

УДК 004.42

DOI <https://doi.org/10.32782/2663-5941/2022.6/19>**Марчук Д.К.**

Державний університет «Житомирська політехніка»

**Левківський В.Л.**

Державний університет «Житомирська політехніка»

**Марчук Г.В.**

Державний університет «Житомирська політехніка»

**Голенко М.Ю.**

Viseven Ukraine

## СИСТЕМА РОЗПІЗНАВАННЯ ДАКТИЛЬНОЇ МОВИ УКРАЇНСЬКОЇ АБЕТКИ

У даній роботі розглянуто основні етапи розробки нового програмного забезпечення для розпізнавання дактильної мови. Дактильна мова – це можливість спілкування за допомогою жестової мови. Кожна літера зображується певним положенням пальців руки, що позначає дактильний знак – дактилему. Зазвичай, це мова якою спілкуються люди з порушенням слуху чи мовлення. У більшості таких абеток використовується одна рука. Розроблений програмний засіб може проводити розпізнавання букв як правої так і лівої рук. Об'єктом дослідження є методи і процес відстеження та розпізнавання динамічних жестів людини з застосуванням машинного навчання. Для розпізнавання дактилем використано алгоритм на основі скелета – обчислення проводяться за віртуальним скелетом людини, а частини тіла діляться на певні сегменти. Це реалізовано через метод MediaPipe, який є кросплатформним конвеєром машинного навчання, що забезпечує продуктивність у реальному часі. В роботі використано особливий різновид архітектури рекурентних нейронних мереж, здатний до навчання, а саме модель довготривалої короткочасної пам'яті LSTM (Long short-term memory), який довів свою ефективність. Для навчання і тренування мережі створено унікальний набір даних для української дактильної абетки. В роботі також розглянуто результати навчання декількох моделей з різними функціями активації. Проаналізовано та досліджено показники втрат і точності. Проведено опис інтерфейсу при роботі з програмним продуктом і зазначено варіанти та можливості його використання. Було досліджено ефективність роботи програмного забезпечення у реальних умовах, перевірено результат якості та точності розпізнавання української дактильної абетки та створення речень з перекладених знаків.

**Ключові слова:** розпізнавання, дактильна мова, дактилема, нейронна мережа, машинне навчання, RNN.

**Постановка проблеми.** Дактильною мовою користуються люди з особливими потребами для обміну інформацією між собою та іншими людьми. Мова має свою абетку, словник і синтаксис, який суттєво відрізняється від звичайної. Навіть з цією альтернативною формою спілкування дуже складно спілкуватися у широкій громадськості, оскільки розуміння мови жестів потребує тривалого навчання та практики. Як наслідок, цей розрив створює значну соціальну ізоляцію серед людей з вадами слуху та мови в суспільстві, в якому вони співіснують. Щоб подолати цей розрив, потрібна надійна допоміжна технологія, яка може розпізнавати дактильну мову і перекладати її.

Саме тут методи розпізнавання та машинного навчання відіграють важливу роль допомагаючи автоматизувати процес ідентифікації жестів. На даний момент часу систем розпізнавання української дактильної мови та алфавіту у відкритому доступі немає. Так само, як і даних для машинного навчання нейронних моделей.

**Аналіз останніх досліджень і публікацій.** На сьогодні методи штучного інтелекту застосовуються у різних сферах нашого життя. Згорткові нейронні мережі є одним із найпоширеніших типів нейронних мереж, які використовуються в комп'ютерному баченні. Наприклад в роботі [1] використовується згорткова нейронна мережа для розпізнавання рукописного тексту. У статті [2]

розглянуто нову систему для фільтрації спаму та реклами, яка надає засоби для швидкого визначення та фільтрації спам-контенту, реклами та контенту для дорослих поєднуючи класичні методи із сучасними, які базуються на використанні нейронних мереж.

