МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет прикладної математики та інформатики
(повне наймснування назва факультету)

Кафедра дискретного аналізу та інтелектуальних систем
(повна назва кафедри)

Магістерська робота

Застосування методів штучного інтелекту для визначення правильності носіння маски в умовах карантину

Виконав: студент групи ПМіМ-23
спеціальності
122 «Комп'ютерні науки»
(шифр і назва спеціальності)

(підпис) Бать Т.А.
(прізвище та ініціали)

Керівник (підпис) Позднякова І.В.
(прізвище та ініціали)

Рецензент (підпис) Глова А.Р.
(прізвище та ініціали)

Львів – 2022

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА

Факультет Прикладної математики та інформатики
Кафедра Дискретного аналізу та інтелектуальних систем
Спеціальність 122 «Комп'ютерні науки»
(шифр і назва)

«ЗАТВЕРДЖУЮ»
Завідувач кафедри Притула М.М.

"31" серпня 2022 року

З А В Д А Н Н Я

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Батю Тарасу Андрійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи «Застосування методів штучного інтелекту для визначення правильності носіння маски в умовах карантину»

керівник роботи Позднякова Інна Володимирівна, канд. фіз.-мат. наук,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені Вченою радою факультету від "13" вересня 2022 року № 15

2. Строк подання студентом роботи **12.12.2022р**

3. Вихідні дані до роботи Документація по Python, TensorFlow, Keras, OpenCV, Pandas, NumPy. Науково-технічні публікації, інтернет ресурси, відкриті набори даних з цифровими зображеннями. Середовище розробки PyCharm Community 2022.1

4. Зміст магістерської роботи (перелік питань, які потрібно розробити) Огляд існуючих досліджень, огляд наборів даних, реалізація алгоритму, аналіз результатів, висновки.

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)
Приклади зображень із наборів даних, схема запропонованого способу визначення правильності носіння маски, архітектури нейронних мереж, демонстрація роботи програми.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання "31" серпня 2022р

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів магістерської роботи	Строк виконання етапів роботи	Примітка
1	Аналіз існуючих досліджень	01.09.2022-06.09.2022	
2	Аналіз наборів даних	07.09.2022-12.09.2022	
3	Дослідження необхідних технологій	13.09.2022-16.09.2022	
4	Розробка алгоритму	17.09.2022-20.09.2022	
5	Програмна реалізація	21.09.2022-29.09.2022	
6	Тестування отриманих моделей	30.09.2022-02.10.2022	
7	Демонстрація роботи програми	03.10.2022-04.10.2022	
8	Аналіз отриманих результатів	05.10.2022-10.10.2022	
9	Оформлення роботи	10.10.2022-10.11.2022	

Студент _____ Бать Т.А.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Позднякова І.В.
(підпис) (прізвище та ініціали)

Реферат

Магістерська робота складається із огляду існуючих досліджень, огляду наборів даних, опису реалізації алгоритму, аналізу результатів та висновків. Загальний обсяг складає 26 сторінок, список використаних джерел містить 25 найменувань, робота ілюстрована рисунками.

У даній роботі було реалізовано чотири моделі для автоматичного визначення правильності носіння маски. У підсумку отримано нові результати, які демонструють точність застосування розглянутих моделей для розпізнавання різних категорій помилок неправильного носіння масок.

Доведено, що обов'язкове носіння масок запобігає поширенню інфекційних захворювань. Одним із них є коронавірус, який у наш час став глобальною проблемою. Дана робота націлена на підвищення ефективності маскового режиму шляхом виявлення типових помилок під час носіння масок. Дана програма може використовуватись для камер спостережень в метро, торговельних центрах, аеропортах, лікарнях або в інших громадських місцях, щоб запобігти проблемі поширення вірусу, подаючи відповідне сповіщення. Це може бути автоматичне голосове повідомлення, або оператор системи може передавати інформацію до охоронців, які в свою чергу зроблять зауваження порушнику.

Ключові слова: глибинне навчання, класифікація, згорткова нейронна мережа, набір даних, TensorFlow, Keras, маска, коронавірус.

ЗМІСТ

Вступ **6**

1 Огляд літератури **8**

1.1 Огляд існуючих досліджень 8

1.2 Огляд наборів даних 11

2 Реалізація алгоритму **14**

2.1 Попередня обробка 14

2.2 Класифікація 14

2.3 Модель CNN 16

2.4 Модель MobileNetV2 18

2.5 Модель Xception 19

2.6 Модель VGG-16 20

3 Результати **22**

Висновки **26**

Список використаних джерел **28**

ВСТУП

Боротьба з пандемією COVID-19 стала викликом для усього світу. Вплив пандемії не лише економічний, а й соціальний. Кожен із багатьох механізмів боротьби із заразністю хвороби є надзвичайно важливим. Одним із найважливіших механізмів є використання масок. Однак неправильне носіння масок робить цей метод профілактики неефективним. Штучний інтелект та його методи розпізнавання обличчя можна використати для виявлення недотримання маскового режиму, що вплине на зменшення передачі вірусу, особливо в приміщенні. У даній роботі буде представлено метод автоматичного визначення правильності носіння маски. Для цього буде використано згорткові нейронні мережі, щоб виявляти не лише наявність маски на обличчі, але й інші помилки, які зазвичай не враховуються, але які можуть сприяти поширенню вірусу.

У грудні 2019 року в Ухані, провінція Хубей, Китай, з'явилися випадки пневмонії невідомого походження, клінічні симптоми якої були дуже схожі на пневмонію вірусного походження. У першій групі інфікованих було встановлено, що це зоонозна інфекція, тобто передача вірусу відбулась від тварин до людини [1]. Завдяки взяттю зразків із дихальних шляхів за допомогою ПЛР тестів, було проведено детальніше вивчення вірусу. Цей факт дозволив уточнити походження цього вірусу. Даному захворюванню присвоїли назву Новий Коронавірус 2019 (2019-nCoV).

Різні дослідження показали, що SARS-CoV-2 поширюється серед людей повітряно-крапельним шляхом. Крім того, було продемонстровано, що вірус може поширюватися ще до появи симптомів, тобто можливий безсимптомний перебіг хвороби. Це означає, що деякі люди можуть бути носіями вірусу та поширювати його, не проявляючи жодних симптомів захворювання. Цей факт став причиною стрімкого розвитку та поширення пандемії. У зв'язку зі стрімким поширенням вірусу Всесвітня організація охорони здоров'я оголосила COVID-19 надзвичайною ситуацією в галузі охорони здоров'я. За даними ВООЗ, станом на 2 травня 2021 року було зареєстровано 151 803 822 підтверджених випадків COVID-19, у тому числі 3 186 538 летальних.

Через передачу вірусу повітряно-крапельним шляхом багато країн запровадили масковий режим для захисту від коронавірусної інфекції. Завдяки використанню масок, захворюваність на COVID-19 знизились. Крім того, використання масок разом із дотриманням соціальної дистанції змогли значно знизити статистику захворюваності від епідемії. Основна функція масок - це запобігання передачі вірусних частинок, які знаходяться в повітрі, між людьми під час кашлю, розмови, чхання тощо. Також, використання масок дуже допомогло

в запобіганні заражень безсимптомними пацієнтами решти населення. Останні дослідження показують, що хірургічні маски ефективно захищають від коронавірусу.

Рекомендації Центрів з контролю та профілактики захворювань свідчать [2], що під час носіння хірургічної маски, на видиху повітря з носа та рота виходить із великою швидкістю та спрямовується фронтально. Частинки вірусу відносно товсті, від 3 до 8 мікрон, вдаряються безпосередньо об внутрішню частину маски. Навіть якщо повітря виходить через краї, бактерії чи інші частинки не виходять, оскільки через свій розмір, вони не можуть рухатись разом із потоком повітря, яке виходить через краї, якщо маска правильно одягнута.

Актуальність теми. Саме через велике значення використання масок для безпеки, необхідно контролювати їхнє правильне використання. Цей факт призводить до збільшення методів контролю правильності носіння маски, які дозволяють зменшити передачу вірусу. З цієї причини, методи, засновані на штучному інтелекті, набули великої актуальності, адже вони дозволяють більш якісно контролювати носіння масок у громадських місцях.

