**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп’ютерних наук та кібернетики

**Звіт**

до дослідження лабораторної роботи

з дисципліни

**«**Штучний інтелект: принципи та методи**»**

Виконав студент 1-го курсу

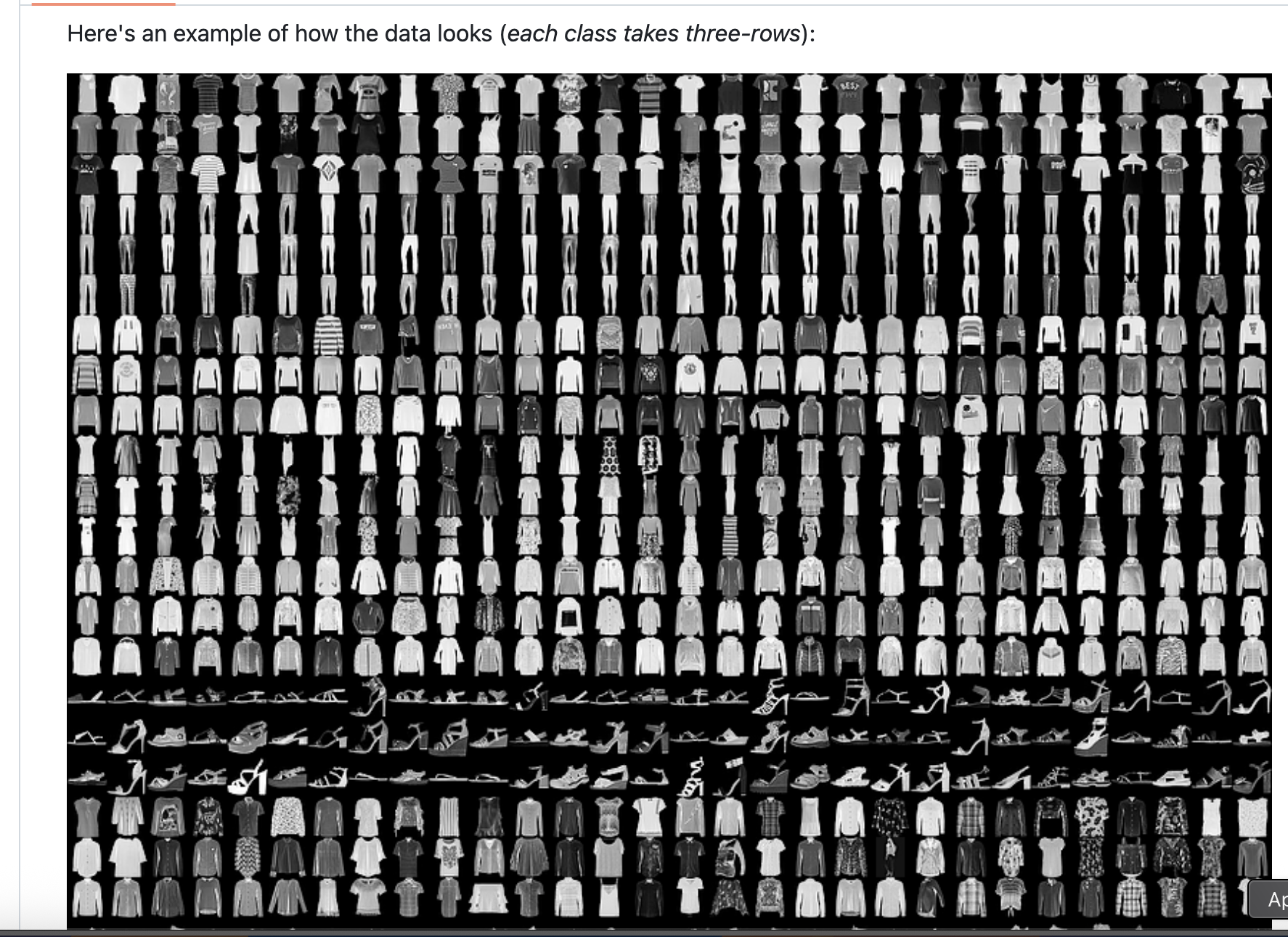
ОНП “Штучний інтелект”

Ходаков Максим Олегович

Київ – 2025

**Звіт-дослідження: класифікація зображень одягу на датасеті Fashion-MNIST**

**Fashion-MNIST** — це публічний набір зображень товарів Zalando, задуманий як «drop-in» заміна MNIST для бенчмаркингу алгоритмів машинного навчання. Він містить 70 000 зображень у градаціях сірого розміром 28×28: 60 000 у тренувальній частині та 10 000 у тестовій; кожне зображення належить до однієї з 10 категорій (від “T-shirt/top” до “Ankle boot”). Формат і розбиття ідентичні MNIST, завдяки чому датасет підтримують більшість бібліотек «з коробки». Офіційний репозиторій Zalando на GitHub надає опис, прямі посилання на файли та навіть автоматичний бенчмарк класичних методів; короткі енциклопедичні довідки доступні на Wikipedia, а також на сторінках агрегаторів на зразок Kaggle.