Рекурентні нейронні мережі мають багато застосувань. У роботі [3] запропоновано програмну реалізацію з використанням методів штучного інтелекту для пошуку вільного місця на парковці. Розроблена система надає змогу користувачам авто здійснювати пошук вільних місць, витрачаючи при цьому мінімум часу. У статті [4] описано основні алгоритми аналізу потоку кадрів відеоданих, що надходять з камер міста. Основною метою дослідження є мінімізація часу на пошук вільного місця для паркування автомобіля.

В роботі [5] проведено аналіз нейронної мережі Хопфілда, мереж Елмана і Джордана, ESN-мережі, рекурсивної мережі і рекурентної мережі з довгою короткостроковою пам'яттю для безпосереднього визначення оптимальної архітектури мережі для обробки великих текстових даних

Наприклад, рекурентні нейронні мережі використовувалися для прогнозування курсу акцій [6], машинного перекладу [7] і створення музики [8]. Месюра В. та Гранік М. [9] застосовують методи штучного інтелекту для аналізу новин та доводять ефективність у вирішенні проблеми виявлення фейкових новин.

**Формулювання цілей статті.** На основі аналізу предметної області постає необхідність дослідження методів розпізнавання дактильної мови та розробці програмного забезпечення. Для реалізації цієї мети були поставлені такі завдання:

- провести аналіз методів відстеження та розпізнавання динамічних жестів людини.
- створити набір даних для розпізнавання літер дактильної абетки;
- розробити обчислювально-ефективну модель і алгоритм розпізнавання динамічних жестів людини.
- провести експерименти по оцінці надійності і працездатності системи в реальному часі.

**Виклад основного матеріалу.** Українська жестова мова – природна візуально-жестова мовна система з власною лексико-граматичною структурою, що сформувалась еволюційним шляхом і використовується як основний або один з основних засобів спілкування жестомовних осіб, які постійно проживають або впродовж тривалого часу проживали на території України. Крім цього, зазначають, що українська жестова мова є мовою

спільноти жестомовних осіб, до яких відносять людей із порушеннями слуху чи мовлення.

Українська дактильна абетка допоміжна система української жестової мови, в якій кожному жесту однієї руки (зазвичай нерухомою правою, зігнутою в лікті) відповідає літера української мови; багато дактилем зовні схожі на відповідні літери української абетки.

Сучасна українська дактильна абетка нараховує 33 дактильних знаки, тобто стільки ж, скільки літер в українській абетці [10]. Кожна дактилема відповідно позначає графему і відтворюється згідно з мовленнєвою нормою. Залежно від типу вхідних даних, підхід до інтерпретації жесту може здійснюватися різними способами. Однак більшість методів спираються на ключові покажчики, представлені в тривимірній системі координат. На основі їх відносного руху жест можна виявити з високою точністю, залежно від якості введення та обраному методу.

Щоб інтерпретувати рухи тіла, потрібно класифікувати їх за загальними властивостями та повідомленням, яке рухи можуть виражати. Наприклад, на мові жестів кожен жест представляє слово чи фразу.

У літературі розрізняють 2 різні підходи до розпізнавання жестів: на основі 3D-моделі та на основі зовнішнього вигляду. Перший метод використовує тривимірну інформацію про основні елементи тіла, щоб отримати кілька важливих параметрів, як положення долоні або орієнтація суглобів руки. З іншого боку, такі системи на основі зовнішнього вигляду використовують зображення або відео для безпосередньої його інтерпретації.

Для точності будь якої нейронної мережі необхідна велика кількість даних для її навчання. Нажаль, для української жестової мови такі дані відсутні у вільному доступі. Для реалізації програмного продукту створимо власний набір даних для розпізнавання.

