

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО»

Кваліфікаційна наукова
праця на правах рукопису

Надеран Марьям

УДК 004.8:616-006

ДИСЕРТАЦІЯ

**ГІБРИДНА ЗГОРТКОВА МЕРЕЖА ДЛЯ ОБРОБКИ
ЗОБРАЖЕНЬ ТА МЕДИЧНОЇ ДІАГНОСТИКИ**

Шифр і назва спеціальності: 122 Комп'ютерні науки

Галузь знань: 12 Інформаційні технології

Дисертація містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело.

_____ М. Надеран

Подається на здобуття наукового ступеня доктора філософії.

Науковий керівник: **Зайченко Юрій Петрович**, доктор технічних
наук, професор

Київ – 2021

АНОТАЦІЯ

Надеран М. Гібридна згорткова мережа для обробки зображень та медичної діагностики. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії зі спеціальності 122 «Комп'ютерні науки». – Національний Технічний Університет України «Київський Політехнічний Інститут Імені Ігоря Сікорського», 2021.

Метою дисертаційного дослідження є розробити нову архітектуру гібридної згорткових мережі і модель класифікації для підвищення якості розпізнавання раку молочної залози і зниження часу на її навчання.

Дослідження обумовлена потребою в розробці нових і вдосконаленні існуючих моделей і методів для обробки зображень і медичної діагностики. Гібридна згорткова мережа повинна забезпечувати виділення інформативних ознак, що дозволяє підвищити критерій якості моделі для задач діагностування рак молочної залози.

На сьогодні прогрес в напрямку підвищення якості та розширення можливостей сучасних моделей, оптимального оцінювання станів, а також діагностування раку молочної залози на основі мамографічних сканів неможливо без застосування сучасних методів і моделей для попередньої обробки і виділення інформативних ознак.

Розроблена гібридна згорткова мережа на основі енкодера, яка дозволяє підвищити якість класифікації раку молочної залози і зокрема досягти мінімального відсотка помилково негативної помилки (англ. False negative, FN), що і є одним з ключових завдань поточного етапу розвитку методів. Для навчання моделі згорткового енкодера можна використовувати зображення, які не мають мітки, що зменшує розмірність зображення і видаляє шуми, в результаті чого залишаються важливі інформативні ознаки в меншій розмірності. Також, згорткова мережа з

меншою кількістю навчальних параметрів, відповідно менш схильна до явища перенавчання. З цих причин у запропонованій моделі було використано згортковий автоенкодер для підвищення критерій якості моделі при меншому часі навчання.

У запропонованій гібридній згортковій моделі, згортковий автокодер було використано для пошуку інформативних ознак, а згорткова нейронна мережа DenseNet – для класифікації. Для навчання запропонованої моделі, необхідно виконати:

- Попередню обробку вхідних даних. Для попередньої обробки медичних зображень було застосовано: нормалізацію, скорочення розмірів зображення і аугментацію даних. В аугментації даних застосовувались такі параметри, як обертання, розмір зображення та горизонтальний зсув.

- Оптимізацію згорткової мережі автокодера шляхом мінімізації функції втрат.

- відрегулювати гіперпараметри за допомогою експериментального аналізу для підвищення продуктивності моделі.

- Шляхом трансферного навчання, навчити згортковій мережі DenseNet.

- Підключити вихід енодера до навченої згорткової мережі DenseNet для завдання класифікації.

- На останньому етапі необхідно оптимізувати гібридну згорткову мережу за допомогою мінімізації критерія Перехресна ентропія:

$$H_p = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(P(y_i)) + (1 - y_i) \cdot \log(1 - P(y_i)),$$

де,

N - кількість спостережень;

y - бінарний індикатор (0 або 1) того, чи є мітка класу правильної класифікації для спостереження;

p - прогнозована ймовірність моделі.

Виконано експериментальні дослідження розробленої моделі розпізнавання раку молочної залози, отримані показники: чутливість, точність (precision), F1-Score і точність (accuracy) моделі при цьому становить 93,5%, 93,2%, 93,3% і 93%, відповідно, що значно більше ніж у відомих згорткових мереж, які були застосовані для цієї задачі.

Наукова новизна дисертаційної роботи полягає:

- Запропоновано модель, яка на відміну від існуючих моделей, дозволяє діагностування раку молочної залози за мінімальний час в порівнянні з відомими методами.
- Розроблена гібридна згорткова мережа на основі енкодера, яка дозволяє підвищити якість класифікації раку молочної залози і зокрема досягти мінімального відсотка помилково негативної помилки (англ. False negative, FN) в порівнянні з відомими роботами в задачах класифікації раку молочної залози. Даний показник є одним з основних критеріїв для діагностики ракових захворювань.
- Проведена модифікація архітектури моделі Inception V3 шляхом розширення числа повнозв'язаних шарів.

Запропоновані в дисертаційній роботі моделі забезпечують розпізнавання раку молочної залози з найкращою ефективністю в порівнянні з відомими методами. Загальний час навчання моделі становить приблизно 13 годин, що є найменшим часом навчання порівняно з іншими згортковими мережами.

Ключові слова: гібридні згорткові мережі, класифікація ракових захворювань молочної залози, згортковий автокодер, класифікація зображень, , алгоритми глибокого навчання.

ABSTRACT

Naderan M. Hybrid convolution network for image processing and medical diagnostics. - Qualifying scientific work on the rights of the manuscript.

Dissertation for the degree of Doctor of Philosophy in the specialty 122 "Computer Science". - National Technical University of Ukraine "Kyiv Polytechnic Institute named after Igor Sikorsky", 2021.

The aim of the dissertation research is to develop a new architecture of a hybrid convolutional network and a classification model to improve the quality of breast cancer recognition and reduce the time for its training.

The study is driven by the need to develop new and improve existing models and methods for image processing and medical diagnostics. The hybrid convolutional network provides the selection of informative features, which allows to increase the productivity of the model for the tasks of diagnosing breast cancer.

Nowadays, progress towards improving the quality and capabilities of modern models, optimal assessment of conditions, as well as the diagnosis of breast cancer based on mammography scans is impossible without the use of modern methods and models for pre-processing input data and selection of informative features.

Development of a hybrid convolutional network based on an encoder, which allows to improve the quality of classification of breast cancer and to achieve a minimum percentage of false negative error (FN), which is one of the key steps of the current stage of development. For training the model of the convolutional encoder, images that have no label can be used. It reduces the dimensionality of the image and removes noise, which leaves important informative attributes in a smaller dimensionality. Also, convolutional autoencoder has less learning parameters, respectively, less sensitive to be

overtraining. For these reasons convolutional autoencoder were used in the proposed model to increase performance model with less training time.

In the proposed hybrid convolutional model, the convolutional autocoder was used as informative features extraction, and the convolutional neural network was used as a classifier. To learn the proposed model

- First is needed to pre-process the input data. As pre-processing medical images, normalization, image reduction and data augmentation were used. In the date of augmentation, parameters like rotation, image resizing, and horizontal shift were applied.

- Optimize the convolutional network of the autoencoder by minimizing the loss function.

- Adjust the hyperparameters by experimental analysis to increase the performance of the model.

- Fine tune the DenseNet convolutional networks through transfer training.

- Connect the output of the encoder to the trained convolutional network DenseNet for the classification task. The last step is necessary to optimize the hybrid convolutional network by minimizing cross entropy:

$$H_p = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(P(y_i)) + (1 - y_i) \cdot \log(1 - P(y_i))$$

Where,

N - number of observations.

Y - is a binary indicator (0 or 1) of whether there is a class label of the correct classification for observation.

P - is the predicted probability of the model.

Experimental studies of the developed model of breast cancer recognition were performed, the following indicators were obtained: sensitivity, precision, F1-Score and accuracy of the model are 93.5%, 93.2%, 93.3% and 93%, respectively.

The scientific novelty of the dissertation is:

— Proposed a model that, unlike existing models, allows diagnosing breast cancer in the shortest time compared to known methods.

— Developed a hybrid convolutional network based on an encoder, which improves the quality of breast cancer classification and in particular to achieve a minimum percentage of false negative error (False negative, FN) compared to known works in the classification of breast cancer. This indicator is one of the main criteria for the diagnosis of cancer.

— Modified the architecture of the Inception V3 model by expanding the number of fully connected layers.

The models proposed in the dissertation provide recognition of breast cancer with the best efficiency in comparison with known methods. The total training time of the model is approximately 13 hours, which is the shortest training time compared to other convolutional networks.

Key words: hybrid convolutional networks, classification of breast cancer, convolutional autocoder, image classification, deep learning algorithms.

ЗМІСТ

ВСТУП	12
РОЗДІЛ 1	17
АНАЛІЗ СУЧАСНИХ МОДЕЛЕЙ ДЛЯ ДІАГНОСТИКИ РАКОВИХ ЗАХВОРЮВАНЬ	17
1.1 Аналіз показників захворюваності та смертності від раку молочної залози	17
1.2 Фактори ризику розвитку раку молочної залози	19
1.3 Основні проблеми, характерні для аналізу медичних зображень	21
1.3.1 Висока роздільна здатність зображень.	22
1.3.2 Обмежений розмір навчальної вибірки.	24
1.3.3 Проблема масштабування.	26
1.3.4 Різні варіації кольору та артефакти.	27
1.4 Огляд існуючих робіт на тему діагностики раку молочної залози	29
1.5 Постановка задачі дисертаційного дослідження	36
1.6 Висновки до розділу	37
РОЗДІЛ 2	38
АНАЛІЗ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ЗАДАЧ ДІАГНОСТИКИ ЗАХВОРЮВАНЬ	38
2.1 Основні положення теорії штучних нейронних мереж	38
2.2 Згорткові нейронні мережі	44
2.2.1 Згорткові шари.	45
2.2.2 Пулінг або шар субдискретизації.	48
2.2.3 Виключення (Dropout).	50

	10
2.2.4 Шар нормалізації.....	52
2.3. Блоки згорткових нейронних мереж (CNN).....	54
2.3.1 Залишкові блоки.....	54
2.3.2 Агрегаційні блоки.....	55
2.3.3 Щільні блоки.....	56
2.3.4 Перехідні шари.....	60
2.4 Автоенкодер.....	60
2.5 Згортковий автоенкодер.....	63
2.6. Методи аналізу згорткових мереж для задач діагностики раку молочної залози.....	65
2.6.1 Матриця помилок.....	65
2.6.2 Криві валідації: точність, втрати та інші показники.....	67
2.7 Алгоритм градієнтного спуску для навчання згорткових мереж.....	71
2.8 Висновки до розділу.....	77
РОЗДІЛ 3.....	78
РОЗРОБЛЕНІ МОДЕЛІ ТА МЕТОДИ РОЗПІЗНАВАННЯ РАКУ НА ОСНОВІ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ.....	78
3.1 Оброблення вхідних даних для згорткових мереж.....	78
3.1.1 Нормалізація (оброблення зображень).....	79
3.1.2 Перетворення колірного простору.....	80
3.1.3. Декорелювальне перетворення за методом аналізу нульових компонентів (ZCA).....	82
3.2. Застосування згорткових мереж Inception-v3, DenseNet до задачі діагностики раку молочної залози.....	85

3.3 Гібридний сверточних мережу і її навчання для завдання діагностики раку молочної залози.....	88
3.4 Нормалізація плями (англ. Stain normalization, SN).....	89
3.5 Методи оптимізації згортальних нейронних мереж	92
3.6 Критерії для аналізу мережі для класифікації типів ракових захворювань	96
3.7 Висновки до розділу.....	98
РОЗДІЛ 4.....	99
ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ РОЗРОБЛЕНИХ МЕТОДІВ РОЗПІЗНАВАЬННЯ РАКОВИХ ЗАХВОРЮВАНЬ	99
4.1 Експериментальні дослідження згорткових мереж Inception- v3, DenseNet.....	99
4.2 Застосування згорткової мережі DenseNet до задачі розпізнання раку молочної залози	104
4.3 Застосування згорткового автоенкодера для задачі діагностики раку молочної залози	107
4.4 Експериментальні дослідження згорткових автоенкодерів.....	113
4.5 Експериментальні дослідження іншого машинного навчання для розпізнавання раку	119
4.6 Порівняльний аналіз методів діагностики раку молочної залози	125
4.7 Висновки до розділу.....	126
ВИСНОВКИ ДО ДИСЕРТАЦІЇ.....	127
СПИСОК ДЖЕРЕЛ	129

ВСТУП

Актуальність теми. Рак молочної залози – дуже поширений вид раку у жінок віком від 35 до 55 років і другий за поширеністю у світі (близько 1 152 161 нових випадків за рік). Діагностика раку молочної залози часто розглядається як проблема класифікації у нейронних мережах. Виявлення та діагностика раку молочної залози на ранніх стадіях має вирішальне значення для порятунку життя.

Його виявлення на ранніх стадіях може посприяти запобіганню поширення раку на інші органи, що дасть змогу лікарям допомогти пацієнтці до того, як стане запізно. Раннє виявлення потребує систематичних і надійних методів, що дадуть змогу медичним працівникам точно розрізнити доброякісні і злоякісні пухлини. З огляду на це точне виявлення і класифікація раку молочної залози надзвичайно важливі для охорони здоров'я суспільства і для збереження життя онкохворих.

Оскільки завдання розпізнавання та діагностики раку молочної залози на основі медичних зображень є завданням класифікації та розпізнавання, його доцільно вирішувати, ґрунтуючись на методах інтелектуального аналізу даних, зокрема за допомогою нейронних мереж. Серед вітчизняних вчених, які зробили значний внесок у розвиток методів штучного інтелекту і теорії нейронних мереж, можна відзначити таких: Івахненка О.Г., Шлезінгера М.І., Зайченка Ю.П., Резніка О.М., Руденка О.Г.; на проблематику розроблення та реалізації систем діагностики раку молочної залози зверталася увага у працях зарубіжних учених: Бенедиктсона Дж.А (Benediktsson J.A.), Фієра І.С. (Fear E.C.), Келі К.М. (Kelly K.M.), Клема М. (Klemm M.), Крадока І. (Craddock I.), Пога Б.Ві.

(Pogue V.W.), Дехгані Х. (Dehghani H.), Карпентера К.М. (Carpenter C.M.) та ін.

Віднедавна активного використання знає апарат штучних нейронних мереж з метою вирішення різних завдань класифікації. Перевага штучних нейронних мереж – здатність до навчання й узагальнення знань. Інша важлива властивість нейронних мереж – це паралелізм роботи, завдяки якому можна досягнути збільшення швидкості оброблення даних. Оскільки мамографічні знімки – це вхідні дані, а згорткові нейронні мережі є найбільш надійними мережами для оброблення зображень, тому у цій праці проаналізовано різні згорткові мережі для діагностування ракових захворювань.

Технічних проблем в задачах діагностики раку, являються виділенні інформативні ознаки і швидкість навчання моделі. Дана дисертаційна робота було концентровано на рішення цих проблем.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертаційну роботу здобувачем виконано на кафедрі математичних методів системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» відповідно до ініціативної науково-дослідної роботи: “ Розроблення та дослідження методів обробки, розпізнавання, захисту та зберігання медичних зображень в розподілених комп'ютерних системах” (ДР 0117U004267), що виконувалась в 2017-2019рр.. А також НДР (тема № 2304) «Математичні та програмні методи оброблення мультимодальних даних моніторингу медико-біологічних об'єктів для діагностики стану здоров'я пацієнтів», яка виконується в КПІ в 2020-2022рр.

Мета і завдання наукового дослідження. Мета дисертаційної роботи – розроблення нових методів діагностики раку молочної залози з метою покращення ефективності і пришвидшення процесу діагностики.

Поставлена мета визначила необхідність вирішення таких завдань:

— Проаналізувати існуючі роботи на тему розпізнавання раку молочної залози.

— Розробити модель попереднього оброблення медичних зображень таких, як перетворення колірного простору, стандартизація, скорочення розмірів, і т. д.

— Розробити модифікації структури згорткових мереж для вирішення поставленого завдання.

— Розробити модель класифікації медичних зображень молочної залози на основі запропонованої структури згорткового енкодера.

— Виконати експериментальне дослідження отриманих моделей та провести порівняльний аналіз якості розпізнавання раку молочної залози з відомими моделями та методами.

Об'єктом дослідження є процеси діагностики ракових захворювань.

Предметом дослідження є технології та методи глибокого навчання, зокрема, згорткові мережі в задачах діагностики раку молочної залози.

Методи дослідження. У праці використано теорію штучних нейронних мереж, сучасні методи класифікації зображень.

У ході дослідження застосовувались фреймворки машинного навчання Keras і tensorflow, мова програмування python.

Наукова новизна роботи. Наукова новизна дисертаційної роботи полягає:

- Запропоновано модель, який, на відміну від існуючих моделей, дозволяє діагностування раку молочної залози за мінімальний час в порівнянні з відомими методами.
- Розроблено гібридна сверточное мережу на основі енкодера, яка дозволяє підвищити якість класифікації раку молочної залози і зокрема досягти мінімального відсотка помилково негативної помилки (англ. False negative, FN) в порівнянні з відомими роботами в задачах класифікації раку молочної залози, і даний фактор є одним з основних критеріїв для діагностики ракових захворювань.
- Проведена модифікації архітектура моделі Inception V3 шляхом розширення числа пов'язаних шарів.

Практичне значення отриманих результатів. Запропоновані у дисертаційній роботі методи реалізовано як інформаційну технологію, що збільшує якість діагностики раку молочної залози, а саме підвищує точність класифікації раку молочної залози, та зменшує час класифікації в порівнянні з відомими методами та моделями на основі згорткових мереж різних класів.

Розроблено систему комп'ютерної діагностики раку молочної залози на основі обробки та аналізу медичних зображень молочної залози.

Особистий внесок здобувача. Усі наукові положення і результати, що охоплюють основний зміст роботи, здобуті автором самостійно. У працях, написаних у співавторстві, здобувачеві належать: порівняльний аналіз згорткової мережі для задач класифікації раку молочної залози [1], оцінювання ефективності застосування згорткової мережі автоенкодера для поставленого завдання [2], порівняльний аналіз алгоритму глибокого навчання для діагностики раку [3], удосконалення методу класифікації згорткового автоенкодера [164].

Апробація результатів дисертації. Основні положення роботи доповідалися і обговорювалися на наукових конференціях:

— 20-й Міжнародній науково-технічній конференції «SAIT-2018», м. Київ, 2018 р.

— 13-й Міжнародній конференції «International Symposium on Applied Informatics and Related Areas», м. Спекесфегервар, Угорщина, 2018 р.

— 5-й Міжнародній науково-технічній конференції «Обчислювальний інтелект 2019», м. Ужгород, 2019 р.

— 2-й Міжнародній конференції «IEEE International Conference on System Analysis & Intelligent Computing (SAIC)», 05–09 жовтня 2020 р.

— 22-й Міжнародній конференції «Joint International Scientific Events on Informatics «Information Theories and Applications»», 2020 р.

Публікації результатів наукових досліджень. Результати теоретичних і експериментальних досліджень відображено у чотирма опублікованих працях: із них 3 статті у виданні, яке входить до переліку видань, затверджених ВАК України, одна стаття опублікована в міжнародних виданнях.

Структура та обсяг роботи. Дисертація складається із анотації, вступу, переліку умовних скорочень, чотирьох розділів і списку використаних джерел. Обсяг роботи — 149 сторінки, містить 35 рисунків і 19 таблиць.

РОЗДІЛ 1

АНАЛІЗ СУЧАСНИХ МОДЕЛЕЙ ДЛЯ ДІАГНОСТИКИ РАКОВИХ ЗАХВОРЮВАНЬ

1.1 Аналіз показників захворюваності та смертності від раку молочної залози

Рак молочної залози є найбільш поширеним раком у жінок як у розвинених, так і в менш розвинених країнах. Згідно із статистикою 2019 р. захворюваність на рак молочної залози серед усіх видів раку становить 11,6%, що спричиняє 6,5% смертельних випадків у всьому світі [42]. Також за даними «ГЛОБОКАН» за 2018 р. захворюваність на рак молочної залози збільшиться з двох мільйонів пацієнтів у 2018 р. до більш ніж трьох мільйонів у 2046 р., показавши зростання на 46% [42], [43]. Через його невиліковний характер і тривалий процес впливу рак молочної залози є тягарем як для пацієнтів, так і для осіб, які доглядають за хворими [44]–[46]. Крім цього, наслідок раку молочної залози – великі витрати для системи охорони здоров'я. Дані свідчать про те, що тільки в США на лікування раку молочної залози щороку витрачається понад 16 млн доларів [46]. Доведено, що депресія – один з найбільш поширених психіатричних симптомів у пацієнтів із раком молочної залози. Однак якщо рак молочної залози діагностується і лікується вчасно, якість життя пацієнтів і їхній прогноз захворювання можуть поліпшитись [47].

У 2017 р. виконано дослідження, яке показало, що смертність від раку становила 9,56 млн випадків, серед яких близько 611 625 жінок померли від раку молочної залози [4]. На жаль, цей відсоток щороку збільшується. Наприклад, за оцінками [11] в усьому світі понад 508 000 жінок у 2011 р.

померли від раку молочної залози. У 2012 р. рак молочної залози став причиною 522 000 смертей у всьому світі разом із 1,68 млн нових випадків захворювання [37]. Вважається, що рак молочної залози є захворюванням розвиненого світу, однак майже 50% випадків раку молочної залози і 58% випадків смерті трапляються в менш розвинених країнах [12].

Показники захворюваності в усьому світі істотно відрізняються: від 19,3% на 100 000 жінок у Східній Африці до 89,7% на 100 000 жінок у Західній Європі. У більшості регіонів, що розвиваються, показники захворюваності менші за 40% на 100 000 жінок [12]. Найменші показники захворюваності спостерігаються у більшості африканських країн, але і в них рівень захворюваності на рак молочної залози поступово збільшується.

Показники виживання від раку молочної залози у всьому світі дуже відрізняються: від 80% і більше у Північній Америці, Швеції та Японії до близько 60% у країнах із середнім рівнем доходу, а також менше ніж 40% у країнах із низьким рівнем доходу [13]. Малі показники виживання у менш розвинених країнах пояснюються здебільшого відсутністю програм раннього виявлення, що призводить до збільшення частки жінок із пізньою стадією захворювання, а також відсутністю належних засобів діагностики і лікування [4]. Кількість померлих від раку молочної залози в 2017 р. показано на рис. 1.1.

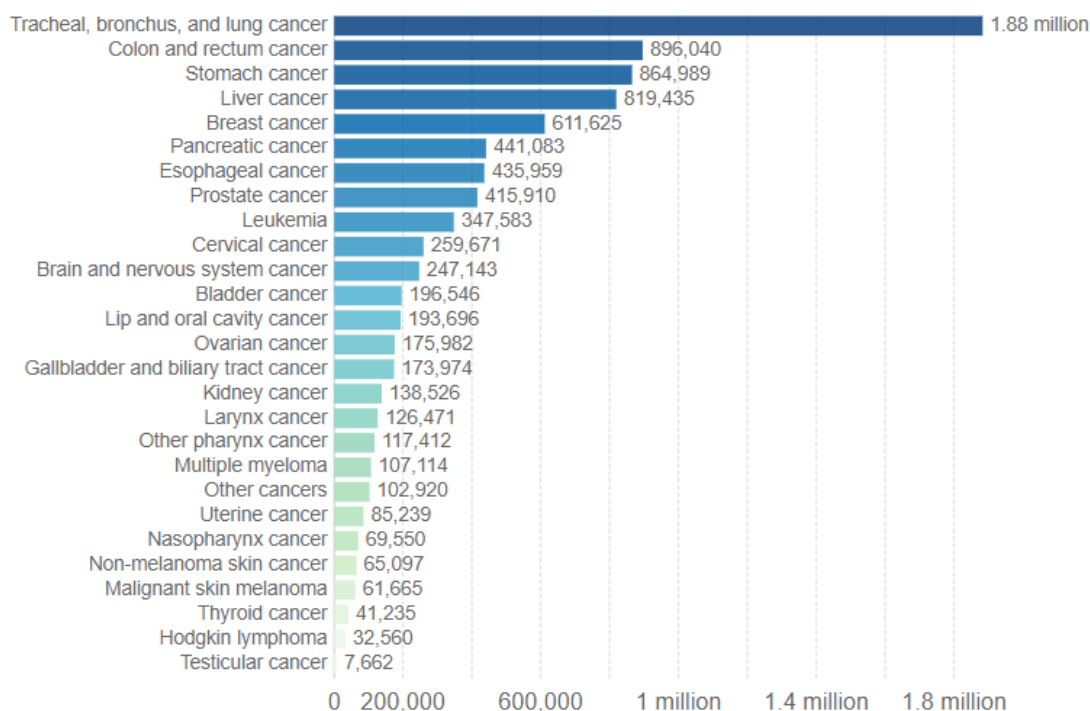


Рис. 1.1. Показник померлих у 2017 р. від різних ракових захворювань

1.2 Фактори ризику розвитку раку молочної залози

Декілька факторів ризику, які спричиняють рак молочної залози, вже задокументовано. Однак для більшості жінок із раком молочної залози неможливо визначити конкретні чинники ризику, які спричинили хворобу [5], [6].

Випадки у сімейному анамнезі раку молочної залози збільшує ризик появи захворювання вдвічі або втричі. Деякі мутації, особливо в BRCA1, BRCA2 і p53, збільшують ризик розвитку раку молочної залози. Однак ці мутації рідкісні і становлять невелику частину від загальної кількості випадків раку молочної залози.

Репродуктивні чинники, пов'язані з тривалим впливом ендогенних естрогенів, такі як рання менархе, пізня менопауза, пізній вік перших

пологів, є одними з найбільш істотних факторів ризику розвитку раку молочної залози. Екзогенні гормони також збільшують ризик розвитку раку молочної залози. Користувачі пероральної контрацепції та замісної гормональної терапії більше схильні до ризику, ніж некористувачі. Грудне вигодовування має захисну дію [5], [6].

Внесок різних модифікованих факторів ризику, крім репродуктивних факторів, у загальну ситуацію з раком молочної залози розраховано Danaei G. et al. [7]. Автори зробили висновок, що 21% від усіх випадків смерті від раку молочної залози в світі пов'язані з вживанням алкоголю, надмірною вагою й ожирінням, а також відсутністю фізичної активності (ця частка більша в країнах із високим рівнем доходу – 27%), але найважливішим фактором була надлишкова вага й ожиріння. У країнах із низьким і середнім рівнем доходу частка раку молочної залози, пов'язана із зазначеними факторами ризику, становила 18%, відсутність фізичної активності була найважливішим визначальним фактором – 10%.

Відмінності у захворюваності на рак молочної залози між розвиненими країнами і країнами, що розвиваються, можуть частково пояснюватись дієтними ефектами в поєднанні з більш пізніми першими пологами, коротшим терміном грудного вигодовування [8]. Зростання поширення західного способу життя в країнах із низьким і середнім рівнем доходу є важливим фактором, який і спричиняє збільшення кількості випадків захворювання на рак молочної залози в цих країнах.

За даними Всесвітньої організації охорони здоров'я (англ. World Health Organization, WHO) [51] мамографічні скани ефективні для аналізу наявності раку молочної залози у пацієнтів. Хоча мамографія високовартісна процедура, але є єдиним методом для скринінгу раку грудей, який підтвердив свою ефективність. У праці [52] автори вказали,

що цифрова мамографія є кращим медичним зображенням і наполегливо рекомендується для автоматизованої діагностики.

1.3 Основні проблеми, характерні для аналізу медичних зображень

Необхідно зрозуміти унікальні характеристики аналізу патологічних зображень і обчислювальні методи їхнього лікування [50]. Табл. 1.1 демонструє огляд робіт, присвячених проблемам і рішенням аналізу патологічних зображень.

Таблиця 1.1

Огляд робіт, присвячених проблемам і рішенням для гістопатологічного аналізу зображень

Проблема	Рішення	Робота
Велика розмірність зображень	Класифікація на рівні кейса, що підсумовує класифікацію на рівні патча або на рівні об'єкта	Марковська мережа [53], Bag of Words of local structure [54], Випадковий ліс [55]–[57],
Недостатня кількість маркованих зображень	Навчання з частковим залученням учителя;	Множинне навчання (Manifold learning) [58], SVM [59], Feature

	трансферне навчання (Transfer learning)	extraction [60], точне налаштування (fine-tuning) [61]–[63]
Різні рівні збільшення призводять до різних рівнів інформації	Багатомасштабний аналіз	CNN [64], dictionary learning [65], texture features [66]
Варіації кольору та артефакти	Видалення ефекту зміни кольору	Нормалізація кольору [67]–[70], посилення кольору (color augmentation) [71],[72]

1.3.1 Висока роздільна здатність зображень.

Якщо такі зображення, як, наприклад, собаки чи будинки, класифікуються за допомогою глибокого навчання, як вхідні дані часто використовують зображення невеликого розміру, такі як 256×256 пікселів. Зображення великого розміру часто потребують зміни розміру на менший, оскільки збільшення розміру вхідного зображення призводить до збільшення оцінюваного параметра, необхідної обчислювальної потужності і пам'яті. Навпаки, повне слайдове зображення (англ. whole slide image, WSI) містить багато пікселів, (може складатися з десятків

мільярдів) і, як правило, важко піддається аналізу. Однак зміна розміру всього зображення на менший розмір, наприклад, 256×256 пікселів, призведе до втрати інформації на клітинному рівні, у результаті чого помітно знизиться точність ідентифікації. Тому все повнослайдове зображення зазвичай ділиться на ділянки розміром близько 256×256 пікселів («патчі»), і кожен патч аналізується незалежно, як, наприклад, виявлення регіону інтересу (англ. Region of interest, ROI). Завдяки досягненням у галузі обчислювальної потужності і пам'яті розмір патча збільшується (наприклад, 960×960 пікселів), що, як очікується, сприятиме підвищенню точності.

Є можливість удосконалити метод інтеграції результату кожного патча. Наприклад, оскільки усе повнослайдове зображення може містити сотні тисяч патчів, хибно позитивні результати з високою ймовірністю можуть з'явитися, навіть якщо окремі патчі класифіковані точно. Одним із можливих рішень для цього є усереднення по регіонах інтересу, а регіони класифікуються як регіони інтересу тільки тоді, коли вони поширюється на кілька патчів. Однак за такого підходу можливі хибно негативні результати, що зумовлено відсутністю малого ROI, наприклад, в ізольованих пухлинних клітинах [56].

В імуногістохімічному дослідженні для інсценування метастазів у лімфатичних вузлах пацієнтів і встановлення діагнозу раку простати за шкалою Глісон із декількома регіонами в межах одного слайда потрібні складніші алгоритми для інтеграції рішень на рівні патчів або на рівні об'єктів [53]–[56],[73],[74]. Наприклад, для патологічної класифікації метастатичного раку молочної залози кілька команд-учасниць і [50] застосовували випадковий ряд класифікаторів імовірностей на рівні пікселів або патчів, які оцінюються шляхом глибокого навчання з

використанням різних характеристик, зокрема оцінювального розміру пухлини [56].

1.3.2 Обмежений розмір навчальної вибірки.

