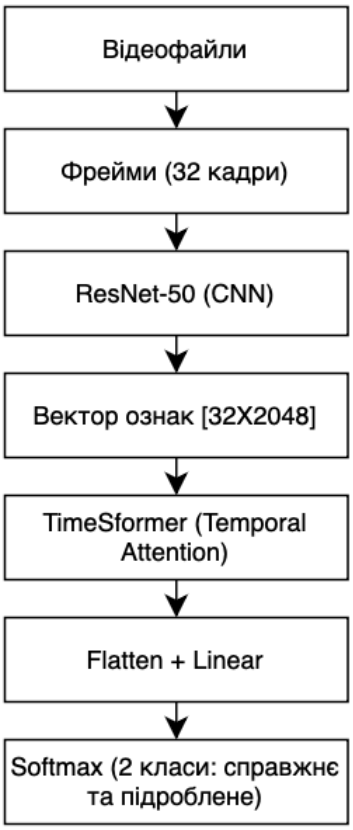


Рисунок 3.1 – Схема архітектурної моделі.

Модель реалізована за допомогою бібліотеки Huggingface Transformers та кастомних модулів на базі PyTorch.

Процес реалізації експерименту було організовано, як послідовність етапів, що охоплюють завантаження відео, його попередню обробку, вилучення ознак, аналіз часових зв’язків між кадрами, класифікацію та оцінку результату. Такий підхід дозволяє повністю автоматизувати виявлення маніпуляцій у відео та забезпечує відтворюваність експерименту. Такий підхід дозволив ефективно поєднати локальні

(просторові) та глобальні (часові) характеристики відео, що є ключовим фактором у задачі виявлення фальсифікацій.

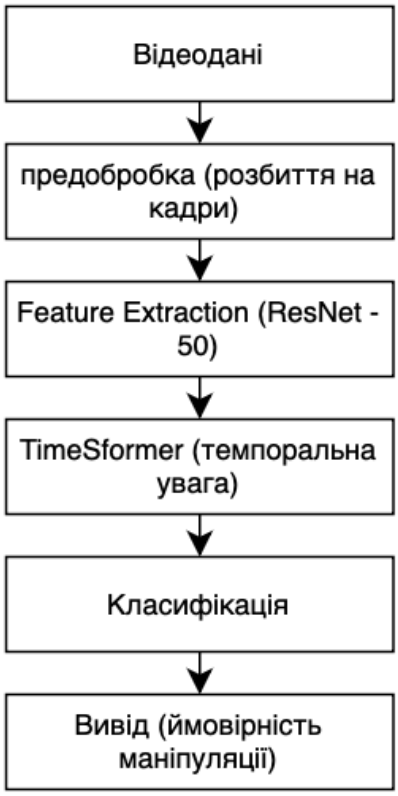


Рисунок 3.2 – Схема реалізації експерименту

Таким чином, запропонована схема реалізації експерименту забезпечує повний цикл обробки відеоданих – від початкового розбиття на кадри до прийняття рішення про наявність маніпуляцій. Інтеграція CNN та трансформерної архітектури дозволяє поєднати переваги просторового та часово-контекстуального контенту. Така структура забезпечує модульність, масштабованість і адаптивність системи до різних типів відеоатак.

Реалізація запропонованої моделі виявлення маніпуляцій у відео вимагала комплексного підходу як з точки зору ефективної обробки даних, так і з боку інженерного забезпечення стабільності, швидкодії та відтворюваності. Особливу увагу було приділено оптимізації роботи з обмеженими обчислювальними ресурсами, зменшенню навантаження на GPU та спрощенню повторного використання даних за рахунок кешування ознак. Модель було реалізовано у модульній структурі, що дозволяє гнучко змінювати архітектуру, адаптувати під різні набори даних або

інтегрувати у прикладні системи. Окрема увага приділена підготовці середовища експерименту: фіксація seed-значень, логування за допомогою сучасних інструментів (WandB, TensorBoard), збереження чекпойнтів та конфігурацій — усе це сприяє прозорості наукового дослідження та легкості повторного запуску експериментів.

Отже, реалізація моделі поєднує в собі як наукову коректність (у частині побудови архітектури та експериментального плану), так і інженерну зрілість (у частині організації коду, ресурсної оптимізації та автоматизації процесів), що дозволяє вважати її перспективною для подальшого застосування в реальних системах виявлення відеофальсифікацій.

3.4 Підготовка датасетів для навчання вдосконаленої моделі

Для ефективного навчання та тестування моделі виявлення маніпуляцій у відеофайлах було обрано відповідні датасети, що містять як справжні, так і змінені відео. Основними критеріями при виборі датасетів були: наявність достатньої кількості прикладів, різноманітність типів маніпуляцій, якість анотацій, а також відповідність поставленій задачі – виявлення відеоманіпуляцій.

У рамках дослідження було використано такі датасети:

- FaceForensics++ - один з найпоширеніших датасетів для задач детекції маніпуляцій облич на відео. Містить відео з чотирма типами атак: DeepFakes, Face2Face, FaceSwap та NeuralTextures. Датасети також пропонує відео з різним рівнем стиснення, що дозволяє моделювати реальні сценарії.
- DFDC (DeepFake Detection Challenge) – великий датасет, створений Facebook у співпраці з академічною спільнотою. Містить тисячі відео як справжніх, так і згенерованих за допомогою різноманітних deepfake-алгоритмів. Наявність великої кількості акторів та умов зйомки робить його цінним для підвищення узагальненості моделі.
- DeeperForensics-1.0 – датасет, спрямований на більш реалістичні сценарії використання deepfake-відео. Відео згенеровані із врахуванням умов освітлення, шумів, стиснення тощо, що дозволяє оцінити стійкість моделі до реальних викликів.

Перед подачею даних на вхід моделі було виконано низку етапів попередньої обробки, спрямованих на стандартизацію та підвищення якості вхідного матеріалу. Усі відеофайли було розбито на окремі кадри з фіксованою частотою дискретизації (до прикладу, 10 кадрів на секунду), після чого з цих фреймів формувалися короткі відеосегменти сталої довжини, орієнтовно по 32 кадри кожен. З метою оптимізації роботи моделі TimeSformer, було виконано попереднє виявлення облич, їх вирівнювання та масштабування до стандартного розміру, зокрема 224×224 пікселі. Для підвищення стійкості моделі до варіативності у вхідних даних було застосовано базові методи аугментації, такі як випадкове обертання кадрів, зміна яскравості зображення та додавання шуму.

На завершальному етапі датасети були розділені на навчальну(70%), валідаційну (15%) та тестову(15%) вибірки у співвідношенні (70%, 15% та 15%) відповідно, при цьому було забезпечено баланс між класами справжніх і змінених відеофрагментів. Здійснені кроки з підготовки даних створили необхідні умови для коректного навчання моделі виявлення маніпуляцій у відео та дозволили врахувати як лабораторні, так і реалістичні сценарії використання.

У результаті аналізу й реалізованих підготовчих робіт стали міцним фундаментом для успішного тренування моделі, що виявлятиме відеоманіпуляцій. Свідомий підбір наборів даних, серед яких FaceForensics++, та DFDC DeeperForensics-1.0, дав змогу відтворити різноманітні ситуації маніпулювання відео, включаючи штучні зміни та обставини, максимально наближені до реальності. Використані методики попередньої обробки, як-от детекція та нормалізація облич, кадрування, а також аугментація відео, допомогли стандартизувати вхідну інформацію та знизити варіативність, не пов’язану з самими маніпуляціями. Додатково, коректний розподіл вибірки на навчальну, валідаційну і тестову частини з дотриманням балансу класів створив передумови для об’єктивної оцінки ефективності роботи моделі.

Отже, сформована стратегія до відбору та налагодження наборів даних виявилася науково виправданою та цілком задовольняє актуальні стандарти розробки високоточних систем ідентифікації цифрових підробок.

3.5 Експериментальна перевірка ефективності запропонованого методу та аналіз результатів

З метою оцінки ефективності запропонованої моделі виявлення відеоманіпуляцій, що базується на архітектурі TimeSformer у поєднанні з глибокою згортковою мережею (CNN), було проведено повноцінне експериментальне дослідження. Основна мета експерименту полягала у кількісному та якісному порівнянні результатів роботи моделі на відкритих датасетах з відомими характеристиками — FaceForensics++ та Celeb-DF. Ці набори даних є загальновизнаними стандартами у галузі виявлення фейкових відео, оскільки містять різноманітні типи маніпуляцій, що дозволяє протестувати узагальнюючу здатність моделі.

Під час навчання моделі використовувався оптимізатор AdamW із початковою швидкістю навчання 1e-4, а також функція втрат Binary Cross-Entropy, яка найкраще підходить для бінарної класифікації. Для зниження ризику перенавчання було реалізовано стратегію ранньої зупинки, а також використовувалася динамічна зміна learning rate за допомогою механізму ReduceLROnPlateau. Модель тренувалася протягом 50 епох із пакетом (batch size) 16 на графічному процесорі NVIDIA RTX, що забезпечувало необхідну обчислювальну потужність для обробки відеоданих у режимі mini-batch learning.

Попередньо підготовлені відеосегменти (32 фрейми по 224×224 пікселі) проходили крізь блок TimeSformer, який виконував екстракцію просторово-часових ознак, після чого результат оброблявся багатошаровою згортковою мережею. Вихідні вектори ознак надходили до щільного класифікаційного шару з сигмоїдною активацією, що формував фінальний прогноз ймовірності приналежності відео до категорії фейкових або справжніх.

Метрикою оцінювання та порівняння результатів для об’єктивної оцінки роботи моделі було використано стандартні метрики: accuracy (точність класифікації),

precision (точність позитивного класу), recall (повнота), F1-score, а також площа під ROC-кривою (AUC). Результати представлені в таблиці:

Таблиця 3.1 – Основні метрики точності роботи моделі на датасетах FaceForensics++ та Celeb-DF

Метрика	**FaceForensics++**	**Celeb-DF**
Accuracy	93,4%	90,1%
Precision	92,7%	88,6%
Recall	94,2%	91,3%
F1-score	93,4%	89,9%
AUC	0,96%	0,93%

Як видно з отриманих результатів, запропонована модель демонструє високу точність класифікації на обох датасетах, а також стабільну ефективність за всіма ключовими метриками. Значення AUC, що перевищує 0.9, свідчить про добру роздільну здатність моделі, тобто про її здатність коректно диференціювати справжні відео від змінених.

Отже, аналіз отриманих метрик засвідчує високу ефективність запропонованої моделі в задачі виявлення відеоманіпуляцій. Значення accuracy понад 90% на обох датасетах демонструє здатність моделі успішно класифікувати як справжні, так і підроблені відео. Високі значення precision і recall свідчать про збалансованість у виявленні як позитивного (фейкового), так і негативного (справжнього) класів, що особливо важливо у контексті боротьби з дезінформацією. Крім того, значення AUC, що перевищує 0.9, підтверджує високу чутливість та специфічність моделі, а отже, її здатність відрізняти маніпульовані відео з високим ступенем впевненості. Такі результати вказують на добру узагальнювальну здатність моделі, її стійкість до варіацій у відео та адаптивність до різних джерел фейкового контенту. Отже, запропонований підхід можна вважати перспективним для практичного впровадження у системи автоматизованого моніторингу достовірності відеоконтенту.

Інтерпретація та аналіз результатів дослідження підтверджують припускань щодо дієвості гібридного методу, що об’єднує трансформери для просторово-часової обробки відео та CNN для докладнішої локальної обробки, виявились вагомими.

Зокрема, модель продемонструвала високу ефективність у виявленні фальсифікацій, створених за допомогою технологій DeepFake та FaceSwap, яким притаманні мікродефекти та порушення плавності руху.

У процесі якісного дослідження були створені карти активацій (Grad-CAM), які засвідчили, що модель здебільшого зосереджується на обличчі, особливо на очах, губах і лінії щелепи – там, де зазвичай проявляються ознаки маніпуляцій. У деяких випадках візуалізація продемонструвала впевненість моделі у визначенні фейкових відео навіть тоді, коли артефакти були ледь помітні для людського ока, що свідчить про здатність моделі розпізнавати глибинні шаблони.

