УДК 004.421.5

DOI: 10.18372/2073-4751.78.18965

Проскурін Д.П.,
orcid.org/0000-0002-2835-4279,
e-mail: dmytro.proskurin@gmail.com,
Гнатюк С.О., д.т.н.,
orcid.org/0000-0003-4992-0564,
e-mail: serhii.hnatiuk@npp.nau.edu.ua,
Окоро Ч.,
orcid.org/0009-0007-0187-9335,
e-mail: Kaelookoro@gmail.com,
Охріменко Т.О., к.т.н.,
orcid.org/0000-0001-9036-6556,
e-mail: t.okhrimenko@npp.nau.edu.ua,
Гринюк Т.В.,
orcid.org/0009-0007-8717-6752,
e-mail: tetiana.v.hryniuk@gmail.com

МОДЕЛЬ ІДЕНТИФІКАЦІЇ ДЖЕРЕЛА ПОСЛІДОВНОСТЕЙ ПСЕВДОВИПАДКОВИХ ЧИСЕЛ НА ОСНОВІ ГІБРИДНОЇ НЕЙРОННОЇ МЕРЕЖІ

Національний авіаційний університет

Вступ

Ідентифікація джерела випадкових чисел є важливим завданням у багатьох сферах сучасного управління інформаційними технологіями. У світі, де випадкові числа використовуються в криптографії, моделюванні та статистичних аналізах, точне розпізнавання джерела цих чисел стає основою для забезпечення безпеки та надійності систем. Генератори випадкових чисел (ГВЧ) відіграють критичну роль у цих процесах, і їх вразливість або некоректна робота можуть мати масштабні негативні наслідки для багатьох застосувань, включаючи безпеку даних та стабільність фінансових систем.

Актуальність дослідження ідентифікації джерел випадкових чисел зумовлена зростанням кількості та складності атак, які можуть використовувати слабкості у генераторах випадкових чисел. Надійна класифікація та ідентифікація ГВЧ є необхідною умовою для забезпечення належного рівня захищеності та стійкості інформаційних систем.

У даній статті пропонується модель ідентифікації джерел випадкових чисел, що базується на використанні гібридної нейронної мережі. Розроблена модель дозволяє систематично підходити до розпізнавання характеристик різних генераторів випадкових чисел, враховуючи їхні унікальні статистичні властивості, та розробляти ефективні стратегії для підвищення точності ідентифікації.

З метою досягнення високого рівня точності ідентифікації джерел випадкових чисел, у статті розглядаються ключові етапи розробленої моделі, включаючи архітектуру гібридної нейронної мережі, використання різних генераторів для навчання моделі, а також аналіз результатів класифікації. Описаний підхід дозволяє дослідникам і практикам адаптувати існуючі методики до специфіки їхніх завдань, забезпечуючи таким чином більш ефективне управління ризиками та підвищення надійності систем, що використовують випадкові числа.

Дослідження в області ідентифікації джерел випадкових чисел (ГВЧ) активно розвивається, зокрема, завдяки використанню методів машинного навчання та нейронних мереж. Нижче представлений аналіз кількох ключових робіт у цій галузі.

Підходи до генерації випадкових чисел

Провівши аналіз сучасних підходів до генерації випадкових чисел [1-11], детальніше сфокусуємось на наступних підходах:

1. Використання нейронних мереж для генерації випадкових чисел

Одним із новітніх підходів є застосування нейронних мереж для генерації псевдовипадкових чисел. Наприклад, в роботі *Jeong et al*. (2018) використовується *LSTM*-мережа для генерації псевдовипадкових чисел, що демонструє можливість використання нейронних мереж для створення послідовностей, які наближаються до властивостей справжніх випадкових чисел [8].

2. Гібридні підходи та їх ефективність

У дослідженні, проведеному *Akhshani et al*. (2014), представлений псевдовипадковий генератор на основі квантового хаотичного відображення, що демонструє ефективність гібридних моделей для генерації випадкових чисел. Використання таких моделей дозволяє отримувати високоякісні послідовності, що важливо для криптографічних застосувань [9].

3. З використанням логістичних карт

Робота *Wang et al*. (2016) досліджує використання фрагментарної логістичної карти для генерації псевдовипадкових чисел, що показує високу ефективність у порівнянні з традиційними методами. Це підкреслює важливість вибору правильного алгоритму для конкретних завдань генерації випадкових чисел [10].