Імовірно найбільша проблема у ході аналізу патологічних зображень з використанням машинного навчання полягає в тому, що є лише невелика кількість навчальних даних. Перевагою глибокого навчання в загальному завданні розпізнавання образів є те, що навчальних даних досить багато. Хоча для більшості задач цифрової патології, зокрема діагностики, потрібне маркування інформації на рівні патчів або пікселів (наприклад, внутрішня/зовнішня межа ракових ділянок), більшість міток у повнослайдовому здебільшого трапляються на рівні конкретних випадків. Інформацію про маркування в загальному аналізі зображень можна легко отримати з інтернету, а також використати краудсорсингове маркування, оскільки будь-яка людина може ідентифікувати об'єкти і виконувати маркування. Проте тільки патології можуть точно маркуватися на патологічному зображенні, а маркування на регіональному рівні великого повнослайдового зображення передбачає істотніший обсяг роботи.

Можливе повторне використання загальнодоступних і готових до аналізу даних як навчальних у машинному навчанні, в ImageNet [75], у природних зображеннях, реалізації міжнародного співробітництва у галузі макроскопічної діагностики шкіри, зокрема її візуалізації [76]. У цифровій патології існують деякі загальнодоступні набори даних, які містять гістопатологічні зображення, отримані вручну, такі як BreakHis [77],[78], MITOS-ATYPIA-14 [79] та інші.

Багато досліджень спрямовано на вирішення цієї проблеми. Більшість із них належать до однієї з таких категорій:

- Ефективне збільшення даних за мітками.
- Використання слабких міток або немаркованої інформації;

— Використання моделей/параметрів для інших завдань.

Ефективне збільшення даних за мітками. Один із способів збільшення системи адаптації – скорочення робочого часу патологоанатомів за вказівкою ROI у WSI. Прості у використанні інструменти графічного інтерфейсу користувача допомагають патологоанатомам ефективно маркувати більшу кількість зразків за коротші проміжки часу [80], [81].

Інший підхід, який використовує метод машинного навчання, – активне навчання [81]–[86]. Активне навчання – це метод, який застосовують у навчанні під наглядом, він автоматично вибирає найбільш цінну немарковану вибірку (тобто ту, яка, як очікується, поліпшить ідентифікаційні характеристики класифікаторів за правильного маркування і використання як навчальних даних) і відображає її для маркування з допомогою патологоанатомів.

Використання слабких міток або немаркованої інформації. Навіть якщо точне розміщення ROI у WSI невідоме, можливо, що інформація про наявність/відсутність ROI у WSI доступна з патологічного діагнозу, визначеного на WSI або мітках WSI. Ці так звані слабкі мітки легше отримати порівняно з маркуванням на рівні патчів. WSI вважається «пакетом», виготовленим із великої кількості патчів у налаштуванні машинного навчання. За діагностики раку WSI маркується як рак, якщо хоча б один патч містить ракову тканину, або як не рак, якщо жоден із патчів не містить ракову тканину. Ця ситуація є проблемою багаторазового навчання [88], [89] або навчання без учителя (англ. weakly-supervised learning) [85], [90]. У типовій багаторазовій задачі навчання позитивні пакети містять принаймні один позитивний екземпляр, а негативні пакети не містять жодних. Мета багаторазового навчання на прикладах –

маркування примірника на основі навчальної вибірки. Для гістопатологічного аналізу зображень застосовано різні методи навчання, зокрема, метод бустінгу [85], векторно-машинний метод підтримання [91] і метод глибокого навчання [90].

На відміну від цього у напівавтоматичному навчанні (англ. semi-supervised learning, SSL) [58],[59],[92],[93] використовуються як марковані, так і немарковані дані. Немарковані дані використовуються для оцінювання дійсного розподілу помаркованих даних.

Повторне використання параметрів з інших задач. Проведення навчання з учителем із використанням невеликої кількості даних для навчання призведе до недостатньої узагальнювальної ефективності. Це особливо помітно за глибокого навчання, де кількість досліджуваних параметрів дуже велика. У такому випадку замість навчання всієї моделі з нуля навчання часто починається з використання (частини) параметрів попередньо підготовленої моделі, оптимізованої в іншій подібній задачі. Такий метод навчання називається трансферним навчанням (англ. transfer learning, TL). Останні шари у згорткових нейронних мережах (повнозв'язні шари) також вважаються екстракторами інформативних ознак. Повнозв'язні шари часто замінюються новою мережею, що підходить для цільової задачі. Параметри в ранніх шарах (згорткових) можуть використовуватись як у праці [60], і як початкові параметри, після чого мережа частково або повністю навчається на основі навчальних даних цільової задачі [61]–[63] (так зване точне налаштування (англ. fine-tuning)).

1.3.3 Проблема масштабування.

Тканини, як правило, складаються з клітин і мають виражені клітинні особливості. Інформація про форму клітин добре захоплюється в надпотужних мікроскопічних зображеннях поля, але структурна

інформація, така як залозиста структура, що складається з багатьох клітин, краще захоплюється в низькопотужних полях (рис. 1.2). Оскільки ракові тканини мають як клітинну, так і структурну атипію, знімки роблять у багаторазовому збільшенні, кожен із них міститиме важливу інформацію. Патологоанатоми діагностують захворювання шляхом отримання різної інформації від клітинного до тканинного рівня за допомогою зміни збільшення мікроскопа. Навіть у машинному навчанні існують дослідження з використанням зображень за різних збільшень [64]–[66].

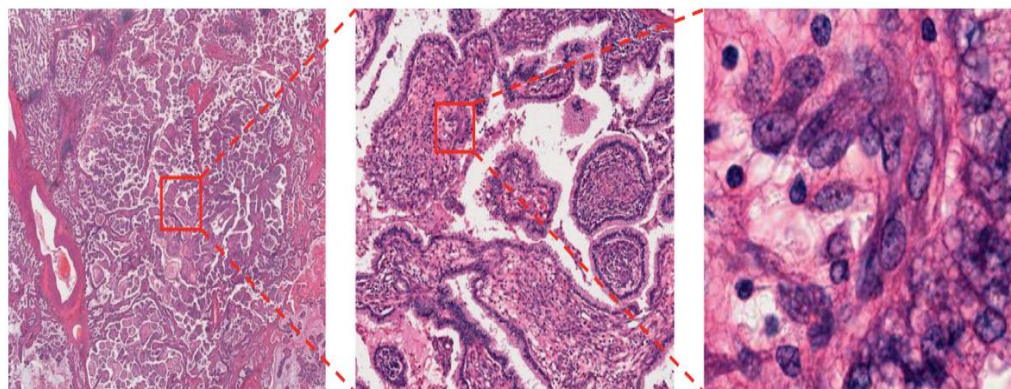


Рис. 1.2. Багаторазове збільшення того самого гістопатологічного зображення. На знімку справа показана збільшена ділянка, позначена червоним квадратом зліва

1.3.4 Різні варіації кольору та артефакти.

Ще одним серйозним артефактом є зміна кольору, як зображено на рис. 1.3. Джерелами варіації є різні партії або виробники фарбувальних реагентів, товщина перетинів тканин, умови фарбування і моделі сканерів. Навчання без урахування колірних варіацій може погіршити роботу алгоритму машинного навчання. Якщо включити достатню кількість даних, отриманих кожним сканером, за кожною пофарбованою тканиною,

то вплив колірних змін на точність класифікації може стати незначним, проте наразі це здається малоімовірним.

Для вирішення цієї проблеми запропоновано різні методи, включно з переходом на відтінки сірого, нормалізацією кольору [67]–[70] і збільшенням кольору [71], [72]. Перетворення у відтінки сірого – найпростіший спосіб, але тоді ігнорується важлива інформація про колір подання, що регулярно використовується патологами. Натомість нормалізація кольору намагається налаштувати значення кольору зображення за пікселями таким чином, щоб колірний розподіл вихідного зображення збігався з кольоровим розподілом еталонного зображення. Але оскільки компоненти і співвідношення складу клітин або тканин в цільових і еталонних зображеннях у цілому відрізняються, часто потрібне попереднє оброблення, таке як ядерне виявлення з використанням спеціального алгоритму коригування компонента. З огляду на це нормалізація кольору видається доречною, коли аналізовані у задачах WSI містять принаймні частково подібні склади клітин або тканин.

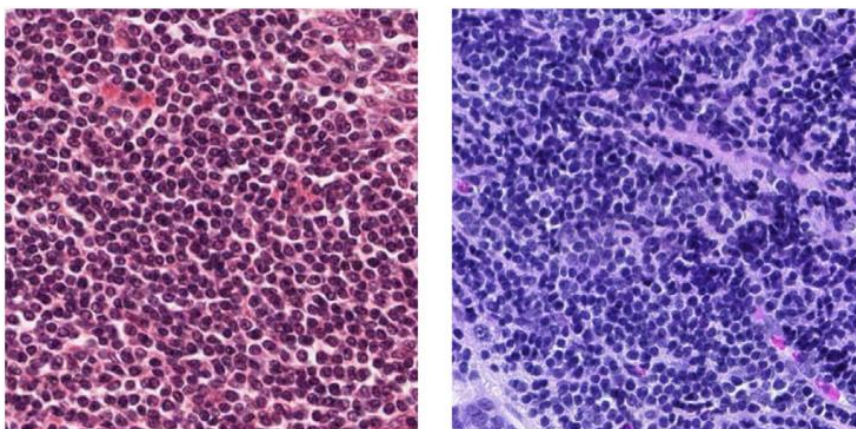


Рис. 1.3. Колірна варіація гістопатологічних зображень

1.4 Огляд існуючих робіт на тему діагностики раку молочної залози

Систематичний огляд виконано у праці [21]. У ній автори проілюстрували важливість вибору правильних медичних образів. Цифрова мамографія є найрезультативнішим способом медичного скринінгу. Вона допомагає виявити пухлину до її подальшого розвитку. Крім цього, автори порівнюють різні методи, які використовувалися у 2011–2017 рр., і доходять висновку, що метод опорних векторів (англ. support vector machines, SVM) дає найкращий результат, ґрунтуючись на [22]–[25].

Натепер мало праць, у яких розглядається можливість використання згорткових нейронних мереж для розпізнавання раку молочної залози. У [26] порівняно методи машинного навчання, такі як метод опорних векторів (SVM), дерева прийняття рішень (C4.5), наївний баєсів класифікатор (англ. Naive Bayes, NB) і метод k-найближчих сусідів (англ. k-Nearest Neighbor, k-NN) за лейкоцитами. На підставі результатів виконаних експериментів SVM має кращу точність – 97,13%. Однак для діагностики раку точність і чутливість варто розглядати як оцінювальні показники.

У праці [27] запропоновано модифіковану архітектуру згорткової нейронної мережі (англ. convolutional neural network, CNN) із дев'ятьма шарами. Із цих шарів шість були згорткові і пулінгові, а три – повнозв'язні. Результат дає точність 69,99% і 81,44% у класифікації раку і виявленні некрозу відповідно.

Автори [28] розглядають такі фактори, як справді позитивний, хибно позитивний, точність і чутливість для класифікації з використанням гібридного предиктора рецидиву раку молочної залози

(англ. hybrid predictor of breast cancer, HPBCR). Чутливість моделі становила 77%, точність – 95%, ті ж значення для SVM і дерева рішень – 67%, 78% і 75%, 77% відповідно.

Автори праці [29] пропонують end-to-end підхід, за якого модель для класифікації локальних знімків проходить попередню підготовку з використанням повністю анотованого набору даних із ROI. Результати експерименту показують, що ResNet 50 має кращу точність (97%) порівнянно з VGG (84%). End-to-end підхід є одним із видів глибокого процесу навчання, коли всі параметри навчаються спільно, а не покроково. Недолік полягає в тому, що автори [29] використовують просту валідацію. Проста валідація означає, що набір даних для навчання розділений на набори даних для навчання і для валідації. Наприклад, 70% даних використовуються для навчання і 30% – для тестування. Замість цього К-фолд валідація (англ. K-fold cross validation) ділить дані на k частини (folds), і кожна частина у якийсь момент використовується як тестова вибірка. Використання крос-валідації допомагає запобігти перенавчанню без втрати даних [30],[31].

У праці [32] для вилучення інформаційних ознак використано метод головних компонент (англ. principal component analysis, PCA). Оскільки за використанням PCA багато ознак, які можуть бути інформативними, не розглядалися у ході дослідження, автори запропонували для аугментації ознак додати рандомно такі ознаки, які вилучено зі згорткової мережі. Як об'єкт дослідження використано дані експресії генів.

У праці [33] запропоновано метод глибокого навчання для виявлення мітозів молочної залози. У запропонованій моделі у згорткову мережу додано шар «краудсорсинг»; значення F1-score становило 61,33%.

Запропонований у [34] метод глибокого навчання для діагностики раку молочної залози дав точність 87,3%. Також виконано порівняльний аналіз із SVM, де точність на 1,5% менша, ніж у запропонованому методі. Для навчальної вибірки використано 1000 мамографічних сканів, для тестової вибірки – 204 скани.

У праці [35] застосовано глибокі нейронні мережі з точністю 85% для задачі діагностики раку молочної залози. Також здійснено порівняння зі звичайною нейронною мережею (англ. artificial neural network, ANN), де точність становила 79%. Автори виконали порівняльний аналіз методів, які застосовувались для задачі діагностики раку молочної залози. Табл. 1.2 показує точність кожного методу.

Таблиця 1.2

Порівняльний аналіз існуючих методів для діагностики раку
молочної залози

Рік	Моделі	Кількість навчальної вибірки	AUC
2015	GoogLeNet	2250	0,85
2015	AlexNet	2250	0,84
2015	VGG-Net 16	2250	0,82
2016	ViDi Red	286	0,81
2016	4Convolutional ANN and 1 fully connected layer	840	0,70

2016	Convolutional Sparse Autoencoder	2244	0,61
------	--	------	------

У праці [36] запропоновано модифіковану згорткову мережу, у якій як навчальну вибірку використано набір даних BreakHist (7909 мамографічних сканів). Виконано операцію аугментації навчальної вибірки, оскільки набір даних є незбалансованим. Також автори проводили експериментальний аналіз гіперпараметра швидкості навчання (англ. learning rate, LR), оптимальне значення якого становило 0,001. Також автори показали, що за навчальної ітерації, яка дорівнює 5000, запропонована модель показує точність для навчальної іа валідаційної вибірок.

Є багато досліджень, у яких розглядається виявлення раку молочної залози за допомогою CNN. Однак більшість із них у своїх експериментах покладаються на точність, але точність у виявленні раку молочної залози не є єдиним достовірним фактором, який треба враховувати [9]. У цих задачах варто враховувати чутливість моделі, щоб зрозуміти, скільки разів модель класифікувала рак неправильно.

У праці [10] запропоновано сучасну згорткову нейронну мережу (DenseNet) для виявлення раку молочної залози із зображень гістології раку молочної залози (BACH) із точністю до 85,6%. Коефіцієнт помилкової класифікації класу раку становив у середньому 14,4%. У зазначеній праці чутливість (у середньому за чотирма класам) ResNet 50 порівняно із запропонованою ними CNN – 76% і 79% відповідно.

Порівняно з раніше існуючими моделями CNN (VGG-16, VGG19, Xception, Resnet, Inception) із 80% точністю у багатокласовій класифікації автори праці [14] запропонували модель, де точність

становила 83,97% у середньому для двох класів (Benign і Malignant). Запропонована модель є комбінацію Зачаття і Мережі з використанням набору даних BreakHist, який містить 7909 мамографічних сканів із чотирма збільшеннями (зуми 40X, 100X, 200X і 400X).

У праці [15] стверджується, що завдяки архітектурі DenseNet, у якій всі шари повністю з'єднані з кожним попереднім шаром і короткою зв'язкою між цими шарами поблизу входу і виходу, модель може бути навчена більш ефективно і точно.

У праці [16] запропоновано мережу DenseNet із часом навчання 11 годин. Автори стверджують, що вперше використовували ваги з Imagenet і допрацювали модель для тренування DenseNet. Усі згорткові шари сету були заморожені і тільки повнозв'язний шар був навчений. У праці [17] автори використовували atrous DenseNet, яка досягає багатомасштабного виділення ознак шляхом інтеграції атронних конволюцій у щільний блок. Автори порівнювали дві бази даних, VASIS і CCG, у яких середня точність для двох класів (рак/не рак) запропонованої моделі становила 82,50% і 87,05% відповідно для кожної бази даних.

Нову модель згорткової нейронної мережі запропоновано у праці [18], де автори використовували 400 зображень із 40-кратним збільшенням для навчальної вибірки і 200 зображень для тестової вибірки. Також оцінено три різні архітектури ConvNet:

- 1) тришарова архітектура ConvNet;
- 2) чотиришарова архітектура ConvNet;
- 3) більш глибока шестишарова архітектура ConvNet.

Тришарова архітектура ConvNet містить одну згортку, один пулінг або шар субдискретизації і один повнозв'язний шар. Чотиришарова архітектура мала два згорткові і два пулінгові шари, а

останній шар був повнозв'язний. Шестишарова архітектура ConvNet складається з чотирьох згорткових і пулінгових шарів з 16 одиницями, повністю з'єднаних між собою. Згідно з результатами у [18] глибока архітектура (шеститишарова архітектура ConvNet) дає кращий результат із точністю до 1,06% порівняно з чотиришаровою архітектурою.

Автори [19] запропонували навчання з частковим залученням учителя (напіваавтоматичне навчання, SSL) із використанням згорткових нейронних мереж. Точність розробленої моделі становила 82,43%, а площа під ROC-кривою, що спостерігається у ході дослідження, – 88,18%. В експериментах використовувалися 1874 пари зображень мамограм. Крім цього, автори розробили три рівняння зважування даних, використовуючи експонентну функцію, гаусівську функцію і функцію Лапласа. За результатами [19] у порівнянні з двома іншими ваговими рівняннями експоненціальна функція дала кращі результати для точності маркування, чутливості і специфічності – 82,43%, 81,00% і 72,26% відповідно.

У праці [20] автори застосували метод головних компонент (PCA) для Hybrid Fuzzy CNN Network. Ідея використання PCA полягала у зменшенні кількості видобутих інформаційних ознак. Також автори запропонували модель, у якій CNN VGG 16 використовувалась для отримання інформаційних ознак, а FNN NEFClass – для задач класифікації.

У [39] використано набір даних ImageNet для навчання згорткових нейронних мереж (CNN). У процесі дослідження автори вилучали ділянки розмірами 32x32 і 64x64 із зображень, щоб навчити CNN. У результатах автори показали, що точність CNN зменшується зі збільшенням масштабування.

У праці [40] запропоновано модифікацію мережі, де для вилучення інформаційних ознак використано CNN і як класифікатор застосовано SVM, щоб унормувати мамографічні скани за допомогою методу, який запропоновано у праці [41].

Результат експерименту у [38] показав, що якщо кількість класу для класифікації дорівнює чотирьом, то точність моделі становить 85%, а для двох класів – 93%. Для експерименту автори використовували набір даних Breast Cancer Histology Challenge (BACH), де кожен скан має розмірність 2048×1536 пікселів.

У [48] проведено експериментальний аналіз між існуючими згортковими мережами LeNet, AlexNet, GoogleNet і запропонованою автором модифікованою згортковою мережею. На основі експериментів встановлено, що запропонована архітектура згорткової нейронної мережі [48] перевершує три інші мережі в класифікації медичних зображень. Точність запропонованої моделі становила 81% і є кращою порівняно з іншими згортковими мережами, а саме LeNet, AlexNet і GoogleNet. Архітектура запропонованої моделі складалась із чотирьох згорткових, двох пулінгових і двох повнозв'язних шарів. Як навчальну вибірку використано 37198 МРТ-сканів і 500 МРТ-сканів – як тестову вибірку. На рис. 1.4 проілюстровано архітектуру згорткової мережі, яку запропоновано у праці [48].

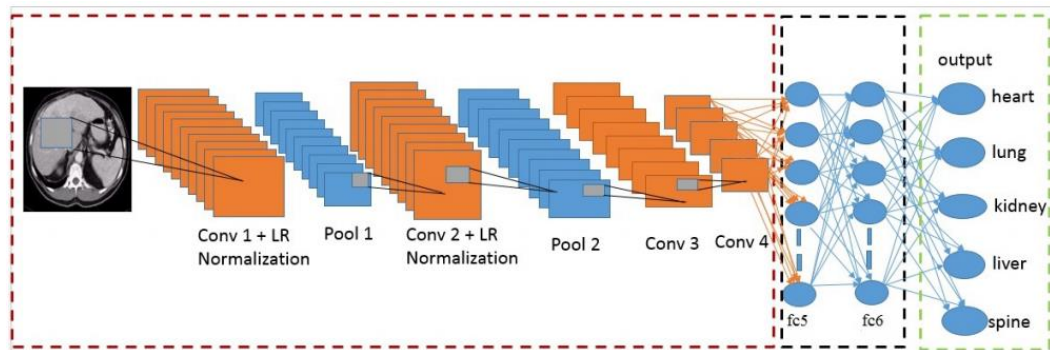


Рис. 1.4. Архітектура згорткової мережі у [48]

У праці [49] запропоновано модель, де згорткова мережа Xception застосовується для вилучення інформаційних ознак, а мережа Long short-term memory (LSTM) – як класифікація. Також автори порівняли різні класифікатори, експеримент показав, що кращий результат буде за використання LSTM із точністю 90,97%.

1.5 Постановка задачі дисертаційного дослідження

1. Аналіз існуючих робіт на тему розпізнавання раку молочної залози і деменції.
2. Розроблення методів оброблення медичних зображень таких, як перетворення колірного простору, стандартизація, скорочення розмірів, і т. д..
3. Проведення експериментального дослідження існуючих згорткових мереж.
4. Розроблення модифікації моделі згорткових мереж для поставленої задачі.
5. Проведення експериментальних досліджень отриманої моделі.
6. Реалізація запропонованого методу розпізнавання раку на основі медичних зображень.

1.6 Висновки до розділу

- Виконано аналіз поширення раку молочної залози, встановлено, що рак молочної залози є одним з найнебезпечніших видів раку з великою імовірністю смерті. Але його розпізнавання на ранніх етапах може сприяти зменшенню кількості летальних випадків.
- Розглянуто процес розпізнавання раку молочної залози. Виконано огляд існуючих систем розпізнавання раку молочної залози. Показано, що натеper найбільш поширеними методами розпізнавання і діагностики раку молочної залози є: метод опорних векторів і методи на основі штучних нейронних мереж. На жаль, більшість із них у діагностиці раку молочної залози ґрунтувались на точності моделі. Однак не менш важливим фактором для визначення кращої моделі є чутливість, оскільки цей фактор визначає на скільки правильно модель визначила злоякісні пухлини.
- Проаналізовано основні методи для розпізнавання раку молочної залози.
- Розглянуто навчання згорткових мереж для розпізнавання раку на медичних зображеннях на основі оброблення вхідних даних. Для цього потрібно нормалізувати дані. Також за необхідності збільшити кількість навчальних даних за допомогою аугментації.

РОЗДІЛ 2

АНАЛІЗ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ЗАДАЧ ДІАГНОСТИКИ ЗАХВОРЮВАН

2.1 Основні положення теорії штучних нейронних мереж

Штучні нейронні мережі (ШНМ) виникли на основі знань про функціонування нервової системи живих істот. Нервова клітина, нейрон, – це особлива біологічна клітина, яка обробляє інформацію. Варто пам'ятати, що у словосполученні «нейронна мережа» наголос ставити на слові «мережа», а не на слові «нейронна». Сенс цього зауваження полягає в тому, що те, як штучні нейрони з'єднані або об'єднані в мережу набагато важливіше за те, як кожен нейрон виконує свою просту операцію, для якої він призначений.

Штучний нейрон має імітувати дію біологічного нейрона, тобто приймати безліч різних сигналів, x_i , від багатьох сусідніх нейронів і обробляти їх визначеним простим способом. Залежно від результату цього оброблення, нейрон j вирішує, чи вилучати вихідний сигнал y_i , чині. Вихідний сигнал (якщо він спрацював) може бути або 0, або 1, або мати будь-який числове значення між 0 і 1 (рис. 2.1) залежно від того, чи мається справа з «бінарними» або «реально значущими» штучними нейронами.

Переважно з історичного погляду функція, що обчислює вихід з m -мірного вхідного вектора X , $f(X)$, розглядається як така, що складається з двох частин. Перша частина оцінює так званий «чистий вхід», друга «передає» чистий вхід нелінійним способом на вихідне значення y .

Перша функція є лінійною комбінацією вхідних змінних, x_1, x_2, \dots, x_m , помножену на коефіцієнти, w_{ij} , звані вагами, а друга функція слугує передавальною функцією, оскільки передає сигнал(и) через аксон

нейрона на дендрити інших нейронів. Для розрахунку значення виходу обчислюється u_j на j -му нейроні. Спочатку обчислюється чистий вхід за рівнянням (2.1)

$$Net_j = \sum_{i=1}^m w_{ij} x_i , \quad (2.1)$$

а потім Net_j використовується як аргумент у передавальній функції (2.2)

$$y_j = out_j = \frac{1}{1+e^{-\alpha_{ij}(Net_j+\theta_{ij})}} . \quad (2.2)$$

Ваги w_{ij} у штучних нейронах є аналогами реальних напружень нейронного синапсу між аксонами, що подають сигнали, і дендритами, які приймають ці сигнали (рис. 2.1). Кожна сила синапсу між аксонами і дендритами визначає, яка частка вхідного сигналу передається в тіло нейронів.

Вважається, що знання у мозку здобуваються завдяки постійній адаптації синапсів до різних вхідних сигналів, що призводить до кращого виходу сигналів, тобто таких сигналів, які викличуть правильну реакцію або відповіді тіла. Результати постійно повертаються як нові вхідні дані. Аналогічно до реального мозку штучні нейрони намагаються імітувати адаптацію синапсиса за допомогою ітераційної адаптації ваг w_{ij} у нейронах відповідно до відмінностей між реально отриманими результатами y_j та бажаними відповідями або цілями t_j .

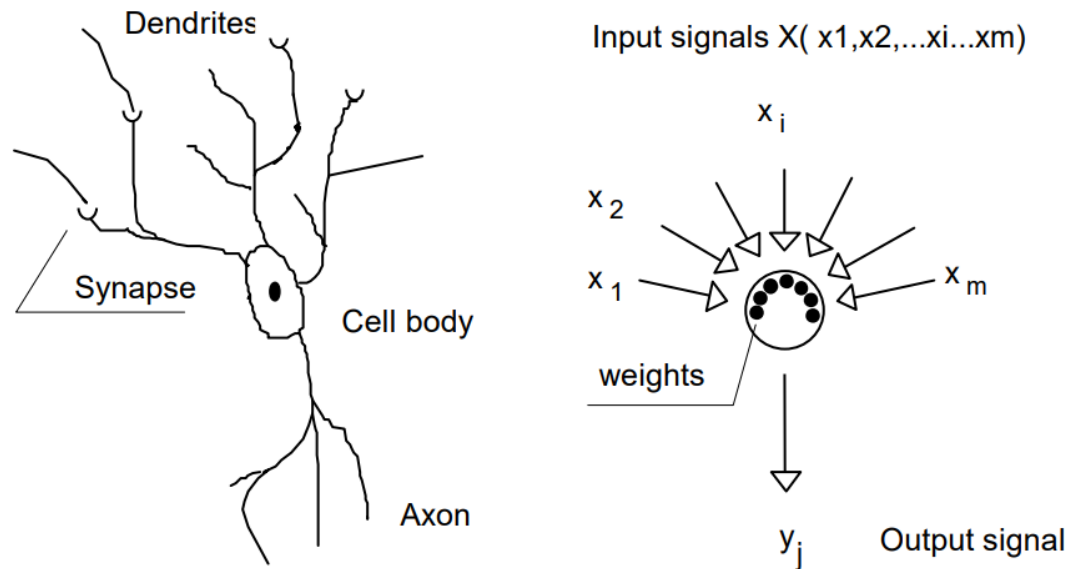


Рис. 2.1. Порівняння між біологічним і штучним нейроном.

Коло, що імітує тіло клітини нейрона, є простою математичною процедурою, яка робить один вихідний сигнал y_j із заданих вхідних сигналів, відображених через вектор X

Якщо вихідна функція $f(Net_j)$ є двійковою пороговою функцією (рис. 2.2,(а)), то вихід y_j одного нейрона має тільки два значення: 0 або 1. Однак здебільшого за ШНМ-підходу мережі нейронів складаються з нейронів, що мають передавальну функцію сигмоїдальної форми (формула 2.2, рис. 2.2,(б)). У деяких випадках передавальна функція в штучних нейронах може мати так звану кругову форму:

$$y_j = out_j = e^{-[(\alpha_j Net_j + \theta_j)]^2} . \quad (2.3)$$

Деякі можливі форми передавальної функції зображено на рис. 2.2. Форма передавальної функції після того, як вона буде обрана, полягає в подальшому використанні для всіх нейронів у мережі незалежно від того, де вони розміщені і яким чином вони пов'язані з іншими нейронами. Що

змінюється під час навчання (бо навчання це не функція) – це ваги та параметри функції, які керують положенням порогового значення θ_j та нахилом передавальної функції α_j (рівняння 2.2 та 2.3).

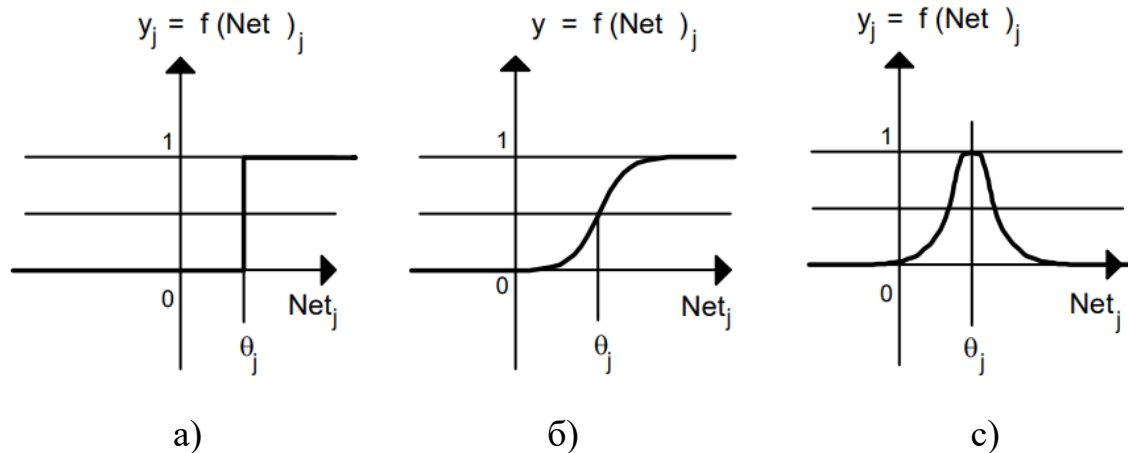


Рис. 2.2. Три різні передавальні функції: порогова – а), сигмовидна – б), радіальна – с). Параметр θ_j в усіх трьох функціях визначає значення Net_j , навколо якого нейрон найбільш вибірковий. Другий параметр, α_j , що розглядається у рівняннях 2.1 і 2.2, впливає на нахил передавальної функції (не застосовується у разі, якщо а) перебуває на околиці θ_j).