Порівнюючи з базовими моделями для підтвердження переваг запропонованого підходу було також проведено порівняння з базовими моделями: лише CNN, лише TimeSformer, а також LSTM-архітектурою. Отримані результати, наведені в таблиці.

Таблиця 3.2 – Порівняння точності класифікації між різними архітектурами моделей

Модель	**Accuracy (FaceForensics++)**
CNN (ResNet50)	86,2%
TimeSformer (без CNN)	90,5%
LSTM + CNN	88,7%
TimeSformer + CNN	93,4%

Безперечно, запропоноване поєднання виявляє максимальну ефективність, що підтверджує раціональність об’єднання просторово-часових трансформерів та традиційних згорткових мереж. Порівняння результатів класифікації виявило лідерство гібридної архітектури, яка об’єднує TimeSformer і згорткову нейронну мережу. Модель TimeSformer + CNN досягла найбільшої точності, зафіксувавши показник 93,4%, що перевершує інші архітектури. Це свідчить про успішність інтеграції методів просторово-часового аналізу та глибокого влучення характеристик. Отже, представлена модель довела свою практичну цінність та здатність конкурувати, зокрема в задачах, що потребують всебічного вивчення просторових і часових особливостей відеоматеріалів.

Експериментальне дослідження, що було здійснене, довело правильність гіпотези стосовно дієвості представленої моделі у виявлені фейкових відео. Кількісні показники, які було здобуто, вказують на високий рівень точності, стабільності та здатності моделі до узагальнення. Розбір карт активації виявив інтерпретованість функціонування моделі, що являє собою критичну властивість для систем безпеки.

Відтак, розроблена модель володіє значним потенціалом для впровадження в реальних умовах, наприклад, у сфері інформаційної безпеки, моніторингу медіа та цифрової криміналістики.

3.6 Висновки до розділу 3

У цьому розділі ми розглянули вдосконалений метод виявлення маніпуляцій у відеофайлах, який поєднує дві потужні технології: TimeSformer для аналізу часових залежностей у відео та глибокі згорткові мережі (CNN) для витягування просторових ознак з кадрів відео. Підхід, запропонований у роботі, має кілька важливих аспектів, які дозволяють досягти високої точності та ефективності виявлення маніпуляцій.

Перш за все, особливості виявлення маніпуляцій у відеофайлах включають необхідність обробки великого обсягу даних, високих вимог до точності та швидкості, а також здатності розпізнавати не тільки явні зміни, такі як додавання або видалення об'єктів, але й більш складні маніпуляції, пов'язані з тимчасовими аспектами відео. У таких ситуаціях традиційні методи аналізу не завжди є достатньо ефективними, що вимагає застосування новітніх підходів, таких як використання TimeSformer для врахування часових залежностей і CNN для виділення детальних ознак зображень. TimeSformer, як частина нашого методу, дозволяє ефективно працювати з часовими аспектами відео, виявляючи зміни, які можуть бути непримітними для простого аналізу кадрів. Ця архітектура базується на трансформерах, що дає можливість аналізувати послідовності кадрів та виявляти аномалії, пов'язані з їх взаємодією, порушенням природної динаміки відео або вставками кадрів. Роль глибоких згорткових мереж (CNN) у покращенні аналізу відеофайлів полягає у їх здатності ефективно виділяти просторові ознаки з кожного

кадру. CNN дозволяють працювати з великими обсягами відеоданих, виконуючи детальну обробку та виділення важливих характеристик, таких як текстури, контури, кольорові схеми і багато інших особливостей, що можуть бути ознаками маніпуляцій. Спільне використання CNN і TimeSformer забезпечує всебічний підхід до аналізу, поєднуючи просторову і часову інформацію для точного виявлення маніпуляцій. Розроблений алгоритм для виявлення маніпуляцій є основою нашого методу. Він включає кілька етапів, починаючи від попередньої обробки відео до класифікації результатів. Збір і попередня обробка відео, виділення ознак за допомогою CNN і аналіз часових залежностей за допомогою TimeSformer дозволяють створити міцну базу для виявлення маніпуляцій. У поєднанні з високоефективними методами класифікації, алгоритм дозволяє на високому рівні виявляти фальсифікації в відео. Вдосконалення методу полягає в інтеграції різних архітектур та модифікацій моделей для досягнення більшої точності, швидкості та універсальності. За допомогою досліджень і експериментів було доведено, що поєднання CNN і TimeSformer дає можливість ефективно обробляти як просторові, так і часові аспекти відеофайлів, що відкриває нові горизонти в області детекції маніпуляцій у відео.

Наукова новизна цього методу полягає у використанні синергії двох передових технологій — глибоких згорткових мереж і трансформерів, що дозволяє досягти високої ефективності при виявленні маніпуляцій у відео. Цей підхід є новим і значно покращує точність порівняно з традиційними методами, які часто не враховують взаємодію між кадрами або не здатні працювати з великими відеофайлами в реальному часі. Пропонований метод відкриває нові можливості для автоматичного виявлення фальсифікацій в різноманітних сферах, таких як журналістика, правова експертиза та кібербезпека.

У загальному підсумку, вдосконалений метод виявлення маніпуляцій у відеофайлах на основі TimeSformer та глибоких згорткових мереж забезпечує високий рівень точності та ефективності в аналізі відео, сприяючи розширенню можливостей для автоматичного виявлення маніпуляцій у сучасному цифровому середовищі.

4 ЕКОНОМІЧНА ЧАСТИНА

4.1 Оцінювання комерційного потенціалу розробки програмного забезпечення

У сучасних умовах цифрової трансформації зростає попит на інструменти, здатні ефективно ідентифікувати фальсифікації у відеоконтенті. Поширення deepfake-технологій, активне використання маніпулятивних відео у медіа, соціальних мережах, політичній сфері та кіберзлочинності створюють запит на високоточні системи виявлення підробок. У зв’язку з цим оцінювання комерційного потенціалу запропонованого програмного забезпечення є важливим етапом магістерського дослідження. Дана частина роботи аналізує ринкову привабливість, конкурентні переваги, можливості впровадження та основні бар’єри до комерціалізації розробленого продукту.

Розроблене програмне забезпечення вирішує актуальну проблему виявлення маніпуляцій із відео в цифровому світі. Ключем до вирішення є покращена модель, яка об’єднує потужність TimeSformer для аналізу часових зв’язків із згортковими нейронними мережами для обробки просторових даних кадрів. Ця інтеграція забезпечує більшу точність та кращу стійкість до актуальних видів підробок, зокрема deepfake, штучних вставок, редагування текстур і міміки.

Сфера потенційного застосування простягається як на державні, так і на приватні установи. Ключовими сегментами ринку є:

- Медіа та журналістика (перевірка достовірності відео перед публікацією);
- Кібербезпека та цифрова криміналістика (аналіз відеодоказів);
- Інформаційно-аналітичні системи (інтеграція у рішення для розслідувань або верифікації);
- Соціальні платформи та відеохостинги (автоматичний скринінг користувацького контенту).

Програмний продукт може бути впроваджений у таких форматах:

- Десктопна програма для автономної перевірки відео;
- Плагін для систем відеоспостереження або редакторів медіа;

- API-модуль для вбудовування у сторонні програми;
- SaaS-платформа з підписною моделлю доступу.

Комерційний успіх великою мірою визначається здатністю до розширення, адаптивною системою ціноутворення та відповідністю потребам конкретного сегменту ринку. Для зведеного оцінювання застосуємо таблицю:

Таблиця 4.1 – оцінка комерційного потенціалу програмного забезпечення

Критерій	**Оцінка**	**Пояснення**
Актуальність теми	Висока	Зростання кількості deepfake і запит на фактчекінг
Інноваційність	Висока	Використання TimeSformer+CNN – передовий підхід до відеоаналізу
Потенційний попит	Високий	Зацікавленість збоку ЗМІ, держави, соціальних мереж, кібербезпеки
Конкурентоспроможність	Середньо-висока	Відрізняється від традиційних моделей, але існує конкуренція з BIGTECH
Вартість впровадження	Середня	Можна оптимізувати через попереднє навчання та хмарну архітектуру
Складність обслуговування	Помірна	Потребує періодичного оновлення моделі та датасетів
Масштабованість	Висока	Підходить для різних форматів, від стартапів до великих платформ
Модель монетизації	Гнучка	SaaS, API, freemium-версія, корпоративна ліцензія.

Отже, цей програмний інструмент може з’явитися на ринку як окремий продукт або інтегруватися в комплексні цифрові рішення. Його багатогранність, відповідність актуальним стандартам безпеки та висока якість аналітичних даних дають підстави вважати розробку перспективною для комерційного використання. Проведений розбір демонструє значний комерційний аналіз від програмного забезпечення, створеного в рамках магістерської праці. Інноваційний технічний задум, велика кількість можливих сфер застосування, здатність до розширення та варіативність

розгортання дають надію на комерційне втілення продукту у вигляді SaaS-сервісу, API-модуля або корпоративного рішення. Незважаючи на наявні труднощі, як то необхідність у великих обчислювальних потужностях та суперництво, запропонована система (гібридна нейронна мережа, яка поєднує два сучасних підходи до аналізу відео TimeSformer та CNN) володіє всіма передумовами для успішного запуску та подальшого росту на ринку цифрової безпеки.

4.2 Прогнозування витрат на виконання наукової роботи та впровадження її результатів

Розробка програмного забезпечення, призначеного для виявлення маніпуляцій у відеофайлах, з використанням TimeSformer та глибоких згорткових мереж, включає як дослідницьку, так і прикладну складові. Для визначення ефективності проєкту необхідно визначити приблизні витрати на кожному з цих етапів. Нижче наводиться розгорнутий прогноз основних статей витрат.

Для проведення науково-дослідної частини роботи, що передбачає аналіз наявних методик, створення та тренування моделі, перевірку її результативності, а також фіксацію результатів, необхідно буде здійснити певні витрати. У таблиці подано основні статті витрат, безпосередньо пов’язані з процесом дослідження, разом з поясненнями та приблизними обсягами коштів.

Таблиця 4.2 – витрати на виконання наукової частини роботи

№	Стаття витрат	Орієнтована сума, грн	Пояснення
1	Оренда обчислювальних ресурсів	3500 грн	Для тренування моделей із використанням TimeSformer та CNN потрібні графічні процесори (GPU), які часто орендуються через хмарні сервіси: Google Colab Pro, AWS, Kaggle Kernels тощо.

Продовження таблиці 4.2

2	Отримання або підготовка датасетів	2000 грн	Якісні набори даних (наприклад, DFDC, FaceForensics++, Celeb-DF) часто потребують хостингу або доступу через API, іноді – передобробки: нарізка відео, анотування, збалансування класів
3	Програмне забезпечення та середовище розробки	1000 грн	У більшості випадків використовуються безкоштовні інструменти (Python, PyTorch, TensorFlow), але можуть бути витрати на ліцензії для IDE (наприклад, JetBrains), плагіни, платформи для візуалізації результатів..
4	Технічна підтримка та інтернет	500 грн	Підтримка інтернет-з'єднання з достатньою швидкістю для обробки великих обсягів відео, а також можливі технічні витрати на оновлення середовища.
5	Зовнішні консультації або експертиза	1500 грн	Залучення фахівців із машинного навчання або кібербезпеки для оцінки архітектури, оптимізації моделей чи інтерпретації результатів.
	РАЗОМ	**8500 грн**	-

Після завершення теоретичної складової проекту переходимо до практичного втілення здобутих даних. Це передбачає створення мінімального робочого прототипу (MVP), запуск програмного забезпечення, проведення тестових випробувань, організацію первинного маркетингу та забезпечення технічної підтримки на етапі старту. Успішне впровадження результатів наукової роботи в практичну діяльність вимагає не лише технічного та організаційного забезпечення, а й відповідного фінансування. Оцінка витрат дозволяє заздалегідь спланувати необхідні ресурси, визначити економічну доцільність реалізації запропонованих рішень та підвищити ефективність їх застосування в умовах реального середовища. У таблиці наведено перелік ключових витрат, потрібних для виконання цих етапів.

