

DOI: 10.20535/kpissn.2021.3.251462

УДК 004.85

Н.І. Недашківська*, О.О. Сапельніков
КПІ ім. Ігоря Сікорського, Київ, Україна
*corresponding author: n.nedashkivska@gmail.com

ПОРІВНЯЛЬНИЙ АНАЛІЗ МОДЕЛЕЙ ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ПРОГНОЗУВАННЯ ДІАГНОЗУ COVID-19

Received: 2 Oct. 2021. Accepted: 7 Dec. 2021.

Проблематика. Ситуація з коронавірусом COVID-19 станом на середину 2021 року знову викликає занепокоєння у всьому світі через появу нового небезпечнішого штаму “Дельта”. Науковці намагаються побудувати математичні моделі опису поширення захворюваності. У Китаї, США й інших країнах розробляють інструменти штучного інтелекту, спрямовані на прогнозування діагнозу COVID-19, і впроваджують їх у лікарнях.

Мета дослідження. Порівняти розпізнавання різними моделями глибоких згорткових нейронних мереж хвороби COVID-19 на основі рентген-зображень грудної клітки людини.

Методика реалізації. Моделі побудовано на основі архітектур мереж VGG16, ResNet50 і SqueezeNet із додаванням шарів для регуляризації цих мереж. Перед останнім, вихідним шаром мережі додавали шар дропауту. Навчання моделей здійснювали з використанням технології передачі знань transfer learning. Якість моделей оцінювали за матрицею помилок (confusion matrix), показниками точності (precision), повноти (recall/sensitivity), специфічності (specificity) і F-міри на валідаційній/тестовій множині рентген-зображень.

Результати дослідження. Побудовано кілька моделей нейронних мереж, що поділяють рентген-зображення грудної клітки людини на класи: хворий на COVID-19 (клас 1) або здоровий (клас 2). Модель на основі архітектури ResNet50 отримала досить велике значення precision 96 % для класу 1 і загальне значення specificity 96,14 %. Recall виявлення захворювання на основі цієї моделі склала 88 %. Модель на основі архітектури VGG-16 правильно класифікувала 100 % (експеримент 1) / 96 % (експеримент 2) хворих на COVID-19 із тестової вибірки. Значення precision для класу хворих і значення загальної specificity на основі цієї моделі, однак, були меншими й склали 89 % / 92 % і 88,46 % / 92,31 % відповідно для двох експериментів. Значення F-міри для моделей ResNet50 і VGG-16 були досить значними й рівними 92 і 94 %.

Висновки. Моделі глибоких згорткових нейронних мереж показали обнадійливі результати прогнозування діагнозу COVID-19 і потребують досліджень. Прогнозування діагнозу за допомогою розробленого програмного продукту на основі цих моделей триває до однієї хвилини – це швидше, ніж тести RT-PCR (ПЛР із зворотною транскрипцією), які нині використовують для підтвердження захворювання.

Ключові слова: згорткові нейронні мережі; глибинне навчання; класифікація зображень; прогнозування діагнозу COVID-19.

Вступ

Всесвітньою організацією охорони здоров'я в березні 2020 року проголошено пандемію коронавірусу COVID-19 [1]. Ця хвороба має досить великий потенціал розвитку та поширення, з'являються нові небезпечніші штами коронавірусної інфекції. Новий тип вірусу і завдає шкоди здоров'ю населення всіх країн світу, і серйозно впливає на національні економіки. Станом на середину вересня 2021 року на COVID-19

захворіли понад 226 мільйонів людей, померли від цього вірусу вже понад 4,5 мільйони [2]. Найбільша кількість смертей станом на середину вересня 2021 року – у США, Індії та Великій Британії [3]. В Україні на час написання статті лабораторно підтверджено понад 2 мільйони інфікованих на COVID-19, із них майже 55 тисяч смертельних випадків [4]. Науковці розробляють методи безпосереднього порівняння розвитку епідемії між країнами з різними стратегіями її стримування [5]. За результатами наукових роз-

Рекомендуємо цитувати цю статтю так: Н.І. Недашківська, О.О. Сапельніков, “Порівняльний аналіз моделей глибоких нейронних мереж для прогнозування діагнозу COVID-19”, *Наукові вісті КПІ*, № 3, с. 46–56, 2021. doi: 10.20535/kpissn.2021.3.251462.

Please cite this article as: N.I. Nedashkovskaya, and O.O. Sapelnikov, “Comparative analysis of deep neural network models for prediction of COVID-19 diagnosis”, *KPI Science News*, no. 3, pp. 46–56, 2021. doi: 10.20535/kpissn.2021.3.251462.

відок, раннє стримування є ключовим фактором у згладжуванні епідемічної кривої [5].

Протягом останнього року робляться спроби побудувати моделі поширення коронавірусу COVID-19 в окремих країнах і в усьому світі [6]–[14]. На основі моделі ARIMA в [6] спрогнозовано кількість підтверджених випадків, смертей і одужань від COVID-19 у Пакистані. Гібридні моделі ARIMA та вейвлет-аналізу застосовано для прогнозування кількості смертей [7] і щоденних випадків захворюваності [8] на COVID-19 у різних країнах світу. На основі SIR-моделі в [9] проаналізовано та спрогнозовано розповсюдження COVID-19 у Китаї, Італії та Франції. У [10] на основі модифікованої SEIR-моделі зроблено спробу спрогнозувати потреби в приміщеннях і лікарняних ліжках в Чилі за використання різних стратегій втручання з боку держави, як-от: блокування та поєднання стратегій часткової ізоляції, закриття шкіл й університетів, домашнього карантину та соціальної відстані. Модель SARIqSq, яка описує динаміку поширення COVID-19 [11], також базується на епідеміологічній SEIR-моделі. На основі цієї моделі [11] прогнозували терміни закінчення COVID-19 у кожній із 17 провінцій Індії. Аналіз і прогнозування поширення COVID-19 в Індії виконували також за допомогою генетичного алгоритму [12]. У [13] для апроксимації відносної кількості щоденних нових заражень в Італії в момент часу t використано розподіл Пуассона та розраховано ймовірне число нових заражень на наступний місяць. Узагальнену логістичну модель зростання пандемії COVID-19 в Азії запропоновано в [14].