Два згадані параметри, α_j і θ_j , можуть бути оброблені (скориговані у ході навчання) у всіх ШНМ, що мають таку передавальну функцію, так само, як і вагові коефіцієнти w_{ij} .

Спочатку запишемо аргумент u_j функції $f(u_j)$ (рівняння 2.3) у розширеному вигляді, включивши в нього рівняння 2.1:

$$u_j = \alpha_j (\text{Net}_j + \theta_j) \quad (2.4)$$

$$= \sum_{i=1}^m \alpha_j w_{ji} x_i + \alpha_j \theta_j = \alpha_j w_{j1} x_1 + \alpha_j w_{j2} x_2 + \dots + \alpha_j w_{ji} x_i + \dots + \alpha_j w_{jm} x_m + \alpha_j \theta_j \quad (2.5)$$

На початку навчального процесу жодна з констант, α_j , w_{j1} або θ_j , не відома. Отже, не має значення, чи написаний такий добуток, як $\alpha_j w_{ji}$ і $\alpha_j \theta_j$ як одне постійне значення, або як добуток двох констант. Тому якщо просто записати добуток двох невідомих як тільки один невідомий параметр, можна отримати рівняння (2.5) у вигляді

$$u_j = w_{j1} x_1 + w_{j2} x_2 + \dots + w_{ji} x_i + \dots + w_{jm} x_m + w_{jm+1} \quad (2.6)$$

Тепер, додавши одну змінну x_{m+1} до всіх вхідних векторів X , можна отримати так званий розширений вектор ($X' = x, x_2, \dots, x_i, \dots, x_m, x_{m+1}$). Після того, як ця додаткова змінна буде встановлена на 1, у всіх без винятку випадках можна записати всі доповнені вхідні вектори X' як $(x, x_2, \dots, x_i, \dots, x_m, 1)$.

Тепер можна об'єднати останні ваги w_{jm+1} із продукту $\alpha_j \theta_j$ в один новий термін Net_j , що містить всі параметри (ваги, поріг і нахил) для адаптації (навчання) нейрона у тій самій формі.

Додаткова вхідна змінна, яка завжди дорівнює 1, називається зміщенням. Зміщення робить нейрони істотно гнучкішими й адаптивнішими. Рівняння для отримання вихідного сигналу від такого нейрона тепер виходить шляхом об'єднання рівняння 2.6 із будь-яким із рівнянь (2.2) або (2.3) залежно від бажаної форми передавальної функції:

$$y_j = out_j = \frac{1}{1 + e^{-Net_j}}$$

$$y_j = out_j = e^{-Net_j^2},$$

де
$$Net_j = \sum_{i=1}^{m+1} w_{ji} x_i$$

Коли розглянуто штучні нейрони і те, як вони керують вхідними сигналами, можна розглянути великі ансамблі нейронів. Великі ансамблі нейронів, у яких більшість нейронів взаємопов'язані, називають нейронними мережами. Якщо окремі нейрони моделюються комп'ютерами, то можна моделювати й ансамблі нейронів. Ансамблі нейронів, які моделюються на комп'ютерах, називаються штучними нейронними мережами (ШНМ).

ШНМ можуть складатися з різної кількості нейронів. У хімічних додатках розміри ШНМ, тобто кількість нейронів, варіюється від десятків тисяч до всього лише десяти. Усі нейрони у ШНМ можуть поміщатися в один чи два шари, або утворювати три і навіть більше. На рис. 2.3 показано відмінність між одношаровою та багатшаровою структурою ШНМ.

На рис. 2.3 одношарова мережа має чотири нейрони (іноді їх називають вузлами), кожен з яких має чотири ваги. Усього у цій одношаровій мережі є 16 ваг. Кожен із чотирьох нейронів приймає всі вхідні сигнали, а також додатковий вхід від зміщення, який завжди дорівнює 1. Однак той факт, що вхід дорівнює 1, не перешкоджає зміні ваг, що ведуть від зміщення до вузлів.

На рис. 2.3 нейрони зображено у вигляді кіл, а ваги – у вигляді ліній, що з'єднують кола (вузли) у різних шарах. Як видно, лінії (ваги) між вхідними змінними і першим шаром не мають ні вузлів, ні нейронів, позначених зверху. Для подолання цієї непослідовності входом вважається неактивний шар нейронів, який слугує тільки для розподілу.

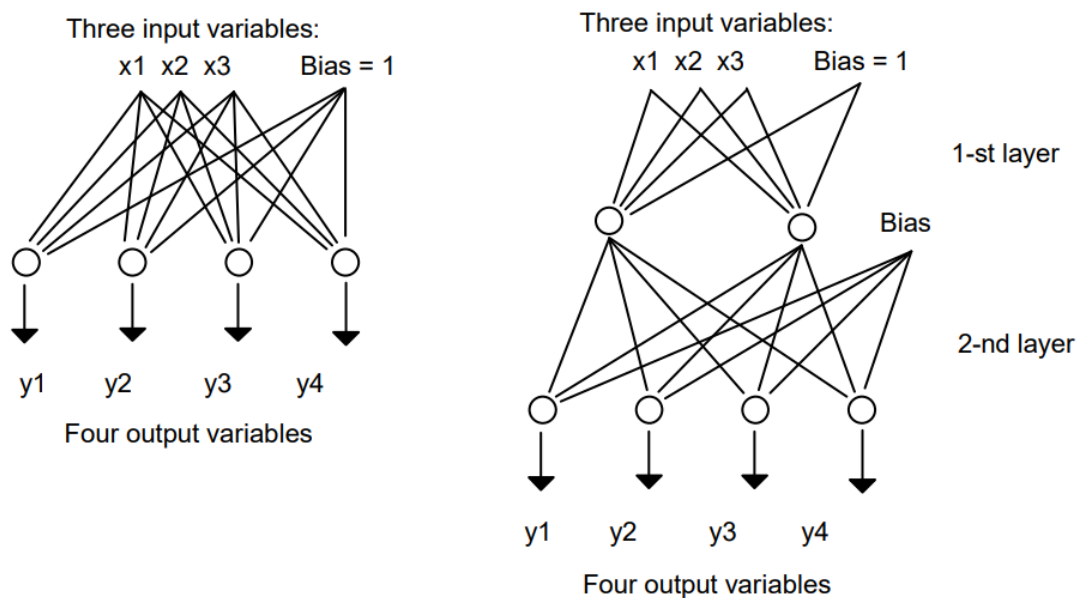


Рис. 2.3. Одношарові (зліва) та двошарові (справа) ШНМ

2.2 Згорткові нейронні мережі

Згорткова нейронна мережа (англ. convolutional neural network, CNN) – це мережа глибокого навчання, яка може приймати вхідне зображення, надавати важливість вагам і спотворенням, а також різним аспектам/об’єктам зображення з можливістю їх розрізнення.

Якщо ідея глибокого багатошарового перцептрона (англ. multilayer perceptron, MLP) полягає в тому, що ієрархії ознак легше захоплюють важливі частини вхідного сигналу, то ідея CNN – трансляційна інваріантність: багато ознак у зображенні є трансляційно інваріантними. Наприклад, якщо розроблено автомобіль, його можна спробувати виявити за його деталями [103]. Але тоді існує багато положень, при яких можуть бути колеса. У ході їх комбінування бажано захопити ознаки низького рівня, трансляційно інваріантні ознаки на нижніх шарах ШНМ, і

високорівневі ознак, які є комбінаціями низькорівневих ознак, на верхніх шарах.

Також моделі мають зважати на те, що пікселі зображень є впорядкованими. Один із способів використання цього методу – навчання фільтрів зображень у так званих згорткових шарах.

Якщо MLP векторизує вхід, то вхід шару у CNN є картою ознак. Карта ознак – це матриця $m \in R^{w \times h}$, але зазвичай ширина дорівнює висоті ($w = h$). Для зображення у форматі RGB кількість карт ознак становить $d = 3$. Кожен кольоровий канал – це карта характеристик.

Оскільки AlexNet [104] майже вдвічі зменшила помилку в задачі ImageNet, CNN є найсучаснішими в різних задачах комп'ютерного зору.

Традиційні CNN мають три важливі інструменти побудови:

- 1) згорткові шари з нелінійною функцією активації;
- 2) пулінгові шари;
- 3) шари нормалізації.

2.2.1 Згорткові шари.

Згорткові шари беруть кілька карт характеристик як вихідні і виробляють n карт ознак як вихідні, де n – кількість фільтрів у згортковому шарі. Ваги фільтрів лінійної згортки є параметрами, які адаптовані до навчальних даних. Кількість n фільтрів, а також розмір фільтра $k_w \times k_h$ є гіперпараметрами згорткових шарів. Фільтри мають розмір $k_w \times k_h \times d(i-1)d^{(i-1)}$, де $d^{(i-1)}$ – кількість карт характеристик (ознак) вхідного шару $(i-1)$.

Іншим гіперпараметром шарів згортки є крок $s \in N \geq 1$ і прокладка (англ. padding). Padding (зазвичай нульова [105]–[107]) використовується для того, щоб розмір карт ознак не змінювався.

Гіперпараметрами згорткових шарів є:

- кількість фільтрів $n \in N \geq 1$;
- $k_w, k_h \in N \geq 1$ розміру фільтра $k_w \times k_h \times d^{(i-1)}$;
- функція активації шару;
- крок $s \in N \geq 1$.

Типовими варіантами є: $n \in \{32, 64, 128\}$, $k_w = k \in \{1, 3, 5, 11\}$ (наприклад, у [108]–[110]), активація випрямленого лінійного блоку (ReLU) та $s = 1$.

Концепція розподілу ваг має вирішальне значення для CNN. Цю концепцію подано у [111]. За розподілу ваг фільтри можна навчати методом стохастичного градієнтного спуску (англ. stochastic gradient descent, SGD) так само, як і багатошарові перцептрони. Фактично кожна CNN має еквівалентний багатошаровий перцептрон, який обчислює ту ж саму функцію, якщо порівнювати тільки сплющений вихід.

Це легше побачити, коли операцію фільтрації виражено

$$o^i(x) = b + \sum_{j=1}^k w_{ij} \cdot x_j \text{ и } i \in \{1, \dots, w\} \times \{1, \dots, h\} \times \{1, \dots, d\}$$

$$o^{(x,y,z)}(I) = b + \sum_{i_x=1-\lfloor \frac{k_w}{2} \rfloor}^{\lfloor \frac{k_w}{2} \rfloor} \sum_{i_y=1-\lfloor \frac{k_h}{2} \rfloor}^{\lfloor \frac{k_h}{2} \rfloor} \sum_{i_c}^d F_z(i_x, i_y, i_c) \cdot I(x + i_x, y + i_y, i_c)$$

де $b \in R$, $X \in \{1, \dots, w\}$, $y \in \{1, \dots, h\}$ та $z \in \{1, \dots, d\}$.

Видно, що більшість ваг еквівалентного багатошарового перцептрона дорівнюють нуль, а частина еквівалентні. Отже, перевага CNN порівняно з MLP – зниження параметрів. Ефект від меншої кількості параметрів полягає в тому, що для отримання потрібних їм оцінок потрібно менше навчальних даних. Це означає, що MLP здатний обчислити ті ж

функції, що і CNN, але, швидше за все, матиме гірші результати на тому ж самому наборі даних, якщо архітектура CNN підходить для цього набору даних.

Застосування згорткового шару візуалізовано на рис. 2.4.

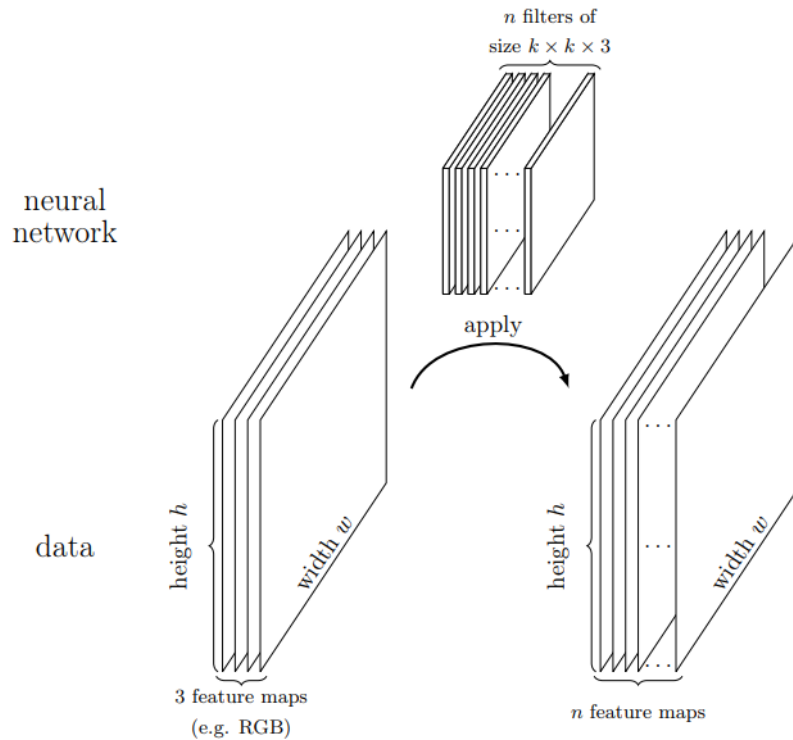


Рис. 2.4. Застосування одного згорткового шару з n фільтрами розміру $k \times k \times 3$ із кроком $s = 1$ для уведення даних розміру ширини на висоту з трьома каналами (RGB)

Згортковий шар з n фільтрами розміру $k_w \times k_h$ і такою ж прокладкою після $d^{(i-1)}$ карт ознак розміру $s_x \times s_y$ має $n \cdot d^{(i-1)} \cdot (k_w \cdot k_h)$ параметрів, якщо не використовується зміщення. І навпаки, повнозв'язний шар, який виробляє той самий вихідний розмір і не використовує зміщення, матиме $n \cdot d^{(i-1)} \cdot (s_x \times s_y)^2$ параметрів. Це означає, що згортковий шар має значно менше параметрів. З одного боку, це значить, що він може вивчити менш складні межі рішень, а з іншого – що потрібно вивчати менше параметрів,

тобто процедура оптимізації потребує менше прикладів і мета оптимізації простіша.

Необхідно відзначити, що навіть згортковий шар фільтрів 1×1 дійсно враховує лінійну комбінацію карт вхідних ознак d . Це можна використати для зменшення розмірності, якщо в згортковому шарі менше фільтрів 1×1 , ніж у картах вхідних ознак.

Повнозв'язний шар із входами $d \in \mathbb{N}_{\geq 1}$ і $n \in \mathbb{N}_{\geq 1}$ можна інтерпретувати як згортковий шар із входами форми $1 \times 1 \times d$ і n фільтрів розміру 1×1 . У результаті виходить вихідна форма $1 \times 1 \times n$. Кожен вихід з'єднаний з усіма входами.

Коли за згортковим шаром іде повнозв'язний, необхідно векторизувати його на карті ознак. Якщо до векторизованого виходу застосувати згортковий фільтр 1×1 , то він повністю еквівалентний повнозв'язному шару. Однак векторизацію можна пропустити, якщо на виході застосувати згортковий шар без прокладки і розмір фільтра дорівнює розміру карти ознаки. Це використано у [112].

2.2.2 Пулінг або шар субдискретизації.

Пулінг підсумовує розмірність $p \times p$ на вхідній карті ознак. Як і згорткові шари, пулінг можна використовувати з кроком $s \in \mathbb{N}_{> 1}$. Оскільки $s \geq 2$ є звичайним вибором, пулінгові шари іноді називають субдискретизацією.

Як правило, $p \in \{2, 3, 4, 5\}$ і $s = 2$, наприклад, для AlexNet [113] і VGG-16 [114].

Вид зведення для набору активацій A варіюється між функціями, перерахованими в табл. 2.1, просторовим об'єднанням пулінгу, поданим у [116], і узагальнювальними функціями пулінгу, поданими в [117].

Таблиця 2.1

Типи пулінгів для безлічі A активацій $a \in R$

Назва	Визначення	Використаний
Max Pooling	$\max \{a \in A\}$	[117],[118]
Average/Mean Pooling	$\frac{1}{ A } \sum_{a \in A} a$	LeNet-5 [119], [120]
l_2 Pooling	$\sqrt{\sum_{a \in A} a^2}$	[121]
Stochastic Pooling	* ¹	[122]

Пул застосовується з трьох причин, щоб:

- 1) отримати локальну трансляційну інваріантність;
- 2) отримати інваріантність за незначних локальних змін i ;
- 3) скоротити обсяг даних до $\frac{1}{s^2}$ за допомогою кроків $s > 1$.

Максимальний пулінг зображено на рис. 2.5.

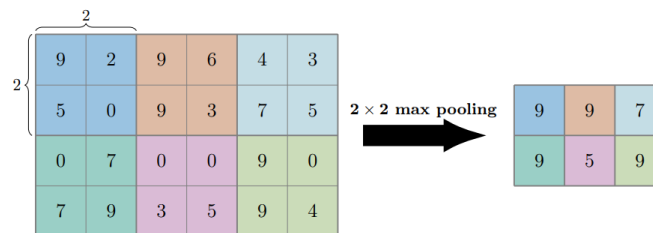


Рис. 2.5. 2×2 максимальний пулінг, нанесений на карту ознак розміром 6×4 із кроком $s = 2$ і padding

¹ Для стохастичного пулінгу кожне зі значень активації $p \times p$ a_i на ділянці пулінгу вибирається з імовірністю $p_i = \frac{a_i}{\sum_{a_j \in A} a_j}$. Це передбачає, що активації a_i невід'ємні.

Середнє значення пулінгу $p \times p$ площ із кроком s може замінюватись згортковим шаром. Якщо на вході пулінгового шару є карти ознак $d^{(i-1)}$, то згортковий шар має мати фільтри $d^{(i-1)}$ розміром $p \times p$ і кроком s . Перший фільтр має значення

$$\begin{pmatrix} \frac{1}{p^2} & \cdots & \frac{1}{p^2} \\ \vdots & \ddots & \vdots \\ \frac{1}{p^2} & \cdots & \frac{1}{p^2} \end{pmatrix}$$

для вимірювання i та нульової матриці,

$$\begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

для решти вимірів $i = 1, \dots, d^{(i-1)}$.

2.2.3 Виключення (Dropout).

Виключення (англ. dropout) – метод, який використовується для запобігання перенавчання і коадаптації нейронів шляхом установлення виходу будь-якого нейрона на нуль з імовірністю p , де перенавчання – це коли модель добре працює для навчальної вибірки, але не для тестової вибірки; і недонавчання – коли модель працює погано і для навчальної, і для тестової вибірок. Метод dropout подано у праці [123] і добре описано у [124].

Виключений шар можна реалізувати так: для входу будь-якої форми s вибирається тензор тієї ж форми $B \in \{0,1\}^s$, при цьому кожен елемент

d_i вибирається незалежно від розподілу Бернуллі. Результати множаться на елементи, щоб обчислити вихід із шару виключення:

$$Out = D \cdot in \text{ с } d_i \sim B(1, p) \quad ,$$

де \cdot – добуток Адамара,

$$(A \cdot B)_{ij} := (A)_{ij} \cdot (B)_{ij} .$$

Отже, кожне значення входу встановлюється на нуль з імовірністю виключення p . Зазвичай використовується dropout із $p = 0,5$. Шари, розміщені ближче до входу, зазвичай мають меншу ймовірність виключення, ніж більш глибокі шари. Для того, щоб зберегти очікуваний результат на тому ж значенні, дані шару виключення множаться на $\frac{1}{1-p}$, якщо виключення активоване [125], [126]. Під час виведення виключення деактивується.

Виключення зазвичай застосовується тільки після повнозв'язних шарів, а не після згорткових, оскільки це зазвичай збільшує помилку тесту, як зазначено у [127].

Моделі, які використовують dropout, можуть інтерпретуватися як ансамбль моделей із різною кількістю нейронів у кожному шарі, але також із поділом ваг.

Концептуально схожими є DropConnect і мережі зі стохастичною глибиною. DropConnect [128] є узагальненням dropout, яке встановлює ваги на нуль, на відміну від установки виходу нейрона на нуль. У мережах зі стохастичною глибиною, уведеною у [129], виключаються тільки повнозв'язні шари. Це можна зробити за наявності залишкових мереж, які

мають одне ідентифікаційне з'єднання і одне залишкове. Таким чином, залишкові ознаки можуть бути виключені, а ідентифікаційне з'єднання залишиться.

2.2.4 Шар нормалізації.

Однією з проблем навчання глибоких нейронних мереж є внутрішнє коваріантне зрушення: коли параметри близьких до виходу шарів адаптовано до деяких входів, виробленими нижніми шарами, параметри нижніх шарів також є адаптованими. Це призводить до того, що параметри верхніх шарів гірші. Низька швидкість навчання (англ. learning rate, LR) обирається для пристосування до радикальних змін входних ознак із часом.

Одним із способів вирішення цієї проблеми є нормалізація мініпартій (англ. mini-batches), як описано у [130]. Пакетна нормалізація (англ. batch normalization, BN) з d -мірним входом $x = (x^{(1)}, \dots, x^{(d)})$ спочатку нормалізується за точкою

$$\hat{x}^{(k)} = \frac{x^{(k)} - \bar{x}^{(k)}}{\sqrt{s'[x^{(k)}]^2 + \varepsilon}}$$

При цьому $\bar{x}^{(k)} = \frac{1}{m} \sum_{i=1}^m x^{(k)}_i$ – середнє значення вибірки, а $s'[x^{(k)}]^2 = \frac{1}{m} \sum_{i=1}^m (x^{(k)}_i - \bar{x}^{(k)})^2$ – дисперсія вибірки, де $m \in \mathbb{N} \geq 1$ – кількість навчальних зразків на mini-batch, що є малою константою для запобігання поділу на нуль, $x^{(k)}_i$ – активація нейрона k для навчальної вибірки i .

Крім цього, для кожної активації $x^{(k)}$ вводяться два параметри, $\gamma^{(k)}$, $\beta^{(k)}$, які масштабують і зрушують ознаки:

$$y^{(k)} = \hat{x}^{(k)} \cdot \gamma^{(k)} + \beta^{(k)}$$

У разі повнозв'язних шарів параметри застосовуються для активації до застосування нелінійності. Якщо параметри застосовується після активації, то на ранніх стадіях навчання вони завдають шкоди. Для згортки на карті ознак вивчається тільки один γ і один β .

Одним із особливих випадків є $\gamma^{(k)} = \sqrt{s'[x^{(k)}]^2 + \varepsilon}$ і $\beta^{(k)} = \bar{x}^{(k)}$, який зробить шар BN шаром ідентичності.

Під час проведення оцінювання² очікуване значення і дисперсія розраховуються один раз для повного набору даних. Використовується об'єктивне оцінювання емпіричної дисперсії.

Питання, де мають застосовуватися шари пакетної нормалізації і з яких причин, ще відкрите. Для Dropout не має значення, чи застосовується він до або після функції активації. З огляду на це можливі варіанти для послідовності:

1. CONV / FC \rightarrow BN \rightarrow activation function \rightarrow Dropout \rightarrow . . .
2. CONV / FC \rightarrow activation function \rightarrow BN \rightarrow Dropout \rightarrow . . .
3. CONV / FC \rightarrow activation function \rightarrow Dropout \rightarrow BN \rightarrow . . .
4. CONV / FC \rightarrow Dropout \rightarrow BN \rightarrow activation function \rightarrow . . .

де CONV – операції згортки, FC – повнозв'язний шар.

Автори у [130] пропонують використовувати BN перед функцією активації, як у варіантах 1 і 4. BN після активації дає кращі результати (див. на сайті <https://github.com/ducha-aiki/caffenet-benchmark/blob/master/batchnorm.md>).

Інший рівень нормалізації – це локальна нормалізація реакції, як описано у [131], яка містить l_2 нормалізацію, як описано у [132]. Ці два шари нормалізації однак замінені на BN.

² Також называемое время вывода

2.3. Блоки згорткових нейронних мереж (CNN)

2.3.1 Залишкові блоки.

Залишкові блоки (англ. residual blocks) у тому вигляді, як їх подано у [133], є важливим розділом у галузі комп'ютерного зору. Вони дали змогу спільноті комп'ютерного зору перейти від приблизно 16 шарів, як у VGG 16-D, до декількох сотень шарів. Ключова ідея глибоких залишкових мереж (англ. Residual Network, ResNets), яку описано у [133], полягає в додаванні ідентифікаційного з'єднання, яке пропускає два рівні. Це ідентифікаційне з'єднання додає карти характеристик до інших карт характеристик i , таким чином, потребує, щоб вихідний шар вхідного шару залишкового блоку був того ж розміру, що і його останній шар.

Це також можна описати так: якщо x_i – це карти ознак після шару i , а x_0 – вхідне зображення, то H – це нелінійне перетворення карт ознак. Тоді

$$Y = H(x)$$

описує традиційну CNN. Це може бути кілька шарів. Залишковий блок, як зображено на рис. 2.6, описується так

$$Y = H(x) + x$$

У [107] автори використовували тільки залишкові скіпові з'єднання для пропускання двох шарів. Отже, якщо $conv_i(x_i)$ описує застосування згорткового шару i до входу x_i без нелінійності, то таким залишковим блоком є

$$x_{i+2} = conv_{i+1} \left(ReLU(conv_i(x_i)) \right) + x_i$$

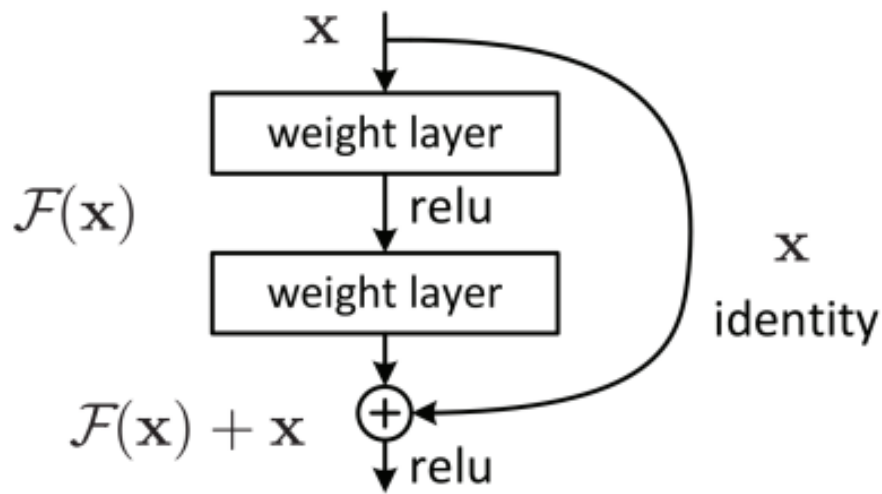


Рис. 2.6. Модуль ResNet

2.3.2 Агрегаційні блоки.

Два найпоширеніші способи додавання більшої кількості параметрів у нейронні мережі: 1) збільшення їх глибини шляхом додавання більшої кількості шарів; 2) збільшення їх ширини шляхом додавання більшої кількості нейронів/фільтрів. У працях [134], [135] описано нову ідею агрегаційних блоків “ResNeXt block: Підвищення кардинальності”

$$C \in \mathbb{N} \geq 1$$

Під кардинальністю автори розуміють концепцію наявності невеликих конволюційних мереж C із однаковою топологією, але різними вагами. Цю концепцію зображено на рис. 2.7.

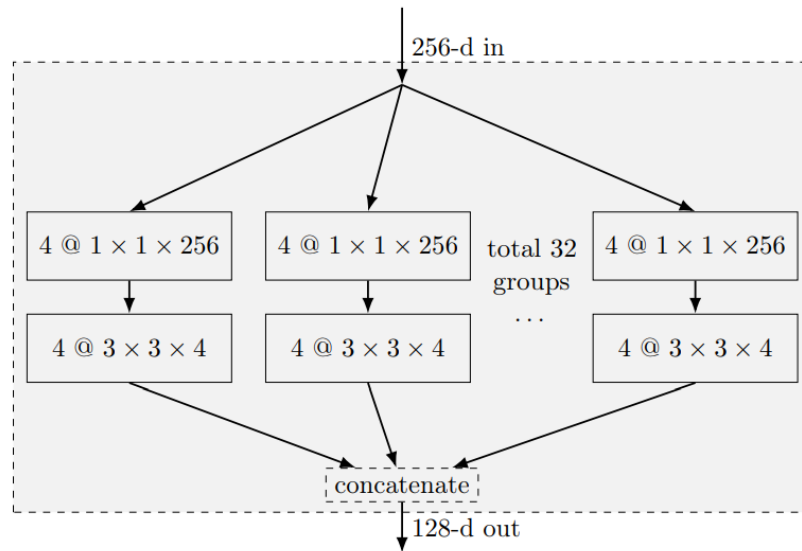


Рис. 2.7. Агрегаційний блок із кардинальністю $C = 32$. Кожна з 32 груп є двошаровою конволюційною мережею. Перший шар отримує 25 карт ознак і застосовує до них чотири фільтри 1×1 . Другий шар застосовує чотири фільтри 3×3 . Хоча кожна група має однакову топологію, навчена вага різна. Виходи груп зводяться воедино

Гіперпараметрами блоку агрегації є:

- Топологія членів групи.
- Кардинальність $C \in \mathbb{N} \geq 1$. Кардинальність $C = 1$ у кожному аспекті еквівалентна використанню групової мережі без блоку агрегування.

2.3.3 Щільні блоки.

Щільні блоки (англ. dense blocks) – набори згорткових шарів, які уведено у [136]. Ідея полягає в тому, щоб з'єднати кожен конволюційний шар безпосередньо з наступними конволюційними шарами. Традиційні CNN з L шарами і одним вхідним шаром мають з'єднання L між шарами, але щільні блоки мають з'єднання $\frac{L(L+1)}{2}$ між шарами. Вхідні карти ознак

зібрані воедино в глибину. На думку авторів [136], це запобігає вивільненню ознак і дозволяє використовувати набагато меншу кількість фільтрів на згортковому шарі. Там, де AlexNet і VGG-16 мають кілька сотень фільтрів на згортковому шарі (табл. 2.2 і 2.3), використано лише близько 12 карт об'єктів на шар.

На рис. 2.8 зображено щільний блок.

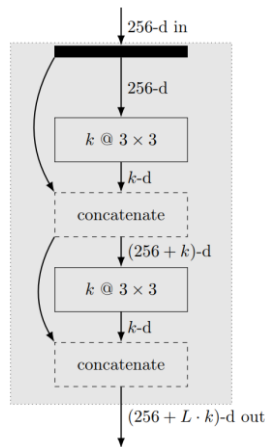


Рис. 2.8. Щільний блок з $L = 2$ шарами і коефіцієнтом росту k

Щільний блок має такі гіперпараметри:

- Використовується функція активації. Автори використовують ReLU.
- Розмір $k_w \times k_h$ фільтрів. Автори використовують $k_w = k_h = 3$.
- Кількість шарів L , де $L = 2$ – простий згортковий шар.
- Кількість k фільтрів, доданих на кожний шар.

Для зменшення кількості карт ознак $K \cdot L$ може знадобитися використання згортки 1×1 .