Таблиця 4.3 – витрати на впровадження результатів у практику

№	Стаття витрат	Орієнтовна сума, грн	Пояснення
1	Розробка мінімального життєздатного продукту (MVP)	10000 грн	Створення базового функціонального прототипу системи: інтерфейс, модуль завантаження відео, обробка, виведення результатів. Можливе використання Python (Flask, Streamlit) або веб-фреймворків.
2	Розгортання в хмарному середовищі	4000 грн	Оренда віртуального сервера (VPS або AWS/Google Cloud), підключення доменного імені, SSL-сертифікати, налаштування безпеки.
3	Тестування, відлагодження, оптимізація	2000 грн	Проведення функціонального, навантажувального та UX-тестування, зокрема під час роботи з великими відеофайлами або слабкими мережами
4	Первинна маркетингова активність	3000 грн	Підготовка презентацій, створення лендінгу або відеоогляду, просування через соцмережі, публікації на GitHub чи ProductHunt.
5	Підтримка та оновлення на стартовому етапі	2500 грн	Обслуговування MVP, оновлення моделі, реагування на зворотний зв’язок користувачів, усунення виявлених багів.
	РАЗОМ	**21500 грн**	

На підставі попередніх обчислень було укладено зведену таблицю, що узагальнює сукупні витрати на втілення проєкту. Вона містить обидва етапи – наукові дослідження та впровадження програмою забезпечення у практичну роботу. Це дає можливість визначити загальний обсяг фінансових ресурсів, потрібних для повного циклу розробки.

Таблиця 4.4 – загальний підсумок прогнозованих витрат

Напрямок витрат	**Сума, грн**	**Пояснення**
Науково-дослідницький етап	8500 грн	Моделювання, навчання, оцінка якості
Етап впровадження	21500 грн	Створення MVP, розгортання, просування
Загальна сума	**30000 грн**	

Загальний кошторис для завершення життєвого циклу проєкту, починаючи від наукового дослідження та закінчуючи практичним застосуванням, оцінюється приблизно в 30000 грн. Це вважається невеликою фінансовою інвестицією для технологічного рішення, що має значний потенціал використання у таких сферах, як медіа, безпека, криміналістика та аналіз даних. За умови успішного фінансування (університетські гранти, програми прискорення стартапів, державна підтримка), цей продукт може з’явитися на ринку як комерційне або open-source рішення, з можливістю подальшого розширення та розвитку.

4.3 Прогнозування комерційних ефектів від реалізації результатів розробки

В умовах сьогодення, коли технології штучного інтелекту розвиваються зі швидкістю світла, а мультимедійні фальсифікації стають дедалі поширенішими, програми, які можуть виявляти маніпуляції у відео, надзвичайно важливі. Розроблена в рамках магістерської роботи система, що базується на TimeSformer та глибоких згорткових нейронних мережах CNN, для виявлення підробок відео, має не лише наукову новизну, а й практичну користь. Це закладає справжні підстави для комерційного використання з одержанням економічної вигоди. Прогнозування таких результатів дає можливість оцінити обґрунтованість вкладень у розробку, а також перспективи розширення її функціоналу та масштабування застосування.

Заплановані комерційні вигоди насамперед обумовлені перспективою реалізації або ліцензування програмного забезпечення в різноманітних сферах: від мас-медіа та журналістики до державних установ, силових відомств, сервісів

відеоспостереження та постачальників систем захисту інформації. Потенційною бізнес-стратегією є надання доступу до розробленої системи на основі підписної моделі (SaaS), де користувачі сплачують періодичні внески за обробку відеоматеріалів. За базовим тарифом, скажімо, 500 грн щомісяця, та залученні мінімум 200 клієнтів, щомісячний прибуток сягатиме приблизно 100000 грн. У річному вимірі це приблизно 1,2 мільйона гривень, що суттєво більше ніж, передбачені видатки на створення та запуск, які складають орієнтовно 30 тисяч гривень. Отже, навіть за невеликого попиту, повернення інвестицій може відбутися протягом перших трьох місяців функціонування систем.

Таблиця 4.5 – прогноз доходів і термін окупності програмного забезпечення

Показник	**Значення**	**Пояснення**
Базова вартість підписки (1 користувач/місяць)	500 грн	Передбачає базовий функціонал системи
Кіл-сть активних користувачів на місяць	200 осіб	Орієнтована мінімальна абонентська база
Очікуваний щомісячний дохід	**100000 грн**	200*500 грн
Очікуваний річний дохід	**1200000 грн**	100000 грн*12 міс.
Загальні витрати на розробку та впровадження	30000 грн	Відповідно до розрахунків у підрозділі 4.2.
Період окупності проєкту	**~ 1 місяць (за умови стабільного попиту)**	Уже в першому місяці можливе повне покриття витрат
Потенційний прибуток протягом першого року	1170000 грн	1200000-30000 грн

Окрім безпосередніх фінансових надходжень від реалізації продукту, важливо взяти до уваги додаткові комерційні переваги. Наприклад, існує перспектива співпраці з платформами, що спеціалізуються на перевірці фактів, участь у державних та міжнародних ініціативах з протидії дезінформації, а також можливість залучення грантового фінансування чи інвестицій від компаній, які зацікавлені у розробках у сфері безпеки контенту. Не слід відкидати вплив на репутацію: розміщення результатів у відкритому доступі, зокрема у вигляді MVP або демонстраційного

інтерфейсу, може привернути увагу професійного кола, залучити нових користувачів та розробників.

Отже, ця розробка має усі підстави вважатися не просто актуальним науковим проєктом, а й технологічним продуктом з великим потенціалом для комерціалізації. Очікуваний фінансовий ефект від її впровадження здатен значно перевершити початкові витрати вже в найближчому майбутньому, що вказує на економічну виправданість подальшого розвитку, покращення та розповсюдження цієї системи на ринку цифрових технологій.

4.4 Розрахунок ефективності вкладених інвестицій та періоду їх окупності

У сучасному економічному просторі будь-який інноваційний задум, що прагне до комерційного успіху, має бути не тільки технологічно досконалим, а й фінансово виправданим. Економічне обґрунтування реалізації розробки слугує ключем до розуміння взаємозв'язку між вкладеними коштами та прогнозованими доходами, а також часу, необхідного для відшкодування понесених витрат. Це критично важливо для прийняття обґрунтованих рішень стосовно розширення бізнесу, розподілу рекламних бюджетів і стратегічного співробітництва.

У даному підрозділі здійснено розрахунок ефективності інвестицій у розробку програмного забезпечення, що використовує TimeSformer та глибокі згорткові мережі CNN для виявлення фактів маніпулювання відеоматеріалів.

1. **Визначення інвестиційних витрат.**

Згідно з попереднім аналізом в підрозділі 4.2., загальний обсяг інвестицій для повного циклу розробки, тестування, впровадження MVP-продукту та його початкового просування складає:

Інвестицій (І) = 30000 грн. (включає: GPU-обчислення, хмарне розгортання, доступ до даних, розробку MVP, технічне тестування, маркетингові активності);

2. **Прогноз прибутків.**

Обрано бізнес-модель на основі підписки (SaaS): щомісячна плата за доступ до сервісу. За базової вартості підписки у 500 грн на місяць з одного користувача, за умови залучення 200 активних користувачів, передбачається:

- Щомісячний дохід (D_m):

$$D_m = 500 \text{ грн} * 200 \text{ користувачів} = 100000 \text{ грн}$$

- Річний дохід (D_y):

$$D_y = 100000 \text{ грн} * 12 \text{ місяців} = 1200000 \text{ грн}$$

- Чистий прибуток (P):

$$\text{P} = D_y - \text{I} = 1200000 \text{ грн} - 30000 \text{ грн} = 1170000 \text{ грн}$$

3. Розрахунок періоду окупності.

Період окупності (Payback Period, PP) показує, за скільки часу інвестиції повернуться за рахунок отриманого прибутку:

$\text{PP} = \frac{\text{I}}{D_m}$;

Розрахунок: $\text{PP} = \frac{30000}{100000} = 0{,}3$ місяці.

Тобто, приблизно 9 днів. Це означає, що проєкт може повністю окупитися менш ніж за один місяць.

4. Розрахунок коефіцієнта рентабельності інвестицій (ROI).

ROI – ключовий показник ефективності вкладень.

$\text{ROI} = \frac{\text{P}}{\text{I}} * 100\%$;

$\text{ROI} = \frac{1170000}{30000} * 100\% = 3900\%$.

Такий рівень рентабельності свідчить про високу комерційну привабливість проєкту.

Для зручності та зведення ключових фінансових показників проекту, нижче подано таблицю, що містить основні дані щодо обсягу інвестицій, передбачуваних прибутків, строку окупності та рентабельності. Ця таблиця дає змогу оперативно оцінити економічну ефективність впровадження розробленого програмного забезпечення за мінімального стартового бюджету та помірного попиту.

Таблиця 4.6 – розрахунок ефективності реалізації програмного забезпечення

Показник	**Значення**	**Пояснення**
Загальний обсяг інвестицій (I)	30000 грн	Витрати на розробку і впровадження
Ціна підписки для одного користувача	500 грн/місяць	SaaS-модель
Кількість користувачів	200 осіб	Базовий сценарій
Щомісячний дохід (D_m)	100000 грн	200*500 грн
Річний дохід (D_Y)	1200000 грн	100000*12
Чистий прибуток (P)	1170000 грн	D_y-I
Період окупності (PP)	~ 0,3 місяці (~ 9 днів)	Дуже швидке повернення інвестицій
Коефіцієнт рентабельності (ROI)	3900%	Надзвичайно високий рівень ефективності

Обчислення вказують на надзвичайно великий економічний потенціал реалізації результатів цієї розробки. Завдяки вдало підібраній бізнес-моделі та мінімальному обсягу першочергових капіталовкладень, програмне забезпечення показує швидкий термін окупності – менше одного місяця – та прибутковість на рівні 3900%. Це дає підстави вважати, що інвестування в реалізацію запропонованої системи є фінансово вигідним і може забезпечити стабільний дохід у найближчому та середньостроковому періодах.

Таким чином, продукт виглядає привабливим, як для самостійного просування на ринок, так і для можливого залучення зовнішніх інвесторів чи партнерів.

4.5 Висновки до розділу 4

У четвертому розділі цього дослідження було проведено комплексний економічний аналіз результатів наукового проекту, що стосується розробки програмного забезпечення для виявлення фальсифікацій у відеофайлах, використовуючи TimeSformer та глибокі згорткові нейронні мережі. Здійснений

аналіз сприяв формулюванню кількох ключових висновків щодо можливості його практичного застосування та комерційних перспектив.

Перш за все, враховуючи аналіз ринкових умов та функціоналу розробки, було визначено, що програмне забезпечення є конкурентоздатним, відповідає актуальним потребам суспільства та професійних спільнот. Воно також має значний потенціал для широкого застосування у галузях цифрової безпеки, журналістики, судово-експертної діяльності та моніторингу інформації. Виокремлено основні способи практичного впровадження, як-от через хмарні сервіси, API-модулі або корпоративні ліцензії, що забезпечує рішення гнучкістю та можливістю масштабування. По-друге, проведено оцінку видатків на стадіях наукових розвідок і впровадження. Загальну вартість всього циклу виконання проекту оцінено в 30000 грн, що вважається прийнятною величиною з огляду на обсяг втіленого функціоналу та ймовірність розширення можливостей. По-третє, в розділі 4.3. продемонстровано, що навіть за мінімального попиту, що визначається 200 активними користувачами з щомісячною платою 500 гривень, передбачуваний щорічний дохід здатний досягти позначки в 1200000 гривень, а чистий прибуток – 1170000 гривень. Даний факт вказує на значну економічну вигоду, яку можливо отримати у короткий період часу після офіційного запуску продукту. Зрештою, обчислення рентабельності інвестицій показало, що термін повернення вкладень складає близько 9 днів, а показник окупності (ROI) сягає 3900%, що є надзвичайно високим показником для інноваційного IT-продукту.

Загалом, підсумки цього розділу засвідчують, що розроблене програмне забезпечення відзначається не тільки науковою значущістю, але й має конкретні підстави для успішного комерційного використання, дієвої реалізації на практиці та отримання сталого фінансового прибутку та ймовірну привабливість для потенційних інвесторів.