Нині багато дослідників активно шукають нові та швидкі способи діагностики захворювання на COVID-19 [15]–[26]. Один із таких способів полягає у розпізнаванні ознак хвороби на рентген-зображеннях грудної клітки [20]–[26]. Для цього, зокрема, створюють моделі глибокого навчання [22]–[24]. Хоча рентген поступається в точності діагностики комп'ютерній томографії, він має ряд переваг, які обґрунтовано в [16]. Рентген грудної клітки розглядають як інструмент першої лінії для визначення патологій легень у пацієнтів із симптомами COVID-19; портативні рентгенографічні апарати пропонується встановлювати у відділеннях, призначених для приймання та лікування пацієнтів із підозрою на COVID-19 [20]. Китайські компанії розробляють інструменти штучного інтелекту, спрямовані на прогнозування діагнозу COVID-19, і розгортають їх у лікарнях [15].

Постановка задачі

На основі навчальної вибірки рентген-зображень здорової та хворої на COVID-19 людини побудувати моделі класифікації зображень на два класи: здоровий або має вказану хворобу. Для побудови моделей використати різні архітектури глибоких згорткових нейронних мереж. Обрати найкращу модель класифікації за множиною метрик якості. Розробити інструментарій для прогнозування ймовірності діагнозу COVID-19 для індивідуального рентген-зображення на основі обраної моделі.

Початкові дані для навчання моделей

Існують великі загальнодоступні набори даних рентген-зображень грудної клітки зі Стенфорда [27], Массачусетського технологічного інституту [3] та Національного інституту охорони здоров'я США [28]. Для нашого дослідження необхідно мати колекції рентген-зображень або комп'ютерної томографії грудної клітки ураженої COVID-19 людини, адаптованих для використання в машинному навчанні. Нами використано базу даних [25], [26], яка є у вільному доступі та містить рентген-зображення грудної клітки здорових і хворих на COVID-19 людей, а також людей з іншими легеневиими захворюваннями, спричиненими MERS, SARS й ARDS. До розгляду й аналізу відібрано рентген-зображення двох класів: здорових і хворих на COVID-19 людей. Множину зображень розбито на підмножини: навчальну та валідаційну/тестову.

Коронавірус COVID-19, що з'явився у 2019 році, має кілька особливостей [16], [17]. Відомо, що діагноз підтверджують за допомогою полімеразної ланцюгової реакції (ПЛР), а зображення рентгенограм грудної клітки пацієнтів, інфікованих COVID-19, характеризуються малюнком, який людське око виявляє лише невпевнено [18]. Пацієнти з COVID-19 мають аномалії на зображеннях комп'ютерної томографії грудної клітки, частіше з двостороннім враженням [19].

Архітектури глибоких згорткових нейронних мереж

Згорткові мережі – це різновид нейронних мереж, які складаються з фільтрів або згорток (рис. 1). Кожен такий фільтр унаслідок навчання нейронної мережі вчиться розрізняти певні образи. Традиційно згорткові мережі застосовують саме для класифікації зображень, а резуль-

татом останнього шару мережі є ймовірності належності зображення до кожного з класів. Перша згорткова мережа LeNet запропонована Я. Лекуном у 1989 році для розпізнавання рукописних символів [29]. Сучасний стрімкий розвиток згорткових нейронних мереж розпочався у 2012 році, коли зросли обчислювальні потужності комп'ютерів для швидкої роботи мереж і стали доступними більше даних для навчання. У 2012 році з'явилася мережа AlexNet, на основі якої у 2014 році створено мережі Network-in-Network, VGG і першу архітектуру Inception під назвою GoogLeNet [30].



Рис. 1. Приклад структури згорткової мережі, адаптовано за [30]

У мережі VGG у кожному згортковому шарі вперше використано малі фільтри 3×3 й об'єднано ці шари у послідовність згорток [30], [31]. Авторами VGG показано, що послідовність двох шарів із фільтрами 3×3 еквівалентна одному шару з фільтрами 5×5 , і ця послідовність економить 22 % параметрів мережі. Послідовність із трьох шарів із фільтрами 3×3 замінює один шар із фільтрами 7×7 , а це економить 55 % параметрів.

У 2014 році мережа VGG16 брала участь у змаганні ILSVRC із розв'язання великомасштабних задач візуального розпізнавання та ввійшла до п'ятірки найточніших моделей [31]. Ідеї, що лежать в основі VGG16, були використані в мережах наступного покоління, зокрема в GoogLeNet і ResNet. Модель VGG16 успішно використано в багатьох практичних задачах, а нині вона є однією з найвідоміших моделей. Її недоліки — це повільна швидкість навчання та значні обчислювальні витрати, пов'язані з великою кількістю властивостей у багатьох шарах мережі.

У 2015 році компанія Microsoft запропонувала мережу ResNet (Residual Network), що містить тисячі шарів, і нову ідею — використати обхідний зв'язок двох згорткових шарів (рис. 2) [32], [33]. Проблема наявних натоді згорткових мереж полягала в тому, що зі збільшенням кількості шарів точність спочатку збільшується, а потім стрімко зменшується. Точність

знижується як на валідаційній, так і на навчальній множині. Щоб подолати цю проблему, Microsoft увела глибоку “залишкову” (residual) структуру навчання.

Нехай оригінальна мережа має обчислювати функцію $H(x)$. Її залишкова функція визначена як $F(x) = H(x) - x$. Зроблено припущення, що $F(x)$ простіше навчається мережею, ніж початкова функція $H(x)$. Формулювання $F(x) + x$ реалізовано мережею ResNet із застосуванням обхідного зв'язку чи з'єднання для швидкого доступу (shortcut connection) (див. рис. 2).