Метою роботи є дослідження методів машинного навчання, для створення застосунку, який буде визначати чи носить людина маску належним чином. Застосунок повинен враховувати не лише наявність чи відсутність маски, але й можливість розрізняти, коли вона правильно одягнена, а коли ні.

Досягнення зазначеної мети передбачає виконання таких ***завдань***:

1) Ознайомлення з наявними алгоритмами.

2) Підготовках тестових наборів даних.

3) Реалізація нейронної мережі для визначення правильності носіння маски.

Об'єкт дослідження – визначення правильності носіння маски.

Предмет дослідження – системи комп'ютерного зору.

1 Огляд літератури

1.1 Огляд існуючих досліджень

Через серйозну загрозу пандемії COVID-19 тема автоматизації визначення правильності носіння маски набула актуальності. Штучний інтелект є одним із найкращих методів для вирішення таких проблем, як класифікація зображень, для різних випадків. У цьому розділі буде розглянуто найактуальніші роботи з галузі штучного інтелекту та COVID-19.

Застосунки штучного інтелекту та нейронних мереж здатні вирішувати проблеми, виявляючи шаблони на зображеннях. Такі автори, як Чанг та інші, у [3], розробили програму для мобільного телефону, здатну розпізнавати та класифіковувати рослини з зображень за допомогою нейронної мережі InceptionV3 [4]. Вони також реалізували прототип для ідентифікації порід дерев, для яких класифікація відбувається в режимі реального часу.

Однак ситуація в світі, спричинена SARS-CoV-2, у якій ми опинилися, змусила авторів спробувати покращити контроль над поширенням вірусу шляхом розробки додатків для контролю використання масок. У попередні роки такі автори, як Нієто-Родрігез та інші вже розробили такий тип застосунку для боротьби з іншими епідеміями. Вони у 2015 році розробили систему що дозволяє визначати обличчя без маски на відстані 5 метрів. Система була розроблена для контролю використання хірургічних масок медичним персоналом в операційних. Це була свого роду сигналізація, яка спрацьовувала, коли медичний персонал проходив без масок повз камери, оскільки їхнє використання є обов'язковим.

Тим не менш, за останні два роки кількість публікацій, пов'язаних із цими системами ШІ, різко зросла через критичну ситуацію, з якою ми стикаємося. Потреба контролювати скупчення людей, які не носять маски в громадських місцях, набула особливого значення.

Награт та інші в [5] розробили додаток, який визначає наявність маски на обличчі. Щоб виявляти маски в режимі реального часу, вони використали архітектуру класифікатора із підходом SSDMNV2. Для цього вони застосували глибинне навчання TensorFlow, Keras і OpenCV. Точність цього дослідження становить 92,6%.

У дослідженні, проведеному Джаухарі у [6], мета полягала в виявленні ключових точок на обличчі, щоб мати можливість виявити наявність масок на зображеннях. Додаток був розроблений для одноплатного комп'ютера Raspberry Pi. Для отримання ефективних, швидких і точних результатів була використана система розпізнавання обличчя на основі методу Віоли Джонс. Цей метод дозволяє налаштувати каскадний класифікатор для визначення області обличчя на зображенні. Точність цього дослідження становить 90,9%. Сен в [7] розробив си-

стему, здатну відрізняти людей, які носять маски, від тих, хто не носить. Він використав модель PyTorch Python і OpenCV для виявлення масок. Модель має точність 79,24%. Крім того, цей застосунок має систему оповіщення, яка видає звуковий сигнал, для попередження людини, яка не має одягнутої маски. Камера Raspberry-PI використовується для знімання відео та перетворення його на зображення для подальшої обробки.

Наступні дослідження показують приклади застосування цих типів програм. Курлекар та інші в [8] розробили систему, яка може бути інтегрована в офіси, аеропорти та громадські місця в цілому. За допомогою даної системи надається можливість виявляти маски для обличчя на статичних зображеннях, а також на відео в режимі реального часу. Для цього вони використовували концепції комп'ютерного зору та глибинного навчання, використовуючи OpenCV і TensorFlow. Сакші в [9] розробив детектор маски на обличчі за допомогою Keras та TensorFlow. Модель було навчено з кількома варіаціями вхідних даних, щоб гарантувати, що система зможе ідентифікувати маски на обличчях в режимі реального часу за допомогою відео або статичних зображень. Кінцева мета даних досліджень полягає в тому, щоб за допомогою комп'ютерного зору реалізувати працездатну модель, яка працюватиме з високою кількістю людей в одному кадрі, для галузей охорони здоров'я, навчальних закладів тощо.

У 2020 році Ченг та інші в [10] довели, що виявлення використання масок було важливим для припинення поширення вірусу. Завдяки використанню алгоритму YOLO-v3 він довів придатність виявлення маски в режимі реального часу. Крім того, додаток получився невеликим за розміром, швидким і підходящим для розгортання на мобільному обладнанні. Лой та інші в [11] розробили гібридну систему для виявлення масок на обличчі. Вони вибрали три набори даних. Набір даних із змодельованими обличчями в масці (за допомогою PhotoShop), набір із людьми в масках у реальному світі та набір фотографій з камер спостережень в метро. В основі цього дослідження лежить нейронна мережа Resnet-50 для виявлення ключових точок обличчя. Наступний компонент це класифікатор на основі методу опорних векторів і загального алгоритму класифікації правильності носіння маски. Точність системи становить 89,49%, 89,64% і 88% відповідно для кожного з досліджуваних наборів даних.

Крім визначення наявності маски, важливою є можливість виявлення, коли вона одягнута не правильно. Неправильне використання значно знижує ефективність маски проти вірусу. З цієї причини кілька авторів, окрім визначення наявності чи відсутності маски, зосередилися на виявленні правильного або неправильного розміщення. У 2020 році Рудрарагі та інші в [12] розробили програму, яка базується на двох етапах. На першому етапі виявляється наявність або відсутність маски. Після виявлення маски аналізується правильне чи неправильне її розміщення. Для цього автор покладається на туманні обчислення. Для ви-

явлення облич використовуються каскадні класифікатори Хаара, а для класифікації модель MobileNet. Потокове передавання даних відбувається локально без використання Інтернету. Точність цієї системи становить близько 90%.

Ванг в [13], також запропонував виявляти правильність носіння масок за допомогою двоетапного підходу, використовуючи методи гібридного машинного навчання. На першому етапі користувач, який носить маску, виявляється за допомогою моделей Faster RCNN і InceptionV2. Наступний етап спрямований на перевірку правильності розміщення, реалізований класифікатором з адаптивною системою навчання. Загальна точність для простих сценаріїв становить 97,32%, а для більш складних сценаріїв – 91,13%.

Квін та ін. у [14], використовуючи систему підвищення якості зображень і мережі класифікації SRCNet, за допомогою навчального набору даних Medical Masks, розробили метод ідентифікації наявності або відсутності маски. Крім того, даний метод здатний виявити найчастішу помилку неправильного використання - носіння маски під носом. Алгоритм базується на наступних чотирьох кроках: попередня обробка зображення, розпізнавання та обрізання обличчя, підвищення якості зображення та розпізнавання маски. Точність, досягнута за допомогою цієї методики, становить 88,70%. У таблиці 1 узагальнено різні роботи, представлені в цьому розділі.

Табл. 1: Підсумки результатів існуючих досліджень.

Автор	**Дата**	**Модель**	**Бібліотека ПЗ**	**Точність**
Награт	2021р	SSDMNV2	TensorFlow, OpenCV	92.6%
Джаухарі	2021р	AdaBoost	Python, Keras	90.9%
Сен	2021р	DenseNet169	PyTorch, OpenCV	79.2%
Курлекар	2021р	CNN	TensorFlow, OpenCV	82.5%
Сакші	2021р	MobileNetV2	TensorFlow, Keras	94.4%
Ченг	2020р	CNN + SVM	YOLO v3	73,5%
Лой	2021р	Resnet50	Fast RCNN	89.6%
Рудрагарі	2020р	MobileNet	OpenCV, Keras	90.6%
Ванг	2021р	InceptionV2	OpenCV, Matlab	91.1%
Квін	2020р	SRCNet	Matlab	88.7%

1.2 Огляд наборів даних

Набір даних є однією з найважливіших частин для навчання нейронної мережі. У цьому розділі представлені загальні набори даних порівняльного аналізу, які використовуються в наукових дослідженнях для визначення правильності носіння маски. Набір даних Synthetic CelebFaces Attributes (Synthetic CelebA) [15] складається з 10 000 загальнодоступних синтетичних зображень. CelebA [16] — це великомасштабний набір даних облич з понад 200 000 зображеннями знаменитостей. Він був побудований з використанням 50 видів синтетичних масок різних розмірів, форм, кольорів і структур. При створенні синтетичних зразків, маска була нанесена на обличчя за допомогою програми Adobe Photoshop.