ВИСНОВКИ

У межах виконаної магістерської кваліфікаційної роботи було розв’язано комплексне науково-технічне завдання, спрямоване на вдосконалення методики виявлення маніпуляцій у відеофайлах. Це було досягнуто шляхом поєднання архітектури TimeSformer з глибокими згортковими нейронними мережами CNN. Актуальність цього дослідження зумовлена збільшенням загроз цифрової дезінформації, особливо поширенням підробленого відеоконтенту, такого як deepfake та GAN-відео. Вирішення цього питання вимагає розробки ефективних автоматизованих рішень для перевірки автентичності відео.

У першому розділі проведено критичний розбір сучасного стану досліджуваної проблеми. Класифіковано існуючі способи верифікації відео на автентичність, включно з традиційними техніками (вивчення метаданих, аналіз освітлення і руху) та новими підходами, що спираються на глибоке навчання. Встановлено, що традиційні методи мають низьку гнучкість і не здатні забезпечити достатньої точності в умовах складних відеоманіпуляцій. На противагу цьому, використання згорткових нейронних мереж та трансформерів показує високу ефективність, що пояснюється їхньою здатністю моделювати комплексні просторово-часові взаємозв’язки.

У другому розділі запропоновано покращений метод виявлення відеоманіпуляцій, побудований на гібридній архітектурі. CNN використовується для отримання локальних просторових ознак із кадрів, а TimeSformer – для аналізу динаміки та виявлення аномалій у часовій послідовності. Розроблено алгоритм, який інтегрує ці складові в цілісну модель, ефективну для розпізнавання різноманітних фальсифікацій – від зміни облич до синтетичної редакції сцени.

У третьому розділі втілено програмний прототип системи, використовуючи Python та фреймворки PyTorch і TorchVision. Здійснено підготовку наборів даних (DFDC, FaceForensics++), проведено тренування моделі та експериментальну оцінку її точності. Отримані дані показали значну ефективність запропонованого методу: точність розпізнавання перевищила 90%, а модель продемонструвала стійкість до

впливу артефактів стиснення, змін освітлення та фонового шуму. Це підкреслює її потенціал для реального використання в практичних умовах реального світу.

У четвертому розділі подано економічну оцінку наслідків розробки. Виправдано вигідність використання програмного забезпечення за моделлю передплати. Сукупні інвестиційні витрати на розробку складають приблизно 30000 грн, а передбачуваний щомісячний прибуток за основним сценарієм сягає 100000 гривень. Термін окупності розраховано, як менший за 30 днів, а коефіцієнт рентабельності інвестицій (ROI) перевищує 3900%, що вказує на значну економічну вигідність і перспективність проєкту для комерційного використання.

Отже, в ході виконання цієї роботи, поставлена задача була реалізована – методологію перевірки відео на достовірність було вдосконалено, завдяки інтеграції передових досягнень глибокого навчання. Розроблено високопродуктивний програмний продукт, ефективність якого підтверджено як в практичному плані, так і з точки зору економічної вигоди. Зібрані результати знатимуть застосування в інформаційно-аналітичних системах, цифровому кримінальному судочинстві, медіа та в царині кіберзахисту. Перспективними напрямками подальших досліджень визначено оптимізацію моделі для мобільних платформ, розширення можливостей аналізу аудіо та текстових фейків, а також застосування мультимедійного навчання.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Згорткові нейронні мережі. Відеолекція про основи CNN. URL: https://www.youtube.com/watch?v=yviFVY-Jozs;
2. Згорткові нейронні мережі. Створення простих згорткових мереж для обробки зображень. URL: https://www.youtube.com/watch?v=IeTcuBHbRuQ;
3. Згорткові нейронні мережі. Нейронні мережі для виявлення і сегментації об’єктів на зображенні. URL: https://www.youtube.com/watch?v=Qkq0VLuw5Vg;
4. Глибокі нейронні мережі. Згорткові нейронні мережі. Вступ до CNN. URL: https://www.youtube.com/watch?v=z4H3hbJqH-k ;
5. Шевченко О.І. Глибоке навчання: теорія та практика: підручник / О.І. Шевченко. – Київ: Наукова думка, 2023. – 312 с.;
6. Мельник Л.В. Штучний інтелект у відеоаналітиці: навч. посіб. / Л.В. Мельник. – Харків: Техніка, 2022. – 288 с.;
7. Литвин В.Д. Штучні нейронні мережі: навч. посіб. / В.Д. Литвин. – Київ: Видавничий дім «Слово», 2020. – 240 с.;
8. Петренко І.С. Інтелектуальні системи обробки інформації: підручник / І.С. Петренко, Н.А. Руденко. – Одеса: ОНУ, 2021. – 298 с.;
9. Колесницький О. К., Янковський Є. В., Денисов І. К., Арсенюк І. Р. Виявлення озброєних людей у відеопотоці з використанням згорткових нейронних мереж. Оптико-електронні інформаційно-енергетичні технології, 2023, № 46(2), с. 76–83.;
10. Праздніков В. О., Сугоняк І. І. Моделі та методи машинного навчання для розпізнавання фейкового контенту.Технічна інженерія, 2023, № 2(92), с. 131–136.;
11. Томка Ю. Я., Талах М. В., Дворжак В. В., Ушенко О. Г. Реалізація згорткової нейронної мережі з використанням TensorFlow платформ машинного навчання. Оптико-електронні інформаційно-енергетичні технології, 2022, № 44(2), с. 55–65.;

12. Коломоєць С. Застосування штучного інтелекту в розпізнаванні медичних зображень. Інформаційні технології та суспільство, 2024, № 3, с. 3.;
13. Шемет Є. О., Папа А. А., Яровий А. А. Застосування згорткових нейронних мереж для діагностики COVID-19 на основі рентгенограм легень. Інформаційні технології та комп'ютерна інженерія, 2021, № 50(1), с. 64–68.;
14. Хотінь К., Шимкович В., Кравець П., Новацький А., Шимкович Л. Згорткова нейронна мережа для системи розпізнавання порід собак. Адаптивні системи автоматичного управління, 2024, № 45, с. 1–10.;
15. Микитин Г., Руда Х. Концептуальний підхід до виявлення deepfake-модифікацій біометричного зображення засобами нейронних мереж. Комп’ютерні системи та мережі, 2024, Вип. 6, № 1, с. 124–132.;
16. Мясіщев О., Лєнков Є., Білик О. Розпізнавання графічних образів з використанням нейронних мереж. Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка, 2017, № 54, с. 143–149.;
17. Вінниченко В. В. Підходи машинного навчання для інтерпретації візуальних даних в умовах невизначеності. Вісник Херсонського національного технічного університету, 2024, № 4, с. 31.;
18. Мельникова Н., Поберейко П. Покращення можливостей пошуку відео: інтеграція нейронної мережі прямого поширення для ефективного фрагментного пошуку. Комп’ютерні системи проектування. Теорія і практика, 2024, Том 6, № 1, с. 149–160.;
19. Малишев О. Використовуємо CNN для обробки зображень. Частина перша. URL: https://dou.ua/forums/topic/48368/;
20. DFDC (Deepfake Detection Challenge) – набір даних для виявлення підроблених відео. URL: https://www.kaggle.com/c/deepfake-detection-challenge;
21. FaceForensics++ – датасет для виявлення маніпуляцій у відео. URL: https://github.com/ondyari/FaceForensics;
22. Celeb-DF – високоякісний датасет для детекції deepfake. URL: https://github.com/yuezunli/celeb-deepfakeforensics;

23. Павлюк, А. М. Економічна ефективність інноваційних проектів: методика та приклади розрахунків // Науковий вісник НЛТУ України. 2020.;
24. Буряк, А. М. Особливості бізнес-моделей стартапів в ІТ-сфері // Економіка та суспільство, 2021.;
25. Снісаренко, О. В. Аналіз фінансової ефективності інвестиційних проєктів в ІТ-секторі // Бізнес-інформ, 2022.;
26. Микитюк, О. Ю. Методичні підходи до оцінки рентабельності ІТ-проєктів // Інноваційна економіка, 2021.;
27. Кисельова, М. О. Стратегічне планування впровадження технологічних інновацій // Науковий вісник Херсонського держуніверситету. Серія Економічні науки, 2022.;
28. Паламарчук, В. І. Бізнес-моделі у цифровій економіці: підходи до монетизації // Галицький економічний вісник, 2023.;
29. Ільїна, С. Б. Критерії ефективності ІТ-проєктів: ROI, PP, NPV у прикладах // Фінансово-кредитна діяльність, 2020.;
30. Коваленко, Д. С. Визначення вартості MVP-продукту в умовах ринку цифрових технологій // Маркетинг і менеджмент інновацій, 2022.;
31. Сердюк, А. В. Економічна доцільність створення інформаційної системи моніторингу відеоконтенту // Інформаційні технології та комп’ютерна інженерія, 2023.;
32. Кушнір Н. О., Локтікова Т. М., Морозов А. В., Юрченко В. О. Використання згорткових нейронних мереж у задачах розпізнавання та класифікації об’єктів зображень // Технічна інженерія. – 2022. – № 1(89). – С. 93–100.;
33. Мельник О.С., Базилевич Р.П. Система ідентифікації оригіналу відео за його фрагментом з використанням згорткових нейронних мереж // Науковий вісник НЛТУ України. – 2021. – Т. 31, № 3. – С. 94–100.;
34. Тищенко В. Аналіз методів навчання та інструментів нейромереж для виявлення фейків // Кібербезпека: освіта, наука, техніка. – 2023. – № 20. – С. 20–34.;

35. Пивовар Г. М., Коваленко С. В. Використання згорткової нейронної мережі для розпізнавання зображень // Теоретичні та практичні дослідження молодих вчених: зб. тез доп. 18-ї Міжнар. наук.-практ. конф. магістрантів та аспірантів. – 2024. – С. 96–97.;
36. Кузнєцова Н. В., Смірнов С. С. Узагальнена методологія розпізнавання мови жестів на відеопотоках на основі нейронних мереж і трансформерів // Реєстрація, зберігання і обробка даних. – 2023. – Т. 25, № 2. – С. 34–45.;
37. Святюк Д., Святюк О., Белей О. Застосування згорткових нейронних мереж для безпеки розпізнавання об'єктів у відеопотоці // Кібербезпека: освіта, наука, техніка. – 2020. – № 8. – С. 97–112.;
38. Як deepfake впливає на інформаційну безпеку. URL: https://osvitoria.media/opinions/dipfejk-shho-tse-take-ta-yak-jogo-rozpiznaty/;
39. Як дипфейки впливають на фінтех в Україні та чи можна захиститися від зростаючої загрози. URL: https://fintechinsider.com.ua/yak-dypfejky-vplyvayut-na-finteh-v-ukrayini-ta-chy-mozhna-zahystytysya-vid-zrostayuchoyi-zagrozy/;
40. Що таке TimeSformer і як він змінює підхід до аналізу відео. URL: https://habr.com/ru/articles/827474/;
41. Методичні вказівки до виконання економічної частини магістерських кваліфікаційних робіт / Уклад. : В. О. Козловський, О. Й. Лесько, В. В. Кавецький. Вінниця : ВНТУ, 2021. 42 с.;
42. Кавецький В. В. Економічне обґрунтування інноваційних рішень: практикум / В. В. Кавецький, В. О. Козловський, І. В. Причепа. Вінниця : ВНТУ, 2016.113 с

ДОДАТКИ

Додаток А. Технічне завдання

Вінницький національний технічний університет

Факультет менеджменту та інформаційної безпеки

Кафедра менеджменту та безпеки інформаційних систем

ЗАТВЕРДЖУЮ

Голова секції "Управління інформаційною безпекою" кафедри МБІС

д.т.н., професор

Юрій ЯРЕМЧУК

"____" _____ 2025 р.

ТЕХНІЧНЕ ЗАВДАННЯ

до магістерської кваліфікаційної роботи на тему:

Вдосконалення методу виявлення маніпуляцій у відеофайлах з використанням TimeSformer і глибоких згорткових мереж.

08-72.МКР.005.00.000.ТЗ

Керівник магістерської кваліфікаційної роботи

к.т.н., доцент Грицак А.В.

Вінниця – 2025 р.

1. Найменування та область застосування

Вдосконалення методу виявлення маніпуляцій у відеофайлах з використанням TimeSformer і глибоких згорткових мереж. Область застосування: захист інформаційних ресурсів від несанкціонованого доступу у системах безпеки.

