p-ISSN 1607-3274
e-ISSN 2313-688X

# Радіоелектроніка
# Інформатика
# Управління

# Radio Electronics
# Computer Science
# Control

# Радиоэлектроника
# Информатика
# Управление

42
9 771607 327005

## 2024/2

Національний університет «Запорізька політехніка»

# Радіоелектроніка, інформатика, управління

Науковий журнал
Виходить чотири рази на рік
№ 2(69) 2024
Заснований у 1998 році, видається з 1999 року.
Засновник і видавець – Національний університет «Запорізька політехніка».
ISSN 1607-3274 (друкований), ISSN 2313-688X (електронний).
Запоріжжя
НУ «Запорізька політехніка»
2024

National University Zaporizhzhia Polytechnic

# Radio Electronics, Computer Science, Control

The scientific journal
Published four times per year
№ 2(69) 2024
Founded in 1998, published since 1999.
Founder and publisher – National University Zaporizhzhia Polytechnic.
ISSN 1607-3274 (print), ISSN 2313-688X (on-line).
Zaporizhzhia
NU Zaporizhzhia Polytechnic
2024

Национальный университет «Запорожская политехника»

# Радиоэлектроника, информатика, управление

Научный журнал
Выходит четыре раза в год
№ 2(69) 2024
Основан в 1998 году, издается с 1999 года.
Основатель и издатель – Национальный университет «Запорожская политехника».
ISSN 1607-3274 (печатный), ISSN 2313-688X (электронный).
Запорожье
НУ «Запорожская политехника»
2024

**Науковий журнал «Радіоелектроніка, інформатика, управління»** (скорочена назва – РІУ) видається Національним університетом «Запорізька політехніка» (НУ «Запорізька політехніка») з 1999 р. періодичністю чотири номери на рік.

Зареєстровано у Міністерстві юстиції України 19.11.2019 р. (Свідоцтво про державну реєстрацію друкованого засобу масової інформації серія КВ № 24220-14060 ПР.)

**ISSN** 1607-3274 (друкований), **ISSN** 2313-688X (електронний).

Наказом Міністерства освіти і науки України № 409 від 17.03.2020 р. «Про затвердження рішень Атестаційної колегії Міністерства щодо діяльності спеціалізованих вчених рад від 06 березня 2020 року» **журнал включений до переліку наукових фахових видань України в категої «А» (найвищій рівень),** в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук і доктора філософії (кандидата наук).

Журнал включений до польського Переліку наукових журналів та рецензованих матеріалів міжнародних конференцій з присвоєною кількістю балів (додаток до оголошення Міністра науки та вищої освіти Республіки Польща від 31 липня 2019 р.: № 16981).

В журналі безкоштовно публікуються наукові статті англійською, російською та українською мовами.

**Правила оформлення статей** подано на сайті: http://ric.zntu.edu.ua/information/authors.

Журнал забезпечує **безкоштовний відкритий он-лайн доступ** до повнотекстових публікацій.

Журнал дозволяє авторам мати авторські права і зберігати права на видання без обмежень. Журнал дозволяє користувачам читати, завантажувати, копіювати, поширювати, друкувати, шукати або посилатися на повні тексти своїх статей. Журнал дозволяє повторне використання його вмісту у відповідності Creative Commons ліцензією CC BY-SA..

Опублікованим статтям присвоюється унікальний ідентифікатор цифрового об'єкта DOI.

**Журнал входить до наукометричної бази Web of Science.**

**Журнал реферується та індексується** у провідних міжнародних та національних реферативних журналах і наукометричних базах даних, а також розміщується у цифрових архівах та бібліотеках з безкоштовним доступом у режимі on-line, повний перелік яких подано на сайті: http://ric.zntu.edu.ua/about/editorialPolicies#custom-0.

**Журнал розповсюджується** за Каталогом періодичних видань України (передплатний індекс – 22914).

**Тематика журналу**: телекомунікації та радіоелектроніка, програмна інженерія (включаючи теорію алгоритмів і програмування), комп'ютерні науки (математичне і комп'ютерне моделювання, оптимізація і дослідження операцій, управління в технічних системах, міжмашинна і людино-машинна взаємодія, штучний інтелект, включаючи системи, засновані на знаннях, і експертні системи, інтелектуальний аналіз даних, розпізнавання образів, штучні нейронні і нейро-нечіткі мережі, нечітку логіку, колективний інтелект і мультиагентні системи, гібридні системи), комп'ютерна інженерія (апаратне забезпечення обчислювальної техніки, комп'ютерні мережі), інформаційні системи та технології (структури та бази даних, системи, засновані на знаннях та експертні системи, обробка даних і сигналів).

Усі статті, пропоновані до публікації, одержують **об'єктивний розгляд**, що оцінюється за суттю без урахування раси, статі, віросповідання, етнічного походження, громадянства або політичної філософії автора(ів).

Усі статті проходять двоступінчасте закрите (анонімне для автора) **резензування** штатними редакторами і незалежними рецензентами – провідними вченими за профілем журналу.

## РЕДАКЦІЙНА КОЛЕГІЯ



## РЕДАКЦІЙНО-КОНСУЛЬТАТИВНА РАДА



Рекомендовано до видання Вченою радою НУ «Запорізька політехніка», протокол № 10 від 04.06.2024.

Журнал зверстаний редакційно-видавничим відділом НУ «Запорізька політехніка».

**Веб-сайт журналу**: http://ric.zntu.edu.ua.

**Адреса редакції:** Редакція журналу «РІУ», Національний університет «Запорізька політехніка», вул. Жуковського, 64, м. Запоріжжя, 69063, Україна.

Тел: (061) 769-82-96 – редакційно-видавничий відділ Факс: +38-061-764-46-62

E-mail: rvv@zntu.edu.ua

**The scientific journal Radio Electronics, Computer Science, Control** is published by the National University «Zaporizhzhia Polytechnic» NU «Zaporizhzhia Polytechnic» since 1999 with periodicity four numbers per year.

The journal is registered by the Ministry of Justice of Ukraine in 19.11.2019. (State Registration Certificate of printed mass media series КВ № 24220-14060 ПР).

**ISSN** 1607-3274 (print), **ISSN** 2313-688X (on-line).

By the Order of the Ministry of Education and Science of Ukraine from 17.03.2020 № 409 "On approval of the decision of the Certifying Collegium of the Ministry on the activities of the specialized scientific councils dated 06 March 2020" **journal is included in the list of scientific specialized periodicals of Ukraine in category "A" (highest level),** where the results of dissertations for Doctor of Science and Doctor of Philosophy may be published.

The journal is included to tne Polish List of scientific journals and peer-reviewed materials from international conferences with assigned number of points (Annex to the announcement of the Minister of Science and Higher Education of Poland from July 31, 2019: Lp. 16981).

The journal publishes scientific articles in English, Russian, and Ukrainian free of charge.

The **article formatting rules** are presented on the site: http://ric.zntu.edu.ua/information/authors.

The journal provides policy of **on-line open (free of charge) access** for full-text publications. The journal allow the authors to hold the copyright without restrictions and to retain publishing rights without restrictions. The journal allow readers to read, download, copy, distribute, print, search, or link to the full texts of its articles. The journal allow reuse and remixing of its content, in accordance with Creative Commons license CC BY-SA.

Published articles have a unique digital object identifier (DOI).

**The journal is included into Web of Science.**

**The journal is abstracted and indexed** in leading international and national abstractig journals and scientometric databases, and also placed to the digital archives and libraries with a free on-line access, full list of which is presented at the site: http://ric.zntu.edu.ua/about/editorialPolicies#custom-0.

**The journal is distributed**:by the Catalogue of Ukrainian periodicals (the catalog number is 22914).

**The journal scope:** telecommunications and radio electronics, software engineering (including algorithm and programming theory), computer science (mathematical modeling and computer simulation, optimization and operations research, control in technical systems, machine-machine and man-machine interfacing, artificial intelligence, including data mining, pattern recognition, artificial neural and neuro-fuzzy networks, fuzzy logic, swarm intelligence and multiagent systems, hybrid systems), computer engineering (computer hardware, computer networks), information systems and technologies (data structures and bases, knowledge-based and expert systems, data and signal processing methods).

All articles proposed for publication receive an **objective review** that evaluates substantially without regard to race, sex, religion, ethnic origin, nationality, or political philosophy of the author(s).

All articles undergo a two-stage **blind peer review** by the editorial staff and independent reviewers – the leading scientists on the profile of the journal.

## EDITORIAL BOARD



## EDITORIAL-ADVISORY COUNCIL



Recommended for publication by the Academic Council of NU «Zaporizhzhia Polytechnic», protocol № 10 dated 04.06.2024.

The journal is imposed by the editorial-publishing department of NU Zaporizhzhia Polytechnic.

**The journal web-site** is http://ric.zntu.edu.ua.

**The address of the editorial office:** Editorial office of the journal Radio Electronics, Computer Science, Control, National University Zaporizhzhia Polytechnic, Zhukovskiy street, 64, Zaporizhzhia, 69063, Ukraine.

Tel.: +38-061-769-82-96 – the editorial-publishing department. Fax: +38-061-764-46-62

E-mail: rvv@zntu.edu.ua

# ЗМІСТ

# CONTENTS

# РАДІОЕЛЕКТРОНІКА ТА ТЕЛЕКОМУНІКАЦІЇ

# RADIO ELECTRONICS AND TELECOMMUNICATIONS

УДК 681.3.06

## УДОСКОНАЛЕНИЙ МЕТОД ОЦІНЮВАННЯ НАДІЙНОСТІ ОБ'ЄКТІВ ЗІ ЗМІННОЮ СТРУКТУРОЮ

**Бабій О. С.** – старший викладач кафедри військово-технічної підготовки, Київський національний університет імені Тараса Шевченка, Київ, Україна.

**Сакович Л. М.** – канд. техн. наук, доцент, доцент спеціальної кафедри Інституту спеціального зв'язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

**Слюсарчук О. О.** – канд. військових наук, старший науковий співробітник, старший дослідник Науково-дослідного інституту воєнної розвідки Головного управління розвідки міністерства оборони України, Київ, Україна.

**Єлісов Ю. М.** – канд. техн. наук, науковий співробітник Науково-дослідного інституту воєнної розвідки Головного управління розвідки міністерства оборони України, Київ, Україна.

**Курята Я. Е.** – начальник науково-організаційного відділу науково-дослідного центру, Інститут спеціального зв'язку та захисту інформації Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

### АНОТАЦІЯ

**Актуальність.** Основна ідея – врахування можливості впливу прихованих дефектів на надійність багаторежимних радіоелектронних засобів зі змінною структурою, що не враховують відомі методи розрахунку показників надійності. Пропонується кількісна оцінка параметра потоку відмов виробів з врахуванням впливу на його значення накопичення прихованих дефектів.

**Мета.** Удосконалення методу оцінки надійності об'єктів зі змінною структурою врахуванням можливості виникнення прихованих дефектів під час використанням за призначенням в окремих режимах роботи.

**Метод.** Використовується методологія оцінки значень показників надійності складних технічних систем. Метод, що розробляється, є розвитком алгоритму оцінки показників надійності багаторежимних об'єктів в напрямку врахування можливості появи і накопичення прихованих дефектів в підмножинах елементів об'єкту, які при його роботі в окремих режимах не використовують.

**Результати.** Отримані функціональні залежності часткових і комплексних показників надійності багаторежимних об'єктів від накопичення прихованих дефектів, які проявляються тільки під час технічного обслуговування або зміни режимів роботи. Рішення формалізовано у вигляді алгоритму, що використовує результати дослідної експлуатації виробів в якості вихідних даних.

**Висновки.** Наукова новизна полягає у розробці наступних інноваційних рішень: 1) вперше запропоновано враховувати наявність прихованих дефектів при оцінці надійності багаторежимних об'єктів зі змінною структурою; 2) вперше отримано і досліджено функціональні залежності впливу наявності прихованих дефектів на значення часткових та комплексних показників надійності. Практичне значення результатів полягає в тому, що це дозволяє на етапі дослідної експлуатації радіоелектронних засобів зі змінною структурою об'єктивно оцінити відповідність розрахунків необхідним значенням показників надійності завдяки врахуванню можливості появи прихованих дефектів.

**КЛЮЧОВІ СЛОВА:** оцінювання показників надійності, об'єкти зі змінною структурою, приховані дефекти.

### АБРЕВІАТУРИ

БРО – багаторежимні об'єкти;
ЗВТ – засіб вимірювальної техніки;
ОЗС – об'єкти зі змінною структурою;
РЕЗ – радіоелектронні засоби;
УАД – умовний алгоритм діагностування.

### НОМЕНКЛАТУРА

$A$ – коефіцієнт готовності об'єкта;

$A'$ – коефіцієнт готовності об'єкта без врахування прихованих дефектів;

$\alpha_i$ – коефіцієнт прихованих відмов в підмножині елементів $L_i$;

$\alpha$ – середнє значення коефіцієнту прихованих відмов в об'єкті;

$L_i$ – підмножина елементів, що використовується в режимі роботи $i$;

$Q_i$ – кількість визначених прихованих відмов в режимі роботи $i$;


DOI 10.15588/1607-3274-2024-2-1

$Q$ – загальна кількість відмов в об'єкті за час роботи $T_р$;

$K$ – середня кількість перевірок при пошуку дефекту;

$K_i$ – середня кількість перевірок при відмові об'єкту в режимі роботи i;

$N_i$ – коефіцієнт врахування впливу прихованих дефектів на параметр потоку відмов;

$P$ – ймовірність вірної постановки діагнозу;

$P(\tau)$ – ймовірність безвідмовної роботи засобів вимірювальної техніки в міжперевірочний період τ;

$p$ – ймовірність вірної оцінки результату виконання перевірки;

$R$ – кількість підмножин елементів залежно від режимів роботи;

$T$ – наробіток об'єкту на відмову;

$T'$ – наробіток об'єкту на відмову без врахування прихованих дефектів;

$T_в$ – середній час відновлення об'єкту;

$T_в'$ – середній час відновлення об'єкту без врахування прихованих дефектів;

$T_{вр}$ – розрахунковий час відновлення об'єкту при відмові;

$T_{вр}'$ – розрахунковий час відновлення об'єкту без врахування прихованих дефектів;

$T_{вп}$ – припустиме значення середнього часу відновлення виробу після відмови;

$T_i$ – загальний час роботи об'єкта в режимі i;

$T_n$ – припустиме значення наробітку виробу на відмову;

$T_p$ – загальний час роботи об'єкта;

$t$ – середній час виконання перевірки;

$t_y$ – середній час усунення несправності;

$U$ – коефіцієнт неготовності об'єкта;

$U'$ – коефіцієнт неготовності об'єкта без врахування прихованих дефектів;

$u_i$ – відносний час роботи підмножини елементів $L_i$;

$Z$ – параметр потоку відмов об'єкта;

$Z'$ – параметр потоку відмов об'єкта без врахування прихованих дефектів;

$Z''$ – параметр потоку відмов об'єкта від прихованих дефектів;

$Z_i$ – параметр потоку відмов об'єкта в режимі роботи $i$;

$\eta_T$ – ефект від уточнення значення наробітку об'єкту на відмову;

$\eta_U$ – ефект від уточнення оцінки значення коефіцієнту неготовності.

## ВСТУП

Мотивація дослідження продиктована тим, що сучасні радіоелектронні засоби (РЕЗ) відносяться до багаторежимних об'єктів (БРО), які під час використання за призначенням застосовують різні сукупності елементів, тобто є об'єктами зі змінною структурою (ОЗС). При цьому в частині об'єкту, яка не використовується, можливе поява і накопичення прихованих дефектів. Їх визначення можливе тільки при зміні режимів роботи, коли підключаються нові підмножини елементів, або під час технічного обслуговування РЕЗ з повною перевіркою працездатності в усіх можливих режимах роботи і інструментальної перевірки значень параметрів засобами вимірювальної техніки (ЗВТ). Ця обставина впливає на реальні значення показників надійності РЕЗ, але в відомих методах їх кількісної оцінки не враховано, що визначає актуальність роботи.

Об'єкт дослідження – процес кількісної оцінки значень часткових та комплексних показників надійності сучасних РЕЗ.

Предмет дослідження – удосконалення методу оцінювання значень показників надійності ОЗС з врахуванням можливості появи і накопичення прихованих дефектів.

Мета дослідження – удосконалення математичного апарату і формалізація процесу оцінки значень показників надійності ОЗС з врахуванням можливості появи прихованих дефектів в підмножинах елементів, які в окремих режимах роботи не використовують.

## 1 ПОСТАНОВКА ЗАДАЧИ

Сучасні програмно-керовані РЕЗ відрізняються багатофункціональністю та багаторежимністю, при цьому залежно від режиму роботи використовують різні сукупності елементів. Тобто вони відносяться до ОЗС, значення показників надійності яких суттєве залежать від відносного часу роботи виробу в конкретному режимі. Вочевидь, що ресурс окремих підсистем при цьому використовується не рівномірно. Крім того, відомо, що під час короткочасного і довгострокового зберігання РЕЗ в них накопичуються приховані дефекти, які неможливо виявити й усунути до включення апаратури та перевірки її працездатності. Також це має місце і під час роботи РЕЗ в підмножинах елементів, які при цьому не використовують. Приховані дефекти, які накопичуються, також можливо виявити тільки при зміні режиму роботи виробу. Але ця обставина в відомих методах оцінки надійності БРО, які відносяться до ОЗС, до цього часу не враховується. Тому виникає проблема оцінки значень показників надійності ОЗС з врахуванням властивості появи прихованих дефектів в підсистемах, які не використовують в даний момент часу.

Вирішується проблема кількісної оцінки впливу прихованих дефектів на значення параметра потоку відмов окремих підмножин елементів ОЗС. Необхідно визначити і кількісно оцінити вплив збільшення параметра потоку відмов окремих підмножин елементів на значення часткових (наробіток на відмову і середній час відновлення РЕЗ) і комплексних (коефіцієнти готовності і неготовності) ОЗС [1], які в відомих методах [2] не враховано. Для практичного використання результатів дослідження доцільно формалізувати їх у вигляді алгоритму реалізації.


DOI 10.15588/1607-3274-2024-2-1

Тобто, задача полягає у: 1) розробці структурної схеми реалізації удосконаленого методу; 2) кількісної оцінки впливу прихованих дефектів на часткові та комплексні показники надійності; 3) формалізації процесу оцінки значень показників надійності; 4) перевірці можливості реалізації; 5) дослідження отриманих функціональних залежностей; 6) оцінці ефекту від впровадження результатів дослідження.

## 2 ОГЛЯД ЛІТЕРАТУРИ

В останні роки отримано і досліджено нові моделі і методи оцінки надійності ОЗС, які підвищують точність розрахунків в порівнянні з відомими [3–10], але вони не враховують можливість появи в ОЗС прихованих дефектів.

Аналіз останніх досліджень та публікацій свідчить про розвиток в галузі технічної надійності складних технічних об’єктів і систем. Під час проектування нових зразків РЕЗ кількісно оцінюють часткові та комплексні показники надійності [1, 2]. Але при цьому властивість зміни структури об’єкту під час використання за призначенням не враховують [3–8]. Встановлено, що на значення показників надійності РЕЗ впливають окремі види надлишковості [9, 10]. Крім того, на надійність програмно-керованих РЕЗ суттєво впливає якість їх програмного забезпечення [11, 12]. В [13] показано, що в дійсний час відсутні не тільки інженерні методи, але і теоретична розробка аналізу надійності технічних систем зі змінною структурою. Це завдання успішно вирішено в [14–16], де запропоновано нові модель і метод оцінки значень показників надійності ОЗС. При використанні існуючих методик оцінки значень показників надійності без врахування властивості зміни структури [1–10] отримували занижені значення показників наробітку на відмову і коефіцієнту готовності, що вимагало для забезпечення вимог використовувати більш надійну елементну базу або резервування, тобто додаткову залишковість виробу, яка веде до збільшення його вартості та масогабаритних характеристик. Практика експлуатації ОЗС показала, що значення їх наробітку на відмову в рази перевищує розрахункове. Тому в подальшому запропоновано при використанні аналогічної елементної бази й умов експлуатації результати розрахунків збільшувати на поправочний коефіцієнт, визначений експериментально при експлуатації аналогічних об’єктів [17].

На підмножини елементів, що не використовуються при роботі об’єкту в даному режимі, також впливають кліматичні умови, нагрівання від працюючої частини об’єкту, а також механічні впливи на мобільних засобах. Все це у сукупності веде до виникнення потоку відмов у відключених підмножинах елементів. В залежності від умов експлуатації (стаціонарні або польові) параметр потоку відмов непрацюючої частини об’єкту збільшується на $0{,}01 \le \alpha \le 0{,}1$ від розрахункового значення при його використанні за призначенням [17]. Це також має місце і під час короткочасного зберігання техніки. Приховані дефекти проявляються тільки при зміні режимів роботи або виконання щомісячного технічного обслуговування, коли виконують повну перевірку працездатності виробу в усіх режимах роботи. Так, наприклад, для ЗВТ коефіцієнт прихованих відмов складає $0{,}1 \le \alpha \le 0{,}24$ [18], що суттєве впливає на якість робіт під час технічного обслуговування і поточного ремонту внаслідок помилкової оцінки результату окремих перевірок і технічного стану виробу в цілому. Явні дефекти усуваються в процесі поточного ремонту після відмови об’єкту, а приховані – під час щомісячного технічного обслуговування [15–17].

Проведений аналіз показує актуальність і необхідність врахування можливості виникнення прихованих дефектів під час оцінки значень показників надійності ОЗС.

## 3 МАТЕРІАЛИ ТА МЕТОДИ

Метод призначений для уточнення значень показників надійності об’єктів зі зміною структурою. Сутність удосконалення методу та його відмінність від відомих полягає в урахуванні можливості появи прихованих дефектів, які неможливо виявити під час роботи об’єкту в даному режимі. В відомих роботах ця обставина не враховується, що дає завищену оцінку значень показників надійності [14–16].

Структура методу, його призначення і сутність, вихідні дані, обмеження і припущення, а також результати від використання приведено на рис. 1. Обмеження і припущення відповідають реальним умовам використання за призначенням і відновлення працездатності РЕЗ.

Математичний апарат заснований на використанні коефіцієнту прихованих відмов α, значення якого визначають фіксацією відмов під час технічного обслуговування об’єкту, а також при зміні режиму роботи

$$\alpha_i = Q_i / Q,$$

де $Q_i$ – кількість визначених прихованих відмов в підмножині елементів $L_i$, $Q$ – загальна кількість відмов в об’єкті за час роботи $T_p$.

Зважаючи на то, що елементи виробу розміщені у одному приміщенні або апаратній, тобто працюють в ідентичних умовах, можна вважати

$$\alpha = \frac{1}{R}\sum_{i=1}^{R}\alpha_i ,$$

де $R$ – кількість підмножин елементів об’єкту.


DOI 10.15588/1607-3274-2024-2-1

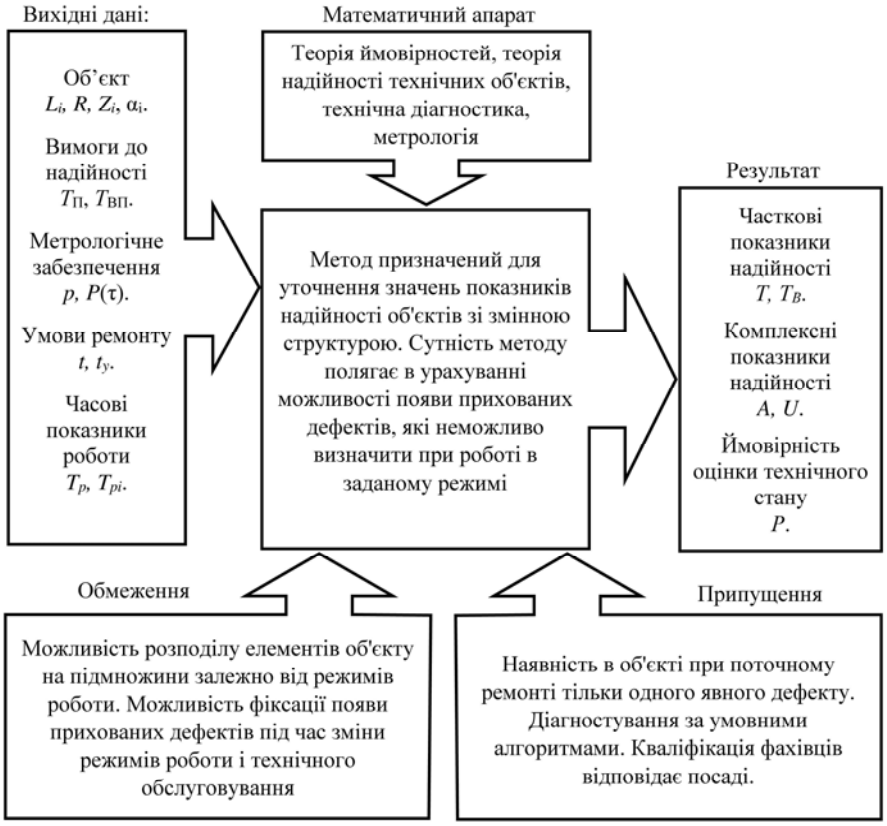



Рисунок 1 – Структурна схема реалізації вдосконаленого методу оцінки значень показників надійності об'єктів зі змінною структурою

Параметр потоку відмов об'єкту дорівнює

$$Z = \sum_{i=1}^{R} u_i Z_i + \sum_{i=1}^{R} \alpha_i (1-u_i) Z_i ,$$

де $u_i$ – відносний час роботи підмножини елементів і $u_i = T_i / T_p$.

В такому разі отримаємо

$$Z = \sum_{i=1}^{R} Z_i (u_i (1-\alpha_i) + \alpha_i) ,$$

наробіток на відмову виробу $T = 1 / Z$.

Ймовірність відмови об'єкту внаслідок появи дефектів серед підмножини елементів $L_i$

$$\frac{Q_i}{Q} = \frac{u_i (u_i Z_i + \alpha_i (1-u_i) Z_i)}{Z} .$$

Розрахунковий час відновлення об'єкту дорівнює при використанні умовних алгоритмів діагностування (УАД) при пошуку дефектів

$$T_{BP} = \frac{t \cdot \sum_{i=1}^{R} Z_i (u_i (1-\alpha) + \alpha_i) \cdot \log_2 L_i}{Z} + t_y ,$$

де $t$ – середній час виконання перевірки, $t_y$ – середній час усунення несправності.

Середня кількість перевірок під час пошуку дефектів за УАД

$$K = \frac{1}{R} \sum_{i=1}^{R} \log_2 L_i ,$$

а ймовірність вірного визначення діагнозу

$$P = p^K,$$

де $p$ – ймовірність вірної оцінки результату виконання перевірки.

З врахуванням метрологічної надійності ЗВТ остаточно отримаємо середній час відновлення об'єкту при поточному ремонті

$$T_B = \frac{T_{BP}}{P \cdot P(\tau)},$$

де $P(\tau)$ – ймовірність безвідмовної роботи ЗВТ в міжперевірочний період $\tau$ [18].


DOI 10.15588/1607-3274-2024-2-1

Отримані функціональні залежності приведено у табл. 1.

Блок-схема алгоритму реалізації методу наведена на рис. 2, де $T_{П}$, $T_{ВП}$ – припустимі значення.

Таблиця 1 – Оцінка значень показників надійності об’єктів зі змінною структурою

| Показник | Функціональні залежності | |
|---|---|---|
| | Відомі [13–15] | Запропоновані |
| Параметр потоку відмов | $Z' = \sum_{i=1}^{R} u_i Z_i$ | $Z = \sum_{i=1}^{R} Z_i (u_i (1-\alpha_i) + \alpha_i)$ |
| Напрацювання на відмову, годин | $T' = 1 / Z'$ | $T = 1 / Z$ |
| Кількість відмов в підмножині $L_i$ за час $T_р$ | $Q' = T_P u_i Z_i$ | $Q_i = T_P \ Z_i (u_i (1-\alpha_i) + \alpha_i)$ |
| Загальна кількість відмов об’єкту за час $T_р$ | $Q' = T_P \sum_{i=1}^{R} u_i Z_i$ | $Q = T_P \sum_{i=1}^{R} Z_i (u_i (1-\alpha_i) + \alpha_i)$ |
| Розрахунковий середній час відновлення, хвилини | $T'_{BP} = t_y + \frac{t}{Z'} \cdot \sum_{i=1}^{R} u_i \cdot Z_i \cdot \log_2 L_i$ | $T_{BP} = t_y + \frac{t}{Z} \cdot \sum_{i=1}^{R} Z_i \cdot (u_i \cdot (1-\alpha) + \alpha_i) \cdot \log_2 L_i$ |

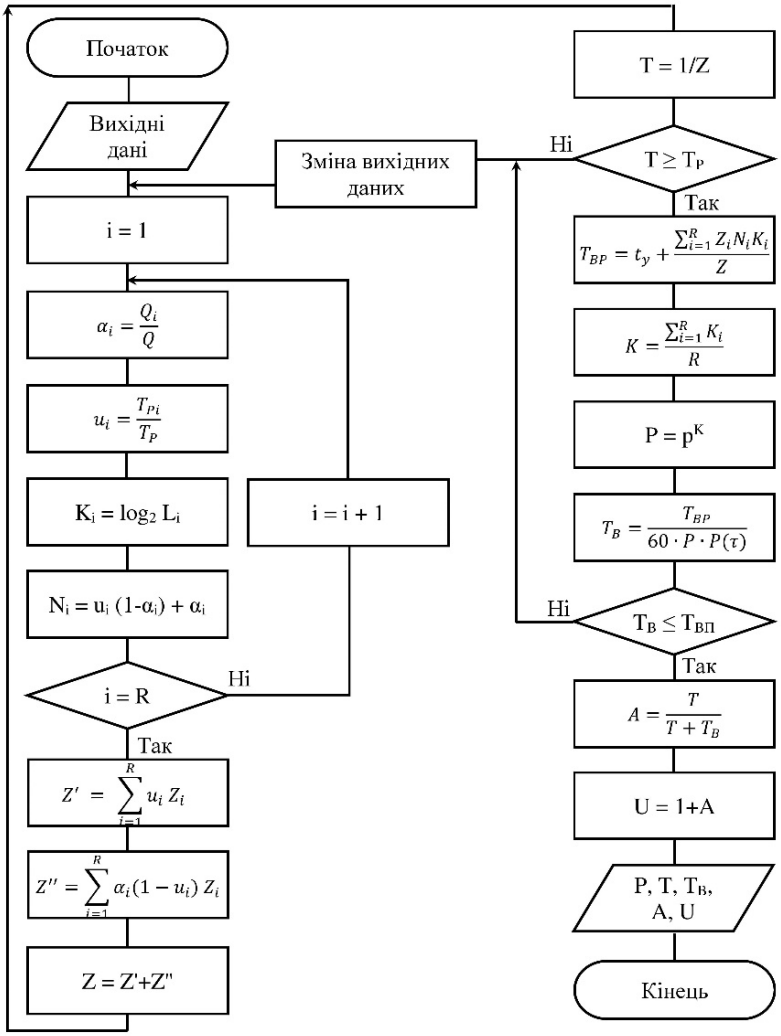



Рисунок 2 – Блок-схема алгоритму розрахунку значень показників надійності об’єктів зі змінною структурою і врахуванням прихованих дефектів


DOI 10.15588/1607-3274-2024-2-1

## 4 ЕКСПЕРИМЕНТИ

Верифікацію запропонованого удосконалення методу виконуємо порівнянням α=0 у правому рядку табл. 1, що веде до відомих результатів [14–16]. Розглянемо використання методу на прикладі радіостанції при наступних вихідних даних [14–15]:
$L_1$=1024; $Z_1$=307 $10^{-6}$ год$^{-1}$; $K_1$=10; $u_1$=1– $u_2$;
$L_2$=3072; $Z_2$=532 $10^{-6}$ год$^{-1}$; $K_2$=11,5; $0 \le u_2 \le 1$;
$L_3$=512; $Z_3$=154 $10^{-6}$ год$^{-1}$; $K_3$=9; $u_3$=1;
$p$=0,997; $P(\tau)$=0,96; $t$=3,5 хв; $t_y$=8 хв.

