

# Нейромережеві методи автоматичного визначення пози людини, яка виконує вправи з художньої гімнастики

**Анастасія Нескородєва**

ORCID: 0000-0002-8591-085X

a.neskorodieva@donnu.edu.ua

Донецький національний університет імені Василя Стуса

## Ключові слова:

трансформер,  
класифікація,  
художня гімнастика,  
перцептрон,  
розпізнавання пози,  
відеопотік.

У художній гімнастиці оцінювання виступів спортсменок є складним суб'єктивним завданням через необхідність врахувати технічні характеристики динамічних вправ, та естетичну досконалість окремих сцен і композиції загалом. У статті представлено результати дослідження, яке спрямоване на розробку програми автоматичного визначення поз спортсменок під час виконання вправ із художньої гімнастики. Розробці моделі передуює вивчення поточних правил підрахунку балів з урахуванням складності рухів спортсменки. Для дослідження створено спеціальний набір даних із визначення поз гімнасток, який включає оцінювані елементи та неоцінювані позиції. Набір даних містить стоп-кадри з відео змагань та фотозвітів. Для визначення поз використовуються 2 методи комп'ютерного зору: MediaPipe Pose та ViTPose. MediaPipe Pose завдяки високій продуктивності дає змогу у реальному часі визначати 33 3D-орієнтири. ViTPose забезпечує високу точність завдяки використанню зорових трансформерів. Здійснено порівняльний аналіз цих двох методів із виділенням сильних сторін та обмежень кожної з них. Для класифікації спортивних елементів на основі розпізнаних поз розроблено модель типу "багатошаровий перцептрон". Навчання моделі здійснено за зворотним розповсюдженням помилки на основі методу градієнтного спуску Адама. Після навчання моделі на авторському наборі даних досягається висока точність класифікації поз гімнасток. Проведене дослідження є інтелектуальним складником майбутньої системи автоматичного визначення в реальному часі пози спортсменки для більш достовірного та оперативного оцінювання виступів у художній гімнастиці. Запропонований підхід до поєднання методів комп'ютерного зору та машинного навчання може бути поширений для вдосконалення спортивного аналізу і в інших споріднених дисциплінах.

DOI: 10.31558/2786-9482.2023.1.4

## Вступ

Спорт завжди був важливою частиною людського життя та культури, слугуючи засобом розваги, змагання та фізичних навантажень. Останніми роками галузь спортивного аналізу стрімко розвивається завдяки прогресу в технологіях і аналітиці даних. Відеоаналіз, зокрема, став важливим інструментом для тренерів, спортсменів і аналітиків для отримання інформації щодо дій окремих гравців та всієї команди.

Художня гімнастика – це вид спорту, який поєднує в собі елементи балету, танцю та гімнастики. В художній гімнастиці спортсмени виконують вправи за допомогою ручних снарядів, як-от скакалка, м'яч, обруч, булави або стрічка. Це красивий і елегантний вид спорту, який вимагає сили, гнучкості, координації та музичності. Однак оцінювання виступів з художньої гімнастики є складним завданням з високим впливом суб'єктивності.

Одним з основних факторів, який впливає на достовірність оцінювання художньої гімнастики, є високий ступінь складності цього виду спорту. Виступи гімнасток можуть включати заплутані та складні рухи, які потребують великої вправності та точності. Спортсмени повинні виконувати різноманітні рухи тіла – стрибки, рівноваги та кидки з одночасним маніпулюванням снарядом. Складність програм може сильно відрізнятися для спортсменів, і судді, виставляючи бали, повинні враховувати рівень складності.

Специфікою художньої гімнастики є суб'єктивний характер цього виду спорту. Естетична привабливість художньої гімнастики відіграє значну роль у системі підрахунку балів. Спортсменів оцінюють не лише за технічними навичками, а й за граціозністю, художньою інтерпретацією та експресією. Судді під час оцінювання враховують музику, хореографію, костюм та обладнання. Ці суб'єктивні фактори ускладнюють стандартизацію системи підрахунку балів і можуть призвести до розбіжностей між оцінками різних суддів.

Швидкість виконання вправ і плавність переходу між елементами може ускладнити суддям можливість зафіксувати кожен рух і деталь у режимі реального часу. Відеоповтори можуть допомогти суддям зробити більш точні оцінки, але це сповільнює темп змагань. Тому створення програми для оцінювання виступів спортсменів є актуальним. Передумовою такої автоматизації є побудова нейромережевого методу для автоматичного визначення поз людини, яка виконує вправи з художньої гімнастики. Саме це і є *метою* дослідження.