Таблиця 2.2

Архітектура AlexNet: особливим випадком AlexNet є угруповання згорток за допомогою обчислювальних обмежень на момент її розроблення. Це також зменшує кількість параметрів і дозволяє проводити паралельні обчислення на окремих GPU

#	Type	Filters @ Patch size / stride	Parameters	FLOPs	Output size
	Input				3 @ 224 × 224
1	Convolution	96 @ 11 × 11 × 3 / 4	34 944	211 M	96 @ 55 × 55
	LCN			12 M	96 @ 55 × 55
2	Max pooling	3 × 3 / 2	0	301 k	96 @ 27 × 27
3	Convolution	256 @ 5 × 5 × 48 / 1	307 456	448 M	256 @ 13 × 13
	LCN			3 M	256 @ 13 × 13
4	Max pooling	3 × 3 / 2	0	50 k	256 @ 13 × 13
5	Convolution	384 @ 3 × 3 × 256 / 1	885 120	299 M	384 @ 13 × 13
7	Convolution	384 @ 3 × 3 × 192 / 1	663 936	224 M	384 @ 13 × 13
9	Convolution	256 @ 3 × 3 × 192 / 1	442 624	150 M	256 @ 13 × 13
10	Max pooling	3 × 3 / 2	0	50 k	256 @ 6 × 6
11	FC	4096 neurons	37 752 832	75 M	4096
12	FC	4096 neurons	16 781 312	34 M	4096
13	FC	1000 neurons	4 097 000	8 M	1000
Σ			60 965 224	3300 M	1 122 568

Таблиця 2.3

Архітектура VGG-16: автори вирішили надати число тільки тим шарам, у яких є впізнавані параметри. Усі конвекції мають нульову прокладку для запобігання зміни розміру і використовують функції активації ReLU

#	Type	Filters @ Patch size / stride	Parameters	FLOPs	Output size
	Input				3 @ 224 × 224
1	Convolution	64 @ 3 × 3 × 3 / 1	1 792	186 M	64 @ 224 × 224
2	Convolution	64 @ 3 × 3 × 64 / 1	36 928	3712 M	64 @ 224 × 224
	Max pooling	2 × 2 / 2	0	2 M	64 @ 112 × 112
3	Convolution	128 @ 3 × 3 × 64 / 1	73 856	1856 M	128 @ 112 × 112
4	Convolution	128 @ 3 × 3 × 128 / 1	147 584	3705 M	128 @ 112 × 112
	Max pooling	2 × 2 / 2	0	1 M	128 @ 56 × 56
5	Convolution	256 @ 3 × 3 × 128 / 1	295 168	1853 M	256 @ 56 × 56
6	Convolution	256 @ 3 × 3 × 256 / 1	590 080	3703 M	256 @ 56 × 56
7	Convolution	256 @ 3 × 3 × 256 / 1	590 080	3703 M	256 @ 56 × 56
	Max pooling	2 × 2 / 2	0	<1 M	256 @ 28 × 28
8	Convolution	512 @ 3 × 3 × 256 / 1	1 180 160	1851 M	512 @ 28 × 28
9	Convolution	512 @ 3 × 3 × 512 / 1	2 359 808	3701 M	512 @ 28 × 28
10	Convolution	512 @ 3 × 3 × 512 / 1	2 359 808	3701 M	512 @ 28 × 28
	Max pooling	2 × 2 / 2	0	<1 M	512 @ 14 × 14
11	Convolution	512 @ 3 × 3 × 512 / 1	2 359 808	925 M	512 @ 14 × 14
12	Convolution	512 @ 3 × 3 × 512 / 1	2 359 808	925 M	512 @ 14 × 14
13	Convolution	512 @ 3 × 3 × 512 / 1	2 359 808	925 M	512 @ 14 × 14
	Max pooling	2 × 2 / 2	0	<1 M	512 @ 7 × 7
14	FC	4096 neurons	102 764 544	206 M	4096
	Dropout		0	0	4096
15	FC	4096 neurons	16 781 312	34 M	4096
	Dropout		0	0	4096
16	FC	1000 neurons	4 097 000	8 M	1000
Σ			138 357 544	31 000 M	15 245 800

Після порівняння табл. 2.2 і 2.3 можна зробити висновок, що час для навчання AlexNet значно менший, ніж для VGG-16, тому що кількість параметрів істотно менша. Отже, час обчислень скорочується.

2.3.4 Перехідні шари.

Перехідні шари використовуються для подолання обмежень, що виникають через обмеженість ресурсів або вибір архітектурної моделі. Одним із обмежень є кількість карт ознак. Для того щоб зменшити кількість карт ознак, зберігаючи при цьому якомога більше актуальної інформації в мережі, додається згортковий шар i з k_i -фільтрами форми $1 \times 1 \times k_{i-1}$. Кількість фільтрів k_i безпосередньо контролює кількість згенерованих карт ознак.

Щоб зменшити розмірність (ширину і висоту) карт об'єктів, зазвичай застосовується пулінг.

Глобальний пул – це ще один тип перехідного шару. Він застосовує об'єднання в пул усіх карт ознак, щоб зменшити вхід з постійною 1×1 карти ознак i , отже, дозволяє одній мережі мати різні розміри входу.

2.4 Автоенкодер

Автоенкодер (АЕ) – це нейронна мережа, яка за певних обмежень відновлює вхідний сигнал на виході. Інакши кажучи, для заданого вхідного вектора x , АЕ намагається апроксимувати [102]:

$$hw(x) \approx x$$

У 2006 р. запропоновано алгоритм попереднього навчання без учителя для визначення ваг і спотворень автоенкодера [97], який високоефективний за наявності лише невеликої кількості маркованих навчальних вибірок [96], [98]. АЕ реалізує попереднє навчання без учителя спочатку кодує, а потім декодує входи.

Для заданого вхідного вектора x кодувальник (англ. encoder) розраховує нелінійне відображення входів у вигляді:

$$e_i = \sigma(W x_i + b)$$

Тут σ позначає нелінійну функцію активації, W – ваги, b – спотворення кодувальника. Потім закодовані функції декодуються (англ. decoder) для відновлення заданого вхідного вектора x з використанням:

$$z_i = \sigma(\tilde{w} e_i + \tilde{b})$$

Тут \tilde{w} та \tilde{b} позначають ваги і спотворення декодера. Під час попереднього навчання без учителя мережа намагається мінімізувати помилку реконструкції,

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N (x_i - z_i)^2$$

регулюючи її ваги та спотворення $\theta = [W, b, \tilde{w}, \tilde{b}]$. Щоб мережа не могла вивчити функцію ідентифікації, до функції витрат додається параметр розрідженості (англ. sparsity). Цей параметр змушує мережу вивчати кореляцію між заданими вхідними векторами [99]. Після додавання параметра розрідженості функція витрат, таким чином, стає:

$$\arg \min_{\theta} J(\theta) = \frac{1}{N} \sum_{i=1}^N (x_i - z_i)^2 + \beta \sum_{j=1}^h KL(p \| p_j)$$

Тут h позначає кількість нейронів у прихованому шарі, β – пропорцію розрідженості, $\sum_{j=1}^h KL(p \| p_j)$ – розходження Куллбек-Лейблера (KL) між випадковими змінними Бернуллі із середнім p та p_j , відповідно. Розходження KL між двома випадковими змінними подано так:

$$KL(p||p_j) = p \log\left(\frac{p}{P_j}\right) + (1-p) \log\left(\frac{1-p}{1-P_j}\right)$$

де P_j – активація j -го нейрона в прихованому шарі, p – бажане середнє значення активації. Оскільки h – число нейронів у прихованому шарі, то розходження KL викликає активацію прихованої одиниці поблизу p .

Після підготовки без учителя декодер видаляється з мережі, а інші компоненти кодувального пристрою навчаються з учителем, додаючи після кодувального пристрою softmax-класифікатор з K нейронами, де K – це кількість класів. Класифікатор softmax є багатосторонньою версією логістичної регресії. Для заданого входу x_i функція softmax оцінює імовірність того, що $P(Y_k|x_i)$ для $k = 1, 2, \dots, K$, де K позначає кількість класів. Інакше кажучи, оцінюється імовірність того, що вхід x_i належить мітці класу Y_k . Математично клас імовірності p_k можна подати так:

$$P(y = k|x_i) = \frac{e^{\theta_k x_i}}{\sum_{k=1}^K e^{\theta_k x_i}}$$

Ваги і спотворення мережі θ можна оптимізувати, мінімізувавши таку функцію вартості:

$$J(\theta) = - \sum_{i=1}^N \sum_{k=1}^K 1\{y^i = k\} \log \frac{e^{\theta_k x_i}}{\sum_{k=1}^K e^{\theta_k x_i}} \quad (2.2)$$

де $1\{.\}$ – індикаторна функція, N – кількість помічених прикладів. Рівняння (2.2) можна розв'язати градієнтним алгоритмом. Цей процес називається точним налаштуванням (англ. fine-tuning), за якого мережа навчається з

учителям. Під час fine-tuning замість середньої квадратної помилки як опція втрат вибирається категорична перехресна ентропія збитку.

АЕ можна укласти ієрархічно так, щоб верхні шари отримували входи від виходів нижчих шарів. Функція активації ReLU використовується для нелінійності. Розбіжність KL для регуляризації розрідженості (англ. sparsity regularization) вибирається як 2, а β – як 0,1. Оптимізацію як попереднього навчання без учителя, так і fine-tuning обчислюють із використанням алгоритму оцінювання адаптивного моменту (англ. Adaptive Moment Estimation, ADAM) [100] зі швидкістю навчання 0,001. Таким чином, алгоритм допомагає моделі запобігти перенавчання (англ. overfitting). Перенавчання – це коли модель добре працює на навчальній вибірці, але не може також добре працювати на тестовій вибірці. Загальну архітектуру автоенкодера зображено на рис. 2.9.

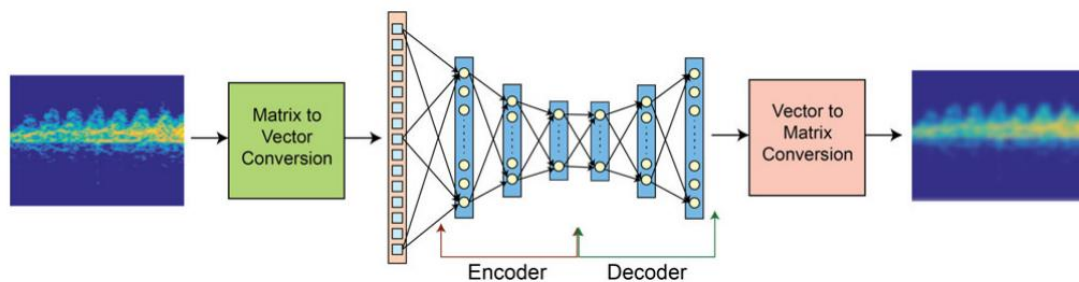


Рис. 2.9. Загальна архітектура АЕ

2.5 Згортковий автоенкодер

Згортковий автоенкодер поєднує в собі переваги згорткової фільтрації в CNN із попередньою підготовкою АЕ без учителя. Однак на відміну від топології АЕ замість повнозв'язних шарів енкодер містить згорткові шари, а декодер – зворотні згорткові шари. Крім цього, за кожним зворотним згортковим шаром має йти зворотний пулінговий шар [101]. Операція unpooling здійснюється шляхом збереження максимальних

значень під час пулінгу, збереження цих значень під час депулінгу і обнулення інших.

Просторова локалізація зберігається завдяки включенню операції згортки на кожному нейроні. Таким чином, для заданого входу матриці P кодувальний пристрій розраховує:

$$e_i = \sigma (P * F^n + b),$$

де σ – функція активації, $*$ – двомірна згортка, F^n – 2-D конволюційний фільтр, b – зміщення енкодера пристрою. Для збереження просторової роздільної здатності до вхідної матриці P застосовується нульове доповнення. Потім можна отримати реконструкцію, використовуючи:

$$z_i = \sigma(e_i * \tilde{F}^n + b),$$

де z_i – реконструкція входу матриці, \tilde{F}^n – n -ий 2-D згортковий фільтр у декодері, b – зміщення декодеру. Попереднє навчання без учителя може застосовуватись до мережі, яка спрямована на мінімізацію такого рівняння:

$$E(\theta) = \sum_{i=1}^m (x_i - z_i)^2.$$

Після попереднього навчання без учителя, unpooling зворотних згорткових шарів декодувальна частина мережі видаляється і в кінці мережі додаються повнозв'язні шари, а також softmax-класифікатор. Потім мережа може бути точно налаштована шляхом оптимізації (2.1). Як і в CNN, в АЕ використовується функція активації ReLU, а також алгоритм ADAM для оптимізації двох повнозв'язних шарів.

2.6. Методи аналізу згорткових мереж для задач діагностики раку молочної залози

Для CNN є десятки гіперпараметрів і способів їх налаштування. Хоча існують автоматичні методи, такі як випадковий пошук [137], пошук за мережею [138], градієнтна гіперпараметрична оптимізація [139] і гіперсмуга [140], деякі дії потребують «ручного» дослідження для поліпшення якості моделі.

Для покращення якості моделі є такі можливості:

- Змінити визначення проблеми (наприклад, класи, які мають бути виділені).
- Отримати більше даних для навчання.
- Підготувати дані для навчання [145]–[147].
- Змінити попереднє оброблення.
- Аугментація навчальної вибірки.
- Змінити налаштування навчання [148], [149].
- Змінити модель.

У сучасних архітектурах метод попереднього оброблення даних зазвичай не змінюється. Однак залишаються різні способи поліпшення класифікатора. Зміна навчального налаштування і моделі має забагато можливих варіантів, щоб повністю їх навчити. Таким чином, технології необхідні для того, щоб спрямувати до змін для поліпшення моделі.

2.6.1 Матриця помилок.

Матриця помилок (англ. confusion matrix) – це матриця $(C_{ij}) \in \mathbb{N}_{\geq 0}^{k \times k}$, де $k \in \mathbb{N}_{\gg 2}$ – кількість класів, що містять всі правильні і неправильні класифікації. Елемент C_{ij} – кількість разів, коли елементи класу i класифікувалися як клас j . Це означає, що правильна класифікація

розміщена по діагоналі C_{ij} , а всі неправильні класифікації – не по діагоналі.

Сума $\sum_{i=1}^k \sum_{j=1}^k C_{ij}$ – це загальна кількість оцінених зразків, $\frac{\sum_{i=1}^k C_{ij}}{\sum_{i=1}^k \sum_{j=1}^k C_{ij}}$ – це точність.

Суми $r(i) = \sum_{j=1}^k C_{ij}$, кожного класу i варто досліджувати, оскільки вони показують, що класи перекошені. Якщо кількість прикладів одного класу домінує в наборі даних, то класифікатор може отримати високу точність. Якщо точність класифікатора наближена до теоретичної ймовірності найбільш поширеного класу, можуть допомогти методи роботи з перекошеними класами.

Автоматичний критерій для перевірки цієї проблеми:

$$accuracy \leq \frac{\max(\{r(i) | i = 1, \dots, k\})}{\sum_{i=1}^k r(i)} + \varepsilon$$

де ε – невелике значення, щоб компенсувати той факт, що деякі приклади можуть бути випадково правильними.

Інше значення, яке треба перевіряти, – це класова чутливість:

$$s(k) = \frac{\# \text{ коректно ідентифіковані випадки класу } k}{\# \text{ випадки класу } k} = \frac{C_{kk}}{r(k)} \in [0,1]$$

Якщо $s(i)$ істотно менше, ніж $s(j)$, то це показник того, що для $s(i)$ необхідно більше навчальних даних.

Класова помилка

$$f_{\text{помилка}}(k_1, k_2) = \frac{C_{k_1 k_2}}{\sum_{j=1}^k C_{k_1 j}}$$

вказує на те, чи часто клас k_1 класифікується як клас k_2 . Найбільші значення можуть вказувати на необхідність злиття двох класів, або ж спеціалізована модель для поділу цих класів може поліпшити загальну систему.

2.6.2 Криві валідації: точність, втрати та інші показники .

Криві валідації відображають гіперпараметр (наприклад, тренувальна епоха) на горизонтальній осі і метрику якості на вертикальній осі. Точність, похибка = (1 - точність) або втрати є типовими метриками якості. Інші метрики можна знайти у [141], наприклад, чутливість (англ. sensitivity або recall).

У разі, якщо як досліджуваний гіперпараметр використовується кількість тренувальних епох, криві валідації дають показник про те, що якщо тренування триваліше, воно покращує роботу моделі.

Із побудови графіка помилки на навчальній вибірці, а також помилки на валідаційній вибірці, також можна оцінити, чи може перенавчання стати проблемою. Приклад цього зображено на рис. 2.10.

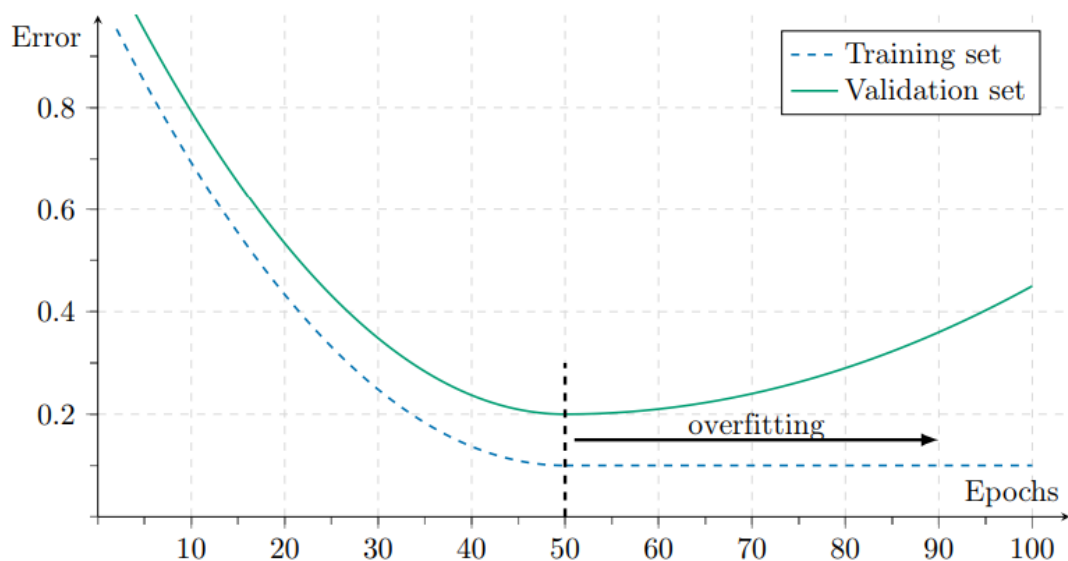


Рис. 2.10. Типова крива валідації (англ.validation curve)

У цьому випадку гіперпараметр – це число епох, а метрика якості – похибка (1 - точність). Чим довше навчалась мережа, тим краще вона потрапляє на тренувальний майданчик. У якийсь момент мережа занадто добре орієнтується в даних тренування і втрачає здатність до узагальнення. У цей момент крива якості набору для навчання і набору для перевірки розходяться. Коли класифікатор все ще удосконалюється щодо навчальної вибірки, валідація і тестова вибірка стають гіршими.

Коли на кривій валідації втрат за епоху є плато, як на рис. 2.10, це означає, що процес оптимізації не покращився за декілька епох. Три можливі способи зменшити проблему плато:

- 1) зміна вагової ініціалізації, якщо плато перебувало на початку;
- 2) регуляризація моделі;
- 3) зміна алгоритму оптимізації.

Функція втрат. Функція втрат (також зветься функцією помилки) – це функція, яка присвоює реальне значення складної події, такої як прогнозований клас ознакового вектора. Вона використовується для визначення об'єктної функції. Для класифікаційних задач функція втрат, як правило, є крос-ентропією з регуляризації l_1 або l_2 , як це описано у [142]:

$$E_{CE}(w) = - \sum_{x \in X} \sum_{k=1}^K [t_k^x \log(o_k^x) + (1 - t_k^x) \log(1 - o_k^x)] + \lambda_1 \cdot \sum_{w \in W} |w| + \lambda_2 \cdot \sum_{w \in W} w^2$$

де W – ваги, X – навчальний набір даних, $K \in \mathbb{N} \geq 0$ – число класів, t_k^x указує, чи належить навчальна вибірка x до класу k , o_k^x – вихід алгоритму

класифікації, який залежить від ваг, $\lambda_1, \lambda_2 \in [0, \infty)$ – це ваги регуляризації, які зазвичай менші за 0,1.

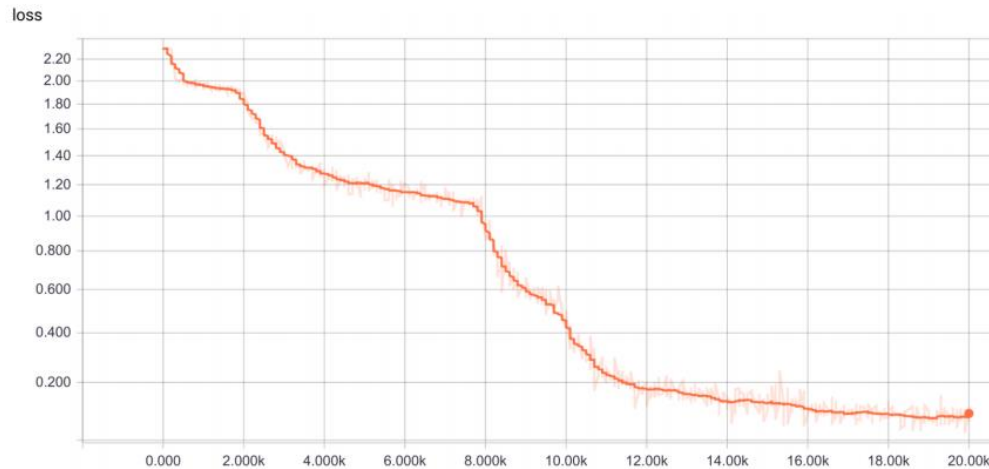


Рис. 2.11. Приклад кривої валідації (побудована функція втрат) із плато. Крива темно-оранжевого кольору згладжена, а світло-помаранчевого незгладжена

У кривій валідаційних втрат є три умови, які можуть поліпшити мережу:

- 1) якщо втрати не зменшуються протягом декількох епох, то швидкість навчання може бути занадто повільною. Процес оптимізації також може застрягти в локальному мінімумі;
- 2) втрата NaN може зумовлюватись надто високою швидкістю навчання. Інша причина – ділення на нуль або прийняття логарифма нуля. В обох випадках додавання маленької константи типу 10^{-7} вирішує проблему;
- 3) якщо крива валідаційних втрат має плато на початку, то ініціалізація ваг може бути поганою.

Критерії якості. Існує кілька критеріїв якості для моделі визначення задач класифікації. Більшість критеріїв якості засновані на матриці

помилку C , яка позначає в C_{ij} кількість разів, коли істинний клас i і j спрогнозовано. Це означає, що діагональ містить кількість правильних прогнозів. Далі нехай $t_i = \sum_{j=1}^k C_{ij}$ буде кількістю навчальних вибірок для класу i . Найпоширеніший критерій якості – точність:

$$accuracy(c) = \frac{\sum_{i=1}^k c_{ii}}{\sum_{i=1}^k t_i} \in [0,1]$$

Однією з проблем точності як критерію якості є перекошені класи. Якщо один клас є більш поширеним, ніж інші, то найпростіший спосіб отримати високу точність – це завжди класифікувати все як найпоширеніший клас.

Щоб вирішити цю проблему, можна використовувати середню точність:

$$mean - accuracy(c) = \frac{1}{k} \cdot \sum_{i=1}^k \frac{c_{ii}}{t_i} \in [0,1]$$

Для двокласових задач існують інші метрики, такі як precision, recall і F1-score. Критерії якості семантичної сегментації пояснено у [143].

Існують також інші методи аналізу згорткових мереж для задач діагностики раку молочної залози, які описано у [144], такі як криві навчання, метод argmax, реконструкції карт ознак і т.ін.

2.7 Алгоритм градієнтного спуску для навчання згорткових мереж

У згорткових нейронних мережах є два поширення: пряме і зворотне. Через ініціалізацію зваженої функції спочатку у процесі навчання CNN не доведено, чи задовольняє зважена функція запит точності структури. Функцію треба виправляти знову і знову. Метод зворотного поширення помилки призначений для передавання виправлення помилки з верхніх шарів у нижні шари. Потім нижні шари можуть підтримувати зважені функції шляхом виправлення помилок. За допомогою методу градієнтного спуску можна знайти виправлені помилки.

Метод градієнтного спуску (англ. gradient descent, GD) є одним із найбільш поширених алгоритмів оптимізації в галузі машинного навчання. Як відомо, більшість математичних моделей мають помилки через оцінювальні фактори. Є поняття, що називається функцією втрат, для опису помилки між зразком і виходом функції. Можна припустити, що математична модель є рівнянням лінійної регресії і надати їй гіпотетичну функцію $h_{\theta}(x) = \theta_0 + \theta_1 x$, де θ_0 і θ_1 – фактори рівняння. Його зразок (x_i, y_i) ($i = 1, 2, \dots, n$), де кожний x_i відповідає y_i . Тоді функція втрати гіпотези функції має бути

$$J(\theta_0, \theta_1) = \sum_{i=1}^m (h_{\theta}(x_i) - y_i)^2$$

Метод градієнтного спуску спрямований на оптимізацію математичної моделі і функції втрат, що означає зміну коефіцієнтів вихідних функцій і зменшення значення функції втрат.

Тепер можна ускладнити гіпотезу, наприклад: $h_{\theta}(x_1, x_2, \dots, x_n) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$ [$\theta_i = (i = 1, 2, \dots, n)$], можна додати $x_0 = 1$, а потім спростити її як

$$h_{\theta}(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \theta_i x_i$$

Тоді функція втрати набуде вигляду:

$$J(\theta_0, \theta_1, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x_0, x_1, \dots, x_n) - y_i)^2$$

Оскільки необхідно зменшити значення функції втрат, спочатку потрібно зробити її мікрокосмічною для аналізу змін функцій. Часткові похідні є вдалим методом. Тут слід згадати деякі поняття для ініціалізації методу градієнтного спуску.

Градієнт, який вимірює тенденцію цільової функції, розглядається як градієнт часткового похідного результату функції втрат, а рівняння має вигляд $\frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_n)$. Відстань між рекурсивними терміналами ϵ , як правило, є точністю функції і є різницею між виходом математичної моделі і зразком. Коли різниця менша або дорівнює відстані до рекурсивного терміналу, зважені функції в моделі відповідають запиту процесу. Тоді процес може зупинитися. За розміру кроку α фактор контролює, скільки відсотків градієнта використовується для відновлення нових θ_i . Вираз функції поновлення $\theta_{iNew} = \theta_{iOld} - \alpha \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_n)$. Якщо цільова функція є опуклою, то її можна оптимізувати. Розмір кроку також є важливим елементом програмування. Коли розмір кроку занадто малий, процес може обчислити найбільш оптимальний варіант, однак швидкість

збіжності істотно мала. Великий розмір кроку може суттєво поліпшити швидкість збіжності, але не може гарантувати отримання оптимального варіанту, тому що один великий розмір кроку іноді покриває маленький простір конвергенції. Тоді це перешкоджає визначенню оптимальної точки в цьому просторі конвергенції. На рис. 2.12 проілюстровано порівняння ситуацій конвергенції функцій втрат із різними розмірами кроку.

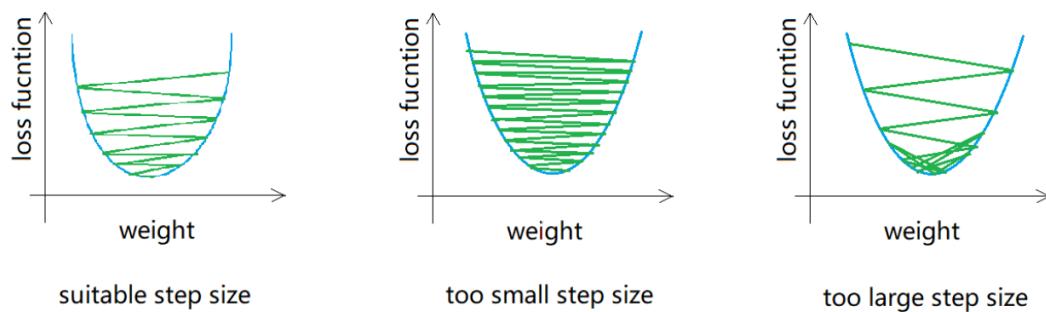


Рис. 2.12. Порівняння різного розміру кроку

Тому загальний алгоритм градієнтного спуску такий:

- 1) ініціалізувати $\theta_0, \theta_1, \dots, \theta_n$ у цільовій функції відстань до рекурсивного терміналу ε , розмір кроку α ;
- 2) обчислити градієнт, часткову похідну функції втрати функції прицілювання $\frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_n)$;
- 3) якщо $J(\theta_0, \theta_1, \dots, \theta_n) \leq \varepsilon$, зупинити алгоритм, в іншому випадку продовжити процес;
- 4) використовувати градієнт, що множиться на розмір кроку α ;
- 5) оновити всі θ по $\theta_i = \theta_i - \alpha \frac{\partial}{\partial \theta_i} J(\theta_0, \theta_1, \dots, \theta_n)$, а потім повернутися до другого етапу.

Метод зворотного поширення помилки. Крім оновлення вагових функцій, виправлені помилки треба пропустити через шари в CNN. Це поєднання алгоритму градієнтного спуску, правила ланцюжка і методу зворотного поширення помилки.

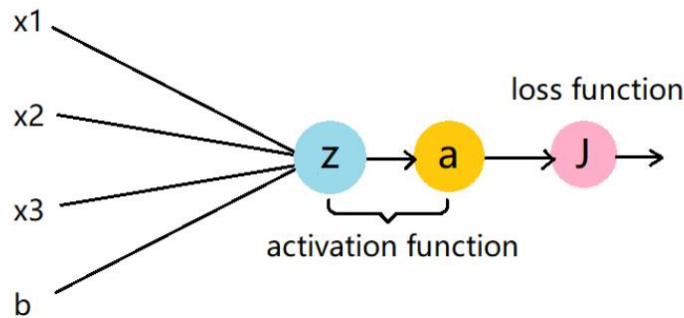


Рис. 2.13. Приклад зворотного поширення помилки в одиничному обчислювальному блоці

Рис. 2.13 є діаграмою однієї нейронної клітини. Можна припустити, що навчальний зразок є (x, y) , x – вхід, який проходить через функцію активації $h_{w,b}(x)$. Мережа отримує вихід «а», потім результат «а» проходить через функцію втрат, виходом функції втрат є вартість J .

