http://www.cnews.ru/news/line/oracle_predstavila_novyj_produkt_dlya. 9. Надежная защита критически важных баз данных в реальном времени [Електронний ресурс]. Режим доступу до ресурсу : https://www.mcafee.com/ru/products/data-center-securi. 10. Система защиты баз данных [Електронний ресурс]. Режим доступу до ресурсу : http://elvis.ru/services/application/database/. 11. Безопасность баз данных под контролем [Електронний ресурс]. Режим доступу до ресурсу : http://www.mfisoft.ru/direction/ib/garda-bd/.

УДК 519.7

В. Хома[1,2], Ю. Хома[1], В. Герасименко[1], Д. Сабодашко[1]
[1]Національний університет "Львівська політехніка",
кафедра захисту інформації,
кафедра інформаційно-вимірювальних технологій,
[2]Політехніка Опольська (Польща)

ЕКГ-ІДЕНТИФІКАЦІЯ З ВИКОРИСТАННЯМ ГЛИБИННИХ НЕЙРОННИХ МЕРЕЖ

Описано розроблення нової біометричної системи на основі електрокардіограми людини (ЕКГ) як альтернативного джерела біометричної інформації. Для здійснення ідентифікації запропоновано використати передові алгоритми машинного навчання (глибинні нейроні мережі) у поєднанні з техніками цифрового опрацювання сигналів. Всі експерименти проводились на самостійно зібраному наборі даних з використанням вбудованої електроніки з мінімізованою аналоговою частиною.

Ключові слова: біометрія, нейронні мережі, ідентифікація, електрокардіограма.

This paper is focus on developing novel biometric system based on humans' electrocardiogram (ECG) as alternative type of biometric information. To perform identification advanced machine learning algorithms (deep neural networks) combined with digital signal processing techniques. All experiments were done on self-collected data set using the embedded electronics with a minimized analog front end.

Keywords: biometrics, neural networks, identification, electrocardiogram.

Вступ

Багато аспектів нашого повсякденного життя потребують автоматичного і точного підтвердження ідентичності особистості. Широке впровадження механізмів розпізнавання, побудованих на основі сутностей (USB-брелок, ID-картка) або на основі знань (PIN-код, пароль), викликає побоювання щодо безпеки через ризик крадіжки ідентифікаторів.

Процес розпізнавання ідентифікатора користувача називається ідентифікацією. Автентифікація являє собою процес підтвердження достовірності заявленого користувачем ідентифікатора. Рівень безпеки визначається рівнем прийняття системи.

За високого рівня безпеки (наприклад, на військовому об'єкті) не має бути жодних помилкових валідацій невідомих персон, навіть якщо це збільшує кількість помилок під час перевірки акредитованих користувачів.

Найкращим способом перевірити ідентичність суб'єкта є використання його біометричних даних. Використання цих даних (характеристик або ознак) у методах ідентифікації та/або автентифікації

називається *біометрією*. Перевагою біометричних систем є повна залежність від індивідуальних характеристик людини. Біометричні дані можна поділити на два основні класи: фізіологічні та поведінкові. Фізіологічні належать до форми тіла (наприклад, відбитки пальців, ДНК, долоня руки, райдужка ока). Поведінкові пов'язані з поведінкою людини (наприклад, хода, мова).

Одним з основних недоліків більшості біометричних систем є простота фальсифікації облікових даних. Наприклад, фотографією можна підробити, райдужку ока фальсифікувати контактними лінзами, і навіть відбитки пальців можна змінити за допомогою гелевих накладок на пальці [1]. Тому постає необхідність використання внутрішніх фізіологічних характеристик. Електрокардіограма (ЕКГ) використовується протягом багатьох років як медичні діагностичні дані, і сьогодні вона є основою сучасних біометричних систем.

Обґрунтованість використання ЕКГ для біометричного розпізнавання підтверджується тим фактом, що фізіологічні і геометричні відмінності серця у різних людей відображають певну унікальність їх ЕКГ.