### **Постановка завдання**

За сучасними правилами під час виступу художня гімнастка рівня майстер спорту має право виконати максимум 9 елементів, якими може бути стрибок, обертання та рівновага. У правилах Міжнародної федерації гімнастики (франц. *Fédération Internationale de Gymnastique*, скор. *FIG*) [1] описано, які саме елементи і у який спосіб дозволено виконувати, а також зазначена кількість балів, яку отримує спортсменка за правильне виконання кожного елемента. Змагання під егідою FIG, наприклад, чемпіонати Європи, Африки, Північної Америки, Кубок Світу, олімпіади проходять за правилами FIG [2]. Особливістю художньої гімнастики є те, що раз на 4 роки після кожної літньої олімпіади FIG змінює правила оцінювання виступів спортсменок. У правила додаються нові елементи, відбувається переоцінка наявних елементів та вимог до того, що гімнастка повинна виконати за півторихвилинний виступ. Під час останньої зміни також була додана нова бригада суддів, яка оцінює технічні помилки за допомогою зниження балів та визначає загальне зниження балів незалежно та без консультації з іншими суддями.

У роботі [3] наведено алгоритм присвоєння суддівських оцінок рухам художньої гімнастики. У статті [4] представлена нова гібридна динамічно-статична мережа уваги з урахуванням контексту для оцінювання дій у довгих відео. Але результати цих досліджень не вирішують проблему оцінювання за зміни правил. У статті [5] досліджено проблеми оцінювання пози людини та класифікації діяльності за допомогою підходу глибокого

навчання. Оцінювання пози людини є однією із складних задач комп'ютерного зору, яка базується на детектуванні та аналізі положень суглобів. Наявні набори даних для вивчення поз виявилися недостатньо повними з погляду різноманітності поз, оклюзії об'єктів і думок. Такий набір даних дає змогу зробити процес анотації пози відносно простим, але і обмежує застосування навчених моделей на практиці. У статті [6] автори досліджують наявні набори даних для класифікації поз людини та пропонують новий набір даних. Для більшої різноманітності поз людини автори пропонують концепцію дрібнозернистої ієрархічної класифікації поз та пропонують набір даних Yoga-82 для розпізнавання поз йоги з використанням 82 класів. Yoga-82 складається зі складних поз, де точні анотації можуть бути неможливими. Щоб вирішити цю проблему, автори надають ієрархічні мітки для поз йоги на основі конфігурації тіла пози. Набір даних містить трьохрівневу ієрархію, включно з положеннями тіла, варіаціями положень тіла та фактичними назвами поз. У статті зазначена точність класифікації поз з Yoga-82 за умови використання згорткових нейронних мереж. Також представлено кілька ієрархічних варіантів DenseNet для використання ієрархічних міток.

Внаслідок аналізу сучасних релевантних досліджень встановлено, що класичні набори даних для визначення поз людини не підходять для випадків зі складним положенням людини у просторі, якими є елементи у художній гімнастиці. Тому вирішено розробити власний набір даних для визначення поз людини, яка виконує вправи з художньої гімнастики.

### **Розробка власного набору даних**

Набір даних для навчання моделі класифікації спортивних елементів складається з тестового та тренувального наборів. Класами рішень обрано 5 найбільш популярних за виконанням елементів та 2 додаткові класи, які характеризують найчастіших положення спортсменок і не є оцінюваними елементами. На рис. 1 зображено 5 оцінюваних елементів, з першого до п'ятого класу, та 2 неоцінювані положення – шостий та сьомий класи. Приклади виконання елементів взяті з офіційних правил FIG.

У наборі даних усі зображення індивідуальних спортсменок у повний зріст займають не менше 50% довжини зображення. Усі частини тіла потрапляють на зображення. Мінімальна якість усіх зображень є достатньою для точного визначення положення скелета спортсмена; вона становить 240–360 пікселів на дюйм.

Для формування набору даних використано такі ресурси:

- стоп-кадри виконання окремих елементів з відеозаписів кваліфікаційних та фінальних виступів спортсменок на змаганнях XXXII Олімпійських ігор у Токіо [7, 8];
- фотозвіти зі змагань офіційного фотографа Національної збірної України з художньої гімнастики Марії Музиченко [9];
- фотозвіти фотографа Ігоря Сахацького [10].

Для тестування використано відео виступів Вікторії Онопрієнко з м'ячем у кваліфікації на чемпіонаті світу у м. Кітакюсю у 2021 [11]. У тренувальному наборі даних зображення з Вікторією Онопрієнко відсутні. На кожен клас припадає приблизно 90 фото у тренувальному наборі та 6 фото – у тестовому наборі.

