

Є. Ю. Вайвала, Н. В. Цьопа

Національний технічний університет України «КПІ імені Ігоря Сікорського», Київ, Україна

## ПІДВИЩЕННЯ ТОЧНОСТІ РОЗПІЗНАВАННЯ ОБІГОВИХ МОНЕТ ЗА РАХУНОК ВИКОРИСТАННЯ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ З ДЕКІЛЬКОМА ВИХОДАМИ

**Анотація:** У роботі розглянуто проблему розпізнавання зображень обігових монет за допомогою згорткових нейронних мереж. Описано традиційний підхід до вирішення задачі розпізнавання зображень, який передбачає використання звичайної згорткової нейронної мережі з одним виходом, показано та проаналізовано отримані результати. Для підвищення точності розпізнавання було застосовано архітектуру згорткової нейронної мережі з декількома виходами. Отримані результати були порівняні із результатами роботи звичайної мережі, були наведені причини відмінностей у результатах та переваги і недоліки кожного із розглянутих підходів.

**Ключові слова:** згорткова нейронна мережа, нейронна мережа з декількома виходами, розпізнавання зображень, машинне навчання.

### Вступ

**Постановка проблеми.** Навіть на сьогоднішній час, за умови стрімкого розвитку безготівкового розрахунку, випуск металевих монет як для звичайного користування, так і виключно для колекціонування не припиняється. Тому необхідність опізнати ту чи іншу монету може виникнути як у колекціонерів, так і у зацікавлених людей. Для початківців та інших малодосвідчених осіб така задача може стати суттєвою проблемою. У такому випадку найлегшим рішенням є розпізнавання монети за її зображеннями із допомогою програмних засобів, що нададуть користувачеві базову інформацію про монету за її зображеннями з двох боків.

Сьогодні найкращим підходом до розпізнавання зображень вважається використання згорткової нейронної мережі. У даній роботі представлено результати порівняння роботи традиційної загорткової нейронної мережі та згорткової нейронної мережі з декількома виходами для розпізнавання монет на прикладі датасету зображень, що містить фотографії сучасних обігових монет деяких країн Європи.

**Аналіз останніх досліджень і публікацій.** Традиційний підхід до розпізнавання зображень полягає у використанні звичайної згорткової нейронної мережі, яка приймає на вхід певне зображення та повертає вектор чисел, що відображають впевненість нейронної мережі у відповідності зображення кожному

із заданих класів [1-6]. При такому підході кожне зображення у тренувальному датасеті визначається одним і лише одним класом, який повинен повною мірою охарактеризувати дане зображення і визначити відмінності від інших. Для розпізнавання обігових монет кожному зображенню монети ставиться у відповідність характеристика, що містить номінал, грошову одиницю та країну монети. Таким чином отримані класи виду «1 копійка, Україна» для кожного із зображень. Для країн Європи можна отримати декілька сотень таких класів, що при невеликій кількості зображень у датасеті призведе до досить малого обсягу зображень у кожному окремому класі.

Традиційна згорткова нейронна мережа (рис. 1) складається із послідовного чергування шарів згортки (convolutional layer) із функцією активації ReLU, які являють собою набір фільтрів, що після проходження по зображенню формують карти ознак для кожного фільтра, і агрегування (pooling), які виконують підвибірку і слугують для скорочення розміру представлення. На виході такої мережі розміщено один або декілька звичайних повноз'язних нейронних шарів для класифікації зображення [7].

Для більшості задач розпізнавання зображень звичним підходом є використання готової нейронної мережі, наприклад VGG16 або ResNet50, що була навчена на наборі зображень ImageNet, для переносу навчання з великого датасету на менший [8].