Порівняно з популярними комерційними біометричними системами – такими, як відбитки пальців, геометрія долоні, розпізнавання обличчя – електрокардіограма має такі переваги [2–4]:

- надійність, оскільки це внутрішня біометрична характеристика людини, і тому її важче фальсифікувати;
- висока точність навіть в умовах, відмінних від нормальних, низька чутливість до шуму;
- проста для вимірювання: електрокардіограму можна виміряти з пальців і долонь одним сенсором або використовуючи текстильні електроди;
- можливість вимірювання у режимі реального часу.

Процес ідентифікації ЕКГ складається з таких етапів: збирання даних, попереднє опрацювання, класифікація [4–6]

Збір даних

Для вимірювання ЕКГ використовують плати Arduino Uno та e-Health Sensor Platform V2.0. Arduino Uno – це плата на основі мікроконтролера ATmega328. До її складу входить все необхідне для зручної роботи з мікроконтролером: 14 цифрових входів / виходів (з них 6 можна використовувати як ШІМ-виходи), 6 аналогових входів, кварцовий резонатор на 16 МГц, роз'єм USB, роз'єм живлення і кнопка скидання. Основні технічні характеристики наведено в табл. 1.

Таблиця 1

Характеристики Arduino Uno

Мікроконтролер	ATmega328
Робоча напруга	5В
Напруга живлення	7-12В
Flash-пам'ять	32 КБ (ATmega328)
SRAM	2 КБ (ATmega328)
EEPROM	1 КБ (ATmega328)
Тактова частота	16 МГц

Плата e-Health Sensor Shield V2.0 дає змогу використовувати Arduino в біометричних і медичних цілях. Моніторинг тіла можна здійснювати за допомогою 10 різних датчиків: пульсу, рівня кисню в крові, витрати повітря, температури тіла, вимірюючи електрокардіограму (ЕКГ), рівень глюкози в крові, шкірно-гальванічну реакцію, артеріальний тиск (тонометром), положення пацієнта (акселерометр) і м'язову активність (ЕМГ) [7].

Збирали дані для бази електрокардіограм за вимірювальною схемою з різницевим підсилювачем та 8-бітний АЦП з частотою дискретизації 277 Гц. Дані з АЦП передавали на ПК через COM-порт, використовуючи бібліотеку PySerial. Кожне вимірювання тривало близько 10 секунд, відповідно кожний запис містить від 10-ти ударів серця.

Для формування навчальної та тестової бази електрокардіограм було вирішено вимірювати І відведення Ейнтговена. Це підвищує комфортність використання реалізованої системи, оскільки для зняття ЕКГ необхідно розмістити на електродах пальці правої та лівої руки (рис. 1).

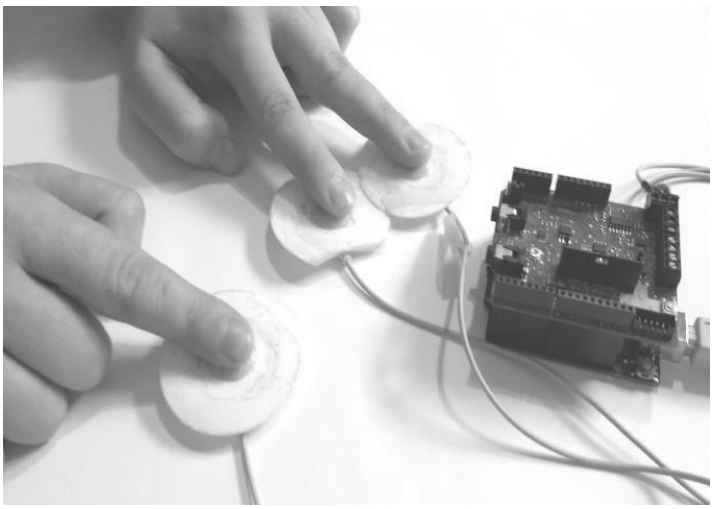

Рис. 1. Вимірювання електрокардіограми з першого відведення

Попереднє опрацювання даних

Виміряний ЕКГ-сигнал попередньо опрацювували, що передбачало фільтрацію, сегментацію та нормалізацію. Попереднє опрацювання передує безпосередній класифікації даних нейронною мережею.