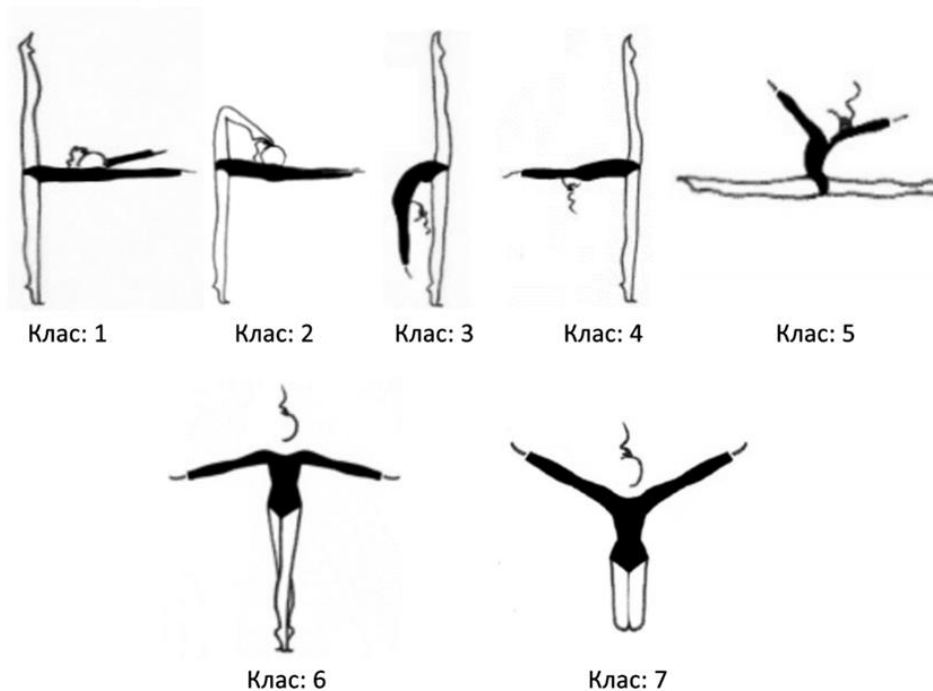


Рисунок 1. Візуалізація класів

### Знаходження ключових точок скелета людини за методом MediaPipe

MediaPipe пропонує кросплатформні настроювані рішення машинного навчання для живих і потокових медіа [12]. Бібліотека містить усе необхідне для налаштування та розгортання на мобільних пристроях (Android, iOS), в інтернеті, настільних комп'ютерах, периферійних пристроях та IoT без особливих зусиль. Цей набір гнучких інструментів створено на основі бібліотеки TensorFlow Lite. Метод забезпечує високу продуктивність у реальному часі на сучасних мобільних телефонах, настільних комп'ютерах, ноутбуках та не вимагає наявності відеокарти. Програмний інтерфейс додатка на мовах програмування Python та JavaScript дає змогу використовувати метод для розв'язання власних задач. У роботі використовується MediaPipe Pose – рішення методом машинного навчання для відстеження пози тіла, яке визначає 33 3D-орієнтири (рис. 2) та маску сегментації фону на всьому тілі з відеокадрів RGB.

Результатом обробки одного зображення за методом MediaPipe Pose є вектор зі 132 значень, де в послідовності від першої ключової точки до 33-ї записані координати  $x$ ,  $y$  та  $z$  і видимість. Якщо видимість має значення, близьке до 1, тоді частину тіла видно якісно, чітко та однозначно. Якщо значення ближче до 0.5, тоді аналізовану частину тіла, може перекривати інший об'єкт, але сусідні скелетні точки дають змогу визначити ймовірне розташування цієї ключової точки. Якщо значення наближається до 0, тоді частину тіла не видно зовсім або вона навіть не потрапляє на зображення.

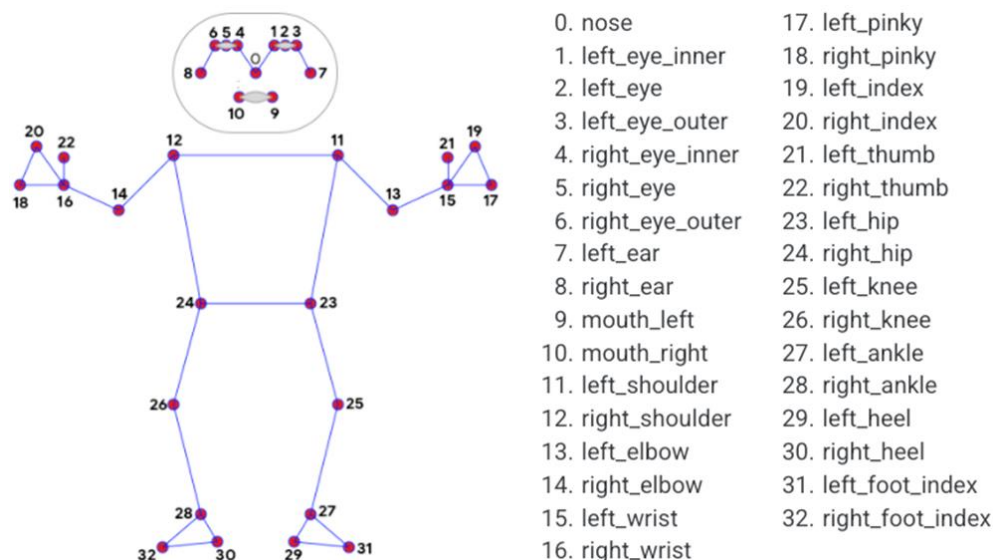


Рисунок 2. Порядок позначень ключових точок у моделі MediaPipe [16]

Обробивши за методом MediaPipe Pose власний набір даних маємо набір векторів, які описують 7 класів – у середньому по 96 векторів для кожного класу. Далі це буде використано для навчання моделі класифікації елементів.