Synthetic Face-Occluded Dataset [17] було створено з використанням загальнодоступних наборів даних CelebA та CelebA-HQ. CelebA-HQ — це великомасштабний набір даних з понад 30 000 зображень знаменитостей. Кожне зображення обличчя обрізане та вирівняне за положенням очей. Оклюзії були синтезовані п'ятьма популярними об'єктами, які перекривають обличчя: руками, маскою, сонцезахисними окулярами, окулярами та мікрофоном. Було використано понад 40 різноманітних видів кожного об'єкта з різними розмірами, формами, кольорами та структурами. Крім того, предмети накладались на обличчя випадковим чином.

The Masked Face Detection Dataset (MFDD), Real-World Masked Face Recognition Dataset (RMFRD), та Masked Face Recognition Dataset (SMFRD) були представлені в [18]. MFDD містить 24 771 зображень облич в масках. RMFRD містить 5000 зображень 525 людей в масках і 90 000 зображень цих самих людей без масок. Цей набір даних є найбільшим набором даних для систем розпізнавання правильності носіння маски. Щоб зробити набір даних більш різноманітним, дослідники представили SMFRD, який складається з 500 000 зображень облич 10 000 людей з інтернету, на які наклали маски у Adobe Photoshop.

Набір даних Masked Face Segmentation and Recognition (MFSR) [19] складається з двох частин. Перша частина включає 9742 зображень облич в масках, зібраних з інтернету. Друга частина включає 11 615 зображень 1004 людей, де 704 з них зібрані в реальному світі, а решта зібрана з інтернету. Кожна персона має принаймні одне зображення в масці і одне звичайне. Celebrities in Frontal-Profile (CFP) [20] включає обличчя 500 знаменитостей у портретному вигляді.

AgeDB [21] — це перший набір даних, зібраний вручну. Він включає 16 488 зображень від 568 знаменитостей різного віку. Набір містить чотири протоколи верифікації, де порівнювані обличчя мають різницю у віці 5, 10, 20 і 30 років. У [22] вони створили новий набір даних, узгодивши свої дані з 3D-моделлю, яка може змінюватися. Новий набір складається з 3D-

Рис. 1: Приклади зображень із наборів даних

сканувань облич 100 жінок і 100 чоловіків.

Набір даних MS1MV2 [23] є удосконаленою версією набору даних MS-Celeb1M. MS1MV2 містить 58 мільйонів зображень 85 000 різних людей. Бутрос та ін. [24] створили нову версію з медичними масками MS1MV2, позначену як MS1MV2-Masked. Тип і колір маски були випадково вибрані для кожного зображення, щоб згенерувати більше варіацій. MFR2 містить 269 зображень 53 осіб, взятих з Інтернету. Отже, зображення в наборі даних MRF2, можна вважати знятими в повсякденному житті. MRF2 містить зображення з масками та без масок, в середньому п'ять зображень на особу.

Табл. 2: Підсумки наборів даних.

Набір даних	**Кількість зображень**	**Кількість осіб**	**Тип даних**
RMFRD	95,000	525	Real-world
SMFRD	500,000	10,000	Synthetic
MFSR	11,615	1004	Real-world/synthetic

MFR2	269	53	Synthetic
MFD	990	45	Synthetic
CelebA	>200 K	10,177	Synthetic
CelebA-HQ	>30 K	307	Synthetic
MS1MV2-Masked	5.8 M	85,000	Synthetic
CFP-FP	7000	500	Synthetic
CFP-FF	7000	500	Synthetic
AgeDB	16,488	568	Real-World

2 Реалізація алгоритму

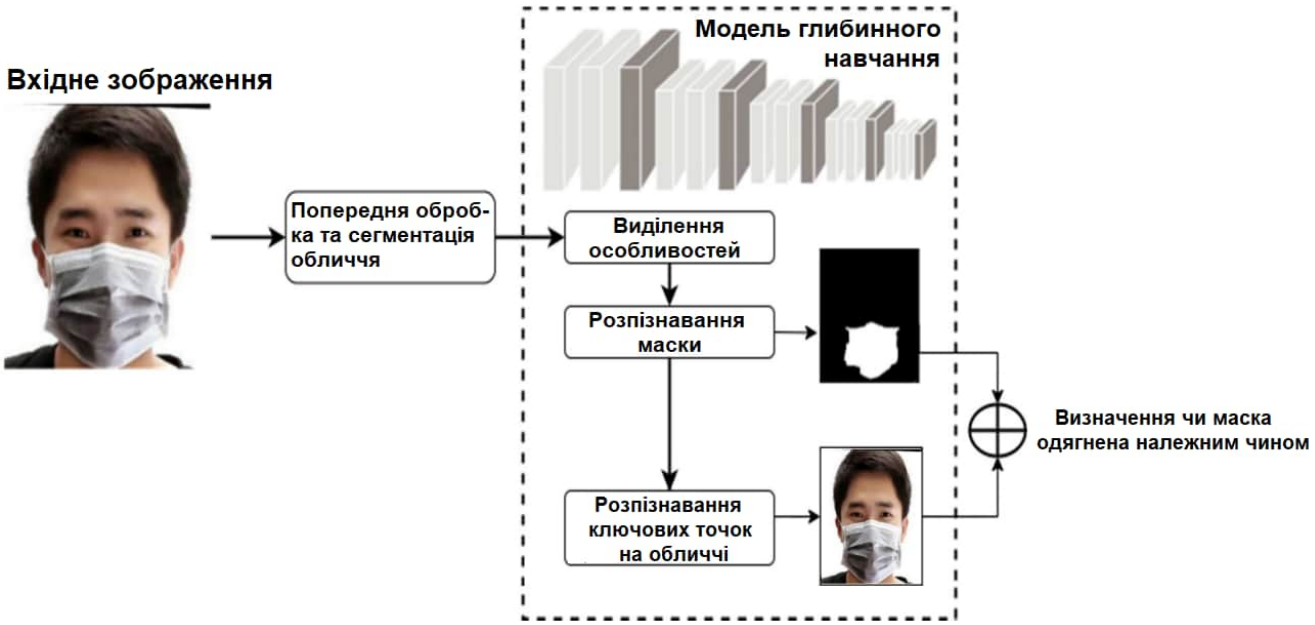


Рис. 2: Огляд запропонованого способу

На рисунку 2 представлений огляд запропонованого способу. Він складається з наступних етапів:

2.1 Попередня обробка

Не всі зображення в використаних наборах даних були обрізані навколо обличчя, тому було прийняте рішення виконати розпізнавання обличчя для усунення зайвих країв, щоб локалізувати обличчя з кожного зображення. Для цього було використано бібліотеку Mediapipe, яка добре визначає області обличчя. Також було потрібно виправити поворот обличчя, щоб ефективно визначати неправильне розміщення маски. Для цього виявляється 68 ключових точок на обличчі за допомогою бібліотеки з відкритим кодом Keras, представленої в [25]. Відповідно до розташування очей застосовується 2D поворот, щоб вирівняти їх горизонтально, як показано на рисунку 3.

2.2 Класифікація

Нещодавно машинне навчання пережило прорив завдяки появі глибинного навчання. Згорткові нейронні мережі лежать в основі цієї революції. Згорткова нейронна мережа структурована ієрархічно. Перші шари відповідають за отримання загальних характеристик зображення, таких як краї або текстури. Наступні рівні використовують ці попередні хара-

Рис. 3: Горизонтальне вирівнювання обличчя

ктеристики для пошуку більш конкретних характеристик. Даний процес триває протягом кількох ітерацій, доки не стане можливим виявлення характеристик з високою семантичною цінністю, наприклад виявлення очей або носа. На завершальному етапі для виконання класифікації використовується звичайна нейронна мережа.