На сигнал ЕКГ впливає багато різних факторів, тому під час вимірювання до корисного сигналу потрапляє також багато шуму, наприклад:

- низькочастотний шум, спричинений потовиділенням, що впливає на імпеданс електродів, а також дихання, рухи тіла;
- навантаження від мережі може спричинити синусоїдний шум 50 Гц;
- шуми, спричинені м’язовими зусиллями, наприклад, диханням, тремтінням рук;
- шум, спричинений електричними елементами, які використовуються у схемі вимірювання.

Інформаційна частина ЕКГ-сигналу знаходиться у проміжку 5–35 Гц. Для фільтрації шуму використовували смуговий рекурсивний фільтр на основі полінома Батерворта.

Відфільтрований сигнал було нормалізовано і приведено до діапазону від -1 до 1. Після цього сигнал розбивався на окремі проміжки (удари серця). Для визначення R-піків (максимального значення при кожному ударі) використовували алгоритм Гамілтона з бібліотеки BioSPPy [8].

Класифікація

Особливістю методу навчання з учителем, який був обраний нами, є використання попередньо зібраних записів ЕКГ для встановлення залежності між вхідними і вихідними даними. Після того, як мережа, “навчаючись”, підібрала найкращі вагові коефіцієнти, можна використовувати її для передбачення результатів поданих на вхід невідомих доти записів. Існує велика кількість алгоритмів для класифікації. Конкретну архітектуру можна обрати залежно від характеристик попередньо зібраних даних, особливостей поставленої задачі та областей використання отриманої моделі.

Базовою архітектурою для навчання було обрано нейронну мережу прямого поширення (Feedforward neural network). Нейронна мережа складається з вхідного шару (270 нейронів, які відповідають 270 попередньо опрацьованим вибіркам сигналу), трьох прихованих шарів із 70, 50 та 30 нейронами, а також 18 нейронів вихідного шару (відповідає кількості користувачів). Передавальною функцією для прихованих шарів було вибрано Rectified Linear Unit, для вихідного

шару – softmax. Алгоритм навчання – Adagrad, кількість циклів навчання – 3000, швидкість навчання – 0,05, L1 коефіцієнт регуляризації – 0,001.

Після етапу навчання необхідно оцінити продуктивність натренованої мережі. Для цього попередньо зібрані дані було розділено на тренувальну (навчальну) та тестову бази. Щоб уникнути проблеми неповного представлення класу, розділення проводили для кожного класу окремо, щоб представити його у кожній із баз пропорційно.

Експерименти та результати

Під час досліджень використовували дані з Lviv Biometric Data Set, яка в момент підготування статті містила 147 ЕКГ-записів 18-ти осіб. Мінімальна кількість записів на людину – 3. Навчальна база містить 88 записів, тестова – 49 [9].

Дослідження проводили за допомогою мови програмування Python 2.7. Також використовували такі фреймворки та бібліотеки, як skflow, scipy, numpy, matplotlib, sci-kit learn. Всі алгоритми глибинного навчання виконано у поєднанні з фреймворком Tensorflow. Вихідний код можна знайти за посиланнями [10, 11].

Для досліджень було використано параметри та конфігурації з секції “Класифікація”. Етап навчання (тренування) займав близько 10-ти хвилин на ПК з такими характеристиками: CPU – Intel Core i7-5500, операційна система – Ubuntu 14.04, 8 GB RAM.

Мета першого експерименту – перевірити стабільність результатів класифікації. Передумовою цього є випадкова ініціалізація вагових коефіцієнтів штучної нейронної мережі (ШНМ). Відповідно кожен цикл навчання дасть різні результати навіть за однакових вихідних умов (попередньо зібрані дані та гіперпараметри навчання). З метою зменшення впливу випадкової ініціалізації вибрані архітектури ШНМ навчалися з ітерацією 100 разів. Результати показано у табл. 2. Наведені результати вказують на те, що всі архітектури мають значне відхилення. Отже, найобґрунтованішим рішенням буде навчати модель певну кількість разів, але використовувати її найкращу версію. Точність класифікації не зазнає відчутних змін при модифікації типу архітектури. Для подальших досліджень використовували три приховані шари з 70, 50 та 30 нейронами.