Для підключення цього методу у проєкт необхідно його спочатку встановити як бібліотеку, виконавши команду “pip install mediapipe” у терміналі проєкту. Далі безпосередньо у файлі для виконання на мові Python необхідно під’єднати бібліотеку й викликати необхідні для роботи методи для обробки зображень та візуалізації результатів обробки. Відповідний фрагмент коду наведено нижче:

```
import mediapipe as mp
mp_pose = mp.solutions.pose
mp_drawing = mp.solutions.drawing_utils
mp_drawing_styles = mp.solutions.drawing_styles
```

### Знаходження ключових точок скелета людини за методом ViTPose

Другим рішенням для тестування обрано методом ViTPose, який розроблено на базі трансформерів [13]. У цій статті показано хороші можливості трансформерів простого зору для оцінювання пози з різних аспектів: простота структури моделі, масштабованість розміру моделі, гнучкість парадигми навчання та можливість передачі знань між моделями за допомогою простої базової моделі під назвою ViTPose. Зокрема, ViTPose використовує звичайні та неієрархічні трансформери зору як магістралі для отримання характеристик аналізованого екземпляра людини та полегшений декодер для оцінювання пози. Кількість параметрів можна суттєво збільшити завдяки масштабованій ємності моделі та високому паралелізму трансформерів. Тоді буде встановлено новий фронт Парето між пропускнуою здатністю та продуктивністю. ViTPose є дуже гнучким щодо типу уваги, роздільної здатності введення, попереднього навчання та стратегії тонкого налаштування, а також роботи з кількома завданнями поз у вхідних даних, стратегії попереднього навчання та налаштування, а також може працювати з декількома завданнями. Автори також емпірично демонструють, що знання великих моделей ViTPose можна легко передати малим за допомогою простого

маркера знань, що знання великих моделей ViTPose можна легко переносити на малі моделі за допомогою простого токена знань. Експериментальні результати показують, що базова модель ViTPose перевершує конкурентні методи на складному тесті MS COCO Keypoint Detection.

На рис. 3 зображено загальну та деталізовані структури, які використовуються в методі ViTPose. Для забезпечення роботи цього методу у реальному часі необхідна відеокарта типу Nvidia RTX 2080. У методі ViTPose використовується інша анотація скелета людини за 17-ма ключовими точками (рис. 4). Внаслідок обробки зображення для кожного побудованого скелета видається вектор, у якому для кожної з 17-ти ключових точок вказується координати  $x$  і  $y$  та точність виявлення.

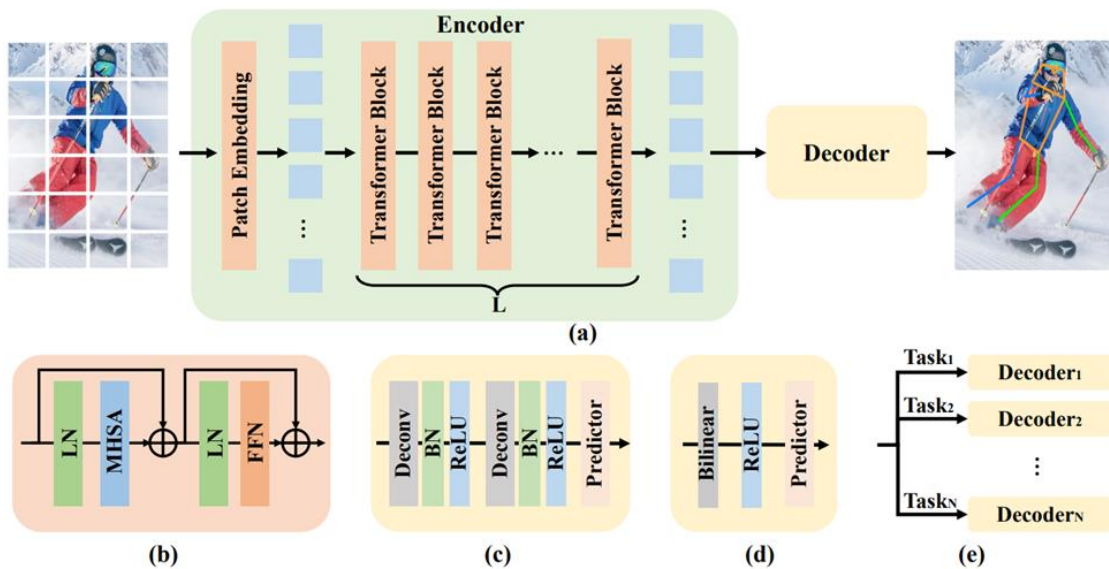


Рисунок 3. Структури в методі ViTPose [13]:

- (a) – загальна структура; (b) – блок трансформера; (c) – класичний декодер;
- (d) – простий декодер; (e) – декодери для кількох наборів даних