Не зважаючи на те, що згорткові нейронні мережі широко використовуються в задачах обробки мовлення або аудіо, велика кількість досліджень показує, що їхнє використання дає найкращі результати в задачах розпізнавання об'єктів на зображеннях. Дані результати роблять очевидним вибір використання згорткових нейронних мереж для задачі визначення правильності носіння маски.

Тим не менш, у машинному навчанні, отримання набору даних є однією з найскладніших частин. Найпоширенішою проблемою є мала кількість, необхідних для розв'язку задачі, зображень. Зазвичай набори даних містять зображення для розв'язку узагальнених задач. У випадку для задачі визначення правильності носіння маски, існує декілька загальних наборів даних, в більшості з яких зображення поділяються на два типи: з людьми в масках, та без масок. Тому доводиться вручну підбирати зображення для класу неправильно одягнених масок. Для отримання максимальної користі від набору даних також можна збільшити кількість необідних даних. Це можна зробити шляхом внесення невеликих змін до зображень, таких як невеликі повороти, віддзеркалення та масштабування вхідних зображень для збільшення варіативності навчального набору.

Коли доступно небагато навчальних зразків, дуже корисною є техніка «Трансферного навчання». Цей метод базується на використанні моделі, яка була попередньо натренована на великому наборі даних, зазвичай у великомасштабному завданні класифікації зображень.

Трансферне навчання застосовується у два етапи. Для початку, використовуються згорткові шари з оригінальної моделі для виділення загальних ознак. Останні шари, де класифі-

куються зображення, модифікуються для розв'язання конкретних задач. На першому етапі згорткові шари залишаються фіксованими, навчаються лише шари класифікації. На другому етапі, відомому як тонке налаштування, усі рівні розблоковуються, а система перенавчається. Таким чином, модель стає відповідною для виконання поставленої задачі.

Зараз у розпорядженні є велика різноманітність згорткових мереж із десятками навчених шарів. Можна виділити CNN, VGG16, MobileNetV2 та Xception. На основі цих згорткових мереж буде розв'язано задачу визначення правильності носіння маски в даній роботі. Усі ці мережі були навчені за допомогою набору даних ImageNet із понад 14 мільйонами зображень, де було розпізнано 20 000 різних об'єктів.

2.3 Модель CNN

CNN - це класична згорткова нейронна мережа, яка використовується для класифікації зображень. Згорткова нейронна мережа CNN складається з чотирьох компонентів: шарів згортання, нелінійної функції активації, шарів об'єднання та класифікатора. Типову архітектуру CNN представлено на рисунку 4.

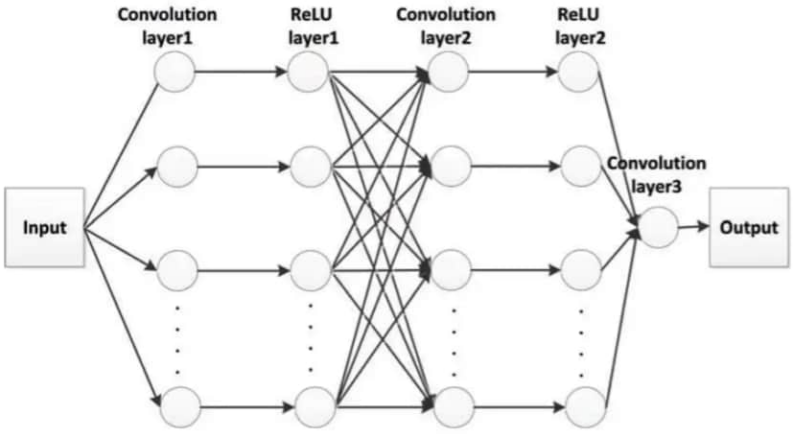


Рис. 4: Архітекрута моделі CNN

Метою згортки є локальне виділення ознак об'єкта на зображенні. Це означає, що мережа вивчатиме певні шаблони всередині зображення та зможе їх розпізнавати. Згортка — це поелементне множення. Мережа сканує частину зображення, зазвичай розміром 3 × 3, і множить її на фільтр. Результат поелементного множення називають картою ознак. Цей крок повторюється, доки не буде відсканована все зображення. Після кожної ітерації згортки розмір зображення зменшується. Схематично згортка представлена на рисунку 5

Після операції згортання до карти ознак застосовується нелінійна функція активації. Стандартною функцією активації для CNN є Relu. Крім того, усі пікселі з від'ємним значе-

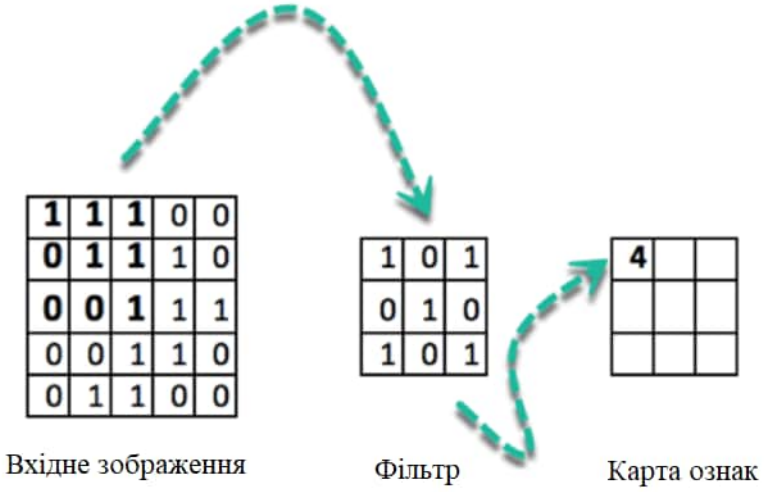


Рис. 5: Візуалізація згортки

нням будуть замінені на нуль.

Наступним етапом є операція об'єднання. Метою об'єднання є зменшення розмірності вхідного зображення. Це робиться для зменшення обчислювальної складності операції. Зменшуючи розмірність, мережа має змогу простіше та ефективніше коригувати ваги. На цьому етапі потрібно визначити розмір і крок. Стандартним способом об'єднання вхідного зображення є використання максимального значення карти ознак. Схематично це представлено на риску 6. Об'єднання екранує чотири підматриці карти ознак 4×4 і повертає максимальне значення. Наприклад, перша підматриця [3,1,3,2], після об'єднання поверне максимум, який дорівнює 3.

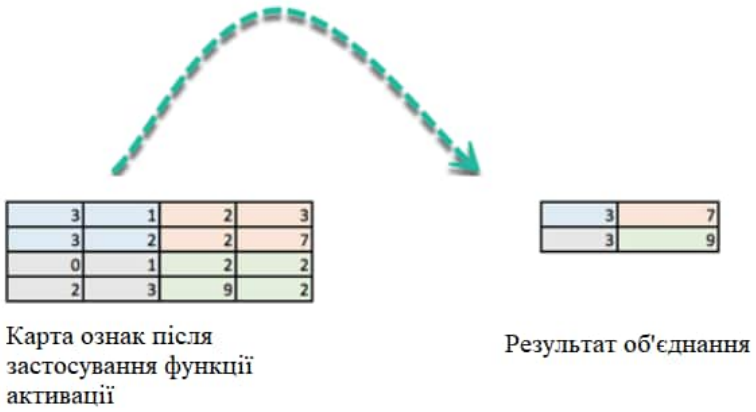


Рис. 6: Візуалізація об'єднання

Останній етап полягає в класифікації. Карта ознак надсилається до основного повністю зв'язаного шару, та за допомогою функції Softmax здійснюється класифікація.

2.4 Модель MobileNetV2

Основний блок цієї мережі загалом схожий на попереднє покоління MobileNetV1, але має низку ключових особливостей. Як і в MobileNetV1, тут є блоки згортки з кроком 1 і з кроком 2, це можна побачити на рисунку 7. Блоки з кроком 2 призначені для зниження просторової розмірності тензора.