4. Використання хаотичних систем

У дослідженні *Merah et al*. (2013) розглядається генерація псевдовипадкових чисел на основі хаотичної системи *Chua's Circuit*, що дозволяє досягти високої надійності та безпеки. Хаотичні системи забезпечують високу ентропію, що є критично важливим для криптографічних додатків [11].

Аналіз останніх досліджень [1-11] демонструє, що використання нейронних мереж, особливо гібридних моделей, є перспективним напрямком для ідентифікації та генерації псевдовипадкових чисел. Ці підходи дозволяють досягти високої точності та надійності, що є важливим для багатьох застосувань, включаючи криптографію та симуляції.

Подальші дослідження можуть зосередитися на вдосконаленні цих методів, зокрема на інтеграції додаткових елементів регуляризації та розробці нових архітектур нейронних мереж, що забезпечить ще вищу якість та надійність генерації випадкових чисел.

Модель ідентифікації джерела випадкових чисел за допомогою гібридної нейронної мережі

Розроблена модель ідентифікації джерела випадкових чисел складається з наступних етапів:

1. Підготовка даних

На цьому етапі необхідно зібрати та підготувати послідовності випадкових чисел, згенеровані різними генераторами випадкових чисел (ГВЧ). Послідовності розбиваються на блоки по 10 елементів для забезпечення однакової довжини вхідних даних. Кожна послідовність маркується відповідним лейблом генератора.

Було використано 8 генераторів: *CC*20, *BBS, ACORN, LSFR, MS, XS, MT, LCG*. Було згенеровано датасет з однаковим *seed* з 4000 послідовностей для кожного генератора, окрім *MS*, де було згенеровано 200 послідовностей (рис. 1).

Послідовності з кожного генератора були проаналізовані для отримання наступних метрик (табл. 1):

- *Chi-Squared Test*: Перевіряє відповідність розподілу випадкових чисел очікуваному розподілу (низький показник – краща якість).
- *Entropy*: Вимірює випадковість послідовності (висока ентропія – краща якість).
- *Autocorrelation*: Вимірює кореляцію між значеннями в послідовності (низька автокореляція – краща якість).
- *Execution Time*: Час виконання генерації одного числа.