Вираз функції активації або гіпотези має вигляд:

$$h_{w,b}(x) = a = \text{sigmoid}(z) = \text{sigmoid}(\sum_{i=0}^n (x_i w_i + b)) \quad (2.3)$$

Рівняння функції втрат:

$$J(W, b, x, y) = \frac{1}{2} \|y - h_{wb}(x)\|^2$$

У функції активації (рівняння (2.3)) сигмовидна функція $\sigma(z) = \frac{1}{1+e^{-z}}$, є гіпотетичною функцією. Як функції активації можна використовувати й інші функції. Сигмоїдальна функція має властивість, потрібну для обчислення:

$$a = \sigma(z); \frac{\partial}{\partial z} = a(1 - a).$$

Під час навчання необхідно обчислити градієнти вартості J за частковими похідними. Є два градієнти: один для фактора W , другий для фактора b . Потім береться часткова похідна, що стосується W і b . Однак неможливо обчислити часткову похідну від вартості J до W і b . Правило ланцюжка допомогло б розв'язати цю задачу. Воно може відобразитися як $\frac{\partial}{\partial W} J = \frac{\partial}{\partial z} \frac{\partial}{\partial W}$. Тепер потрібно спочатку вирахувати часткову похідну від вартості J до проміжної змінної a і z :

$$\delta^{(a)} = \frac{\partial}{\partial z} J(W, b, x, y) = -(y - a)$$

$$\delta^{(a)} = \frac{\partial}{\partial z} J(W, b, x, y) = \frac{\partial J}{\partial a} \frac{\partial a}{\partial z} = \delta^{(a)} a(1 - a)$$

Потім згідно з правилом ланцюжка отримаємо градієнти коефіцієнтів W і b :

$$\nabla_W J(W, b, x, y) = \frac{\partial}{\partial W} J = \frac{\partial J}{\partial z} \frac{\partial z}{\partial W} = \delta^{(z)} X^T$$

$$\nabla_b J(W, b, x, y) = \frac{\partial}{\partial b} J = \frac{\partial J}{\partial z} \frac{\partial z}{\partial b} = \delta^{(z)}$$

Під час описаних процесів спочатку береться $\frac{\partial J}{\partial a}$, потім обчислюється $\frac{\partial J}{\partial z}$, далі отримуємо $\frac{\partial}{\partial w}$ і $\frac{\partial}{\partial b}$. Таким чином, можна визначити, що процедура, яка поширює збільшення ∂J функції втрат зі зворотного боку, називається зворотним поширенням.

2.8 Висновки до розділу

1. Оглянуто основні положення теорії штучних нейронних мереж, теорії згорткових мереж і алгоритмів навчання мереж.
2. Розглянуто види згорткових мереж, таких як Inception-v3 і DenseNet, і методи, які існують для навчання даних цих мереж.
3. Встановлено різні критерії для аналізу мережі для задач класифікації раку молочної залози, одним з яких є аналіз матриці помилок. Така матриця показує, на скільки правильно модель прогнозує існуючі класи (рак/не рак).
4. Розглянуто параметри, які використовуються для перевірки моделі, зокрема точність, чутливість. Для задач класифікації раку більш правильним є врахування саме їх, оскільки з їх допомогою можна визначити кількість правильно класифікованих класів рак/не рак.

РОЗДІЛ 3

РОЗРОБЛЕНІ МОДЕЛІ ТА МЕТОДИ РОЗПІЗНАВАННЯ РАКУ НА ОСНОВІ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

3.1 Оброблення вхідних даних для згорткових мереж

Попереднє оброблення має велике значення у машинному навчанні. Однак за наявності наборів даних із сотнями прикладів на клас і можливості CNN самостійно вивчати ознаки більшість моделей покладаються на необроблені значення пікселів. Єдиним поширеним методом попереднього оброблення є нормалізація розмірів. Для отримання фіксованого вхідного розміру CNN можна застосовувати таку процедуру:

- 1) брати одну або кілька ділянок зображення з потрібним співвідношенням сторін;
- 2) змінити масштаб частини зображення (кадру) до потрібного розміру.
- 3) для навчання всі кадри можна використовувати незалежно один від одного;
- 4) у тестуванні всі кадри можна пропускати через мережу, а розподіл імовірностей на виході можна звести воедино, наприклад, шляхом усереднення.

Інші методи попереднього оброблення:

1. Перетворення колірного простору (RGB, HSV і т.ін.).
2. Середнє віднімання.
3. Стандартизація значень пікселів до $[0, 1]$ шляхом ділення на 255 (застосовується у [136]).
4. Скорочення розмірів:

- а) аналіз основних компонентів (англ. principal component analysis, PCA);
- б) автоенкодер для скорочення кількості ознак;
- в) лінійний дискримінантний аналіз (англ. linear discriminant analysis, LDA).

5. Декорелювальне перетворення за методом аналізу нульових компонентів (англ. zero-phase component analysis, ZCA) (застосовується у [150]).

3.1.1 Нормалізація (оброблення зображень).

В обробленні зображень нормалізація – це процес, який змінює діапазон значень інтенсивності пікселів. Як додатки можна використовувати фотографії зі слабким контрастом. Нормалізацію іноді називають контрастним розтягуванням або розтягуванням гистограми. У більш загальних галузях оброблення даних, таких як цифрове оброблення сигналів, її називають розширенням динамічного діапазону [151].

Нормалізація перетворює n -мірне зображення в шкалу сірого (англ. grayscale)

$$I: \{X \subseteq \mathbb{R}^n\} \rightarrow \{Min, \dots, Max\}$$

зі значеннями інтенсивності в діапазоні (Min, \dots, Max) у нове зображення

$$I_N: \{X \subseteq \mathbb{R}^n\} \rightarrow \{newMin, \dots, newMax\}$$

зі значеннями інтенсивності в діапазоні $(newMin, \dots, newMax)$.

Лінійна нормалізація цифрового зображення в шкалу сірого виконується за формулою

$$I_N = (I - Min) \frac{newMax - newMin}{Max - Min} + newMin$$

Наприклад, якщо діапазон інтенсивності зображення становить від 50 до 180, а бажаний діапазон – від 0 до 255, то процес призводить до віднімання 50 від кожного пікселя інтенсивності, що робить діапазон від 0 до 130. Потім інтенсивність кожного пікселя множиться на 255/130, тоді діапазон набуде значень від 0 до 255.

У цій праці $Min = 0$ та $Max = 255$.

$$I_N = (I - Min) \frac{255}{Max - Min}$$

3.1.2 Перетворення колірному простору.

В обробленні зображень комп'ютерні алгоритми використовують для виконання операцій із цифровими зображеннями з метою аналізу, поліпшення, модифікації і оброблення цифрових зображень. Оброблення зображень уможлиблює використання широкого спектра алгоритмів для оброблення вхідних цифрових даних зображень і сприяє уникненню вбудованих спотворень сигналу і шумів на період оброблення зображень.

Поліпшення кольорових зображень є дуже важливим і необхідним процесом оброблення зображень. Поліпшення якості зображень відіграє важливу роль в галузі медичної візуалізації, космічної науки і т.ін. Покращені медичні зображення більше підходять для аналізу і встановлення правильного діагнозу.

Далі показано різні колірні простори або колірні моделі, які використовують для подання цифрових зображень, такі як HIS, YIQ, RGB, YCbCr та ін. Також показано, як перетворити шумове цифрове RGB

зображення в інший колірний простір, такий як HIS, YIQ, і YCbCr відповідно, і після цього фільтрує ці зображення один за одним, перетворюючи в RGB колірний простір, а не обчислює PSNR, який показує пікове відношення шуму сигналу для різних колірних просторів.

Колірний простір RGB

У колірній просторовій моделі RGB (red, green, blue – червоний, зелений, синій) використовується декартова координатна схема. У ній осями є три основні кольори світла (червоний, зелений і синій), зазвичай нормалізовані до діапазону (рис. 3.1). В отриманому кубі вісім вершин відповідають трьом основним кольорам світла і трьом вторинним кольорам (блакитному, жовтому, пурпурному), чистому білому і чистому чорному. Значення червоного, зеленого і синього кольорів для кожної з восьми вершин зображено на рис. 3.1.

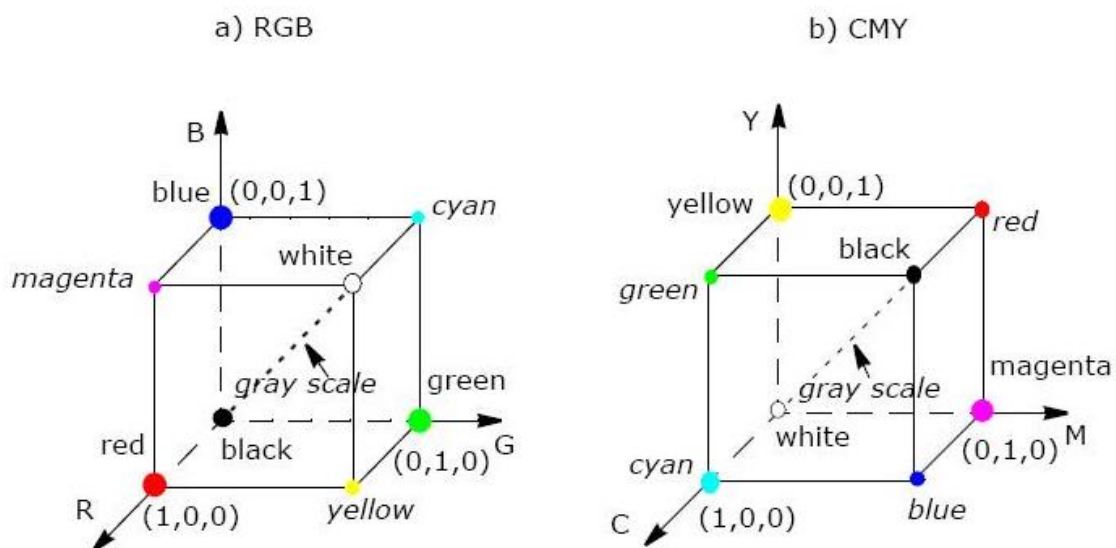


Рис. 3.1. Куб RGB–CMY

Колірні моделі CMY і CMYK

У моделі CMY використовуються три основні кольори пігментів. Це блакитний, пурпурний і жовтий. Переконвертація з RGB на CMY модель має вигляд

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Колірна модель HSV

Колірна модель HSV (Hue, Saturation, Value – Відтінок, Насиченість, Значення) досягається шляхом перетину на колірному кубі RGB збоку його головної діагональної і сірої осі, що призводить до появи шестигранної кольорової палітри [152].

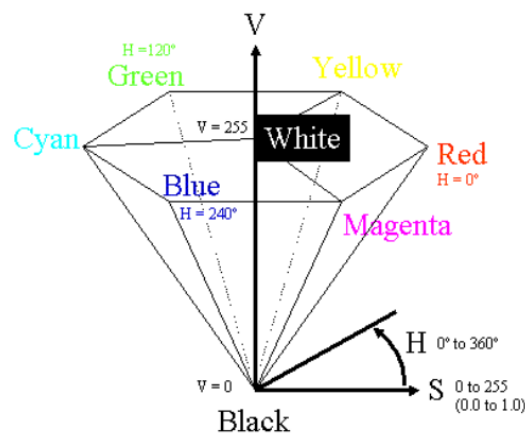


Рис. 3.2 Схема HSV

3.1.3. Декорелювальне перетворення за методом аналізу нульових компонентів (ZCA)

Нехай x – деяка випадкова величина [153].

$$x \sim F(\mu, \Sigma),$$

$$\mu = E[x], \Sigma = cov(x, x)$$

де Σ – коваріаційна матриця.

Наступне перетворення називається декорелювальним:

$$z = \mu, \Sigma^{-\frac{1}{2}} (x - \mu)$$

Його властивості:

$$E[z] = 0,$$

$$cov(z, z) = I$$

У праці використано три набори даних: BreaKHis, Breast histology і Kaggle [154]–[156]. У наборі даних BreaKHis 5500 зображень використано для навчання і 1500 зображень – для тестування; у Breast histology 200 зображень – для тренування і 37 – для тестування. Набір даних в Kaggle числовий, 500 даних використано для навчання і 70 – для тестування.

На рис. 3.3 показано, що невеликий набір даних може вплинути на точність, і згорткові нейронні мережі не можуть бути добре навчені, що призводить до перенавчання (англ. *overfitting*) або недонавчання (англ. *underfitting*).

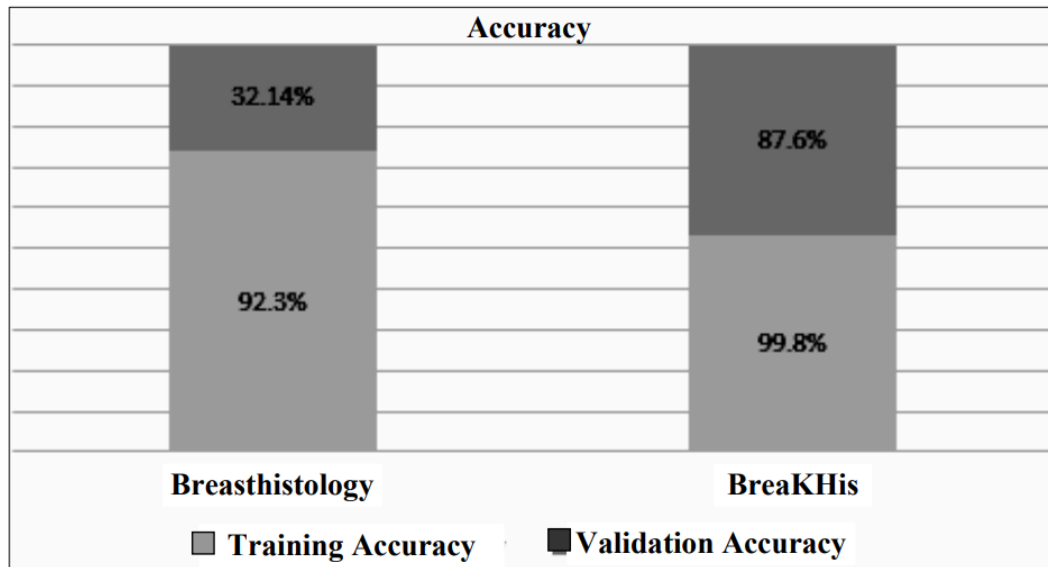


Рис. 3.3. Порівняння точності для навчальних та тестувальних наборів

Згідно з рис. 3.3 точність для валідаційних даних, установлена в наборі даних Breast histology, значно менша, ніж у наборі даних BreaKHis. Як згадувалось раніше, кількість мамографічних сканів у BreaKHis для навчання моделі було 5500, коли для навчання моделі за допомогою датасету Breast histology використано 200 сканів. У цьому випадку модель зазнає перенавчання. Тобто модель добре працює на навчальній вибірці, але не з новими для неї даними (валідаційний набір). Таким чином, використання великого набору даних запобігає перенавчанню моделі під час навчання і точного налаштування.

За даними Всесвітньої організації охорони здоров'я [157] мамографія економічно виправдана процедура для встановлення наявності раку молочної залози у пацієнтів. Вартість мамографії дорога, але вона є одним із найкращих методів для скринінгу раку грудей, який довів свою ефективність.

3.2. Застосування згорткових мереж Inception-v3, DenseNet до задачі діагностики раку молочної залози

У навчанні згорткової мережі необхідно навчити мережі так, щоб не було перенавчання.

Є два способи запобігти перенавчанню моделі:

- 1) навчати мережу на більшій кількості прикладів;
- 2) змінити складність мережі.

Перевагою глибоких нейронних мереж є те, що їхня ефективність поліпшується залежно від того, як вони отримують більше наборів даних. Також потрібно враховувати, що можна збільшити набори даних тільки до певної кількості, – і продуктивність мережі збільшиться відповідно. Потім робота мережі не буде змінюватись, і в такому випадку потрібно зупинити додавання навчальної вибірки, оскільки модель може зазнати перенавчання.

Пропускна здатність нейромережевої моделі, її складність визначаються як її структурою в аспекті вузлів і шарів, так і параметрами в аспекті ваг. Таким чином, можна зменшити складність нейромережі для запобігання перенавчання одним із двох способів:

- 1) змінити її структуру (кількість ваг);
- 2) змінити параметри мережі (значення ваг).

Далі наведено перелік найпоширеніших додаткових методів регуляризації:

1. Виключення (Dropout): можливість видалити входи під час тренування.
2. Шум (Noise): додати статистичний шум на входи під час навчання.

3. Рання зупинка (Early stopping): моніторинг якості моделі на валідаційному наборі і зупинка тренування за зниження продуктивності.

Виключення

Під час навчання нейронних мереж за допомогою параметрів виключення можна виключити певний процент випадкових нейронів, які перебувають як у прихованих, так і у видимих шарах різних ітерацій (epoch) навчання.

Шум

Коли модель має здатність навчатися з вхідними зображеннями і шумом, тоді вона може добре працювати і на валідаційному, і на тестовому наборах даних.

Рання зупинка

За допомогою ранньої зупинки можна зупинити навчання моделі, коли її продуктивність на валідаційних даних не змінюється протягом кількох epoch. Шляхом експерименту можна визначити скільки потрібно часу, щоб зупинити навчання після певної кількості epoch, коли точність на валідаційних даних не змінюється.

Є два види навчання згорткової мережі:

1. Навчання з нуля (англ. learning from scratch);
2. Трансферне навчання (англ. transfer learning).

Для навчання з нуля необхідна велика кількість набору даних для хорошого результату. Щоб мати доступ до більшої кількості наборів даних, можна використати аугментацію для їх збільшення, оскільки володіти великою кількістю наборів даних складно.

Перевага використання вже навченої моделі полягає в тому, що така модель вже навчалась на великій кількості даних мережі ImageNet. У базі даних ImageNet нараховується 14 197 122 зображень, поділених на 21 841

категорію. Таким чином, модель може добре навчатися на цій базі даних. Коли використовується модель, яка навчається на ImageNet, а оскільки у цій базі даних немає маммографічних сканів, необхідна заміна останніх шарів (повнозв'язних шарів/повністю підключених) на нову відповідну задачу класифікації раку молочної залози, де є два класи – рак/не рак.

На рис. 3.4 показано приклад архітектури згорткової мережі Inception-v3, де використано маммографічний скан, який застосовується у вхідних даних.

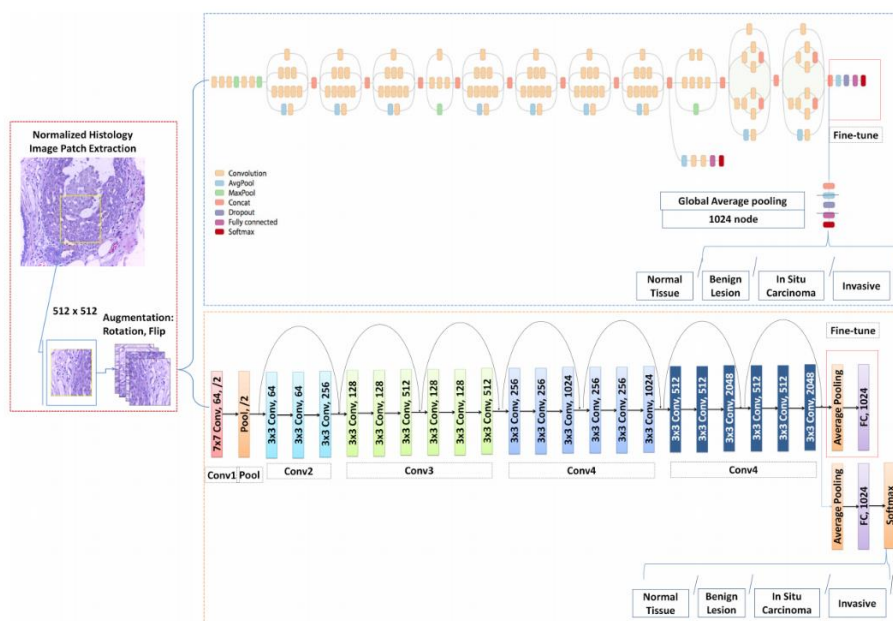


Рис. 3.4. Процес класифікації зображень гістологічних груп за допомогою відрегульованої мережевої архітектури Google Inception-v3 і ResNet50. У блоці зліва подано етапи попереднього оброблення, у блоках справа – мережеві архітектури Inception-v3 (зверху) і ResNet50 (знизу)

Модель Inception-v3 від Google – це третя версія із серії глибоких навчань згорткових архітектур. Її освоєно за допомогою набору даних 1000 класів із оригінального набору даних ImageNet, навченого на понад одному

мільйони навчальних зображень. Inception-v3 розроблено для розв'язання задач розпізнавання у мережі ImageNet, де вона посіла перше місце.

3.3 Гібридний сверточних мережу і її навчання для завдання діагностики раку молочної залози

Запропонований гібридний сверточних мережу, складається з двох моделі сверточного автоенкодера як виділ інформативних ознак, і сверточних мережі в якості класифікатора.

Для навчання запропонованого моделі, необхідно:

1) Оптимізувати сверточних мережі автоенкодера без вчителя за допомогою мінімізації функції втрат:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - Y^{\wedge}_i)^2$$

2) Також, необхідно налаштувати гіперпараметри шляхом експериментальних аналіз таким чином, що збільшити критерій якості моделі.

3) Шляхом трансформного навчання, навчити сверточних мережі DenseNet.

4) Підключити вихід енкодера до навчений сверточних мережі DenseNet для завдання класифікації.

5) Оптимізувати гібридний сверточних мережу за допомогою мінімізації Перехресних ентропія:

$$H_p = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(P(y_i)) + (1 - y_i) \cdot \log(1 - P(y_i)),$$

де,

N - кількість спостережень;

y - бінарний індикатор (0 або 1) того, чи є мітка класу правильної класифікації для спостереження;

p - прогнозована ймовірність моделі.

3.4 Нормалізація плями (англ. **Stain normalization, SN**)

Перед використанням медичних зображень їх необхідно нормалізувати. Одним із поширених способів аналізу зразка тканин є оброблення його плямами, що мають вибіркоче споріднення з різними біологічними речовинами. Більшість плям поглинають лише світло, тому зафарбовані слайди проглядаються за допомогою світлового мікроскопа, що освітлює зразок знизу. Якщо його не буде, все світло проникатиме крізь нього і ставатиме яскравим білим кольором. Зони, де пляма прилипла до речовини в тканині, поглинатимуть частину світла. Кількість поглиненого світла залежить від багатьох факторів. Пляма поглинатиме певну кількість світла у кожному спектрі. У разі мультиспектральної візуалізації цей процес може бути досить складним. У цій праці основну увагу звернено на використання стандартних 24-бітових RGB-камер для отримання зображень, тому методологію обмежено трьома довжинами хвиль світла. Пропорція кожної поглиненої довжини хвилі утворює вектор плями. Вектор плями не тільки істотно варіюється між різними плямами, а й може значно відрізнитися від тієї самої плями залежно від таких факторів, як виробник, умови зберігання до використання і спосіб нанесення [158], [159]. На рис. 3.5 зображено два приклади слайдів гістології шкіри, оброблені в різний час у різних лабораторіях. Абсолютні значення кольору слайда піддаються багатьом впливам, але лише один із них може вплинути на біологічний компонент, який потрібно отримати. Цей біологічний

компонент є фактичною кількістю клітинної речовини, до якої буде прикріплена певна пляма.

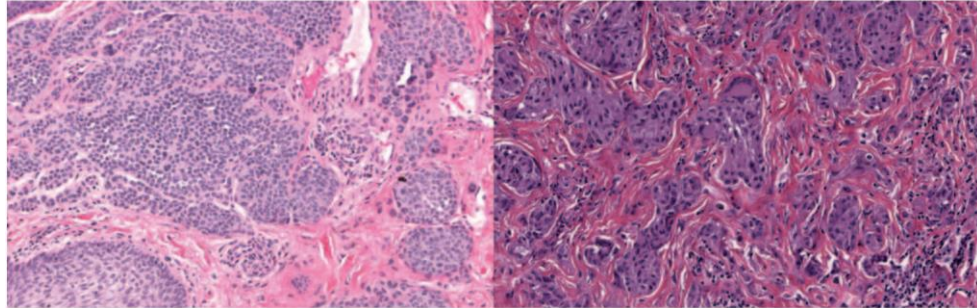


Рис. 3.5. Приклад двох гістопатологічних слайдів меланом, обидва забарвлені гематоксиліном й еозином, але з проявами, що істотно різняться

Відповідно до стандартної практики [160] усі значення кольорів, розглянуті у праці, перетворюються у відповідні значення оптичної щільності (англ. optical density, OD):

$$OD = \log_{10}(I),$$

де I – колірний вектор RGB, кожен компонент якого нормований до $[0,1]$. Це перетворення надає простір, у якому лінійна комбінація плям призведе до лінійної комбінації значень OD [160]. Взаємозв'язок між інтенсивністю й OD зображено на рис. 3.6 (а) і 3.6 (б) із використанням даних, отриманих на забарвлених гематоксиліном й еозином знімках меланом.

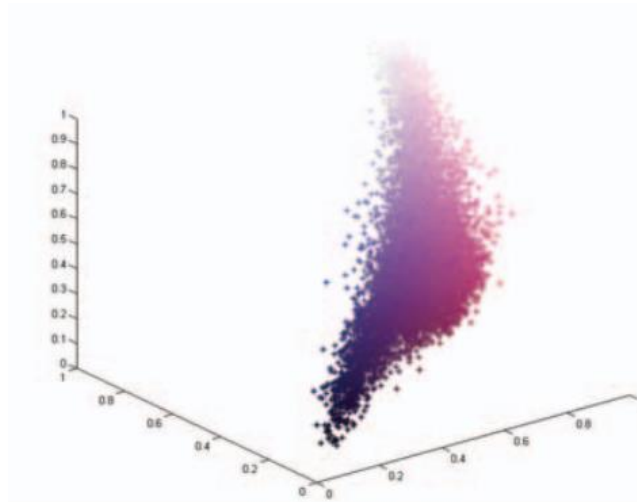


Рис. 3.6 (а). Синій і рожевий кольори здаються роздільними, але вигнутими

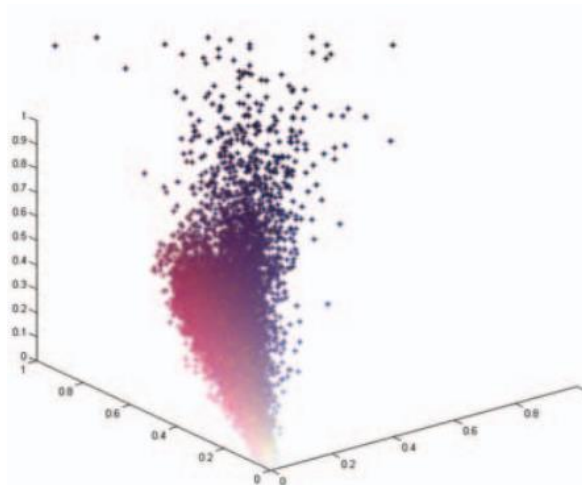


Рис. 3.6(б). Ці пікселі перетворено в простір оптичної щільності. Синій і рожевий пікселі тепер відокремлюються по прямій лінії від початку

Після визначення правильних векторів будь-яким методом використовується проста схема колірної деконволюції, аналогічно до [161], для перетворення значень кольору в кількісні величини, які цікавлять:

$$OD = VS \Rightarrow S = V^{-1}OD,$$

де OD – спостережуване значення оптичної щільності, V і S – матриці векторів плям і насиченість кожної плями відповідно.

Інтенсивність конкретної плями залежить від початкової концентрації плями, процедури фарбування, скільки вицвітання сталося з моменту першого оброблення зразка і, зрештою, від того, скільки клітинної речовини, яка цікавить, є в матеріалі [41].

3.5 Методи оптимізації згортальних нейронних мереж

Оптимізатори – це алгоритми або методи, які застосовують для зміни атрибутів нейромережі, таких як ваги і швидкість навчання, з метою зменшення втрат. Існує кілька алгоритмів оптимізації [162] нейронних мереж: Adam (англ. Adaptive moment estimation), RMSProp (англ. Root mean square propagation), SGD (англ. Stochastic gradient descent).

Adam

Adam – один з оптимізаційних алгоритмів. Він поєднує в собі ідею накопичення руху, а також ідею слабшого поновлення ваг для типових ознак.

На рис. 3.7 зображено короткий опис алгоритму Adam.

```

Require:  $\alpha$ : Stepsize
Require:  $\beta_1, \beta_2 \in [0, 1)$ : Exponential decay rates for the moment estimates
Require:  $f(\theta)$ : Stochastic objective function with parameters  $\theta$ 
Require:  $\theta_0$ : Initial parameter vector
 $m_0 \leftarrow 0$  (Initialize 1st moment vector)
 $v_0 \leftarrow 0$  (Initialize 2nd moment vector)
 $t \leftarrow 0$  (Initialize timestep)
while  $\theta_t$  not converged do
   $t \leftarrow t + 1$ 
   $g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1})$  (Get gradients w.r.t. stochastic objective at timestep  $t$ )
   $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$  (Update biased first moment estimate)
   $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$  (Update biased second raw moment estimate)
   $\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$  (Compute bias-corrected first moment estimate)
   $\hat{v}_t \leftarrow v_t / (1 - \beta_2^t)$  (Compute bias-corrected second raw moment estimate)
   $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$  (Update parameters)
end while
return  $\theta_t$  (Resulting parameters)

```

Рис. 3.7. Алгоритм Adam

Для реалізації алгоритму необхідно використати підхід експоненціально затухального ковзного (рухомого) середнього для градієнтів цільової функції і їх квадратів:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t,$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2,$$

де m_t – оцінка перших кроків (середнє градієнтів); v_t – оцінка другого моменту (середня нецентрована дисперсія градієнтів).

Однак у такому випадку є проблема тривалого накопичення m_t та v_t на початку роботи алгоритму, особливо коли коефіцієнти збереження β_1 і β_2 наближені до 1.

Щоб ліквідувати цю проблему і не вводити нові гіперпараметри, оцінки першого і другого моментів зазнають незначних видозмін:

$$m_t^{\wedge} = \frac{m_t}{1 - \beta_1^t}$$

$$v_t^{\wedge} = \frac{v_t}{1 - \beta_2^t}$$

Тоді вираз для поновлення параметрів:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t^{\wedge}} + \varepsilon} m_t^{\wedge}$$

RMSProp

RMSProp – це алгоритм оптимізації, розроблений паралельно з AdaDelta і є його складовою. Обидва алгоритми створено для вирішення основної проблеми AdaGrad: неконтрольного накопичення квадратів градієнтів, яке зрештою призводило до паралічу процесу навчання.

Ідея RMSProp така: замість повної суми оновлень G_t використовуватиметься усереднений за історією квадрат градієнта. Метод нагадує принцип, який використовується в MomentumSGD, – метод експоненціально затухального ковзного середнього.

Уведемо позначення $E[g^2]_t$ – ковзне середнє квадрата градієнта у момент часу t . Формула для його обчислення:

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1 - \gamma)g^2_t$$

Тоді, підставивши $E[g^2]_t$ замість G_t у формулу поновлення параметрів для AdaGrad, отримаємо (матричні операції пропущені для спрощення):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \varepsilon}} g_t$$

Знаменник – корінь із середнього квадрата градієнта, тобто root mean square, RMS:

$$RMS[g]_t = \sqrt{E[g^2]_t + \varepsilon}$$

SGD

Цей модуль реалізовує принцип роботи стохастичного градієнтного спуску (англ. stochastic gradient descent, SGD).