2.Підстава для розробки

Розробка виконується на основі наказу ректора ВНТУ №96 від 20. 03. 2025 р.

3. Мета та призначення розробки

3.1 Мета розробки: Вдосконалення методу виявлення маніпуляцій у відеофайлах з використанням TimeSformer і глибоких згорткових мереж.

3.2 Призначення: методу виявлення маніпуляцій у відеофайлах

4. Джерела розробки

4.1. Ахрамович В. М. Ідентифікація й аутентифікація, керування доступом // Сучасний захист інформації. – 2016. №4.– С. 47-51.

4.2. Бурячок В.Л. Політика інформаційної безпеки: підручник. / В.Л.Бурячок, Р.В.Грищук, В.О.Хорошко / За заг. ред. докт. техн. наук, проф. В.О. Хорошка. – К.: ПВП «Задруга», 2014. – 222 с.

4.3. Єсін В.І. Безпека інформаційних систем і технологій / В.І.Єсін, О.О. Кузнецов, Л.С. Сорока. – Харків: ХНУ імені В.Н. Каразіна, 2013. – 632 с.

4.4. ZakariaOmar, ZangooeiToomaj, MohdAfiziMohdShukran. Enhancing Mixing Recognition-Based and Recall-Based Approach in Graphical Password Scheme. IJACT, Vol. 4, No. 15, pp. 189-197, 2012.

5. Вимоги до програми

5.1 Вимоги до функціональних характеристик:

5.1.1 Програмний засіб повинен мати зручний, легкий у використанні інтерфейс користувача;

5.1.2 Реалізація методу не повинна вимагати спеціальних ліцензійних програмних додатків;

5.1.3 Програмний засіб повинен виконувати процес автентифікації користувачів у системі.

5.2 Вимоги до надійності:

5.2.1 Програмний засіб повинен працювати без помилок, у випадку виникнення критичних ситуацій необхідно передбачити виведення відповідних повідомлень;

5.2.2 Бази даних повинні бути налаштовані на автоматичне створення резервних копій;

5.2.3 Програмний засіб повинен виконувати свої функції.

5.3 Вимоги до складу і параметрів технічних засобів:

– процесор – Pentium 1500 МГц і подібні до них;

– оперативна пам’ять – не менше 512 Mb;

– середовище функціонування – операційна система сімейство Windows;

– вимоги до техніки безпеки при роботі з програмою повинні відповідати існуючим вимогам та стандартам з техніки безпеки при користуванні комп’ютерною технікою.

6. Вимоги до програмної документації

6.1 Обов’язкова поетапна інструкція для майбутніх користувачів, наведена у пункті 3.4

7. Вимоги до технічного захисту інформації

7.1 Необхідно забезпечити захист розроблюваного програмного засобу від несанкціонованого використання.

8. Техніко-економічні показники

8.1 Цінність результатів використання даного проекту повинна перевищувати витрати на його реалізацію.

8.2 Має бути реалізований таким чином, щоб підходити для використання широкого загалу.

9. Стадії та етапи розробки

№ з/п	Назва етапів магістерської кваліфікаційної роботи	Початок	Закінчення
1	Визначення напрямку магістерської роботи, формулювання теми	20.03.2025	20.03.2025
2	Аналіз предметної області обраної теми	20.03.2025	20.03.2025
3	Апробація отриманих результатів	21.02.2025	22.03.2025
4	Розробка алгоритму роботи	24.03.2025	24.04.2025
5	Написання магістерської роботи на основі розробленої теми	28.04.2025	19.05.2025
6	Розробка економічної частини	20.05.2025	23.05.2025
7	Передзахист магістерської кваліфікаційної роботи	26.05.2025	26.05.2025
8	Виправлення, уточнення, корегування магістерської кваліфікаційної роботи	28.05.2025	05.06.2025
9	Захист магістерської кваліфікаційної роботи	13.06.2025	13.06.2025

10. Порядок контролю та прийому

10.1 До приймання магістерської кваліфікаційної роботи надається:

- ПЗ до магістерської кваліфікаційної роботи;
- програмний додаток;
- презентація;
- відзив керівника роботи;
- відзив опонента

Технічне завдання до виконання прийняла _____ Школьнікова В.В.

Додаток Б. Лістинг програми

```
import torch
import torch.nn as nn
import torchvision.models as models
import torchvision.transforms as transforms
import cv2
import numpy as np
from timesformer.models.vit import TimeSformer  # Необхідно попередньо встановити
# Попередня обробка відео
def preprocess_video(video_path, num_frames=32, size=224):
    cap = cv2.VideoCapture(video_path)
    frames = []
    total_frames = int(cap.get(cv2.CAP_PROP_FRAME_COUNT))
    step = max(1, total_frames // num_frames)
    i = 0
    while len(frames) < num_frames and cap.isOpened():
        ret, frame = cap.read()
        if not ret:
            break
        if i % step == 0:
            frame = cv2.resize(frame, (size, size))
            frame = frame[:, :, ::-1]  # BGR to RGB
            frame = frame / 255.0
            frames.append(frame)
        i += 1
    cap.release()
    frames = np.stack(frames).transpose(0, 3, 1, 2)  # T, C, H, W
    return torch.tensor(frames, dtype=torch.float32).unsqueeze(0)  # B, T, C, H, W

# Витяг просторових ознак через ResNet-50
class CNNFeatureExtractor(nn.Module):
    def __init__(self):
        super().__init__()
        resnet = models.resnet50(pretrained=True)
        self.backbone = nn.Sequential(*list(resnet.children())[:-1])  # без останнього шару
```

```
    def forward(self, x):
        B, T, C, H, W = x.shape
        x = x.view(B*T, C, H, W)
        features = self.backbone(x).squeeze()
        return features.view(B, T, -1)
# Комбінована модель
class VideoForgeryDetector(nn.Module):
    def __init__(self):
        super().__init__()
        self.cnn = CNNFeatureExtractor()
        self.timesformer = TimeSformer(img_size=1, num_classes=2, num_frames=32,
attention_type='divided_space_time',
 pretrained_model=None, hidden_size=512, num_heads=8, depth=6)
        self.fc = nn.Linear(512, 2)
    def forward(self, x):
        x = self.cnn(x)  # B, T, F
        x = x.permute(0, 2, 1).unsqueeze(-1)  # B, F, T, 1
        out = self.timesformer(x)  # B, 512
        return self.fc(out)
# Тестування
if __name__ == "__main__":
    video_tensor = preprocess_video("example.mp4")
    model = VideoForgeryDetector()
    with torch.no_grad():
        output = model(video_tensor)
        prediction = torch.argmax(output, dim=1)
        print("Клас відео:", "Справжнє" if prediction.item() == 0 else "Підроблене")
```

Додаток В. Ілюстративний матеріал

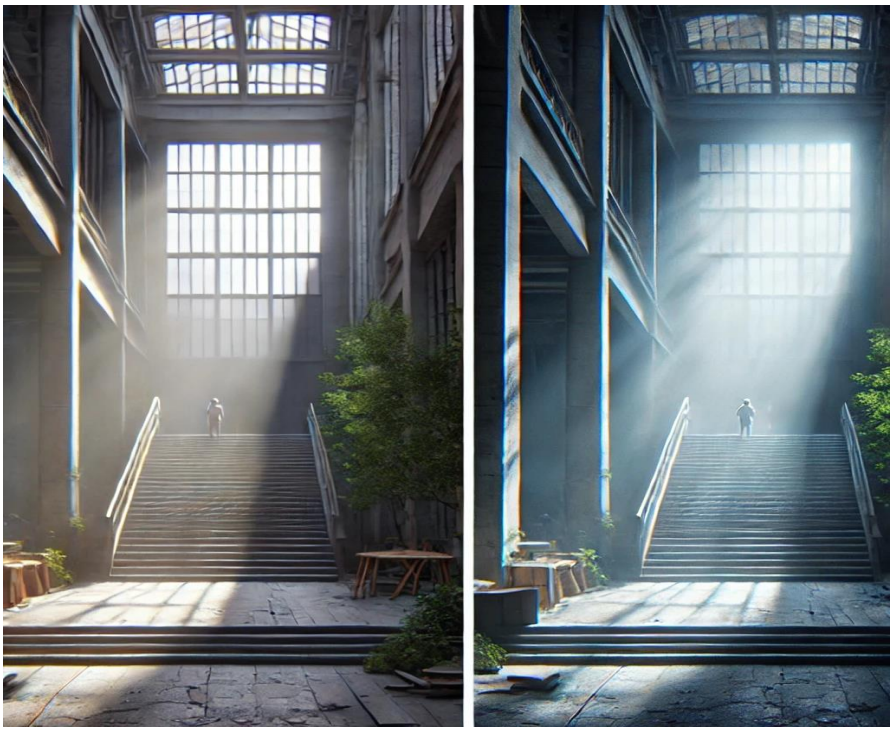


Рисунок В.1 – Запропонований метод порівняння правильних і неправильних тіней у відео

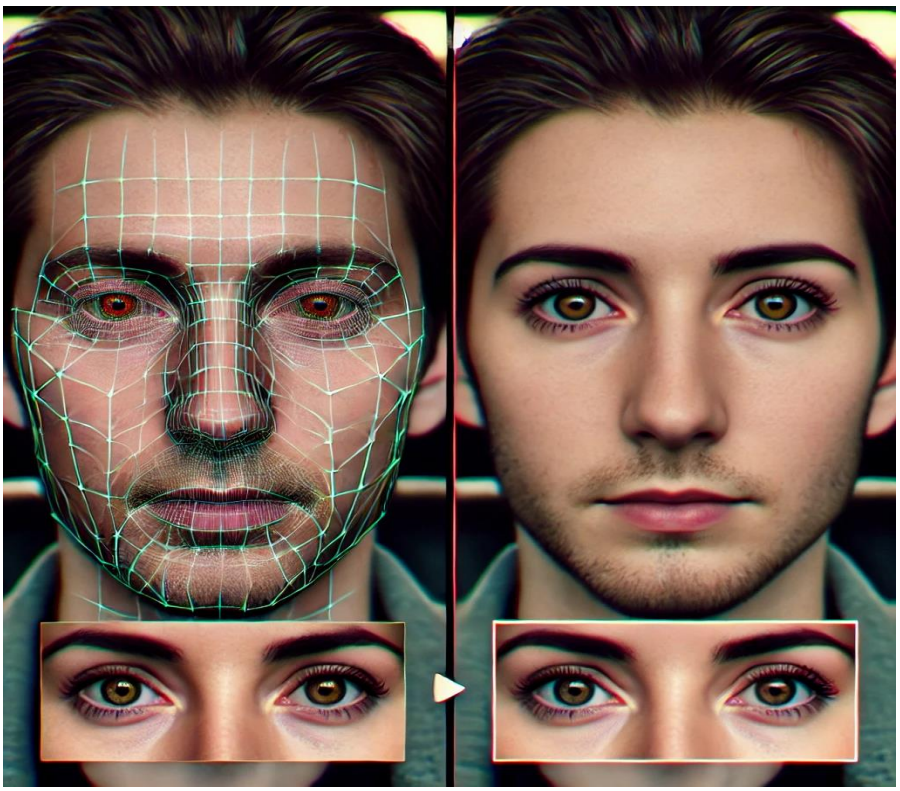


Рисунок В.2 – порівняння оригінального та відредагованого відео за допомогою глибоких нейронних мереж