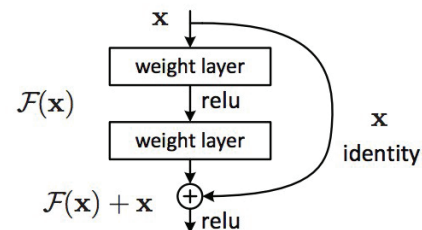


Рис. 2. З'єднання для швидкого доступу в мережі ResNet [34]

Мережа ResNet — це система одночасно паралельних і послідовних модулів: у багатьох модулях сигнал проходить паралельно, а вихідні сигнали кожного модуля з'єднуються послідовно [34]. Цікаво, що мережу ResNet можна вважати ансамблем, який складається з менших залишкових нейронних мереж, де ефективна глибина збільшується в процесі навчання [35].

Використання мережі ResNet дає змогу вирішити кілька проблем [34]. По-перше, її відносно легко оптимізувати. По-друге, достатньо легко збільшити точність результатів шляхом збільшення глибини цієї мережі, чого з іншими мережами домогтися складніше.

Розвиток мобільних технологій і поширеність смартфонів зумовили створення мереж SqueezeNet і MobileNet [36]. SqueezeNet складається з багатьох шарів, кожен з яких містить малі згортки. Так кількість параметрів зменшується в декілька разів у порівнянні з AlexNet і VGG, а показники точності класифікації зберігаються.

Нейронні мережі з великою кількістю шарів сформували окремий напрямок — “глибине навчання”. У 2015 році глибокі згорткові нейронні мережі перевершили рівень класифікації зображень, якого досягла людина [36]. VGG16, SqueezeNet і ResNet50 — одні з найбільш значущих архітектур глибоких згорткових мереж для задачі класифікації зображень. Вони дали змогу значно покращити точність розпізнавання

та досягнути результатів, не доступних класичним методам комп'ютерного зору.

Побудова моделей прогнозування діагнозу коронавірус COVID-19

Моделі прогнозування діагнозу коронавірусної інфекції побудовано на основі відомих архітектур глибоких згорткових нейронних мереж VGG16, ResNet50 і SqueezeNet із додаванням спеціальних шарів для їх регуляризації. Остання є одним із рішень для перенавчання нейронної мережі [36]. Серед різних методів регуляризації нейронних мереж для цієї задачі обрано метод виключення чи дропауту (dropout), оскільки він мав найкращі показники при розв'язанні задач класифікації зображень. Застосування дропауту до мережі рівнозначне навчання ансамблю нейронних мереж з наступним усередненням отриманих результатів. Мережі ансамблю будують виключенням нейронів із мережі: за будь-яких вхідних даних або параметрах нейрон повертає 0 із певною ймовірністю p . Ймовірність збереження нейрона в мережі становить $q = 1 - p$. Виключені нейрони не враховують у процесі навчання, тому виключення хоча б одного з нейронів рівнозначно навчання нової нейронної мережі.

Нехай $h(x)$ – вихід лінійної частини нейрона, коли на його вхід подають вектор x ; d_n – вимірність простору вихідних значень цього нейрона; $a(h)$ – функція активації. Застосування дропауту на етапі навчання представляють зміненою функцією активації $F(h) = D \odot a(h)$, $D = (x_1, \dots, x_{d_n})$ – це $(d_n - 1)$ -вимірний вектор випадкових величин, розподілений за законом Бернуллі.

У роботі побудовано та досліджено такі моделі.

Модель 1 заснована на архітектурі мережі VGG16 із додаванням шару дропауту перед останнім, вихідним шаром мережі.

Модель 2 заснована на архітектурі VGG16 без додавання регуляризації.

Моделі 3 і 4 побудовані на архітектурі SqueezeNet і відрізняються додаванням або виключенням шарів дропауту.

Моделі 5 і 6 базуються на архітектурі ResNet50 відповідно з додаванням шару дропауту перед останнім, вихідним шаром мережі (модель 5) і без регуляризації (модель 6).

Як функцію втрат в усіх побудованих моделях обрано бінарну перехресну ентропію, оскільки маємо задачу класифікації на два класи. Кожну модель навчали по черзі двома алгоритмами

оптимізації: Адам і стохастичного градієнтного спуску (SGD). Це зроблено задля оцінювання впливу алгоритму оптимізації на результати навчання моделей і вибору кращого алгоритму для цієї задачі.

Навчання моделей здійснювали з використанням технології передання знань (transfer learning). Її суть у тому, що для кожної нової практичної задачі немає потреби навчати нейронну мережу “з нуля”, тобто починаючи з випадкових початкових значень ваг. Натомість доцільно використати результати попереднього навчання мережі для певної близької задачі. Це значно скорочує витрати в часі та кількості обчислень, і немає суттєвої втрати точності для багатьох практичних задач. Так, у цій роботі використано ваги мереж VGG16 і ResNet50, попередньо отримані внаслідок навчання цих мереж на надвеликому наборі зображень ImageNet. На основі цих ваг, що є у вільному доступі, проведено донавчання мереж VGG16 і ResNet50 на обраному наборі рентген-зображень грудної клітки для виявлення характеристик рентген-зображень хворих на COVID-19 і їх використання у поставленій задачі прогнозування діагнозу.

Якість моделей оцінювали за значеннями функції втрат (loss) і таких метрик: матриця помилок (confusion matrix), точність (precision), повнота (recall/sensitivity), специфічність (specificity), F-міра (F1-score) і правильність (accuracy), розрахованих на валідаційній/тестовій множині [37]–[41]. Нехай TP – істиннопозитивний результат: людина, дійсно інфікована COVID-19, правильно визначається моделлю як хвора на COVID-19; TN – істиннонегативний результат: людина правильно визначається моделлю як така, що не хворіє на COVID-19; FP – хибнопозитивний: хибне виявлення, коли у здорової людини модель виявляє позитивний результат на COVID-19; FN – хибнонегативний: хибне виявлення в разі, якщо людина, інфікована COVID-19, визначається моделлю як здорова.

Точність (precision) моделі визначають за формулою

$$\text{precision} = TP / (TP+FP),$$

вона вказує частку насправді хворих пацієнтів, у яких модель діагностувала хворобу.

Повнота (recall) або чутливість (**sensitivity**):

$$\text{recall} = TP / (TP+FN)$$

показує частку істиннопозитивних класифікацій, а саме частку хворих пацієнтів, правильно виявлених класифікатором.