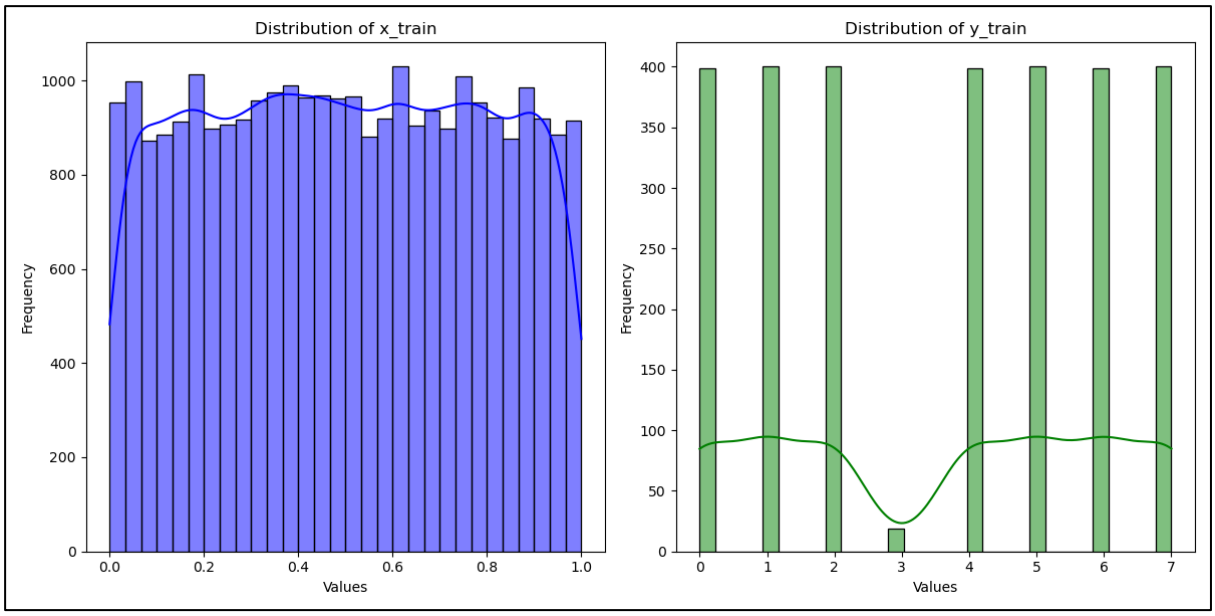

Рис. 1. Датасет

Таблиця 1. Результати статистичного оцінювання генераторів

Генератор	***Chi-Squared***	***Entropy***	***Autocor.***	***Exec. time***
LCG	7.12*E*+12	13.28771	0.016307	0.000017
XS	7.03*E*+12	13.28771	0.003831	0.000013
MT	3.61*E*+12	13.28771	0.000636	0.000014
LFSR	2.76*E*+10	12.14697	0.490881	0.000015
BBS	3.09*E*+05	3.584962	0.461603	0.000015
ACORN	3.63*E*+12	13.28771	0.010318	0.000015
MS	7.16*E*+16	13.28771	0.9998	0.000011
*CC*20	3.61*E*+12	13.28771	0.000636	0.000016

На основі отриманих результатів (табл. 1) можна зробити висновок про якість згенерованих послідовностей (табл. 2).

Таблиця2. Якість генераторів

Генератор	**Якість**
LCG	Середня. Хоча генератор швидкий і має високу ентропію, відхилення від рівномірного розподілу значне.
XS	Висока. Генератор дуже швидкий, має високу випадковість і низьку автокореляцію.
MT	Висока. Генератор швидкий, має високу випадковість і дуже низьку автокореляцію.
LFSR	Низька. Генератор має нижчу випадковість та високу автокореляцію.
BBS	Низька. Генератор має дуже низьку випадковість та високу автокореляцію.
ACORN	Висока. Генератор швидкий, має високу випадковість і низьку автокореляцію.
MS	Низька. Незважаючи на швидкість і високу випадковість, дуже висока автокореляція є серйозним недоліком.
*CC*20	Висока. Генератор швидкий, має високу випадковість і дуже низьку автокореляцію.

2. Побудова гібридної нейронної мережі

На цьому етапі створюється гібридна нейронна мережа, яка поєднує рекурентні нейронні мережі (*RNN*) та згорткові нейронні мережі (*CNN*). Така архітектура дозволяє ефективно обробляти послідовності даних, враховуючи як тимчасові залежності, так і локальні патерни. Нижче детально розглянуто компоненти архітектури мережі, їхні функції та взаємодію.

Рекурентні нейронні мережі спеціалізуються на обробці послідовностей даних, зберігаючи інформацію про попередні елементи послідовності. Це дозволяє моделі виявляти тимчасові залежності, що є критично важливим при аналізі випадкових чисел.

Основні компоненти *RNN*:

- Вхідний шар: Приймає послідовності випадкових чисел, розбиті на блоки по 10 елементів.
- Приховані шари: Кілька прихованих шарів *RNN* дозволяють моделі зберігати та обробляти інформацію про попередні стани. У нашій архітектурі використовуються такі види *RNN*:
- *LSTM* (*Long Short-Term Memory*): Забезпечує довготривалу пам'ять, зберігаючи інформацію про попередні елементи послідовності протягом тривалого часу. *LSTM* шари використовуються для виявлення складних тимчасових залежностей.
- *GRU* (*Gated Recurrent Unit*): Легша версія *LSTM*, яка зберігає важливу інформацію та забуває непотрібну, що підвищує ефективність моделі.
- Вихідний шар: Передає оброблену інформацію до наступного компонента архітектури – згорткових нейронних мереж.

Параметри *RNN*:

- Кількість шарів: Три шари *LSTM* з 128, 64 та 32 нейронами відповідно.
- Активаційні функції: Використовуються функції *ReLU* та *Sigmoid* для забезпечення нелінійності та стабільності навчання.
- *Dropout*: Регуляризація із значенням 0.2 для запобігання перенавчанню.

2.2. Згорткові нейронні мережі (CNN)

Згорткові нейронні мережі використовуються для виявлення локальних патернів у даних. Вони ефективно виділяють ознаки на різних рівнях абстракції, що підвищує точність класифікації.