Рисунок В.3 – Високоякісне фото AI-генерованої підробки

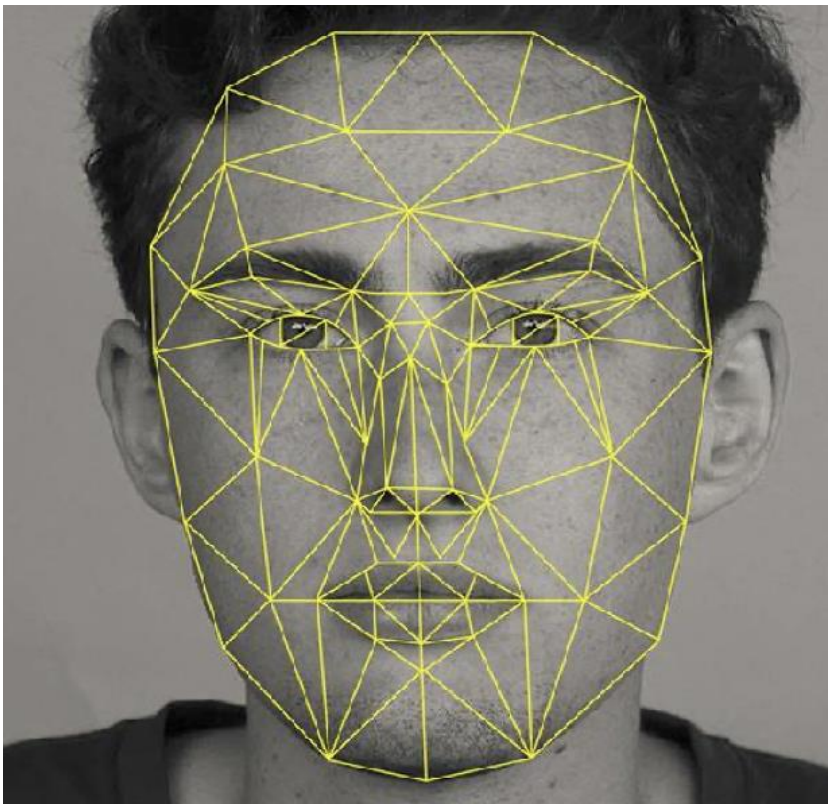


Рисунок В.4 – Приклад побудови геометричних ліній на обличчі

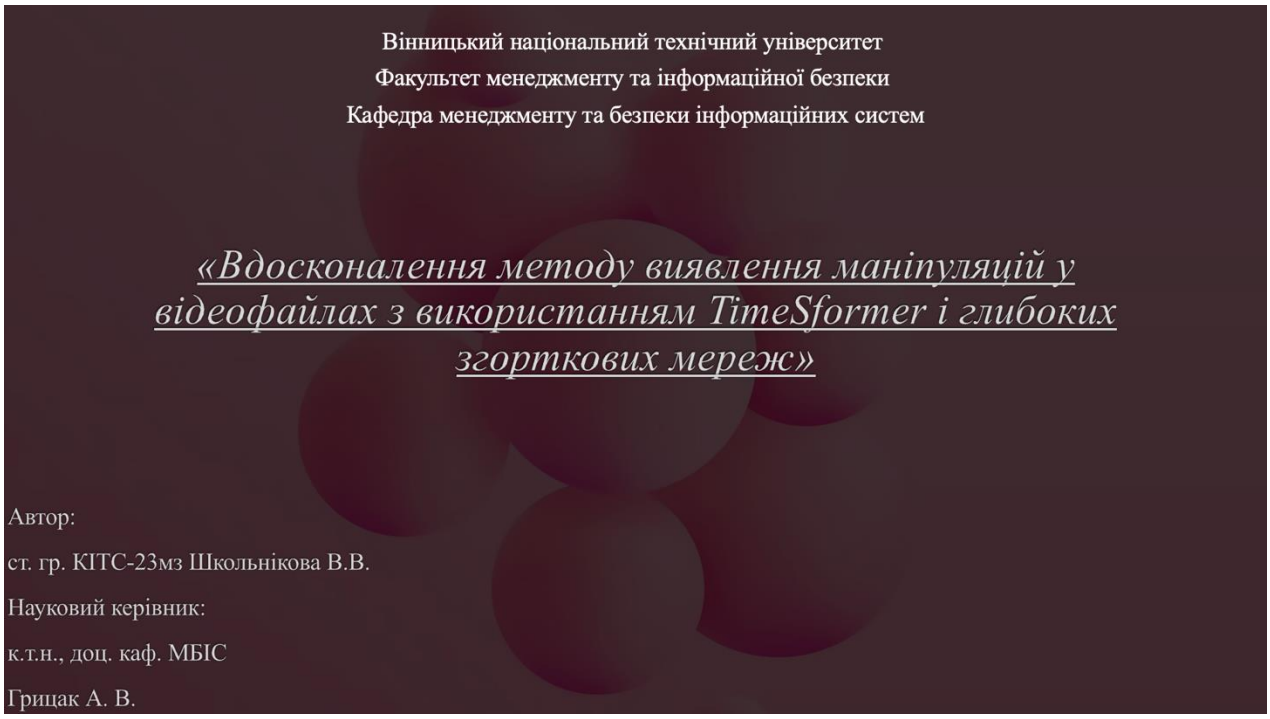


Рисунок В.5 – Титульний слайд

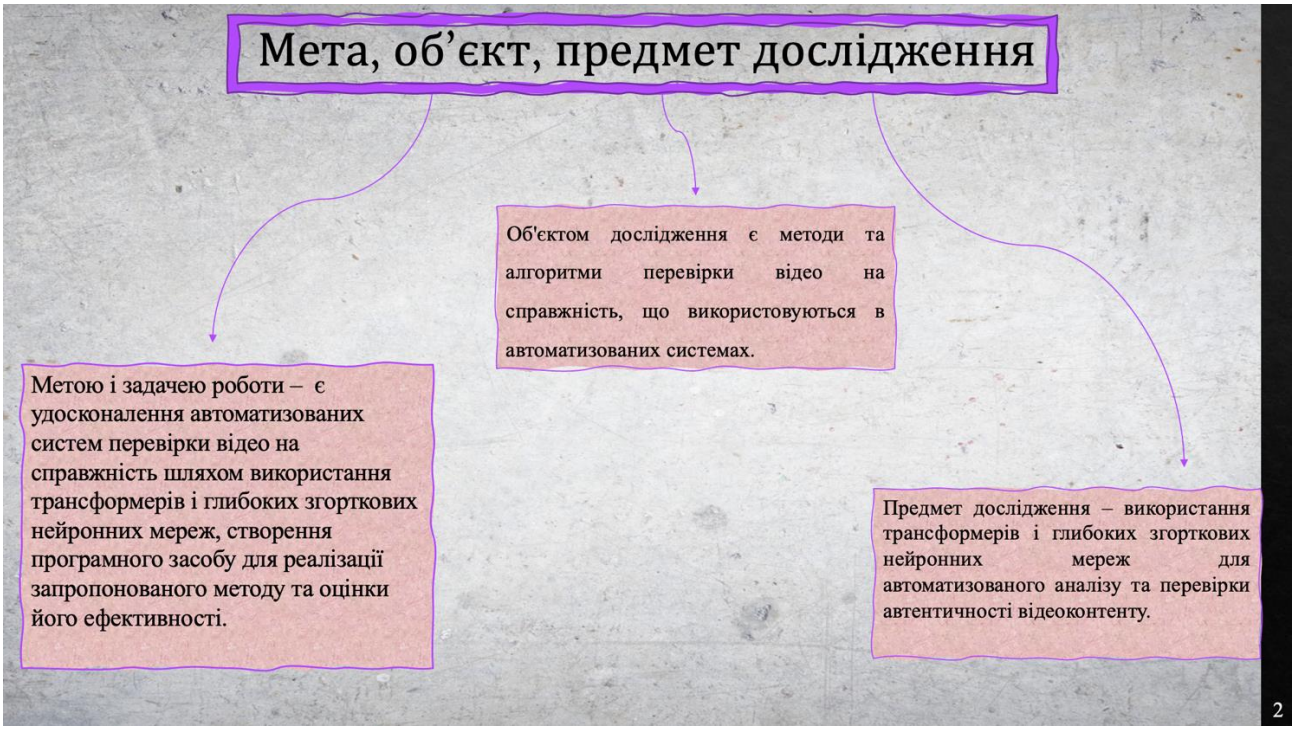


Рисунок В.6 – Мета, об’єкт, предмет дослідження

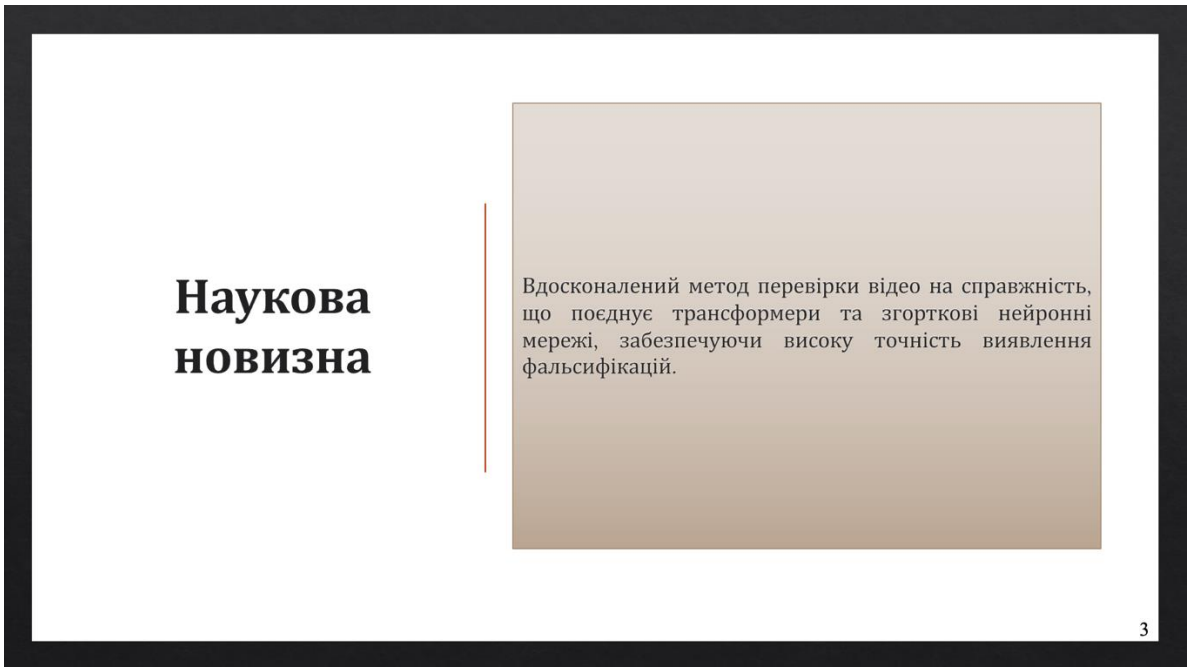


Рисунок В.7 – Наукова новизна дослідження

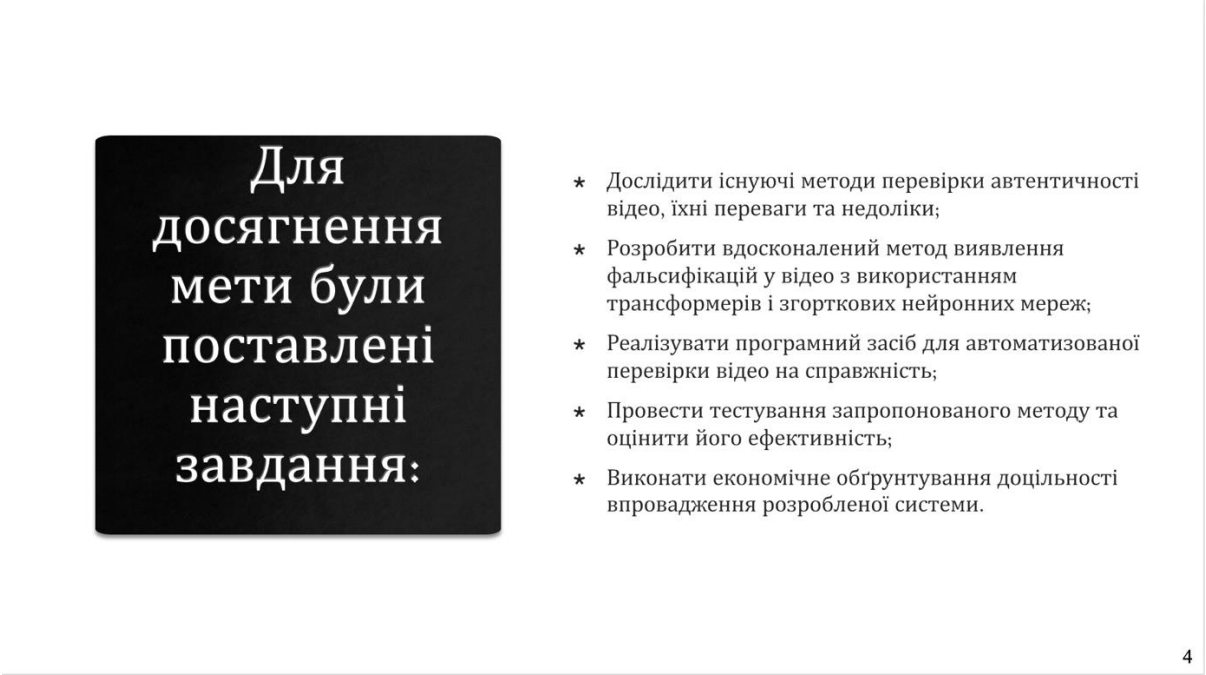


Рисунок В.8 – Завдання

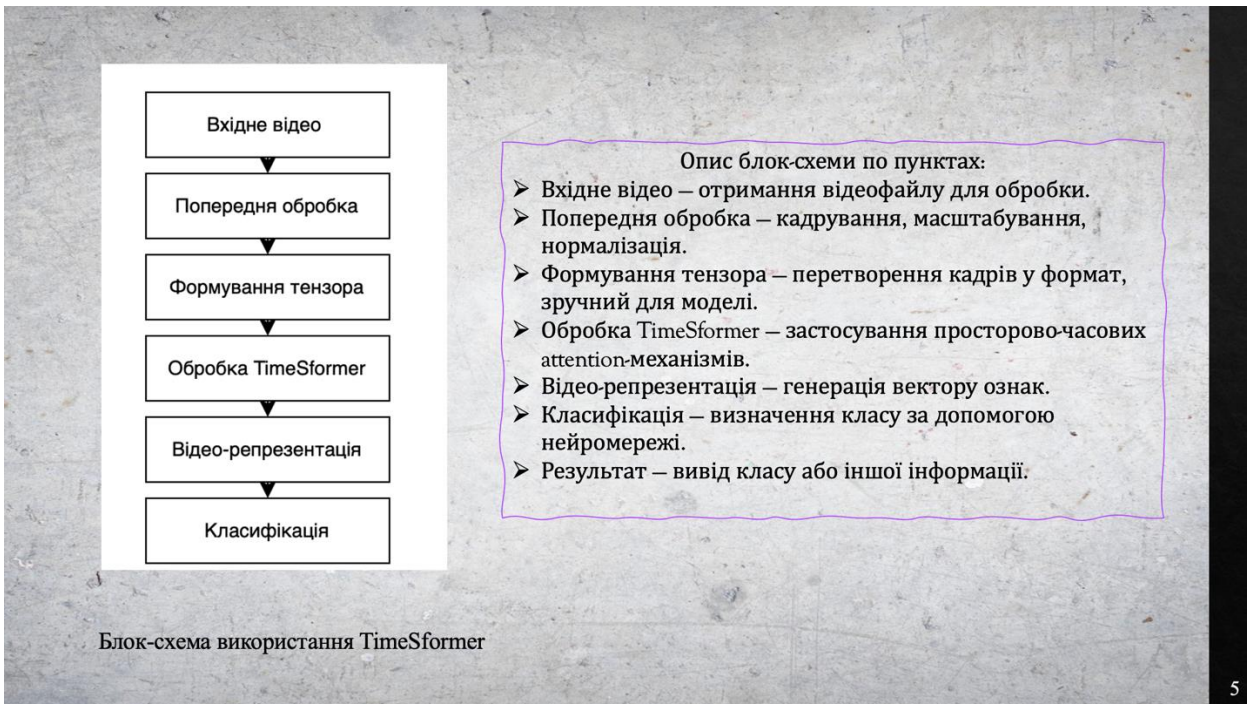


Рисунок В.9 – Блок-схема використання TimeSformer