Взаємодію підмножин елементів радіостанції відображено на рис. 3, де $L_3$ – загальна частина, що працює у всіх режимах (електроживлення, антена, генератор, управління функціонуванням).

Ефект від впровадження запропонованого методу оцінюється згідно табл. 2.

Результати розрахунків залежно від відносного часу роботи радіостанції в режимі «прийом» приведено на рисунках:

– рис. 4 – залежності наробітку на відмову від наявності прихованих дефектів;

– рис. 5 – залежності коефіцієнту неготовності радіостанції від наявності прихованих дефектів;

– рис. 6 – уточнення наробітку на відмову радіостанції від відносного часу роботи в режимі «прийом»;

– рис. 7 – уточнення коефіцієнту неготовності радіостанції від відносного часу роботи в режимі «прийом».

Таблиця 2 – Оцінка ефективності використання запропонованого методу

| Показник | Функціональні залежності | |
|---|---|---|
| | Відомі [13–15] | Запропоновані |
| Середній час відновлення, годин | $T'_B = \frac{T'_{BP}}{60 \cdot P \cdot P(\tau)}$ | $T_B = \frac{T_{BP}}{60 \cdot P \cdot P(\tau)}$ |
| Коефіцієнт готовності об'єкту | $A' = \frac{T'}{T' + T'_B}$ | $A = \frac{T}{T + T_B}$ |
| Коефіцієнт неготовності об'єкту | $U' = 1 - A'$ | $U = 1 - A$ |
| Ефект від уточнення значень показників надійності | $\eta_\tau = \frac{T' - T}{T} 100\%$; $\eta_U = \frac{U' - U}{U} 100\%$ | |

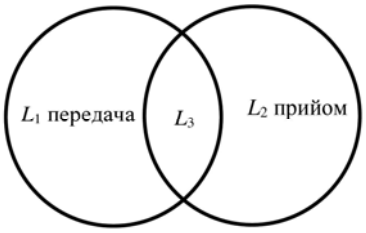



Рисунок 3 – Взаємодія підмножини елементів радіостанції залежне від режиму роботи

При заданих значеннях $T \ge 1000$ годин і $T_B \le 1$ година результати оцінки показників надійності задовольняють вимогам.

## 5 РЕЗУЛЬТАТИ

У роботі отримані і досліджені нові функціональні залежності часткових і комплексних показників надійності ОЗС від кількості прихованих дефектів. Формалізовано порядок використання удосконаленого методу у вигляді алгоритму. Показано приклад використання і встановлено, що врахування можливості появи прихованих дефектів в ОЗС дозволяє уточнити значення показників надійності РЕЗ в цілому. Встановлено, що урахування впливу можливості появи прихованих дефектів в підмножинах елементів об'єкту, які не використовують при його роботі у заданому режимі, уточнює розрахункові значення показників надійності.

Ефект від застосування отриманих результатів полягає в уточненні значень показників надійності у порівнянні відносного збільшення наробітку на відмову при α=0,1 з α=0 (рис. 4), відносного уточнення наробітку на відмову від врахування впливу прихованих дефектів (рис. 6), зменшення відносного уточнення коефіцієнту неготовності ($\eta_U$) виробу від врахування прихованих дефектів згідно табл. 2, а саме:

$$\eta_U = \frac{U' - U}{U} 100\%,$$

де $U'$ – розраховується за умов α=0, $U$ – розраховується за умов α=0,1 (рис. 5).

Отримані результати відносного уточнення коефіцієнту неготовності радіостанції від врахування впливу прихованих дефектів наведені на рис. 7.

## 6 ОБГОВОРЕННЯ

Порівняння результатів використання методу з прототипом [14–16], показує, що врахування можливості появи прихованих дефектів в підмножинах елементів, що не використовуються, радіостанції у розглянутому прикладі дозволяє на 4,5–12% підвищити точність оцінки коефіцієнту неготовності, в чому і полягає ефект від впровадження запропонованого методу.

Аналіз отриманих результатів показує, що оцінка впливу прихованих дефектів зніжує час наробітку на відмову (рис. 4) і підвищує реальні значення коефіцієнту неготовності радіостанції (рис. 5), що в усіх випадках веде до їх відносного уточнення не залежно від режиму роботи (рис. 6, 7).


DOI 10.15588/1607-3274-2024-2-1

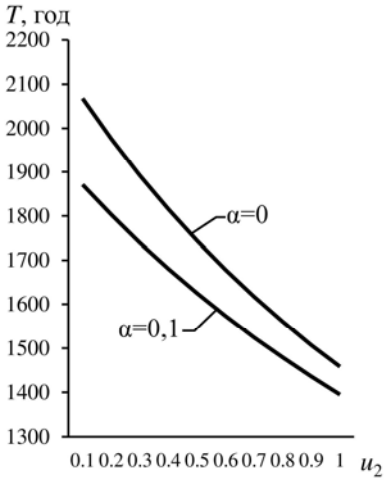



Рисунок 4 – Функціональні залежності $T$ (α, $u_2$)

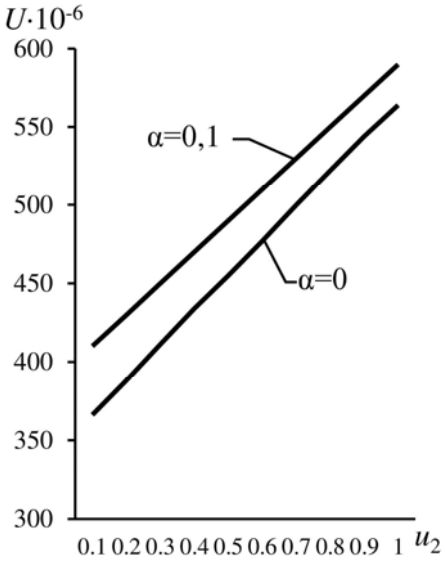



Рисунок 5 – Функціональні залежності $U$ (α, $u_2$)

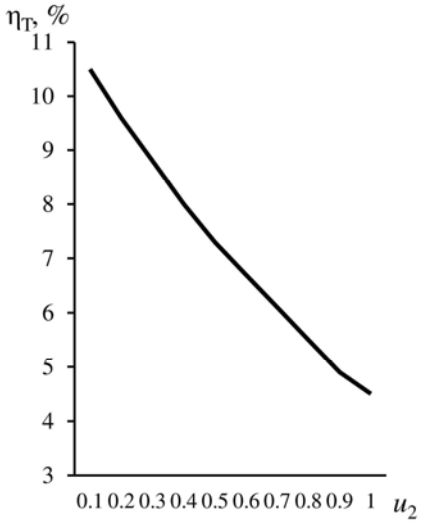



Рисунок 6 – Відносне уточнення наробітку на відмову радіостанції від врахування впливу прихованих дефектів

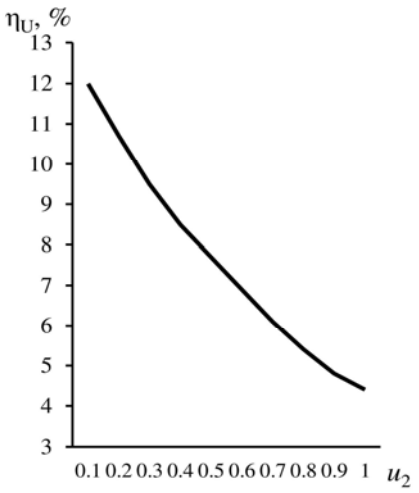



Рисунок 7 – Відносне уточнення коефіцієнту неготовності радіостанції від врахування впливу прихованих дефектів

Застосування запропонованого удосконаленого методу не потребує економічних витрат і додаткової підготовки персоналу та може бути використано під час дослідної експлуатації перспективних зразків РЕЗ різноманітного призначення. Це дозволяє забезпечити потреби вимог щодо надійності виробів з мінімальною вартістю їх елементів.

Переваги запропонованого удосконаленого методу оцінювання надійності ОЗС полягають у наступному:

1) під час дослідної експлуатації перспективних зразків РЕЗ можливо встановити значення коефіцієнту прихованих дефектів, що відсуне у відомих методах;

2) використання цих результатів дозволяє уточнити значення показників надійності БРО в цілому;

3) якщо вони не задовольняють вимогам, то доцільно підвищити якість діагностичного ($K_i$, $K$) і метрологічного ($p$, $P(\tau)$) забезпечення підвищити рівень кваліфікації фахівців ремонтного органу ($t$, $t_y$) і якщо цього не достатньо – замінити елементну базу на більш надійну або удосконалити конструкцію РЕЗ для зниження перегріву непрацюючої частини БРО і підвищення стійкості до механічних перевантажень ($Z_i$, $Z$);

4) використання запропонованих пропозицій виключає серійне виробництво РЕЗ з недостатнім рівнем надійності.

Перспективи подальших досліджень: подальші дослідження доцільно направити на удосконалення метрологічного забезпечення ОЗС, а саме – обґрунтування мінімально необхідних метрологічних характеристик ЗВТ для забезпечення вимог до поточного ремонту під час експлуатації і після короткочасного зберігання РЕЗ з метою зниження вартості ЗВТ. Також доцільно удосконалювати діагностичне забезпечення ОЗС при наявності прихованих дефектів. Цей напрямок не потребує додаткових економічних витрат, а необхідний ефект зниження середнього часу відновлення досягається


DOI 10.15588/1607-3274-2024-2-1

тільки використання УАД спеціальної форми і усіченої процедури пошуку кратних дефектів [15]. Крім того для ОЗС великої розмірності (наприклад апаратних зв'язку) після їх короткочасного зберігання доцільно використовувати груповий пошук дефектів бригадою фахівців [15].

## ВИСНОВКИ

На основі аналізу існуючих методів оцінки значень показників надійності ОЗС встановлено, що в неробочій частині виробу можлива поява прихованих дефектів, але цю обставину відомі джерела не враховують. Отримані і досліджені функціональні залежності впливу прихованих дефектів на значення показників надійності, що дозволило до 10–12% уточнити значення часткових і комплексних показників надійності. Ці результати доцільно використовувати після обробки статистичних даних щодо показників надійності під час дослідної експлуатації перспективних зразків радіоелектронних ОЗС.

Наукова новизна отриманих результатів полягає у розробці наступних інноваційних рішень:

1) вперше запропоновано врахувати наявність прихованих дефектів в непрацюючій частині БРО при оцінці значень показників надійності ОЗС;

2) вперше отримано і досліджено функціональні залежності впливу наявності прихованих дефектів на значення часткових і комплексних показників надійності ОЗС;

3) формалізовано у вигляді алгоритму процес оцінки значень показників надійності ОЗС з врахуванням можливості появи прихованих дефектів.

Практична значимість дослідження полягає в тому, що це дозволяє на етапі дослідної експлуатації РЕЗ зі змінною структурою об'єктивно оцінити відповідність результатів розрахунків необхідним значенням показників надійності завдяки врахуванню появи прихованих дефектів в процесі експлуатації або під час короткочасного зберігання.

## ЛІТЕРАТУРА

1. Надійність техніки. Терміни та визначення: ДСТУ 2860-94. – [Чинний від 1996-01-01]. – К. : Держстандарт України, 1994. – 88с.
2. Надійність техніки. Методи розрахунку показників надійності. Загальні вимоги: ДСТУ 2862-94. – [Чинний від 1997-01-01]. – К. : Держстандарт України, 1995. – 39с.
3. Kuo W. Optimal Reliability Modeling: Principles and Applications / W. Kuo, M. J. Zuo. – New York : John Wiley & Sons, Inc., 2003. – 544 p.
4. Rausand M. System reliability theory: models, statistical methods, and applications / M. Rausand, A. Barros, A. Høyland. – Hoboken : John Wiley & Sons, Inc., 2021. – 813 p.
5. Zio E. Reliability Engineering: Old Problems and New Challenges/ E. Zio // Reliability Engineering and System Safety. – 2009. – Vol. 94. – P. 125–141. https://doi.org/10.1016/j.ress.2008.06.002
6. Uvarov B. M. Radioelectronic Apparatus Design with Optimal Reliability Indicators / B. M. Uvarov, A. V. Nikitchuk // Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia. – 2018. – №75. – P. 48–53. https://doi.org/10.20535/RADAP.2018.75.48-53
7. Kharchenko V. A. Problems of reliability of electronic components / V. A. Kharchenko // Modern Electronic Materials. – 2015. – Vol. 1, Issue 3. – P. 88–92. https://doi.org/10.1016/j.moem.2016.03.002
8. Swingler J. Reliability Characterisation of Electrical and Electronic Systems / J. Swingler. – Cambridge, UK: Woodhead Publishing, 2015. – 274 p. https://doi.org/10.1016/C2013-0-16487-2
9. Reliability Basics of Information Systems / [A. Petrov, V. Khoroshko, L. Scherbak et al.]. – Krakow : Wydawnictwa AGH, 2016. – 246 p.
10. Reliability of Redundant Telecommunications Equipment Advanced Model Considering Failures and Refusals of Structure Elements / [D. Mogylevych, I. Kononova, B. Kredenzer, O. Oksiiuk] // International Conference on Advanced Trends in Information Theory, Kyiv, 18–20 December 2019 : proceedings. – Kyiv: IEEE, 2019. – P. 238–243. https://doi.org/10.1109/ATIT49449.2019.9030502
11. Yamada S. OSS Reliability Measurement and Assessment / S. Yamada, Y. Tamura. – Springer, 2016. – 185 p. https://doi.org/10.1007/978-3-319-31818-9_1
12. Maintenance for Industrial Systems / [R. Manzini, A. Regattieri, H. Pham, E. Ferrari]. – London : Springer, 2010. – 497 p. https://doi.org/10.1007/978-1-84882-575-8.
13. Gurov S. V. Safety Analysis of a Multi-phased Control System / S. V. Gurov, S. P. Habarov, L. V. Utkin // Microelectronics Reliability. – 1997. – Vol. 37, Issue 2. – P. 243–254. https://doi.org/10.1016/S0026-2714(96)00088-1
14. Evaluation of Reliability of Radio-Electronic devices with Variable Structure / [Y. V. Ryzhov, L. N. Sakovich, O. O. Puchkov, Y. E. Nebesna] // Radio Electronics, Computer Science, Control. – 2020. – № 3. – P. 31–41. https://doi.org/10.15588/1607-3274-2020-3-3
15. Технічна експлуатація засобів та систем зв'язку [Електронний ресурс] / [Л. М. Сакович, В. П. Романенко, І. М. Гиренко та ін.]. – Київ : НТУУ КПІ ім. І. Сікорського, 2021. – 176 с. Режим доступу: https://ela.kpi.ua/bitstream/123456789/57827/1/Tekhnichna_ekspluatatsiia_zasobiv_ta_system_zviazku.pdf.
16. Assessing the Reliability of Complex Systems Under Uncertainty in the Context of Ensuring National Resilience / [S. I. Pyrozhkov, O. O. Reznikova, S. Ye. Gnatiuk, Ya. E. Kuryata] // Science and Innovation. – 2023. – № 19 (4). – P. 3–15. https://doi.org/10.15407/scine19.04.003
17. Ayers M. L. Telecommunications System Reliability Engineering, Theory, and Practice / M. L. Ayers. – Hoboken : John Wiley & Sons, 2012. – 256 p. https://doi.org/10.1002/9781118423165.scard
18. Основи експлуатації засобів вимірювальної техніки військового призначення в умовах проведення АТО / [В. Б. Кононов, С. В. Водолажко, С. В. Коваль та ін.]. – Харьков : ХНУПС, 2017. – 288 с.

Стаття надійшла до редакції 13.11.2023.
Після доробки 07.04.2024.

UDC 681.3.06

**IMPROVED METHOD FOR ASSESSING THE RELIABILITY OF OBJECTS WITH A VARIABLE STRUCTURE**

**Babii O. S**. – Senior Lecturer at the Department of Military and Technical Training, Kyiv National Taras Shevchenko University, Kyiv, Ukraine.


DOI 10.15588/1607-3274-2024-2-1

**Sakovych L. M.** – PhD, Associate Professor of the Special Department of the Institute of Special Communication and Information Security, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

**Sliusarchuk O. O.** – PhD, Senior researcher, Senior Researcher of the Military Intelligence Research Institute of the Defence Intelligence of Ukraine, Kyiv, Ukraine.

**Yelisov Y. M.** – PhD, Research Fellow of the Military Intelligence Research Institute of the Defence Intelligence of Ukraine, Kyiv, Ukraine.

**Kuryata Y. E.** – Head of the Scientific and Organizational Department of the Research Center, Institute of Special Communication and Information Security, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

## ABSTRACT

**Context.** The main idea is to take into account the possibility of the influence of hidden defects on the reliability of multi-mode radio-electronic equipment with a variable structure, which do not take into account the known methods for calculating reliability indicators. A quantitative assessment of the parameter of the failure flow of products is proposed, taking into account the impact of the accumulation of hidden defects on its value.

**Objective.** Improvement of the method for assessing the reliability of objects with a variable structure, taking into account the possibility of hidden defects when used for their intended purpose in certain operating modes.

**Method.** The methodology for assessing the values of reliability indicators of complex technical systems is used. The method being developed is the development of an algorithm for assessing the reliability indicators of multi-mode objects in the direction of taking into account the possibility of appearance and accumulation of hidden defects in subsets of elements of the object, which are not used when it operates in separate modes.

**Results.** Functional dependencies of partial and complex indicators of reliability of multi-mode objects on the accumulation of hidden defects, which manifest themselves only during maintenance or change of operating modes, are obtained. The solution is formalized in the form of an algorithm that uses the results of trial operation of products as initial data.

**Conclusions.** The scientific novelty lies in the development of the following innovative solutions: 1) for the first time it is proposed to take into account the presence of hidden defects when assessing the reliability of multi-mode objects with a variable structure; 2) for the first time, functional dependencies of the influence of the presence of hidden defects on the values of partial and complex reliability indicators were obtained and studied. The practical significance of the results lies in the fact that it allows, at the stage of trial operation of radio-electronic equipment with a variable structure, to objectively assess the compliance of calculations with the required values of reliability indicators by taking into account the possibility of hidden defects.

**KEYWORDS:** assessment of reliability indicators, objects with variable structure, hidden defects.

## REFERENCES

1. Nadiinist tekhniky. Terminy ta vyznachennia: DSTU 2860-94. [Chynnyi vid 1996-01-01]. Kyiv, Derzhstandart Ukrainy, 1994, 88 p.
2. Nadiinist tekhniky. Metody rozrakhunku pokaznykiv nadiinosti. Zahalni vymohy: DSTU 2862-94. [Chynnyi vid 1997-01-01]. Kyiv, Derzhstandart Ukrainy, 1995, 39 p.
3. Kuo W., Zuo M. J. Optimal Reliability Modeling: Principles and Applications. New York, John Wiley & Sons, Inc., 2003, 544 p.
4. Rausand M., Barros A., Høyland A. System reliability theory: models, statistical methods, and applications. Hoboken, John Wiley & Sons, Inc., 2021, 813 p.
5. Zio E. Reliability Engineering: Old Problems and New Challenges, *Reliability Engineering and System Safety*, 2009, Vol. 94, pp. 125–141. https://doi.org/10.1016/j.ress.2008.06.002
6. Uvarov B. M., Nikitchuk A. V. Radioelectronic Apparatus Design with Optimal Reliability Indicators, *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*, 2018, №75, pp. 48–53. https://doi.org/10.20535/RADAP.2018.75.48-53
7. Kharchenko V. A. Problems of reliability of electronic components, *Modern Electronic Materials*, 2015, Vol. 1, Issue 3, pp. 88–92. https://doi.org/10.1016/j.moem.2016.03.002
8. Swingler J. Reliability Characterisation of Electrical and Electronic Systems. Cambridge, UK, Woodhead Publishing, 2015, 274 p. https://doi.org/10.1016/C2013-0-16487-2
9. Petrov A., Khoroshko V., Scherbak L., Petrov A., Aleksander M. Reliability Basics of Information Systems. Krakow, Wydawnictwa AGH, 2016, 246 p.
10. D. Mogylevych, I. Kononova, B. Kredenzer, O. Oksiiuk Reliability of Redundant Telecommunications Equipment Advanced Model Considering Failures and Refusals of Structure Elements, *International Conference on Advanced Trends in Information Theory, Kyiv, 18–20 December 2019, proceedings.* Kyiv, IEEE, 2019, pp. 238–243. https://doi.org/10.1109/ATIT49449.2019.9030502
11. Yamada S., Tamura Y. OSS Reliability Measurement and Assessment. Springer, 2016, 185 p. https://doi.org/10.1007/978-3-319-31818-9_1
12. Manzini R., Regattieri A., Pham H., Ferrari E.Maintenance for Industrial Systems. London, Springer, 2010, 497 p. https://doi.org/10.1007/978-1-84882-575-8.
13. Gurov S. V., Habarov S. P., Utkin L. V. Safety Analysis of a Multi-phased Control System, *Microelectronics Reliability*, 1997, Vol. 37, Issue 2, pp. 243–254, https://doi.org/10.1016/S0026-2714(96)00088-1
14. Ryzhov Y. V., Sakovich L. N., Puchkov O. O., Nebesna Y. E. Evaluation of Reliability of Radio-Electronic devices with Variable Structure, *Radio Electronics, Computer Science, Control*, 2020, №3, pp. 31–41. https://doi.org/10.15588/1607-3274-2020-3-3
15. Sakovych L. M., Romanenko V. P., Hyrenko I. M., Kuriata Ya. E., Myroshnychenko Yu. V. Tekhnichna ekspluatatsiia zasobiv ta system zviazku [Elektronnyi resurs]. Kyiv, NTUU KPI im. I. Sikorskoho, 2021, 176 p. Rezhym dostupu https://ela.kpi.ua/bitstream/123456789/57827/1/Tekhnichna_ekspluatatsiia_zasobiv_ta_system_zviazku.pdf.
16. Pyrozhkov S. I., Reznikova O. O., Gnatiuk S. Ye., Kuryata Ya. E. Assessing the Reliability of Complex Systems Under Uncertainty in the Context of Ensuring National Resilience, *Science and Innovation*, 2023, № 19 (4), pp. 3–15. https://doi.org/10.15407/scine19.04.003
17. Ayers M. L. Telecommunications System Reliability Engineering, Theory, and Practice. Hoboken, John Wiley & Sons, 2012, 256 p. https://doi.org/10.1002/9781118423165.scard
18. Kononov V. B., Vodolazhko S. V., Koval S. V., Naumenko A. M., Kondrashova I. I. Osnovy ekspluatatsii zasobiv vymiriuvalnoi tekhniky viiskovoho pryznachennia v umovakh provedennia ATO. Kharkov, KhNUPS, 2017, 288 p.


DOI 10.15588/1607-3274-2024-2-1

UDC 621.372.852

# OPTIMAL SYNTHESIS OF STUB MICROWAVE FILTERS

**Karpukov L. M.** – Dr. Sc., Professor, Professor of the Department of Information Security and Nanoelectronics, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine.

**Voskoboynyk V. O.** – PhD, Associate Professor, Professor of the Department of Information Security and Nanoelectronics, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine.

**Savchenko Iu. V.** – PhD, Associate Professor, Associate Professor of the Department of Cybersecurity and Information Technologies, University of Customs and Finance, Dnipro, Ukraine.

## ABSTRACT

**Context.** Microwave stub filters are widely used in radio engineering and telecommunication systems, as well as in technical information protection systems due to simplicity of design, possibility of realization in microstrip design and manufacturability in mass production. For synthesis of stub filters nowadays traditional methods based on transformation of low-frequency prototype filters on *LC*-elements into filtering structures on elements with distributed parameters are used. The transformations used are approximate and provide satisfactory results for narrowband stub filters. In this connection there is a necessity in development of direct synthesis methods for stub filters, excluding various approximations and providing obtaining of amplitude-frequency characteristics with optimal shape for any bandwidths.

**Objective.** The purpose of the study is to develop a method for direct synthesis of stub band-pass filters and low-pass filters with Chebyshev amplitude-frequency response in the passband.

**Method.** The procedure of direct synthesis includes the formulation of relations for filter functions of plume structures, selection of approximating functions of Chebyshev type for filter functions and formation of a system of nonlinear equations for calculation of parameters of filter elements.

**Results.** A method for the direct synthesis of stub bandpass and lowpass filters with Chebyshev response is developed.

**Conclusions.** Scientific novelty of the work consists in the development of a new method of direct synthesis of l stub filters. The method, in contrast to approximate traditional methods of synthesis of microwave filters, is exact, and the obtained solutions of synthesis problems are optimal.

The experiments confirmed the performance of the proposed method and the optimality of the obtained solutions. Prospects for further research suggest adapting the method to the synthesis of filter structures with more complex resonators compared to stubs.

**KEYWORDS:** synthesis, band-pass filter, low-pass filter, plume, scattering parameters, filter function, approximation.

## ABBREVIATIONS

BPF is a band pass filter;
LPF is a low-pass filter;
FF is a filter function;
PB is a passband
SB is a stopband;
AFR is an amplitude-frequency response.

## NOMENCLATURE

$\Theta$ is an electric length, rad;
$\rho_s$ is a stub impedance, Ω;
$\rho_c$ is a wave impedance of transmission lines, Ω;
$\rho_0$ is a port impedance, Ω;
$R_i$ is a normalized wave impedance;
$S_{ik}$ is a scattering parameter of the filter structure element;
$\hat{S}_{ik}$ is a filter scattering parameter;
$\alpha$ is a attenuation, dB;
$f_0$ is a center frequency, GHz;
$f_c$ is a ripple cutoff frequency, GHz;
$T$ is a transmission matrix.

## INTRODUCTION

Microwave filters are the most important component of modern microwave systems of information processing, transmission, reception and protection. Among the designs of microwave filters a special place is occupied by stub filter structures, due to the simplicity of their structure and the possibility of its realization in planar microstrip design.

Synthesis of microstrip stub filters is usually carried out by the traditional method using low-frequency *LC*-prototype filters [1–6]. To transform the prototype filter into a filtering structure composed of elements with distributed parameters, approximate procedures are used, which do not provide optimal frequency characteristics of the filters.

**The object of study.** The object of research are procedures and methods of synthesis of filtering structures on elements with distributed parameters.

**The subject of study.** The subject of the study is methods of synthesizing stub filters with Chebyshev characteristic.

**The purpose of the work.** Based on the initial data for the synthesis, including the values of the center frequency $f_0$, relative bandwidth $2\Delta f/f_0$, ripple level $\alpha_c$ in the passband and attenuation level $\alpha_s$ in the fence band, it is required to determine the values of the wave impedances $\rho_i$ of the filter elements, at which the ideal equal-wave AFR with a given level of ripple is ensured and the technological requirements for the value of the wave impedances are fulfilled.

## 1 PROBLEM STATEMENT

Realization of the method of direct synthesis of stub filters determines the necessity of development of a number of computational procedures, providing the


DOI 10.15588/1607-3274-2024-2-2

compilation of the transfer function of the filter in a form convenient for synthesis, selection of the approximating function for the amplitude-frequency response and calculation of the parameters of the filter elements on the basis of the condition of physical realizability.

## 2 REVIEWS OF THE LITERATURE

The traditional method of synthesizing microwave stub filters is based on the transformation of a low-pass prototype filter composed of *LC*-elements into a structure consisting of elements with distributed parameters [1–6]. The prototype filter has a ladder structure, characterized by the cutoff frequency $f_c$=1 Hz and *g*-parameters representing the values of inductances and capacitances normalized by the resistance of the loads.

The transition from the low-frequency prototype filter to the filter, the scheme and values of *LC*-elements of which provide the cutoff frequency and bandwidth required by the synthesis task, is carried out according to the known formulas of frequency conversion [1–5].

Fulfillment of the requirement for physical realizability of the filter in the form of a planar structure is carried out by modifying the prototype filter circuit by introducing into it inverters that convert branches included in series into branches included in parallel.

To transform the prototype filter circuit into a filtering structure on elements with distributed parameters, the Richards frequency transformation [1–5] is used, according to which the inductance *L* is replaced by short-circuited and capacitance *C* by open-circuited sections of the transmission line on the basis of the relations $j\omega L=j\Omega\mathrm{tg}(\theta)$, $j\omega C=j\Omega\mathrm{tg}(\theta)$, где $\Omega=\mathrm{tg}(\theta)$ a frequency variable introduced by Richards, $\theta=\omega l/v$– electrical length of the transmission line segment, *l* – section length, ν – wave propagation velocity in the line. In turn, the concentrated *LC*-resonators are replaced by distributed resonators by equating their goodness-of-fit or admeasurement steepness [1–5]. Quarter-wave sections of transmission lines are used as inverters [1–5]. Equivalent transformations in the transmission line of serial branches into parallel ones are realized by means of Kuroda identities [1–6].

It should be noted that the frequency characteristics of resonators on short-circuited and open-circuited transmission line segments approximate with sufficient accuracy the characteristics of concentrated *LC*-resonators only near resonance, while quarter-wave line segments are characterized by perfect inversion only at the center frequency corresponding to the $\theta=\pi/2$. Thus, the conventional synthesis procedure is approximated by matching the performance of the prototype low-pass *LC*-filter and the corresponding stub filter in a narrow frequency band.

In [7–11], digital filters were proposed to be used as prototype filters for the synthesis of microwave filters. Type of transfer functions of digital filters $K_d(z)$, $z=\exp(j\omega\tau)$, composed by means of the discrete Laplace transform is identical to the representation of transfer functions of circuits composed of transmission line segments of the same length. In [9, 10] the method of calculation of circuit parameters on commensurate segments of the transmission line by the function $K_d(z)$ digital prototype filter [12]. The method includes selecting a suitable circuit configuration for realizing the function $K_d(z)$ and compilation of an autoregressive model for estimating the values of the coefficients of the transfer function of the circuit by the values of the coefficients of the function $K_d(z)$ of the prototype filter. The estimation is performed by the method of least squares. This synthesis method is an approximation. It should also be noted that the circuit configuration chosen for synthesis may not be optimal and may contain redundant elements in its structure.

Thus, there is a need to develop methods for direct synthesis of stub filters.

For stub filter structures with Chebyshev characteristic the method of direct synthesis is proposed in [14, 15]. The method includes the main stages of classical synthesis of electric filters [13]: selection of approximating function for a given filter structure; compilation of Hurwitz polynomial for physically realizable transfer function; determination of type and parameters of filter elements by transfer function. The stage of filter realization in this method is rather complicated and time-consuming. Therefore, there is a need to improve the efficiency of this method by simplifying the computational procedure at the stage of filter realization.

## 3 MATERIALS AND METHODS

Fig. 1 shows the investigated filter structures composed of combinations of quarter-wave stubs with wave impedance $\rho_s$ and quarter-wave sections of transmission lines with wave impedance $\rho_c$. The input and output ports of the filters have wave impedance $\rho_0$. BPFs use short-circuited stubs, LPFs use open-circuited stubs. Filter structures have symmetry.