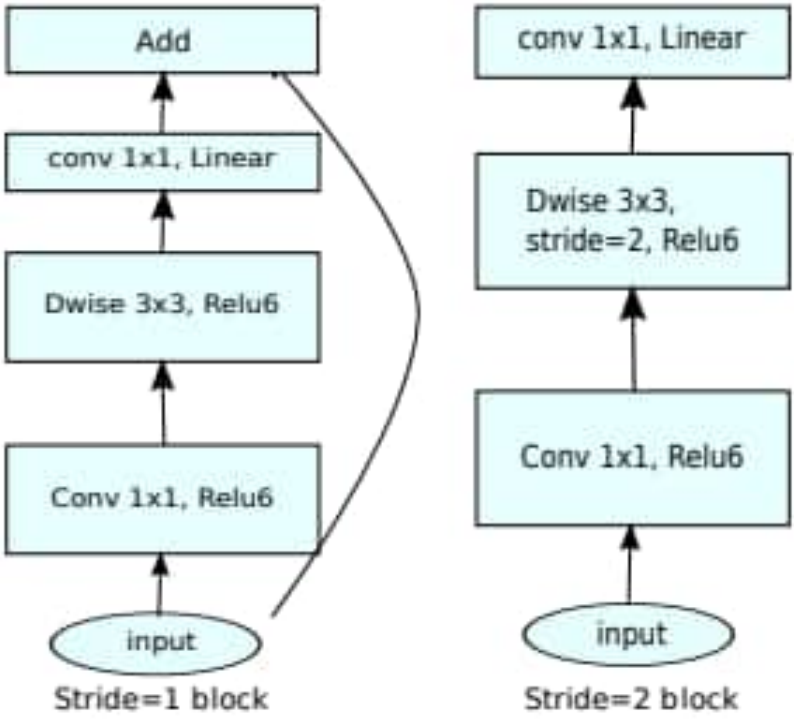


Рис. 7: Архітектура моделі MobileNetV2

Згорткова нейронна мережа MobileNetV2 складається з трьох компонентів: розширювального шару, шару поглибленої згортки та класифікатора.

На вході розширювальний шар приймає тензор розмірності $D_f * D_f * C_{in}$, а на виході видає тензор $D_f * D_f * (t * C_{in})$, де t - новий гіперпараметр, названий рівнем розширення. Автори рекомендують задавати цьому гіперпараметру значення від 5 до 10, менші значення краще працюють з малими мережами, а більші великими. Цей шар створює відображення вхідного тензора у просторі великої розмірності. Автори називають таке відображення "цільовим різноманіттям".

Наступним йде шар поглибленої згортки з ReLU6 - активацією. На вході цей шар приймає тензор розмірності $D_f * D_f * (t * C_{in})$, а на виході видає тензор $(D_f/s) * (D_f/s) * (t * C_{in})$, де s - це крок згортки. Даний шар не змінює кількість каналів.

В кінці йде 1x1 - згортка з лінійною функцією активації, що зменшує кількість каналів. На цьому етапі «цільове різноманіття» високої розмірності, отримане після попередніх кроків,

вкладається в підпростір меншої розмірності без втрати корисної інформації та відбувається класифікація. Схематично це представлено на рисунку 8.

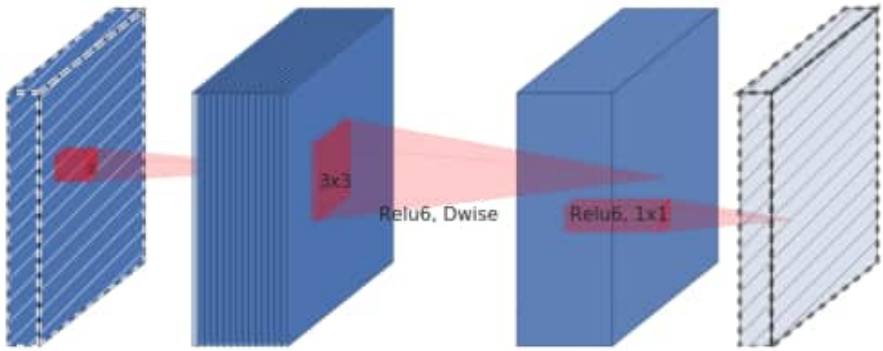


Рис. 8: Структура шарів моделі MobileNetV2

2.5 Модель Xception

Xception - це згорткова нейронна мережа, особливістю якої є те, що згортки розділяються по глибині. Дану мережу розробили дослідники компанії Google. Дана мережа складається з трьох блоків: вхідного, середнього та вихідного. Дані спочатку проходять через вхідний блок, потім через середній, який повторюється вісім разів, та класифікуються в вихідному блоці. Після кожного шару згортки виконується пакетна нормалізація. Архітектура мережі зображена на рисунку 9.

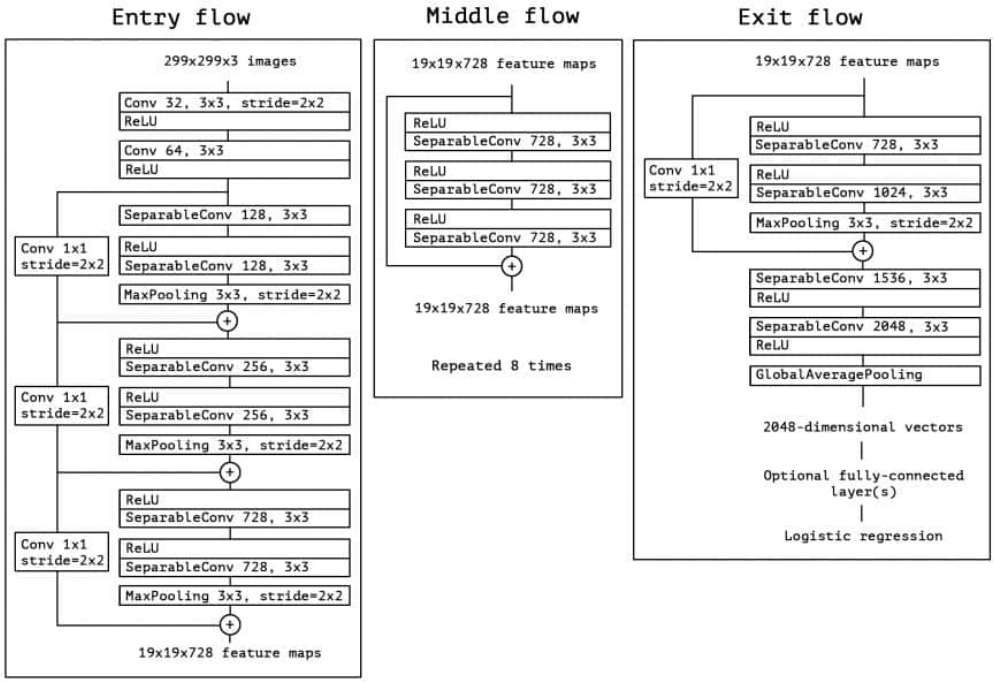


Рис. 9: Архітектура моделі Xception

Згортка по глибині – це просторова згортка n×n по блоку. Вона виконується послідовно в два етапи, як зображено на рисунку 10. Під час першого етапу вихідний тензор 1x1 згортається і в результаті отримується тензор $M * M * C_2$. Ця операція називається початковою згорткою. На другому етапі кожен канал згортається окремою 3x3 згорткою, при цьому розмірність не змінюється, оскільки згортаються не всі канали разом, як у звичайному згортковому шарі. Ця операція називається глибинно-просторовою згорткою.