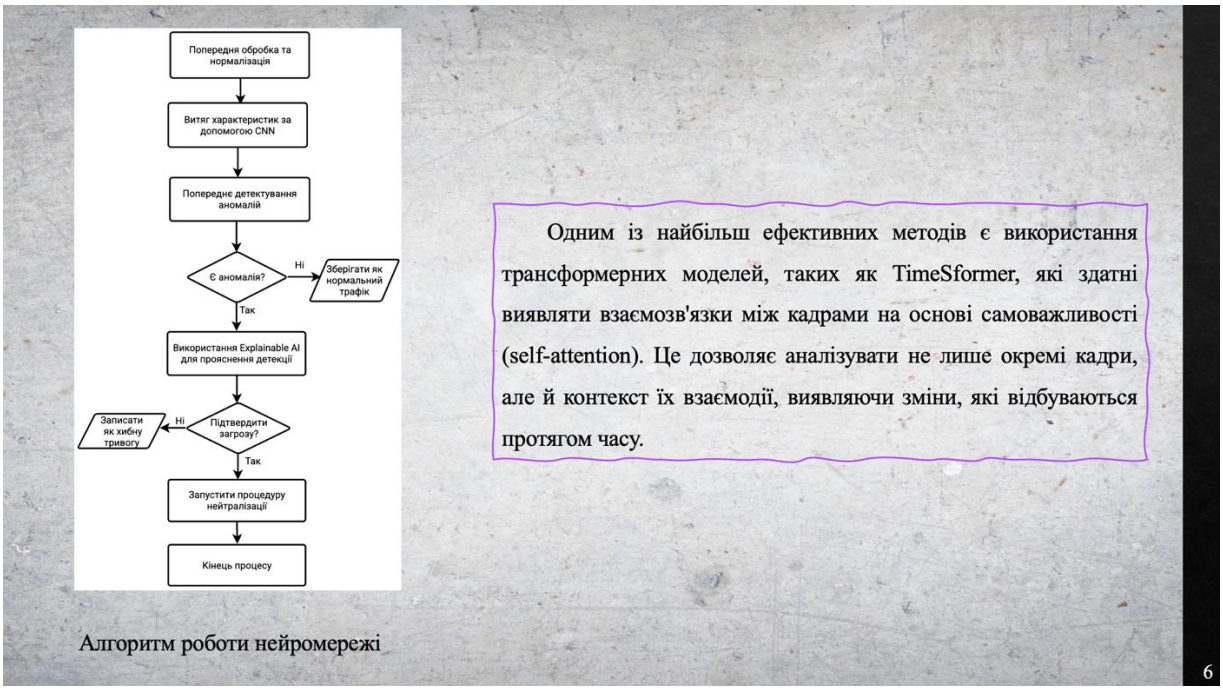


Рисунок В.10 – Алгоритм роботи нейромережі

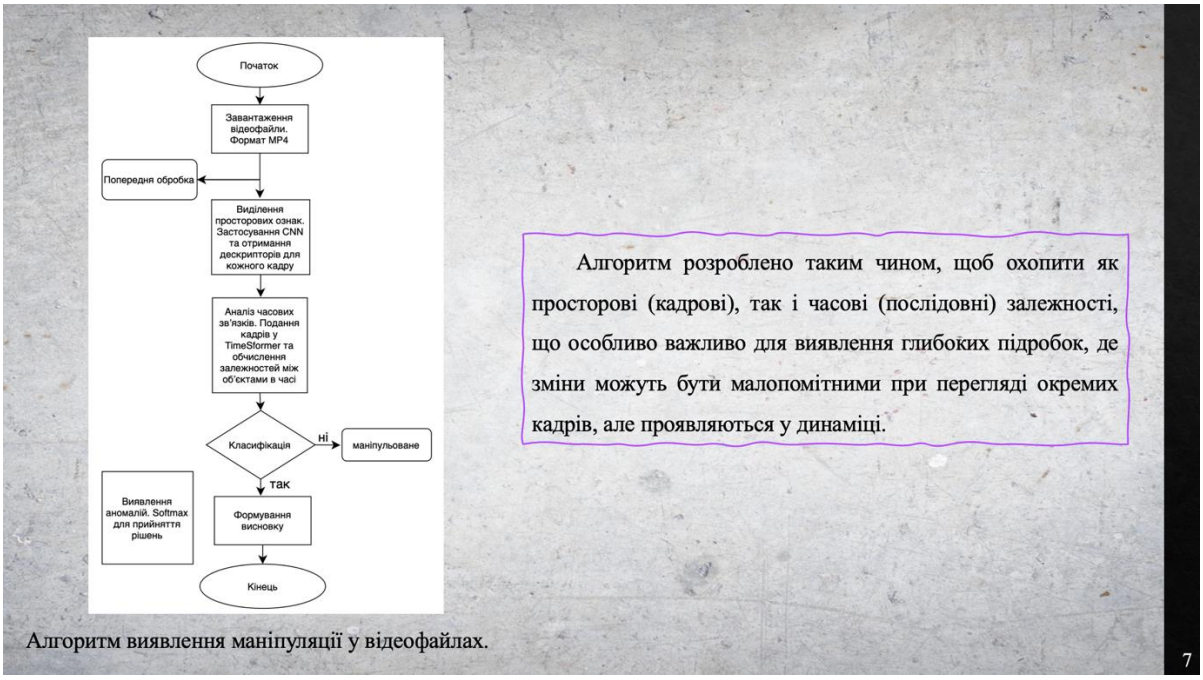


Рисунок В.11 – Алгоритм виявлення маніпуляції у відеофайлах

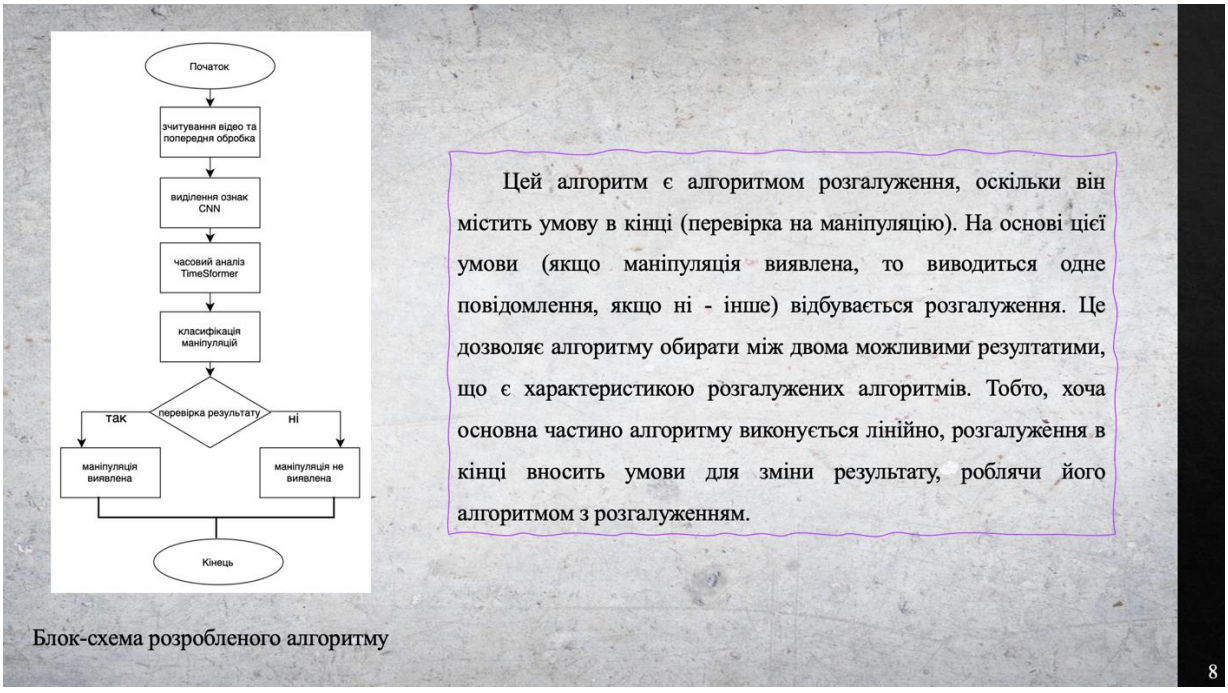


Рисунок В.12 – Блок-схема розробленого алгоритму

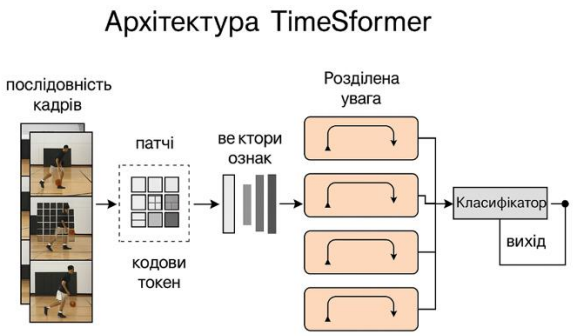
Порівняння запропонованої архітектури

Метод	Точність (%)	FPS	Переваги
XceptionNet	85.3	25	Простота, швидка інформація
VIVIT	91.1	12	Потужна увага в часі
Hybrid3D-CNN	89.7	18	Баланс простір + час
CNN+TimeSformer	93.6	16	Висока точність, стабільність