Специфічність (specificity):

$$\text{specificity} = \text{TN} / (\text{FP} + \text{TN}).$$

Кращим є класифікатор із високими значеннями precision, recall і specificity. Для задач медичної діагностики саме показник recall має особливе значення, оскільки показує кількість хибнонегативних випадків (FN) виявлення хвороби. Тому в цій задачі висновки зроблено переважно на основі аналізу співвідношення між показниками precision і recall.

F-міра (F1-score) – агрегований показник – є гармонійним середнім двох попередніх метрик:

$$\text{F1-score} = 2 * \text{precision} * \text{recall} / (\text{precision} + \text{recall}).$$

Гармонійне середнє надає низьким значенням більшу вагу, на відміну від середнього арифметичного. Як наслідок, F-міра набуде великого значення лише за умови, що й precision, і recall одночасно мають великі значення.

Правильність (accuracy) – це частка правильних відповідей у загальній кількості відповідей:

$$\text{accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}).$$

Показник accuracy є ефективною мірою якості класифікації у задачах зі збалансованими класами. Цей показник потрібно обережно використовувати у задачах діагностики захворювань: якщо кількість об'єктів одного класу (здорових) значно перевищує кількість об'єктів іншого класу (хворих), то високі значення показника accuracy не є правдивими.

Розроблено програмний модуль у середовищі python3 (рис. 3), який має функції побудови моделей, вибору найкращої моделі та прогнозування діагнозу COVID-19 на основі цієї моделі. При розробці програмного забезпечення використано сучасні open-source бібліотеки Tensorflow [37]–[42] і Keras [43].



Рис. 3. Схема розробленого програмного продукту

Результати експериментів

Виконаємо порівняльний аналіз наведених вище моделей 1–6. На рис. 4 наведено графіки значень train_loss – функції втрат на навчальному наборі, val_loss – функції втрат на валідаційному/тестовому наборі, train_acc і val_acc – значення показника правильності accuracy на навчальному та валідаційному наборах для моделі 1 залежно від кількості епох навчання. Як видно з рис. 4, значення функції втрат на навчальній і перевірчій множинах наближаються одне до одного та в середньому зменшуються зі збільшенням кількості епох. Значення accuracy на навчальній і валідаційній множинах також є близькими між собою і набувають достатньо великих значень зі збільшенням кількості епох. Тому на основі рис. 4 зроблено висновок про відсутність перенавчання моделі 1. Матриця помилок confusion matrix (рис. 5) показує, що всі рентген-зображення хворих людей із тестової множини були правильно класифіковані моделлю як хворі; цей висновок маємо й на основі значення повноти recall/sensitivity, що дорівнює 1,00 (див. рис. 5). Однак повнота виявлення здорових людей була меншою і склала 88 %, а 12 % здорових людей були хибно віднесені моделлю до хворих. Середнє значення accuracy моделі 1 дорівнює 94,12 %. Показник F1-score на валідаційній множині зображень набуває доволі великого значення 94 % (див. рис. 5).

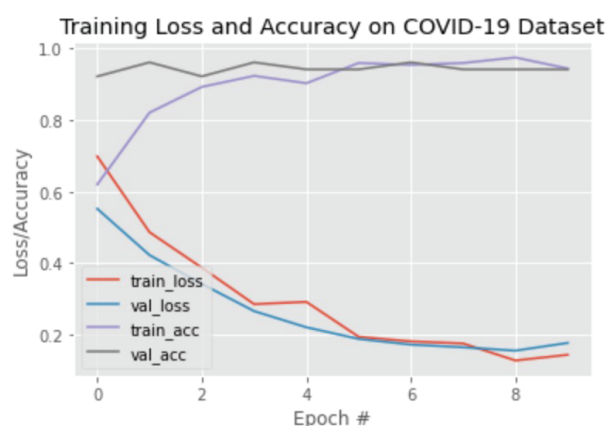


Рис. 4. Результати навчання та валідації моделі 1

У моделі 2 не спостерігалось перенавчання (рис. 6), і значення повноти recall були більш збалансованими в порівнянні з моделлю 1. Матриця помилок confusion matrix (рис. 7)

показала, що лише 96 % хворих на COVID-19 були класифіковані безпомилково, а у 4 % випадків модель помилилася: людина, інфікована COVID-19, визначалася моделлю як здорова. 92 % здорових людей були правильно класифіковані як здорові, 8 % здорових були віднесені моделлю до хворих. Значення F1-score було таким самим, як для моделі 1, і дорівнювало 94 % (див. рис. 7).

| confusion matrix | precision | recall | f1-score |
|---------------------|-------------------|--------|----------|
| [[25 0] [3 23]] | | | |
| acc: 0.9412 | covid 0.89 | 1.00 | 0.94 |
| sensitivity: 1.0000 | normal 1.00 | 0.88 | 0.94 |
| specificity: 0.8846 | accuracy 0.94 | | 0.94 |
| | macro avg 0.95 | 0.94 | 0.94 |
| | weighted avg 0.95 | 0.94 | 0.94 |

Рис. 5. Матриця помилок і значення метрик якості на валідаційній множині для моделі 1

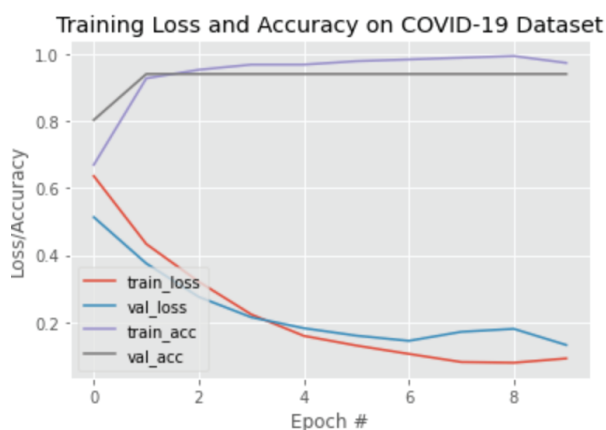


Рис. 6. Результати навчання та валідації моделі 2

| confusion matrix | precision | recall | f1-score |
|---------------------|-------------------|--------|----------|
| [[24 1] [2 24]] | | | |
| acc: 0.9412 | covid 0.92 | 0.96 | 0.94 |
| sensitivity: 0.9600 | normal 0.96 | 0.92 | 0.94 |
| specificity: 0.9231 | accuracy 0.94 | | 0.94 |
| | macro avg 0.94 | 0.94 | 0.94 |
| | weighted avg 0.94 | 0.94 | 0.94 |