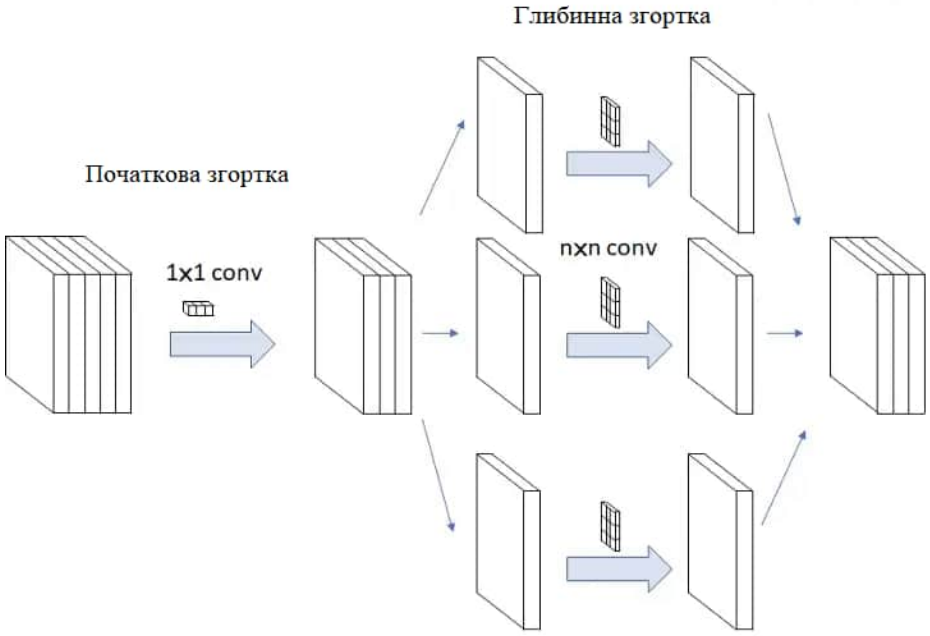


Рис. 10: Структура шарів моделі Xception

2.6 Модель VGG-16

VGG16 — це згорткова нейронна мережа, яка вважається однією з найкращих моделей комп'ютерного зору на сьогоднішній день. В даній мережі використана архітектура з малими 3×3 фільтрами згортки, що показало значне покращення конфігурацій. В VGG-16 є 16 шарів, які містять ваги, 13 згорткових шарів та 5 шарів об'єднання. Архітектура моделі зображена на рисунку 11.

Вхідними даними для VGG-16 є зображення фіксованого розміру 224 x 224 із трьома каналами – R, G і B. Єдиною попередньою обробкою є нормалізація значень RGB для кожного пікселя. Це досягається шляхом віднімання середнього значення від кожного пікселя.

Зображення проходить через перший стек з 2 шарів згортки дуже малого розміру 3 x 3, після чого застосовується функція активації ReLU. Кожен із цих двох шарів містить 64 фільтри. Крок згортки фіксується на рівні 1 пікселя. Ця конфігурація зберігає просторову

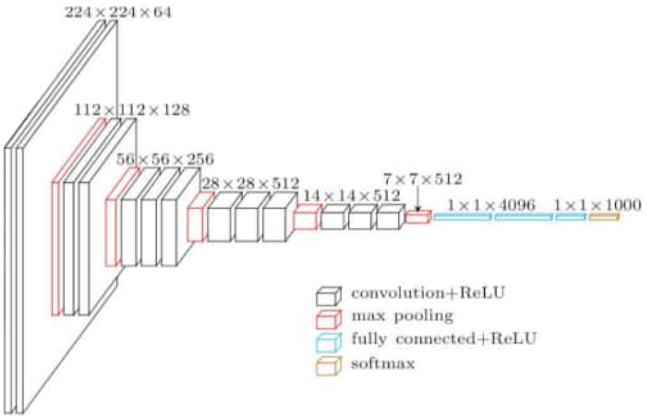


Рис. 11: Архітектура моделі VGG-16

роздільну здатність, при цьому розмір вихідної карти особливостей такий самий, як розміри вхідного зображення. Потім карта особливостей проходить через просторове об'єднання розміром 2 x 2 пікселя з кроком у 2 пікселі. Це удвічі зменшує застосування функції активації. Таким чином, карта особливостей в кінці першого стека становить 112 x 112 x 64.

На наступному етапі карта особливостей проходить через аналогічний другий стек, але зі 128 фільтрами. Отже, розмір після другого стека стає 56 x 56 x 128. За ним слідує третій стек із трьома згортковими шарами та шаром об'єднання. На цьому етапі застосовується 256 фільтрів, що надає вихідний стек розміром 28 x 28 x 256. Далі йдуть два стеки з трьох згорткових шарів, кожен з яких містить 512 фільтрів. Результатом, наприкінці обох цих стеків, буде карта особливостей розміром 7 x 7 x 512.

За стеками згорткових шарів слідують три повнозв'язні шари. Перші два шари мають по 4096 нейронів кожен, а останній повнозв'язний шар служить вихідним рівнем і містить таку кількість нейронів, що і кількість можливих класів класифікації. За вихідним рівнем слідує рівень активації Softmax, який використовується для класифікації. Дану схему візуалізовано на рисунку 12.

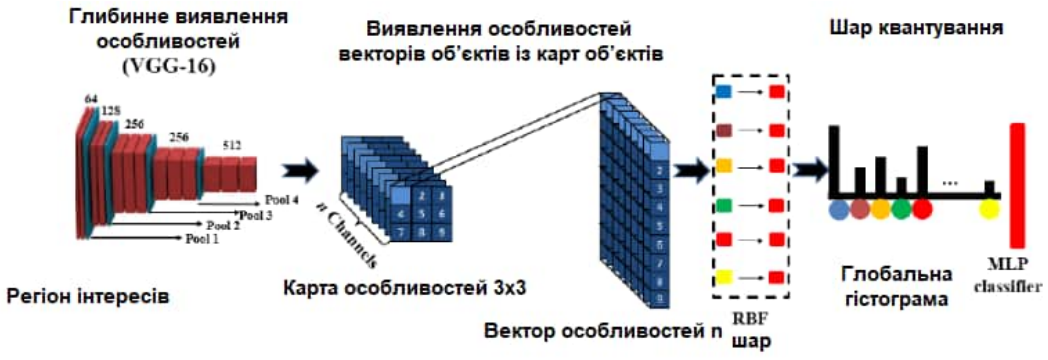
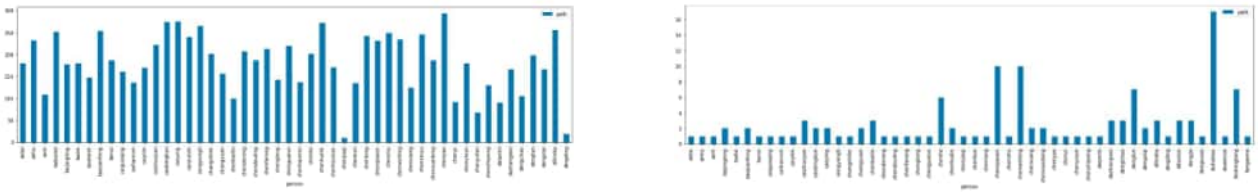


Рис. 12: Структура шарів моделі VGG-16

3 Результати

Для тренування було використано відкритий набір даних RMFRD. В даному наборі даних окремо розділені зображення людей в масках, та без масок. В кожної особи є декілька десятків фотографій в масці з різних ракурсів, та декілька фотографій без маски 13.



а) З масками б) Без масок

Рис. 13: Аналіз розподілення даних в наборі

Для того щоб проілюструвати переваги використання трансферного навчання, в задачі розпізнавання правильності носіння маски, порівняємо використані моделі MobileNetV2, Xception, CNN та VGG16. Порівняємо такі характеристики, як розмір мережі, кількість параметрів, споживання пам'яті та точність. Результати представлені в таблиці 3. Для тренування нейронних мереж було виконано по 40 епох для кожної. Тренування однієї епохи триває в середньому 75 хвилин.

Табл. 3: Підсумки результатів існуючих досліджень.

Модель	**Точність**	**Розмір**	**Кількість параметрів**	**Споживання пам'яті**
CNN	87.67%	27.1MB	2,203,557	72.58MB
VGG16	93.08%	62.4MB	288,357	18.06 MB
MobileNetV2	97.33%	20.8MB	1,094,373	226.67 MB
Xception	98.33%	96.6MB	1,074,789	368.18 MB

Аналізуючи отримані результати, можна зауважити, що в кожної моделі є свої сильні сторони. Модель VGG16 використовує найменшу кількість параметрів 288,357, серед інших порівнюваних моделей, та має найменше споживання оперативної пам'яті - всього 18.06MB. Модель MobileNetV2 має високу точність 97.33% та найменший розмір 20.8MB. Найкращу точність показала модель Xception 98.33%, проте недостатками цієї моделі є великий розмір 96.6MB та велике споживання оперативної пам'яті 368.18 MB.