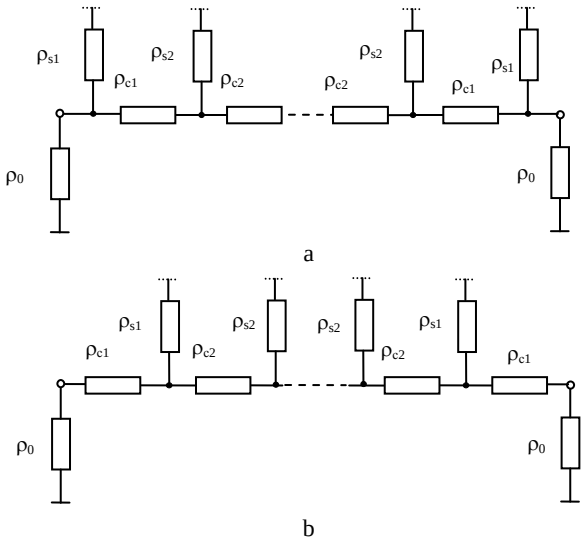



Figure 1 – Stub filter structures


DOI 10.15588/1607-3274-2024-2-2

It is rational to analyze the n-cascade filter structure using the modified transfer matrix (T-matrix) by the formulas [14]:

$$\left.\begin{aligned}\hat{T} &= \begin{bmatrix} \hat{B} & -\hat{A}_{11} \\ \hat{A}_{11} & -\hat{A} \end{bmatrix} = \prod_{i=1}^{n} \begin{bmatrix} B_i & -A_{11_i} \\ A_{11_i} & -A_i \end{bmatrix}, \\ \hat{A}_{21} &= \prod_{i=1}^{n} A_{21_i}, \end{aligned}\right\} \quad (1)$$

where $\hat{A} = \dfrac{\hat{A}_{11}^2 - \hat{A}_{21}^2}{\hat{B}}$, $A_i = \dfrac{A_{11_i}^2 - A_{21_i}^2}{B_i}$.

On the basis of (1) the scattering parameters are determined $\hat{S}_{ik} = \hat{A}_{ik} / \hat{B}$ filter and its components $S_{ik_n} = A_{ik_n} / B_n$, $i,k$=1,2.

Let us represent the transfer matrices of the filter elements as follows:

– for the transmission line segment

$$\left.\begin{aligned} T_c(\theta) &= \cos(\theta) E + j \frac{\sin(\theta)}{2R_c} \begin{bmatrix} R_c^2 + 1 & -\left(R_c^2 - 1\right) \\ R_c^2 - 1 & -\left(R_c^2 + 1\right) \end{bmatrix}, \\ A_{21_c} &= 1; \end{aligned}\right\} \quad (2)$$

– for open-circuited stubs

$$\left.\begin{aligned} T_s(\theta) &= \cos(\theta) E + j \frac{\sin(\theta)}{2R_s} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \\ A_{21_s} &= \cos(\theta); \end{aligned}\right\} \quad (3)$$

– for short-circuited stubs

$$\left.\begin{aligned} T_s(\theta) &= \sin(\theta) E - j \frac{\cos(\theta)}{2R_s} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \\ A_{21_s} &= \sin(\theta). \end{aligned}\right\} \quad (4)$$

Here $j = \sqrt{-1}$ – imaginary unit, θ – electrical length of transmission line segments and stubs, $R = \rho / \rho_0$ – normalized wave impedance of lines, $E$ – unit matrix.

Selective properties of the filter are characterized by its transfer function

$$\left|\hat{S}_{21}(\theta)\right|^2 = \frac{1}{1 + \left|F(\theta)\right|^2} \quad (5)$$

Here $F(\theta)$ is FF. For symmetric structures it is imaginary and is defined by the relation:

$$F(\theta) = j \hat{A}_{11}(\theta) / \hat{A}_{21}(\theta). \quad (6)$$

Analysis by formulas (1), (2), (4) of the filter structures in Fig. 1, consisting of $n_s$ closed stubs and $n_c$ of transmission lines gives the following relation for BPF:

$$\left.\begin{aligned} \hat{A}_{11}(\theta) &= j \sin(\theta)^{n_s - 1} \gamma \sum_{k=0}^{m} a_k \cos(2k\theta), \\ \hat{A}_{21}(\theta) &= \sin(\theta)^{n_s}, \end{aligned}\right\} \quad (7)$$

where $m=n_c/2$, $\gamma=\cos(\theta)$ at $n_c$ even; $m=(n_c+1)/2$, $\gamma=1$ at $n_c$ odd.

Accordingly, FF for BPF will take the following form:

$$F(\theta) = \frac{j \gamma \sum_{k=0}^{m} a_k \cos(2k\theta)}{\sin(\theta)}. \quad (8)$$

It should be noted that the function (8) does not depend on the type of circuits of the structures in Fig. 1, but is determined by the number of transmission lines in these structures.

For LPFs whose structures are composed of $n_s$ of open stubs and $n_c$ of transmission lines the result of analysis by (1), (2), (3) is the formula

$$\left.\begin{aligned} \hat{A}_{11}(\theta) &= j \sin(\theta) \sum_{k=0}^{m} a_k \cos(2k\theta), \\ \hat{A}_{21}(\theta) &= \cos(\theta)^{n_s}, \end{aligned}\right\} \quad (9)$$

where $m=(n_s+n_c-1)/2$.

According to (9) FF for LPF will take the following form:

$$F(\theta) = \frac{j \sin(\theta) \sum_{k=0}^{m} a_k \cos(\theta)^{2k}}{\cos(\theta)^{n_S}}. \quad (10)$$

One of the main tasks of the synthesis is to compose an approximating function on FF of the investigated structure, providing the given requirements on AFR of the filter. For BPF with FF (8) the approximating function can be a Chebyshev function of the following form

$$T(\theta) = \operatorname{ch}\left[ n_s \operatorname{arch}\left( \frac{\operatorname{ctg}(\theta)}{\operatorname{ctg}(\theta_c)} \right) + n_c \operatorname{arch}\left( \frac{\cos(\theta)}{\cos(\theta_c)} \right) \right], \quad (11)$$

where $\theta_c$ – angle at the filter bandwidth boundary in terms of ripple level.

Let us represent this function as follows:

$$T(\theta) = \frac{A^{(+)}(\theta) + A^{(-)}(\theta)}{2 \sin(\theta)}, \quad (12)$$


DOI 10.15588/1607-3274-2024-2-2

where $A^{(\pm)}(\theta)=\left(\frac{S}{C}\cos(\theta)\pm Q(\theta)\right)\left(\frac{\cos(\theta)}{C}\pm Q(\theta)\right)^{n_c}$,

$Q(\theta)=\sqrt{\cos(\theta)^2/C^2-1}$, $S=\sin(\theta_c)$, $C=\cos(\theta_c)$.

Ratio (12) can be written in a form similar to FF (8)

$$T(\theta)=\frac{\beta\sum_{k=0}^{m}\alpha_k\cos(2k\theta)}{\sin(\theta)}, \qquad (13)$$

where $m=n_c/2$, $\beta=\cos(\theta)/C^{n_c+1}$ at $n_c$ even; $m=(n_c+1)/2$, $\beta=0,5/C^{n_c+1}$ at $n_c$ odd.

Table 1 presents a series of relations for the coefficients in (13) for a range of values of $n_c$.

For LPF with FF (10) the Chebyshev approximating function can be written in the form

$$T(\theta)=\mathrm{ch}\left[n_s\,\mathrm{arch}\left(\frac{\mathrm{tg}(\theta)}{\mathrm{tg}(\theta_c)}\right)+n_c\,\mathrm{arch}\left(\frac{\sin(\theta)}{\sin(\theta_c)}\right)\right]. \qquad (14)$$

Let us represent this function as follows

$$T(\theta)=\frac{A^{(+)}(\theta)+A^{(-)}(\theta)}{2\cos(\theta)^{n_s}}, \qquad (15)$$

where $A^{(\pm)}(\theta)=\left(\frac{C}{S}\sin(\theta)\pm Q(\theta)\right)^{n_s}\left(\frac{\sin(\theta)}{S}\pm Q(\theta)\right)^{n_c}$,

$Q(\theta)=\sqrt{\sin(\theta)^2/S^2-1}$.

Transforming (15), we obtain

$$T(\theta)=\frac{\sin(\theta)\sum_{k=0}^{m}\alpha_k\cos(2k\theta)}{\cos(\theta)^{n_s}}, \qquad (16)$$

where $m=(n_s+n_c-1)/2$.

In Table 2, where $\eta=2S^{n_c+n_s}$, for a range of values $n_c$, $n_s$ are presented expressions for the coefficients of the rows in (16).

The process of filter synthesis begins with determining the number of elements of the selected circuit on the basis of the following initial data: $f_0$ – center frequency corresponding to the electric length $\theta_0=\pi/2$; $\alpha_c$ – attenuation in terms of ripple level in the passband, *dB*; $f_c$ – bandwidth limit frequency; $\alpha_s$ – attenuation at frequency $f_s$ in the stopband, dB.

The number of elements is determined by the formulas derived from the approximating functions (13), (16):

– for BPF

$$n_c=\frac{\mathrm{arch}\left(\sqrt{\frac{10^{\frac{\alpha_s}{10}}-1}{10^{\frac{\alpha_c}{10}}-1}}\right)-\mathrm{arch}\left(\frac{\mathrm{ctg}(\theta_s)}{\mathrm{ctg}(\theta_c)}\right)}{\mathrm{arch}\left(\frac{\cos(\theta_s)}{\cos(\theta_c)}\right)}, \qquad (17)$$

Table 1 – Coefficients of the Chebyshev function of the stub BPF

| $n_c$ | $n_s$ | $\alpha_0$ | $\alpha_2$ | $\alpha_4$ | $\alpha_6$ |
|---|---|---|---|---|---|
| 1 | 2 | $-1+S+2S^2$ | $1+S$ | | |
| 2 | 1 | $-1+2S^2+S^3$ | $1+S$ | | |
| | 3 | | | | |
| 3 | 2 | $S^2+3S^3+2S^4$ | $-1+S+5S^2+3s^3$ | $1+S$ | |
| | 4 | | | | |
| 4 | 3 | $1-2S^2+2S^3+4S^4+S^5$ | $-2+6S^2+4S^3$ | $1+S$ | |
| | 5 | | | | |
| 5 | 4 | $S^2+5S^3+7S^4+5S^5+2S^6$ | $2S^2+10S^3+13S^4+5S^5$ | $-1+S+7S^2+5S^3$ | $1+S$ |
| | 6 | | | | |
| 6 | 5 | $1+4S^2+3S^3+S^4+6S^5+6S^6+S^7$ | $2-6S^2+6S^3+19S^4+9S^5$ | $-2+8S^2+6S^3$ | $1+S$ |
| | 7 | | | | |

Table 2 – Coefficients of the Chebyshev function of the stub LPF

| $n_c$ | $n_s$ | $\alpha_0$ | $\alpha_2$ | $\alpha_4$ | $\alpha_6$ |
|---|---|---|---|---|---|
| 1 | 2 | $-(1+C)\,(1+C-4C^2)/\eta$ | $-(1+C)^2/\eta$ | | |
| 2 | 1 | $(-2+4C^2+2C^3)/\eta$ | $-2\,(1+C)/\eta$ | | |
| 2 | 3 | $(1+C)2(3-3C-8C2+14C3)/2\eta$ | $(1+C)2\,(2-C-4C2-C3)/\eta$ | $(1+C)3/2\eta$ | |
| 3 | 2 | $(1+C)^2(2-4C+C^2+4C^3)/\eta$ | $(1+c)^2\,(3-4C-3C^2)/\eta$ | $(1+C)^2/\eta$ | |
| 3 | 4 | $(1+C)^3(-7+17C+C^2-39C^3+48C^4)/4\eta$ | $-(1+C)^4(11-32C+20C^2+16C^3)/4\eta$ | $(1+C)^4(-5+8C+3C^2)/4\eta$ | $-(1+C)^4/4\eta$ |
| 4 | 3 | $(-2+7C^2-5C^4+13C^5+20C^6+7C^7)/\eta$ | $-(1+C)^3(7-18C+6C^2+18C^3+2C^4)/2\eta$ | $(1+C)^3(-2+3C+2C^2)/\eta$ | $-(1+C)^3/2\eta$ |


DOI 10.15588/1607-3274-2024-2-2

– for LPF

$$n_c = \frac{\operatorname{arch}\left(\sqrt{\frac{10^{\frac{\alpha_s}{10}}-1}{10^{\frac{\alpha_c}{10}}-1}}\right) - \delta \operatorname{arch}\left(\frac{\operatorname{tg}(\theta_s)}{\operatorname{tg}(\theta_c)}\right)}{\operatorname{arch}\left(\frac{\operatorname{tg}(\theta_s)}{\operatorname{tg}(\theta_c)}\right) + \operatorname{arch}\left(\frac{\sin(\theta_s)}{\sin(\theta_c)}\right)}, \qquad (18)$$

where $\theta_c = \pi f_c / 2 f_0$, $\theta_s = \pi f_s / 2 f_0$, $\delta = \pm 1$.

According to the values calculated by (17), (18) $n_c$ and $n_s$=$n_c$+δ the coefficients of polynomials $\alpha_k$ in (13), (16) are calculated.

Wave impedances of the filter circuit elements are determined by solving the system of nonlinear equations

$$a_k(R) - \varepsilon \alpha_k = 0, k = 1,...m, \qquad (19)$$

where $a_k(R)$ – coefficients of the polynomials in (8), (10); $R$ – vector formed from the sought wave impedances, $\varepsilon = \sqrt{10^{\alpha_c/10} - 1}$ – ripple value in the filter passband.

The system (19) is solved by Newton's method. At each iteration the coefficients $a_k(R)$ are calculated by decomposition $\hat{A}_{11}(\theta) = T_{21}(\theta)$ from (1) into Fourier series by cos(2$k$θ). To solve the system (19) can also be used methods of solving optimization problems with constraints in case of the need to take into account technological tolerances on the value of line impedances.

## 4 EXPERIMENTS

For approbation and substantiation of reliability and efficiency of the proposed method, the results of synthesis of broadband BPF and LPF are presented and the obtained results are compared with the data of calculations by methods based on prototype filters.

Stub BPF synthesis is performed for the following data: $f_0$=2 GHz; $\alpha_c$=0.1 dB (ε=0.1526); $\alpha_s$=40 dB; $f_s$=3,5 GHz and for two variants of calculation: $f_c$=1.5 GHz at relative bandwidth 2Δ$f/f_0$=0.5 (50%) and $f_c$=1.0 GHz, at 2$\Delta f/f_0$=1.0 (100%).

For these data, the type and number of elements in the circuits are determined by (17): for the circuit in Fig. 1a, four transmission lines are obtained $n_c$=4 and five stubs' $n_s$=5; for the scheme in Fig. 1b – $n_c$=4, $n_s$=3. For $f_c$=1,5 GHz for the specified filter structures the coefficients of the rows in (13) took the following values: α=543.100; 764.645; 234.412.

For the circuit in Fig. 1a the required vector of normalized resistances is: $R$=$R_{s1}$, $R_{c1}$, $R_{s2}$, $R_{c2}$, $R_{s3}$. From the solution of the system (19) we can obtain a number of normalized values of $R$, depending on the initial approximation $R_0$. For example, for the given values of the coefficients of the polynomial, the following is obtained:

– $R$=0.4889; 1.0160; 0.5417; 1.3839, 0.7353 to an initial approximation $R_{0k}$=0.5;

– $R$=0.4745; 1.0847; 0.6855; 1.3575, 0.5407 for $R_{0k}$=1.0;

– $R$=0.4471; 1.2613; 1.0229; 1.8692, 0.7695 for $R_{0k}$=1.5, $k$=1,...,5.

For the circuit in Fig. 1, b the required resistances are: $R$=$R_{c1}$, $R_{s1}$, $R_{c2}$, $R_{s2}$. At the initial approximation $R_{0k}$=0.5 these resistances obtained the following values: $R$=0.3301; 0.0378; 0.2412; 0.6866.

By the traditional synthesis method using the formulas [1, 4] compiled when converting the LC prototype filter to a stub filter, the calculation for the circuit in Fig. 1a with the $n_c$=4, $n_s$=5 gave the following values for the resistances: $R$=0.5674; 0.7732; 0.2883; 0.7175; 0.2931

Fig. 2 shows the AFR BPF in the form of dependencies $S_{21}(f)_{dB} = -20\,lg\left|\hat{S}_{21}(f)\right|$. In this figure and in the following figures, curve 1 corresponds to the calculation by the developed method, curve 2 – to the traditional methods. AFRs in the filter bandwidth are given in the center of the figures.

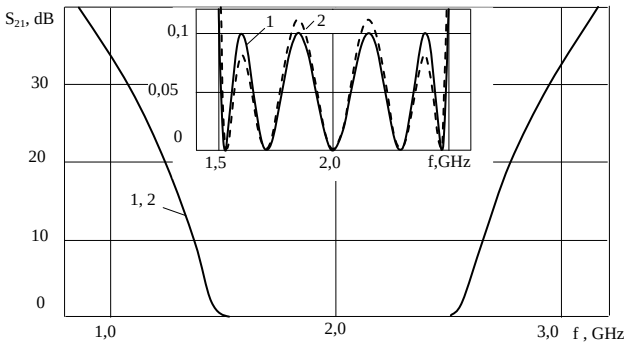



Figure 2 – AFR BPF with relative bandwidths 50%

In Fig. 3 shows the AFRs for $fc$=1.0 GHz at relative bandwidth 2Δ$f/f0$=1.0. For these data the coefficients of the rows in (13) took the following values: α=10.6568; 13.6568; 9.6568, For these values from the solution (19) at initial approximation $R_{0k}$=0.5 received:

– $R$=2.4384; 0.7909; 1.4946; 0.6894; 1.0649 for the scheme in Fig. 1a;

– $R$=0.5972; 0.6199; 0.3776; 0.500 for the scheme in Fig. 1b.

Calculation by the traditional method according to formulas from [1,4] for the scheme in Fig. 1a gives $R$=2.2977; 0.7732; 1.1815; 0.7175; 1.2161.


DOI 10.15588/1607-3274-2024-2-2

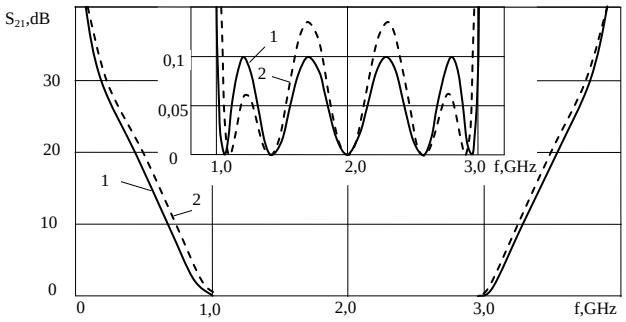



Figure 3 – AFR BPF with relative bandwidths 100%

Fig. 4 shows the results of BPF synthesis with relative bandwidth $2\Delta f/f0=0.4$ and attenuation $\alpha_c=0.5$ dB in the passband. In this figure, curve 2 is plotted based on the transfer function of the prototype digital filter when the stub filter structure formed by of eight stubs and seven transmission lines transmission lines was selected for synthesis [9]. In the simulation, this structure was represented by seven sections. Each section consisted of two stubs connected by a section of transmission line. As a result of the synthesis, the resistances of the elements of the sections, represented as follows ($\rho_{s1}$, $\rho_{s2}$, $\rho_{c1}$), took the following values at port resistance 50 Ω: (27.1; 98.6; 49.5), (98.9; 96.3; 89.5), (47.5; 53.7; 92.0), (56.3; 77.1; 64.8), (75.8; 94.0; 31.7), (85.9; 59.9; 27.3) [9].

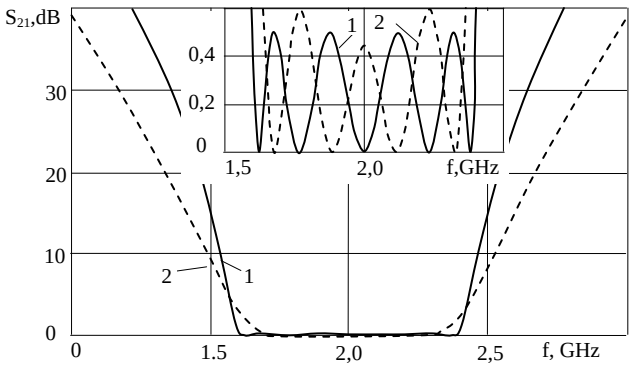



Figure 4 – AFR BPF with relative bandwidths 40%

AFR LPF with relative bandwidth $2\Delta f/f_0=1.0$ c are shown in Fig. 5.

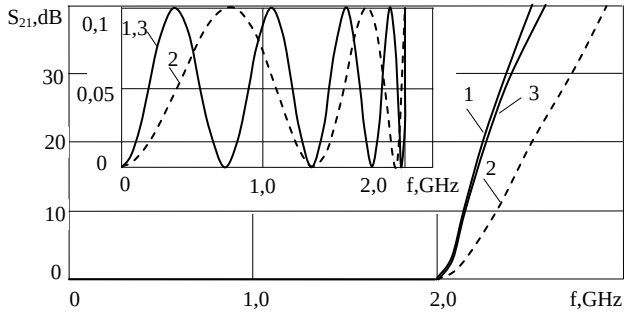



Figure 5 – AFR LPF with relative bandwidths 100%

Curve 1 in Fig. 4 corresponds to the calculation developed by the direct synthesis method for a structure of five stubs and four transmission lines with normalized impedances $R$=0.229; 0.89; 0.334; 1.05; 0.249.

When synthesizing LPF with $f_0$=2 GHz; $\alpha_c$=0.1 dB, coefficients of the rows in (16) at $2\Delta f/f_0=1.0$ take the following values:

– α=118.1684; –203.4698; 138.4215 –62.9662; 20.5030. for the scheme in Fig. 1a at $n_c$=4, $n_s$=5;

– α=142.9481; –246.6830; 166.6154; –75.8302; 24.0208 for the scheme on 1b at $n_c$=5, $n_s$=4.

The corresponding values of resistances obtained from (19):

– $R$=0.8417; 1.8135; 0.4828; 1.9622; 0.4622 for the scheme in Fig. 1a;

– $R$=1.7378; 0.8095; 2.5761; 0.6376; 2.6621 for the scheme on 1b.

By the traditional method using Richards transform, inverters and Kuroda identities it is obtained for scheme 1a: $R$=2.872; 1.5342; 0.6241; 1.837; 0.5063 [1, 6].

Curves 1,3 in the figure are obtained by the developed synthesis method, curve 1 corresponds to the scheme in Fig. 1a, curve 3 corresponds to the scheme in Fig. 1b.

## 5 RESULTS

As follows from Figs. 2–4, curves 1 obtained by the developed synthesis method have an ideal equal-wave shape and fully satisfy the synthesis assignment.

AFR BPF of schemes in Fig. 1a and 1b, at equal number of connecting transmission lines are identical, their frequency characteristics at $n_c$=4 are presented in Fig. 2–3 by one curve 1.

The shape of curves 2 in Figs. 2–3, obtained by the conventional synthesis method according to the low-pass *LC*-prototype filter using ideal inverters, does not correspond to the equal-wave characteristic, and the attenuation in the passband exceeds the specified value. The mismatch grows with the increase of the passband, at the same time its narrowing occurs.

AFR BPF, presented in Fig. 4, curve 2, is calculated for the stub filter structure of 15 elements by the transfer function of the prototype digital filter. This characteristic does not fully satisfy the conditions of the synthesis task. These conditions, curve 1, were met when applying the developed synthesis method to the filter structure with a much smaller number of elements, equal to 9 elements.

The AFR LPFs in Fig. 4, calculated for the schemes in Figs. 1a and 1b by the developed method (curves 1 and 3), have an equal-wave shape in the passband, matching the graphical accuracy. However, outside the passband, the scheme in Fig. 1a has a higher steepness of the AFR decay in the fence band.

Curve 2 in Fig. 4, corresponding to the traditional synthesis method using Kuroda identities, also has an equal-wave AFR in the passband, but since only stub resonators were involved in the formation of the AFR, this characteristic has a smaller number of ripples in the passband and a much smaller slope steepness in the obstruction bandwidth.


DOI 10.15588/1607-3274-2024-2-2

## 6 DISCUSSIONS

The presented calculation results confirm the validity and efficiency of the developed direct method of loop filter synthesis. Unlike approximate traditional synthesis methods, the results of synthesis by the developed method fully correspond to the technical specification, providing a strictly equal-wave frequency response with a given level of ripple in the passband and a given value of attenuation in the barrier band for filters with narrow and wide passbands. It should also be noted that in the developed method all elements of filtering structures are involved in the formation of the AFR of filters, so at the stage of realization of filters their microstrip structures will be more compact in comparison with the structures realized according to the results of traditional synthesis procedures.

From the presented results of synthesis by the developed method, it follows that the BPF based on the schemes in Fig. 1a and 1b have matching AFR at the same number $n_c$ of transmission lines. The advantage of the scheme in Fig. 1b is the smaller number of elements. However, the values of wave impedances of elements in this scheme may not meet the conditions of technological realizability, in particular, for narrowband filters. For microstrip structures at wave impedance of ports $\rho_0$=50 Ω the permissible values of wave impedances of lines lie in the range of 15–150 Ω. These conditions are largely satisfied by the values of the wave impedances of the elements obtained as a result of synthesis for the circuit in Fig. 1, a.

Frequency response of the LPF based on the schemes in Fig. 1a and 1b with the same number of elements have an equal-wave shape in the passband, which coincides with the graphical accuracy. As for the fence band, the AFR of the scheme in Fig. 1a has a steeper decline. This scheme also corresponds more to the technological conditions on the realized in the process of synthesis value of wave impedances of its elements. Thus, when choosing a scheme for the BPF or LPF, the scheme in Fig. 1a should be preferred.

The proposed method is simple in program implementation, since the basis of the program algorithm is formed by known, well-developed computational procedures from the mathematical support of the Fourier transform and the solution of systems of nonlinear equations [16]. If necessary, in the program algorithm at the stage of parametric optimization, it is easy to add restrictions on the values of wave impedances of the elements of the synthesized filter circuits when composing the target function.

## CONCLUSIONS

The actual problem on improvement of design methods of microwave stub filters is solved.

**The scientific novelty.** Scientific novelty of the obtained results consists in the fact that for the first time a method of direct synthesis is proposed, which allows obtaining optimal amplitude-frequency characteristics of stub filters directly from the filtering function by means of simple computational procedures of decomposition of functions into Fourier series and solving systems of nonlinear equations.

**The practical significance.** The practical value of the obtained results consists in the fact that the filter structures synthesized by the developed method are optimal both by the number of elements contained in them and by compliance with the conditions of the technical task for synthesis. The program realization of the developed method does not cause difficulties, and its use in computational experiments requires insignificant computational and time resources. Application of the developed method and program in computer-aided design systems will allow to improve the quality of design solutions in the synthesis of stub filters.

**Prospects forfurther research.** Prospects for further research consist in extending the proposed approach to the solution of problems of synthesizing a wide range of diverse filtering structures composed of commensurate sections of transmission lines.

## ACKNOWLEDGEMENTS

The work was carried out in accordance with the theme of research work “Construction of a high-speed information and telecommunication system using alternative types of routing”, State registration number 0123U105266, in the framework of joint research work of the Department of Information Security and Nanoelectronics of the National University Zaporizhzhia Polytechnic, Zaporizhian, and the Department of Cybersecurity and Information Technology of the University of Customs and Finance, Dnipro, on the basis of the agreement on creative cooperation between the departments.

## REFERENCES

1. Hong J.-S. Microstrip filters for RF microwave applications. New York, John Wiley & Sons, 2nd ed, 2011, 635 p. DOI: 10.1002/9780470937297.
2. Pramanick P., Bhartia P. Modern RF and Microwave Filter Design. Boston, London, Artech House, 2016, 435 p.
3. Cameron R. J., Kudsia C. M., Mansour R. R. Microwave filters for communication systems: fundamentals, design, and applications. New York, Wiley & Sons, Inc., 2007, 771 p. DOI: 10.1002/9781119292371.
4. Mattaei G., Young L., Jones E. M. T. Microwave Filters, Impedance-Matching Networks and Coupling Structures. Boston, Artech House, 1985, 1096 p.
5. Pozar D. M. Microwave Engineering. New York, Wiley& Sons, 4nd ed, 2012, 756 p.
6. Kuroda K. General properties and synthesis of transmission-line net-works, *Microwave Filters and Circuits, Eds.: A. Matsumoto.* New York, Academic, 1970, Vol. 22, pp. 13–60.
7. Pan T.-W., Hsue C.-W. Modified transmission and reflection coefficients of nonuniform transmission lines and their applications, *IEEE Trans. Microwave Theory Tech.*, 1998, Vol. 46, № 12, pp. 2092–2097. DOI: 10.1109/22.739287.
8. Tascone R., Savi P., Trinchero D., Orta R. Scattering matrix approach for the design of microwave filters, *IEEE Trans.*


DOI 10.15588/1607-3274-2024-2-2

*Microwave Theory Tech.*, 2000, Vol. 48, № 3, pp. 423–430. DOI: 10.1109/22.826842.
9. Chang D.-C., Hsue C.-W. Design and Implementation of Filters Using Transfer Functions in the Z Domain, *IEEE Trans. Microwave Theory Tech.*, 2001, Vol. 49, № 5, pp. 979–985. DOI: 10.1109/22.920157.
10. Tsai L.-C., Hsue C.-W. Dual-Band Band-Pass Filters Using Equal-length Coupled-Serial-Shunted Lines and Z-Transform Technique, *IEEE Trans. Microwave Theory Tech.*, 2004, Vol. 52, № 4, pp. 1111–1117. DOI: 10.1109/tmtt.2004.825680.
11. Toya S., Huang T., Mohan A. S., Yakabe T. Application of particle swarm optimization for the design of UWB bandpass filters, *2007 IEEE Antennas and Propagation Society International Symposium, 9–15 June 2007: proceedings.* Honolulu, HI, USA, 2007, pp. 1617–1620. DOI: 10.1109/APS.2007.4395820.
12. Oppenheim A. V., Schafer R. W. Discrete-Time Signal Processing. Englewood Cliffs, NJ, Prentice-Hall, 1989, 893 p.
13. Temes G. C., LaPatra J. W. Introduction to Circuit Synthesis and Design. McGraw Hill Book Company, 1977, 596 p.
14. Karpukov L. M., Korolkov R. Yu. Direct synthesis of low-pass stub filters with the Chebyshev characteristic, *Radio Electronics, Computer Science, Control*, 2014, №1, pp. 35–39.
15. Karpukov L. M., Voskoboynik V. O., Savchenko I. V., Kozina G. L. Method of increasing information security in communication channels using microwave filters with attenuation poles at predetermined frequencies, *Systems and Technologies*, 2022, №2 (64), pp. 64–74. DOI: 10.32782/2521-6643-2022.2-64.9.
16. Kiusalaas J. Numerical Methods in Engineering with MATLAB. New York, Cambridge University Press, 3nd ed, 2015, 446 p. DOI: 10.1017/CBO97 81316341599.

Received 01.02.2024.
Accepted 25.04.2024.

УДК 621.372.852

## ОПТИМАЛЬНИЙ СИНТЕЗ ШЛЕЙФНИХ МІКРОХВИЛЬОВИХ ФІЛЬТРІВ

**Карпуков Л. М.** – д-р техн. наук, професор, професор кафедри інформаційної безпеки та наноелектроніки Національного університету «Запорізька політехніка», м. Запоріжжя, Україна.
**Воскобойник В. А.** – канд. техн. наук, доцент, професор кафедри інформаційної безпеки та наноелектроніки Національного університету «Запорізька політехніка», м. Запоріжжя, Україна.
**Савченко Ю. В.** – канд. техн. наук, доцент, доцент кафедри кібербезпеки та інформаційних технологій Університету митної справи та фінансів, м. Дніпро, Україна.

### АНОТАЦІЯ

**Актуальність.** Мікрохвильові шлейфні фільтри знаходять широке застосування в радіотехнічних і телекомунікаційних системах, а також у системах технічного захисту інформації завдяки простоті конструкції, можливості реалізації в мікросмужковому виконанні та технологічності при масовому виготовленні. Для синтезу шлейфових фільтрів наразі використовують традиційні методи, засновані на перетворенні низькочастотних фільтрів-прототипів на LC-елементах на фільтрувальні структури на елементах із розподіленими параметрами. Використовувані перетворення є наближеними і забезпечують задовільні результати для вузькосмугових шлейфових фільтрів. У зв'язку з цим виникає необхідність у розробленні для шлейфних фільтрів методів прямого синтезу, що виключає різного роду наближення і забезпечує отримання амплітудно-частотних характеристик з оптимальною формою для будь-яких смуг пропускання.