Таблиця 2

Аналіз точності для різних архітектур ШНМ

Кількість нейронів прихованих рівнів	Середнє значення точності, %	Стандартне відхилення
[100 70 50 30]	88,50	3,35
[70 50 30]	88,97	2,81
[50 50 20]	84,14	5,28
[70 20]	88,84	4,23

Другий експеримент був спрямований на дослідження впливу кількості класів на точність класифікації. Результати, наведені у табл. 3, демонструють, що додаткові класи не суттєво впливають на точність системи.

Таблиця 3

Залежність точності класифікації від кількості робочих класів

Кількість класів	3	5	7	10
Точність, %	97,03	94,69	93,26	88,97
Кількість класів	12	14	16	18
Точність, %	88,97	92,78	88,43	88,97

За замовчуванням класифікатор ШНМ віднесе записи невідомого користувача до одного з існуючих класів, що є абсолютно неприпустимим для цілей ідентифікації. Надійність такої системи

буде низькою. Щоб впоратися з цією проблемою, необхідно додати опцію відмови, яка дасть змогу відмінити ідентифікацію, якщо ймовірність правильного “вгадування” мережею є нижчою за певний встановлений поріг. На жаль, це призведе до відкидання мережею деяких записів відомого системі користувача (наприклад, через низьку якість запису). Метою третього експерименту було підібрати найкращий поріг відмови для отримання оптимального відношення між точністю класифікації та коефіцієнтом ідентифікації. Результати наведено на рис. 2. Як показано на графіку, оптимальним пороговим значенням є 70 %, за цього значення точність класифікації становить ~96 % за достатньо високої точності ідентифікації ~90 %.

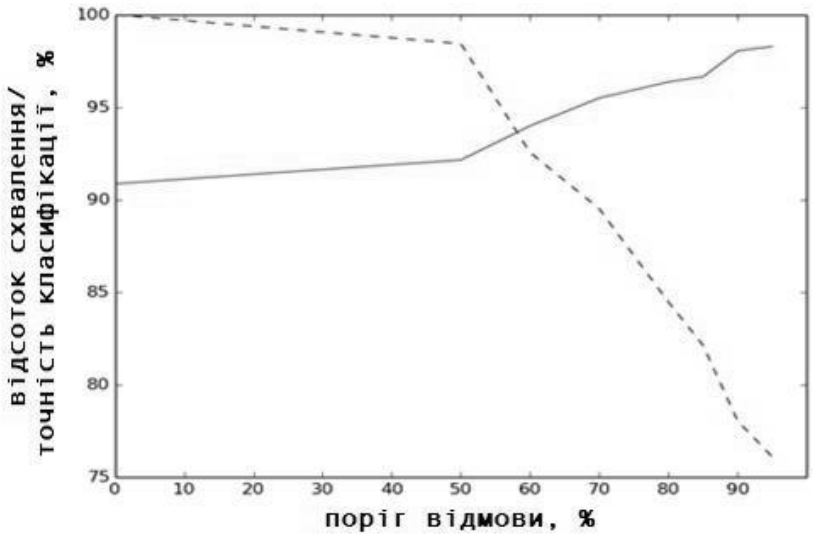

Рис. 2. Точність ідентифікації (суцільна лінія)
та точність ідентифікації (штрихова)

Висновки

Біометрична ідентифікація є ключем до вирішення багатьох проблем у сфері інформаційної безпеки та контролю доступу, автентифікації, цифрових та онлайн-операцій. Використання біосигналів додаткові до загальних біометричних методик, як, наприклад, відбиток пальця, розпізнавання райдужної оболонки та обличчя для багатофакторної ідентифікації є достатньо перспективним підходом. Метою статті було дослідження ідентифікації людини за її електрокардіограмою.

За останні роки глибинне навчання показало вражаючі результати у вирішенні комерційних та наукових проблем. Метою роботи було поєднати методики глибинного навчання з ідентифікацією людини за електрокардіограмою. Цей підхід є перспективним з двох причин:

1. Глибинне навчання зазвичай перевершує більшість інших методів класифікації. У результаті подальші дослідження можуть дати надійніші результати ідентифікації;

2. Глибинне навчання дає змогу виконувати ідентифікацію, використовуючи сирі дані ЕКГ, в обхід етапу підготування ознак, який є невід’ємним за багатьома іншими методиками, а також доволі складним у алгоритмічному і обчислювальному сенсах.