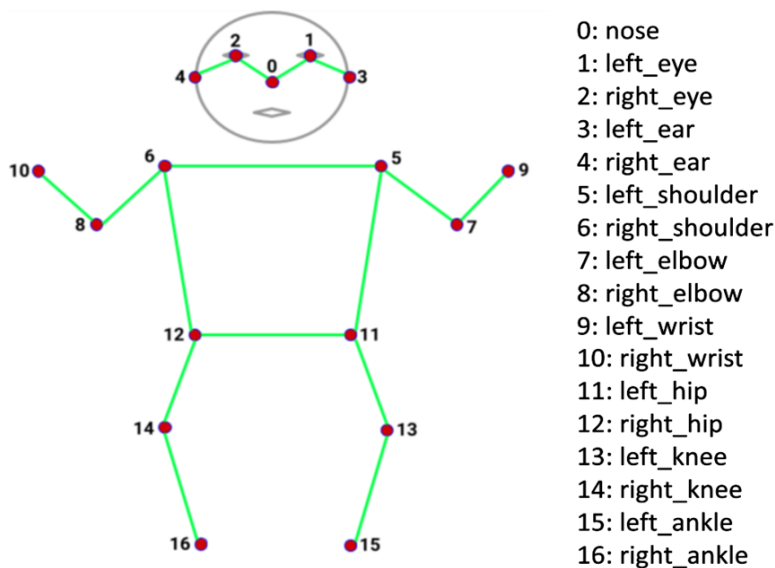


Рисунок 4. Порядок ключових точок у моделі ViTPose [13]

Для використання рішення ViTPose необхідно виконати такі команди для клонування з репозиторію github та встановлення відповідних додаткових бібліотек:

```
git clone https://github.com/open-mmlab/mmcv.git
cd mmcv
git checkout v1.3.9
MMCV_WITH_OPS=1 pip install -e .
cd ..
git clone https://github.com/ViTAE-Transformer/ViTPose.git
cd ViTPose
pip install -v -e .
```

Після виконання цих команд у теці проєкту з'явиться дві нові теки mmcv та ViTPose. Використовуючи методи, що реалізовані у цих бібліотеках, розроблено 2 алгоритми. Перший алгоритм для обробки власного набору даних, а другий – для обробки відеопотоку виступів спортсменів. Алгоритм обробки набору даних знаходить координати ключових точок скелета людини на зображенні та додає їх до набору даних, який далі буде використаний для навчання моделі класифікації.

### Порівняння методів MediaPipe та ViTPose

Для порівняння методів обрано 4 зображення (рис. 5). У верхньому ряді зображення оброблені за методом MediaPipe Pose, в нижньому – за методом ViTPose. На першому зображенні, обробленому методом MediaPipe Pose, зміщені ключові точки стегон, а також некоректно визначено точки на ногах. На другому зображенні, яке оброблене цим же методом, точка правого коліна позначена ближче до стегна. На третьому зображенні помилково визначено лише пальці правої стопи. На четвертому зображенні, де спортсменка виконує елемент, який належить до 3 класу, під час визначення методом MediaPipe Pose було переплутано ключові точки тулуба тіла. Через це було хибно визначено праву ногу, як праву руку, праву руку, як ліву ногу, та ліву ногу, як праву ногу. Недоліком методу MediaPipe Pose є те, що якість відстеження скелета людини під час виконання елементів, пов'язаних з обертанням та групуванням у просторі, є низькою. Наприклад, під час виконання обертів модель починає плутати праві та ліві частини тіла. Ці недоліки ускладнюють подальший розвиток проєкту на базі цього методу.

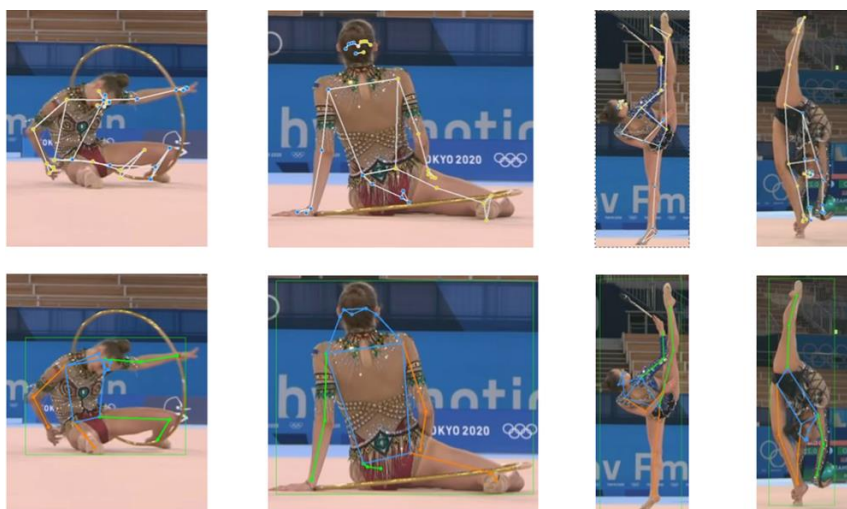


Рисунок 5. Обробка зображень за MediaPipe Pose (зверху) та за ViTPose (знизу)

Результати обробки зображень з рис. 5 методом ViTPose є кращими; критичних помилок не виявлено. Але цей метод вимагає більш потужного обладнання, а моделі займають більший об'єм пам'яті.