**Мета.** Метою дослідження є розробка методу прямого синтезу шлейфних смугово-пропускних фільтрів і фільтрів нижніх частот із чебишевською амплітудно-частотною характеристикою в смузі пропускання.

**Метод.** Процедура прямого синтезу містить у собі складання співвідношень для функцій фільтрацій шлейфових структур, підбір апроксимувальних функцій чебишевського типу для функцій фільтрацій і формування системи нелінійних рівнянь для обчислення параметрів елементів фільтрів.

**Результат.** Розроблено метод прямого синтезу шлейфних смугово-пропускних і фільтрів нижніх із чебишевською характеристикою.

**Висновки.** Наукова новизна роботи полягає в розробленні нового методу прямого синтезу шлейфних фільтрів. Метод, на відміну від наближених традиційних методів синтезу мікрохвильових фільтрів, є точним, а одержувані розв'язки задач синтезу – оптимальними.

Проведені експерименти підтвердили працездатність пропонованого методу й оптимальність одержуваних рішень. Перспективи подальших досліджень передбачають адаптацію методу на синтез фільтрувальних структур зі складнішими порівняно зі шлейфами резонаторами.

**КЛЮЧОВІ СЛОВА:** синтез, смугово-пропускний фільтр, фільтр нижніх частот, шлейф, параметри розсіювання, функція фільтрації, апроксимація.

### ЛІТЕРАТУРА

1. Hong J.-S. Microstrip filters for RF microwave applications / J.-S. Hong. – New York : John Wiley & Sons, 2nd ed, 2011. – 635 p. DOI: 10.1002/9780470937297.
2. Pramanick P. Modern RF and Microwave Filter Design / P. Pramanick, P. Bhartia. – Boston, London : Artech House, 2016. – 435 p.
3. Cameron R. J. Microwave filters for communication systems: fundamentals, design, and applications / R. J. Cameron, C. M. Kudsia, R. R. Mansour. – New York : Wiley & Sons, Inc., 2007. – 771 p. DOI: 10.1002/9781119292371.


DOI 10.15588/1607-3274-2024-2-2

4. Mattaei G. Microwave Filters, Impedance-Matching Networks and Coupling Structures / G. Mattaei, L. Young, E. M. T. Jones. – Boston : Artech House, 1985. – 1096 p.
5. Pozar D. M. Microwave Engineering / D. M. Pozar. – New York : Wiley& Sons, 4nd ed, 2012. – 756 p.
6. Kuroda K. General properties and synthesis of transmission-line net-works / K. Kuroda // Microwave Filters and Circuits / Eds.: A. Matsumoto. – New York : Academic, 1970. – Vol. 22. – P. 13–60.
7. Pan T.-W. Modified transmission and reflection coefficients of nonuniform transmission lines and their applications / N.-W. Pan, C.-W. Hsue // IEEE Trans. Microwave Theory Tech. – 1998. – Vol. 46, № 12. – P. 2092–2097. DOI: 10.1109/22.739287.
8. Scattering matrix approach for the design of microwave filters / [R. Tascone, P. Savi , D. Trinchero, R. Orta] // IEEE Trans. Microwave Theory Tech. – 2000. – Vol. 48, № 3. – P. 423–430. DOI: 10.1109/22.826842.
9. Chang D.-C. Design and Implementation of Filters Using Transfer Functions in the Z Domain / D.-C. Chang, C.-W. Hsue // IEEE Trans. Microwave Theory Tech. – 2001. – Vol. 49, № 5. – P. 979–985. DOI: 10.1109/22.920157.
10. Tsai L.-C. Dual-Band Band-Pass Filters Using Equal-length Coupled-Serial-Shunted Lines and Z-Transform Technique / L. Tsai, C.-W. Hsue // IEEE Trans. Microwave Theory Tech. – 2004. – Vol. 52, № 4. – P. 1111–1117. DOI: 10.1109/tmtt.2004.825680.
11. Application of particle swarm optimization for the design of UWB bandpass filters / [S. Toya, T. Huang, A. S. Mohan, T. Yakabe] // 2007 IEEE Antennas and Propagation Society International Symposium, , 9–15 June 2007: proceedings. – Honolulu, HI, USA, 2007. – P. 1617–1620. DOI: 10.1109/APS.2007.4395820.
12. Oppenheim A. V. Discrete-Time Signal Processing / A. V. Oppenheim, R. W. Schafer. – Englewood Cliffs, NJ : Prentice-Hall, 1989. – 893 p.
13. Temes G. C. Introduction to Circuit Synthesis and Design / G. C. Temes, J. W. LaPatra. – McGraw Hill Book Company, 1977. – 596 p.
14. Карпуков Л. М. Прямий синтез шлейфних фільтрів нижніх частот із чебишевською характеристикою / Л. М. Карпуков, Р. Ю. Корольков // Радіоелектроніка, інформатика, управління. – 2014. – № 1. – С. 35–39.
15. Метод підвищення захищеності інформації в каналах зв'язку застосуванням мікрохвильових фільтрів із полюсами загасання на заданих частотах / [Л. М. Карпуков В. О. Воскобойник, Ю. В. Савченко, Г. Л. Козіна] // Системи та технології. – 2022. – № 2 (64). – С. 64–74. DOI: 10.32782/2521-6643-2022.2-64.9.
16. Kiusalaas J. Numerical Methods in Engineering with MATLAB / J. Kiusalaas. – New York : Cambridge University Press, 3nd ed, 2015. – 446 p. DOI: 10.1017/CBO97 81316341599.


DOI 10.15588/1607-3274-2024-2-2

UDC 621.396.962

# MATHEMATICAL MODEL OF CURRENT TIME OF SIGNAL FROM SERIAL COMBINATION LINEAR-FREQUENCY AND QUADRATICALLY MODULATED FRAGMENTS

**Kostyria O. O.** – Dr. Sc., Senior Research, Leading Research Scientist, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

**Hryzo A. A.** – PhD, Associate Professor, Head of the Research Laboratory, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

**Khudov H. V.** – Dr. Sc., Professor, Head of Department, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

**Dodukh O. M.** – PhD, Leading Research Scientist, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

**Solomonenko Y. S.** – PhD, Deputy Head of the Faculty of Educational and Scientific Work, Ivan Kozhedub Kharkiv National Air Force University, Kharkiv, Ukraine.

## ABSTRACT

**Context.** One of the methods of solving the actual scientific and technical problem of reducing the maximum level of side lobes of autocorrelation functions of radar signals is the use of nonlinear-frequency modulated signals. This rounds the signal spectrum, which is equivalent to the weight (window) processing of the signal in the time do-main and can be used in conjunction with it.

A number of studies of signals with non-linear frequency modulation, which include linearly-frequency modulated fragments, indicate that distortions of their frequency-phase structure occur at the junction of the fragments. These distortions, depending on the type of mathematical model of the signal – the current or shifted time, cause in the generated signal, respectively, a jump in the instantaneous frequency and the instantaneous phase or only the phase. The paper shows that jumps occur at the moments when the value of the derivative of the instantaneous phase changes at the end of the linearly-frequency modulated fragment. The instantaneous signal frequency, which is the first derivative of the instantaneous phase, has an interpretation of the rotation speed of the signal vector on the complex plane. The second derivative of the instantaneous phase of the signal is understood as the frequency modulation rate.

Distortion of these components leads to the appearance of an additional component in the linear term of the instantaneous phase, starting with the second fragment. Disregarding these frequency-phase (or only phase) distortions causes distortion of the spectrum of the resulting signal and, as a rule, leads to an increase in the maxi-mum level of the side lobes of its autocorrelation function. The features of using fragments with frequency modulation laws in complex signals, which have different numbers of derivatives of the instantaneous phase of the signal, were not considered in the known works, therefore this article is devoted to this issue.

**Objective.** The aim of the work is to develop a mathematical model of the current time of two-fragment nonlinear-frequency modulated signals with a sequential combination of linear-frequency and quadratically modulated fragments, which provides rounding of the signal spectrum in the region of high frequencies and reducing the maximum level of side lobes of the autocorrelation function and increasing the speed of its descent.

**Method.** Nonlinear-frequency modulated signals consisting of linearly-frequency and quadratically modulated fragments were studied in the work. Using differential analysis, the degree of influence of the highest derivative of the instantaneous phase on the frequency-phase structure of the signal was determined. Its changes were evaluated using time and spectral correlation analysis methods. The parameters of the resulting signal evaluated are phase and frequency jumps at the junction of fragments, the shape of the spectrum, the maximum level of the side lobes of the autocorrelation function and the speed of their descent.

**Results.** The article has further developed the theory of synthesis of nonlinear-frequency modulated signals. The theoretical contribution is to determine a new mechanism for the manifestation of frequency-phase distortion at the junction of fragments and its mathematical description. It was found that when switching from a linearly-frequency modulated fragment to a quadratically modulated frequency-phase distortion of the resulting signal, the third derivative of the instantaneous phase becomes, which, by analogy with the theory of motion of physical bodies, is an acceleration of frequency modulation. The presence of this derivative leads to the appearance of new components in the expression of the instantaneous frequency and phase of the signal. The compensation of these distortions provides a decrease in the maximum level of the side lobes by 5 dB and an increase in its descent rate by 8 dB/deck for the considered version of the non-linear-frequency modulated signal.

**Conclusions.** A new mathematical model of the current time has been developed for calculating the values of the instantaneous phase of a nonlinear-frequency modulated signal, the first fragment of which has linear, and the second – quadratic frequency modulation. The difference between this model and the known ones is the introduction of new components that provide compensation for frequency-phase distortions at the junction of fragments and in a fragment with quadratic frequency modulation. The obtained oscillogram, spectrum and autocorrelation function of one of the synthesized two-fragment signals correspond to the theoretical form, which indicates the adequacy and reliability of the proposed mathematical model.

**KEYWORDS:** mathematical model; linear and quadratic frequency modulation; maximum level of side lobes.

## ABBREVIATIONS

ACF is an autocorrelation function;

SL is a side lobe;

MOP is a modulation on pulse or internal pulse frequency modulation;

WP is a weight processing;


DOI 10.15588/1607-3274-2024-2-3

ML is a main lobe;
RCS is a radar cross section;
QFM is a quadratic frequency modulation;
RSFM is a root-square frequency modulation;
LFM is a linear frequency modulation;
MM is a mathematical model;
MPSLL is a maximum peak side lobe level;
NLFM is a non-linear frequency modulation;
FMA is a frequency modulation acceleration;
PSLL is a peak side lobe level;
FM is a frequency modulation;
FMR is a frequency modulation rate.

## NOMENCLATURE

$C_1$ is a integration constant;
$F(\cdot)$ is a functional dependency;
$f_0$ is a initial signal frequency, Hz;
$f_n(t)$ is a instantaneous frequency *n*-th fragment NLFM, Hz;
*n*=1, 2 is a serial number signal fragment LFM;
*t* is a current time, s;
$T_n$ is a duration of the *n*-th signal fragment NLFM, s;
$\alpha_2$ is a FMA *n*-th signal fragment NLFM, Hz/s$^2$;
$\beta_n$ is a FMR *n*-th signal fragment NLFM, Hz/s;
$\Delta f_n$ is a frequency deviation of the *n*-th signal fragment NLFM, Hz;
$\delta f_{12}$ is a frequency jump when moving from 1-th signal fragment NLFM to 2-th, Hz;
$\delta\varphi_{12}$ is a phase jump when moving from 1-th до 2-th signal fragment NLFM, rad;
$\varphi(t)$ is a instantaneous phase of synthesized NLFM, rad;
$\varphi_n(t)$ is a instantaneous phase *n*-th signal fragment NLFM, rad.

## INTRODUCTION

Modern methods and means of digital synthesis of radar signals open wide prospects for the formation of probing signals with improved spectral characteristics and correlation properties [1–7].

The combination of signals with different MOP laws and the synthesis of new types of signals of this type was restrained by the insufficient depth of the theoretical analysis of processes occurring at the moments of change of the FM law.

In works [8–10] on the example of NLFM signals with LFM fragments it is shown that frequency and phase jumps occur at their joints or only phases depending on the time representation of MM – in the current or shifted time. Approximation of the FM law to the S-shape provides a smoother rounding of the signal spectrum, which helps to reduce MPSLL. To achieve this, the authors [11–16] combined LFM with non-linear MOP in the lower and upper frequency regions of the NLFM signal. This method reduces MPSLL only for a limited number of parameter sets.

During the studies, it was found that frequency and phase jumps at the junction of LFM and QFM fragments were not previously considered and did not have a formalized description. The presence of such frequency-phase distortions leads to distortions in the temporal and spectral structure of the signal.

The development of an MM signal of the LFM-QFM type for the current time, which allows to take into account a new source of frequency-phase jumps and determine the consequences of these jumps, is devoted to this article. Such an MM two-fragment NLFM signal is selected to provide better visibility of the studies and their illustrative nature.

**The object of study** is the process of synthesis and optimal processing of two-fragment NLFM signals.

**The subject of study** is the NLFM MM signal, the first fragment of which has LFM, and the second – QFM.

The known sampling methods [2–23] are highly iterative and low speed, as well as characterized by the uncertainty of quality criteria of formed subsample.

**The purpose of the work** is to develop MM of the current time of two-phase NLFM signals with frequency-phase distortion compensation for the case when the first fragment has LFM, and the second – QFM. This rounds the signal spectrum in the high frequency region, thereby reducing MPSLL and increasing its decay rate.

## 1 PROBLEM STATEMENT

For further presentation of the material, we justify the feasibility of using the concepts of FMR and FMA. According to the general theory of radio oscillations, the full foray of the signal phase is interpreted as the area under the signal curve, and therefore the first derivative of the instantaneous phase is interpreted as the path that passes the end of the phase vector and this derivative is called the signal frequency. In the theory of motion of physical bodies, the derivative of the path is the speed, and in the presence of frequency modulation for the derivative of the frequency, the use of the concept of FMR instead of “scan speed”, which is reasonably used in the English literature, is justified. Since the derivative of the velocity of the body is acceleration, then in the work we will use the concept of FMA.

As the initial, we will use the known MM of the current time of the NLFM signal consisting of two LFM fragments with compensation of frequency and phase jumps at their junction for the increasing law of frequency change [10]:

$$\varphi(t)=\begin{cases}2\pi\left[f_0 t+\dfrac{\beta_1 t^2}{2}\right], 0\le t\le T_1;\\ 2\pi\left[(f_0-\delta f_{21})t+\dfrac{\beta_2 t^2}{2}\right]-\delta\varphi_{12}, T_1\le t\le T_{12};\end{cases} \quad (1)$$

where $\beta_1=\dfrac{\Delta f_1}{T_1}$; $\beta_2=\dfrac{\Delta f_2}{T_2}$.


DOI 10.15588/1607-3274-2024-2-3

Input parameters of MM are $f_0$, $T_n$, $\Delta f_n$. Based on its output parameter $\varphi(t)$ the values of the instantaneous amplitude of the NLFM signal, its spectrum and ACF are calculated from the known relations [1, 2].

## 2 REVIEW OF THE LITERATURE

The analysis of the studies indicates the widespread use of NLFM signals in various fields of technology [1, 2, 8–19], in particular, radar, sonar, flaw detection, lidar systems, communications, etc. NLFM signals in comparison with LFM signals reduce MPSLL while providing the required range, range measurement accuracy and Doppler frequency shift [11–23].

Many researchers note that there is a potential to further improve the achievable characteristics of NLFM signals due to the structural-parametric optimization of their frequency-time parameters, which can be carried out using the selection method, gradient methods and methods of evolutionary optimization [24–35].

So, in [36], the use of the NLFM signal in a combined system is considered, which involves the simultaneous use of a dedicated frequency range for sensing airspace and communication. The NLFM signal is described as a pulse whose phase change occurs according to a second-order polynomial law. Signal parameters are optimized by selecting phase polynomial coefficients in order to minimize MPSLL. It is noted that the NLFM signal allows you to get better performance by measuring the range and speed of the target compared to the LFM signal and LFM signal with additional weight processing.

In [37], the use of an NLFM signal with a quadratic frequency function is considered in relation to the problem of ultrasonic diagnosis of defects. The results of the experiment showed that the selected NLFM signal provides an optimal ratio of the level of the side lobes and the width of the main lobe of the ACF.

In [38] discusses a hyperbolic frequency modulation signal that is insensitive to Doppler frequency shift. The final parameters of the NLFM signal are obtained using the method of iterative optimization, which is based on the selection of parameters, namely the central frequency and the spectrum width of their sequence.

Various approximations of FM functions (or instantaneous phase changes) are used to conveniently record NLFM MM signals. In [39] discloses a method for describing an NLFM signal based on the use of a stepwise approximation of a phase curve of a hyperbolic frequency modulated signal. To reduce MPSLL simultaneously with the selection of approximation parameters, window processing is used.

A similar approach is used in the study [40] in relation to magnetic-acoustic-electric tomography. A piecewise linear FM function is used, the parameters of which are optimized based on a genetic algorithm.

Another method for reducing MPSLL is based on the synthesis of signals with the desired spectral characteristics, for example, in [41] it is proposed to apply a modified Chebyshev window function to adjust the spectral power density of the signal, which ensures the reduction of MPSLL without high computational complexity.

Reduction of MPSLL can be achieved by synthesizing a new type of signal, as described in [10]. The authors proposed a two-fragment NLFM signal, the first fragment of which has QFM.

## 3 MATERIALS AND METHODS

We will carry out the development of the MM two-fragment NLFM signal, which consists of LFM and QFM fragments, that is, for the first fragment, the signal frequency changes along the linear. $f_1(t) = F(t)$, for the second – according to the quadratic law $f_2(t) = F(t^2)$.

If the instantaneous frequency of the second portion of the NLFM signal changes in a quadratic law, then the FMR changes linearly in time. For the instantaneous phase, we apply the concept of FMA, which in this case is a constant value:

$$\alpha_2 = \frac{d^3\varphi_2(t)}{dt^3}. \tag{2}$$

Accordingly, the instantaneous frequency of the second fragment:

$$\begin{aligned} f_2(t) &= f_0 + \beta_1 T_1 + \iint_t \alpha_2 dt^2 = \\ &= f_0 + \beta_1 T_1 + \alpha_2\left(\frac{t^2}{2} - T_1 t + \frac{T_1^2}{2}\right). \end{aligned} \tag{3}$$

The resulting expression (3) contains the integration constant, which we find by the relation:

$$C_1 = f_0 + \beta_1 T_1 + \alpha_2 \frac{T_1^2}{2}.$$

Determination of the integration constant is mandatory, since its third component is actually equal to the frequency jump at the junction of the NLFM signal fragments.

In the next step, by integrating (3), we find the instantaneous phase of the second fragment.

$$\varphi_2(t) = 2\pi\left[(f_0 + \beta_1 T_1)t + \alpha_2\left(\frac{t^3}{6} - \frac{T_1}{2}t^2 + \frac{T_1^2}{2}t - \frac{T_1^3}{6}\right)\right]. \tag{4}$$

FMA $\alpha_2$ is determined using MM input data (1). Based on the fact that:

$$\Delta f_2 = \alpha_2 \iint_{T_2} dt,$$

– FMA $\alpha_2$ define as:


DOI 10.15588/1607-3274-2024-2-3

$$\alpha_2 = \frac{2\Delta f_2}{T_2^2}. \quad (5)$$

The value (5) is calculated for all expressions.

The considered MM (1) does not take into account the appearance of FMA $\alpha_2$ at the junction of fragments and its influence on the change in the frequency-time characteristics of the second fragment.

To simplify mathematical expressions, we apply symbols similar to those used in (1):

$$\delta f_{12} = \alpha_2 \frac{t^2}{2}\bigg|_{t=T_1} = \alpha_2 \frac{T_1^2}{2}. \quad (6)$$

Frequency jump (6) causes instantaneous phase jump at fragment junction:

$$\delta\varphi_{12} = 2\pi\alpha_2 \frac{T_1^3}{6}. \quad (7)$$

Like (1) and using (6), (7) we obtain the MM of the current time for the instantaneous phase of the NLFM signal as part of the first LFM and second QFM fragments:

$$\varphi(t) = \begin{cases} 2\pi\left[f_0 t + \frac{\beta_1 t^2}{2}\right], 0 \le t \le T_1; \\ 2\pi\left[\left(f_0 + \beta_1 T_1 - \delta f_{12}\right)t + \frac{\alpha_2 t^2 (t - 3T_1)}{6}\right] - \delta\varphi_{12}, \\ T_1 \le t \le T_{12}. \end{cases} \quad (8)$$

The synthesized MM (8) provides finding the instantaneous phase of the NLFM signal in the current time for the case when the first fragment has LFM and the second QFM. This MM implements an increasing law of frequency change, compensates for phase and frequency jumps at the junction of fragments due to the presence of FMA.

## 4 EXPERIMENTS

The operability test of the proposed MM two-segment NLFM signals with the first LFM and the second QFM fragments was carried out in the MATLAB software environment.

For experimental studies, the frequency-time parameters of signals that provide observation of their fine structure, namely, instantaneous phase jumps and frequency in oscillograms, were chosen; the nature and number of pulsations of the top of the spectrum, the presence of pulsations on its slopes; MPSLL, the nature of the change in frequency and shape of the SL, the shape and width of the ML ACF.

Fine structure analysis was performed based on the following patterns. The number of ripples of the top of the spectrum and the ripples of the ACF are proportional to the product of the signal deviation by its duration and their values allow us to assess the correspondence of the obtained graphic material to the given parameters. Spectrum breaks, a sharp change in the frequency of pulsations SL ACF and oscillation on the oscillogram indicate the presence of jumps in the instantaneous frequency of the signal. Violation of the oscillogram smoothness, dips in the spectrum, pulsations of its side slopes, changes in the level of SL ACF are a sign of the presence of instantaneous phase jumps of radio frequency oscillations [8, 9].

Based on the results of the fine structure analysis, it is concluded that the obtained results correspond to the specified input parameters of the MM, the adequacy of its work and the reliability of the obtained results.

At the first stage of experimental studies, one-fragment (LFM, QFM) and two-fragment (LFM-LFM, LFM-QFM) signals are compared with equivalent initial frequency, frequency deviation and duration.

Comparative analysis was carried out based on the results of the obtained MRLSS values, its decay rate and width ML ACF of the generated signal at the level of 0.707 of its maximum value.

To determine the advantages of the proposed signal, a group of five LFM-QFM signals was investigated. Since the frequency-time parameters of this group of signals differ, the width of the ML ACF was not measured.

## 5 RESULTS

The results of modelling single-fragment LFM, QFM and two-fragment LFM-LFM, LFM-QFM signals with equivalent frequency-time parameters are summarized in Table 1.

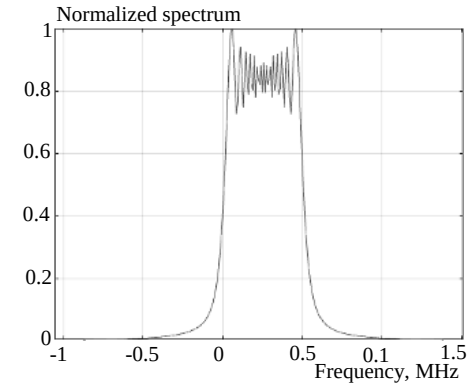
Table 1 – Results of experimental studies of single-fragment and two-fragment signals

| Kind of sign. | $T_1$. μs | $T_2$. μs | $\Delta f_1$. kHz | $\Delta f_2$. kHz | MPSL. dB | Rapid decline. SLL. dB/dec | ML ACF. μs |
|---|---|---|---|---|---|---|---|
| LFM | 120 | – | 500 | – | –13.47 | 19.75 | 1.77 |
| QFM | 120 | – | 500 | – | –9.08 | 16.70 | 1.73 |
| LFM-LFM | 20 | 100 | 150 | 350 | –18.15 | 16.0 | 1.98 |
| LMF-QFM | 60 | 60 | 100 | 400 | –19.16 | 25.0 | 2.17 |

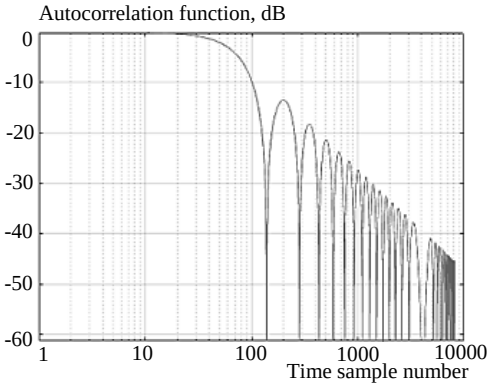
Analysis of the results of Table 1 indicates that the single-fragment QFM signal has no advantages over the LFM signal. But its useful property is the rounding of the spectrum in the high frequency region.

The above shows Fig. 1 and Fig. 2, which show the spectrum of LFM (Fig. 1a) and QFM (Fig. 2a) signals. The corresponding ACFs are shown in Figure 1b and Figure 2b.


DOI 10.15588/1607-3274-2024-2-3

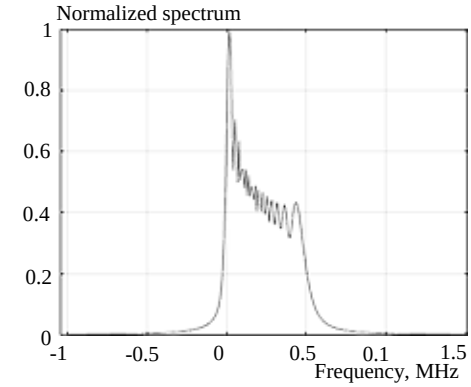



a

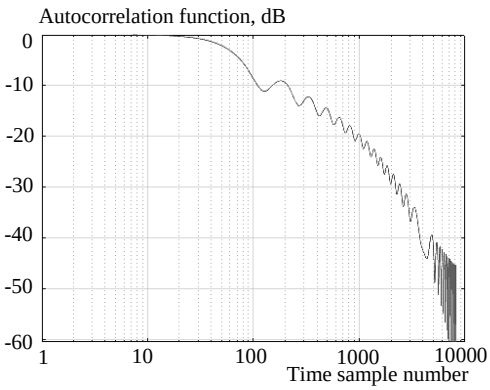



b

Figure 1 – LFM signal: a – spectrum, b – ACF




a




b

Figure 2 – QFM signal: a – spectrum, b – ACF

With regard to two-fragment signals, it should be noted that the main advantage of an NLFM signal with a QFM fragment is a decrease in MPSLL and an increase in the ML ACF decay rate.

The proposed LFM-QFM signal among those shown in Table 1 has the lowest MRSLL, in addition, the width ML of its ACF at the level of 0.707 decreased by 7.5% in relation to the LFM-LFM signal, this leads to an improvement in the range distinguishing ability.

Table 2 shows the results of modelling LFM-QFM signals (8) with different values of time-frequency parameters and corresponding estimates of the obtained indicators.

Table 2 – Results of experimental studies of two-fragment LFM-QFM signals

| No. | $T_1$, μs | $T_2$, μs | $\Delta f_1$, kHz | $\Delta f_2$, kHz | MPSL, dB | Rapid decline,SLL dB/dec |
|---|---|---|---|---|---|---|
| 1. | 60 | 60 | 100 | 350 | –19.12 | 25.0 |
| 2. | 60 | 60 | 100 | 500 | –20.0 | 26.0 |
| 3. | 60 | 60 | 100 | 600 | –20.46 | 27.0 |
| 4. | 80 | 80 | 100 | 600 | –20.03 | 28.0 |
| 5. | 90 | 90 | 100 | 650 | –20.41 | 29.0 |

Analysis of the results of Table 2 indicates that the use of the proposed signal provides a decrease in MPLSS (over 2 dB) compared to the LFM-LFM signal. With an increase in the QFM fragment deviation relative to the LFM fragment deviation, with a simultaneous increase in their duration, an increase in MPSLL (by about 1 dB) and an increase in the SL ACF decay rate (by 6 dB/dec) are observed.

Analysis of the fine structure of the LFM-QFM signal with the parameters of Table 1, namely, the frequency change graph, oscillogram, spectrum and ACF shown in Fig. 3, indicates their compliance with the input MM data.

There are no jumps in the instantaneous frequency and phase of the signal at the junction of the fragments, which is confirmed by the appearance of the oscillogram and the signal spectrum. This is evidence that the created MM adequately reproduces the physical processes that occur during the synthesis of LFM-QFM signals and compensates for the manifestations of the action of frequency-phase distortion sources.

It is determined experimentally that stable operation of MM (8) is provided by the ratio of duration of LFM and QFM fragments 1:1 with the step of changing the parameter 20, 30 units. In this case, the frequency deviation can change in increments of 50, 100 units. In the case of equality of frequency deviations of fragments or when the deviation of the LFM fragment is greater, the resulting NLFM signal in its properties approaches the usual LFM signal, which is natural.

Note the peculiarity of the shape of the ACF LFM-QFM signal (Fig. 3g) in comparison with the LFM-LFM (Fig. 4), which must be taken into account, in particular in radar applications.


DOI 10.15588/1607-3274-2024-2-3

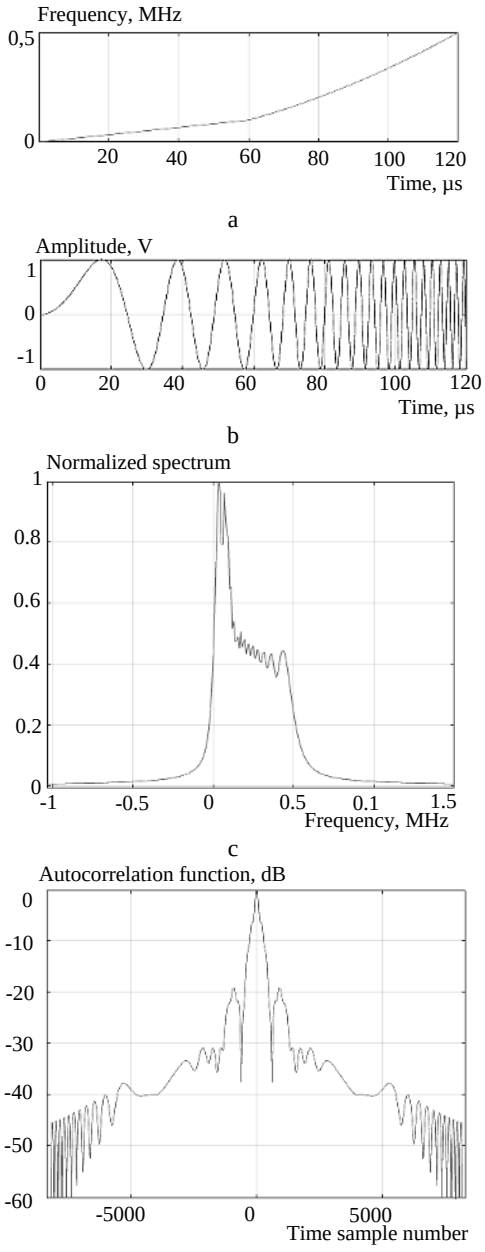



Figure 3 – Signal from LFM-QFM: a – frequency change plot, b – oscillogram, c – spectrum, d – ACF

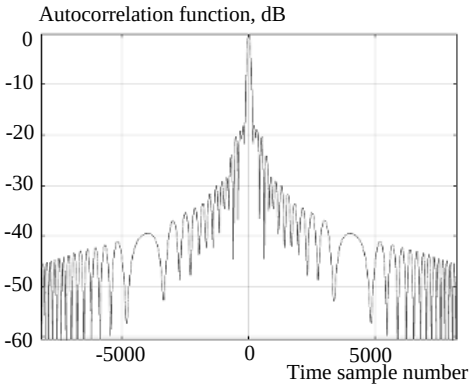



Figure 4 – ACF signal type LFM-LFM

As a result of the merger of ML with near SL, the form of ACF is distorted, at a level of about –10 dB a "pedestal" appears. This form of ACF leads to an increase in the level of the interference background in the presence of long-range interference.