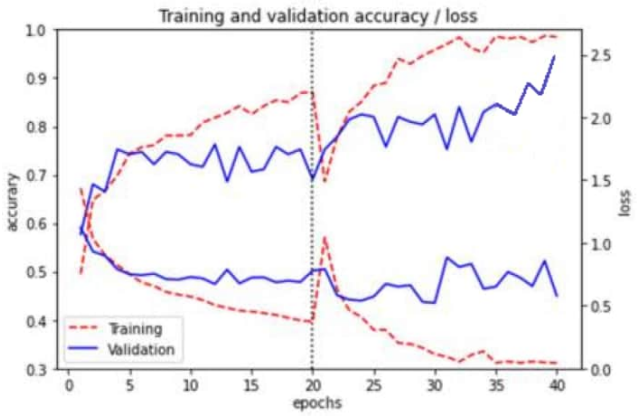


Рис. 14: Xception

Рис. 15: MobileNet V2

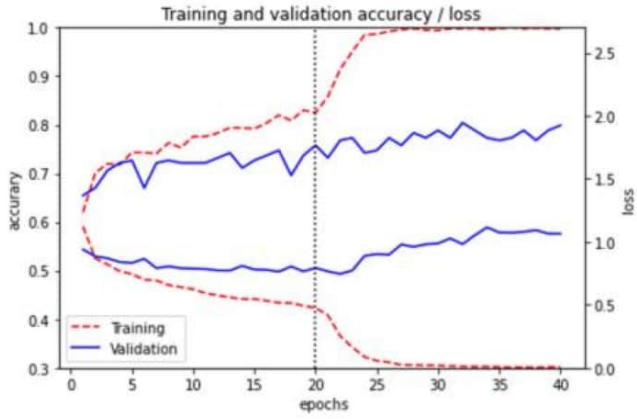


Рис. 16: CNN

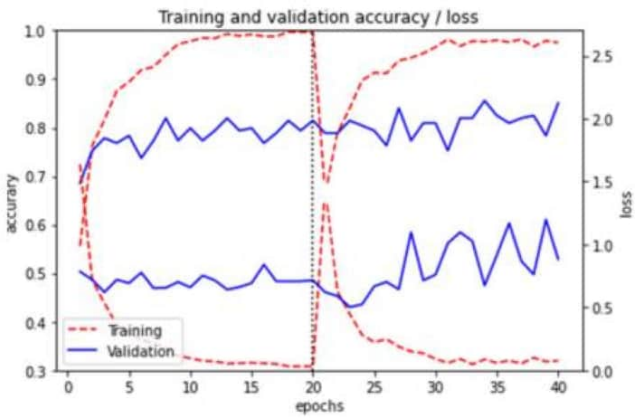


Рис. 17: VGG16

На рисунках 14-17 показано точність моделей під час навчання. На рисунках суміщено графіки функцій точності та логічних втрат. Графік точності показує статистику привильних класифікацій. Графік логічних втрат показує статистику штрафів за неправильні класифікації. Навчання проводилось у два етапи. В перших двадцяти епохах згорткові шари з оригінальної моделі для виділення загальних ознак залишались заблокованими, і навчались лише шари класифікації. У наступних епохах від 21 до 40 виконувалось точне налаштування, під час якого усі шари були розблоковані та вільно перенавчались.

Якщо ми детальніше проаналізуємо ці результати за допомогою матриці помилок на рисунку 18, то побачимо, що деякі типи помилок, такі як glasses (коли маска частково одягнута на окуляри) та side fit (коли маска неправильно одягнута збоку), мають нижчу точність, ніж інші.

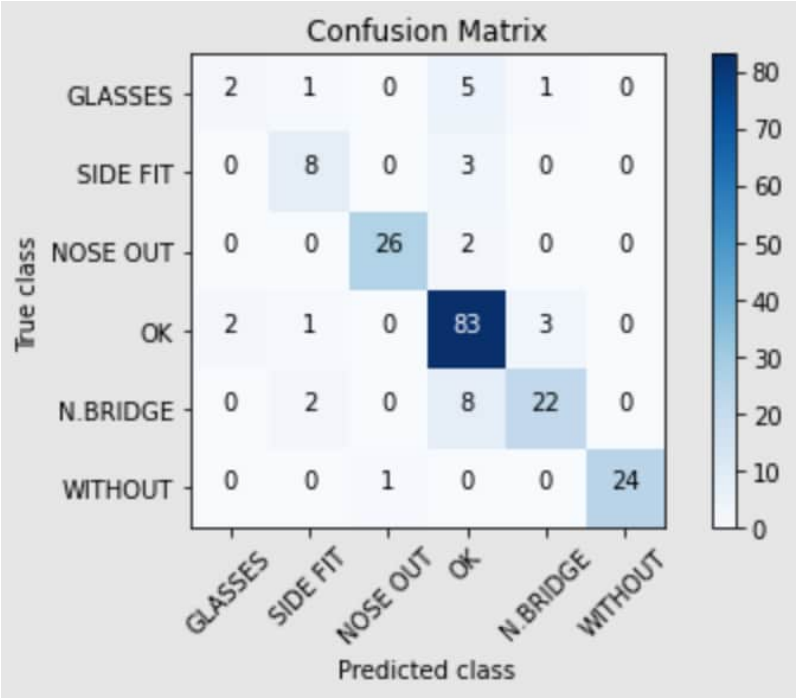


Рис. 18: Матриця помилок

Таблиця 4 показує результати класифікації для кожного з класів за допомогою моделі VGG16. Влучність можна інтерпретувати, як частку об'єктів, які класифікатор розпізнав, як позитивні, і при цьому зробив це правильно. Повнота показує, яку частку об'єктів позитивного класу з усіх об'єктів позитивного класу знайшов алгоритм. F-міра - це середнє арифметичне влучності та повноти.

Табл. 4: Результати класифікації.

Клас	**Влучність**	**Повнота**	**F-міра**
GLASSES	0.56	0.63	0.59
FIT SIDE	0.73	0.67	0.70
NOSE OUT	0.93	0.90	0.91
CORRECT	0.87	0.88	0.87
N.BRIDGE	0.75	0.75	0.75
NO MASK	1.00	1.00	1.00
total	0.84	0.85	0.85

Таку, неідеальну точність, можна пояснити тим, що в наборі даних присутні нечіткі зображення, або зображення на яких обличчя людини частково перекриті. В таких випадках важко визначити чи маска правильно одягнута, навіть людським оком. У всіх інших випадках, система показує стабільно високі результати, які ми можемо побачити на рисунках 19-20.

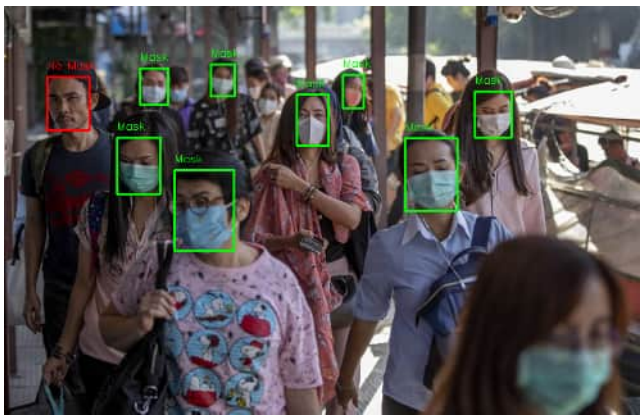


Рис. 19: Приклад роботи моделі CNN

Рис. 20: Приклад роботи моделі VGG16

ВИСНОВКИ

Визначення неправильного носіння маски виявилось достатньо непростою задачею. Основну проблему склала обмежена кількість необхідних зображень в наборах даних. Набори даних, як правило, невеликі та розраховані лише на визначення наявності маски. Для часткового вирішення цієї проблеми кількість необхідних зображень була збільшена за допомогою клонування зображень із внесенням невеликих змін, таких як невеликі повороти, віддзеркалення та масштабування.