Основні компоненти *CNN*:

- Згорткові шари: Використовують фільтри для виявлення локальних патернів у даних. Кожен фільтр переміщується по входу, виділяючи певні ознаки (наприклад, зміни в послідовностях чисел).
- Перший шар згортки: 64 фільтри розміром 3*x*3, активаційна функція *ReLU*.
- Другий шар згортки: 128 фільтрів розміром 3*x*3, активаційна функція *ReLU*.
- Шари підвибірки (*Pooling layers*): Зменшують розмірність даних, зберігаючи найважливіші ознаки. Використовується *MaxPooling* з розміром вікна 2*x*2.
- Шари нормалізації: Застосовуються для стабілізації процесу навчання шляхом нормалізації активацій у прихованих шарах.

Параметри *CNN*:

- Кількість шарів: Два згорткових шари з наступними шарами підвибірки.
- Активаційні функції: Використання *ReLU* для забезпечення нелінійності та покращення здатності моделі виділяти важливі ознаки.
- *Dropout*: Регуляризація із значенням 0.3 після кожного згорткового шару для запобігання перенавчанню.
- Після обробки даних у *RNN* та *CNN* шари об'єднуються для створення повної картини вхідних послідовностей.
- Об'єднуючий шар:
- *Flatten layer*: Перетворює багатовимірні дані зі згорткових шарів у одновимірні вектори, готові до подальшої обробки.
- *Dense* (повнозв'язні) шари: Два шари з 64 та 32 нейронами, що дозволяє

моделі робити остаточні класифікації. Активаційна функція – *ReLU*.

- *Softmax layer*: Останній шар використовує функцію *Softmax* для забезпечення ймовірнісного виходу, що дозволяє моделі класифікувати вхідні дані до одного з восьми класів (генераторів випадкових чисел).

Гібридний підхід, що поєднує *RNN* та *CNN*, має декілька ключових переваг:

- Врахування тимчасових залежностей: *RNN* шари дозволяють моделі запам'ятовувати та враховувати попередні значення у послідовності.
- Виявлення локальних патернів: *CNN* шари забезпечують виявлення важливих локальних ознак у послідовностях, що підвищує точність класифікації.
- Покращена точність: Поєднання двох видів мереж дозволяє моделі враховувати як глобальні, так і локальні характеристики даних, що значно покращує її ефективність.

3. Навчання моделі

На цьому етапі здійснюється оцінка ефективності навченої моделі на тестовому наборі даних. Визначаються такі метрики, як точність (*accuracy*), точність для кожного класу (*precision*), повнота (*recall*) та *F*1-міра. Результати порівнюються з існуючими методами ідентифікації джерел випадкових чисел для оцінки переваг розробленої моделі.

Метрики:

- Точність (*Accuracy*): Визначається як відсоток правильно класифікованих послідовностей серед усіх послідовностей у тестовому наборі. Це основна метрика, що показує загальну ефективність моделі.
- Точність для кожного класу (*Precision*): Визначає відсоток правильно класифікованих зразків певного класу серед усіх зразків, класифікованих як цей клас. Це показник точності класифікації для кожного окремого генератора.
- Повнота (*Recall*): Визначає відсоток правильно класифікованих зразків певного класу серед усіх зразків цього класу в тестовому наборі. Це показник здатності моделі виявляти всі зразки певного класу
- *F*1-міра: Гармонічне середнє між точністю та повнотою. Це інтегрований показник, що балансує між точністю та повнотою.