Для створення набору даних використаємо замість зображень відеозаписи, які в подальшому будуть розбиватись на 30 кадрів. Набір даних буде складатись з 36 класів, з яких 33 це знаки дактильної абетки, які відповідають кількості у звичайному українському алфавіті та 3 спеціальні класи — фон, вставка пробілу між словами і знак видалення літери. В кожному класі буде 100 відеозаписів розбитих на 30 кадрів, що в результаті становитиме 3000 кадрів. Абетка містить як статичні жести, так і динамічні, саме тому буде доречніше використовувати відеозаписи з орієнтирами

MediaPipe для кожного кадру ніж просто набір даних з зображеннями.

Для створення датасету було написано програмний додаток на мові Python з використанням MediaPipe і OpenCV. Під час запису відео MediaPipe малює орієнтири безпосередньо на зображенні.

Процес навчання вимагає створення, класифікації та перемішування даних, щоб вони були подані в конвеєр машинного навчання. Саме тому всі орієнтири з зображення зберігаються у масив бібліотеки NumPy. Після чого такі масиви з координатами орієнтирів для кожного кадру зберігаються у файли з розширенням .NPY для майбутнього навчання моделі. Замість потоку з веб камери OpenCV, на вхід можна передавати і звичайні відео, що суттєво може зменшити витрати у часі. На кожний клас необхідно записати приблизно ~100 відео для забезпечення необхідної кількості даних для навчання моделі.

Записані відео і координати з них мають бути різноманітними та записаними з різних відстаней та положень для покращення кінцевого результату. Відео необхідно записувати окремо для кожної руки, щоб мати можливість розпізнавати жести з різних сторін, що лише збільшує мінімальну необхідну кількість відеозаписів. MediaPipe інколи видає хибні дані у розпізнаванні координат з відео і якщо під час відео трапляються помилки у відображенні орієнтирів, то вони обов'язково залишають відбиток і додають хибних даних у значення записуваного класу.

Для навчання алгоритму розпізнавання була використана модель рекурентної нейронної мережі. Рекурентні нейронні мережі є найкращою підходящою архітектурою, оскільки мова жестів являє собою послідовність рухів рук і алгоритм розпізнавання повинен зберігати та використовувати інформацію послідовно. Вхідними даними для моделі є вектор, що складається з 63 точок даних. Кожна рука має 21 орієнтир, точка з коор-

динатами x, y і z створена для двох рук. Оскільки в деяких відео було видно лише одну руку, ліві точки даних були заповнені нулями.

Спроектвана нейронна мережа складається з 6 шарів, 3 з яких приховані шари. Модель буде використовувати змішані функції активації, адже під час дослідження та експериментів, вони видали найкращий результат порівняно з іншою конфігурацією, коли всі шари використовували функцію активації ReLU. Саме тому перші три шари використовують функції активації з параметром activation=tanh та рекурентну активацію з параметром recurrent\_activation=sigmoid. Модель з даними параметрами не тільки покращила свою роботу в декілька разів за швидкістю та точність, а й ще показала набагато кращий результат порівняно з іншими спробами тренування. На шарах Dense функції активації залишились ReLU і у останньому шарі softmax.

Дані навчання були розділені на дві частини: навчання та перевірка. Використані дані перевірки становили 10% від усіх даних. Остаточне навчання проводилося для 1000 епох, що зайняло приблизно декілька годин з використанням графічного процесору у Google Collabaratory. Остаточна досягнута точність склала 92,54%.

Для дослідження якості та точності навченої моделі використовуємо Tensorboard. Tensorboard дозволяє у реальному часі за допомогою інтерактивного веб-інтерфейсу відстежувати процес навчання. На графіках помаранчева лінія відповідає за тренування моделі, а блакитна за валідацію даних. Графік (рис. 1) відображає зростання точності за кількістю епох, вона без помилок починає впевнено видавати точність після 70–80 епох з точністю до 0.9–0.98.

Графік (рис. 2) відображає після якої епохи втрати під час навчання моделі почали зменшуватись. Як бачимо на рисунку втрати після 70 епох значно зменшились від значення 1.8 до 0.2–0.1.