## 6 DISCUSSION

In works [8, 9], to determine the magnitude of instantaneous phase hopping at the junction of fragments of the NLFM signal, the difference between its value at the beginning of the next fragment and the final value of the previous one was calculated. In this study, a more general approach was used to find analytical expressions of the magnitude of frequency-phase distortions in multifragment NLFM signals, which is based on the calculation of derivatives of the instantaneous phase function of fragments. In this case, the instantaneous frequency (phase) function must have a finite number of derivatives. From this follows the limitation of the list of types of frequency (phase) modulation functions of each of the individual fragments, which is a disadvantage of this approach. Despite this limitation, the application of the proposed approach allowed to substantiate the cause of distortion of the frequency-phase structure at the joints of multifragment NLFM signals. Such a reason is a change in the value of the highest derivative of the instantaneous phase function for fragments with the same FM laws, the appearance of new or possibly the disappearance of existing derivatives if the FM laws of neighboring fragments differ. Thus, the feasibility of further studies is seen, the effect of the disappearance of the existing derivative of the instantaneous phase during the transition to the next fragment on the frequency-phase structure of the resulting signal.

## CONCLUSIONS

**Scientific novelty.** According to the results of the studies, the theory of synthesis of two-fragment signals from NLFM has been developed for the case when the instantaneous phase of one of the signal fragments has the oldest third derivative, and for the other fragment the derivative of the instantaneous phase is limited to the second order. In this case, the source of frequency-phase distortion at the junction of the NLFM signal fragments is FMA (the third derivative of the instantaneous phase).

A new MM NLFM signal of the LFM-QFM type has been developed in the current time, the difference of which is the introduction of an additionals components of the linear and qadratic change in the instantaneous phase of the QFM fragment.

For the synthesized NLFM signal, in comparison with the LFM-LFM signal, the MPSLL value decreased by 2 dB, and its decay rate increased by 9 dB/dec. For the considered group of five such NLFM signals, the MRSLL level varies from –19.12 dB to –20.46 dB, which for two-fragment signals is quite high compared to, for example, [11].

The SLL decay rate of their ACF is within 25 dB/dec – 29 dB/dec, which also exceeds the value of this parame-


DOI 10.15588/1607-3274-2024-2-3

ter for [11], based on the assessment of the given graphic material.

The frequency change graph, oscillogram, spectrum and ACF of the synthesized LFM-QFM signal correspond to the theory and the given parameters, indicating the absence of frequency-phase distortion, which indicates the reliability and adequacy of the developed MM.

**The practical significance** of the obtained results lies in the possibility of using the proposed approach to the analysis and synthesis on its basis of a wide range of NLFM signals for use as probing in radar devices for various purposes. Experimentally obtained versions of the values of the frequency-time parameters of LFM-QFM signals make it possible to use them in radio electronic systems, in which the value of the range discriminating ability is determining, and passive interference is either absent or their influence can be neglected.

**Prospects for further research.** In the future, it is planned, using the developed approach, to synthesize MM three-fragment NLFM signals with the resulting FM law close to *S*-shaped.

## ACKNOWLEDGEMENTS

We thank the management of Ivan Kozhedub Kharkiv National Air Force University for the opportunity to conduct scientific research.

## REFERENCES

1. Levanon N., E. Mozeson Radar Signals [Text]. New York, John Wiley & Sons, 2004, 403 p.
2. Cook C. E., Bernfeld M. Radar Signals: An Introduction to Theory and Application [Text]. Boston, Artech House, 1993, 552 p.
3. Li Tan. Digital Signal Processing. Fundamentals and Applications [Text]. Georgia, Academic Press, 2008, 816 p.
4. Lei T., Liang L. Research and design of digital unit for direct digital frequency synthesizer [Text], *International Conference on Electronic Materials and Information Engineering (EMIE 2021), Journal of Physics: Conference Series*, 2021, Vol. 1907, Article № 012004. DOI:10.1088/1742-6596/1907/1/012004.
5. Genovese M., Napoli E., De Caro D., et al. Analysis and comparison of Direct Digital Frequency Synthesizers implemented on FPGA [Text], *Integration*, 2014, Vol. 47, Issue 2, pp. 261–271.
6. Blackledge J. Digital Signal Processing [Text]. Dublin, Horwood Publishing, second edition, 2006, 840 p.
7. Anil A., Prasad A. FPGA Implementation of DDS for Arbitrary wave generation [Text], *International Journal of Engineering Research and Applications*, 2021, Vol. 11, Issue 7, (Series-II), pp. 56–64. DOI: 10.9790/9622-1107025664.
8. Kostyria O. O., Hryzo A. A., Dodukh O. M. et al. Mathematical model of a two-fragment signal with a non-linear frequency modulation in the current period of time [Text], *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*, 2023, Vol. 92, pp. 60–67. DOI:10.20535/RADAP.2023.92.60-67.
9. Kostyria O. O., Hryzo A. A. , Dodukh O. M. et al. Improvement of mathematical models with time-shift of two- and tri-fragment signals with non-linear frequency modulation [Text], *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*, 2023, Vol. 93, pp. 22–30. DOI: 10.20535/RADAP.2023.93.22-30.
10. Kostyria O. O., Hryzo A. A. , Khizhnyak I. A. et al. Implementation of the Method of Minimizing the Side Lobe Level of Autocorrelation Functions of Signals with Nonlinear Frequency Modulation [Text], *Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia*, 2024, Vol. 95, pp. 16–22. DOI: 10.20535/RADAP.2024.95.16-22.
11. Anoosha Ch. and Krishna B. T. Peak Side Lobe Reduction analysis of NLFM and Improved NLFM Radar signal with Non-Uniform PRI [Text], *Aiub Journal of Science and Engineering (AJSE)*, 2022, Vol. 21, Issue 2, pp. 125–131.
12. Zhaoa Y., Ritchie M., Lua X. et al. Non-continuous piecewise nonlinear frequency modulation pulse with variable sub-pulse duration in a MIMO SAR radar system [Text], *Remote Sensing Letters*, 2020, Vol. 11, Issue 3, pp. 283–292. DOI: 10.1080/2150704X.2019.1711237
13. Valli N. A., Rani D. E., Kavitha C. Modified Radar Signal Model using NLFM [Text], *International Journal of Recent Technology and Engineering (IJRTE)*, 2019, Vol. 8, Issue 2S3, pp. 513–516. DOI:10.35940/ijrte.b1091.0782s319
14. Fan Z., Meng H. Coded excitation with Nonlinear Frequency Modulation Carrier in Ultrasound Imaging System [Text], *2020 IEEE Far East NDT New Technology & Application Forum (FENDT). Kunming, Yunnan province*. China, conference paper, IEEE, 2020, pp. 31–35. https://doi.org/10.1109/FENDT50467.2020.9337517.
15. Valli N. A., Rani D. E., Kavitha C. Performance Analysis of NLFM Signals with Doppler Effect and Background Noise [Text], *International Journal of Engineering and Advanced Technology (IJEAT)*, 2020, Vol. 9 Issue 3, pp. 737–742. DOI: 10.35940/ijeat.B3835.029320
16. Widyantara M. R., Suratman S.-F. Y., Widodo S. et al. Analysis of nonlinear Frequency Modulation (NLFM) Waveforms for Pulse Compression Radar [Text], *Jurnal Elektronika dan Telekomunikasi*, 2018, Vol. 18, Issue 1, pp. 27–34. DOI: 10.14203/jet.v18.27-34
17. Zhang Y., Deng Y., Zhang Z. et al. Analytic NLFM Waveform Design With Harmonic Decomposition for Synthetic Aperture Radar [Text], *IEEE Geoscience and Remote Sensing Letters*, 2022,Vol. 19, Article no 4513405. DOI: 10.1109/lgrs.2022.3204351
18. Ma H., Wang J., Sun X. et al. Joint Radar-Communication Relying on NLFM-MSK Design [Text], *Wireless Communications and Mobile Computing*, 2022, Vol. Article ID 4711132, 26 p. DOI:10.1155/2022/4711132.
19. Dhawan R., Parihar R., Choudhary A. Multiband dual- and cross-LFM waveform generation using a dual-drive Mach-Sender modulator [Text], *Indian Institute of Technology (IIT)*, 2023, 11 p. DOI:10.2139/ssrn.4161569
20. Xu Z., Wang X., Wang Y. Nonlinear Frequency-Modulated Waveforms Modeling and Optimization for Radar Applications [Text], *Mathematics*, 2022, Vol. 10, P. 3939. https://doi.org/10.3390/math10213939.
21. Wang B., Chen X., Li Y.et al. Research on Time Sidelobe Analysis on Pulse Compression Signal [Text] / B. Wang, // *Journal of Physics: 2022 2nd International Conference on Measurement Control and Instrumentation (MCAI 2022), Guangzhou*. China, Conf. Ser., 2022, Vol. 2366. DOI 10.1088/1742-6596/2366/1/012022
22. Saleh M., Omar S.-M., Grivel E. et al. A Variable Chirp Rate Stepped Frequency Linear Frequency Modulation Waveform Designed to Approximate Wideband Non-Linear Radar Waveforms [Text], *Digital Signal Processing*, 2021, Vol. 109, Article №102884. doi 10.1016/j.dsp.2020.102884
23. Swiercz E., Janczak D., Konopko K. Estimation and Classification of NLFM Signals Based on the Tim-Chirp [Text], *Sensors*, 2022,Vol. 22, Issue 21. Article № 8104. DOI: 10.3390/s22218104.
24. Septanto H., Sudjana O., Suprijanto D. A Novel Rule for Designing Tri-Stages Piecewise Linear NLFM Chirp [Text], *2022 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET) 6–7 December*


DOI 10.15588/1607-3274-2024-2-3

*2022, proceedings*. Bandung, Indonesia, IEEE, 2022, pp. 62–67. DOI:10.1109/ICRAMET56917.2022.9991201.
25. Xie Q., Zeng H., Mo Z. et al. A Two-Step Optimization Framework for Low Sidelobe NLFM Waveform Using Fourier Series [Text], *IEEE Geoscience and Remote Sensing Letters*, 2022, Vol. 19, Article no 4020905. DOI: 10.1109/LGRS.2022.3141081.
26. Jiang T., Li B., Li H. et al. Design and implementation of space borne NLFM radar signal generator [Text], *2nd IYSF Academic Symposium on Artificial Intelligence and Computer Engineering, 1 December 2021: proceedings.* Xi'an, China, 2021, Vol. 12079. DOI:10.1117/12.2623222
27. Van Zyl A. C., Wiehahn E. A., Cillers J. E. et al. Optimised multi-parameter NLFM pulse compression waveform for low time-bandwidth radar [Text], *International Conference on Radar Systems (RADAR 2022)*, 2022, pp. 289–294. DOI:10.1049/icp.2022.2332.
28. Singh A. K., Bae K.-B., Park S.-O. NLFM pulse radar for drone detection using predistortion technique [Text], *Journal of Electromagnetic Waves and Applications*, 2021, Vol. 35, Issue 3, pp. 416–429. DOI:10.1080/09205071.2020.1844598
29. Li J. Wang P., Zhang H. et al. A Novel Chaotic-NLFM Signal under Low Oversampling Factors for Deception Jamming Suppression [Text], *Remote Sens.* 2024, Vol. 16, Issue 1. https://doi.org/10.3390/rs16010035
30. Zhuang R., Fan H., Sun Y. et al. Pulse-agile waveform design for nonlinear FM pulses based on spectrum modulation [Text], *IET International Radar Conference*, 2021, pp. 964–969. DOI: 10.1049/icp.2021.0700.
31. Hague D. A. Generating waveform families using multi-tone sinusoidal frequency modulation [Text], *2020 IEEE International Radar Conference (RADAR).* Washington, DC, USA, 2020, pp. 946–951. DOI:10.48550/arXiv.2002.11742
32. Talluri S. R. Effects of High pass Filtering on Transmitted Signals of Non-Linear Frequency Modulated Radar Systems [Text], *International Journal of Advances in Microwave Technology*, 2020, Vol. 4(1), pp. 190–193. DOI:10.32452/IJAMT.2019.190193
33. Shuyi L., Jia Y., Liu Y. et al. Research on Ultra-Wideband NLFM Waveform Synthesis and Grating Lobe Suppression [Text], *Sensors*, 2022, Vol. 22, Issue 24, Article no 9829. https://doi.org/10.3390/s22249829.
34. Mohr C. A., McCormick P. M., Topliff C. A. et al. Gradient-based optimization of PCFM radar waveforms [Text], *IEEE Trans. Aerosp. Electron. Syst.*, 2021, Vol. 57, Issue 2, pp. 935–956. DOI:10.1109/TAES.2020.3037403
35. Roy A., Nemade H. B., Bhattacharjee R. Radar waveform diversity using nonlinear chirp with improved sidelobe level performance [Text], *AEU – International Journal of Electronics and Communications*, 2021, Vol. 136, Article no 153768. DOI: 10.1016/J.AEUE.2021.153768.
36. Mahipathi C., Pardhasaradhi B. P., Gunnery S. et al. Optimum Waveform Selection for Target State Estimation in the Joint Radar-Communication System [Text], *IEEE Open Journal of Signal Processing*, 2024, Vol. 5, pp. 459–477. DOI: 10.1109/OJSP.2024.3359997.
37. Peng S., Jian J., Zhitao L., et al. Nonlinear Frequency Modulation Tfm with Second-Order Tgv and Butterworth Filter for Detection of Cfrp Composites, *Preprint research Paper*, 2024, 12 p. URL: https://ssrn.com/abstract=4747512 or http://dx.doi.org/10.2139/ssrn.4747512.
38. Ping T., Song C., Qi Z. et al. PHS: A Pulse Sequence Method Based on Hyperbolic Frequency Modulation for Speed Measurement [Text], *International Journal of Distributed Sensor Networks*, 2024, Vol. 2024, Article ID 6670576,11 p. DOI:10.1155/2024/6670576
39. Yang J. and Sarkar T. K. A New Doppler-Tolerant Polyphase Pulse Compression Codes Based on Hyperbolic Frequency Modulation [Text], *2007 IEEE Radar Conference: proceedings.* Waltham, MA, USA, IEEE, 2007, pp. 265–270, DOI: 10.1109/RADAR.2007.374225.
40. Cheng Z., Sun Z., Wang J. et al. Magneto-acousto-electrical tomography using nonlinearly frequency-modulated ultrasound [Text], *Physics in Medicine & Biology*, 2024, Vol. 69(8). DOI:10.1088/1361-6560/ad2ee5.
41. Zhang Y., Wang W., Wang R. et al. A Novel NLFM Waveform With Low Sidelobes Based on Modified Chebyshev Window [Text], *IEEE Geoscience and Remote Sensing Letters*, 2020, Vol. 17, No. 5, pp. 814–818. DOI: 10.1109/LGRS.2019.2930817.

Received 06.03.2024.
Accepted 25.04.2024.

УДК 621.396.962

## МАТЕМАТИЧНА МОДЕЛЬ ПОТОЧНОГО ЧАСУ СИГНАЛУ З ПОСЛІДОВНИМ ПОЄДНАННЯМ ЛІНІЙНО-ЧАСТОТНО ТА КВАДРАТИЧНО МОДУЛЬОВАНИХ ФРАГМЕНТІВ

**Костиря О. О.** – д-р техн. наук, с.н.с., провідний науковий співробітник, Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна.

**Гризо А. А.** – канд. техн. наук, доц., начальник науково-дослідної лабораторії, Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна.

**Худов Г. В.** – д-р техн. наук, проф., начальник кафедри Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна.

**Додух О. М.** – канд. техн. наук, провідний науковий співробітник, Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна.

**Соломоненко Ю. С.** – канд. техн. наук, заступник начальника факультету з навчальної та наукової роботи, Харківський національний університет Повітряних Сил імені Івана Кожедуба, Харків, Україна.

### АНОТАЦІЯ

**Актуальність.** Одним з методів вирішення актуальної науково-технічної задачі зниження максимального рівня бічних пелюсток автокореляційних функцій радіолокаційних сигналів є застосування нелінійно-частотно модульованих сигналів. Це забезпечує округлення спектру сигналу, що еквівалентно ваговій (віконній) обробці сигналу у часовій області та може використовуватися спільно з нею. Ряд досліджень сигналів з нелінійною частотною модуляцією, які мають у своєму складі лінійно-частотно модульовані фрагменти, свідчить, що на стику фрагментів виникають спотворення їх частотно-фазової структури. Ці спотворення, у залежності від типу математичної моделі сигналу – поточного або зсунутого часу, викликають у сформованому сигналі відповідно стрибок миттєвої частоти та миттєвої фази або тільки фази. У роботі показано, що стри-


DOI 10.15588/1607-3274-2024-2-3

бки виникають у моменти зміни значення похідної миттєвої фази по закінченні лінійно-частотно модульованого фрагменту. Миттєва частота сигналу, яка є першою похідною миттєвої фази, має тлумачення швидкості обертання вектору сигналу на комплексній площині. Друга похідна миттєвої фази сигналу розуміється як швидкість частотної модуляції. Спотворення цих компонент призводить до появи додаткової складової у лінійному члені миттєвої фази, починаючи з другого фрагменту. Неврахування вказаних частотно-фазових (або тільки фазових) спотворень викликає викривлення спектру результуючого сигналу і, як правило, призводить до зростання максимального рівня бічних пелюсток його автокореляційної функції. Особливості застосування у складних сигналах фрагментів з законами частотної модуляції, які мають різну кількість похідних миттєвої фази сигналу, у відомих роботах не розглядалися, тому дану статтю присвячено цьому питанню.

**Мета роботи –** розробка математичної моделі поточного часу двофрагментних нелінійно-частотно модульованих сигналів з послідовним поєднанням лінійно-частотно та квадратично модульованих фрагментів, що забезпечує округлення спектру сигналу в області верхніх частот та зниження максимального рівня бічних пелюсток автокореляційної функції і збільшення швидкості його спадання.

**Метод.** У роботі досліджувалися нелінійно-частотно модульовані сигнали, які складаються з лінійно-частотно та квадратично модульованого фрагментів. За допомогою диференційного аналізу визначався ступінь впливу старшої похідної миттєвої фази на частотно-фазову структуру сигналу. Її зміни оцінювалися за допомогою методів часового та спектрально-кореляційного аналізу. Показники результуючого сигналу, що оцінювалися, – стрибки фази та частоти на стику фрагментів, форма спектру, максимальний рівень бічних пелюсток автокореляційної функції та швидкість їх спадання.

**Результати.** В статті отримала подальший розвиток теорія синтезу нелінійно-частотно модульованих сигналів. Теоретичний внесок полягає у визначенні нового механізму проявів частотно-фазових спотворень на стику фрагментів та його математичний опис. Встановлено, що при переході від лінійно-частотно модульованого фрагменту до квадратично модульованого першоджерелом частотно-фазових спотворень результуючого сигналу стає третя похідна миттєвої фази, яка за аналогією з теорією руху фізичних тіл є прискоренням частотної модуляції. Наявність цієї похідної призводить до появи нових складових у виразі миттєвої частоти та фази сигналу. Компенсація цих спотворень забезпечує зниження максимального рівня бічних пелюсток на 5 дБ та збільшення швидкості його спадання на 8 дБ/дек для розглянутого варіанту нелінійно-частотно модульованого сигналу.

**Висновки.** Розроблено нову математичну модель поточного часу для розрахунку значень миттєвої фази нелінійно-частотно модульованого сигналу, перший фрагмент якого має лінійну, а другий – квадратичну частотну модуляцію. Відмінністю цієї моделі від відомих є введення нових складових, які забезпечують компенсацію частотно-фазових спотворень на стику фрагментів та у фрагменті з квадратичною частотною модуляцією. Отримані осцилограма, спектр та автокореляційна функція одного з синтезованих двофрагментних сигналів відповідають теоретичному вигляду, що свідчить про адекватність та достовірність запропонованої математичної моделі.

**КЛЮЧОВІ СЛОВА:** математична модель; лінійна та квадратична частотна модуляція; максимальний рівень бічних пелюсток.

## ЛІТЕРАТУРА

1. Levanon N. Radar Signals [Text] / N. Levanon, E. Mozeson. – New York : John Wiley & Sons, 2004. – 403 p.
2. Cook C. E. Radar Signals: An Introduction to Theory and Application [Text] / C. E. Cook, M. Bernfeld. – Boston : Artech House, 1993. – 552 p.
3. Li Tan Digital Signal Processing. Fundamentals and Applications [Text] / Li Tan. – Georgia : Academic Press, 2008. – 816 p.
4. Lei T. Research and design of digital unit for direct digital frequency synthesizer [Text] / T. Lei, L. Liang // International Conference on Electronic Materials and Information Engineering (EMIE 2021), Journal of Physics: Conference Series, 2021. – Vol. 1907. – Article № 012004. DOI:10.1088/1742-6596/1907/1/012004.
5. Analysis and comparison of Direct Digital Frequency Synthesizers implemented on FPGA [Text] / [M. Genovese, E. Napoli, D. De Caro et al.] // Integration. – 2014. – Vol. 47, Issue 2. – P. 261–271.
6. Blackledge J. Digital Signal Processing [Text] / J. Blackledge. – Dublin : Horwood Publishing, second edition, 2006. – 840 p.
7. Anil A. FPGA Implementation of DDS for Arbitrary wave generation [Text] / A. Anil, A. Prasad // International Journal of Engineering Research and Applications. – 2021. – Vol. 11, Issue 7, (Series-II). – P. 56–64. DOI: 10.9790/9622-1107025664.
8. Mathematical model of a two-fragment signal with a non-linear frequency modulation in the current period of time [Text] / [O. O. Kostyria, A. A. Hryzo, O. M. Dodukh et al.] // Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia. – 2023. – Vol. 92. – P. 60–67. DOI:10.20535/RADAP.2023.92.60-67.
9. Improvement of mathematical models with time-shift of two- and tri-fragment signals with non-linear frequency modulation [Text] / [O. O. Kostyria, A. A. Hryzo, O. M. Dodukh et al.] // Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia. – 2023. – Vol. 93. – P. 22–30. DOI: 10.20535/RADAP.2023.93.22-30.
10. Implementation of the Method of Minimizing the Side Lobe Level of Autocorrelation Functions of Signals with Nonlinear Frequency Modulation [Text] / [O. O. Kostyria, A. A. Hryzo, I. A. Khizhnyak et al.] // Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia. – 2024. – Vol. 95. – P. 16–22. DOI: 10.20535/RADAP.2024.95.16-22.
11. Anoosha Ch. Peak Side Lobe Reduction analysis of NLFM and Improved NLFM Radar signal with Non-Uniform PRI [Text] / Ch Anoosha and B. T. Krishna // Aiub Journal of Science and Engineering (AJSE). – 2022. – Vol. 21, Issue 2. – P. 125–131.
12. Non-continuous piecewise nonlinear frequency modulation pulse with variable sub-pulse duration in a MIMO SAR radar system [Text] / [Y. Zhaoa, M. Ritchie, X. Lua et al.] // Remote Sensing Letters. – 2020. – Vol. 11, Issue 3. – P. 283–292. DOI: 10.1080/2150704X.2019.1711237
13. Valli N. A. Modified Radar Signal Model using NLFM [Text] / N. A. Valli, D. E. Rani, C. Kavitha // International Journal of Recent Technology and Engineering (IJRTE). – 2019. – Vol. 8, Issue 2S3. – P. 513–516. DOI:10.35940/ijrte.b1091.0782s319
14. Fan Z. Coded excitation with Nonlinear Frequency Modulation Carrier in Ultrasound Imaging System [Text] / Z. Fan, H. Meng // 2020 IEEE Far East NDT New Technology & Application Forum (FENDT). Kunming, Yunnan province, China : conference paper. IEEE. – 2020. – P. 31–35. https://doi.org/10.1109/FENDT50467.2020.9337517.
15. Valli N. A. Performance Analysis of NLFM Signals with Doppler Effect and Background Noise [Text] / N. A. Valli, D. E. Rani, C. Kavitha // International Journal of Engineering


DOI 10.15588/1607-3274-2024-2-3

and Advanced Technology (IJEAT). – 2020. – Vol. 9, Issue 3. – P. 737–742. DOI: 10.35940/ijeat.B3835.029320
16. Analysis of nonlinear Frequency Modulation (NLFM) Waveforms for Pulse Compression Radar [Text] / [M. R. Widyantara, S.-F. Y. Suratman, S. Widodo et al.] // Jurnal Elektronika dan Telekomunikasi. – 2018. – Vol. 18, Issue 1. – P. 27–34. DOI: 10.14203/jet.v18.27-34
17. Analytic NLFM Waveform Design With Harmonic Decomposition for Synthetic Aperture Radar [Text] / [Y. Zhang, Y. Deng, Z. Zhang et al.] // IEEE Geoscience and Remote Sensing Letters. – 2022. – Vol. 19, Article no 4513405. doi:10.1109/lgrs.2022.3204351
18. Joint Radar-Communication Relying on NLFM-MSK Design [Text] / [H. Ma, J. Wang, X. Sun et al.] // Wireless Communications and Mobile Computing. – 2022. – Vol. Article ID 4711132, 26 p. DOI:10.1155/2022/4711132.
19. Dhawan R. Multiband dual- and cross-LFM waveform generation using a dual-drive Mach-Sender modulator [Text] / R. Dhawan, R. Parihar, A. Choudhary // Indian Institute of Technology (IIT), 2023. – 11 p. DOI:10.2139/ssrn.4161569
20. Xu Z. Nonlinear Frequency-Modulated Waveforms Modeling and Optimization for Radar Applications [Text] / Z. Xu, X. Wang, Y Wang // Mathematics. – 2022. – Vol. 10. – P. 3939. https://doi.org/10.3390/math10213939.
21. Research on Time Sidelobe Analysis on Pulse Compression Signal [Text] / [B. Wang, X. Chen, Y. Li et al.] // Journal of Physics: 2022 2nd International Conference on Measurement Control and Instrumentation (MCAI 2022), Guangzhou, China. Conf. Ser. – 2022. – Vol. 2366. DOI 10.1088/1742-6596/2366/1/012022
22. A Variable Chirp Rate Stepped Frequency Linear Frequency Modulation Waveform Designed to Approximate Wideband Non-Linear Radar Waveforms [Text] / [M. Saleh, S.-M. Omar, E. Grivel et al.] // Digital Signal Processing. – 2021. – Vol. 109. – Article №102884. DOI 10.1016/j.dsp.2020.102884
23. Swiercz E. Estimation and Classification of NLFM Signals Based on the Time-Chirp [Text] / E. Swiercz, D. Janczak, K. Konopko // Sensors. – 2022. – Vol. 22, Issue 21. Article № 8104. doi: 10.3390/s22218104.
24. Septanto H. A Novel Rule for Designing Tri-Stages Piecewise Linear NLFM Chirp [Text] / H. Septanto, O. Sudjana, D. Suprijanto // 2022 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET) 6–7 December 2022: proceedings. – Bandung, Indonesia : IEEE, 2022. – P. 62–67. DOI:10.1109/ICRAMET56917.2022.9991201.
25. A Two-Step Optimization Framework for Low Sidelobe NLFM Waveform Using Fourier Series [Text] / [Q. Xie, H. Zeng, Z. Mo et al.] // IEEE Geoscience and Remote Sensing Letters. – 2022. – Vol. 19, Article no 4020905. DOI: 10.1109/LGRS.2022.3141081.
26. Design and implementation of space borne NLFM radar signal generator [Text] / [T. Jiang, B. Li, H. Li et al.] // 2nd IYSF Academic Symposium on Artificial Intelligence and Computer Engineering, 1 December 2021: proceedings. – Xi'an, China. – 2021. – Vol. 12079. doi:10.1117/12.2623222
27. Optimised multi-parameter NLFM pulse compression waveform for low time-bandwidth radar [Text] / [A. C. van Zyl, E. A. Wiehahn, J. E. Cillers et al.] // International Conference on Radar Systems (RADAR 2022). – 2022. – P. 289–294. DOI:10.1049/icp.2022.2332.
28. Singh A. K. NLFM pulse radar for drone detection using predistortion technique [Text] / A. K. Singh, K.-B. Bae, S.-O. Park // Journal of Electromagnetic Waves and Applications, 2021. – Vol. 35, Issue 3. – P. 416–429. doi:10.1080/09205071.2020.1844598
29. A Novel Chaotic-NLFM Signal under Low Oversampling Factors for Deception Jamming Suppression [Text] / [J. Li, P. Wang, H. Zhang et al.] // Remote Sens. 2024. – Vol. 16, Issue 1. https://doi.org/10.3390/rs16010035
30. Pulse-agile waveform design for nonlinear FM pulses based on spectrum modulation [Text] / [R. Zhuang, H. Fan, Y. Sun et al.] // IET International Radar Conference. – 2021. – P. 964–969. DOI: 10.1049/icp.2021.0700.
31. Hague D. A. Generating waveform families using multi-tone sinusoidal frequency modulation [Text] / D. A. Hague // 2020 IEEE International Radar Conference (RADAR): Washington, DC, USA, 2020. – P. 946–951. DOI:10.48550/arXiv.2002.11742
32. Talluri S. R. Effects of High pass Filtering on Transmitted Signals of Non-Linear Frequency Modulated Radar Systems [Text] / S. R. Talluri // International Journal of Advances in Microwave Technology. – 2020. – Vol. 4(1). – P. 190–193. DOI:10.32452/IJAMT.2019.190193
33. Research on Ultra-Wideband NLFM Waveform Synthesis and Grating Lobe Suppression [Text] / [L. Shuyi, Y. Jia, Y. Liu et al.] // Sensors. – 2022. – Vol. 22, Issue 24. – Article no 9829. https://doi.org/10.3390/s22249829.
34. Gradient-based optimization of PCFM radar waveforms [Text] / [C. A. Mohr, P. M. McCormick, C. A. Topliff et al.] // IEEE Trans. Aerosp. Electron. Syst. – 2021. – Vol. 57, Issue 2. – P. 935–956. DOI:10.1109/TAES.2020.3037403
35. Roy A. Radar waveform diversity using nonlinear chirp with improved sidelobe level performance [Text] / A. Roy, H. B. Nemade, R. Bhattacharjee // AEU – International Journal of Electronics and Communications. – 2021. – Vol. 136. – Article no 153768. DOI: 10.1016/J.AEUE.2021.153768.
36. Optimum Waveform Selection for Target State Estimation in the Joint Radar-Communication System [Text] / [C. Mahipathi, B. P. Pardhasaradhi, S. Gunnery et al.] // IEEE Open Journal of Signal Processing, 2024. – Vol. 5. – P. 459–477. DOI: 10.1109/OJSP.2024.3359997.
37. Nonlinear Frequency Modulation Tfm with Second-Order Tgv and Butterworth Filter for Detection of Cfrp Composites / [S. Peng, J. Jian, L. Zhitao, et al.] // Preprint research Paper, 2024. – 12 p. URL: https://ssrn.com/abstract=4747512 or http://dx.doi.org/10.2139/ssrn.4747512.
38. PHS: A Pulse Sequence Method Based on Hyperbolic Frequency Modulation for Speed Measurement [Text] / [T. Ping, C. Song, Z. Qi et al.] // International Journal of Distributed Sensor Networks. – 2024. – Vol. 2024, Article ID 6670576. – 11 p. doi:10.1155/2024/6670576
39. Yang J. A New Doppler-Tolerant Polyphase Pulse Compression Codes Based on Hyperbolic Frequency Modulation [Text] / J. Yang and T. K. Sarkar // 2007 IEEE Radar Conference: proceedings. – Waltham, MA, USA: IEEE, 2007. – P. 265–270. DOI: 10.1109/RADAR.2007.374225.
40. Cheng Z. Magneto-acousto-electrical tomography using nonlinearly frequency-modulated ultrasound [Text] / [Z. Cheng, Z. Sun, J. Wang et al.]. // Physics in Medicine & Biology. – 2024. – Vol. 69(8). DOI: 10.1088/1361-6560/ad2ee5.
41. A Novel NLFM Waveform With Low Sidelobes Based on Modified Chebyshev Window [Text] / [Y. Zhang, W. Wang, R. Wang et al.] // IEEE Geoscience and Remote Sensing Letters. – 2020. – Vol. 17, No. 5. – P. 814–818. DOI: 10.1109/LGRS.2019.2930817.