Рис. 7. Матриця помилок і значення метрик якості на валідаційній множині для моделі 2

Додавання шару дропауту в архітектурі SqueezeNet (модель 3) призвело в цій задачі до небажаного перенавчання моделі. Тому модель 3 виключено з розгляду. Для моделі 4 на рис. 8 наведено значення функції втрат на навчальному наборі (loss), на валідаційному наборі (val_loss),

показника асигуру на навчальному та валідаційному наборах (acc і val_acc відповідно) для різної кількості епох навчання. Порівняння рис. 5, 7 і 9 свідчить про те, що відсоток виявлення хворих осіб на тестовій множині був таким самим великим, як у попередньої моделі 2, і склав 96 %. Проте за всіма іншими показниками якості моделі 4 виявилася гіршою, як порівняти з попередніми моделями 1 і 2 (рис. 9): середнє значення правильності моделі 4 склало 66 %; показник F1-score становив 74 і 51 % для першого та другого класів. 36 % рентген-зображень здорових людей були класифіковані правильно, 64 % – хибно (див. рис. 9). Тому модель 4 є гіршою за моделі 1 і 2.

| | loss | acc | val_loss | val_acc |
|----------|----------|-------|----------|---------|
| ЕPOCH 1 | 0.452237 | 0.785 | 1.437321 | 0.50 |
| ЕPOCH 2 | 0.413548 | 0.850 | 0.677107 | 0.64 |
| ЕPOCH 3 | 0.420252 | 0.835 | 0.363857 | 0.84 |
| ЕPOCH 4 | 0.348809 | 0.865 | 0.369369 | 0.80 |
| ЕPOCH 5 | 0.319194 | 0.880 | 1.517003 | 0.52 |
| ЕPOCH 6 | 0.315752 | 0.885 | 0.487907 | 0.76 |
| ЕPOCH 7 | 0.245385 | 0.920 | 0.275911 | 0.86 |
| ЕPOCH 8 | 0.231226 | 0.915 | 1.497127 | 0.66 |
| ЕPOCH 9 | 0.202359 | 0.920 | 2.053224 | 0.50 |
| ЕPOCH 10 | 0.282645 | 0.885 | 0.530321 | 0.66 |

Рис. 8. Значення функції втрат loss і показника асигуру в моделі 4

Модель 5 на основі архітектури ResNet50 визнано недостатньо навченою, бо значення асигуру на навчальній множині (train_acc) були нижчими за відповідні значення val_acc на валідаційній множині (рис. 10). Рентген-зображення хворої людини були класифіковані правильно лише у 88 % випадків, а в 12 % хвора

на COVID-19 людина була визнана здоровою. Проте модель 5 краще за всі розглянуті моделі виявила здорових людей: 96 % здорових були правильно класифіковані здоровими, у 4 % випадків модель помилилася (рис. 11). Значення F1-score на валідаційній множині були досить великими,

рівними 92 і 93 % для першого та другого класів, а середнє значення правильності моделі 5 склало 92 %.

| | | | | |
|---------------------|--------------|-----------|--------|----------|
| confusion matrix | | precision | recall | f1-score |
| [[24 1] | | | | |
| [16 9]] | covid | 0.60 | 0.96 | 0.74 |
| acc: 0.6600 | normal | 0.90 | 0.36 | 0.51 |
| sensitivity: 0.9600 | | | | |
| specificity: 0.3600 | accuracy | | | 0.66 |
| | macro avg | 0.75 | 0.66 | 0.63 |
| | weighted avg | 0.75 | 0.66 | 0.63 |

Рис. 9. Матриця помилок і значення метрик якості на валідаційній множині для моделі 4

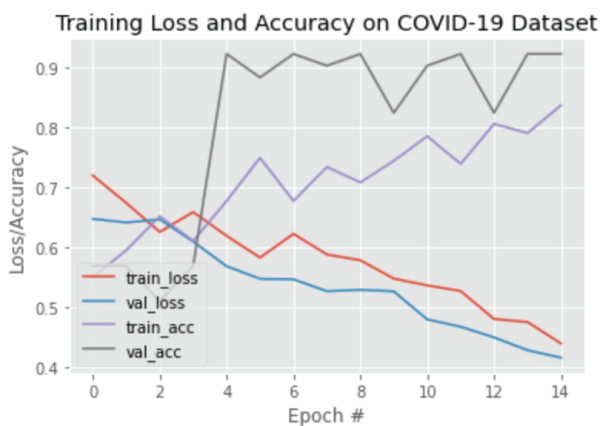


Рис. 10. Результати навчання та валідації для моделі 5

| | | | | |
|---------------------|--------------|-----------|--------|----------|
| confusion matrix | | precision | recall | f1-score |
| [[22 3] | | | | |
| [1 25]] | covid | 0.96 | 0.88 | 0.92 |
| acc: 0.9216 | normal | 0.89 | 0.96 | 0.93 |
| sensitivity: 0.8800 | | | | |
| specificity: 0.9615 | accuracy | | | 0.92 |
| | macro avg | 0.92 | 0.92 | 0.92 |
| | weighted avg | 0.92 | 0.92 | 0.92 |

Рис. 11. Матриця помилок і значення метрик якості на валідаційній множині для моделі 6

Для моделі 6 спостерігався коливальний характер графіків функцій втрат і правильності

за епохами. Тож значення показників якості для моделі 6 були невисокими та наближенішими до значень за моделлю 4, як було наведено на рис. 9. Модель 6 досягла значення ассурагу 93 % на навчальній множині та 89 % – на валідаційній. F-міра (F1-score) набула значення 71 і 62 % для першого та другого класів; матриця помилок була близькою до матриці за моделлю 4 й показала значну кількість хибно класифікованих рентген-зображень здорових людей на тестовій множині.