### Розробка багатшарового перцептрона для класифікації

Для класифікації спортивних елементів розроблено багатшаровий мультикласовий перцептрон [14]. На рис. 6 наведено архітектуру розробленого перцептрона, який є нерекурентною статичною багатшаровою штучною нейронною мережею з двома прихованими шарами. Фрагмент коду на мові Python для створення такого перцептрона є таким:

```
inputs=tf.keras.Input(shape=(51))
embedding=landmarks_to_embedding(inputs)
layer=keras.layers.Dense(128, activation=tf.nn.relu6)(embedding)
layer=keras.layers.Dropout(0.5)(layer)
layer=keras.layers.Dense(64, activation=tf.nn.relu6)(layer)
layer=keras.layers.Dropout(0.5)(layer)
outputs=keras.layers.Dense(len(class_names),
    activation="softmax")(layer)
model=keras.Model(inputs, outputs)
model.summary()
```

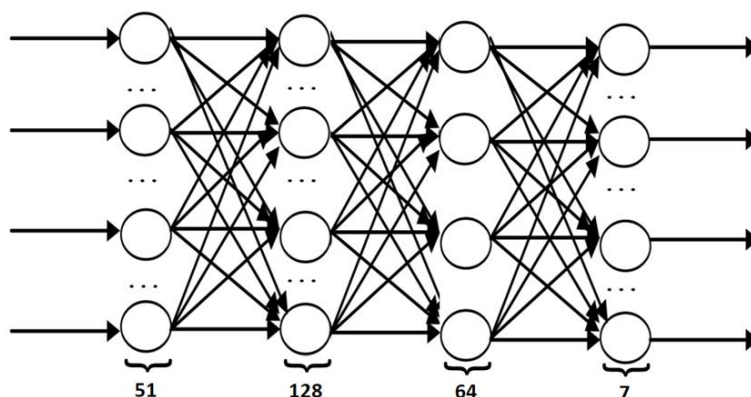


Рисунок 6. Архітектура розробленого тришарового перцептрона

Навчання багатшарового перцептрона здійснюється на основі корекції помилок (навчання з учителем) з використанням методу градієнтного спуску Adam. Після закінчення кожної епохи модель оцінюється на контрольних даних для прийняття рішення про ранню зупинку (early stopping callback) та збереження контрольної точки (checkpoint callback). Безпомилковість навченої моделі на тестовій вибірці становить 0.9048. Це свідчить про те, що модель у більшості випадків правильно класифікує спортивні елементи. Матриця сплутаності (рис. 7) вказує, що найчастіше модель помиляється під час розпізнавання об'єктів першого класу – плутає їх із четвертим. Причиною цього є те, що ці 2 спортивні елементи є однаковими за положенням тіла, а відрізняються лише напрямком тулуба і голови. У класі 1 голова направлена до підлоги, а в класі 4 – до стелі.



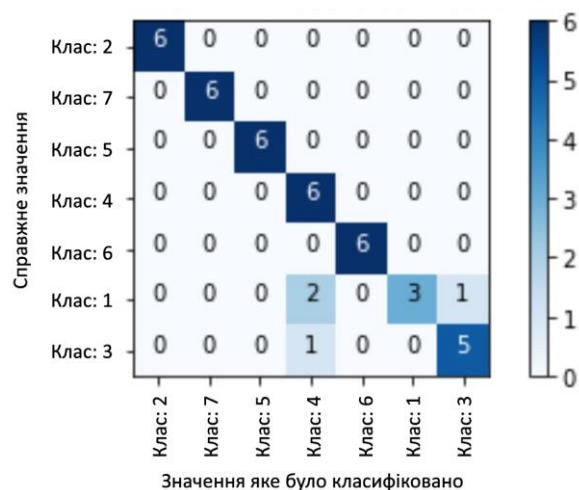


Рисунок 7. Матриця сплутувань

### Програмна реалізація визначення пози спортсменки

Програма працює так:

1. Відео, яке необхідно обробити, програма зчитує покадрово.
2. Кожен кадр передається у першу модель для виявлення людини. Внаслідок цього результати обраний метод повертає координати обмежувальної рамки і точність виявлення кожної рамки, де було виявлено людину.
3. У другу модель для визначення положення скелета передається кадр з відео та відповідна йому анотація від першої моделі. Результати роботи першої моделі необхідні, щоб звузити площу обробки зображення для другої моделі. Тоді модель буде визначати положення скелета людини лише в межах обмежувальних рамок від першої моделі. На виході отримуємо координати ключових точок скелета людини та оцінку їх видимості.
4. Модель класифікації елементів отримує для кожного кадру відповідний йому вектор значень, а результатом обробки є ймовірність обрання кожного із 7 класів. Серед класів обирається той клас, ймовірність до якого є найбільшою і перевищує пороговий рівень впевненості.
5. Програма у верхньому лівому кутку кожного кадру візуалізує відповідне зображення обраного класу. Правіше від зображення класу програма виводить значення впевненості прийняття рішення. Якщо впевненість недостатня, тоді програма залишає початковий кадр без змін.

Приклад обробки кадру програмою зображено на рис. 8. Програмний код захищено авторським свідоцтвом [15].