DOI 10.15588/1607-3274-2024-2-3

p-ISSN 1607-3274 Радіоелектроніка, інформатика, управління. 2024. № 2
e-ISSN 2313-688X Radio Electronics, Computer Science, Control. 2024. № 2

# МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ

# MATHEMATICAL AND COMPUTER MODELING

UDC 621.394.74:519.872

## ANALYSIS OF THE RESULTS OF SIMULATION MODELING OF THE INFORMATION SECURITY SYSTEM AGAINST UNAUTHORIZED ACCESS IN SERVICE NETWORKS

**Ismailov B. G.** – Dr. Sc., Professor of the Department of Computer Systems and Programming, National Aviation Academy, Baku, Azerbaijan.

**ABSTRACT**

**Context.** An analysis of the service network shows that insufficient information security in service networks is the cause of huge losses incurred by corporations. Despite the appearance of a number of works and materials on standardization, there is currently no unified system for assessing information security in the field of information security. It should be noted that existing methods, as well as accumulated experience in this area, do not completely overcome these difficulties. This circumstance confirms that this problem has not yet been sufficiently studied and, therefore, remains relevant. The presented work is one of the steps towards creating a unified system for assessing information security in service networks.

**Objective.** Development of an algorithm and simulation model, analysis of simulation results to determine the key characteristics of the Information Security System, providing the capability for complete closure, through the security system, of all potential threat channels by ensuring control over the passage of all unauthorized access requests through defense mechanisms.

**Method.** To solve the problem, a simulation method was applied using the principles of queuing system modeling. This method makes it possible to obtain the main characteristics of the Information Security System from the unauthorized access with a limited amount of buffer memory.

**Results.** Algorithms, models, and methodology have been developed for the development of Information Security System from unauthorized access, considered as a single-phase multi-channel queuing system with a limited volume of buffer memory. The process of obtaining model results was implemented in the General Purpose Simulation System World modelling system, and comparative assessments of the main characteristics of the Information Security System were carried out for various laws of distribution of output parameters, i.e., in this case, unauthorized access requests are the simplest flows, and the service time obeys exponential, constant, and Erlang distribution laws.

**Conclusions.** The conducted experiments based on the algorithm and model confirmed the expected results when analyzing the characteristics of the Information Security System from the unauthorized access as a single-phase multi-channel queuing system with a limited waiting time for requests in the queue. These results can be used for practical construction of new or modification of existing Information Security System s in service networks of objects of various purposes. This work is one of the approaches to generalizing the problems under consideration for systems with a limited volume of buffer memory. Prospects for further research include research and development of the principles of hardware and software implementation of Information Security System in service networks.

**KEYWORDS:** unauthorized access, information security systems, information security, queuing systems, defense mechanism, simulation modeling.

### ABBREVIATIONS

BM is a Buffer Memory;

DM is a Defense Mechanism;

GPSS World is a General Purpose Simulation System (latest version of GPSS);

ISS is an Information Security System;

QS is a Queuing System;

UA is an Unauthorized Access.

### NOMENCLATURE

AVE.C is an Average Queue Length;

CUM.% is a Cumulative Percentage, expressed as a percentage of the total number of random values;

ENTRIES is a number of requests in DM;

FREQUENCY is the number of random values falling within the given interval;

$L_q$ is an average queue length;

$L^0$ is the permissible limit values $L_q$;

$M$ is a mathematical expectation symbol;

MEAN is a mean value of the corresponding random variable;

$N$ is a number of DMs in ISS;

$N_0$ is the permissible limit values $N$;


DOI 10.15588/1607-3274-2024-2-4

$p_1$ is a probability of blocking UA requests;

$p_2$ is a probability of UA requests bypassing protected resources;

RANGE is a lower and upper bound of the frequency interval;

RETRY is a number of requests waiting for the fulfillment of a specific condition depending on the state of this table;

STD.DEV is a Standard Deviation of the random variable;

$T\_U$ is a time of requests' stay in the system;

$T\_W$ is a time of requests' waiting in the queue;

UTIL is a Utilization Coefficient of DM;

$\lambda$ is an intensity of various threats at the entrance of ISS;

$\lambda_0$ is the permissible limit values $\lambda$;

$\mu$ is an intensity of servicing UA requests;

$\mu_0$ is the permissible limit values $\mu$;

$\tau_0$ is a service delays;

$\rho$ is a normalized intensity.

## INTRODUCTION

This work is dedicated to approaches in researching ISS in service networks, addressing security issues characteristic of systems with limited BM capacities. When addressing security issues in service networks, the primary determinant is the security class of the network, defining a set of DMs that constitute the hardware or software part implemented in the network. In service networks, intentional UA requests are often received alongside regular requests, targeting confidential information from illegal users, which can lead to network disruptions. It should be noted that DMs, influencing the entire information security process, may operate in constant information interaction with other elements of the ISS. The operation of DMs is described by four possible states: operational, non-operational, diagnosed, and restored. In ISS, the possibility of an undesirable event related to the reliability characteristics of DMs, leading to various types of losses, is considered a risk. However, approaches associated with the risk arising from the reliability characteristics of DMs are not considered in this work, i.e., it is assumed that all DMs are reliable.

The task related to the security problem in service networks is addressed by examining the ISS, ensuring the complete closure of all possible channels of threat manifestation through the security system. This is achieved by controlling the passage of all UA requests through DM.

**The object of the study** is an ISS against UA with a limited amount of BM in service networks.

**The subject of the study** is to determine the structure of the object, i.e. determination of the main characteristics of the system – security of information from UA with a limited amount of BM in service networks.

**The goal of the work** is to develop an algorithm and simulation model, analyze the results of the simulation model, allowing us to determine the main characteristics of an ISS against UA with a **limited** amount of BM in service networks.

## 1 PROBLEM STATEMENT

The structure of ISS with limited BM is considered (Fig. 1), where all input streams are directed to DM for servicing. As noted earlier, the security system allows for the complete closure of all possible channels of threat manifestation by controlling the passage of all UA requests through DM. It is assumed that the examined ISS structure ensures maximum information security for service networks. This structure constitutes a hardware and software complex interacting with random event streams, which are conditioned by the actions of attackers, improper access rights distribution, unauthorized software usage, as well as errors in identification and authentication software and technical complexes.

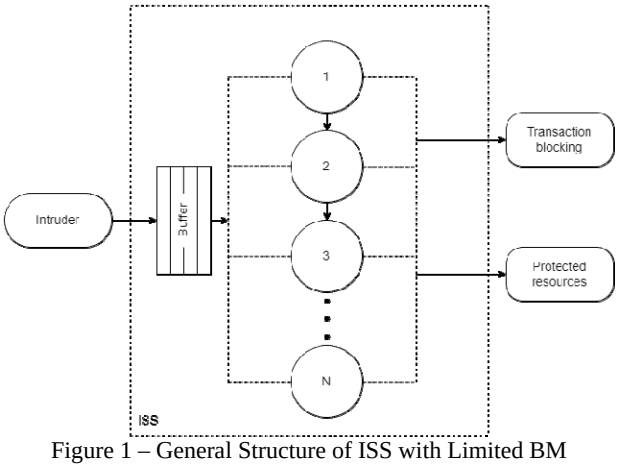



Figure 1 – General Structure of ISS with Limited BM

The assumption is that the intruder (attacker, UA requests) at the system's entrance generates various threats with intensity $\lambda$. The ISS consists of $N$ DMs that introduce delays $\tau_0 = 1/\mu$ in service. If we consider the intruder block as an information source and DMs as devices operating in parallel, the mathematical model of the ISS can be regarded as a single-phase, multi-channel QS with limited BM. Taking into account the complex nature of UA request servicing (filtering UA requests, detection and classification of UA attempts, blocking or allowing UA requests to access protected resources, etc.), Poisson formulas are suggested as the probability loss function for UA requests due to system overload in [1]:

$$P(\lambda, \mu, N) = \sum_{j=N}^{\infty} \left( \rho^j / j! \right) e^{-\rho}.$$

Then, the problem of determining the optimal values of ISS characteristics can be formulated as the minimization of the mathematical expectation of the probability loss function for UA requests due to system overload:


DOI 10.15588/1607-3274-2024-2-4

$$M\left[\sum_{j=N}^{\infty}\left(\rho^{j}/j!\right)e^{-\rho}\right]\rightarrow \min$$

with $\lambda \geq \lambda_0$, $\mu \geq \mu_0$, $N \geq N_0$ , $L_q \leq L^0$,

where $\rho = \lambda/\mu$. Problems related to insufficient information security in service networks, and the task of determining optimal ISS characteristics against UA for various cases, have been considered and analytically solved in [1] and optimal values for QS characteristics with and without waiting requests in the queue have been obtained.

However, for a detailed analysis of ISS characteristics against UA across a wide range of input and output parameters, it is preferable to utilize simulation modeling methods, considering it as a single-phase, multi-channel QS both with and without waiting. Given the volume of obtained results from the simulation model, we will limit the discussion here to the analysis of the simulation model results for a QS with limited waiting requests, encompassing a broad range of input and output parameters.

Thus, based on the presented structure of the ISS, the task in this work is to analyze the results of simulation modeling of a single-phase multi-channel QS with limited BM. To achieve this, using simulation modeling, it is necessary to determine the structural and temporal characteristics of the ISS within the specified values of concurrently operating service devices (DMs).

## 2 LITERATURE REVIEW

Analysis and accumulated experience demonstrate that insufficient information security in service networks leads to significant losses for corporations. This underscores the high importance of the information security problem. An analysis of the current state of the issue in the field of information security and the development of ISS reveals serious challenges, largely stemming from the absence of a unified system for assessing information security. Such a system would enable a quantitative evaluation during the design and operation of service networks [2–8]. It is worth noting, that due to a lack of sufficient experience in designing ISS, tasks related to its construction must be addressed at the early stages of service network design.

Currently, given the increasing number of scientific studies and companies specializing in information security in service networks, this problem is insufficiently explored [4–11, 13]. It should be noted that one of the most obvious causes of ISS violation is intentional UA requests for confidential information by illegal users, followed by undesirable manipulations with this information [1, 2,12]. The effectiveness of information security protection in service networks is primarily determined by the service network's security class [1, 2, 11, 14–16], which defines the set of DMs implemented in the network.

In [1], due to the fact that the security system fails to completely close all possible threat channels, a structure for the ISS was proposed. Unlike existing structures, in this framework, each input stream is provided with a DM for maintenance.

In the work [1], a structure for the ISS with losses is proposed, featuring both limited and unlimited BM. This structure ensures maximum information security in service networks by controlling the passage of all UA requests through DMs. In contrast to [1], an analysis of the simulation model results for ISS with limited BM is conducted here, encompassing a broad range of input and output parameters.

## 3 MATERIALS AND METHODS

To determine the characteristics of the ISS that allow it to operate within limited resources, it is assumed that the input flow of information, i.e., UA requests, is Poisson distributed, and the service time follows exponential, constant, and Erlang distribution laws. Algorithms for the simulation model of the service process have been developed for three cases to adequately describe the operation of the ISS against UA:

1. Incoming requests to the ISS and service time follow an exponential distribution.
2. Incoming requests to the ISS follow an exponential distribution, while the service time follows a uniform distribution.
3. Incoming requests to the ISS follow an exponential distribution, while the service time follows an Erlang distribution.

The developed algorithm for the operation of the ISS against UA includes the following steps:

– setting the minimum permissible limit values for the number of concurrently operating service devices (DMs) and the maximum permissible limit values for the queue length, defining the BM volume;

– to conduct a detailed analysis of the properties of the investigated system, a table structure is organized for queue waiting time and request residence time. An upper limit for the first frequency interval is specified, along with the values for all other frequency intervals and the quantity of frequency intervals. The goal here is to build histograms of the probability density function for the waiting time in the queue and the residence time of requests in the system based on the accumulation of the frequency of occurrence of random variables within the specified frequency intervals;

– when a request is received, the system checks for available places in the queue. If there is no available space in the queue, the request is rejected and exits the system;

– otherwise, if all DMs are occupied, the UA request waits in the system's BM queue until one of the DMs becomes available provided there is free space in the BM.

– upon the release of one of the DMs, the UA request enters this available DM, and the process of filtering UA requests, detecting, and classifying UA attempts takes place. As a result, the initial UA stream is thinned out with certain probabilities $p_1$, $p_2 = 1 - p_1$ forming an output stream, in other words, with a probability of $p_1$ block-


DOI 10.15588/1607-3274-2024-2-4

ing occurs, while with a probability of $p_2$ UA requests are allowed to pass through to the protected resources.

**Note 1.** The values of probabilities $p_1$, $p_2$ are determined based on statistical analysis.

Based on the proposed algorithm covering three cases of ISS functioning against UA as single-phase, multichannel QS with a limited buffer size, simulation models were developed using the GPSS World modeling language. For $N = \overline{2,5}$ during the simulation the model allows you to determine such characteristics as ENTRIES, UTIL, AVE.C, MEAN, STD.DEV, RANGE, RETRY, FREQUENCY, CUM.%.

## 4 EXPERIMENTS

Based on the execution of the simulation model for the average values of real data, with $N = \overline{2,5}$, $\lambda = 1/3500\ ms$ and $\mu = 1/1700\ ms$ results were obtained for three cases:

1. Incoming requests to the ISS and service time follow an exponential distribution.

In the first case, the results of a simulation model of the functioning of the information system were obtained – reports and histograms of the distribution densities of the residence time $T\_U$ and waiting time $T\_W$ of requests, with $N = \overline{2,5}$ (see Appendix A, Fig. A.1–A.8).

Based on the obtained reports, Table 1 was created, providing the dynamics of changes in the number of requests in the DM, the average queue length, and the utilization coefficient of the DM depending on the number of DM ( $N$ ) during the modeling period for the first case.

Table 1 – Dynamics of changes in characteristics depending on the number of DMs for the first case

| The number of DM | The number of requests in the DM | The average queue length | The utilization coefficient of the DM |
|---|---|---|---|
| 2 | 90266 | 1.866 | 0.933 |
| 3 | 99605 | 2.047 | 0.682 |
| 4 | 99989 | 2.056 | 0.514 |
| 5 | 100002 | 2.070 | 0.414 |

The analysis of the dynamics of these parameters shows that with an increase in the number of DM from 2 to 5:

– the number of requests in DM increases, with a difference of 9736 requests;

– the average queue length increases, with a difference of 0.204;

– the utilization coefficient of DM decreases, with a difference of 0.519.

In the models, 10 frequency intervals were chosen for building histograms, and the length of frequency intervals was selected as 0.0004 time units for waiting time in the queue and 0.0008 time units for the service time. The analysis conducted shows that in the first case, with a change in the number of DM from 3 to 5, the characteristics of the density distribution of the residence time $T\_U$ and waiting time $T\_W$ of requests do not change.

**Note 2.** For clarity of histograms, it is desirable to have a large number of frequency intervals. To obtain an objective picture, it is necessary to have a large sample of random variables, which is not always possible and feasible.

**Note 3.** The values of interval lengths and the number of frequency intervals are selected experimentally during several runs of the simulation model or based on assumed values of the mean and standard deviation of the corresponding random variable.

2. The requests entering the ISS follow an exponential distribution, while the service time adheres to a uniform distribution.

In the second case, the results of a simulation model of the functioning of the ISS were obtained – reports and histograms of the distribution densities of the residence time and waiting time of requests, with $N = \overline{2,5}$ (see Appendix B, Fig. B.1–B.8).

Based on the obtained reports, Table 2 was created, providing the dynamics of changes in the number of requests in the DM, the average queue length, and the utilization coefficient of the DM depending on the number of DM ( $N$ ) during the modeling period for the second case.

Table 2 – Dynamics of changes in characteristics depending on the number of DMs for the second case

| The number of DM | The number of requests in the DM | The average queue length | The utilization coefficient of the DM |
|---|---|---|---|
| 2 | 93922 | 1.932 | 0.966 |
| 3 | 99980 | 2.061 | 0.687 |
| 4 | 100002 | 2.056 | 0.514 |
| 5 | 100002 | 2.053 | 0.411 |

The analysis of the dynamics of these parameters shows that with an increase in the number of DM from 2 to 5:

– the number of requests in DM increases, with a difference of 6080 requests;

– the average queue length increases, with a difference of 0.121;

– the utilization coefficient of DM decreases, with a difference of 0.555.

The analysis conducted shows that in the first case, with a change in the number of DM from 3 to 5, the characteristics of the density distribution of residence time $T\_U$ and waiting time $T\_W$ of requests do not change.

3. The incoming requests to the ISS follow an exponential distribution, while the service time follows an Erlang distribution.

In the third case, the results of the ISS simulation model were obtained – reports and histograms of the density distribution of the residence time $T\_U$ and the waiting time $T\_W$ of requests at $N = \overline{2,5}$ (see Appendix C, Fig. C.1–C.8).


DOI 10.15588/1607-3274-2024-2-4

Based on the obtained reports, Table 3 was created, providing the dynamics of changes in the number of requests in the DM, the average queue length, and the utilization coefficient of the DM depending on the number of DM ( $N$ )during the modeling period for the third case.

Table 3 – Dynamics of changes in characteristics depending on the number of DMs for the third case

| The number of DM | The number of requests in the DM | The average queue length | The utilization coefficient of the DM |
|---|---|---|---|
| 2 | 91907 | 1.898 | 0.949 |
| 3 | 99633 | 2.053 | 0.684 |
| 4 | 100003 | 2.061 | 0.515 |
| 5 | 100002 | 2.058 | 0.412 |

The analysis of the dynamics of these parameters shows that with an increase in the number of DM from 2 to 5:

– the number of requests in DM increases, with a difference of 8095 requests;

– the average queue length increases, with a difference of 0.159;

– the utilization coefficient of DM decreases, with a difference of 0.537.

The analysis conducted shows that in the first case, with a change in the number of DM from 3 to 5, the characteristics of the density distribution of the residence time $T\_U$ and waiting time $T\_W$ of requests do not change.

Based on Tables 1–3, the dynamics of changes in the differences in the number of requests in the DM, average queue length, and the utilization coefficient of the DM for three cases with $N = \overline{2,5}$, and the nature of these differences are presented in Fig. 2–4.

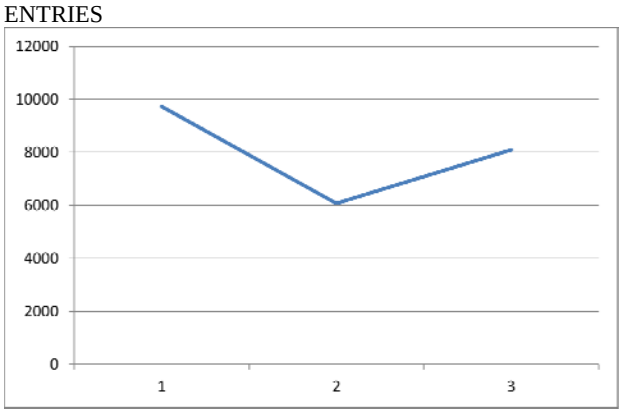



Figure 2 – The nature of the change in the differences in the number of requests in the DM for three cases with $N = \overline{2,5}$

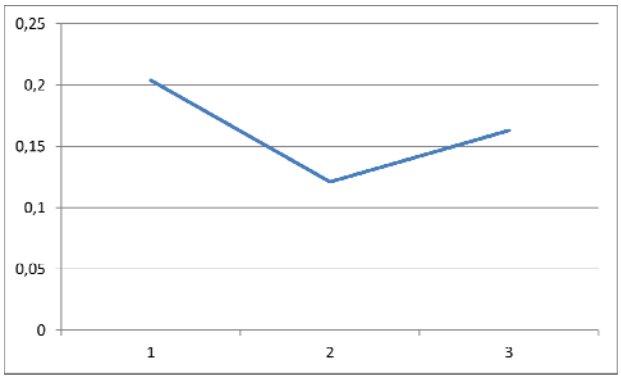



Figure 3 – The nature of changes in the differences in the average queue length for three cases with $N = \overline{2,5}$

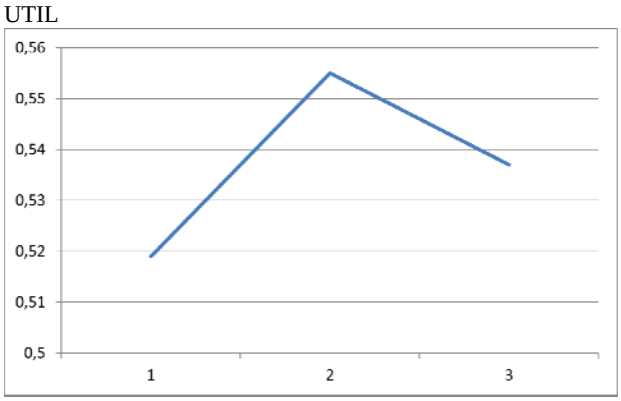



Figure 4 – The nature of changes in the differences in the coefficients of use of DM for three cases with $N = \overline{2,5}$

The results obtained from Table 1–3 and Fig. 2–4 show that with an increase in the number of DM from 2 to 5 in three cases:

– the nature of the change in the differences in the number of requests in the DM is 9736, 6080 and 8095;

– the nature of the change in the differences in the average queue length is 0,204; 0,121 and 0.163;

– the nature of the change in the differences in the utilization coefficient DM is 0.519; 0.555 and 0.537.

## CONCLUSIONS

The current task of developing an algorithm and simulation model, along with the analysis of simulation model results to determine the key characteristics of the ISS, is being addressed. This aims to provide the capability for complete closure, through the security system, of all potential threat channels by ensuring control over the transition of all UA through the DM.

**The scientific novelty** of the obtained results lies in the fact that, for the first time, algorithms, and simulation models, as well as a methodology for developing the ISS, have been proposed and developed based on the analysis of structural and temporal characteristics of ISS from UA. This includes treating it as a single-phase multi-channel queueing system with limited waiting time in the queue across a wide range of input and output parameters. The experiments conducted using the algorithm and model


DOI 10.15588/1607-3274-2024-2-4

confirmed the expected results when analyzing the characteristics of the ISS from UA.

**The practical significance** of the results lies in their applicability for the practical construction of new or modification of existing ISS in networks for various purposes. This work represents one of the approaches to generalizing the considered problems for systems with a limited BM.

**Prospects for** further **research** include the exploration and development of hardware and software implementation principles for ISS from UA with a limited BM in service networks.

## ACKNOWLEDGMENTS

The work was carried out within the framework of the research topic “Mathematical modeling of computer and communication networks” of the Department of Computer Systems and Programming of the National Academy of Aviation (Azerbaijan).

## Appendix A

The first case’s simulation model’s results of the ISS’s operation.

QUEUE

| | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 1 | 90266 | 10069 | 4.826 | 0.002 | 0.002 | 0 |

STORAGE

| | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 2 | 0 | 0 | 2 | 90266 | 1 | 1.866 | 0.933 | 0 | 0 |

TABLE

| | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.002 | 0.001 | | | | 0 | | |
| | | | | _ | 0.000 | | 19250 | 21.33 |
| | | | 0.000 | - | 0.001 | | 9868 | 32.26 |
| | | | 0.001 | - | 0.001 | | 10215 | 43.58 |
| | | | 0.001 | - | 0.002 | | 10616 | 55.34 |
| | | | 0.002 | - | 0.002 | | 10424 | 66.88 |
| | | | 0.002 | - | 0.002 | | 9052 | 76.91 |
| | | | 0.002 | - | 0.003 | | 7398 | 85.11 |
| | | | 0.003 | - | 0.003 | | 5324 | 91.01 |
| | | | 0.003 | - | 0.004 | | 3461 | 94.84 |
| | | | 0.004 | - | _ | | 4657 | 100.00 |
| T_U | 0.002 | 0.001 | | | | 0 | | |
| | | | | _ | 0.001 | | 14294 | 17.58 |
| | | | 0.001 | - | 0.002 | | 17199 | 38.73 |
| | | | 0.002 | - | 0.002 | | 18110 | 61.00 |
| | | | 0.002 | - | 0.003 | | 15313 | 79.83 |
| | | | 0.003 | - | 0.004 | | 9382 | 91.37 |
| | | | 0.004 | - | 0.005 | | 4420 | 96.81 |
| | | | 0.005 | - | 0.006 | | 1704 | 98.90 |
| | | | 0.006 | - | 0.006 | | 629 | 99.68 |
| | | | 0.006 | - | 0.007 | | 206 | 99.93 |
| | | | 0.007 | - | | | 57 | 100.00 |

Figure A.1 – Fragment of the report for the model with $N = 2$

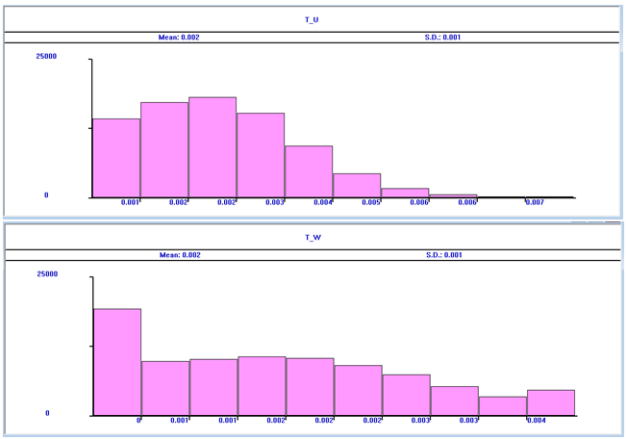



Figure A.2 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 2$

QUEUE

| | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 1 | 99605 | 53200 | 0.951 | 0.000 | 0.001 | 0 |

STORAGE

| | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 3 | 0 | 0 | 3 | 99605 | 1 | 2.047 | 0.682 | 0 | 0 |

TABLE

| | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | | _ | 0.000 | | 75800 | 76.10 |
| | | | 0.000 | - | 0.001 | | 11400 | 87.55 |
| | | | 0.001 | - | 0.001 | | 6100 | 93.67 |
| | | | 0.001 | - | 0.002 | | 3356 | 97.04 |
| | | | 0.002 | - | 0.002 | | 1679 | 98.73 |
| | | | 0.002 | - | 0.002 | | 809 | 99.54 |
| | | | 0.002 | - | 0.003 | | 285 | 99.82 |
| | | | 0.003 | - | 0.003 | | 109 | 99.93 |
| | | | 0.003 | - | 0.004 | | 46 | 99.98 |
| | | | 0.004 | - | _ | | 20 | 100.00 |
| T_U | 0.001 | 0.001 | | | | 0 | | |
| | | | | _ | 0.001 | | 51706 | 57.64 |
| | | | 0.001 | - | 0.002 | | 24423 | 84.87 |
| | | | 0.002 | - | 0.002 | | 9341 | 95.28 |
| | | | 0.002 | - | 0.003 | | 3045 | 98.68 |
| | | | 0.003 | - | 0.004 | | 828 | 99.60 |
| | | | 0.004 | - | 0.005 | | 272 | 99.90 |
| | | | 0.005 | - | 0.006 | | 68 | 99.98 |
| | | | 0.006 | - | 0.006 | | 17 | 100.00 |
| | | | 0.006 | - | 0.007 | | 3 | 100.00 |

Figure A.3 – Fragment of the report for the model with $N = 3$

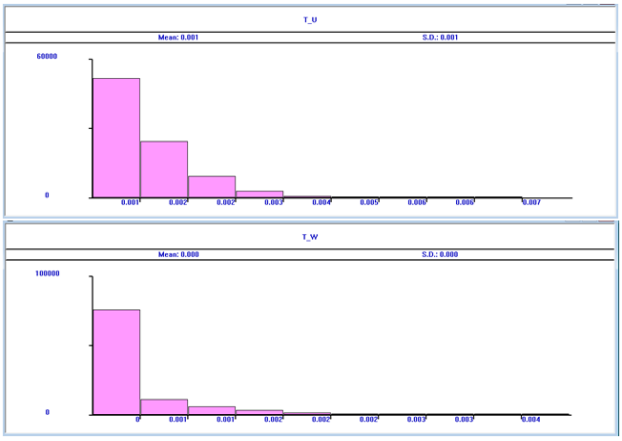



Figure A.4 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 3$

QUEUE

| | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 0 | 99989 | 81309 | 0.200 | 0.000 | 0.000 | 0 |

STORAGE

| | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 4 | 3 | 0 | 4 | 99989 | 1 | 2.056 | 0.514 | 0 | 0 |

TABLE

| | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | | _ | 0.000 | | 94942 | 94.95 |
| | | | 0.000 | - | 0.001 | | 3638 | 98.59 |
| | | | 0.001 | - | 0.001 | | 1052 | 99.64 |
| | | | 0.001 | - | 0.002 | | 258 | 99.90 |
| | | | 0.002 | - | 0.002 | | 80 | 99.98 |
| | | | 0.002 | - | 0.002 | | 16 | 100.00 |
| | | | 0.002 | - | 0.003 | | 3 | 100.00 |
| T_U | 0.001 | 0.001 | | | | 0 | | |
| | | | | _ | 0.001 | | 63515 | 70.46 |
| | | | 0.001 | - | 0.002 | | 19535 | 92.13 |
| | | | 0.002 | - | 0.002 | | 5254 | 97.96 |
| | | | 0.002 | - | 0.003 | | 1343 | 99.45 |
| | | | 0.003 | - | 0.004 | | 366 | 99.85 |
| | | | 0.004 | - | 0.005 | | 106 | 99.97 |
| | | | 0.005 | - | 0.006 | | 25 | 100.00 |
| | | | 0.006 | - | 0.006 | | 1 | 100.00 |
| | | | 0.006 | - | 0.007 | | 1 | 100.00 |

Figure A.5 – Fragment of the report for the model with $N = 4$


DOI 10.15588/1607-3274-2024-2-4

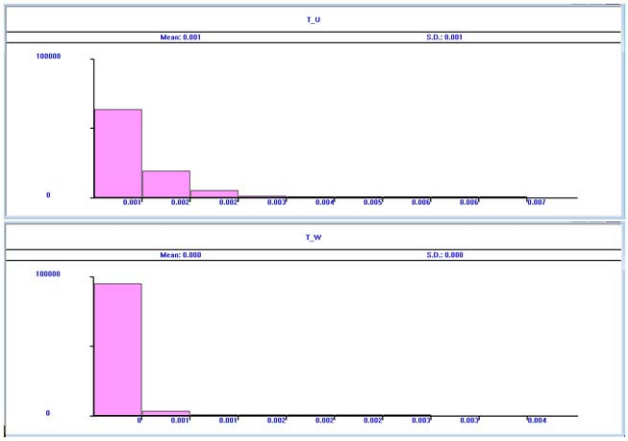



Figure A.6 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 4$

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 9 | 0 | 100002 | 93071 | 0.051 | 0.000 | 0.000 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 5 | 3 | 0 | 5 | 100002 | 1 | 2.070 | 0.414 | 0 | 0 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | | - | 0.000 | | 99001 | 99.00 |
| | | | 0.000 | - | 0.001 | | 817 | 99.82 |
| | | | 0.001 | - | 0.001 | | 167 | 99.98 |
| | | | 0.001 | - | 0.002 | | 15 | 100.00 |
| | | | 0.002 | - | 0.002 | | 2 | 100.00 |
| T_U | 0.001 | 0.001 | | | | 0 | | |
| | | | | - | 0.001 | | 66085 | 73.36 |
| | | | 0.001 | - | 0.002 | | 17898 | 93.22 |
| | | | 0.002 | - | 0.002 | | 4488 | 98.20 |
| | | | 0.002 | - | 0.003 | | 1185 | 99.52 |
| | | | 0.003 | - | 0.004 | | 318 | 99.87 |
| | | | 0.004 | - | 0.005 | | 94 | 99.98 |
| | | | 0.005 | - | 0.006 | | 20 | 100.00 |
| | | | 0.006 | - | 0.006 | | 1 | 100.00 |

Figure A.7 – Fragment of the report for the model with $N = 5$

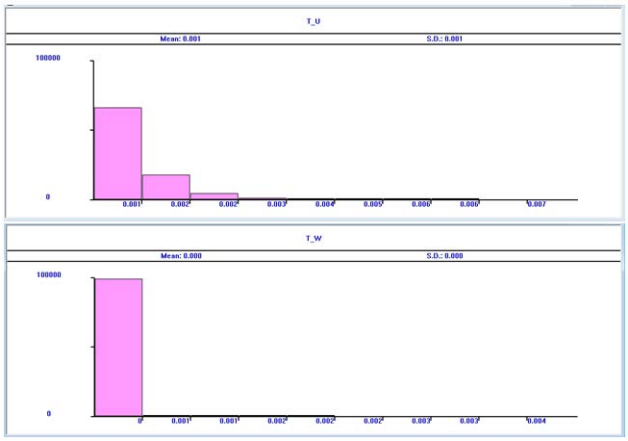



Figure A.8 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 5$

## Appendix B

Results of the second case's simulation model of the ISS's operation.