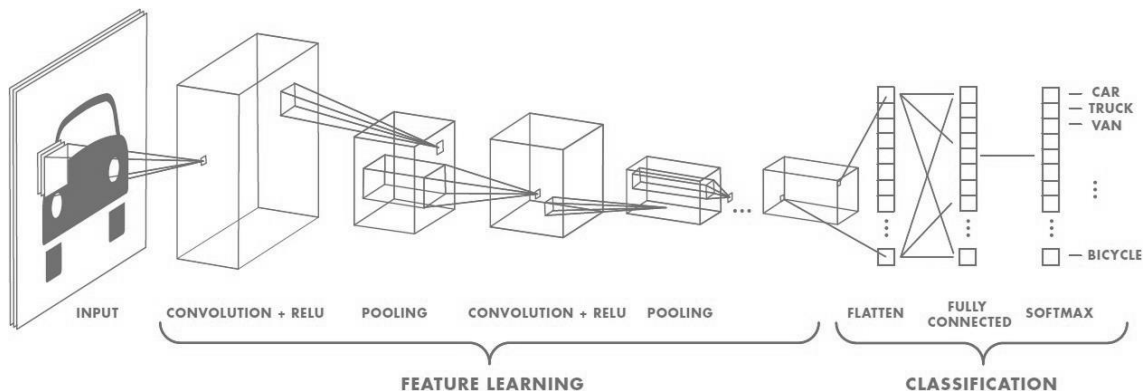


Рис. 1. Традиційна згорткова нейронна мережа

У даній роботі, з огляду на специфіку вхідних даних і обмеженість ресурсів для навчання нейронних мереж, використання таких мереж не виправдує себе і виникає необхідність у проектуванні власної нейронної мережі.

**Формулювання мети статті.** Мета даного дослідження полягає у побудові згорткової нейронної мережі для розпізнавання обігових монет країн Європи. Задля пошуку найкращого рішення розглянуто традиційний підхід до розпізнавання зображень та внесено зміни до моделі нейронної мережі з метою підвищення точності розпізнавання зображень з урахуванням особливостей вхідних даних.

Для якісного навчання будь-якої згорткової нейронної мережі необхідно забезпечити якомога більший тренувальний датасет розмічених зображень, тобто зображень, яким поставлені у відповідність певні класи.

Для переважної більшості задач розпізнавання зображень неважко знайти готові датасети на спеціалізованих сайтах у мережі Інтернет, проте для даного дослідження, за відсутності готового датасету, виникає необхідність власноруч розмітити зображення, що знаходяться у вільному доступі у мережі Інтернет та власні фотографії.

### Виклад основного матеріалу

Для даного дослідження було зібрано датасет, що містить більше ніж 40000 зображень монет, до нього увійшли сучасні обігові монети України, Польщі, Великобританії та країн Євросоюзу (Австрія, Андорра, Бельгія, Ватикан, Греція, Ірландія, Іспанія, Італія, Кіпр, Латвія, Литва, Люксембург, Мальта, Монако, Нідерланди, Німеччина, Португалія, Сан-Марино, Словаччина, Словенія, Фінляндія, Франція, Естонія), усього монети належать до 26 країн. Усі зображення були приведені до спільного вигляду і мають ширину 300 пікселів і висоту 150 пікселів, оскільки кожне з них містить два квадратних (150\*150 пікселів) зображення, на яких подано фотографії монети із двох сторін на білому або світло-сірому фоні.

Датасет було розділено на дві частини, 80% від загальної кількості становлять тренувальні дані, на яких мережа навчатиметься, а 20% – тестові, яких мережа «не бачила», та на яких перевірялася її робота. Під час навчання нейронної мережі тренувальні дані

було знову розділено на 80% виключно для тренування та 20% для валідації під час кожної епохи навчання, що дозволило швидко визначити перенавчання.

Через специфічні особливості вхідних даних, кількість тренувальних зображень для монет кожного класу суттєво залежить від часу, протягом якого вона випускалася та тиражів випуску. Як наслідок, найбільшим недоліком зібраного датасету є нерівномірність розподілення зображень за класами, наприклад, найменше усього зображень зібрано для монет невеликої Андорри, а найбільше – для монет Німеччини, які випускаються одразу на 5 монетних дворах дуже великими тиражами. Загалом, зображення були розподілені на 212 класів виду «номінал, грошова одиниця, країна». У датасеті наявні 8 грошових одиниць: копійка, гривня, грош, злотий, пенні, фунт, євроцент та євро і 8 номіналів: 0.500, 1, 2, 5, 10, 20, 25 та 50.

Також до датасету було включено як звичайні обігові монети, так і пам'ятні монети, що були випущені в обіг меншими тиражами, це такі монети, як, наприклад, пам'ятні українські гривні та пам'ятні монети номіналом 2 євро у країнах Євросоюзу. Останні досить сильно ускладнювали процес розпізнавання, адже усі монети євро мають спільний дизайн реверсу, і, хоча звичайні монети мають індивідуальні дизайни аверсу для монет кожної країни і інколи зазначення назви країни, на деяких пам'ятних монетах вказівки на країну походження є досить непомітними і для людини.