Нехай є деяка модель, що визначається набором параметрів θ (у випадку нейронних мереж – ваги моделі), і навчальна вибірка, що складається з $l + 1$ пар «об’єкт–відповідь» $(x_0, y_0), \dots, (x_l, y_l)$. Також необхідно визначити цільову функцію (функція втрат) у загальному вигляді – $J(\theta)$ або $J(\theta)_i$ для i -го об’єкта вибірки.

Тоді завдання оптимізації у такому випадку формулюється як:

$$\sum_{i=0}^l J(\theta) \rightarrow \min \theta$$

а процес оптимізації параметрів (навчання ваг):

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{1}{l} \sum_{i=0}^l \Delta_{\theta} J_i(\theta_t),$$

де θ_{t+1} – оновлений набір параметрів для наступного кроку оптимізації; θ_t – набір параметрів на поточному кроці; η – швидкість навчання; $\Delta_{\theta} J_i(\theta_t)$ – градієнт функції помилки для i -го об’єкта навчальної вибірки.

У підсумку такий підхід в опуклих функціях помилок завжди призводить до їх глобального мінімуму, а в неопуклих – до локального. Однак, як зазначалось раніше, для виконання лише однієї операції поновлення параметрів доводиться застосовувати градієнт для всієї навчальної вибірки, що може бути досить часо- і ресурсозатратним.

Стохастичний градієнтний спуск існує для того, щоб ліквідувати проблеми такого типу, тому що в такій модифікації градієнтний спуск

застосовується або до випадкових об'єктів вибірки (тобто оптимізація параметрів відбувається після розрахунку градієнта для одного об'єкта), або до мінібатчу – невеликих наборів об'єктів з навчальної вибірки.

Нехай n – розмір мінібатчу, тоді процес оптимізації параметрів матиме такий вигляд:

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{1}{n} \sum_{i=0}^{j+n} \Delta_{\theta} J_i(\theta_t),$$

де j – індекс випадково взятого елемента, причому $j \leq l - n$.

3.6 Критерії для аналізу мережі для класифікації типів ракових захворювань

Для визначення ефективності моделі необхідно врахувати наступні критерії:

- 1) Чутливість моделі:

$$Recall: RE = \frac{tp}{tp + FN}$$

- 2) Точність (англ. Precision):

$$Precision: PR = \frac{TP}{TP + FP}$$

- 3) F1-Score:

$$F1_Score = 2 * \frac{RE * PR}{RE + PR}$$

4) Точність (англ. Accuracy):

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Де:

TP - Істинно-позитивні (англ. True positives)

FP - Помилково-позитивні (англ. False positives)

FN - Помилково-негативні (англ. False negatives)

TN - Істинно-негативні (англ. True negatives)

3.7 Висновки до розділу

1. Розглянуто різні методи для розпізнавання раку і методи для оброблення медичних зображень, такі як нормалізація і декорелювання.
2. Розглянуто згорткові мережі Inception-v3, DenseNet й автоенкодер для розпізнавання раку молочної залози і методи навчання цих згорткових мереж.
3. Показано, як різні набори даних впливають на ефективність моделі. Порівняно набори даних BreaKHis і Breast histology за використання модифікованої згорткової мережі Inception-v3. Результат дослідження показав, що за використання набору даних BreaKHis точність моделі на 7,5% більша, ніж за використання набору даних Breast histology, оскільки кількість медичних зображень у Breast histology обмежена.
4. Розглянуто основні методи для запобігання перенавчання, зокрема застосування Dropout і Early stopping у навчанні згорткових нейронних мереж.
5. Розглянуто різні алгоритми для оптимізації згорткових нейронних мереж, такі як ADAM, RMSProp і SGD.

РОЗДІЛ 4.

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ РОЗРОБЛЕНИХ МЕТОДІВ РОЗПІЗНАВАЇННЯ РАКОВИХ ЗАХВОРЮВАНЬ

4.1 Експериментальні дослідження згорткових мереж Inception-v3, DenseNet

У запропонованій програмі класифікації раку молочної залози використано модифіковану модель Inception-v3, щоб отримати кращі результати порівняно з попередніми працями. Для цього шляхом експериментального аналізу обрано кращі значення гіперпараметрів, що мають істотне значення для моделі. Також перед повнозв'язним шаром додано відсів із $rate = 0,2$, щоб зробити модель більш незалежною від навчальної вибірки. Крім цього, у модифікованій моделі використано два повнозв'язні шари.

Inception-v3 навчалась двома методами: з нуля і за допомогою трансферного навчання. Як видно з рис. 4.1, навчання з нуля дало гірший результат порівняно з використанням точного налаштування моделі. Причина такої поведінки моделі – обмежена кількість навчальної вибірки.

Також потрібно зазначити, що не завжди великий обсяг набору даних для навчання є запорукою кращого результату моделі. Згорткові мережі мають властивість: після певної кількості навчальних вибірок для навчання продуктивність моделі більше не буде змінюватись.

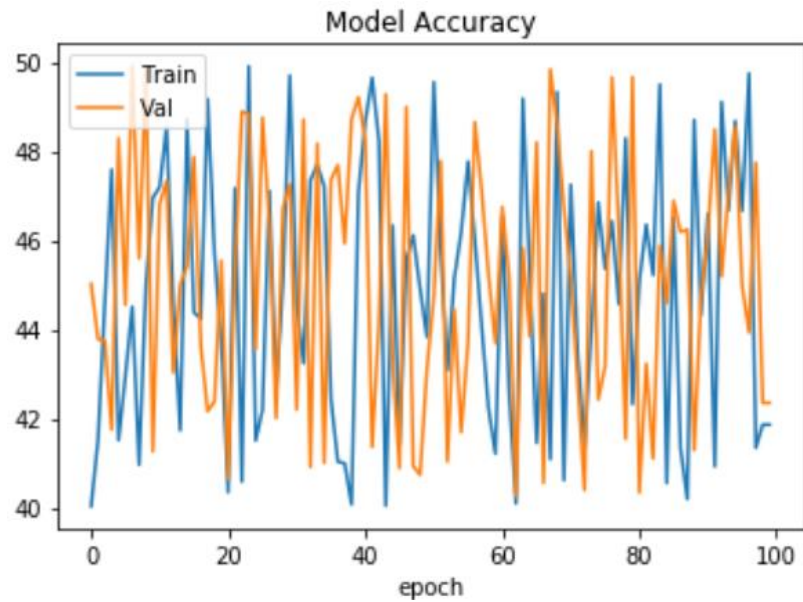


Рис. 4.1 (а). Точність моделі за навчання з нуля за допомогою 5000
мамографічних сканів

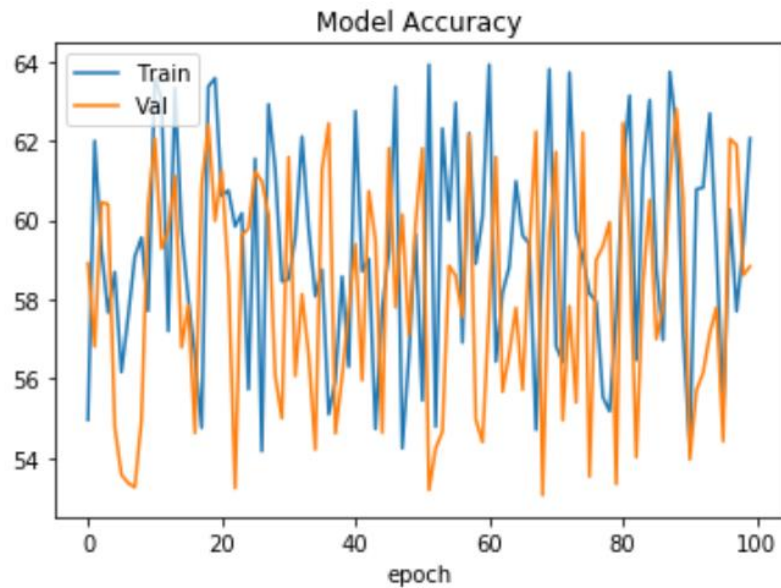


Рис. 4.1 (б). Точність моделі за навчання за допомогою точного
налаштування

Під час трансферного навчання останній шар із попередньої моделі замінюється двома новими шарами, де модель буде навчатися з новим

набором даних. Також нова модель має шар класифікації: рак/не рак молочної залози.

Після того, як експериментом доведено (рис. 4.1), щоб навчати моделі попереднього навчання за допомогою трансферного навчання, необхідно покращувати модель за допомогою вибору кращих гіперпараметрів, таких як, кількість епохи, швидкість навчання, кількість К під час перехресної перевірки, функція оптимізація та ін.

Після попереднього оброблення даних і визначення кращого методу навчання моделі (з нуля/трансферне навчання) необхідно поділити дані на навчальні та валідаційні вибірки. Валідаційні вибірки допомагають моделі бути добре навченою на тих даних, які ніколи їй не траплялись. Є два способи поділу даних на навчання і набори валідації:

- 1) проста валідація;
- 2) К-фолдова перехресна валідація.

Проста валідація ділить дані на два набори, у яких одна частина даних використовується для навчання, а інша – для валідації, коли К-фолдова перехресна валідація ділить набір даних на К частини (англ. K-fold). На першій ітерації перша частина використовується для валідації, (К-1) частини використовуються для навчання. На другій ітерації друга частина використовується для валідації, решта – для навчання. Цей процес повторюється доти, доки кожна папка не використовуватиметься як набір перевірки.

У табл. 4.1 наведено модель, де чутливість з К-fold валідацією краща, ніж із простою валідацією, – 70% і 31,3% відповідно.

Таблиця 4.1 (а)

Якість моделі Inception-v3 за простої валідації (показники точності)

Проста валідація		
Precision	Recall	F1-score
61,59%	64,3%	63,11%

Таблиця 4.1 (б)

Якість моделі Inception-v3 за 3-fold валідації

3-fold валідація		
Precision	Recall	F1-score
65,70%	70,00%	67,78%

Згідно з даними табл. 4.1 (б) продуктивність моделі за 3-fold валідації дає кращий результат порівняно з простою валідацією. Причина кращого результату моделі за K-fold валідації полягає в тому, що під час навчання дані не губляться і модель може навчатися на різних даних.

За використання K-fold валідації необхідно визначити число k . Експериментальний аналіз показує, що модель працює краще, коли

$$k = 5.$$

У табл. 4.2 подано Якість Inception-v3 за різних значень k . На основі експериментального аналізу за $k = 5$ чутливість моделі дає кращий результат – 85,84%. Чутливість є найважливішим фактором, який треба враховувати у навчанні глибоких нейронних мереж для задач розпізнавання раку молочної залози. За допомогою чутливості можна оцінити, чи визначає мережа рак як ракову пухлину. Оскільки у розпізнаванні важливіше правильно визначити рак, ніж не визначити рак як

рак. На жаль, у багатьох працях здебільшого розглядається ассурасу (точність) моделі, у цій праці розглядається recall, precision (чутливість, влучність).

Таблиця 4.2

Якість моделі Inception-v3 за різних чисел k (чутливість)

Recall	Number of k
65,35%	$k = 2$
70,00%	$k = 3$
72,25%	$k = 4$
85,84%	$k = 5$
78,45%	$k = 6$
76,60%	$k = 9$
76,25%	$k = 10$

На основі даних із табл. 4.2 чутливість моделі після $k = 5$ починає зменшуватись. Крім цього, K-fold валідація не лише допомагає розділити набір даних для навчання та валідації, а й налаштувати інші гіперпараметри. Таким чином, продуктивність моделі збільшується.

Після того, як визначено кількість k частин, необхідно визначити кількість епохи. У табл. 4.3 показано, як точність моделі змінюється за різних значень епохи.

Таблиця 4.3

Якість моделі Inception-v3 за різних епох

	Epoch = 50	Epoch = 150	Epoch =250	Epoch =350
Precision	64,15%	64,6%	66,66%	69,45%
Recall	71,85%	79,1%	85,84%	81,6%
F1-Score	67,78%	71,2%	74,99%	75,05%

Як видно з даних табл. 4.3, чутливість моделі за збільшення кількості епохи до 250 покращується, потім поступово погіршується. Це свідчить про те, що модель починає перенавчатись. За значення епохи 350 точність моделі краща, ніж за епохи 250, 69,45% і 66,66% відповідно, але оскільки чутливість є головним фактором у розпізнаванні раку молочної залози, програма буде навчатися за значень епохи 250.

4.2 Застосування згорткової мережі DenseNet до задачі розпізнання раку молочної залози

У праці також виконано модифікацію згорткової мережі DenseNet. Одним із методів її навчання є трансферне навчання (англ. Transfer Learning), де спочатку використовується процес точного налаштування, щоб надати переднавченим вагам змогу скоригуватися и налаштуватися під дані для виявлення раку молочної залози. Потім з обранням кращих ваг й інших гіперпараметрів модель буде заново навчена. Табл. 4.4 показує результати експерименту з використанням моделі DenseNet.

Таблиця 4.4

Якість моделі DenseNet

Precision	Recall	F1-score
91,6%	90,50%	91%

На рис. 4.2 і 4.3 видно, як змінюється точність і функції моделі у навчанні за допомогою навчальної і валідаційної вибірок.

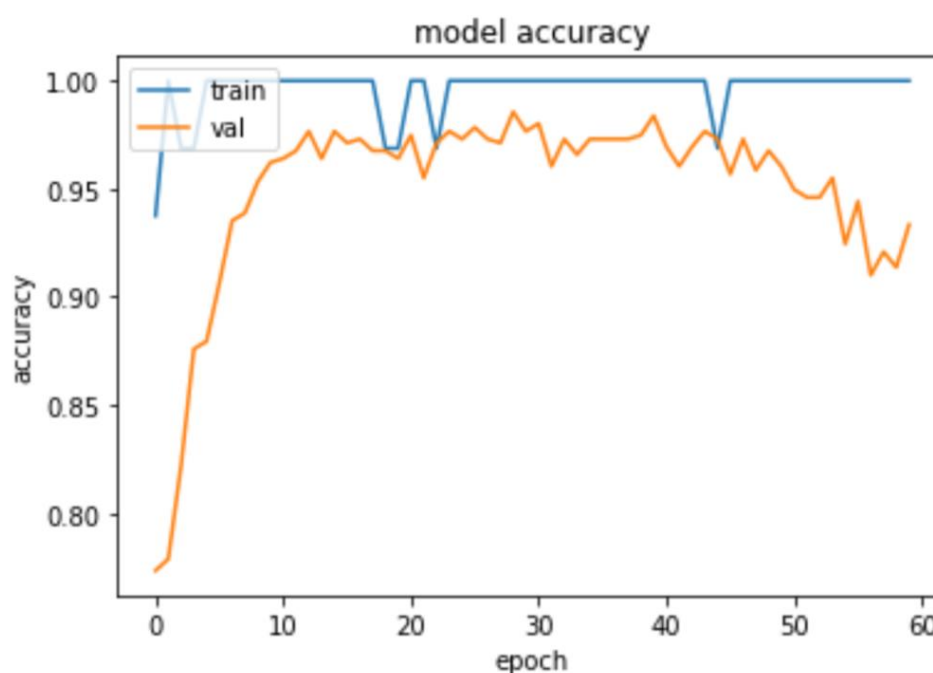


Рис. 4.2. Точність моделі за використання навчальних та валідаційних вибірок

Для навчання моделі використано функцію ранньої зупинки, де вказується, який параметр буде контролюватись і як довго. Інакше кажучи, надається параметр, у цьому випадку функція втрат, для валідаційних даних (англ. `val_loss`), далі спостерігати: якщо протягом певного часу (кількість епохи) `val_loss` не буде зменшуватись, тоді навчання моделі буде зупинено. Таким чином, модель не зазнає перенавчання.

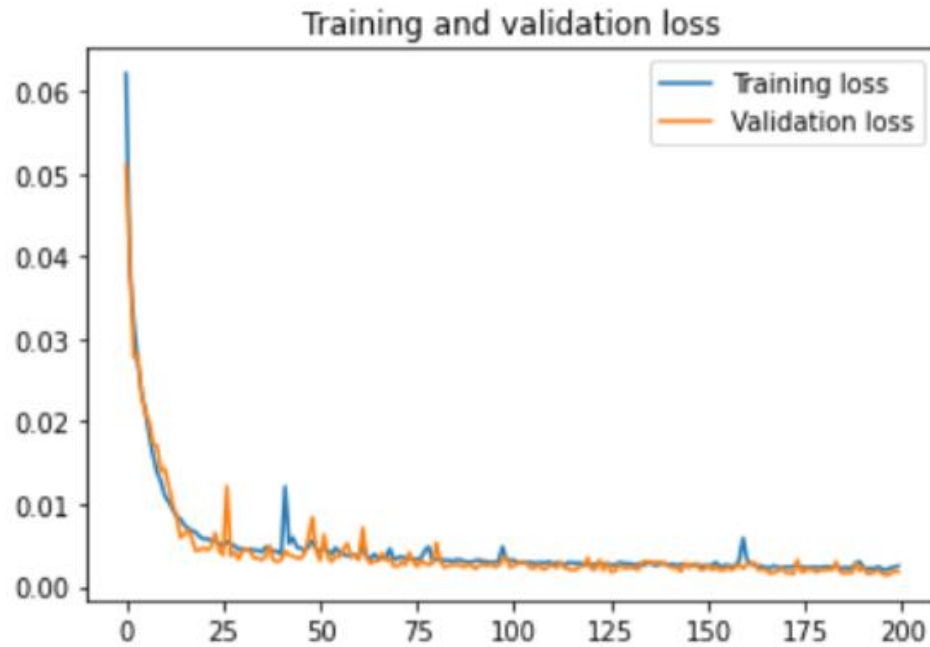


Рис. 4.3. Функція втрат за навчання моделі DenseNet

Також для обрання доречних гіперпараметрів використано K-fold валідацію. Такий підхід допоможе не втратити інформативні зображення для навчання моделі. У табл. 4.5 подано дані щодо чутливості моделі DenseNet за $K=3,5,7,10$.

Таблиця 4.5

Чутливість моделі DenseNet за використання K-fold валідації

Чутливість	K-fold валідація
77.60%	K=3
83.45%	K=5
87.75%	K=7
90.50%	K=10

Як, видно, з табл. 4.5, чутливість моделі досягла максимуму свого значення, а саме 90,50%, за $K = 10$. Експеримент довів, що після $K = 10$ чутливість моделі DenseNet починає зменшуватися.

4.3 Застосування згорткового автоенкодера для задачі діагностики раку молочної залози

Мета використання автоенкодера – навчання стисненого розподіленого подання для заданих даних зазвичай з метою зменшення розмірності. З іншого боку, існує аналіз основних компонентів (РСА) для тієї ж задачі (зменшення розмірності). Однак є деякі переваги використання автоенкодера згорткового типу:

1) може виконувати як лінійні, так і нелінійні перетворення в кодуванні (енкодер), коли РСА може виконувати тільки лінійні перетворення;

2) з погляду параметрів моделі було б ефективніше вивчати кілька шарів за допомогою автоенкодера, а не одне масивне перетворення за допомогою РСА;

3) дає уявлення про більш практичний вихід кожного шару і багаторазове подання різних розмірів.

Одна з причин використання згорткового автоенкодера у ході проведення експериментів полягає в тому, що дуже складно знайти набори даних значного розміру з мітками. Автоенкодер є моделлю, яка не потребує, щоб набір даних був з мітками для навчання частково без учителя (англ. semi-supervised). Ще однією перевагою автоенкодера є те, що модель зменшується. Таким чином, модель матиме менше параметрів, і, як результат, час обчислення і навчання різко скоротиться. Наприклад, у DenseNet 58 420 802 параметри, з яких 7 037 504 не піддаються навчанню.

Однак у запропонованому згортковому автоенкодері 2 940 865 параметрів, з яких лише 3 840 не піддаються навчанню.

У цій праці в ході експерименту використано кілька архітектур згорткового автоенкодера. Згортковий автоенкодер було модифіковано 18 шарами кодування (енкодер) і 14 шарами декодування (декодер). В енкодері використовувалися вісім згорткових і два максимальні пулінгові шари. У декодері було шість згорткових шарів і два деконволюційні шари. Між кожним конволюційним шаром використовувалася нормалізація. Запропонований згортковий автоенкодер навчено таким чином, щоб модель витягувала інформативні ознаки (коди) у процесі кодування, а декодер потім відтворював початкове вхідне зображення енкодера. Модель також відтворює вихідне зображення, незважаючи на те, що на скани нанесено деякі шуми. Порівняння вхідних і реконструйованих зображень проілюстровано на рис. 4.4.

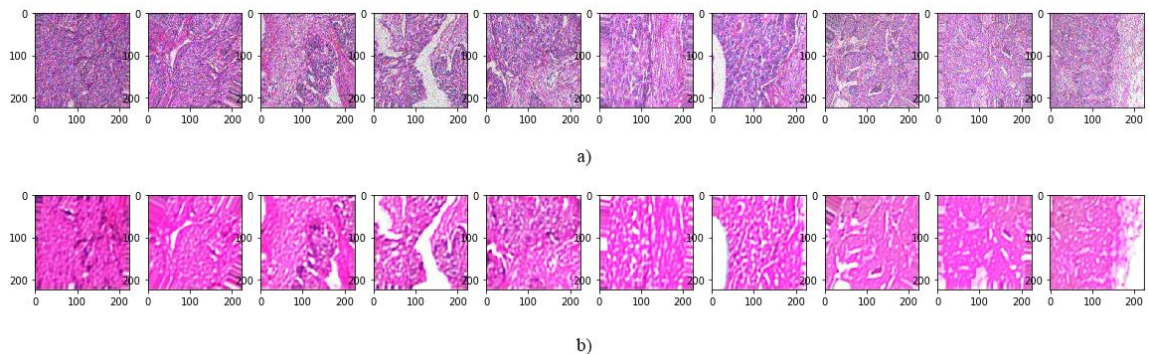


Рис. 4.4. На сканах (а) додано шуми, скани (b) реконструйовано за вихідними тестовими даними

На рис. 4.4 (а) зображено частину навчального набору даних, а на рис. 4.4 (b) – результат роботи згорткового автоенкодера. Точність відтворених тестових даних для згорткового автоенкодера становить 79,38%.

Після створення успішної автокодувальної моделі вихід енкодера буде використано з повнозв'язними шарами для створення повної моделі (згорткового автоенкодера).

На рис. 4.5 зображено модифіковану архітектуру згорткового автоенкодера.

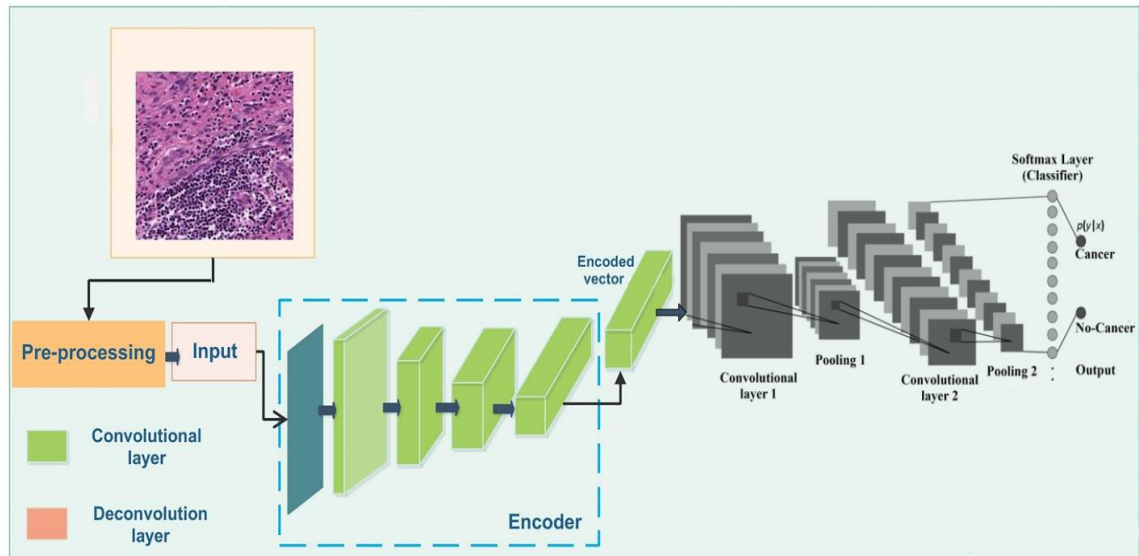
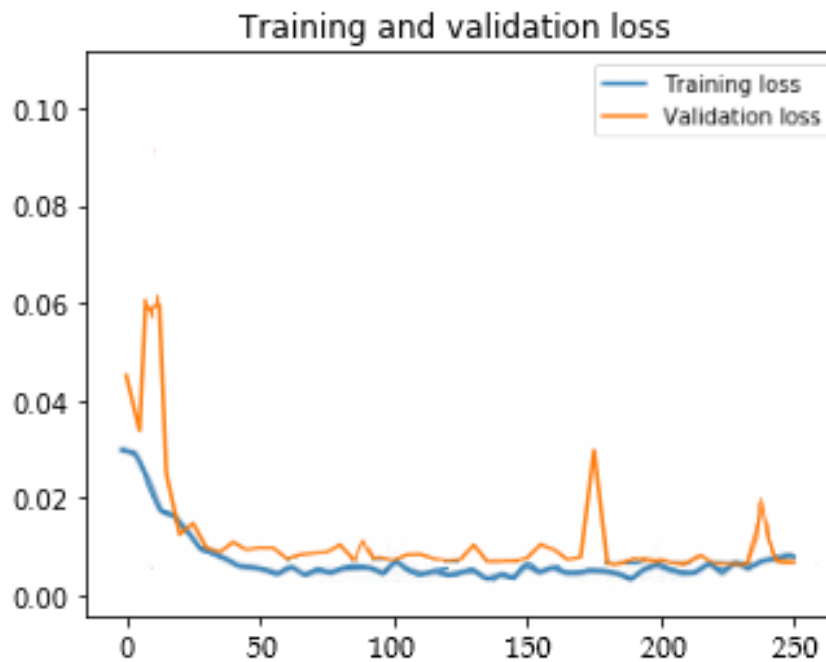


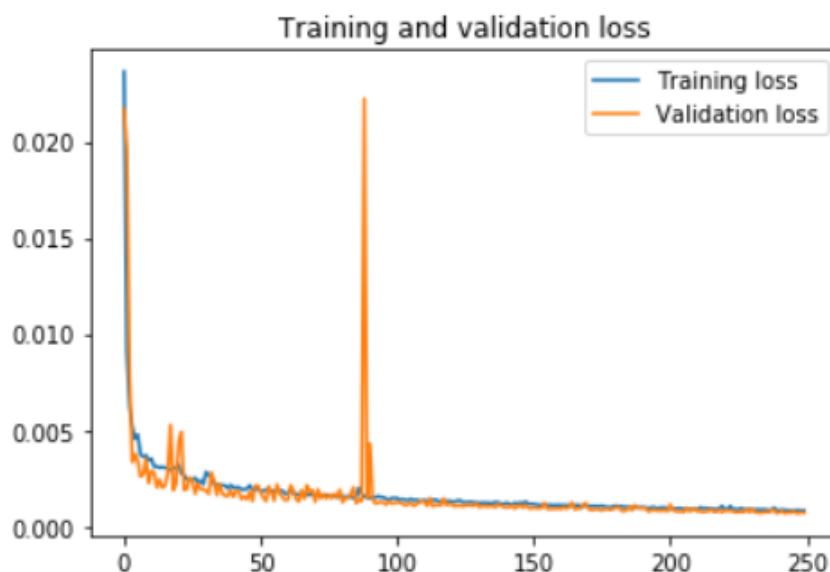
Рис. 4.5. Модифікована архітектура згорткового автоенкодера

Для навчання згорткового автоенкодера є чотири параметри, які необхідно підбирати шляхом експериментального аналізу. Перший – розмір коду. Розмір коду є кількістю вузлів у середньому шарі, менший розмір якого призводить до більшого стиснення. Другим параметром є кількість шарів і архітектура мережі, яка може бути на стільки глибокою, на скільки експерт визначить глибину згортки. Третій параметр – функція втрат. Четвертий – кількість вузлів на кожен шар. Кількість вузлів на шар зменшується з кожним наступним шаром енкодера і збільшується зворотно в декодері. Також декодер симетричний до енкодера в аспекті структури шару.

Оптимізатор Adam зі швидкістю навчання 0,001 використовувався для навчання DenseNet, коли згортковий автоенкодер RMSprop ($lr = 0,0001$) дав кращі результати. На рис. 4.6 (а) і 4.6 (б) зображено, як змінено функцію втрат за навчальної і валідаційної вибірок. Згідно з рис. 4.6 для поточної моделі краще використовувати коефіцієнт навчання 0,0001. Однак для інших моделей цей параметр може відрізнятися.



(a)



(б)

Рис. 4.6. Порівняння втрат навчальної і валідаційної вибірок із коефіцієнтом навчання 0,001 і 0,0001 для (а) і (б) відповідно

Із рис. 4.6 видно, що функція втрат за $lr = 0,0001$ зменшується поступово, але за $lr = 0,001$ графік має значне коливання. Таким чином, для навчання згорткового автоенкодера краще використовувати $lr = 0,0001$.

Експеримент демонструє, що велика кількість навчальної вибірки для навчання автоенкодера може погіршувати працездатність моделі. На рис. 4.7 показано, що збільшення вхідних даних на три рази не поліпшить точність моделі.