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 10 | 93922 | 5263 | 5.047 | 0.002 | 0.002 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 2 | 0 | 0 | 2 | 93912 | 1 | 1.932 | 0.966 | 0 | 10 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.002 | 0.001 | | | | 0 | | |
| | | | | - | 0.000 | | 13253 | 14.11 |
| | | | 0.000 | - | 0.001 | | 10024 | 24.79 |
| | | | 0.001 | - | 0.001 | | 11128 | 36.64 |
| | | | 0.001 | - | 0.002 | | 12288 | 49.72 |
| | | | 0.002 | - | 0.002 | | 13158 | 63.73 |
| | | | 0.002 | - | 0.002 | | 13948 | 78.58 |
| | | | 0.002 | - | 0.003 | | 14840 | 94.39 |
| | | | 0.003 | - | 0.003 | | 5273 | 100.00 |
| T_U | 0.002 | 0.001 | | | | 0 | | |
| | | | | - | 0.001 | | 8029 | 9.49 |
| | | | 0.001 | - | 0.002 | | 18103 | 30.90 |
| | | | 0.002 | - | 0.002 | | 22165 | 57.11 |
| | | | 0.002 | - | 0.003 | | 25273 | 86.99 |
| | | | 0.003 | - | 0.004 | | 11004 | 100.00 |

Figure B.1 – Fragment of the report for the model with $N = 2$

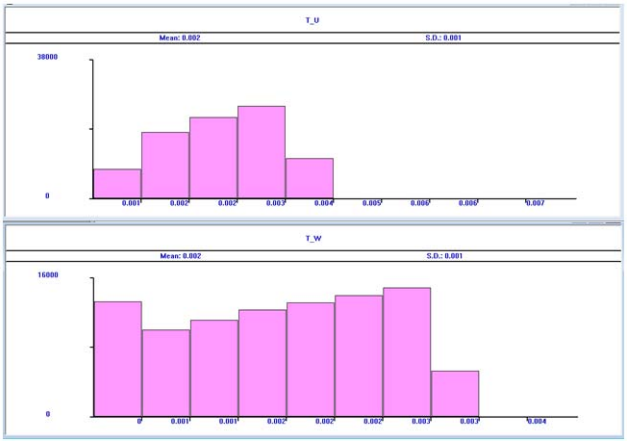



Figure B.2 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 2$

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 0 | 99980 | 54126 | 0.549 | 0.000 | 0.000 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 3 | 2 | 0 | 3 | 99980 | 1 | 2.061 | 0.687 | 0 | 0 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | | - | 0.000 | | 85307 | 85.32 |
| | | | 0.000 | - | 0.001 | | 11147 | 96.47 |
| | | | 0.001 | - | 0.001 | | 2716 | 99.19 |
| | | | 0.001 | - | 0.002 | | 659 | 99.85 |
| | | | 0.002 | - | 0.002 | | 144 | 99.99 |
| | | | 0.002 | - | 0.002 | | 7 | 100.00 |
| T_U | 0.001 | 0.000 | | | | 0 | | |
| | | | | - | 0.001 | | 65090 | 72.35 |
| | | | 0.001 | - | 0.002 | | 23406 | 98.37 |
| | | | 0.002 | - | 0.002 | | 1438 | 99.96 |
| | | | 0.002 | - | 0.003 | | 32 | 100.00 |

Figure B.3 – Fragment of the report for the model with $N = 3$

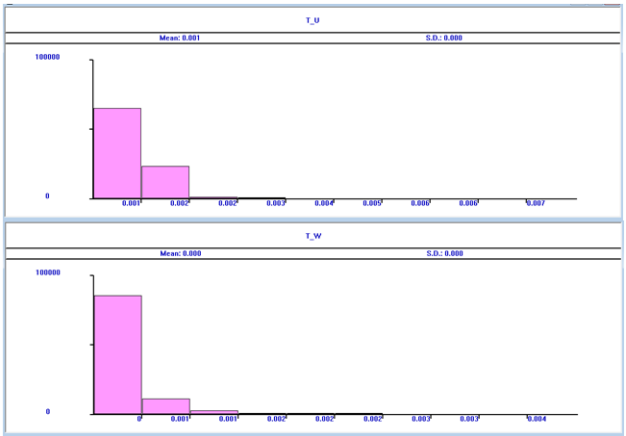



Figure B.4 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 3$


DOI 10.15588/1607-3274-2024-2-4

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 8 | 0 | 100002 | 82358 | 0.112 | 0.000 | 0.000 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 4 | 2 | 0 | 4 | 100002 | 1 | 2.056 | 0.514 | 0 | 0 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | _ | - | 0.000 | | 98663 | 98.66 |
| | | | 0.000 | - | 0.001 | | 1293 | 99.95 |
| | | | 0.001 | - | 0.001 | | 46 | 100.00 |
| T_U | 0.001 | 0.000 | | | | 0 | | |
| | | | _ | - | 0.001 | | 84401 | 93.84 |
| | | | 0.001 | - | 0.002 | | 5542 | 100.00 |
| | | | 0.002 | - | 0.002 | | 3 | 100.00 |

Figure B.5 – Fragment of the report for the model with $N = 4$

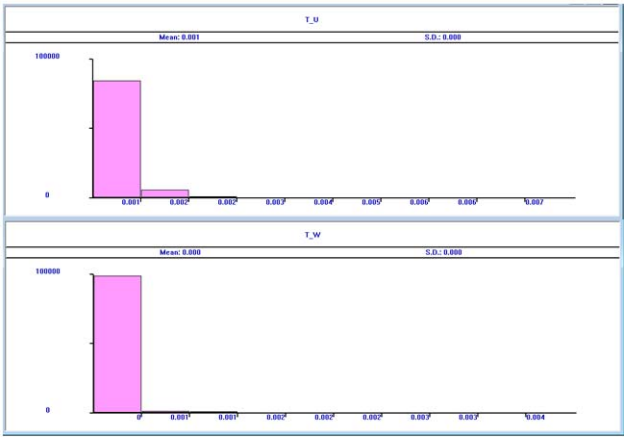



Figure B.6 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 4$

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 7 | 0 | 100002 | 93823 | 0.029 | 0.000 | 0.000 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 5 | 3 | 0 | 5 | 100002 | 1 | 2.053 | 0.411 | 0 | 0 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | _ | - | 0.000 | | 99894 | 99.89 |
| | | | 0.000 | - | 0.001 | | 107 | 100.00 |
| | | | 0.001 | - | 0.001 | | 1 | 100.00 |
| T_U | 0.001 | 0.000 | | | | 0 | | |
| | | | _ | - | 0.001 | | 88876 | 98.71 |
| | | | 0.001 | - | 0.002 | | 1165 | 100.00 |

Figure B.7 – Fragment of the report for the model with $N = 5$

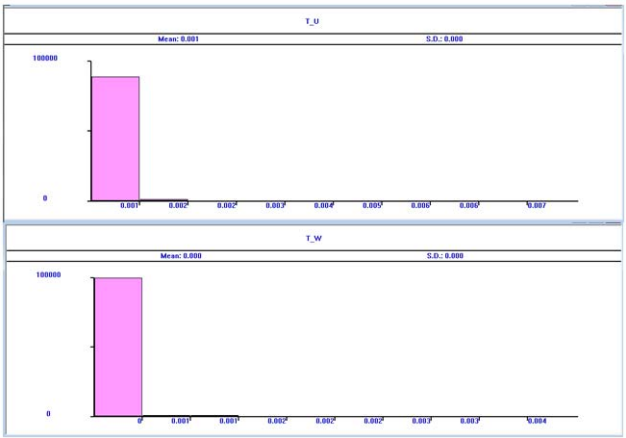



Figure B.8 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 5$

## Appendix C

Results of the third case's simulation model of the ISS's operation.

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 6 | 91907 | 7758 | 4.921 | 0.002 | 0.002 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 2 | 0 | 0 | 2 | 91902 | 1 | 1.898 | 0.949 | 0 | 5 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.002 | 0.001 | | | | 0 | | |
| | | | _ | - | 0.000 | | 16613 | 18.08 |
| | | | 0.000 | - | 0.001 | | 10031 | 28.99 |
| | | | 0.001 | - | 0.001 | | 11184 | 41.16 |
| | | | 0.001 | - | 0.002 | | 11418 | 53.59 |
| | | | 0.002 | - | 0.002 | | 11645 | 66.26 |
| | | | 0.002 | - | 0.002 | | 10743 | 77.95 |
| | | | 0.002 | - | 0.003 | | 8832 | 87.56 |
| | | | 0.003 | - | 0.003 | | 5696 | 93.76 |
| | | | 0.003 | - | 0.004 | | 3276 | 97.32 |
| | | | 0.004 | - | _ | | 2463 | 100.00 |
| T_U | 0.002 | 0.001 | | | | 0 | | |
| | | | _ | - | 0.001 | | 11476 | 13.84 |
| | | | 0.001 | - | 0.002 | | 18300 | 35.92 |
| | | | 0.002 | - | 0.002 | | 20131 | 60.20 |
| | | | 0.002 | - | 0.003 | | 18139 | 82.08 |
| | | | 0.003 | - | 0.004 | | 10430 | 94.66 |
| | | | 0.004 | - | 0.005 | | 3405 | 98.76 |
| | | | 0.005 | - | 0.006 | | 851 | 99.79 |
| | | | 0.006 | - | 0.006 | | 161 | 99.98 |
| | | | 0.006 | - | 0.007 | | 13 | 100.00 |

Figure C.1 – Fragment of the report for the model with $N = 2$

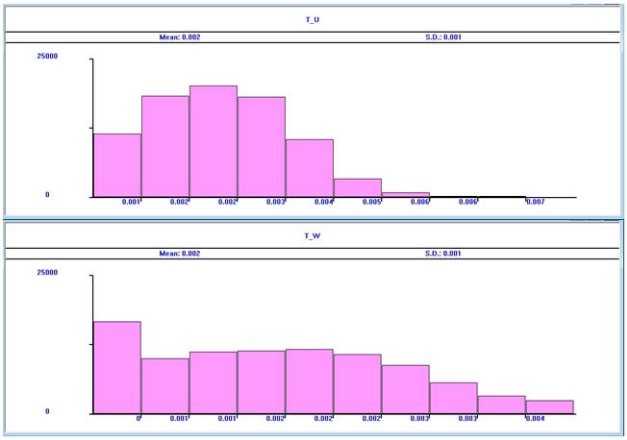



Figure C.2 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 2$

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 0 | 99833 | 53152 | 0.786 | 0.000 | 0.000 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 3 | 1 | 0 | 3 | 99833 | 1 | 2.053 | 0.684 | 0 | 0 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | _ | - | 0.000 | | 78532 | 78.66 |
| | | | 0.000 | - | 0.001 | | 12483 | 91.17 |
| | | | 0.001 | - | 0.001 | | 5268 | 96.44 |
| | | | 0.001 | - | 0.002 | | 2232 | 98.68 |
| | | | 0.002 | - | 0.002 | | 932 | 99.61 |
| | | | 0.002 | - | 0.002 | | 284 | 99.90 |
| | | | 0.002 | - | 0.003 | | 81 | 99.98 |
| | | | 0.003 | - | 0.003 | | 14 | 99.99 |
| | | | 0.003 | - | 0.004 | | 7 | 100.00 |
| T_U | 0.001 | 0.001 | | | | 0 | | |
| | | | _ | - | 0.001 | | 52246 | 58.16 |
| | | | 0.001 | - | 0.002 | | 28786 | 90.21 |
| | | | 0.002 | - | 0.002 | | 7193 | 98.21 |
| | | | 0.002 | - | 0.003 | | 1388 | 99.76 |
| | | | 0.003 | - | 0.004 | | 187 | 99.97 |
| | | | 0.004 | - | 0.005 | | 26 | 100.00 |
| | | | 0.005 | - | 0.006 | | 4 | 100.00 |

Figure C.3 – Fragment of the report for the model with $N = 3$


DOI 10.15588/1607-3274-2024-2-4

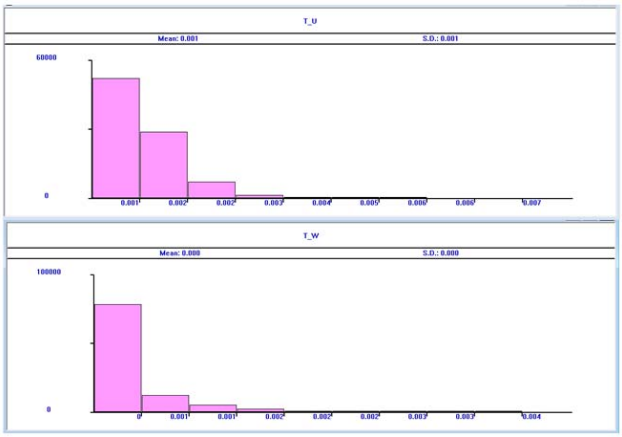



Figure C.4 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 3$

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 3 | 100003 | 81249 | 0.166 | 0.000 | 0.000 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 4 | 0 | 0 | 4 | 100001 | 1 | 2.061 | 0.515 | 0 | 2 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | | _ | 0.000 | | 96173 | 96.17 |
| | | | 0.000 | - | 0.001 | | 3194 | 99.37 |
| | | | 0.001 | - | 0.001 | | 509 | 99.88 |
| | | | 0.001 | - | 0.002 | | 91 | 99.97 |
| | | | 0.002 | - | 0.002 | | 26 | 99.99 |
| | | | 0.002 | - | 0.002 | | 7 | 100.00 |
| T_U | 0.001 | 0.000 | | | | 0 | | |
| | | | | _ | 0.001 | | 64240 | 71.40 |
| | | | 0.001 | - | 0.002 | | 22423 | 96.33 |
| | | | 0.002 | - | 0.002 | | 2962 | 99.62 |
| | | | 0.002 | - | 0.003 | | 308 | 99.96 |
| | | | 0.003 | - | 0.004 | | 32 | 100.00 |
| | | | 0.004 | - | 0.005 | | 2 | 100.00 |

Figure C.5 – Fragment of the report for the model with $N = 4$

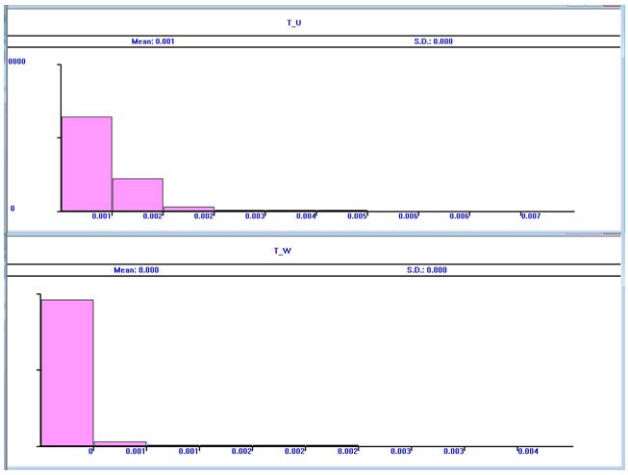



Figure C.6 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 4$

| QUEUE | MAX | CONT. | ENTRY | ENTRY(0) | AVE.CONT. | AVE.TIME | AVE.(-0) | RETRY |
|---|---|---|---|---|---|---|---|---|
| CH_1 | 10 | 0 | 100002 | 93516 | 0.040 | 0.000 | 0.000 | 0 |

| STORAGE | CAP. | REM. | MIN. | MAX. | ENTRIES | AVL. | AVE.C. | UTIL. | RETRY | DELAY |
|---|---|---|---|---|---|---|---|---|---|---|
| UZEL | 5 | 3 | 0 | 5 | 100002 | 1 | 2.058 | 0.412 | 0 | 0 |

| TABLE | MEAN | STD.DEV. | RANGE | | | RETRY | FREQUENCY | CUM.% |
|---|---|---|---|---|---|---|---|---|
| T_W | 0.000 | 0.000 | | | | 0 | | |
| | | | | _ | 0.000 | | 99406 | 99.40 |
| | | | 0.000 | - | 0.001 | | 535 | 99.94 |
| | | | 0.001 | - | 0.001 | | 54 | 99.99 |
| | | | 0.001 | - | 0.002 | | 7 | 100.00 |
| T_U | 0.001 | 0.000 | | | | 0 | | |
| | | | | _ | 0.001 | | 67088 | 74.46 |
| | | | 0.001 | - | 0.002 | | 20334 | 97.03 |
| | | | 0.002 | - | 0.002 | | 2419 | 99.71 |
| | | | 0.002 | - | 0.003 | | 239 | 99.98 |
| | | | 0.003 | - | 0.004 | | 18 | 100.00 |
| | | | 0.004 | - | 0.005 | | 1 | 100.00 |

Figure C.7 – Fragment of the report for the model with $N = 5$

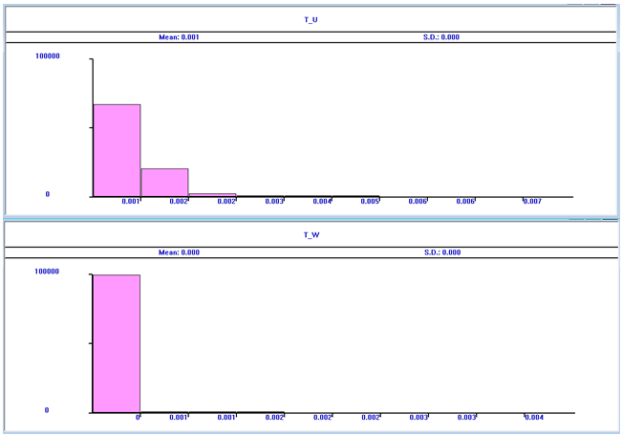



Figure C.8 – Histograms of the probability density functions of the residence time $T\_U$ and waiting time $T\_W$ of requests $N = 5$

## REFERENCES

1. Ismailov B. G. Modelling and analysis of the security system information in service networks, *Problemi informatizatsiĭ ta upravlínnya*, 2022, Vol. 1, № 69, pp. 46–53. DOI: 10.1 837 2/2073-4751.6 9.16812
2. Fan L., Wang Y., Cheng X., Li J., Jin S. Privacy theft malware multi-process collaboration analysis, *Security and Communication Networks*, 2013, No. 8 (1), pp. 51–67. DOI:10.10 02/sec. 705
3. Gordon L. A., Loeb M. P. The Economics of Information Security Investment, *ACM Transactions on Information and System Security*, 2002, No. 5 (4), pp. 438–457. DOI:10.1145/58 127 1.5812 74. S2CID 1 500 788
4. Fienberg S. E., Slavković A. B. Data Privacy and Confidentiality, *International Encyclopedia of Statistical Science*, 2011, pp. 342–345, DOI:10.1007/978-3-642-04898-2_202
5. Pevnev V. Model Threats and Ensure the Integrity of Information, *Systems and Technologies*, 2018, No. 2 (56), pp. 80–95. DOI:10.32836/2521-66 43-2018.2-56.6
6. Loukas G., Oke G. Protection Against Denial of Service Attacks: A Survey, *Comput. J.*, 2012, No. 53 (7), pp. 1020–1037. Archived from the original on March 24, Retrieved August 28, 2015. DOI: 10.1 093/ com jnl/bxp078
7. Fowler Kevvie Developing a Computer Security Incident Response Plan, *Data Breach Preparation and Response, Elsevier*, 2016, pp. 49–77. retrieved June 5, 2021. DOI:10.1016/b978-0-12-803451-4.00003-4
8. Parker D. B. A Guide to Selecting and Implementing Security Controls, *Information Systems Security*, 1994, No. 3 (2), pp. 75–86. DOI:10.1080/10658989 4093 42459


DOI 10.15588/1607-3274-2024-2-4

9. Venter H. S., Eloff J. H. P. A taxonomy for information security technologies, *Computers & Security*, 2003, No.22 (4), pp. 299–307. DOI: 10. 1016/S0167-4048(03)00406-1
10. McDermott B. E.,Geer D. Information security is information risk management, *In Proceedings of the 2001 Workshop on New Security Paradigms NSPW'01*, pp. 97–104. ACM. DOI:10.1 145/ 5081 71. 508187
11. Authorization and approval program, *Internal Controls Policies and Procedures*. Hoboken, NJ, US, John Wiley & Sons, Inc., October 23, 2015, pp 69–72, retrieved June1, 2021. DOI:10.1002/9781119 20 39 64.ch10
12. Almehmadi A., El-Khatib Kh. Authorized! Access denied, unauthorized! Access granted, *Proceedings of the 6th International Conference on Security of Information and Networks*. *Sin '13.US: ACMPress.* New York, 2013, pp. 363–367. DOI:10.1145/2 52 3514.25 23612
13. Joshi Ch., Singh U. K. Information security risks management framework A step towards mitigating security risks in university network, *Journal of Information Security and Applications*. August, 2017, No. 35, pp. 128–137. DOI:10.1016/ j.jisa.2017.06.006
14. Randall A. Harm, risk and threat, Risk and Precaution. Cambridge, Cambridge University Press, 2011, pp. 31–42, retrieved May29, 2021. DOİ:1 0.1017/ cbo97805 1197455 7.0 03
15. Boritz J. E. IS Practitioners' Views on Core Concepts of Information Integrity, *International Journal of Accounting Information Systems*. Elsevier, 2005, No. 6 (4), pp. 260–279. DOI:10.1016/j.accinf. 2005. 07.001
16. Keyser T. Security policy, *The Information Governance Toolkit*. CRC Press, April 19, 2018, pp. 57–62, retrieved May 28, 2021. DOI:10.1 201/978 1315385488-13

Received 09.01.2024.
Accepted 25.04.2024.

УДК 621.394.74:519.872

## АНАЛІЗ РЕЗУЛЬТАТІВ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ СИСТЕМИ БЕЗПЕКИ ІНФОРМАЦІЇ ВІД НЕСАНКЦІОНОВАНОГО ДОСТУПУ У МЕРЕЖАХ ОБСЛУГОВУВАННЯ

**Ісмайлов Б. Г. –** д-р техн. наук, професор кафедри комп'ютерні системи та програмування, Національна Академія Авіації, Баку, Азербайджан.

### АНОТАЦІЯ

**Актуальність.** Аналіз мережі обслуговування показує, що недостатня захищеність інформації в мережах обслуговування є причиною великих збитків, завданих корпораціями. Незважаючи на появу низки робіт, та матеріалів зі стандартизації, в даний час єдина система оцінки захищеності інформації в галузі інформаційної безпеки відсутня. Слід зазначити, що існуючі методи, і навіть накопичений досвід у цій галузі неможливо повністю подолати ці труднощі. Ця обставина **підтверджує**, що дана проблема ще недостатньо вивчена і, отже, зберігає свою актуальність. Представлена робота є одним із кроків на шляху створення єдиної системи оцінки безпеки інформації у мережах обслуговування.

**Мета роботи.** Розробка алгоритму та моделі імітації, аналіз результатів моделі імітації для визначення основних характеристик системи безпеки інформації, що надають можливість повного закриття, за допомогою системи безпеки, всіх можливих каналів прояву загроз, шляхом забезпечення контролю переходу всіх запитів несанкціонованого доступу через механізм захисту.

**Метод.** Для вирішення поставленого завдання було застосовано метод імітаційного моделювання з використанням принципів моделювання системи масового обслуговування. Даний метод дозволяють отримати основні характеристики системи безпеки інформації від несанкціонованого доступу з обмеженим обсягом буферної пам'яті.

**Результати.** Розроблено новий алгоритм, моделі та методологію розробки системи безпеки інформації від несанкціонованого доступу, що розглядається як однофазна багатоканальна системи масового обслуговування з обмеженим обсягом буферної пам'яті. Процес одержання результатів моделі було реалізовано системі моделювання General Purpose Simulation System World і проведено порівняльні оцінки основних характеристик системи безпеки інформації щодо різних законів розподілу вихідних параметрів, тобто. при цьому запити несанкціонованого доступу є найпростішими потоками, а час обслуговування підпорядковується експоненційному, постійному та законам Ерлангового розподілу.

**Висновки.** Проведені експерименти на основі алгоритму та моделі підтвердили очікувані результати при аналізі характеристик системи безпеки інформації від несанкціонованого доступу як однофазної багатоканальної системи масового обслуговування з обмеженим часом очікування запитів у черзі. Ці результати можуть бути використані для практичної побудови нових або модифікації існуючих системи безпеки інформації в мережах обслуговування об'єктів різного призначення. Дана робота є одним з підходів до узагальнення розглянутих проблем для систем з обмеженим обсягом буферної пам'яті. Перспективи подальших досліджень включають в себе дослідження та розробку принципів апаратно-програмної реалізації системи безпеки інформації в мережах обслуговування.

**КЛЮЧОВІ СЛОВА:** несанкціонований доступ, системи безпеки інформації, інформаційна безпека, системи масового обслуговування, механізм захисту, імітаційна моделювання.


DOI 10.15588/1607-3274-2024-2-4

**ЛІТЕРАТУРА**

1. Ismailov B. G. Modelling and analysis of the security system information in service networks / B. G. Ismailov // Problemi ínformatizatsíЇ ta upravlínnya. – 2022. – Vol. 1, №69. – P. 46–53. DOI:10.1 837 2/2073-4751.6 9.16812
2. Privacy theft malware multi-process collaboration analysis / [L. Fan, Y. Wang, X.Cheng et al.] // Security and Communication Networks. – 2013. – No. 8 (1). – P. 51–67. DOI:10.1002/sec.705
3. Gordon L. A. The Economics of Information Security Investment / L. A. Gordon, M. P. Loeb // ACM Transactions on Information and System Security. – 2002. – No. 5 (4). – P.438–457. DOİ:10.1145/58 127 1.5812 74. S2CID 1 500 788
4. Fienberg S. E. Data Privacy and Confidentiality / S. E. Fienberg, A. B. Slavković // Inter national Encyclopedia of Statistical Science, – 2011. – P. 342–345. DOI:10.1007/978-3-642-04898-2_202
5. Pevnev V. Model Threats and Ensure the Integrity of Information / V. Pevnev // Systems and Technologies. – 2018. – No. 2 (56). – P. 80–95. DOİ:10.32836/2521-66 43-2018.2-56.6
6. Loukas G. Protection Against Denial of Service Attacks: A Survey / G. Loukas, G. Oke // Comput. J. – 2012. – No. 53 (7). – P. 1020–1037. Archived from the original on March 24, Retrieved August 28, 2015. DOİ: 10.1 093/ com jnl/bxp078
7. Fowler Kevvie Developing a Computer Security Incident Response Plan / K. Fowler // Data Breach Preparation and Response, Elsevier. – 2016 – P. 49–77, retrieved June 5, 2021.DOİ:10.1016/b978-0-12-80 3451-4.00003-4
8. Parker D. B. A Guide to Selecting and Implementing Security Controls / D. B. Parker // Information Systems Security. – 1994. – No. 3 (2). – P. 75–86. DOI: 10.1080/10 658989409342459
9. Venter H. S. A taxonomy for information security technologies. / H. S. Venter, J. H. P. Eloff // Computers & Security. – 2003. – No. 22 (4). – P. 299–307. DOİ: 10. 1016/S0167-4048(03)00406-1
10. McDermott B. E. Information security is information risk management / B. E. McDermott, & D. Geer // In Proceedings of the 2001 Workshop on New Security Paradigms NSPW'01. – P. 97–104. ACM. DOI:10.1145/508171. 508187
11. Authorization and approval program // Internal Controls Policies and Procedures, Hoboken, NJ, US: John Wiley & Sons, Inc. – October 23, 2015. – P. 69–72, retrieved June1, 2021.DOİ:10.1002/9781119 20 39 64.ch10
12. Almehmadi A. Authorized! Access denied, unauthorized! Access granted / A. Almehmadi, Kh. El-Khatib // Proceedings of the 6th International Conference on Security of Information and Networks. Sin '13.US: ACMPress. New York, 2013. – P. 363–367. DOİ:10.1145/2 52 3514.25 23612
13. Joshi Ch. Information security risks management framework A step towards mitigating security risks in university network / Ch. Joshi, U. K. Singh // Journal of Information Security and Applications. – August 2017. – No. 35. – P. 128–137. DOİ:10.1016/ j.jisa.2017.06.006
14. Randall A. Harm, risk and threat, Risk and Precaution / A. Randall. – Cambridge : Cambridge University Press, 2011. –P.31–42. retrieved May29, 2021. DOİ:1 0.1017/ cbo978051197455 7.0 03
15. Boritz J. E. IS Practitioners' Views on Core Concepts of Information Integrity / J. E. Boritz // International Journal of Accounting Information Systems. Elsevier. – 2005. – No. 6 (4). – P. 260–279. DOİ:10.1016/j.accinf. 2005. 07.001
16. Keyser T. Security policy / T. Keyser // The Information Governance. – Toolkit, CRC Press. – April 19, 2018. – P. 57–62, retrieved May 28, 2021. DOI:10.1 201/978 1315385488-13


DOI 10.15588/1607-3274-2024-2-4

UDC 004.852

# ANALYSIS OF DATA UNCERTAINTIES IN MODELING AND FORECASTING OF ACTUARIAL PROCESSES

**Panibratov R. S.** – Postgraduate student of the Institute for Applied System Analysis, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

## ABSTRACT

**Context.** Analysis of data uncertainties in modeling and forecasting of actuarial processes is very important issue because it allows actuaries to efficiently construct mathematical models and minimize insurance risks considering different situations.

**Objective.** The goal of the following research is to develop an approach that allows for predicting future insurance payments with prior minimization of possible statistical data uncertainty.

**Method.** The proposed method allows for the implementation of algorithms for estimating the parameters of generalized linear models with the preliminary application to data of the optimal Kalman filter. The results demonstrated better forecast results and more adequate model structures. This approach was applied successfully to the simulation procedure of insurance data. For generating insurance dataset the next features of clients were used: age; sex; body mass index (applying normal distribution); number of children (between 0 and 5); smoker status; region (north, east, south, west, center); charges. For creating the last feature normal distribution with known variance and a logarithmic function, exponential distribution with the identity link function and Pareto distribution with a known scale parameter and a negative linear function were used.

**Results.** The proposed approach was implemented in the form of information processing system for solving the problem of predicting insurance payments based on insurance data and with taking into account the noise of the data.

**Conclusions.** The conducted experiments confirmed that the proposed approach allows for more adequate model constructing and accurate forecasting of insurance payments, which is important point in the analysis of actuarial risks. The prospects for further research may include the use of this approach proposed in other fields of insurance related to availability of actuarial risk. A specialized intellectual decision support system should be designed and implemented to solve the problem by using actual insurance data from real world in online mode as well as modern information technologies and intellectual data analysis.

**KEYWORDS:** actuarial risk, generalized linear models, optimal Kalman filter, exponential family of distributions, simulation, iterative-recursive weighted least squares method, Adam method, Monte Carlo for Markov chains.