Для роботи із нейронною мережею було застосовано середовище Google Colab, яке надає потужності для створення, навчання, тестування та використання нейронних мереж. Блокноти Google Colab надають можливість працювати з відкритою бібліотекою для машинного навчання TensorFlow та Keras – надбудовою над нею. TensorFlow підтримує декілька популярних мов програмування, проте найбільш зручним та оптимальним вибором є саме мова Python, яка зазвичай використовується при розробці нейронних мереж.

Враховуючи специфіку вхідних даних, які передбачають прямокутне (300\*150 пікселів) зображення замість традиційно квадратного та вузько стандартизоване подання зображення (дві сторони монети на білому або світло-сірому фоні), і обмеженість ресурсів було вирішено застосувати власну нейронну мережу, що задається кодом на рис. 2.

```
model = keras.models.Sequential([
    keras.layers.experimental.preprocessing.Rescaling(1./255, input_shape=(150, 300, 3)),
    keras.layers.Conv2D(64, 3, activation='relu'),
    keras.layers.MaxPooling2D(),
    keras.layers.Conv2D(128, 3, activation='relu'),
    keras.layers.MaxPooling2D(),
    keras.layers.Conv2D(256, 3, activation='relu'),
    keras.layers.MaxPooling2D(),
    keras.layers.Flatten(),
    keras.layers.Dense(512, activation='relu'),
    keras.layers.Dense(256, activation='relu'),
    keras.layers.Dense(212)
])
```

Рис. 2. Задання нейронної мережі

Дана нейронна мережа є досить простою і містить три пари шарів згортки, що містять 64, 128 та 256 фільтрів розміру 3\*3 і чергуються з шарами максимізаційного агрегування (max pooling), що містять фільтри розміром 2\*2, які віднаходять і передають на наступний шар максимальне значення. На виході нейронної мережі розміщено 2 повноз'язних шари з функцією активації ReLU та кількістю нейронів 512 і 256 і один повноз'язний шар із 212 нейронів для класифікації. Граф отриманої нейронної мережі подано на рис. 3.

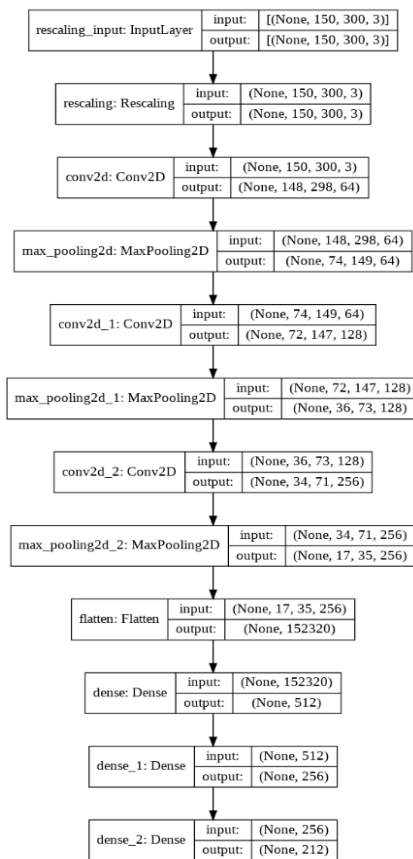


Рис. 3. Граф нейронної мережі

Результати роботи мережі, отримані після навчання протягом 10 епох у вигляді графіків точності розпізнавання та функції втрат на тренувальних та валідаційних даних (рис. 4).

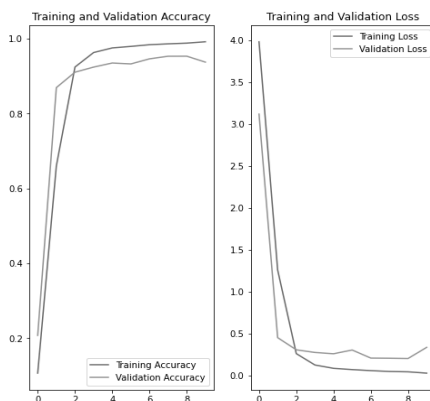


Рис. 4. Графік точності розпізнавання та функції втрат

За характером графіків валідаційних даних можна зробити висновок, що збільшення кількості епох призведе лише до перенавчання, коли мережа починає «завчати» тренувальні дані, збільшуючи при цьому точність їх розпізнавання, при майже незмінній точності на валідаційних даних. Точність розпізнавання на тестових даних становила 0,9365 (рис. 5).

```

    254/254 [=====] -
    18s 69ms/step - loss: 0.3274 - accuracy: 0.9365
    [0.32737427949905396, 0.9365196824073792]
  
```

Рис. 5. Результати тестування нейронної мережі

Протягом роботи над нейронною мережею було проведено декілька експериментів з її ускладненням шляхом додавання додаткових шарів та нейронів, проте таке ускладнення себе не виправдало, адже точність розпізнавання тестових даних не зазнала суттєвого збільшення, залишаючись у межах 93-94%, при цьому значно збільшився обсяг необхідних обчислювальних ресурсів та часу на навчання.