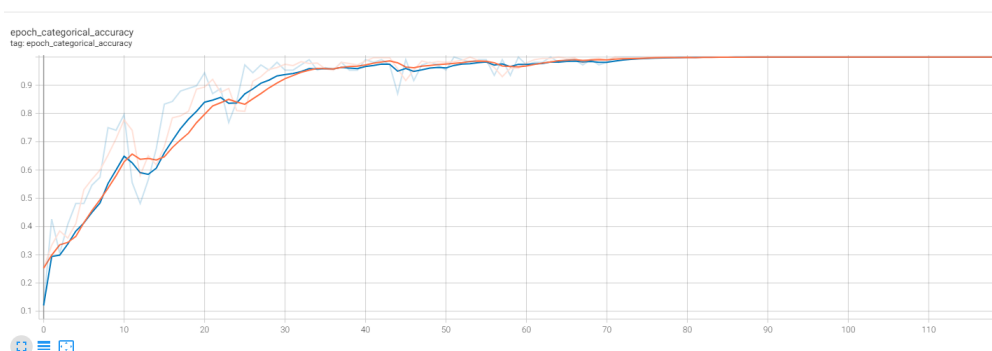


Рис. 1. Результати останньої навченої моделі зі зростанням точності після 70-80 епох

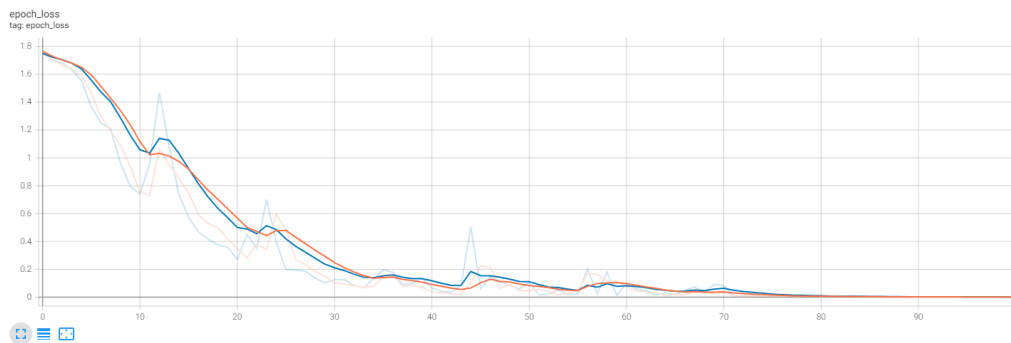


Рис. 2. Результати останньої навченої моделі зі зменшенням втрат після 70-80 епох

Наступні два графіки (рис. 3 та рис. 4) відображають оцінки точності та втрат навченої моделі. На них можна помітити, що стабільна точність та зменшення втрат було досягнуто після 300 епох.

За результатами навчання можна зробити висновок, що завдяки вибору коректних функцій активації точність та швидкість моделі навчання значно зросла.

На основі створеної моделі було розроблено програмне забезпечення для розпізнавання української дактильної мови. Початкове меню про-

грами надає користувачам можливість вибрати необхідний варіант роботи з програмою в залежності від їх потреб – створити власний набір даних або одразу перейти до розпізнавання дактильної абетки (рис. 5).

При натисканні кнопки створення датасету відкривається вікно, у якому відображається зображення з веб-камери. На камеру потрібно показати дактильний знак, для його запису для майбутнього розпізнавання. Після успішного запису необхідної кількості відео, вікно автоматично закривається