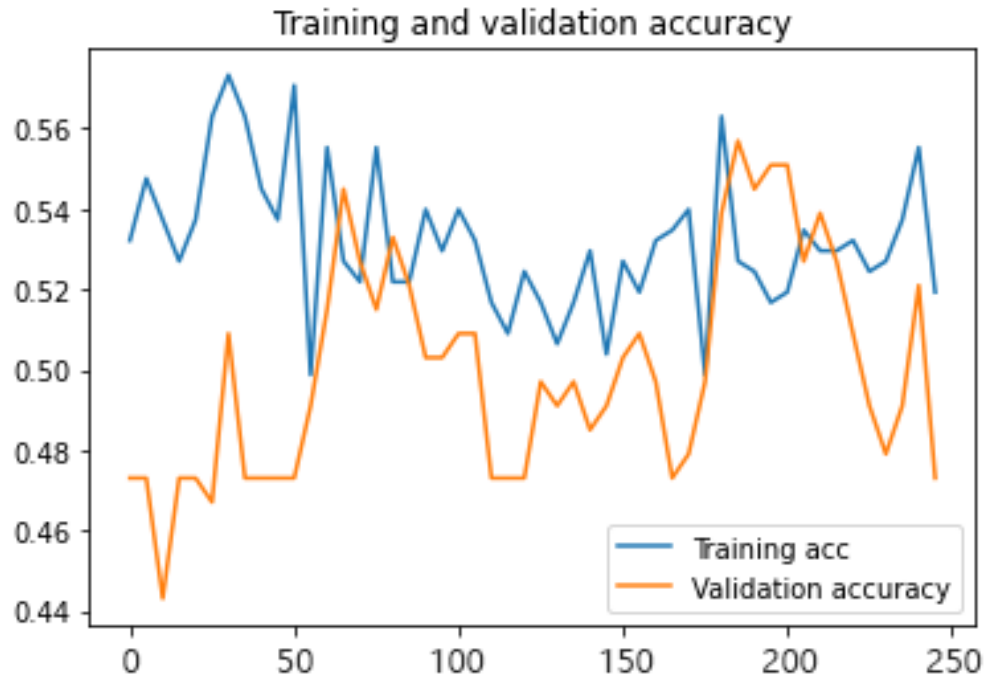


Рис. 4.7 (а). Точність моделі для навчальних і валідаційних даних з розміром виборки 5500

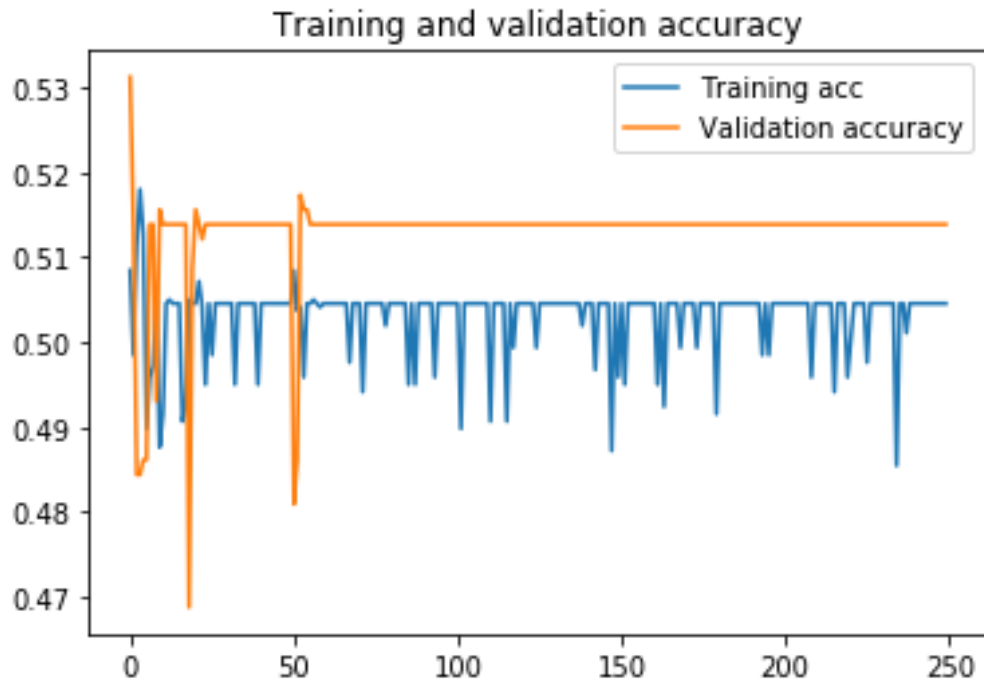


Рис. 4.7 (б). Точність моделі для навчальних і валідаційних даних за збільшення на три рази обсягу навчальної вибірки

Для того щоб отримати велику кількість набору даних, використано Image Data Generator. Ця функція має кілька параметрів, які допомагають з одного зображення зробити кілька. Також необхідно бути уважним у використанні Image Data Generator, зокрема з параметром «shear_range», де задається діапазон зміщення пікселів зображення. Оскільки маммографічні скани мають дуже чутливі дані, параметр shear_range може псувати зображення.

4.4 Експериментальні дослідження згорткових автоенкодерів

Для демонстрації того, як впливає кількість наборів даних на результат моделі згорткового автоенкодера, застосовано різний обсяг датасету. Із рис 4.8. видно, як змінюється точність моделі за різної кількості вхідні даних для навчання моделі.

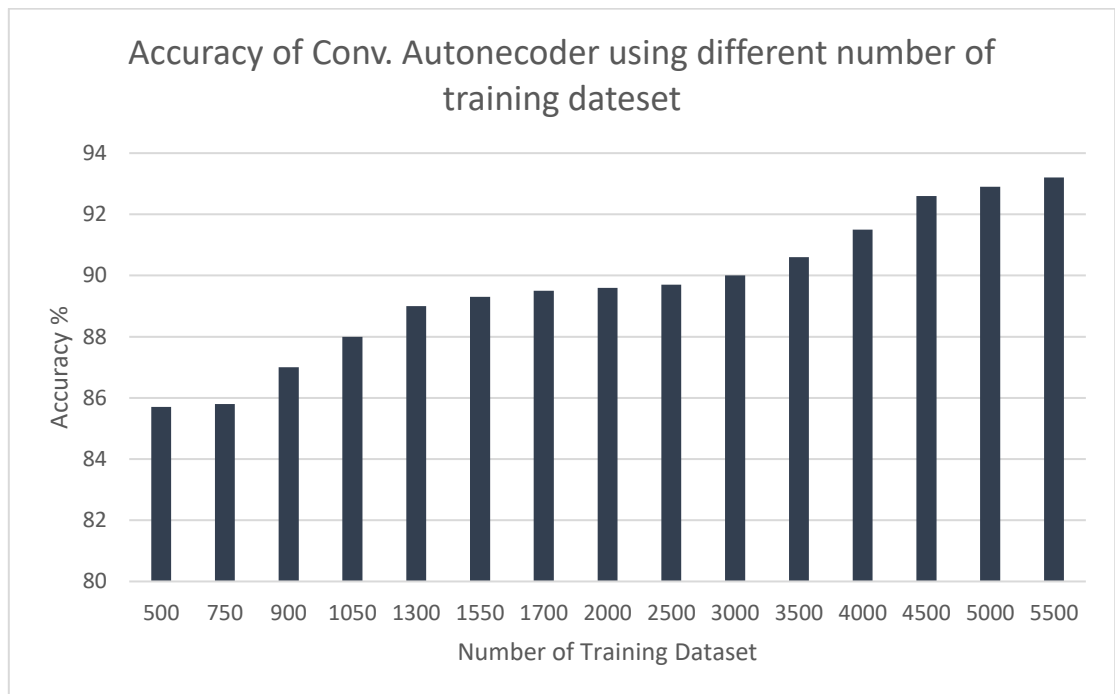


Рис. 4.8. Точність згорткового автоенкодера за різної кількості наборів даних

Експеримент доводить, що для навчання згорткового автоенкодера з нуля не потрібно мати великі набори даних. Натомість для навчання з нуля згорткових мереж DenseNet і Inception-v3 необхідна їх велика кількість. Таким чином, згортковий автоенкодер має спрощену модель, і час навчання значно зменшується порівняно з DenseNet або Inception-v3.

Також вигляд і якість зображення вхідних даних істотно впливають на продуктивність моделі. Для порівняння використано набори даних BreaKHis і Breast histology. У табл. 4.6 показано продуктивність моделі за різних вхідних даних.

Таблиця 4.6

Порівняння якості моделі згорткового автоенкодера для різних датасетів

Фактори / Датасети	BreaKHis	Breast histology
Accuracy	93%	90.5%
Precision	93.2%	91.6%
Recall	93.5%	92.40%
F1-Score	93.3%	92%

Як уже зазначалось у попередніх розділах, для аналізу моделі фактор точності є недостатнім фактором, який необхідно розглядати, бо також потрібно враховувати чутливість моделі.

Результат застосування згорткового автоенкодера подано у табл. 4.7.

Таблиця 4.7

Результати розпізнавання раку молочної залози з використанням згорткового автоенкодера

	Precision	Recall	F1-Score	Support
Class 0	90%	95%	93%	747
Class 1	95%	90%	92%	1626
Macro avg	93%	93%	92%	2373
Weighted avg	93%	93%	92%	2373

На основі даних із табл. 4.7 можна зробити висновок, що чутливість моделі (recall) за використання згорткового автоенкодера дає кращий результат порівняно з Inception-v3. Також потрібно зазначити, що тільки 5% з класу 0, де є рак, було неправильно класифіковано (англ. false negative).

Для того, щоб переконатись у тому, що ваги енкодера аналогічні до ваг, завантажених у кодувальну функцію класифікаційної моделі, треба розглядати той самий шар ваг обох моделей. Якщо вони не подібні, то немає сенсу використовувати класифікаційну функцію автоенкодера.

Рис. 4.9 ілюструє ваги для автоенкодера і ваги, аналогічні до кодувальної функції класифікаційної моделі. Як видно з рис. 4.9 (а) і 4.9 (б), ваги для першого шару аналогічні. Таким чином, автоенкодер працює правильно з мінімальним відхиленням, як зображено на рис. 4.9.

```
autoencoder.get_weights()[0][1]
```

```
array([[ [ 1.44881129e-01, -9.59705189e-02, -1.34551078e-01,
          8.66952688e-02,  7.09979534e-02,  1.23780712e-01,
          1.30307212e-01,  2.60467571e-03, -6.07582228e-03,
          1.93942040e-02,  9.58929062e-02, -1.34063229e-01,
          1.25064701e-01, -3.54465097e-02, -4.99075726e-02,
          2.40283143e-02,  1.17727511e-01,  1.22193985e-01,
          1.90873832e-01, -2.05673128e-02,  1.33657113e-01,
          -3.08714025e-02,  8.70231166e-02,  1.17021656e-04,
          -5.49316816e-02, -1.40636414e-01, -4.69613038e-02,
          -8.57445002e-02,  7.13037327e-02,  9.49074626e-02,
          1.01033375e-01,  1.36659533e-01]],

        [ [ 1.26553744e-01, -1.24628186e-01, -3.28947455e-02,
          -8.59334096e-02,  9.32449698e-02,  5.32429554e-02,
          8.29490498e-02, -1.24924853e-01, -1.35583133e-01,
          1.28295138e-01,  5.83672710e-02, -1.69967674e-02,
          1.00611210e-01, -5.01187481e-02, -1.10651806e-01,
          5.35668992e-02, -1.26742691e-01, -4.88741621e-02,
          1.54765040e-01,  4.46427837e-02, -3.37893441e-02,
          2.14125905e-02,  6.31524250e-02,  1.25887901e-01,
          6.89211711e-02,  1.12214573e-02, -4.61534932e-02,
          4.87665497e-02,  8.59796107e-02, -1.32830650e-01,
          -1.07983470e-01,  1.56714499e-01]],

        [ [-3.08928769e-02,  3.79457325e-02, -1.26879126e-01,
          -1.45610413e-02,  8.00895970e-03,  3.87692675e-02,
          4.91082808e-03,  8.51900429e-02, -1.09762155e-01,
          1.62591696e-01, -4.86132130e-02,  8.91351476e-02,
          -1.02033049e-01,  1.27722353e-01, -8.18031058e-02,
          -1.27505705e-01, -5.25270700e-02, -1.08628941e-03,
          -4.77150306e-02, -1.72305685e-02, -1.71014899e-03,
          1.37026384e-01, -9.13645253e-02, -1.58949606e-02,
          -5.20941503e-02,  1.28007784e-01, -2.08689719e-02,
          -3.90539095e-02,  6.41393811e-02,  1.05605297e-01,
          2.34576613e-02, -8.87834430e-02]]], dtype=float32)
```

Рис. 4.9 (а). Ваги енкодера автоенкодера для першого шару

```
full_model.get_weights()[0][1]
```

```
array([[ 1.44881129e-01, -9.59705189e-02, -1.34551078e-01,
        8.66952688e-02,  7.09979534e-02,  1.23780712e-01,
        1.30307212e-01,  2.60467571e-03, -6.07582228e-03,
        1.93942040e-02,  9.58929062e-02, -1.34063229e-01,
        1.25064701e-01, -3.54465097e-02, -4.99075726e-02,
        2.40283143e-02,  1.17727511e-01,  1.22193985e-01,
        1.90873832e-01, -2.05673128e-02,  1.33657113e-01,
       -3.08714025e-02,  8.70231166e-02,  1.17021656e-04,
       -5.49316816e-02, -1.40636414e-01, -4.69613038e-02,
       -8.57445002e-02,  7.13037327e-02,  9.49074626e-02,
        1.01033375e-01,  1.36659533e-01]],

      [[ 1.26553744e-01, -1.24628186e-01, -3.28947455e-02,
       -8.59334096e-02,  9.32449698e-02,  5.32429554e-02,
        8.29490498e-02, -1.24924853e-01, -1.35583133e-01,
        1.28295138e-01,  5.83672710e-02, -1.69967674e-02,
        1.00611210e-01, -5.01187481e-02, -1.10651806e-01,
        5.35668992e-02, -1.26742691e-01, -4.88741621e-02,
        1.54765040e-01,  4.46427837e-02, -3.37893441e-02,
        2.14125905e-02,  6.31524250e-02,  1.25887901e-01,
        6.89211711e-02,  1.12214573e-02, -4.61534932e-02,
        4.87665497e-02,  8.59796107e-02, -1.32830650e-01,
       -1.07983470e-01,  1.56714499e-01]],

      [[-3.08928769e-02,  3.79457325e-02, -1.26879126e-01,
       -1.45610413e-02,  8.00895970e-03,  3.87692675e-02,
        4.91082808e-03,  8.51900429e-02, -1.09762155e-01,
        1.62591696e-01, -4.86132130e-02,  8.91351476e-02,
       -1.02033049e-01,  1.27722353e-01, -8.18031058e-02,
       -1.27505705e-01, -5.25270700e-02, -1.08628941e-03,
       -4.77150306e-02, -1.72305685e-02, -1.71014899e-03,
        1.37026384e-01, -9.13645253e-02, -1.58949606e-02,
       -5.20941503e-02,  1.28007784e-01, -2.08689719e-02,
       -3.90539095e-02,  6.41393811e-02,  1.05605297e-01,
        2.34576613e-02, -8.87834430e-02]]], dtype=float32)
```

Рис. 4.9 (б). Ваги аналогічні до кодувальної функції класифікаційної моделі для першого шару

Однією з переваг автоенкодера є те, що модель проста і не потребує великої кількості даних для навчання. Те, як функція втрат змінюється за різної кількості навчальної вибірки, зображено на рис. 4.10.

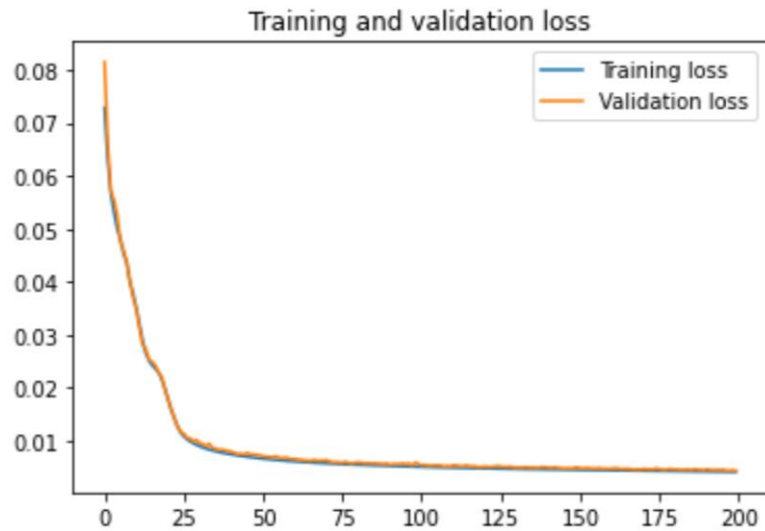


Рис. 4.10 (а). Функція втрат для навчальної і валідаційної вибірок за 1187 зображень

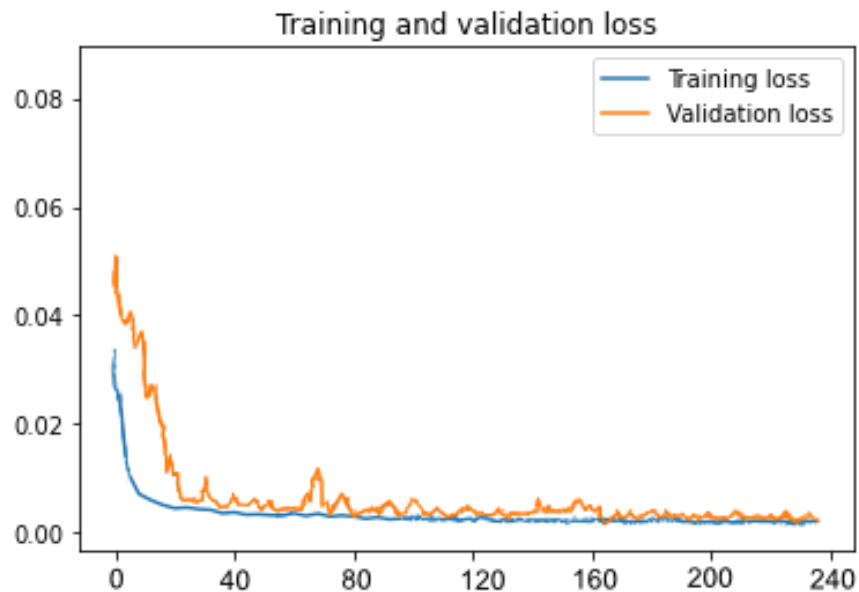


Рис. 4.10 (б). Функція втрат для навчальної і валідаційної вибірок за 5500 зображень

Як видно з рис. 4.10, щоб навчати згортковий автоенкодер немає необхідності в великій набору даних. У цьому випадку модель відтворює отримані зображення з мінімальним відхиленням.

Інша перевага згорткового автоенкодера – час навчання значно менший, ніж в інших згорткових мережах, таких як Inception-v3 або DenseNet. У табл. 4.8 наведено час навчання для різних моделей.

Таблиця 4.8

Час навчання для трьох моделей навчання

Модель	Час навчання
Inception-v3	~27h
DenseNet	~24h
Згортковий автоенкодер	~13h

Навчання мережі відбувається на платних серверах. Таким чином, чим менше часу потребує модель для навчання, тим вигіднішою вона виходить. Згортковий автоенкодер продемонстрував, що потребує найменше часу для навчання порівняно з іншими моделями.

4.5 Експериментальні дослідження іншого машинного навчання для розпізнавання раку

У табл. 4.9 подано порівняння алгоритмів машинного навчання (англ. machine learning, ML) на числових наборах даних [156].

Таблиця 4.9

Порівняння різних алгоритмів машинного навчання на числових наборах даних

ML Алгоритми	Precision	Recall	F1-Score
LR	84,25	80,3	82,23
KNN	81,5	83,6	82,54
SVM	86,8	84	85,37705
NB	79,45	80,1	79,78
DT	80,2	78,7	79,44292
RF	85,7	84,8	85,25

На підставі результатів, наведених у табл. 4.9, метод опорних векторів (SVM) дав кращі результати порівняно з іншими методами. Із використанням SVM досягнуто recall і precision 84% і 86,8% відповідно.

Оскільки SVM і RF дали кращі результати порівняно з іншими методами, ці алгоритми також використано для розпізнавання хвороби Альцгеймера. Вхідними даними для алгоритмів SVM і RF є цифрові дані. Перед навчанням моделі необхідно визначити основні інформативні ознаки. У табл. 4.10 подано всі ознаки, які використовувалися в наборі даних.

Таблиця 4.10

Інформаційні ознаки для розпізнавання хвороби Альцгеймера

Ознаки	Визначення
Subject ID	Ідентифікація об'єкта
MRI ID	Ідентифікація МРТ
Group	Клас
Visit	Відвідування
MR Delay	Магнітний контраст
M/F	Стать
Hand	Домінантна рука
EDUC	Рівень освіти
SES	Тест пізнання SES
MMSE	Тест пізнання MMSE
CDR	Тест пізнання CDR
eTIV	Розрахунковий внутрішньочерепний об'єм
nWBV	Стандартизований об'єм мозку
ASF	Атласна шкала коефіцієнта

У табл. 4.11 подано частину набору даних, які використано для розпізнавання хвороби Альцгеймера на МРТ сканах.

Таблиця 4.11

Дані деяких пацієнтів під час медичного обстеження

Group	Visit	MR Delay	M/F	Hand	Age	EDUC	SES	MMSE	CDR	cTIV	nWBV	ASF
Nondemented	1	0	M	R	87	14	2	27	0	1987	0,696	0,883
Demented	1	0	M	R	75	12		23	0,5	1678	0,736	1,046
Nondemented	1	0	F	R	88	18	3	28	0	1215	0,71	1,444
Nondemented	1	0	M	R	80	12	4	28	0	1689	0,712	1,039
Demented	1	0	M	R	71	16		28	0,5	1357	0,748	1,293
Nondemented	1	0	F	R	93	14	2	30	0	1272	0,698	1,380
Demented	1	0	M	R	68	12	2	27	0,5	1457	0,806	1,205
Demented	1	0	F	R	66	12	3	30	0,5	1447	0,769	1,213
Nondemented	1	0	F	R	78	16	2	29	0	1333	0,748	1,316
Nondemented	1	0	F	R	81	12	4	30	0	1230	0,715	1,427
Demented	1	0	M	R	76	16	3	21	0,5	1602	0,697	1,096
Demented	1	0	M	R	88	8	4	25	0,5	1651	0,66	1,063
Nondemented	1	0	M	R	80	12	3	29	0	1783	0,752	0,985
Converted	1	0	F	R	87	14	1	30	0	1406	0,715	1,248
Converted	1	0	M	R	80	20	1	29	0	1587	0,693	1,106

Хвороба Альцгеймера діагностується на основі значень деяких ключових показників. Аналіз даних із табл. 4.11 ілюструє деякі ознаки із середнім значенням для кожного пацієнта зі слабоумством і без слабоумства (Demented, Nondemented).

Таблиця 4.12

Середнє значення кожної ознаки, яка може спричинити хворобу
Альцгеймера.

AVE	Age	EDUC	SES	MMSE	CDR	eTIV	nWBV	ASF
Female Demented	76,10714	12,89286	3	25,46429	0,625	1373,929	0,728786	1,281036
Female Non-Demented	75,68	15,22	2,38	29,34	0	1402,56	0,75072	1,26152
Male Demented	74,33333	14,30556	2,5625	25,22222	0,583333	1555,278	0,720972	1,135361
Male Non-Demented	74,86364	15,04545	2,5	28,86364	0	1656,364	0,735682	1,070909

Грунтуючись на дані із табл. 4.12, можна зробити висновок, що середній рівень освіти пацієнтів зі слабоумством (Demented) нижчий, ніж у пацієнтів без слабоумства (Nondemented). Тимчасом як середній вік пацієнтів зі слабоумством вищий, ніж у пацієнтів без слабоумства (як жінок, так і чоловіків).

Під час навчання із використанням алгоритму випадкового лісу (англ. random forest algorithm, RFA) окремо порівняно всі ознаки. Після деяких експериментів [158] визначено стать, вік і рівень освіти, які є найбільш цінними атрибутами для класифікації. Коефіцієнт значущості кожної ознаки помножено на 100, як видно на рис. 4.11.

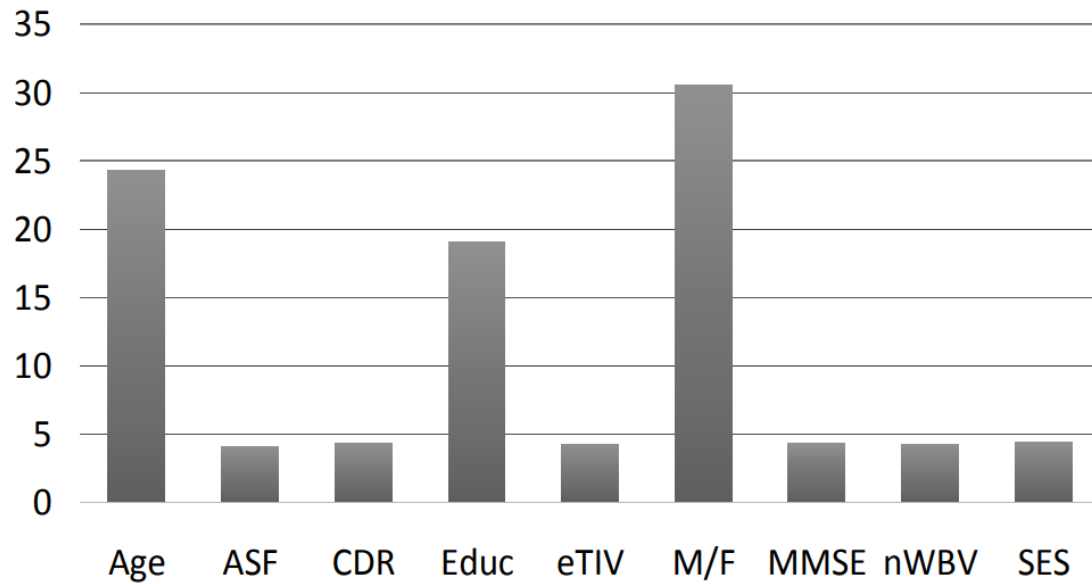


Рис. 4.11. Важливість ознак

У табл. 4.13 порівняно результати двох алгоритмів для класифікації.

Таблиця 4.13

Працездатність моделі за використання двох класифікаторів
SVM і RF [163]

Класифікатор	Recall	Precision	F1-Score
SVM	89%	88%	88,5%
RF	89%	89%	89%

Із використанням алгоритму випадкового лісу визначаємо найбільш інформативні ознаки, такі як стать, вік і рівень освіти, які можуть істотно вплинути на хворобу Альцгеймера. Потім ці інформативні ознаки будуть входом на останній шар згорткової мережі для класифікації раку.

4.6 Порівняльний аналіз методів діагностики раку молочної залози

У праці [15] для виявлення раку молочної залози запропоновано модифіковану Inception-v3. У [1] доопрацьовано DenseNet121 і запропоновано модифікований конволюційний автоенкодер. У табл. 4.14 подано порівняльний аналіз методів, застосованих для розпізнавання раку молочної залози.

Таблиця 4.14

Порівняльний аналіз моделей, застосованих для розпізнавання раку молочної залози

Model	Inception-v3	DenseNet 121	Hybrid CNN [164]
Precision	66,66%	91,6%	93,2%
Recall	85,84%	90,50%	93,5%
F1-Score	74,99%	91%	93.3%
Accuracy	74,78%	89,67%	93%
Time	~27h	~24h	~13h

На основі експериментальних аналізів модифікований згортковий автоенкодер дає кращий результат за усіма параметрами порівняно з Inception-v3 і DenseNet 121. Також суттєво скоротився час навчання для цієї мережі, оскільки модель проста і має в собі значно менше параметрів.

4.7 Висновки до розділу

1. Запропоновано модифікований гібридний згортковий мереж для розпізнавання раку молочної залози, заснований на використанні операції згортки для виділення інформативних ознак.
2. Виконано порівняльний аналіз способу виділення інформативних ознак на основі різних моделей згортки і кількох алгоритмів машинного навчання, таких як SVM і RF.
3. Розроблено метод розпізнавання раку молочної залози, заснований на використанні згорткових моделей як класифікатора. Виконано експерименти зі згортковими моделями: Inception-v3, DenseNet121 й автоенкодер. У ході експериментальних досліджень найкращу якість розпізнавання раку молочної залози на мамографічних зображеннях отримано за навчання згорткової нейронної мережі автоенкодера, у якій відсоток помилково класифікованих зразків становить 6,82.
4. Описано запропонований метод розпізнавання раку молочної залози, що складається з трьох етапів: нормалізації, реконструкції і розпізнавання. Виконано експериментальні дослідження розробленого методу розпізнавання раку молочної залози, отримані показники: чутливість, точність (precision), F1-Score і точність (accuracy) моделі при цьому становлять 93,5%, 93,2%, 93,3% і 93% відповідно.

ВИСНОВКИ ДО ДИСЕРТАЦІЇ

У праці розглянуто і досліджено проблему розпізнавання раку за допомогою медичних зображень з метою зменшення часу навчання моделі. Рак молочної залози – дуже поширений вид раку у жінок і другий за поширеністю у світі. Діагностика раку молочної залози часто розглядається як проблема класифікації у нейронних мережах. При цьому обмежений доступ до маркованих даних є додатковою проблемою у навчанні нейронних мереж із високою ефективністю. Виявлення та діагностика раку молочної залози на ранніх стадіях має вирішальне значення для порятунку життя. У роботі показано, що за використання згорткового автоенкодера модель не потребує великих наборів даних для досягнення високої ефективності порівняно з відомими методами.

У ході проведення дисертаційних досліджень отримано такі наукові результати:

1. Запропоновано спосіб знаходження інформативних ознак для розпізнавання раку, що ґрунтується на використанні операції згортки для виділення інформативних ознак.
2. Розроблено новий метод розпізнавання раку молочної залози, заснований на використанні згорткової мережі автоенкодера як класифікатора.
3. Встановлено, що для поліпшення якості розпізнавання раку доцільно застосовувати стохастичний градієнтний спуск, алгоритм навчання параметрів функції приналежності на етапі первинного навчання системи і алгоритм сполучених градієнтів на етапі тонкого навчання (fine tuning) системи для поліпшення часових показників.

4. Запропоновано модифікований гібридний згортковий мереж, який ґрунтується на знаходженні найкращих інформативних ознак для розпізнавання раку.
5. Запропоновані в дисертаційній роботі методи забезпечують розпізнавання раку молочної залози з найкращою ефективністю в порівнянні з відомими методами. Загальний час навчання моделі становить приблизно 13 годин, що є найменшим часом навчання порівняно з іншими згортковими мережами. Чутливість, точність (precision), F1-Score і точність (accuracy) моделі при цьому становить 93,5%, 93,2%, 93,3% і 93%, відповідно, що значно більше ніж у відомих згорткових мереж, які були застосовані для цієї задачі.

У дисертації здобувачем запропоновано модифіковану архітектуру згорткової мережі, що полягає у поєднанні згорткового автоенкодера з двома згортковими шарами і класифікатора softmax.

Практична цінність результатів полягає в:

- Запропоновані у дисертаційній роботі методи реалізовані у якості інформаційної технології, що збільшує ефективність діагностики раку молочної залози, а саме підвищує точність класифікації та зменшує час класифікації в порівнянні з відомими методами та моделями на основі згорткових мереж.
- Розроблено систему комп'ютерної діагностики раку молочної залози на основі обробки та аналізу медичних зображень молочної залози.

СПИСОК ДЖЕРЕЛ

1. M. Naderan, Yu. Zaychenko, and A. Napoli, “USING CONVOLUTIONAL NEURAL NETWORKS FOR BREAST CANCER DIAGNOSING”, *System Research & Information Technologies*, no. 4, pp. 85–93, 2019.
2. M. Naderan and Yu. Zaychenko, “Convolutional autoencoder application for breast cancer classification”, *SAIC IEEE*, Lviv, Kyiv 2020.
3. M. Naderan and Yu. Zaychenko, “A comparative examination of convolutional autoencoder and DenseNet applications for breast cancer classification”, *Information Theories and Applications*, vol. 27, no. 1, 2020.
4. *Cancer*. Available: <https://ourworldindata.org/cancer>
5. *World cancer report 2008*. Lyon, International Agency for Research on Cancer, 2008.
6. Lacey J.V Jr. et al., “Breast cancer epidemiology according to recognized breast cancer risk factors in the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial Cohort”, *BMC Cancer*, 9, 84, 2009.
7. G. Danaei et al., “Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors”, *Lancet*, vol. 366, pp. 1784–1793, 2005.
8. J. Peto, “Cancer epidemiology in the last century and the next decade”, *Nature*, vol. 411, pp. 390–395, 2005.
9. *Inverse Relationship Between Precision and Recall*. Available: <https://datascience.stackexchange.com/questions/49117/inverserelationship-between-precision-and-recall>.
10. G. Arestaa et al., “BACH: grand challenge on breast cancer histology images”, *Medical Image Analysis*. ELSEVIER, 2019.