При аналізі причин досить низької точності використання класичної згорткової нейронної мережі можна помітити одну із особливостей даних – поточне визначення характеристик обігових монет («номінал, грошова одиниця, країна») надає досить вузьку класифікацію зображень і велику кількість класів (а саме 212), в межах кожного з яких є досить мало зображень, а мала кількість зображень при багатьох класах призводить до зменшення точності навчання.

При подальшому розгляді можна помітити, що кожен клас є складеним і його можна розділити на три окремих характеристики: номінал, грошову одиницю та країну. Кожна із таких характеристик окремо не є достатньою для однозначного визначення певної монети, проте не важко побачити, що при розпізнаванні монет, наприклад, виключно за номіналом, можна отримати значно меншу кількість класів для такої ж кількості тренувальних зображень, що призведе до більшої кількості зображень в межах кожного окремого класу, а отже і до кращих результатів розпізнавання за даною характеристикою. Це стосується і розпізнавання монет виключно за грошовою одиницею чи країною. При цьому, в той час, як розширення датасету за рахунок додавання, наприклад, зображень сучасних обігових монет ще однієї країни, до комбінованої характеристики необхідно додати близько десяти класів, до переліку класів країн можна додати лише один клас, а до переліку класів грошових одиниць та номіналів – від нуля до двох класів.

У випадку, коли невеликий за розмірами датасет містить велику кількість складених класів, можна звернутися до менш поширеної архітектури згорткової нейронної мережі із декількома виходами, що передбачає отримання на вхід зображення і повернення декількох векторів відповідностей, по одному для кожної із декількох класифікацій. Нейронна мережа з декількома виходами реалізується за рахунок розгалуження нейронної мережі на певному етапі, що призводить до утворення двох або більше гілок. Такі розгалуження можна утворювати на будь-якому етапі

нейронної мережі. Зазвичай таку мережу використовують тоді, коли виникає необхідність отримати дані різних форматів за допомогою різних підходів до їх обробки, наприклад, класифікацію та результат регресійного аналізу, проте вона є доречною і для зображень, які можуть мати декілька непов'язаних характеристик.

Під час розпізнавання обігових монет за допомогою згорткової нейронної мережі із декількома виходами можна отримати три вектори відповідностей, по одному для номіналу, грошової одиниці та країни, при цьому точність розпізнавання для кожної із характеристик буде суттєво більшою за точність розпізнавання звичайною мережею із одним виходом з застосуванням складених характеристик у якості класів. Ще однією з переваг такого підходу є можливість більш тонкого налаштування нейронної мережі та економії ресурсів при навчанні, якщо після розгалуження у якості шарів гілок застосувати попередньо навчені моделі, адже інколи моделі для різних харак-

теристик можуть навіть потребувати різної кількості epoch тренування. У даній роботі представлено нейронну мережу із розгалуженням на три гілки на початку мережі, причому кожна із гілок представлена одним шаром, який являє собою звичайну згорткову модель. Для загальної нейронної мережі відмінності між певною послідовністю шарів та моделлю, яка являє собою таку саму послідовність шарів не виникне, проте це дозволить окремо навчити певну гілку та використати отримані ваги у загальній нейронній мережі. Код загальної нейронної мережі подано на рис. 6, а представлення у вигляді графу – на рис. 7.

```
input = keras.layers.Input((150, 300, 3))
value_layer = value_model(input)
currency_layer = currency_model(input)
country_layer = country_model(input)
new_model = keras.models.Model(inputs = input,
input, outputs = [value_layer, currency_layer, country_layer])
```