Під час огляду існуючих досліджень, було з'ясовано, що раніше схожі дослідження не проводились. Інші дослідження були сконцентровані лише на найбільш популярній помилці, яка полягає в носінні маски під носом. Реалізовані моделі в даній роботі здатні виявляти й інші помилки носіння маски, такі як одягання маски поверх окулярів та погане прилягання збоку, які трапляються дуже часто. Однак, через невелику кількість необхідних зображень для навчання, ці типи помилок інколи визначаються неправильно.

Результати підтверджують можливість використання реалізованих моделей для розв'язання реальних задач, що дає змогу запобігти поширенню пандемії. Єдиною складністю з якою можуть стикнутись моделі це різнокольорові маски з малюнками. Останнім часом такі маски набувають популярності, адже вони стали ще одним доповненням до нашого одягу. Багато масок мають блискітки, інші малюнки, починаючи від простих фігур і закінчуючи намальованими посмішками. В такому випадку моделі можна вдосконалити. Удосконалення систем полягатиме у виявленні цих модифікованих масок.

Дані системи можна застосовувати для різних сценаріїв. Наприклад для камер спостережень в метро, торговельних центрах, аеропортах, лікарнях або в інших громадських місцях, щоб запобігти проблемі поширення вірусу, подаючи відповідне сповіщення. Це може бути автоматичне голосове повідомлення, або оператор системи може передавати інформацію до охоронців, які в свою чергу зроблять зауваження порушнику.

В дипломній роботі розглянуто задачу визначення правильності носіння маски в умовах карантину. Стосовно неї отримано такі основні результати:

1. Проаналізовано існуючі дослідження.
2. Проаналізовано відкриті набори даних.
3. Створено програмну реалізацію алгоритму визначення правильності носіння маски.
4. Визначено точність запропонованих моделей та продемонстровано визначення правильності носіння маски на конкретних прикладах.

Отримані результати – краща точність розпізнавання 98% – є основою для подальших досліджень завдання визначення правильності носіння маски в умовах карантину. Спостерігаються неточності в роботі моделей при великому перекритті обличчя. Це пояснюється тим, що ключові точки, по яких відбувається розпізнавання не доступні для порівняння. Іншою проблемою є використання масок з малюнками. Вирішенням даних проблем може стати застосування гібридних нейронних мереж.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Nishiura H., Jung S.M., Linton N.M., Kinoshita R., Yang Y., Hayashi K., Kobayashi T., Yuan B., Akhmetzhanov A.R. Заразність нового коронавірусу в Ухані, China, 2020.

[2] Центри контролю та профілактики захворювань CDC 24/7. Популярні запитання про засоби індивідуального захисту. Доступно за посиланням: https://www.cdc.gov/coronavirus/2019-ncov/hcp/respirator-use-faq.html

[3] Chung Y., Chou C.A., Li C.Y. Нейронна мережа для розпізнавання видів дерев у реальному світі. 2021;18:961.

[4] Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z. Переосмислення початкової архітектури комп'ютерного зору; Матеріали конференції IEEE з комп'ютерного зору та патерни розпізнавання (CVPR); Las Vegas, NV, USA. 27–30 June 2016; ст. 2818–2826.

[5] Nagrath P., Jain R., Madan A., Arora R., Kataria P., Hemanth J. SSDMNV2: Система виявлення маски на обличчі в режимі реального часу на основі DNN, використовуючи детектор одиночної дії Multibox і MobileNetV2. Sustain. Cities Soc. 2021;66:102692.

[6] Jauhari A., Anamisa D.R., Negara Y.D.P. Система виявлення ключових точок обличчя в масках на основі методу Віоли Джонс. J. Phys. Conf. Ser. 2021;1836:012035.

[7] Sen S., Sawant K. Серія конференцій IOP: Матеріалознавство та інженерія. Volume 1070. IOP Publishing; Bristol, UK: 2021. Виявлення масок для пандемії COVID-19 за допомогою pytorch у глибинному навчанні; ст. 12061.

[8] Kurlekar M.S. Система виявлення масок із використанням глибинного навчання. Turk. J. Comput. Math. Educ. Turcomat. ст. 1327–1332

[9] Sakshi S., Gupta A.K., Yadav S.S., Kumar U. Система виявлення масок за допомогою CNN; Greater Noida, India. 4–5 March 2021; ст. 212–216

[10] Cheng G., Li S., Zhang Y., Zhou R. Система виявлення маски на основі Yolov3-Tiny. Front. Soc. Sci. Technol. 2020;2:33–41. doi: 10.25236/FSST.2020.021106.

[11] Loey M., Manogaran G., Taha M.H.N., Khalifa N.E.M. Гібридна модель трансферного навчання з методами машинного навчання для виявлення масок для обличчя в епоху пандемії COVID-19. Measurement. 2021;167:108288. doi: 10.1016/j.measurement.2020.108288.

[12] Rudraraju S.R., Suryadevara N.K., Negi A. Виявлення масок з допомогою Fog Computing Gateway; Матеріали 15-ї конференції IEEE 2020 з комп'ютерних наук та інформаційних систем (FedCSIS); Sofia, Bulgaria. 6–9 September 2020; ст. 521–524.

[13] Wang B., Zhao Y., Chen C.P. Гібридна система трансферного навчання та система Broad Learning для виявлення носіння масок в епоху COVID-19. IEEE Trans. Instrum. Meas. ст. 1–12.

[14] Qin B., Li D. Визначення стану носіння маски за допомогою суперроздільності зображення з мережею класифікації для запобігання поширення COVID-19. Sensors. 2020;20:5236. doi: 10.3390/s20185236.

[15] Din, N.U.; Javed, K.; Bae, S.; Yi, J. Нова мережа на основі GAN для визначення носіння маски. IEEE Access 2020, 8, ст. 44276–44287

[16] Liu, Z.; Luo, P.; Wang, X.; Tang, X. Глибинне вивчення атрибутів обличчя в реальному світі. Матеріали Міжнародної конференції IEEE з комп'ютерного зору, Santiago, Chile, 7–15 Грудня 2015; ст. 3730–3738.

[17] Din, N.U.; Javed, K.; Bae, S.; Yi, J. Ефективне виділення вибраного користувачем об'єкта із зображень обличчя за допомогою нейронної мережі на основі GAN. IEEE Access 2020, 8, ст. 109648–109661

[18] Wang, Z.; Wang, G.; Huang, B.; Xiong, Z.; Hong, Q.; Wu, H.; Yi, P.; Jiang, K.; Wang, N.; Pei, Y.; et al. Набір даних і додаток для розпізнавання обличчя під маскою. arXiv 2020, arXiv:2003.09093

[19] Geng, M.; Peng, P.; Huang, Y.; Tian, Y. Розпізнавання облич в масках із штучним розширенням даних і обмеженим доменним рейтингом. У матеріалах 28-ї міжнародної конференції ACM з мультимедіа, Seattle, WA, USA, 12–16 Жовтень 2020; ст. 2246–2254.

[20] Sengupta, S.; Chen, J.-C.; Castillo, C.; Patel, V.M.; Chellappa, R.; Jacobs, D.W. Фронтальне розпізнавання обличчя. У матеріалах зимової конференції IEEE 2016 із застосування комп'ютерного зору (WACV), Lake Placid, NY, USA, 7–9 Березень 2016; ст. 1–9.

[21] Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; Zafeiriou, S. Agedb: Перший набір зображень із людьми в масках зібраний вручну. У матеріалах конференції IEEE з семінарів з комп'ютерного зору та розпізнавання облич, Honolulu, HI, USA, 21–26 липень 2017; ст. 51–59.

[22] Afzal, H.R.; Luo, S.; Afzal, M.K.; Chaudhary, G.; Khari, M.; Kumar, S.A. 3D-реконструкція обличчя з одного 2D-зображення з використанням характерних ознак. IEEE Access 2020, 8, ст. 180681–180689

[23] Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Додаткові кутові втрати для глибинного розпізнавання обличчя. У матеріалах конференції IEEE/CVF з комп'ютерного зору та розпізнавання облич, Long Beach, CA, USA, 15–20 Червень 2019; ст. 4690–4699.

[24] Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Демаскування облич для точного розпізнавання облич під маскою. arXiv 2021, arXiv:2103.01716

[25] Документація по Keras. Доступно за посиланням: `https://keras.io/api/utils/backend_utils/`