Результати, отримані як нами у цій роботі, так й іншими дослідниками, зведено до таблиць 1 і 2. Модель 1 на основі VGG-16, побудована в цій роботі, мала найбільше серед інших моделей значення повноти recall/sensitivity виявлення хвороби, що становить 100 %. Модель 5 на основі ResNet50 мала найбільше серед інших моделей значення специфічності, рівне 96,14 %. Моделі, побудовані в цій роботі, отримали вище значення F-міри (94 і 92 %), як порівняти з багатьма іншими моделями, дослідженими в [22]–[24], проте “програли” моделі CovXNet (97,1 %), запропонованій у [23].

Результативні значення метрик якості залежать від складу множин: навчальної та перевіркої/ тестової. Автори [22] зазначають, що вони використовували власні підготовлені рентген-зображення. Мережа CovXNet [23] є ансамблем декількох мереж, навчених на рентген-зображен-

нях різної роздільної здатності.

Таблиця 1. Узагальнені за обома класами значення метрик якості для різних моделей

| Метод | Accuracy, % | Recall/sensitivity, % | Specificity, % | F-міра, % |
|-------------------|-------------|-----------------------|----------------|-----------|
| ResNet18 [22] | – | 98 | 90,7 | – |
| ResNet50 [22] | – | 98 | 89,6 | – |
| SqueezeNet [22] | – | 98 | 92,9 | – |
| DenseNet-121 [22] | – | 98 | 75,1 | – |
| CovXNet [23] | 97,4 | 97,8 | 94,7 | 97,1 |

Закінчення таблиці 1.

| Метод | Accuracy, % | Recall/sensitivity, % | Specificity, % | F-міра, % |
|---|--------------|-----------------------|----------------|-----------|
| Residual [23] | 92,1 | 93,4 | 89,2 | 91,9 |
| Inception [23] | 89,5 | 87,7 | 83,2 | 88,4 |
| VGG-19 [23] | 85,3 | 83,9 | 79,9 | 85,1 |
| Модель 1 у цій роботі на основі VGG-16 | 94,12 | 100 | 88,46 | 94 |
| Модель 2 в цій роботі на основі VGG-16 | 94,12 | 96 | 92,31 | 94 |
| Модель 5 у цій роботі на основі ResNet50 | 92,16 | 88 | 96,14 | 92 |

Таблиця 2. Значення метрик якості окремо за класами

| Метод | Правильність accuracy, % | Клас 1 (хворий на COVID-19) | | | Клас 2 (здоровий) | | |
|---|--------------------------|-------------------------------|------------------------|-----------|-------------------------------|------------------------|-----------|
| | | Повнота recall/sensitivity, % | Точність, precision, % | F-міра, % | Повнота recall/sensitivity, % | Точність, precision, % | F-міра, % |
| VGG-19 [24] | 89,47 | 76 | 94 | 84 | 97 | 88 | 92 |
| Модель 1 у цій роботі на основі VGG-16 | 94,12 | 100 | 89 | 94 | 88 | 100 | 94 |
| Модель 2 в цій роботі на основі VGG-16 | 94,12 | 96 | 92 | 94 | 92 | 96 | 94 |
| Модель 5 у цій роботі на основі ResNet50 | 92,16 | 88 | 96 | 92 | 96 | 89 | 93 |

Висновки

У роботі побудовано кілька моделей глибоких згорткових нейронних мереж на основі відомих архітектур VGG16, ResNet50 і SqueezeNet шляхом додавання шарів для регуляризації цих мереж. Моделі використано для класифікації рентген-зображень грудної клітки людини на два класи: хворий на COVID-19 або здоровий, використовуючи відкриті дані [22], [24]. Моделі показали обнадійливі результати прогнозування хвороби та потребують подальших досліджень.

У розробленому програмному продукті реалізовано функцію прогнозування діагнозу COVID-19 на основі обраної моделі при подачі на вхід індивідуального рентген-зображення. Розпізнавання нового рентген-зображення (якого не було в навчальній вибірці) за допомогою цієї програми триває до однієї хвилини — це швидше, ніж тести RT-PCR (ПЛР зі зворотною транскрипцією), які нині використовують

для підтвердження захворювання. Лікарі могли б діяти впевненіше, маючи ще й цифрову думку, що підтверджує їх оцінку стану пацієнта. Розроблений інструмент також може надавати кількісні оцінки для їх використання у подальших дослідженнях.

Надалі доцільно детальніше оцінити якість побудованих у роботі моделей на навчальних/тестових вибірках більшого розміру в міру появи нових рентген-зображень хворих на COVID-19; поряд із задачею бінарної класифікації розглянути задачу з багатьма класами: хворий на COVID-19, хворий на пневмонію, здоровий, що дасть змогу виявити спільні риси COVID-19 і пневмонії; розглянути інші архітектури згорткових нейронних мереж і дослідити точність прогнозування діагнозу COVID-19 на їх основі; дослідити ансамблі моделей для підвищення точності прогнозування; побудувати інструментарій для онлайн-прогнозування діагнозу COVID-19.