Рис. 6. Задання нейронної мережі з трьома виходами

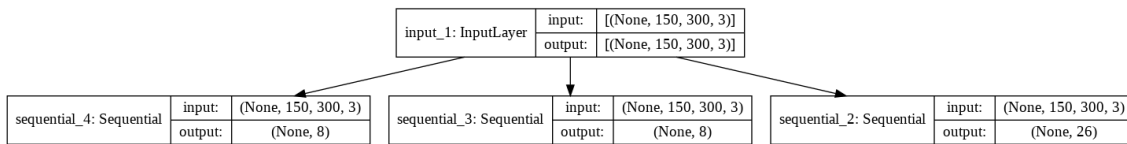


Рис. 7. Граф нейронної мережі з трьома виходами

Моделі `value_model`, `currency_model` та `country_model` використані у якості згрупованих шарів і являються моделями для розпізнавання номіналу, грошової одиниці та країни відповідно. Структура цих моделей подібна до традиційної моделі згорткової нейронної мережі і відрізняються зменшеною кількістю нейронів останнього шару для класифікації – 8 для моделей номіналу та грошової одиниці і 26 для моделі країни. Навчання кожної із моделей відбувалося окремо, причому розподіл даних на трен-

увальні та тестові залишився не змінним, в той час як розподіл даних на виключно тренувальні та валідаційні відрізнявся для різних моделей. Результати навчання моделі для розпізнавання номіналу, грошової одиниці та країни подані на рис. 8 у вигляді графіків точності розпізнавання та функції втрат на тренувальних та валідаційних даних. Точність розпізнавання на тестових даних становила 0,9941 для номіналу, 0,9989 для грошової одиниці та 0,9641 для країни (рис. 9).

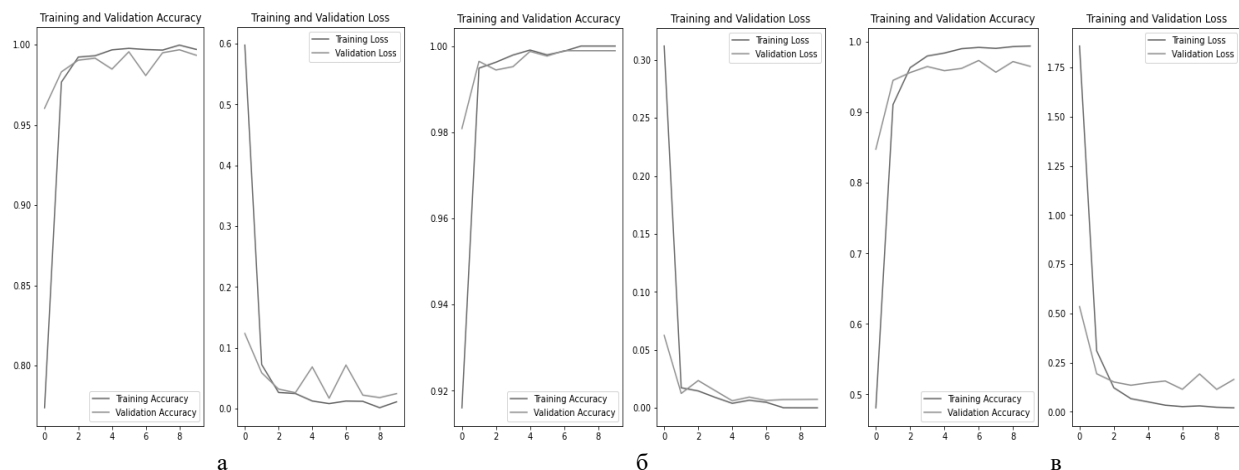


Рис. 8. Графіки тренування моделі розпізнавання: а – номіналу; б – грошової одиниці; в – країни

За отриманими графіками видно, коли для кожної окремої моделі починається перенавчання, що дозволяє визначити необхідну кількість epoch навчання для кожної окремої моделі.

Проаналізувавши отримані результати для тестових даних можна помітити, що точність розпізнавання номіналу та грошової одиниці є досить висо-

кою, приблизно 99%, а точність розпізнавання країн менша, у межах 96-97%. Менша точність пояснюється, зокрема більшою кількістю класів та специфікою вхідних даних, зокрема наявністю у датасеті зображень пам'ятних монет номіналом 2 євро ускладнює визначення приналежності цих монет до певної країни Євросоюзу.