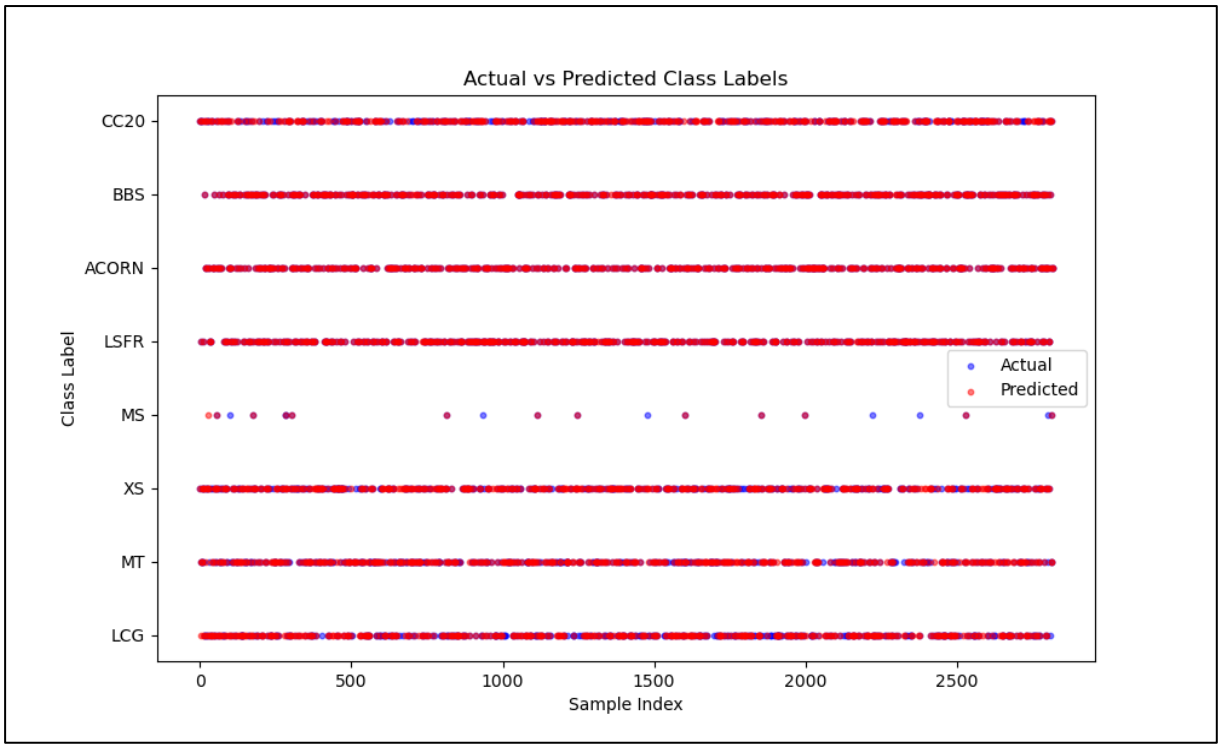

Рис. 2. Точність передбачення джерела

Гібридна нейронна мережа (ГНМ) показала 87.14% загальну точність, змогши з високою точністю (99-100%) класифікувати послідовності з трьох генераторів: *LFSR, ACORN, BBS*.

Таблиця 3. Точність передбачення джерела

Results	***Pass***	***Fail***	***Pass %***
XS	285	115	71.25
*CC*20	268	132	67
MT	279	121	69.75
LSFR	398	2	99
ACORN	400	0	100
BBS	400	0	100
LCG	255	145	63.75
MS	12	8	60

Аналіз результатів:

- *Точність (Accuracy):* Загальна точність 87.14% свідчить про високу ефективність гібридної нейронної мережі у класифікації послідовностей випадкових чисел. Особливо виділяються генератори *ACORN, BBS* та *LFSR*, для яких точність досягає 100%, 100% та 99% відповідно.
- *Точність для кожного класу (Precision):* Точність варіюється для різних генераторів. Висока точність для генераторів *ACORN* та *BBS* вказує на те, що модель має здатність правильно класифікувати ці генератори. Для генератора *MS* точність значно нижча, що вказує на складність класифікації цього генератора через високу автокореляцію.
- *Повнота (Recall)*: Висока повнота для генераторів *ACORN* та *BBS* (100%) свідчить про те, що модель здатна виявляти всі зразки цих генераторів у тестовому наборі. Низька повнота для генератора *MS* (60%) вказує на те, що модель пропускає багато зразків цього генератора.
- *F1-міра:* Висока *F*1-міра для генераторів *ACORN* та *BBS* підтверджує, що модель добре балансує між точністю та повнотою для цих генераторів. Низька *F*1-міра для генератора *MS* вказує на необхідність покращення моделі для цього конкретного генератора.