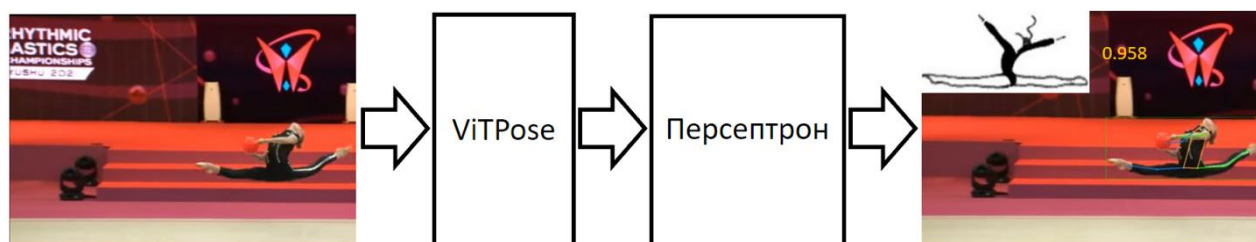


Рисунок 8. Порядок обробки кадру програмою

## Висновки

Зібрано та структуровано набір даних із зображенням виконання спортивних елементів. Виявлено особливості роботи моделей комп'ютерного зору MediaPipe Pose та ViTPose для визначення пози людини, яка виконує вправи з художньої гімнастики, та обрано найкраще рішення для попередньої обробки підготовленого набору. Головним результатом дослідження є модель класифікації спортивних елементів, яка за 7 класами демонструє точність 0.9048.

У наступних дослідженнях планується розширити набір даних, збільшивши кількість класів та екземплярів. Також для покращення результатів класифікації доцільно спробувати використати LSTM та механізм уваги.

## Література

1. *Fédération Internationale de Gymnastique*. URL: <https://www.gymnastics.sport/site/> (дата звернення: 28.05.2023).
2. *2022 – 2024 Code of Points*. URL: [https://www.gymnastics.sport/publicdir/rules/files/en\\_2022-2024%20RG%20Code%20of%20Points.pdf](https://www.gymnastics.sport/publicdir/rules/files/en_2022-2024%20RG%20Code%20of%20Points.pdf) (дата звернення: 27.05.2023).
3. Sierra-Palmeiro, E., Bobo-Arce, M., Pérez-Ferreirós, A., & Fernández-Villarino, M. A. (2019). Longitudinal Study of Individual Exercises in Elite Rhythmic Gymnastics. *Frontiers in Psychology, 10*. DOI: 10.3389/fpsyg.2019.01496.
4. Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic recognition and scoring of Olympic rhythmic gymnastics movements. *Human Movement Science, 34*(1), 63–80. DOI: 10.1016/j.humov.2014.01.001.
5. Bearman, A., & Dong, C. (2015). Human pose estimation and activity classification using convolutional neural networks. *CS231n Course Project Reports*.
6. Verma, M., et al. (2020). Yoga-82: A new dataset for fine-grained classification of human poses. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 1038–1039.
7. *Olympics Gymnastics: Rhythmic Gymnastics – Individual All-Around-Qualification 1&2* (2020). Tokyo 2020. YouTube. URL: <https://www.youtube.com/watch?v=uRzmkLF8MVI> (дата звернення: 27.05.2023).
8. *Olympics: FULL Rhythmic Gymnastics Individual All Around Final at Tokyo 2020* (2020). YouTube. URL: <https://www.youtube.com/watch?v=v6ZuroWdLTs> (дата звернення: 27.05.2023).
9. *Альбоми зі зйомок на спортивних турнірах фотографа Марії Музиченко*. URL: <https://muzychenko.photos/our-services/sports-photography> (дата звернення: 27.05.2023).
10. *Портфолію Ігоря Сахатського*. URL: <https://sakhatskyi.com/portfolio/> (дата звернення: 27.05.2023).
11. *Ukrainian RG Federation: Viktoriia Onopriienko Ball Qual 26,200 – World Championships Kitakyushu 2021* (2021). YouTube. URL: <https://www.youtube.com/watch?v=IKzuWUIe8Rc> (дата звернення: 27.05.2023).

12. GitHub – Google / Media Pipe: Cross-platform, customizable ML solutions for live and streaming media. GitHub. URL: <https://github.com/google/mediapipe> (дата звернення: 27.05.2023).
13. Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in Neural Information Processing Systems*, 35, 38571–38584.
14. Bielecki, A. (2019). Models of neurons and perceptrons: Selected problems and challenges. *Studies in Computational Intelligence*. Vol. 770. Springer Cham, 156 p. DOI: 10.1007/978-3-319-90140-4.
15. А. Р. Нескорородева (Україна). А. С. 116622 Україна, УКРНОІВІ. Комп'ютерна програма "Pose estimation for sports (Rhythmic gymnastics)". № с202300058; заявка 06.01.2023; опубл. 01.03.2023.
16. *ML Kit*. URL: <https://developers.google.com/ml-kit/vision/pose-detection?hl=en> (date of access: 27.05.2023).

Рукопис отримано – 09/06/2023; прийнято до публікації – 26/06/2023.