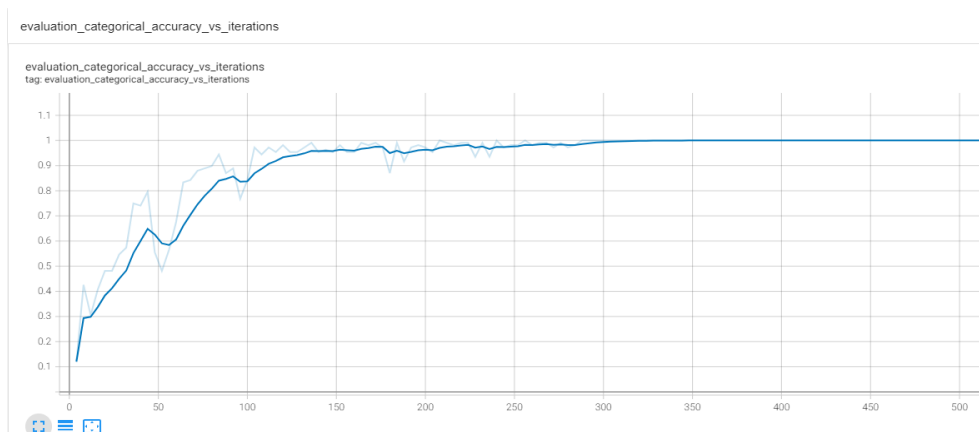


Рис. 3. Результати оцінки останньої навченої моделі зі зменшенням втрат після 250 епох

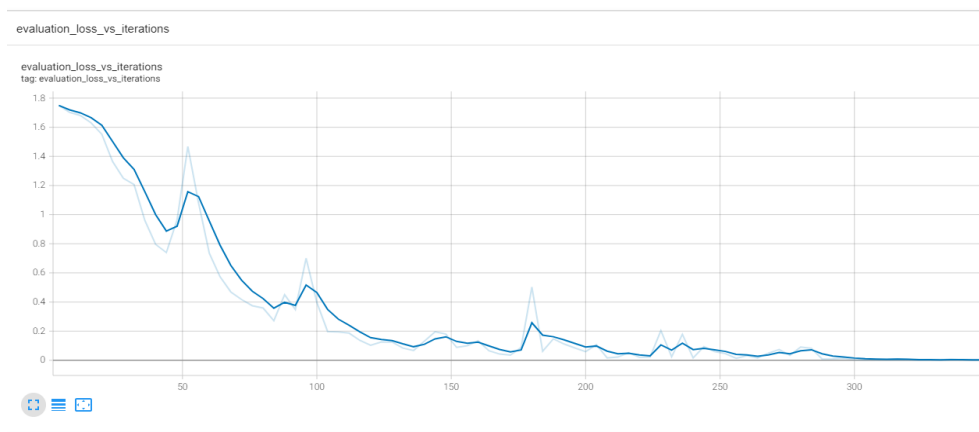


Рис. 4. Результати оцінки останньої навченої моделі зі зменшенням втрат після 290



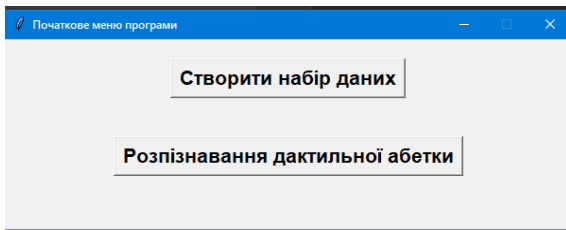


Рис. 5. Початкове меню програми

та програма готова до створення набору даних для іншого класу.

Відкривши програму в режимі розпізнавання дактильної абетки з'явиться головне вікно програми продемонстроване на рисунку 6. Червоною рамкою виділено область, де відображається зображення з веб-камери. У даній області також буде відображатись вірогідність розпізнавання і коректно розпізнаний знак дактильної абетки.