A screenshot of a computer

Description automatically generated

Первісна мотивація авторів — підвищити складність «цифрового» бенчмарку, не змінюючи інтерфейс і формат даних. На відміну від рукописних цифр MNIST, предмети одягу мають значно більшу внутрішньокласову варіативність і міжкласові подібності (наприклад, “Shirt” проти “T-shirt/top”), тож прості моделі рідше «вичерпують» задачу. У статті-запуску (Xiao, Rasul, Vollgraf, 2017) підкреслюється саме ідея «прямої заміни» MNIST для справедливого порівняння підходів; згодом з’явилися й додаткові напрацювання щодо якості даних, зокрема роботи про видалення майже-дублікатів між train/test, які можуть штучно завищувати метрики. [arXiv+1](https://arxiv.org/abs/1708.07747?utm_source=chatgpt.com)

Стандартний протокол оцінювання передбачає тренування на 60 000 прикладах і звітування **accuracy** на офіційних 10 000 тестових зображень. Для глибшого аналізу часто додають precision, recall та F1 (у макро- й зваженому варіантах), а також матрицю плутанини для виявлення систематичних помилок між схожими класами. Базові довідкові цифри щодо структури та метрик узгоджуються між офіційним репозиторієм, енциклопедичними джерелами та супровідними реалізаціями в основних фреймворках. [GitHub+2Wikipedia+2](https://github.com/zalandoresearch/fashion-mnist?utm_source=chatgpt.com)

У класі **традиційних методів** (логістична регресія, SVM, kNN, дерева рішень та ансамблі) Fashion-MNIST використовується переважно як «еталонний базлайн». Команда Zalando підтримує автоматизований бенчмарк для великої колекції таких моделей (близько 129 конфігурацій), що демонструє обмеження класики на складнішій природі даних порівняно з MNIST і мотивує перехід до глибоких підходів. При цьому окремі фіч-інженерні зв’язки на кшталт HOG+SVM у публікаціях зазвичай поступаються компактним згортковим мережам. [GitHub](https://github.com/zalandoresearch/fashion-mnist?utm_source=chatgpt.com)

Найтісніше з датасетом пов’язаний клас **згорткових нейромереж (CNN)**. У рецензованій роботі Sensors (Nocentini та ін., 2022) систематично порівняно кілька CNN-архітектур для класифікації одягу; результати підтверджують, що належно спроєктовані CNN стабільно випереджають класичні ML-алгоритми на Fashion-MNIST. У відкритому доступі на PubMed Central стаття подає деталі архітектур і протоколів, що полегшує відтворення висновків. У ширшому пласті літератури типові компактні CNN без агресивного тюнінгу дають діапазон точності близько 90–95 % на офіційному тесті. [PMC](https://pmc.ncbi.nlm.nih.gov/articles/PMC9739239/?utm_source=chatgpt.com)

Окрема лінія робіт присвячена **трансформерам у комп’ютерному баченні**. Для невеликих 28×28-входів, характерних для Fashion-MNIST, Vision Transformer (ViT) зазвичай вимагає перетворення у RGB та масштабування до вхідної роздільності моделі (наприклад, 224×224) з подальшим донавчанням. У журналі Electronics (MDPI) продемонстровано, що ViT досягає близько **95 %** точності на Fashion-MNIST із порівнюваними значеннями precision/recall/F1, що узгоджується з висновками численних оглядових і прикладних праць про конкурентність трансформерних підходів на простих доменах за умови коректного препроцесингу. Водночас на таких «малих» задачах компактні CNN часто залишаються не гіршими за ViT при рівних ресурсах, тоді як переваги трансформерів сильніше проявляються у масштабуванні на складніші сцени й великі дані. [MDPI](https://www.mdpi.com/2079-9292/12/20/4263?utm_source=chatgpt.com)

A screenshot of a computer

Description automatically generated

За останні роки з’явилися й **амбіційні SOTA-заявки** на основі відносно простих CNN із продуманими аугментаціями. У журналі Mathematics (MDPI, 2024) опубліковано роботу, яка повідомляє **99.65 %** точності на Fashion-MNIST для архітектури «CNN-3-128» за наявності відповідного препроцесингу та підбору гіперпараметрів; у статті також наведено попередні «кращі» орієнтири близько 99.1 %. Такі результати вражають і підкреслюють потребу в прозорих протоколах, публікації коду/конфігурацій та незалежних реплікаціях, оскільки більшість прикладних керівництв і відкритих реалізацій зупиняються суттєво нижче цієї позначки. Для ознайомлення з публікацією можна також використати агрегатори на кшталт DOAJ. [MDPI+1](https://www.mdpi.com/2227-7390/12/20/3174?utm_source=chatgpt.com)

У підсумку Fashion-MNIST залишається зручним і репрезентативним стендом для дослідження моделей класифікації зображень завдяки простому формату, рівномірному класовому балансу й широкій підтримці в бібліотеках. Для коректної порівнянності доцільно дотримуватися стандартного протоколу (60 000/10 000, звітність accuracy на офіційному тесті) та фіксувати всі складові препроцесингу й навчання: перетворення зображень, аугментації, кількість епох, розклад навчання, регуляризації та початкові стани (seed). У сучасній літературі компактні **CNN** демонструють дуже високі результати, тоді як **ViT/трансформери** забезпечують стабільний рівень близько 95 % та хороші перспективи масштабованості. Найвищі заявлені значення понад 99 % потребують особливої уваги до відтворюваності, але водночас слугують корисними «маяками» для подальшого розвитку методів.

Посилання (GitHub: <https://github.com/maksymkhodakov/AILab1>):

Інформація про датасет  
<https://github.com/zalandoresearch/fashion-mnist>  
<https://www.kaggle.com/code/jeegarmaru/fashion-mnist-keras/input>   
<https://en.wikipedia.org/wiki/Fashion_MNIST>

Додаткові ресурси/результати по цьому датасету  
<https://www.mdpi.com/2227-7390/12/20/3174>  
<https://doaj.org/article/7d56b569416f4c28b10598b54bd6e926>  
<https://pmc.ncbi.nlm.nih.gov/articles/PMC9739239/>