- © Донецький національний університет імені Василя Стуса, 2023  
© Анастасія Нескорородева, 2023

## Neural network methods for automatic person pose estimation in rhythmic gymnastics exercises

Anastasiia Neskoro dieva

### Abstract

In rhythmic gymnastics, evaluating the performances of female athletes is a complex subjective task due to the need to consider both the technical characteristics of dynamic exercises and the aesthetic perfection of individual scenes and the composition as a whole. The article presents the results of a study aimed at developing a program for automatically detecting the postures of female athletes during rhythmic gymnastics exercises. The development of the model is preceded by a study of the current scoring rules, considering the complexity of the athlete's movements. For the study, a special dataset was created to determine the gymnasts' poses, which included scored elements and unscored positions. The dataset contains freeze frames from the video of the competition and photo reports. Two computer vision methods are used to identify the poses: MediaPipe Pose and ViTPose. MediaPipe Pose allows for real-time detection of 33 3D landmarks due to its high performance. ViTPose provides high accuracy using visual transformers. A comparative analysis of these two methods is carried out, highlighting the strengths and limitations of each of them. A multilayer perceptron model has been developed to classify sports elements based on recognized poses. The model was trained using back-propagation of error based on Adam's gradient descent method. After training the model on the author's dataset, high accuracy in the classification of gymnasts' poses is achieved. This study is an intellectual component of a future system for automatic real-time detection of an athlete's pose for a more reliable and efficient evaluation of rhythmic gymnastics performances. The proposed approach to combining computer vision and machine learning methods can be extended to improve sports analysis in other related disciplines.

**Keywords:** transformer, classification, rhythmic gymnastics, perceptron, pose estimation, video stream.

### References

1. *Fédération Internationale de Gymnastique*. URL: <https://www.gymnastics.sport/site/> (date of access: 28.05.2023).
2. *2022 – 2024 Code of Points*. URL: [https://www.gymnastics.sport/publicdir/rules/files/en\\_2022-2024%20RG%20Code%20of%20Points.pdf](https://www.gymnastics.sport/publicdir/rules/files/en_2022-2024%20RG%20Code%20of%20Points.pdf) (date of access: 27.05.2023).
3. Sierra-Palmeiro, E., Bobo-Arce, M., Pérez-Ferreirós, A., & Fernández-Villarino, M. A. (2019). Longitudinal Study of Individual Exercises in Elite Rhythmic Gymnastics. *Frontiers in Psychology*, 10. DOI: 10.3389/fpsyg.2019.01496.
4. Díaz-Pereira, M. P., Gómez-Conde, I., Escalona, M., & Olivieri, D. N. (2014). Automatic recognition and scoring of Olympic rhythmic gymnastics movements. *Human Movement Science*, 34(1), 63–80. DOI: 10.1016/j.humov.2014.01.001.
5. Bearman, A., & Dong, C. (2015). Human pose estimation and activity classification using convolutional neural networks. *CS231n Course Project Reports*.
6. Verma, M., et al. (2020) Yoga-82: A new dataset for fine-grained classification of human poses. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 1038–1039.
7. *Olympics Gymnastics: Rhythmic Gymnastics – Individual All-Around-Qualification 1&2 (2020)*. Tokyo 2020. YouTube. URL: <https://www.youtube.com/watch?v=uRzmkLF8MVI> (date of access: 27.05.2023).
8. *Olympics: FULL Rhythmic Gymnastics Individual All Around Final at Tokyo 2020 (2020)*. YouTube. URL: <https://www.youtube.com/watch?v=v6ZuroWdLTs> (date of access: 27.05.2023).
9. *Albomy zi ziomok na sportyvnykh turnirakh fotohrafa Mariyi Muzychenko*. URL: <https://muzychenko.photos/our-services/sports-photography> (date of access: 27.05.2023). [In Ukrainian].
10. Portfolio Yhorya Sakhatskoho. URL: <https://sakhatskyi.com/portfolio/> (date of access: 27.05.2023). [In Ukrainian].
11. *Ukrainian RG Federation: Viktoriia Onopriienko Ball Qual 26,200 - World Championships Kitakyushu 2021 (2021)*. YouTube. URL: <https://www.youtube.com/watch?v=IKzuWUle8Rc> (date of access: 27.05.2023).

12. GitHub – Google / Media Pipe: Cross-platform, customizable ML solutions for live and streaming media. GitHub. URL: <https://github.com/google/mediapipe> (date of access: 27.05.2023).
13. Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in Neural Information Processing Systems*, 35, 38571–38584.
14. Bielecki, A. (2019). Models of neurons and perceptrons: Selected problems and challenges. *Studies in Computational Intelligence*. Vol. 770. Springer Cham, 156 p. DOI: 10.1007/978-3-319-90140-4.
15. A. R. Neskorodieva (Ukraine). A. S. 116622 Ukraine, UKRNOIVI. Kompyuterna prohrama “Pose estimation for sports (Rhythmic gymnastics)” № c202300058; zayava. 06.01.2023; opubl. 01.03.2023. [In Ukrainian].
16. *ML Kit*. URL: <https://developers.google.com/ml-kit/vision/pose-detection?hl=en> (date of access: 27.05.2023).