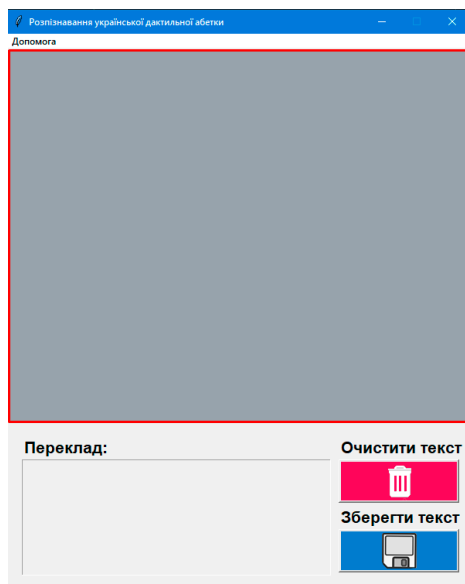


Рис. 6. Вікно розпізнавання дактильної абетки

Основне поле з результатом перекладеного знака дактильної мови у літеру по замовчужанню є заблокованим для користувача і вмикається лише при розпізнаванні і внесенні результату.

Зверху біля назви вікна програми можемо побачити меню з назвою "Допомога". Якщо скористатись ним, то з'явиться контекстне меню з двома параметрами "Дактильна абетка" та "Вийти з програми".

При натисканні на пункт меню "Дактильна абетка" відкривається нове вікно з зображенням всіх літер української дактильної мови для вивчення знаків дактильної абетки.

Було проведено тестування для перевірки ефективності використання додатка. На рисун-

ках 7 та 8 зображено результати розпізнавання деяких літер дактильної абетки.

При послідовному відображенні знаків програма зможе скласти літери в слова, які в подальшому можна зберегти до текстового документу(рис. 9).

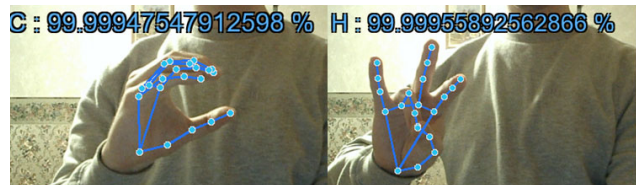


Рис. 7. Результати розпізнавання літер С та Н



Рис. 8. Результати розпізнавання літер В та А

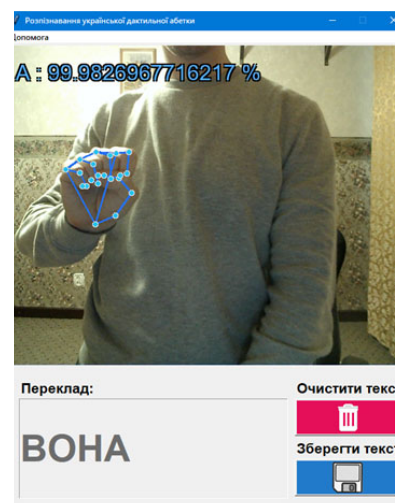


Рис. 9. Результат перекладу та відображення слова «ВОНА»

**Висновки.** В результаті проведених досліджень було:

- проведено аналіз методів відстеження та розпізнавання динамічних жестів людини.
- створено набір даних для розпізнавання літер дактильної абетки;
- розроблено обчислювально-ефективну модель і алгоритм розпізнавання динамічних жестів людини.
- проведено експерименти по оцінці надійності і працездатності системи в реальному часі.

Таким чином отримано програмний продукт Працездатність та ефективність використання було для розпізнавання української дактильної мови. доведено в результаті проведених експериментів.