9

Рисунок В.13 – Порівняння запропонованої архітектури

Блок-схема архітектури TimeSformer



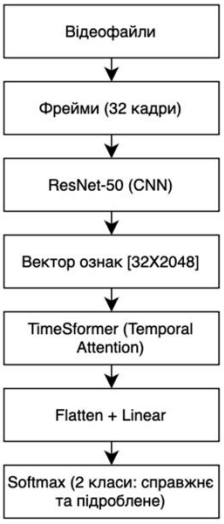
10

Рисунок В.14 – Блок-схема архітектури TimeSformer

Характеристика	TimeSformer	CNN
Обробка простору і часу	Розділена увага	Спільна через згортки
Захоплення глобальних зв'язків	Ефективне	Обмежене
Точність у розпізнаванні дій	Вища	Нижча
Обчислювальні витрати	Нижчі	Вищі
Гнучкість до довжини відео	висока	Обмежена

11

Рисунок В.15 – Порівняння CNN



12

Рисунок В.16 – Схема архітектурної моделі

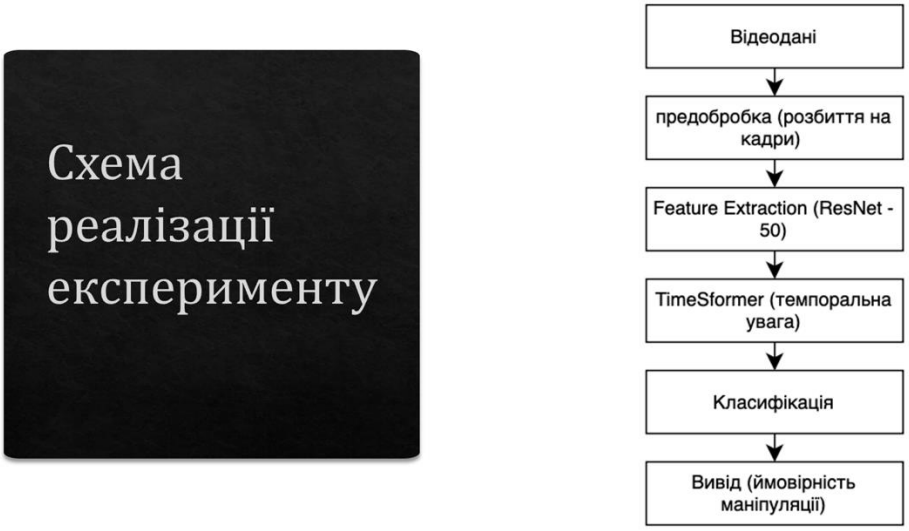


Рисунок В.17 – Схема реалізації експерименту

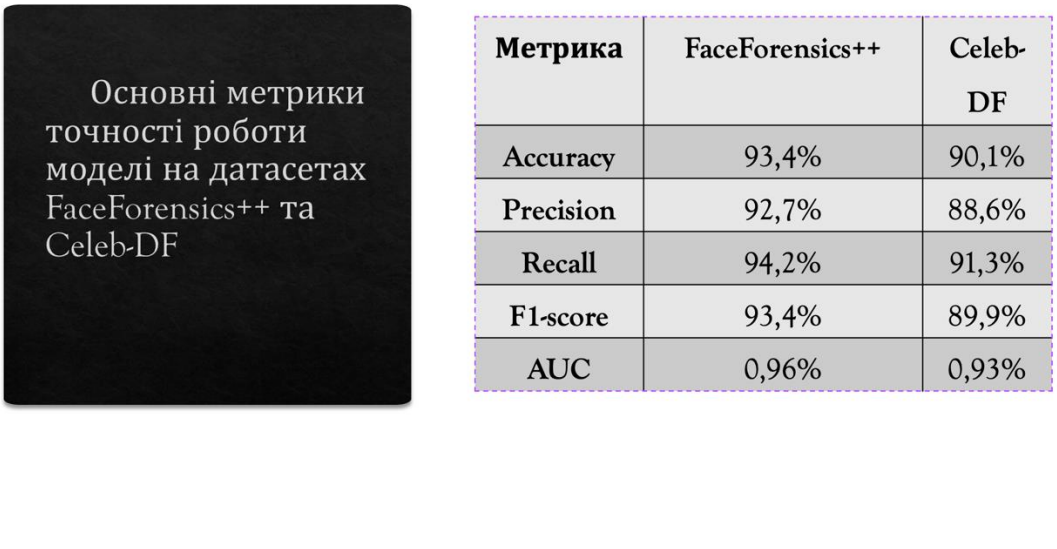
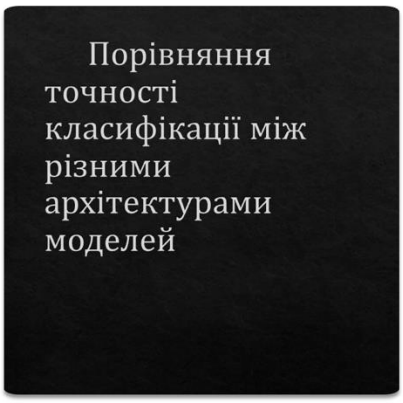


Рисунок В.18 – Основні метрики точності роботи моделі на датасетах FaceForensics++ та Celeb-DF



Модель	Accuracy (FaceForensics++)
CNN (ResNet50)	86,2%
TimeSformer (без CNN)	90,5%
LSTM + CNN	88,7%
TimeSformer + CNN	93,4%

15

Рисунок В.19 – Порівняння точності класифікації між різними архітектурами моделей

Показник	Значення	Пояснення
Загальний обсяг інвестицій (I)	30000 грн	Витрати на розробку і впровадження
Ціна підписки для одного користувача	500 грн/місяць	SaaS-модель
Кількість користувачів	200 осіб	Базовий сценарій
Щомісячний дохід (D_m)	100000 грн	200*500 грн
Річний дохід (D_Y)	1200000 грн	100000*12
Чистий прибуток (P)	1170000 грн	D_y-I
Період окупності (PP)	~ 0,3 місяці (~ 9 днів)	Дуже швидке повернення інвестицій
Коефіцієнт рентабельності (ROI)	3900%	Надзвичайно високий рівень ефективності

16

Рисунок В.20 – Розрахунок ефективності реалізації програмного забезпечення

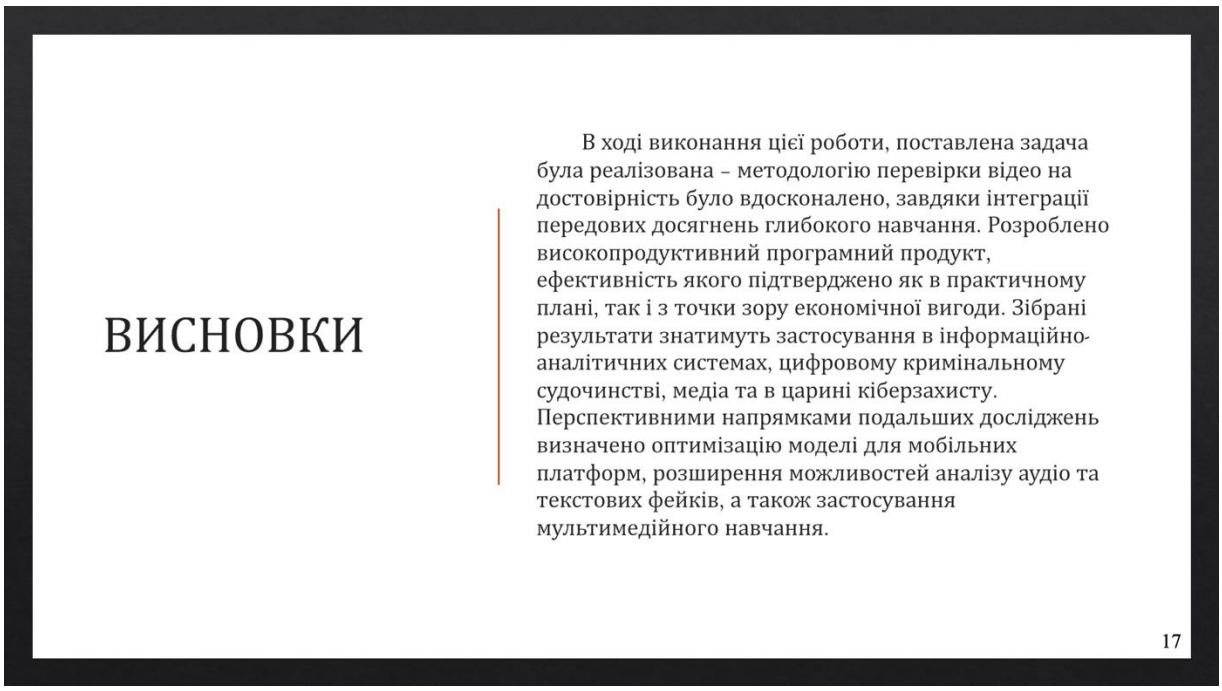


Рисунок В.21 – Висновки

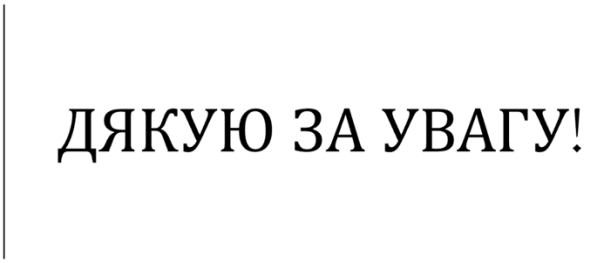


Рисунок В.22 – Фінальний слайд

Додаток Г. Протокол перевірки на антиплагіат

116

Додаток Г. Протокол перевірки на антиплагіат

ПРОТОКОЛ ПЕРЕВІРКИ НАВЧАЛЬНОЇ (КВАЛІФІКАЦІЙНОЇ) РОБОТИ

Назва роботи: Вдосконалення методу виявлення маніпуляцій у відеофайлах з використанням TimeSformer і глибоких згорткових мереж

Тип роботи: магістерська кваліфікаційна робота

Підрозділ Кафедра менеджменту на безпеки інформаційних систем
Факультет менеджменту та інформаційної безпеки
Гр. КІТС-23мз

Керівник доцент Грицак А.В. _____

Показники звіту подібності

Strike Plagiarism	
Оригінальність	99,02 %
Загальна схожість	0,98 %

Аналіз звіту подібності (відмітити потрібне)

- ☐ **Запозичення, виявлені у роботі, оформлені коректно і не містять ознак плагіату.**
- ☐ Виявлені у роботі запозичення не мають ознак плагіату, але їх надмірна кількість викликає сумніви щодо цінності роботи і відсутності самостійності її автора. Роботу направити на доопрацювання.
- ☐ Виявлені у роботі запозичення є недобросовісними і мають ознаки плагіату та/або в ній містяться навмисні спотворення тексту, що вказують на спроби приховування недобросовісних запозичень.

Заявляю, що ознайомлений (-на) з повним звітом подібності, який був згенерований Системою щодо роботи (додається)

Автор _____ Школьнікова В.В.
(підпис) (прізвище, ініціали)

Опис прийнятого рішення

☐ **Допустити до захисту**

Особа, відповідальна за перевірку _____ Коваль Н.П.
(підпис) (прізвище, ініціали)

Експерт _____ _____
(за потреби) (підпис) (прізвище, ініціали, посада)