References

- [1] World Health Organization, *Coronavirus disease (COVID-19) outbreak situation*. Available: <https://www.who.int>. (Accessed: 20.09.2021).
- [2] World Health Organization. *Coronavirus disease (COVID-2019) situation reports*. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (Accessed: 20.09.2021).
- [3] *Coronavirus COVID-19 global cases by the Center for systems science and engineering at Johns Hopkins University*. Available: <https://coronavirus.jhu.edu/map.html>. (Accessed: 20.09.2021).
- [4] *Coronavirus in Ukraine*. Available: <https://index.minfin.com.ua/ua/reference/coronavirus/ukraine/> (Accessed: 20.09.2021).
- [5] R. A. Middelburg and F. R. Rosendaal, "COVID-19: How to make between-country comparison", *Int. J. Infect. Diseases*, vol. 96, pp. 477–481, Jul. 2020, doi: 10.1016/j.ijid.2020.05.066.
- [6] M. Yousaf, S. Zahir, M. Riaz, S. M. Hussain, and K. Shah, "Statistical analysis of forecasting COVID-19 for upcoming month in Pakistan", *Chaos Solit. Fractals*, vol. 138, p. 109926, Sep. 2020, doi: 10.1016/j.chaos.2020.109926.
- [7] S. Singh, K. S. Parmar, J. Kumar, and S. J. S. Makkhan, "Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19", *Chaos Solit. Fractals*, vol. 135, p. 109866, Jun. 2020, doi: 10.1016/j.chaos.2020.109866.
- [8] T. Chakraborty and I. Ghosh, "Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis", *Chaos Solit. Fractals*, vol. 135, p. 109850, Jun. 2020, doi: 10.1016/j.chaos.2020.109850.
- [9] D. Fanelli and F. Piazza, "Analysis and forecast of COVID-19 spreading in China, Italy and France", *Chaos Solit. Fractals*, vol. 134, p. 109761, May 2020, doi: 10.1016/j.chaos.2020.109761.
- [10] G. Rainisch, E. A. Undurraga, and G. Chowell, "A dynamic modeling tool for estimating healthcare demand from the COVID19 epidemic and evaluating population-wide interventions", *Int. J. Infect. Diseases*, vol. 96, pp. 376–383, Jul. 2020, doi: 10.1016/j.ijid.2020.05.043.
- [11] K. Sarkar, S. Khajanchi, and J. J. Nieto, "Modeling and forecasting the COVID-19 pandemic in India", *Chaos Solit. Fractals*, vol. 139, p. 110049, Oct. 2020, doi: 10.1016/j.chaos.2020.110049.
- [12] R. Salgotra, M. Gandomi, A. H. Gandomi, "Time series analysis and forecast of the covid-19 pandemic in India using genetic programming", *Chaos Solit. Fractals*, vol. 138, p. 109945, Sep. 2020, doi: 10.1016/j.chaos.2020.109945.
- [13] N. Chintalapudi, G. Battineni, G. G. Sagaro, and F. Amenta, "COVID-19 outbreak reproduction number estimations and forecasting in Marche, Italy", *Int. J. Infect. Diseases*, vol. 96, pp. 327–333, Jul. 2020, doi: 10.1016/j.ijid.2020.05.029.
- [14] E. Aviv-Sharon and A. Aharoni, "Generalized logistic growth modeling of the COVID-19 pandemic in Asia", *Infect. Disease Modelling*, vol. 5, pp. 505–509, 2020, doi: 10.1016/j.idm.2020.07.003.
- [15] Wired 2020 Chinese hospitals deploy AI to help diagnose Covid-19. Available: <https://www.wired.com/story/chinese-hospitals-deploy-ai-help-diagnose-covid-19>. (Accessed: 20.09.2021)
- [16] Y. Fang *et al.*, "Sensitivity of chest CT for COVID-19: Comparison to RT-PCR", *Radiology*, vol. 296, no. 2, pp. 115–117, Aug. 2020, doi: 10.1148/radiol.2020200432.
- [17] T. Ai *et al.*, "Correlation of Chest CT and RT-PCR Testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 Cases", *Radiology*, vol. 296, no. 2, pp. E32–E40, Aug. 2020, doi: 10.1148/radiol.2020200642.
- [18] M.-Y. Ng *et al.*, "Imaging Profile of the COVID-19 Infection: radiologic findings and literature review", *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, Feb. 2020, doi: 10.1148/ryct.2020200034.
- [19] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China", *Lancet*, vol. 395, no.10223, pp. 497–506, Feb. 2020, doi: 10.1016/S0140-6736(20)30183-5.
- [20] M. Cellina, M. Orsi, T. Toluian, C. Valenti Pittino, and G. Oliva, "False negative chest X-Rays in patients affected by COVID-19 pneumonia and corresponding chest CT findings", *Radiography*, vol. 26, no. 3, pp. e189–e194, Aug. 2020, doi: 10.1016/j.radi.2020.04.017.
- [21] A. England, E. Littler, S. Romani, and P. Cosson, "Modifications to mobile chest radiography technique during the COVID-19 pandemic – implications of X-raying through side room windows", *Radiography*, Available online 3 August 2020. vol. 27, no. 1, pp. 193–199, Feb. 2021, doi: 10.1016/j.radi.2020.07.015.
- [22] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning", *Med. Image Anal.*, vol. 65, p. 101794, Oct. 2020, doi: 10.1016/j.media.2020.101794.
- [23] T. Mahmud, M.A. Rahman, S.A. Fattah, "CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization", *Comput. in Biol. and Medicine*, vol. 122, p. 103869, Jul. 2020, doi: 10.1016/j.compbiomed.2020.103869.
- [24] H. Panwar, P. K. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, and V. Singh, "a deep learning and Grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-scan images", *Chaos Solit. Fractals*, vol. 140, p. 110190, Nov. 2020, doi: 10.1016/j.chaos.2020.110190.