#### Список літератури:

1. Ковальчук, А. М., Марчук, Г. В., Марчук, Д. К. Застосування згорткової нейронної мережі для розпізнавання рукописних символів. Вчені записки Таврійського національного університету імені В.І. Вернадського, Том 30 (69) Ч. 1. № 4. 2019.
2. Suhoniak, I.I., Yefimenko, A.A., Marchuk, G.V., Feschchenko, D.I. (2020). DECISION SUPPORT SYSTEM DEVELOPMENT FOR BLOCKING UNWANTED CONTENT BY NEURAL NETWORKS. Scientific notes of Taurida National V.I. Vernadsky University. Series: Technical Sciences. <https://doi.org/10.32838/2663-5941/2020.5/19>
3. Левківський В.Л., Марчук Г.В., Ципоренко В.В., Марчук Д.К. Комп'ютерна програма «Алгоритмічно-програмне забезпечення обробки та аналізу потоку кадрів відеоданих, що надходять з камер міста», 2021. URL: <http://eztuir.ztu.edu.ua/bitstream/handle/123456789/8019/109822.pdf?sequence=1&isAllowed=y>
4. Levkivskiy V., Marchuk D., Lobanchykova N., Pilkevych I., Salamatov D. Available parking places recognition system. CEUR Workshop Proceedings 4th Workshop for Young Scientists in Computer Science & Software Engineering Volume 3077 (2022). pp.123-134 URL: <http://ceur-ws.org/Vol-3077/paper07.pdf>
5. Nguyen, A. i Sidorov, Y. 2018. Рекурентна нейронна мережа для обробки великих текстових даних. Системи управління, навігації та зв'язку. Збірник наукових праць. 4, 50 (Вер 2018), 135-138. DOI: <https://doi.org/https://doi.org/10.26906/SUNZ.2018.4.135>
6. Sreelekshmy Selvin, R Vinayakumar, EA Gopalakrishnan, Vijay Krishna Menon, and KP Soman. Stock price prediction using lstm, rnn and cnnsliding window model. In 2017 international conference on advances in computing, communications and informatics (icacci), pages 1643–1647. IEEE, 2017.
7. Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1724–1734, Doha, Qatar. Association for Computational Linguistics. URL: <https://aclanthology.org/D14-1179.pdf>
8. Kratarth Goel, Raunaq Vohra, and Jajati Keshari Sahoo. Polyphonic music generation by modeling temporal dependencies using a rnn-dbn. In International Conference on Artificial Neural Networks, pages 217–224. Springer, 2014.
9. M. Granik and V. Mesyura, «Fake news detection using naive Bayes classifier,» 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017, pp. 900-903, URL: <https://doi.org/10.1109/UKRCON.2017.8100379>
10. Лихошерстов Д.О., Лебедев Д.Ю. Аналіз сучасних систем розпізнавання дактильної-жестової мови для систем сурдоперекладу. Вчені записки Таврійського національного університету імені В.І. Вернадського Серія: Технічні науки Том 32 (71) № 6. 2021. Ч. 1. С. 44-48. URL: <https://doi.org/10.32838/2663-5941/2021.6/07>

#### **Marchuk D.K., Levkivskiy V.L., Marchuk G.V., Holenko M.Yu. DACTYLIC LANGUAGE RECOGNITION SYSTEM OF THE UKRAINIAN ALPHABET**

*This work considers the main development stages of new software for dactylic speech recognition. Dactylic language is a process of communication using sign language. Each letter is represented by a certain position of the fingers on the hand, which denotes a dactyl sign – a dactylem. Usually, this is the language used by people with hearing or speech impairments. Most of these alphabets use one hand. The developed software can recognize letters of both right and left hands. The research object is the methods and process of tracking and recognizing dynamic human gestures using machine learning. A skeleton-based algorithm is used for dactyl recognition – calculations are performed on a virtual human skeleton, and body parts are divided into certain segments. This is implemented through the MediaPipe method, which is a cross-platform machine learning pipeline that provides real-time performance. This work uses a special architecture type of recurrent neural networks capable of learning, namely the model of LSTM (Long short-term memory), which has proven its efficiency. A unique dataset for the Ukrainian dactyl alphabet has been created for training the network. The paper also examines the results of training several models with different activation functions. The loss and accuracy indicators were analyzed and investigated. The interface when working with the software product is described, and the options with possibilities of its use are indicated. The effectiveness of the software in real conditions was investigated, the result of the quality and accuracy of recognizing the Ukrainian dactyl alphabet and creating sentences from translated signs was checked.*

**Key words:** recognition, dactyl language, dactylema, neural network, machine learning, RNN.