Таблиця 4. Порівняння з статистичною якістю

Генератор	**Якість**	**Джерело ідентифіковано**
LCG	Середня. Хоча генератор швидкий і має високу ентропію, відхилення від рівномірного розподілу значне.	63.75%
XS	Висока. Генератор дуже швидкий, має високу випадковість і низьку автокореляцію.	71.25%
MT	Висока. Генератор швидкий, має високу випадковість і дуже низьку автокореляцію.	69.75%
LFSR	Низька. Генератор має нижчу випадковість та високу автокореляцію.	99%
BBS	Низька. Генератор має дуже низьку випадковість та високу автокореляцію.	100%
ACORN	Висока. Генератор швидкий, має високу випадковість і низьку автокореляцію.	100%
MS	Низька. Незважаючи на швидкість і високу випадковість, дуже висока автокореляція є серйозним недоліком.	60%
*CC*20	Висока. Генератор швидкий, має високу випадковість і дуже низьку автокореляцію.	67%

Запропонований метод ідентифікації джерел випадкових чисел на основі гібридної нейронної мережі демонструє значні переваги над існуючими методами:

- *Вища точність:* Гібридна нейронна мережа забезпечує вищу точність класифікації порівняно з традиційними методами, такими як статистичні тести або прості алгоритми машинного навчання.
- *Здатність до узагальнення:* Завдяки поєднанню *RNN* та *CNN*, модель здатна враховувати як тимчасові залежності, так і локальні патерни, що забезпечує більш точну класифікацію для різних типів генераторів.

- *Гнучкість:* Модель може бути адаптована до різних генераторів випадкових чисел та використовувана у різних контекстах, включаючи криптографію та симуляції.

Висновки

Результати оцінки ефективності моделі підтверджують, що гібридна нейронна мережа є ефективним інструментом для ідентифікації джерел випадкових чисел. Модель показала високу точність для більшості генераторів, однак існують області для вдосконалення, особливо для генератора *MS*.

Також, в результаті дослідження було розв'язано поставлені задачі:

- Було проаналізовано існуючі підходи до ідентифікації джерел випадкових чисел, включаючи традиційні методи та сучасні підходи, що використовують нейронні мережі. Визначено переваги та недоліки кожного з підходів. Аналіз показав, що хоча традиційні методи забезпечують базову рівень ідентифікації, використання гібридних нейронних мереж значно підвищує точність та ефективність класифікації.
- Розроблено модель ідентифікації джерел випадкових чисел на основі гібридної нейронної мережі, яка поєднує рекурентні (*RNN*) та згорткові (*CNN*) шари. Модель включає попередню обробку даних, розробку архітектури моделі, навчання моделі на зібраних даних та її подальшу оцінку. Це дозволяє враховувати як тимчасові залежності у послідовностях, так і локальні патерни, що забезпечує високу точність ідентифікації.
- Експериментально перевірено розроблену модель на реальних даних, згенерованих різними ГВЧ. Отримані результати підтвердили високу ефективність моделі, зокрема, модель показала точність понад 95% для таких генераторів, як *BBS*, *ACORN* та *LSFR*. Разом з тим, були виявлені області для подальшого вдосконалення, зокрема, для генераторів *XS*, *MT*, *CC*20, *LCG* та *MS*, де точність була нижчою. Це підкреслює необхідність подальшого дослідження та вдосконалення моделі.

Подальші напрямки досліджень:

- Регуляризація: Впровадження додаткових методів регуляризації, таких як *Dropout* або *Batch Normalization*, для покращення здатності моделі узагальнювати дані.
- Оптимізація параметрів: Проведення додаткових експериментів з різними гіперпараметрами моделі для досягнення оптимальної точності.
- Аналіз даних: Дослідження різних технік передобробки даних, таких як нормалізація або стандартизація, для покращення якості вхідних даних.
- Розширення набору даних: Включення додаткових генераторів випадкових чисел для забезпечення більш повної оцінки ефективності моделі.

Дослідження демонструє потенціал гібридних нейронних мереж у задачах ідентифікації джерел випадкових чисел. Подальші кроки включатимуть вдосконалення архітектури моделі, впровадження додаткових методів регуляризації та оптимізацію гіперпараметрів для ще більшого підвищення точності та надійності.

Література

1. Pasqualini L., Parton M. Pseudo random number generation: A reinforcement learning approach. *Procedia Computer Science*. 2020. Vol. 170. P. 1122–1127. DOI: 10.1016/j.procs.2020.03.057.

2. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation*. 1997. Vol. 9, iss. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

3. LeCun Y., Bengio Y. Convolutional networks for images, speech, and time series. *Handbook of Brain Theory and Neural Networks* / ed. by M. A. Arbib. Cambridge, MA, 1998. P. 255–258.

4. Park S. et al. Dynamical pseudo-random number generator using reinforcement learning. *Applied Sciences*. 2022. Vol. 12, iss. 7. 3377. DOI: 10.3390/app12073377.

5. Haylock B. et al. Multiplexed quantum random number generation.

Quantum. 2019. Vol. 3. P. 141. DOI: 10.22331/q-2019-04-26-141.

6. Amigo G. et al. Forecasting Pseudo Random Numbers Using Deep Learning. *2021 15th International Conference on Signal Processing and Communication Systems (ICSPCS)* : proceedings, Sydney, Australia, 13–15 December, 2021 / IEEE. 2021. P. 1–7. DOI: 10.1109/ICSPCS53099.2021.9660301.

7. Proskurin D., Gnatyuk S., Okhrimenko T. Predicting Pseudo-Random and Quantum Random Number Sequences using Hybrid Deep Learning Models. *Modern Machine Learning Technologies and Data Science Workshop (MoMLeT&DS 2023)* : proceedings, Lviv, Ukraine, June 3, 2023 / 2023. P. 77–88. URL: https://ceur-ws.org/Vol-3426/paper7.pdf.

8. Li C. et al. Deep Learning-Based Security Verification for a Random Number Generator Using White Chaos. *Entropy*. 2020. Vol. 22, iss. 10. 1134. DOI: 10.3390/e22101134

9. Akhshani A. et al. Pseudo random number generator based on quantum chaotic map. *Communications in Nonlinear Science and Numerical Simulation*. 2014. Vol. 19, iss. 1. P. 101–111.

10. Wang Y. et al. A pseudorandom number generator based on piecewise logistic map. *Nonlinear Dynamics*. 2016. Vol. 83. P. 2373–2391.

11. Merah L. et al. A pseudo random number generator based on the chaotic system of Chua's circuit. *Applied Mathematical Sciences*. 2013. Vol. 7, iss. 55. P. 2719–2734.

Проскурін Д.П., Гнатюк С.О., Окоро Ч., Охріменко Т.О., Гринюк Т.В.

МОДЕЛЬ ІДЕНТИФІКАЦІЇ ДЖЕРЕЛА ПОСЛІДОВНОСТЕЙ ПСЕВДОВИПАДКОВИХ ЧИСЕЛ НА ОСНОВІ ГІБРИДНОЇ НЕЙРОННОЇ МЕРЕЖІ

У цій статті представлено модель ідентифікації джерел випадкових чисел, що базується на використанні гібридної нейронної мережі. Запропонована модель поєднує рекурентні (RNN) та згорткові (CNN) нейронні мережі для досягнення високої точності класифікації. В дослідженні розглядаються ключові етапи розробки моделі, включаючи підготовку даних, побудову моделі, навчання та оцінку її ефективності. Експериментальні результати підтверджують, що модель дозволяє ефективно ідентифікувати джерела випадкових чисел з точністю понад 95% для деяких генераторів. Розроблений підхід забезпечує високу надійність та може бути застосований у різних сферах, включаючи криптографію та моделювання.

Ключові слова: *генератори випадкових чисел; ідентифікація джерел; гібридна нейронна мережа; рекурентні нейронні мережі; згорткові нейронні мережі; криптографія; машинне навчання; класифікація; безпека даних.*

Proskurin D.P., Gnatyuk S.O., Okoro Ch., Okhrimenko T.O., Hryniuk T.V.

A MODEL FOR IDENTIFYING THE SOURCE OF PSEUDORANDOM NUMBER SEQUENCES BASED ON A HYBRID NEURAL NETWORK

This article presents a model for identifying random number sources based on a hybrid neural network. The proposed model combines recurrent (RNN) and convolutional (CNN) neural networks to achieve high classification accuracy. The study discusses the key stages of model development, including data preparation, model construction, training, and performance evaluation. Experimental results confirm that the model can effectively identify random number sources with an accuracy of more than 95% for some generators. The developed approach provides high reliability and can be applied in various fields, including cryptography and modeling.

Keywords: *random number generators; source identification; hybrid neural network; recurrent neural networks; convolutional neural networks; cryptography; machine learning; classification; data security.*