254/254 [=====] - 18s 69ms/step - loss: 0.0236 - accuracy: 0.9941  
[0.023559754714369774, 0.994071900844574]

а

254/254 [=====] - 18s 70ms/step - loss: 0.0035 - accuracy: 0.9989  
[0.003485651919618249, 0.9988884925842285]

б

254/254 [=====] - 18s 70ms/step - loss: 0.1752 - accuracy: 0.9641  
[0.1752026081085205, 0.9640607833862305]

в

Рис. 11. Результати тестування моделі розпізнавання: а – номіналу; б – грошової одиниці; в – країни

Неважко побачити, що точність розпізнавання номіналу, грошової одиниці та країни окремо вища за результат, отриманий при застосуванні традиційної мережі, при цьому ми із досить високою вірогідністю отримуємо правильно визначений номінал та грошову одиницю.

Якщо визначити загальну точність розпізнавання зображень як добуток результатів кожної з гілок, ми отримаємо підсумкову точність величиною 0,9574, що перевищує точність традиційної згорткової нейронної мережі на 0,0209, що є досить непоганим приростом. При більш гнучкому налаштуванні і тренуванні кожної окремої моделі, що входять до складу згорткової нейронної мережі з трьома виходами, можна і надалі покращити отриманий результат.

### Висновки

Задача розпізнавання монет за зображенням через свою специфіку дозволяє застосувати декілька підходів до розбиття зображень на класи, а отже і до побудови згорткової нейронної мережі. У роботі було наведено традиційний підхід, коли кожне зображення визначається одним складеним класом, що

одночасно вказує на номінал, грошову одиницю та країну і нейронна мережа має один вихід та підхід із використанням нейронної мережі з декількома виходами, коли кожна із характеристик зображення являє собою окрему класифікацію і нейронна мережа має три виходи, по одному для кожної із них.

Через велику кількість малих за обсягом зображень класів при традиційному підході результат розпізнавання виявився не найкращим, 93-94%, при цьому розгалуження нейронної мережі на три гілки для кожної окремої характеристики надало хороший результат у 99% для номіналу та грошової одиниці і 95-96% для країни. Перевагою такого підходу є можливість більш гнучко налаштувати та навчити окремі моделі для кожної характеристики та використати їх у якості шарів окремих гілок нейронної мережі. Недоліком такого підходу є в декілька разів більша кількість параметрів нейронної мережі, що означає суттєве збільшення часу на її навчання та використання ресурсів, при цьому окреме тренування вкладених моделей дозволить зменшити кількість ресурсів, що одночасно задіяні під час навчання мережі.

### СПИСОК ЛІТЕРАТУРИ

1. Хайкин Саймон. Нейронные сети: полный курс, 2-е изд. : пер. с англ. / Саймон Хайкин. – М. : Издательский дом «Вильямс», 2006. – 1104 с.
2. Ямпольский Леонид Стефанович. Нейротехнології та нейрокомп'ютерні системи : підручник / Л.С. Ямпольский, О.І. Лісовиченко, В.В. Олійник. – К. : Дорадо-Друк, 2016. – 576 с.
3. A Survey on Multi-output Learning / [D. Xu, Y. Shi, I. Tsang та ін.]. – 2019. – 21 с.
4. A survey on multi-output regression / H.Borchani, G. Varando, C. Bielza, P. Larranaga. – 2015. – 27 с.
5. CS231n Convolutional Neural Networks for Visual Recognition [Електронний ресурс] – Режим доступу до ресурсу: <https://cs231n.github.io>.
6. Keras API reference [Електронний ресурс] – Режим доступу до ресурсу: <https://keras.io/api/>.
7. LeCun Y. Convolutional Networks for Images, Speech, and Time-Series / Y. LeCun, Y. Benigo. – 1995. – 14 с.
8. Module: tf | TensorFlow Core v2.6.0 [Електронний ресурс] – Режим доступу до ресурсу: [https://www.tensorflow.org/api\\_docs/python/tf?hl=en](https://www.tensorflow.org/api_docs/python/tf?hl=en).

Received (Надійшла) 15.05.2021

Accepted for publication (Прийнята до друку) 21.07.2021

### Improving the accuracy of circulating coins recognition by using a convolutional neural network with multiple outputs

Ye. Vaivala, N. Tsyopa

**Abstract.** The article considers the problem of circulating coins recognition using convolutional neural networks. A traditional approach to solving the problem of image recognition, which involves the use of a regular convolutional neural network with one output, is described, the results are shown and analyzed. To improve the recognition accuracy, the architecture of a convolutional neural network with multiple outputs was used. The results obtained were compared with the results of a regular network, the reasons for the differences in the results and the advantages and disadvantages of each of the considered approaches were given.

**Keywords:** convolutional neural network, multi-output neural network, image recognition, machine learning.