- [25] J. P. Cohen and P. Morrison and L. Dao, “COVID-19 image data collection”, arXiv:2003.11597, 2020, Available: <https://github.com/ieee8023/covid-chestxray-dataset>. (Accessed: 20.09.2021)
- [26] J. P. Cohen *et al.*, “COVID-19 image data collection: prospective predictions are the future”, arXiv:2006.11988, 2020. Available: <https://github.com/ieee8023/covid-chestxray-dataset>. (Accessed: 20.09.2021)
- [27] CheXpert: A large chest X-Ray dataset and competition. Available: <https://stanfordmlgroup.github.io/competitions/chexpert/> (Accessed: 20.09.2021)
- [28] NIH. Open access biomedical image search engine. Available: <https://openi.nlm.nih.gov/> (Accessed: 20.09.2021)
- [29] Y. LeCun *et al.*, “Backpropagation applied to handwritten zip code recognition”, *Neural Comput.*, vol. 1, no. 4, pp. 541–551, Dec. 1989, doi:10.1162/neco.1989.1.4.541.
- [30] Neural network architectures. Available: <https://habr.com/ru/company/nix/blog/430524/> (Accessed: 20.09.2021)
- [31] Convolutional network for image feature extraction VGG16. Available: <https://neurohive.io/ru/vidy-nejrosetej/vgg16-model/>. (Accessed: 20.09.2021)
- [32] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, arXiv:1512.03385, 2015.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, *Proceedings of the 2016 IEEE Conf. on Comput. Vision and Pattern Recognition*, pp. 770–778, Jun. 2016, doi:10.1109/CVPR.2016.90.
- [34] ResNet (34,50,101): “Residual” CNN for image classification. Available: <https://neurohive.io/ru/vidy-nejrosetej/resnet-34-50-101/> (Accessed: 20.09.2021)
- [35] A. Veit, M. Wilber, and S. Belongie, “Residual networks behave like ensembles of relatively shallow networks”, *NIPS’16: Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 550–558, 2016.
- [36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, Cambridge, MA, USA, 2017.
- [37] B. Ramsundar and R. B. Zadeh, *TensorFlow for deep learning: from linear regression to reinforcement learning*. O’Reilly Media, Sebastopol, CA, USA, 2018.
- [38] T. Hope, Y. S. Resheff, and I. Lieder, *Learning Tensorflow: A guide to building deep learning systems*. O’Reilly Media, Sebastopol, CA, USA, 2017.
- [39] A. Gulli, A. Kapoor, and S. Pal, *Deep learning with TensorFlow 2 and Keras*, 2nd ed. Packt Publishing, Birmingham – Mumbai, 2019.
- [40] N. Shukla, *Machine learning with TensorFlow*. Manning Publications, NY, USA, 2018.
- [41] A. Geron, *Hands-on machine learning with Scikit-Learn and TensorFlow*. O’Reilly Media, Sebastopol, CA, USA, 2017.
- [42] *TensorFlow documentation*. Available: www.tensorflow.org (Accessed: 20.09.2021).
- [43] *Keras documentation*. Available: <https://keras.io/guides/> (Accessed: 20.09.2021).

Н.И. Недашковская, А.А. Сапельников

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ДИАГНОЗА COVID-19

Проблематика. Ситуация с коронавирусом COVID-19 по состоянию на середину 2021 года снова вызывает беспокойство во всем мире в связи с появлением нового более опасного штамма “Дельта”. Ученые пытаются построить математические модели описания распространения заболеваемости. В Китае, США и других странах разрабатывают инструменты искусственного интеллекта, направленные на прогнозирование диагноза COVID-19, и внедряют их в больницы.

Цель исследования. Сравнить распознавание различными моделями глубоких сверточных нейронных сетей болезни COVID-19 на основе рентген-изображений грудной клетки человека.

Методика реализации. Модели построены на основе архитектур сетей VGG16, ResNet50 и SqueezeNet с добавлением слоев для регуляризации этих сетей. Перед последним выходным слоем сети добавляли слой дропаута. Обучение моделей проводилось с использованием технологии передачи знаний transfer learning. Качество моделей оценивалось на основе матрицы ошибок (confusion matrix), показателей точности (precision), полноты (recall) и F-меры на валидационном множестве рентген-изображений.

Результаты исследования. Модель на основе архитектуры ResNet50 получила достаточно большое значение precision 96 % для класса 1 и общее значение specificity 96,14 %. Полнота (recall) обнаружения заболевания на основе этой модели составила 88 %. Модель на основе архитектуры VGG-16 правильно классифицировала 100 % (эксперимент 1) / 96 % (эксперимент 2) больных COVID-19 из тестовой выборки. Значение precision для класса больных и значение общей specificity на основе этой модели, однако, были меньшими и составили 89 % / 92 % и 88,46 % / 92,31 % соответственно для двух экспериментов. Значения F-меры для моделей ResNet50 и VGG-16 были довольно значительными и равными 92 и 94 %.

Выводы. Модели глубоких сверточных нейронных сетей показали обнадеживающие результаты прогнозирования диагноза COVID-19 и требуют исследований. Прогнозирование диагноза с помощью разработанного программного продукта на основе этих моделей длится до одной минуты – это быстрее, чем тесты RT-PCR (ПЦП с обратной транскрипцией), которые используют для подтверждения заболевания.

Ключевые слова: сверточные нейронные сети; глубокое обучение; классификация изображений; прогнозирование диагноза COVID-19.

N.I. Nedashkovskaya, O.O. Sapelnikov

COMPARATIVE ANALYSIS OF DEEP NEURAL NETWORK MODELS FOR PREDICTION OF COVID-19 DIAGNOSIS

Background. The situation with the coronavirus COVID-19 in mid-2021 is causing concern around the world again due to the emergence of a new more dangerous strain “delta”. Attempts are being made to build mathematical models for describing the disease spread. China, the United States and other countries are developing artificial intelligence tools to help predict the COVID-19 diagnosis and are deploying them in hospitals.

Objective. The purpose of the paper is to compare how different deep convolutional neural network models deal with COVID-19 disease recognition based on chest x-rays.

Methods. The models are based on the VGG16, ResNet50 and SqueezeNet network architectures with the addition of layers to regularize these networks. The models were trained using the transfer learning technology. The quality of the models was assessed based on the confusion matrix, precision, recall, specificity and F-scores on the validation set of X-ray images.

Results. Several neural network models have been built to classify an X-ray image of the human chest into two classes: person with COVID-19 (class 1) or healthy person (class 2). The ResNet50 model achieved a fairly high precision of 96 % for class 1 and an overall specificity of 96.14 %. Recall disease detection based on this model was 88%. The VGG-16 model correctly classified 100% (Experiment 1) / 96 % (Experiment 2) of COVID-19 patients in the test sample. The patient class precision and overall specificity values based on this model, however, were lower at 89 % / 92 % and 88.46 % / 92.31 %, respectively, for the two experiments. The F-score values for the ResNet50 and VGG-16 models were quite significant and equal to 92 and 94 %.

Conclusions. Deep convolutional neural network models have shown promising results in predicting the COVID-19 diagnosis and require research. Predicting the diagnosis using the developed software based on these models takes up to one minute – this is faster than RT-PCR (reverse transcription PCR) tests, which are used to confirm the disease.

Keywords: convolutional neural networks; deep learning; image classification; predicting the COVID-19 diagnosis.

Рекомендована Радою
Інституту прикладного системного аналізу
КПІ ім. Ігоря Сікорського

Надійшла до редакції
2 жовтня 2021 року

Прийнята до публікації
7 грудня 2021 року