11. “Global Health Estimates”, *WHO*, 2013. Available: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates>.
12. *GLOBOCAN*, 2008. Available: <https://pubmed.ncbi.nlm.nih.gov/21351269>.
13. M.P. Coleman et al., “Cancer survival in five continents: a worldwide population-based study (CONCORD)”, *Lancet Oncol*, vol. 9, pp. 730–756, 2008.
14. *BreakHist-Dataset-Image-Classification*. Available: <https://github.com/Anki0909/BreakHist-Dataset-Image-Classification>.
15. Gao Huang, Zhuang Liu and Kilian Q. Weinberger, “Densely Connected Convolutional Networks”, *arXiv preprint arXiv:1608.06993v3*, 2018.
16. Majid Nawaz, Adel A. Sewissy, and Taysir Hassan A. Soliman, “MultiClass Breast Cancer Classification using Deep Learning Convolutional Neural Network”, *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, 2018.
17. Yuexiang Li, Xinpeng Xie, Linlin Shen, and Shaoxiong Liu, “Reversed Active Learning based Atrous DenseNet for Pathological Image Classification”, *arXiv preprint arXiv:1807.02420*, 2018.
18. A. Cruz-Roa et al., “Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent”, *Sci. Rep*, 7(1): 46450, 2017. doi: 10.1038/srep46450.
19. Sun Wenqing Tseng, Bill Zhang, and Jianying Qian Wei, “Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data”, *Computerized Medical Imaging and Graphics*, 2016. doi: 57. 10.1016/j.compmedimag.2016.07.004.

20. Y. Zaychenko and G. Hamidov, “Hybrid Fuzzy CNN Network in the Problem of Medical Images Classification and Diagnostics”, *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery*, vol. 1074. Springer, Cham, 2020. doi: 10.1007/978-3-030-32456-8_95.
21. N. Yassin, S. Omran, E. El Houbay, and H. Allam, “Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review”, *Computer Methods and Programs in Biomedicine*, vol. 156, pp. 25–45, 2018.
22. S. Sharma and P. Khanna, “Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM”, *Journal of Digital Imaging*, vol. 28 (1), pp. 77–90, 2015.
23. S.M.A. Beheshti et al., “An efficient fractal method for detection and diagnosis of breast masses in mammograms”, *Journal of Digital Imaging*, vol. 27 (5), pp. 661–669, 2014.
24. A.T. Azar and S.M. El-Metwally, “Decision tree classifiers for automated medical diagnosis”, *Neural Comput. Appl.*, vol. 23 (7), pp. 2387–2403, 2013.
25. W. Jian, X. Sun, and S. Luo, “Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform”, *Biomed. Eng. Online*, vol. 11 (1), p. 96, 2012.
26. A. Hiba, M. Hajar, M. Hassan, and N. Thomas, “Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis”, *The 6th International Symposium on Frontiers in Ambient and Mobile Systems. Procedia Computer Science*, 83, pp. 1056–1063, 2016.
27. H. Sharma et al., “Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology”, *Computerized Medical Imaging and Graphics*, vol. 61, pp. 2–13, 2017.

28. M. Mohammad et al., “A Hybrid Computer-aided-diagnosis System for Prediction of Breast Cancer Recurrence (HPBCR) Using Optimized Ensemble Learning”, *Computational and Structural Biotechnology Journal*, vol. 15, pp. 75–85, 2017.
29. Li Shen et al., *Deep Learning to Improve Breast Cancer Early Detection on Screening Mammography*. 2018. doi: 10.1038/s41598-018-22437-z
30. *K-Fold Cross Validation*. Available: <https://medium.com/datadriveninvestor/K-fold-cross-validation-6b8518070833>
31. A. Rakhlin, A. Shvets, and V. Iglovikov, “Deep Convolutional Neural Networks for Breast Cancer Histology Image Analysis”, *ICIAR 2018 Grand Challenge; arXiv preprint arXiv:1802.00752v2*.
32. R. Fakoor, F. Ladhak, A. Nazari, and M. Huber, “Using deep learning to enhance cancer diagnosis and classification”, *Proceedings of the 30 th International Conference on Machine Learning*, Atlanta, Georgia, USA, 2013, vol. 28. JMLR: W&CP.
33. S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci and N. Navab, “AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images”, in *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1313–1321, May 2016.
34. J. Wang et al., “Discrimination of Breast Cancer with Microcalcifications on Mammography by Deep Learning”, *Sci. Rep*, 6, 27327, 2016. Available: <https://doi.org/10.1038/srep27327>.
35. A.S. Becker, M. Marcon, G. Soleen, M.C. Wurnig, T. Frauenfelder, and A. Boss, *Deep Learning in Mammography Diagnostic Accuracy of a Multipurpose Image Analysis Software in the Detection of Breast Cancer*. Wolters Kluwer Health, Inc. 2017. doi:10.1097/RLI.0000000000000358

36. Z. Han et al., “Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model”, *Sci. Rep.*, 7, 4172, 2017. Available: <https://doi.org/10.1038/s41598-017-04075-z>.
37. S. McGuire, World cancer report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015, *Advances in Nutrition*, vol. 7(2), pp. 418–419, 2016.
38. Aditya Golatkar, Deepak Anand, and Amit Sethi, “Classification of Breast Cancer Histology using Deep Learning”, *arXiv preprint arXiv:1802.08080v2*, 2018.
39. F.A. Spanhol, L.S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural networks”, *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2560–2567.
40. T. Ara’ujo et al, “Classification of breast cancer histology images using convolutional neural networks”, *PLoS ONE*, 12(6): e0177544, 2017. doi: 10.1371/journal.pone.0177544.
41. M. Macenko et al., “A method for normalizing histology slides for quantitative analysis”, *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Boston, MA, 2009, pp. 1107–1110. doi: 10.1109/ISBI.2009.5193250.
42. *Observatory TGC (2018) GLOBOCAN. Breast Cancer 2019*. Available: <http://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf>.
43. R.L. Siegel, K.D. Miller, and A. Jemal, “Cancer statistics, 2019”. *CA: A Cancer Journal for Clinicians*, vol. 69(1), pp. 7–34, 2019.
44. E. Grunfeld et al., “Family caregiver burden: results of a longitudinal study of breast cancer patients and their principal caregivers”, *CMAJ*, vol. 170(12), pp. 1795–1801, 2004.

45. H.P. Kale and N.V. Carroll, “Self-reported financial burden of cancer care and its effect on physical and mental health-related quality of life among US cancer survivors”, *Cancer*, vol. 122(8), pp. 283–289, 2016.
46. “The costs of breast cancer in the U.S.”, *care Co*, 2019. Available: <https://costsofcare.org/the-costs-of-breast-cancer-in-the-u-s/>
47. M. Pilevarzadeh et al., “Global prevalence of depression among breast cancer patients: a systematic review and meta-analysis”, *Breast Cancer Res Treat*, vol. 176, pp. 519–533, 2019.
48. S. Khan and S. Yong, “A deep learning architecture for classifying medical images of anatomy object”, *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Kuala Lumpur, 2017, pp. 1661–1668.
49. G. Liang, H. Hong, W. Xie and L. Zheng, “Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification”, in *IEEE Access*, vol. 6, pp. 36188–36197, 2018.
50. D. Komura and S. Ishikawa, “Machine Learning Methods for Histopathological Image Analysis”, *Computational and Structural Biotechnology Journal*, vol. 16, pp. 34–42, 2018.
51. *Breast cancer: prevention and control*. Available: <https://www.who.int/cancer/detection/breastcancer/en/>
52. N. Yassin, S. Omran, E. El Houbay, and H. Allam, “Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review”, *Computer Methods and Programs in Biomedicine*, vol. 156, pp. 25–45, 2018.
53. T. Mungle et al., “MRF-ANN: a machinelearning approach for automated ER scoring of breast cancer immunohistochemical images”, *Journal of Microscopy*, vol. 267, pp. 117–129, 2017. Available: <https://doi.org/10.1111/jmi.12552>.

54. D. Wang, D.J. Foran, J. Ren, H. Zhong, I.Y. Kim, and X. Qi, “Exploring automatic prostate histopathology image gleason grading via local structure modeling”, *37th Annu.Int. Conf. IEEE Eng. Med. Biol. Soc. EMBC*, 2015. pp. 2649–2652. Available: <https://doi.org/10.1109/EMBC.2015.7318936>.
55. D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A.H. Beck, “Deep learning for identifying metastatic breast cancer”, *arXiv preprint arXiv1606.05718*, 2016.
56. Y. Liu, et al., “Detecting cancer metastases on gigapixel pathology images”, *arXiv preprint arXiv1703.02442*, 2017.
57. *CAMELYON17*. Available: <https://camelyon17.grand-challenge.org>
58. R. Sparks and A. Madabhushi, “Out-of-sample extrapolation utilizing semi-supervised manifold learning (OSE-SSL): content based image retrieval for histopathology images”, *Sci. Rep*, vol. 6, 2016. Available: <https://doi.org/10.1038/srep27306>.
59. M. Peikari, J. Zubovits, G. Clarke, and A.L. Martel, “Clustering analysis for semi-supervised learning improves classification performance of digital pathology”, *Mach. Learn. Med. Imaging*. Springer, Cham, 2015. pp. 263–270. Available: https://doi.org/10.1007/978-3-319-24888-2_32.
60. Y. Xu et al., “Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features”, *BMC Bioinformatics*, 18, 281 (2017). Available: <https://doi.org/10.1186/s12859-017-1685-x>
61. B. Kieffer, M. Babaie, S. Kalra, and H.R. Tizhoosh, *Convolutional neural networks for histopathology image classification: training vs. using pre-trained networks*, 2017. arXiv preprint arXiv1710.05726.

62. *Transfer learning for cell nuclei classification in histopathology Images*. Available: https://link.springer.com/chapter/10.1007/978-3-319-49409-8_46
63. B. Wei, K. Li, S. Li, Y. Yin, Y. Zheng, and Z. Han, “Breast cancer multi-classification from histopathological images with structured deep learning model”, *Sci. Rep*, 7,4172(2017). Available: <https://doi.org/10.1038/s41598-017-04075-z>.
64. Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, “Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning”, *IEEE Trans Biomed Eng*, 62, pp. 2421–2433, 2015. Available: <https://doi.org/10.1109/TBME.2015.2430895>.
65. D. Romo, J.D. García-Arteaga, P. Arbeláez, and E. Romero, “A discriminant multi-scale histopathology descriptor using dictionary learning”, *International Society for Optics and Photonics*, vol. 9041, 2014. Available: <https://doi.org/10.1117/12.2043935>.
66. S. Doyle, A. Madabhushi, M. Feldman, and J. Tomaszewski, “A boosting cascade for automated detection of prostate cancer from digitized histology”, *Med image comput comput-assist interv MICCAI int conf med image comput comput-assist interv*, vol. 9, pp. 504–511, 2006.
67. B.E. Bejnordi et al., “Stain specific standardization of whole-slide histopathological images”, *IEEE Trans Med Imaging*, 35, pp. 404–415, 2016. Available: <https://doi.org/10.1109/TMI.2015.2476509>.
68. F. Ciompi et al., “The importance of stain normalization in colorectal tissue classification with convolutional networks”, *arXiv preprint arXiv1702.05931*, 2017.
69. A.M. Khan, N. Rajpoot, D. Treanor, and D. Magee, “A nonlinear mapping approach to stain normalization in digital histopathology images using

- image-specific color deconvolution”, *IEEE Trans Biomed Eng*, 61, pp. 1729–1738, 2014. Available: <https://doi.org/10.1109/TBME.2014.2303294>.
70. H. Cho, S. Lim, G. Choi, and H. Min, “Neural stain-style transfer learning using GAN for histopathological images”, *arXiv preprint arXiv1710.08543*, 2017.
71. M.W. Lafarge, J.P.W. Pluim, K.A.J. Eppenhof, P. Moeskops, and M. Veta, “Domain-adversarial neural networks to address the appearance variability of histopathology images”, *arXiv preprint arXiv1707.06183*, 2017.
72. *ScanNet: a fast and dense scanning framework for metastatic breast cancer detection from whole-slide images - semantic scholar*. Available: <https://www.semanticscholar.org/paper/ScanNet-A-Fast-and-Dense-Scanning-Framework-for-Me-Lin-Chen/9484287f4d5d52d10b5d362c462d4d6955655f8e>
73. *CAMELYON17*. Available: <https://camelyon17.grand-challenge.org>
74. T. Wollmann and K. Rohr, “Automatic breast cancer grading in lymph nodes using a deep neural network”, *arXiv preprint arXiv1707.07565*, 2017.
75. O. Russakovsky et al., “ImageNet large scale visual recognition challenge”, *Int J Comput Vis*, 115, pp. 211–252, 2015. Available: <https://doi.org/10.1007/s11263-015-0816-y>.
76. D. Gutman et al. “Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (ISBI) 2016”, *arXiv preprint arXiv1605.01397*.
77. F.A. Spanhol, L.S. Oliveira, C. Petitjean, and L. Heutte, “Breast cancer histopathological image classification using convolutional neural

- networks”, *2016 Int. Jt. Conf. Neural Netw (IJCNN)*, pp. 2560–2567.
Available: <https://doi.org/10.1109/IJCNN.2016.7727519>
78. *Breast Cancer Histopathological Database (BreakHis) – Laboratório Visão Robótica e Imagens*. Available: <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathologicaldatabase-breakhis/>
79. *MITOS-ATYPIA-14 - Dataset*. Available: <https://mitos-atypia-14.grand-challenge.org/dataset/>
80. R. Marée et al., “Collaborative analysis of multi-gigapixel imaging data using cytomine”, *Bioinformatics*, 32, pp. 1395–1401, 2016. Available: <https://doi.org/10.1093/bioinformatics/btw013>.
81. *Interactive phenotyping of large-scale histology imaging data with HistomicsML*. Available: <http://www.biorxiv.org/content/early/2017/05/19/140236>
82. S. Doyle, J. Monaco, M. Feldman, J. Tomaszewski, and A. Madabhushi, “An active learning based classification strategy for the minority class problem: application to histopathology annotation”, *BMC Bioinformatics*, 12, p. 424, 2011. Available: <https://doi.org/10.1186/1471-2105-12-424>
83. R.K. Padmanabhan et al., “An active learning approach for rapid characterization of endothelial cells in human tumors”, *PLoS ONE*, 9, 2014. Available: <https://doi.org/10.1371/journal.pone.0090495>
84. Y. Zhu, S. Zhang, W. Liu, and D.N. Metaxas, “Scalable histopathological image analysis via active learning”, *Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv*, 17, pp. 369–376, 2014.
85. Y. Xu, J.-Y. Zhu, E.I.-C. Chang, M. Lai, and Z. Tu, “Weakly supervised histopathology cancer image segmentation and classification”, *Med Image*

- Anal*, 18, pp. 591–604, 2014. Available: <https://doi.org/10.1016/j.media.2014.01.010>.
86. S. Tong and D. Koller, “Support vector machine active learning with applications to text classification”, *J Mach Learn Res*, 2, pp.45–66, 2001.
87. D.D. Lewis and W.A. Gale, “A sequential algorithm for training text classifiers”, *SIGIR '94 Proc. Seventeenth Annu. Int. ACM-SIGIR Conf. Res. Dev. Inf. Retr.* Springer-Verlag, London, pp. 3–12, 1994. Available: https://doi.org/10.1007/978-1-4471-2099-5_1.
88. Y. Xu et al., “Parallel multiple instance learning for extremely large histopathology image analysis”, *BMC Bioinformatics*, 18, p. 360, 2017. Available: <https://doi.org/10.1186/s12859-017-1768-8>.
89. T.G. Dietterich, R.H. Lathrop, and T. Lozano-Pérez, “Solving the multiple instance problem with axis-parallel rectangles”, *Artif Intell*, 89, pp. 31–71, 1997. Available: [https://doi.org/10.1016/S0004-3702\(96\)00034-3](https://doi.org/10.1016/S0004-3702(96)00034-3)
90. Z. Jia, X. Huang, E.I.-C. Chang, and Y. Xu, “Constrained deep weak supervision for histopathology image segmentation”, *arXiv preprint arXiv1701.00794*, 2017.
91. A. BenTaieb, H. Li-Chang, D. Huntsman, and G. Hamarneh, “A structured latent model for ovarian carcinoma subtyping from histopathology slides”, *Med Image Anal*, 39, pp. 194–205, 2017. Available: <https://doi.org/10.1016/j.media.2017.04.008>
92. T. Miyato, S. Maeda, M. Koyama, and S. Ishii, “Virtual adversarial training: a regularization method for supervised and semi-supervised learning”, *arXiv preprint arXiv1704.03976*, 2017.
93. A. Rasmus, H. Valpola, M. Honkala, M. Berglund, and T. Raiko, “Semi-supervised learning with ladder networks”, *arXiv preprint arXiv1507.02672*, 2017.

94. H. Sharma et al., “Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology”, *Computerized Medical Imaging and Graphics*, 61, pp. 2–13, 2017.
95. *An Intuitive Explanation of Convolutional Neural Networks*. Available: <https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/>
96. O.E. David and N.S. Netanyahu, “Deeppainter: Painter classification using deep convolutional autoencoders”, *Proc. Int. Conf. Artif. Neural Netw.* Springer, 2016, pp. 20–28.
97. G.E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks”, *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
98. Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives”, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
99. A. Ng, “Sparse autoencoder”, *CS294A Lecture Notes*, vol. 72, pp. 1–19, 2011.
100. D. Kingma and J. Ba Adam, “A method for stochastic optimization”, in *Proc. 3rd Int. Conf. Learning Representations*, San Diego, 2015.
101. M.D. Zeiler, G.W. Taylor, and R. Fergus, “Adaptive deconvolutional networks for mid and high level feature learning”, in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2018–2025, 2011.
102. M. Saygin, A. Murat, and S. Zubeyde, “Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities”, *IEEE transactions on aerospace and electronic systems*, vol. 54, no. 4, Aug. 2018.

103. P.F. Felzenszwalb et al., “Object detection with discriminatively trained part-based models”, *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.
104. A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks”, in *Advances in Neural Information Processing Systems 25 (NIPS)*, pp. 1097–1105, 2012. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-withdeep-convolutional-neural-networks.pdf>
105. P. Sermanet, S. Chintala, and Y. LeCun, “Convolutional neural networks applied to house numbers digit classification,” in *International Conference on Pattern Recognition (ICPR)*, no. 21, pp. 3288–3291, Apr. 2012. Available: <https://arxiv.org/abs/1204.3968>
106. P. Sermanet et al., “Overfeat: Integrated recognition, localization and detection using convolutional networks”, *arXiv preprint arXiv:1312.6229*, Feb. 2013. Available: <https://arxiv.org/abs/1312.6229v4>
107. K. He et al., “Deep residual learning for image recognition”, *arXiv preprint arXiv:1512.03385*, Dec. 2015. Available: <https://arxiv.org/abs/1512.03385v1>
108. A. Krizhevsky, I. Sutskever, and G.E. Hinton, “Imagenet classification with deep convolutional neural networks”, in *Advances in Neural Information Processing Systems 25 (NIPS)*, pp. 1097–1105, 2012. Available: <http://papers.nips.cc/paper/4824-imagenet-classification-withdeep-convolutional-neural-networks.pdf>
109. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, *arXiv preprint arXiv:1409.1556*, Sep. 2014. Available: <https://arxiv.org/abs/1409.1556>

110. C. Szegedy et al., “Going deeper with convolutions”, in *Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 1–9, Sep. 2015. Available: <https://arxiv.org/abs/1409.4842>
111. A. Waibel et al., “Phoneme recognition using time-delay neural networks”, *IEEE transactions on acoustics, speech, and signal processing*, vol. 37, no. 3, pp. 328–339, Aug. 1989. Available: <http://ieeexplore.ieee.org/document/21701/>
112. J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation”, in *Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, pp. 3431–3440, Mar. 2015. Available: <https://arxiv.org/abs/1411.4038v2>
113. AlexNet: The Architecture that Challenged CNNs. Available: <https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951>.
114. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, *arXiv preprint arXiv:1409.1556*, Sep. 2014. Available: <https://arxiv.org/abs/1409.1556>
115. K. He et al., “Spatial pyramid pooling in deep convolutional networks for visual recognition”, in *European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 346–361. Available: <https://arxiv.org/abs/1406.4729>
116. C.-Y. Lee, P. W. Gallagher, and Z. Tu, “Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree”, in *International Conference on Artificial Intelligence and Statistics*, 2016. Available: <https://arxiv.org/abs/1509.08985v2>
117. Y.-L. Boureau, J. Ponce, and Y. LeCun, “A theoretical analysis of feature pooling in visual recognition”, in *International Conference on*

- Machine Learning (ICML)*, no. 27, pp. 111–118, 2010. Available: <http://yann.lecun.com/exdb/publis/pdf/boureau-icml-10.pdf>
118. T. Mesut, E. Burhan, C. Zafer. Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders. *Medical Hypotheses*, Vol. 135, 2020, 109503, ISSN 0306-9877, <https://doi.org/10.1016/j.mehy.2019.109503>.
119. Y. LeCun et al., “Gradient-based learning applied to document recognition”, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998. Available: <http://yann.lecun.com/exdb/publis/pdf/lecun01a.pdf>
120. K. Kavukcuoglu et al., “Learning convolutional feature hierarchies for visual recognition”, in *Advances in Neural Information Processing Systems 23 (NIPS)*, pp. 1090–1098, 2010. Available: <http://papers.nips.cc/paper/4133-learning-convolutional-featurehierarchies-for-visual-recognition.pdf>
121. Q.V. Le, “Building high-level features using large scale unsupervised learning”, in *International conference on acoustics, speech and signal processing, IEEE*, pp. 8595–8598, 2013. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6639343>
122. M.D. Zeiler and R. Fergus, “Stochastic pooling for regularization of deep convolutional neural networks”, *arXiv preprint arXiv:1301.3557*, Jan. 2013. Available: <https://arxiv.org/abs/1301.3557v1>
123. G.E. Hinton et al., “Improving neural networks by preventing co-adaptation of feature detectors”, *arXiv preprint arXiv:1207.0580v1*
124. N. Srivastava et al., “Dropout: a simple way to prevent neural networks from overfitting”, *Journal of Machine Learning Research*, vol.

- 15, no. 1, pp. 1929–1958, 2014. Available: <https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf>
125. *Noise layers*. Jan. 2017. Available: <http://lasagne.readthedocs.io/en/latest/modules/layers/noise.html#lasagne.layers.DropoutLayer>
126. *tf.nn.dropout*. Dec. 2016. Available: https://www.tensorflow.org/api_docs/python/nn/activation_functions_dropout.
127. Y. Gal and Z. Ghahramani, “Bayesian convolutional neural networks with Bernoulli approximate variational inference”, *arXiv preprint arXiv:1506.02158*, Jan. 2016. Available: <https://arxiv.org/abs/1506.02158v6>
128. L. Wan et al., “Regularization of neural networks using dropconnect”, in *International Conference on Machine Learning (ICML)*, no. 30, pp. 1058–1066, 2013. Available: <http://www.matthewzeiler.com/pubs/icml2013/icml2013.pdf>
129. G. Huang et al., “Deep networks with stochastic depth”, *arXiv preprint arXiv:1603.09382*, Mar. 2016. Available: <https://arxiv.org/abs/1603.09382>
130. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, *arXiv preprint arXiv:1502.03167*, Feb. 2015. Available: <https://arxiv.org/abs/1502.03167>
131. E. Ogasawara, L. C. Martinez, O. Daniel, Silva, M. Mattoso and more. Adaptive Normalization: A novel data normalization approach for non-stationary time series. Proceedings of the International Joint Conference on Neural Networks 2018. 1-8. 10.1109/IJCNN.2010.5596746.

132. X. Wang, L. Wang, and Y. Qiao, “A Comparative Study of Encoding, Pooling and Normalization Methods for Action Recognition”, *Lecture Notes in Computer Science*, vol. 7726, pp. 572–585, Nov. 2013. Available: https://wanglimin.github.io/papers/WangWQ_ACCV12.pdf
133. K. He et al., “Deep residual learning for image recognition”, *arXiv preprint arXiv:1512.03385*, Dec. 2015. Available: <https://arxiv.org/abs/1512.03385v1>
134. M.T. Alexander Mordvintsev and Christopher Olah, *Inceptionism: Going deeper into neural networks*, Jun. 2015. Available: <https://research.googleblog.com/2015/06/inceptionism-going-deeperinto-neural.html>
135. S. Xie et al., “Aggregated residual transformations for deep neural networks”, *arXiv preprint arXiv:1611.05431*, Nov. 2016. Available: <https://arxiv.org/abs/1611.05431v1>
136. G. Huang, Z. Liu, and K.Q. Weinberger, “Densely connected convolutional networks”, *arXiv preprint arXiv:1608.06993*, Aug. 2016. Available: <https://arxiv.org/abs/1608.06993v1>
137. J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization”, *Journal of Machine Learning Research*, vol. 13, pp. 281–305, Feb. 2012. Available: <http://jmlr.csail.mit.edu/papers/volume13/bergstra12a/bergstra12a.pdf>
138. Y.A. LeCun et al., “Efficient BackProp”, *Lecture Notes in Computer Science*, vol. 1524, pp. 9–50. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998. Available: <http://dx.doi.org/10.1007/3-540-49430-8>
139. D. Maclaurin, D. Duvenaud, and R. Adams, “Gradient-based hyperparameter optimization through reversible learning”, in

- International Conference on Machine Learning (ICML)*, 2015, pp. 2113–2122.
140. L. Li et al., “Hyperband: A novel bandit-based approach to hyperparameter optimization”, *arXiv preprint arXiv:1603.06560*, Mar. 2016. Available: <https://arxiv.org/abs/1603.06560>
 141. J. Ortigosa-Hernández, I. Inza, and J.A. Lozano, “Towards competitive classifiers for unbalanced classification problems: A study on the performance scores”, *arXiv preprint arXiv:1608.08984*, Aug. 2016. Available: <https://arxiv.org/abs/1608.08984>
 142. S.J. Nowlan and G.E. Hinton, “Simplifying neural networks by soft weight-sharing”, *Neural computation*, vol. 4, no. 4, pp. 473–493, 1992. Available: <https://www.cs.toronto.edu/~hinton/absps/sunspots.pdf>
 143. M. Thoma, “A survey of semantic segmentation”, *arXiv preprint arXiv:1602.06541*, Feb. 2016. Available: <https://arxiv.org/abs/1602.06541>
 144. D. Ciregan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification”, in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Feb. 2012, pp. 3642–3649. Available: <https://arxiv.org/abs/1202.2745v1>
 145. S. Dieleman, J. De Fauw, and K. Kavukcuoglu, “Exploiting cyclic symmetry in convolutional neural networks”, *arXiv preprint arXiv:1602.02660*, Feb. 2016. Available: <https://arxiv.org/abs/1602.02660>
 146. A. G. Howard, “Some improvements on deep convolutional neural network based image classification”, *arXiv preprint arXiv:1312.5402*, Dec. 2013. Available: <https://arxiv.org/abs/1312.5402>

147. K. He et al., “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification”, in *International Conference on Computer Vision (ICCV)*, pp. 1026–1034, Feb. 2015. Available: <https://arxiv.org/abs/1502.01852>
148. N. Zhang et al., “Part-based R-CNNs for fine-grained category detection”, in *European Conference on Computer Vision (ECCV)*. Springer, pp. 834–849, Jul. 2014. Available: <https://arxiv.org/abs/1407.3867>
149. R. Girshick et al., “Rich feature hierarchies for accurate object detection and semantic segmentation”, in *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2014, pp. 580–587. Available: <https://arxiv.org/abs/1311.2524>
150. A. Krizhevsky and G. Hinton, *Learning multiple layers of features from tiny images*, Apr. 2009. Available: <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
151. *Normalization*. Available: [https://en.wikipedia.org/wiki/Normalization_\(image_processing\)#cite_note-1](https://en.wikipedia.org/wiki/Normalization_(image_processing)#cite_note-1)
152. Kamlesh Lakhwani, P.D. Murarka, Mr. Narendra, “Color Space Transformation for Visual Enhancement of noisy color Image”, *International Journal of ICT and Management*, vol. III, iss. 2, October 2015.
153. *Whitening*. Available: [https://learnmachinelearning.wikia.org/ru/wiki/%D0%94%D0%B5%D0%BA%D0%BE%D1%80%D1%80%D0%B5%D0%BB%D0%B8%D1%80%D1%83%D1%8E%D1%89%D0%B5%D0%B5_%D0%BF%D1%80%D0%B5%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_\(Whitening\)](https://learnmachinelearning.wikia.org/ru/wiki/%D0%94%D0%B5%D0%BA%D0%BE%D1%80%D1%80%D0%B5%D0%BB%D0%B8%D1%80%D1%83%D1%8E%D1%89%D0%B5%D0%B5_%D0%BF%D1%80%D0%B5%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5_(Whitening))

154. *Breast cancer dataset from breakhis*. Available: <https://www.kaggle.com/kritika397/breast-cancer-dataset-from-breakhis>
155. *Bioimaging Challenge 2015 Breast Histology Dataset*. Available: <https://rdm.inesctec.pt/dataset/nis-2017-003>
156. *Breast histopathology*. Available: https://www.kaggle.com/paultimothymooney/breast-histopathology-images#IDC_regular_ps50_idx5.zip
157. *Breast cancer: prevention and control*. Available: <https://www.who.int/cancer/detection/breastcancer/en/>
158. K. Glatz-Krieger, U. Spornitz, A. Spatz, M. Mihatsch, and D. Glatz, “Factors to keep in mind when introducing virtual microscopy”, in *Virchows Archiv*, vol. 448, pp. 248–255, 2006.
159. A. Ljungberg and O. Johansson, “Methodological aspects on immunohistochemistry in dermatology with special reference to neuronal markers”, in *The Histochemical J.*, vol. 25, pp. 735–745, 1993.
160. J.K. Fawcett and J.E. Scott, “A rapid and precise method for the determination of urea”, in *J. of Clin. Path.*, vol. 13, pp. 156–159, 1960.
161. A.C. Ruifrok and D.A. Johnston, “Quantification of histochemical staining by color deconvolution”, in *Anal. and Quant. Cytology and Histology*, vol. 23, pp. 291–299, Aug. 2001.
162. Optimizers. Available: <https://docs.puzzlelib.org/documentation/base/optimizers>
163. M. Naderan, “REVIEW METHODS FOR BREAST CANCER DETECTION USING ARTIFICIAL INTELLIGENCE AND DEEP LEARNING METHODS”, *System Research & Information Technologies*, no. 1, pp. 98-101, 2021.

164. M. Naderan, Yu. Zaychenko, “METHODS FOR IMPROVING ACCURACY OF THE DEMENTIA DIAGNOSIS USING FEATURE DIMENSION REDUCTION”, *System Research & Information Technologies*, no. 2, pp. 25–30, 2019.