Для запису ЕКГ використовували прилад на основі вбудованого мікроконтролера та диференційного підсилювача. Опрацювання сигналу та класифікацію проводили на ПК. Для зручності вимірювання схему першого відведення було модифіковано, і показники знімали не з грудної клітки, а з пальців двох рук (двох пальців лівої руки та одного пальця правої). Цей підхід демонструє, що біометрична система ЕКГ може бути суттєво менших розмірів та інтегрована надалі з існуючими біометричними системами чи електронними гаджетами.

Виміряні дані зібрано у загальну базу Lviv Biometric Data Set, яка містить 137 записів ЕКГ 18-ти осіб та доступна в інтернеті.

Основні результати досліджень:

- найбільше впливає на точність ідентифікації порівняно з іншими факторами кількість користувачів (класів), а також кількість нейронів та прихованих рівнів у мережі;
- точність ідентифікації відчутно змінюється для тієї самої структури моделі. Можливо, це відбувається через ініціалізацію вагових коефіцієнтів випадковими значеннями і через невипуклу функцію втрат (non-convex cost function). Щоб отримати якомога більшу точність, експеримент необхідно проводити багато разів та використовувати надалі лише модель з найкращими результатами.
- щоб попередити успішну ідентифікацію невідомого системі користувача, запропоновано поріг відмови для відкидання ідентифікацій з низьким рівнем надійності. Експериментально було визначено оптимальний поріг відмови – 70 %, за якого співвідношення точності класифікації та відсотка схвалення є максимальним.

На жаль, отримані результати не повністю відповідають очікуваним. Можливою причиною може бути спотворення ЕКГ, яке не вдалось повністю усунути фільтруванням, а також невелика кількість записів на клас. Точність ідентифікації можна бути підвищити вибираючи іншу архітектуру ШНМ, дієвіший алгоритм фільтрування, збільшуючи кількість даних з використанням для цього генеративних моделей (генеративно-конкуруюча мережа, варіаційний автоенкодер) чи інших методик. Ці пропозиції потребують подальшого вивчення та досліджень.

1. A. K. Jain, P. Flynn, and A. A. Ross, Handbook of Biometrics. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007. 2. Handbook of Biometrics. Jain A., Flynn P., Ross A. A. (Eds.). Springer, 2008, 564 p. 3. Fratini A., Sansone M., Bifulco P., Cesarel M. Individual identification via electrocardiogram analysis. BioMed Eng OnLine. 2015, pp. 1–23. 4. M. Bassiouni, W.Khalefa, El-Sayed A. El-Dahshan, Abdel-Badeeh. M. Salem. A study on the Intelligent Techniques of the ECG-based Biometric Systems. Recent Advances in Electrical Engineering, pp. 26–31. 5. Kaur G., Singh D., Kaur S. Electrocardiogram (ECG) as a Biometric Characteristic: A Review. International Journal of Emerging Research in Management &Technology, 2015, (Volume-4, Issue-5), pp. 202–206. 6. Matos A. C., Lourenc A., Nascimento J. Embedded system for individual recognition based on ECG Biometrics. In: Proceedings Conference on Electronics, Telecommunications and Computers – CETC 2013, pp. 265–272. 7. e-Health Sensor Platform V2.0 for Arduino and Raspberry Pi [Electronic resource]. – Access mode: https://www.cooking-hacks.com/documentation/tutorials/ehealth-biometric-sensor-platform-arduino-raspberry-pi-medical (last access: 21.03.17). – Title from the screen. 8. BioSPPy – Biosignal Processing in Python [Electronic resource]. – Access mode: https://github.com/PIA-Group/BioSPPy (last access: 21.03.17). 9. Lviv Biometric Data Set [Electronic resource]. – 2017 – Access mode: https://github.com/YuriyKhoma/Lviv-Biometric-Data-Set (last access: 21.03.17). 10. The source code of the project [Electronic resource]. – 2017. – Access mode: https://github.com/YuriyKhoma/ecg-identification (last access: 21.03.17). 11. TensorFlow [Electronic resource]. – Access mode: https://www.tensorflow.org/ (last access: 21.03.17).