## ABBREVIATIONS

NPV is a net present value;
MEBN is a multi-entity Bayesian networks;
EVT is an extreme value theory;
GLM is a generalized linear models;
GLMC is a generalized linear models with credibility;
MCMC is a Markov chain Monte Carlo method;
IRWLS is an iterative-recursive weighted least squares.
MSE is a mean squared error.
RMSE is a root mean squared error.
MAE is a mean absolute error,
Adam is an adaptive moment estimation.

## NOMENCLATURE

$A$ is a system dynamic matrix;
$x(n)$ is a vector of states at time step $n > 0$;
$\tilde{x}(n)$ is a vector estimate of states at time step $n > 0$;
$B$ is a matrix of control coefficients;
$u(n)$ is a vector of controls at time step $n > 0$;
$w(n)$ is a noise vector at time step $n > 0$, which has a normal distribution with mean vector with all zero values and covariance matrix $Q$;
$Q$ is a covariance matrix of state disturbances;
$z(n)$ is a vector of measurements of output variables at time step $n > 0$;
$H$ is a matrix of observation coefficients;
$v(n)$ is a vector of random measurement noise values at time step $n > 0$, which has a normal distribution with mean zero vector and covariance matrix $R$;
$R$ is a matrix of measurement errors;
$P(n)$ is a covariance matrix of errors of state vector estimates at time step $n > 0$;
$K(n)$ is a filter's matrix optimum coefficient at time step $n > 0$;
$I$ is an identity matrix;
$a(\bullet), b(\bullet), c(\bullet,\bullet)$ are functions that are defined at the outset in exponential family of distributions;
$\theta$ is a parameter associated with mean values;
$\varphi$ is a scale parameter associated with variance;
$y$ is a target variable for insurance charges and set of financial processes;
$\eta$ is a linear predictor;
$X$ is a matrix of covariates;
$\beta$ is a estimated parameter of GLM;
$g$ is a link function
$E$ is an expected value;
$x_m$ is a scale parameter for Pareto distribution;
$\sigma$ is a standard deviation for normal distribution.

## INTRODUCTION

The existence of factors that prevent possibilities from having deterministic outcomes is implied by uncertainty,


DOI 10.15588/1607-3274-2024-2-5

and it is unknown to what extent these factors may have an impact on the outcomes.

Either a practical or abstract theoretical study of the circumstances for the presence of uncertainty can be carried out, depending on the decision-making perspective that is applied to a particular case. For instance, several mathematical models are employed at the abstract theoretical level, while an evaluation of the quantity of information needed for decision selection is done at the practical level. Selection of these models considers the likelihood of their development in particular scenarios. Information entropy may be used to estimate the quantity of information needed to characterize the uncertainty of the selection scenario.

The uncertainty category is defined by few variable characteristics that characterize many kinds of uncertainties, such as situational, political, social in nature global, and so on. Determining the degree of analysis and the kinds of uncertainties being taken into account is essential to solving the challenges associated with decision-making in the face of uncertainty.

It should be highlighted that uncertainty is frequently limited to the absence of comprehensive knowledge about a particular object. Indeed, uncertainty is not limited to inadequate understanding about object states. In addition, it is occasionally feasible to take into account the ambiguity of the decision-selection criteria and the objectives.

The amount of alternative possibilities and the variety and quantity of criteria used to evaluate these options define the degree of decision-making complexity in many real-world situations.

Since genuine risks and uncertainty are a part of the past, present, and future of analyzed process development, they must be considered in all actions that have an impact on the goals of the organization. Risk and uncertainty are present in all economic activity in varying amounts, but no matter how thorough the risk management, uncertainty cannot be totally removed. Unexpected circumstances and interdependencies might arise at any time. Such unanticipated occurrences may result in deviations that radically alter the data arrangement. As a result, uncertainty can become a risk factor when it results from incomplete information or from using sources that are frequently at odds with the real circumstances of a company or the competitive market [1].

It should be highlighted that a variety of uncertainties, taken together to produce a specific complex of uncertainties known as systemic uncertainty, are frequently present in actual practical situations involving decision-making.

**The object of study** is the process of search the best approach which allows to analyze actuarial risk more efficiently. It is proposed to generate insurance indicators and target variables randomly with adding noise to simulate real-world data, because they are not always publicly available. Therefore, it is proposed to implement approach, which allows to forecast insurance indicators more efficiently by reducing uncertainty.

**The subject of study** are methods for forecasting insurance data.

**The purpose of the work** is to implement approach, which allows to reduce uncertainty during the solving task of forecasting the insurance indicators.

## 1 PROBLEM STATEMENT

For the class of financial processes $\{y(\bullet)\}$ with a generalized form of the probability distribution:

$$f(y,\theta,\varphi)=\exp\{\frac{y\theta-b(\theta)}{a(\varphi)}+c(y,\varphi),$$

where $a(\bullet),b(\bullet),c(\bullet,\bullet)$ are functions, that correspond to a certain distribution law; $y$ is a dependent variable; $\theta$ is a canonical parameter or function of some parameter of a certain distribution; $\varphi$ is the variance parameter. The following distribution laws are allowed: normal, Poisson, binomial, inverse Gaussian, gamma, exponential.

The function $b(\bullet)$ assumes special significance in generalized linear models, because it describes the relationship between the mean value of $\mu_y$ and the variance of the process $\{y\}:\sigma_y^2$. If $\varphi$ is known, then it is an exponential model with the canonical parameter $\theta$. Also, the exponential distribution can be two-parameter, if $\varphi$ is unknown.

It is proposed to determine and minimize the impact of statistical data regarding the dependent variable $\{y\}$, which lead to deterioration of the results of estimation of the structure and parameters of mathematical models and estimates of forecast calculated on the basis of these models. In this case, the following types of models $\{M_i\}, i=1,2,3,...$ are possible: linear regression, variance and covariance analysis, Log-linear models for the analysis of random tables, probit/logit models, Poisson regression.

## 2 REVIEW OF THE LITERATURE

As of right now, there are no widely applicable methods for accounting for the uncertainties of the majority of types that are now in use that can be effectively used to solve the aforementioned real-world issues. Generally speaking, the current techniques for handling uncertainties allow for the consideration of certain particular kinds of uncertainties to enhance the quality of the outcome. Thus, for instance, optimal Kalman filter allows for optimum estimates of the process's state to be obtained against the backdrop of negative random effects by accounting for and minimizing the influence of state disturbances and measurement noise.

The author of [2] provided evidence that determining the measurement uncertainty of every approved analytical test process should be viewed as a valuable completion that adds value rather than as an extra burden. Evaluation


DOI 10.15588/1607-3274-2024-2-5

and comparison of a result with other outcomes are made possible by measurement uncertainty. Communicating the positive meaning of "measurement uncertainty" to clients and the head of authority is crucial. Declaring an excessive amount of uncertainty based just on conjecture is illogical.

Data mining algorithms such as neural networks, evolutionary methods, informed search and space exploration targeted at solving optimization problems, mathematical logic, decision trees, and some others are very helpful in the fighting against uncertainties. With the option to include expert estimates, Bayesian networks – probabilistic models in the form of directed acyclic graphs with the variables of the processes under investigation at their vertices–are an incredibly powerful tool for data analysis.

Multi-Entity Bayesian Networks (MEBN) were presented by the authors in [3]. Given any consistent collection of finitely many first-order phrases, its logic may assign a conditional probability distribution and, conversely, assign probabilities to any set of sentences in first-order logic in a logically coherent manner. That is, MEBN logic can assign a probability to everything that can be represented in first-order reasoning. It is not easy to obtain complete first-order expressive capability in the Bayesian logic. Representing an unlimited or potentially infinite number of random variables is necessary for this, some of which could have an unbounded or potentially infinite number of possible values. Furthermore, we must be able to express random variables with potentially infinite or unbounded parents as well as recursive definitions. More challenges arise from possible random variables that take values in uncountable sets, like the real numbers.

According to MEBN logic, the environment is made up of entities with qualities and relationships to other entities. The features of entities and the connections between them are represented by random variables. MEBN theories are collections of MEBN fragments arranged to represent knowledge about qualities and connections. Given their parents in the fragment network and the context nodes, a MEBN fragment provides the conditional probability distribution for instances of its resident random variables. Any collection of MEBN pieces that together satisfy consistency conditions guaranteeing the presence of a distinct joint probability distribution over instances of the random variables each MEBN fragment in the collection represents is referred to as a MEBN theory [3].

In [4], the assessment of risky investment choices is predicted on techniques that have developed over time to account for both project risk and flexibility. The first steps in project evaluation were calculating the project's net present value (NPV) using the proper discount rate. In recent times, managers have been able to ascertain the proper modifications in project value estimations that represent flexibility, or the chance to respond to unforeseen circumstances and surprises, thanks to the instruments of decision trees and actual alternatives. These techniques offer a sophisticated manner of appraising the value of this flexibility.

The authors of [5] examined the reliability and precision of forecasts in a wide range of topics in the scientific and social sciences. Because they are subject to human biases and limitations, judgmental predictions are no more reliable than statistical ones. As long as forecasts are independent and gathered from a variety of sources, combining them appears to increase accuracy. This is especially true for judgmental forecasts, where averaging of several forecasts typically yields forecasts that are more accurate than the best individual forecasts while simultaneously lowering the variation of predicting errors. On the other hand, both statistical models and subjective forecasters often grossly underestimate uncertainty. The authors outlined two main categories of forecasting scenarios that call for various methodologies and models. Predicting normal conditions in a steady, stable context with known patterns and linkages is referred to as the first. The second occurs in peculiar circumstances with ephemeral, shifting patterns. It should be underlined that booms in business and economic recessions and crises cannot be viewed as anomalies; instead, they need to be forecasted using a different acceptable methodology and adequate model [6].

Explorers face many challenges and issues as a result of the significant variation in prediction accuracy and uncertainty over different time horizons. Additionally, the degree of ambiguity and precision differs across different fields. Normal-condition forecasting errors are thin-tailed, but unusual-condition forecasting errors exhibit radically different behavior, with fat tails. Extreme Value Theory (EVT) has shown to be a useful tool for scientists in estimating uncertainty and producing realistic risk assessments that account for fat tail errors while avoiding the pitfall of average assessments, which drastically underestimate risk and uncertainty. Their results have a lot of promise today and in the future and can be used in different forecasting contexts [5].

## 3 MATERIALS AND METHODS

In particular, the adaptive Kalman filter is a pretty useful tool for assessing and accounting for statistical uncertainty and allows one to assess and anticipate the status of dynamic processes [7] in real time. In this instance, real-time computed estimates of the covariance matrices of the designated random processes are used to adapt the model to the features of always available random disturbances and measurement noise. The capacity to explicitly consider the statistical properties of measurement noises and state disturbances, the ability to calculate optimal estimates of state variables and their forecasts, the possibility to perform effectively data fusion, the ability to estimate unmeasured components of the state vector, and the capacity to estimate states and some model parameters simultaneously are some of the benefits of available optimal filtering procedures.

In state space format, Kalman filters are used to estimate states based on linear or nonlinear dynamical sys-


DOI 10.15588/1607-3274-2024-2-5

tems. The evolution of the state from time $n-1$ to time $n$ is defined by the process model as follows [8]:

$$x(n) = Ax(n-1) + Bu(n-1) + w(n-1).$$

The process model is paired with the measurement model that describes the relationship between the state and the measurement at the current time step $n$ as:

$$z(n) = Hx(n) + v(n).$$

Given the initial estimate of $x(0)$, the series of measurements, $z(1), z(2), ..., z(n)$, and the details of the system model defined by $A, B, H, Q,$ and $R$, the task of the Kalman filter is to generate an optimal estimate of $x(n)$ at time $n$.

In many real-world applications, the true statistics of the noises are either unknown or not Gaussian, despite the fact that the covariance matrices are meant to represent their statistics. As a result, $Q$ and $R$ are typically employed as tuning parameters, which the user can modify to get the intended filter performance.

The covariance matrix of errors of state vector estimates, which is connected to the state estimate, is also used in this technique. It is denoted as $P(n)$.

Algorithm of Kalman filter consists of the next steps.

1. For the state vector and the covariance matrix of estimate errors $P(0)$, set the initial conditions $x(0)$. Assign values to measurement errors $R$ and state disturbances of covariance matrices $Q$.

2. Determine the filter's matrix optimum coefficient as follows:

$$K(n) = \hat{P}(n-1)H^T [HP(n-1)H^T + R]^{-1}.$$

3. To get the state vector's current estimate, use the new measurements:

$$\tilde{x}(n) = A\tilde{x}(n-1) + K(n)[z(k) - HA\tilde{x}(n-1)].$$

4. For updated estimations, compute the posterior covariance matrix of errors:

$$P(n) = [I - K(n)H]\hat{P}(n).$$

5. Determine the a priori covariance matrix of estimate errors (for the subsequent state vector estimation):

$$\hat{P}(n+1) = AP(n)A^T.$$

then proceed to step 2 (the filter equations subsequent calculation).

The authors of [9] introduced KalmanNet, a hybrid system that combines the traditional model-based extended Kalman filter with deep learning techniques. Their method learns to overcome model mismatches and nonlinearities while enabling real-time state estimation in the same way as model-based Kalman filtering.

The drawbacks of low filtering accuracy and the divergence of conventional nonlinear algorithms in situations when the system noise is unknown can be successfully addressed by the suggested technique in [10]. Additionally, the filter's stability and flexibility are enhanced by the suggested algorithm.

The uncertainty of a financial loss that insurers assess using statistical and mathematical techniques is known as actuarial risk. Actuaries assist insurance firms in correctly setting premiums and reserves by analyzing previous data to estimate future risks. Policyholder protection and financial stability are guaranteed by this delicate balance.

With Generalized Linear Models (GLM), assumptions on the characteristics of the insurance data and how they relate to the anticipated variables can be made explicitly. Moreover, GLM offer statistical diagnostics that support the process of identifying just important variables and validating model hypotheses. This methodology is commonly acknowledged as a conventional approach to insurance pricing across many markets and nations.

As a particular instance among the many models that make up the GLM, there is the linear and nonlinear regression model. Rejecting assumptions for the latter include additive nature of effects, constant variance, and a normal distribution. One possible source for the target variable is an exponential family of distributions [11].

The general form of the exponential family of distributions is as follows:

$$f(y, \theta, \varphi) = \exp\{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi\}.$$

Both the variance and the distribution mean may fluctuate. It is expected that explanatory variables have an additive effect on a different scale. For GLM, the following presumptions are made:

1. Stochastic component: every element that makes up y is independent and comes from the same exponential family distribution.

2. Systematic component: the linear predictor $\eta$ is formed of $p$ covariates, or explanatory variables:

$$\eta = X\beta.$$

3. Link function: a differentiable, monotonic link function establishes the linkage between the random and systematic components.

$$E[y] = \mu = g^{-1}(\eta).$$

Forward stepwise regression produced very good results for the identification of risk variables in [12]. When the technique for selecting risk factors was not used prior


DOI 10.15588/1607-3274-2024-2-5

to inclusion in GLM, it has discovered a number of risk variables for both the frequency and severity of claims, improving the predictive performance of the GLM in comparison to the traditional approach.

It was discovered in [13] that iterative algorithm for generalized linear models with credibility (GLMC) works best when combined with exhaustive variable selection techniques. Its computational efficiency and simplicity enable a rapid estimation of model parameters.

The estimate of GLM parameters is a major issue that has to be given enough consideration in the process of model constructing. The following methods were used successfully to evaluate the parameters: Markov chain Monte Carlo method (MCMC), Adaptive moment estimation (Adam) optimization algorithm, and Weighted least squares iterative-recursive approach (IRWLS).

These algorithms are fully described in the following works [14–16].

## 4 EXPERIMENTS

Since insurance data is not always accessible to the general public, it was chosen to create target variables and insurance indicators at random using simulation approach. The data structure consists of the next features:

– Age is a numerical variable, which was generated in range from 18 to 64;

– Sex is a categorical string variable;

– Body mass index is numerical variable, which was generated by using normal distribution;

– Number of children is a numerical variable, which was generated in range from 0 to 5;

– Smoker is a categorical string variable;

– Region is a categorical string variable, which was generated from sample: “east”, “south”, “west”, “center”, “north”;

– Charges is a numerical variable.

The target is the final variable, and the distribution laws and matching link functions listed below were applied to it:

– a normal distribution with a logarithmic link function and a known variance $\sigma$;

– an exponential distribution using the link function of identity;

– Pareto distribution with a link function of the following type $f(x) = -1 - x$ and a given scale parameter, $x_m$.

The predicted variable was supplemented with Gaussian noise with varying variance, which is a linear function.

## 5 RESULTS

After applying Kalman filter original insurance charges were compared with original values.

Results of applying Kalman filter on charges for different distributions is shown on Figures 1, 2, 3.

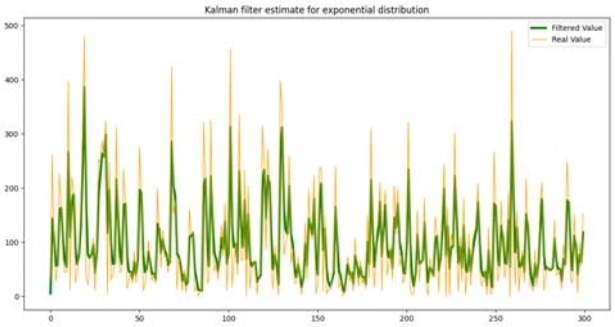



Figure 1 – Results of applying Kalman filter on charges with exponential distribution using the link function of identity

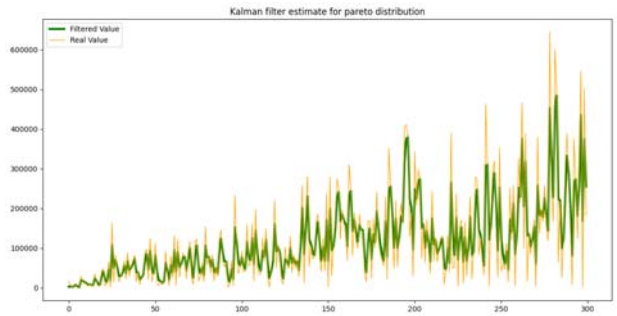



Figure 2 – Results of applying Kalman filter on charges with Pareto distribution with a link function of the negative linear function and a given scale parameter, $x_m$

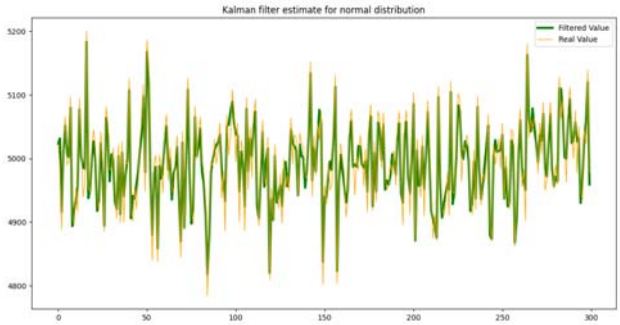



Figure 3 – Results of applying Kalman filter on charges with a normal distribution with a logarithmic link function and a known standard devaition $\sigma$

The models’ quality was assessed by using next forecasting metrics: Mean squared error (MSE), Root mean squared error (RMSE) and mean absolute error (MAE).

Tables 1–6 exhibit the findings of the estimate of GLM parameters using the three approaches (IRWSL, ADAM and MCMC) with and without the Kalman filter for three proposed distribution laws with specialized link functions.

Table 1 – Results of GLM construction for charges with Gaussian distribution with known variance and a logarithmic link function without Kalman filter

| Metric | MCMC | ADAM | IRWLS |
|---|---|---|---|
| MSE | 4408.68 | 684.63 | 2457.11 |
| RMSE | 64.35 | 25.10 | 48.4 |
| MAE | 48.76 | 23.76 | 48.1 |


DOI 10.15588/1607-3274-2024-2-5

Table 2 – Results of GLM construction for charges with Gaussian distribution with known variance and a logarithmic link function with Kalman filter

| Metric | MCMC | ADAM | IRWLS |
|---|---|---|---|
| MSE | 165.33 | 74.5 | 218.68 |
| RMSE | 12.85 | 8.63 | 14.79 |
| MAE | 3.957 | 5.65 | 11.77 |

Table 3 – Results of GLM construction for charges with Pareto distribution with known scale parameter and a negative linear link function without Kalman filter

| Metric | MCMC r | ADAM | IRWLS |
|---|---|---|---|
| MSE | 52724.54 | 88185.03 | 638115.4 |
| RMSE | 228.51 | 286.65 | 797.61 |
| MAE | 153.03 | 205.31 | 771.25 |

Table 4 – Results of GLM construction for charges with Pareto distribution with known scale parameter and a negative linear link function with Kalman filter

| Metric | MCMC r | ADAM | IRWLS |
|---|---|---|---|
| MSE | 7858.747 | 16808.82 | 22900.647 |
| RMSE | 88.65 | 129.65 | 151.33 |
| MAE | 63.513 | 112.12 | 128.532 |

Table 5 – Results of GLM construction for charges with an exponential distribution and a an identity link function without Kalman filter

| Metric | MCMC | ADAM | IRWLS |
|---|---|---|---|
| MSE | 211.38 | 147.9 | 288.51 |
| RMSE | 14.47 | 12.18 | 16.7 |
| MAE | 11.05 | 3.63 | 13.7 |

Table 6 – Results of GLM construction for charges with an exponential distribution and a an identity link function with Kalman filter

| Metric | MCMC | ADAM | IRWLS |
|---|---|---|---|
| MSE | 78.21 | 54.723 | 107.478 |
| RMSE | 8.84 | 7.397 | 10.367 |
| MAE | 4.1 | 2.34 | 7.147 |

## 6 DISCUSSION

Uncertainties for statistical data are factors that have a negative impact on the results of calculations performed at all stages of the data processing process. In this work three approaches were implemented for estimating parameters of GLM with and without the preliminary use of the Kalman filter. It is evident from the GLM building results for the three scenarios mentioned that, for the most part, the Adam technique produced quite decent outcomes. In the case of the Pareto distribution, the MCMC approach also produced positive outcomes. It can be seen that applying Kalman filter for preliminary data processing and fitting model provides for better results of the quality metrics used.

## CONCLUSIONS

The problem of minimizing influence of uncertainties in the process of analysis of actuarial risks regarding forecasting charges is solved in this work.

**The scientific novelty** of obtained results shows that combination of generalized linear models and optimal Kalman filter can be used for building efficient and adequate high quality forecasting models.

**The practical significance** of current work and its results is that future prospects for further research may include the use of this approach in other fields of insurance related to analysis of actuarial risk.

**Prospects for further research** are to study the other approaches that can be used to reduce negative influence of possible data and expert estimates uncertainties related to the analysis of actuarial risks. A specialized decision support system should be designed and implemented to solve the problem.

## ACKNOWLEDGEMENTS

The author of the presented research results wants to appreciate his scientific advisor, Petro I. Bidyuk – Dr. Tech. Sc., Professor at the Department of Mathematical Methods of System Analysis, Institute for Applied Systems Analysis at the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine.

## REFERENCES

1. Toma S. V. Chiriță M., Șarpe D. Risk and uncertainty, *Procedia Economics and Finance*, Vol. 3, pp. 975–980. DOI: https://doi.org/10.1016/S2212-5671(12)00260-2.
2. Meyer V. R. Measurement uncertainty, *Journal of Chromatography A*, 2007, Vol. 1158, № 1–2, pp. 15–24. DOI: https://doi.org/10.1016/j.chroma.2007.02.082.
3. Laskey K. B., da Costa P. C. G., Tolk A., Jain L. C. (eds) Uncertainty Representation and Reasoning in Complex Systems, *Complex Systems in Knowledge based Environments: Theory, Models and Applications*. New York. Springer, 2009, Ch. 2, pp. 7–40. DOI: https://doi.org/10.1007/978-3-540-88075-2_2.
4. Dyer J. S., McDaniel R. R., Driebe D. J. 15 The Fundamental Uncertainty of Business: Real Options, *Uncertainty and Surprise in Complex Systems: : questions on working with the unexpected*. Berlin, Springer-Verlag, 2005, pp. 153–164. DOI: https://doi.org/10.1007/10948637_15.
5. Makridakis S., Bakas N. Forecasting and uncertainty: A survey, *Risk and Decision Analysis*, 2016, Vol. 6, № 1, pp. 37–64. DOI: http://dx.doi.org/10.3233/RDA-150114.
6. Buchanan M. Forecast: what physics, meteorology, and the natural sciences can teach us about economics. USA, Bloomsbury Publishing, 2013, 272 p.
7. Wu X., Kumar V., Ross Quinlan J. et al. Top 10 algorithms in data mining, *Knowledge and Information Systems*, 2008, Vol. 14, pp. 1–37. DOI: https://doi.org/10.1007/s10115-007-0114-2.
8. Urrea C., Agramonte R. Kalman filter: historical overview and review of its use in robotics 60 years after its creation, *Journal of Sensors*, 2021, Vol. 2021, pp. 1–21. DOI: https://doi.org/10.1155/2021/9674015.
9. Revach G., Shlezinger N., Xiaoyong N. et al. KalmanNet: Neural network aided Kalman filtering for partially known dynamics, *IEEE Transactions on Signal Processing*, 2022, Vol. 70, pp. 1532–1547. DOI: https://doi.org/10.48550/arXiv.2107.10043.
10. Xu D., Wang B., Zhang L. et al. A New Adaptive High-Degree Unscented Kalman Filter with Unknown Process Noise, *Electronics*, 2022, Vol. 11, № 12, pp. 1863–1874. DOI: https://doi.org/10.3390/electronics11121863.
11. Anderson D., Feldblum S., Modlin C. et al. A practitioner’s guide to generalized linear models, *Casualty Actuarial Society Discussion Paper Program*, 2004, Vol. 11, Issue 3, pp. 1–116.
12. Omerašević A., Selimović J. Risk factors selection with data mining methods for insurance premium ratemaking, *Zbornik Radova Ekonomski Fakultet u Rijeka*, 2020, Vol. 38, № 2, pp. 667–696. DOI: https://doi.org/10.18045/zbefri.2020.2.667.
13. Campo B. D. C., Antonio K. Insurance pricing with hierarchically structured data an illustration with a workers’ compensation insurance portfolio, *Scandinavian Actuarial Journal*, 2023, Vol. 2023, Issue 9, pp. 853–884. DOI: https://doi.org/10.1080/03461238.2022.2161413.


DOI 10.15588/1607-3274-2024-2-5

14. McCullagh P., Nelder J. Generalized Linear Models. Second edition. London, Chapman & Hall, 1989, 532 p.
15. Akrout M., Tweed D. On a Conjecture Regarding the Adam Optimizer, 2022. [Electronic resource]. Access mode: https://arxiv.org/pdf/2111.08162.pdf.
16. Roy V. MCMC for GLMMs, 2022. [Electronic resource]. Access mode: https://arxiv.org/pdf/2204.01866.pdf.

Received 23.01.2024.
Accepted 20.04.2024.

УДК 004.852

## АНАЛІЗ НЕВИЗНАЧЕНОСТЕЙ ДАНИХ У МОДЕЛЮВАННІ ТА ПРОГНОЗУВАННІ АКТУАРНИХ ПРОЦЕСІВ

**Панібратов Р. С.** – аспірант Інституту прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

### АНОТАЦІЯ

**Актуальність.** Розглянуто задачу аналізу невизначеностей даних у моделюванні та прогнозуванні актуарних процесів. Об'єктом дослідження є задача прогнозування страхових виплат на основі даних про страхових клієнтів з врахуванням можливих ситуацій невизначеності.

**Мета роботи** – розробка підходу, що дозволяє спрогнозувати майбутні страхові виплати з попередньою мінімізацією можливої невизначеності статистичних даних.

**Метод.** Запропоновано метод, що дозволяє реалізувати алгоритми оцінювання параметрів узагальнених лінійних моделей з попереднім використанням оптимального фільтру Калмана. Результати продемонстрували більш якісні результати прогнозу та більш адекватні структури моделі. Даний підхід був успішно застосований на процедурі штучно згенерованих страхових даних. Для генерування страхового набору даних клієнтів були використані наступні показники: вік; стать; індекс маси тіла (використовуючи нормальний закон розподілу); кількість дітей (від 0 до 5); статус курця; регіон (північ, схід, південь, захід, центр); виплати. Для створення останньої величини використовувався нормальний розподіл з відомою дисперсією і логарифмічною функцією зв'язку, експоненційний розподіл з одиничною функцією зв'язку та розподіл Парето з відомим параметром масштабування і від'ємною лінійною функцією зв'язку.

**Результати.** Запропонований підхід реалізований програмно у вигляді системи обробки інформації для розв'язування задачі прогнозування страхових виплат за страховими даними та з урахуванням зашумленості даних.

**Висновки.** Запропонований підхід реалізований програмно для побудування більш адекватних моделей та розв'язування задачі точного прогнозування страхових виплат за страховими даними та врахуванням зашумленості даних. Перспективи подальших досліджень можуть включати використання даного підходу в інших областях застосування, що пов'язані з акутарним ризиком. Необхідно розробити спеціалізовану інтелектуальну систему підтримки прийняття рішень для розв'язування задач з використанням страхових даних реального світу в режимі онлайн, а також сучасних інформаційних технологій та інтелектуального аналізу даних.

**КЛЮЧОВІ СЛОВА:** актуарний ризик, узагальнені лінійні моделі, оптимальний фільтр Калмана, експоненційна множина розподілів, моделювання, ітеративно-рекурентно зважуваний метод найменших квадратів, метод Adam, метод Монте-Карло для марківських ланцюгів.

### ЛІТЕРАТУРА

1. Toma S. V. Risk and uncertainty. / S. V. Toma, M. Chiriță, D. Șarpe // Procedia Economics and Finance. – Vol. 3 – P. 975–980. DOI: https://doi.org/10.1016/S2212-5671(12)00260-2.
2. Meyer V. R. Measurement uncertainty / V. R. Meyer // Journal of Chromatography A. – 2007. – Vol. 1158, № 1–2. – P. 15–24. DOI: https://doi.org/10.1016/j.chroma.2007.02.082.
3. Laskey K. B. Uncertainty Representation and Reasoning in Complex Systems / Laskey K. B., da Costa P. C. G., Tolk A., Jain L. C. (eds) // Complex Systems in Knowledge based Environments: Theory, Models and Applications. – New York : Springer, 2009. – Ch. 2 – P. 7–40. DOI: https://doi.org/10.1007/978-3-540-88075-2_2.
4. Dyer J. S. 15 The Fundamental Uncertainty of Business: Real Options / J. S. Dyer, R. R. McDaniel, D. J. Driebe // Uncertainty and Surprise in Complex Systems: : questions on working with the unexpected. – Berlin : Springer-Verlag, 2005. – P. 153–164. DOI: https://doi.org/10.1007/10948637_15.
5. Makridakis S. Forecasting and uncertainty: A survey / S. Markidakis, N. Bakas // Risk and Decision Analysis. – 2016. – Vol. 6, № 1. – P. 37–64. DOI: http://dx.doi.org/10.3233/RDA-150114.
6. Buchanan M. Forecast: what physics, meteorology, and the natural sciences can teach us about economics. / M. Buchanan – USA : Bloomsbury Publishing, 2013 – 272 p.
7. Top 10 algorithms in data mining / [X. Wu, V. Kumar, J. Ross Quinlan et al.] // Knowledge and Information Systems. – 2008. – Vol. 14. – P. 1–37. DOI: https://doi.org/10.1007/s10115-007-0114-2.
8. Urrea C. Kalman filter: historical overview and review of its use in robotics 60 years after its creation / C. Urrea, R. Agramonte // Journal of Sensors. – 2021. – Vol. 2021. – P. 1–21. DOI: https://doi.org/10.1155/2021/9674015.
9. KalmanNet: Neural network aided Kalman filtering for partially known dynamics / [G. Revach, N. Shlezinger, N. Xiaoyong et al.] // IEEE Transactions on Signal Processing. – 2022. – Vol. 70. – P. 1532–1547. DOI: https://doi.org/10.48550/arXiv.2107.10043.
10. A New Adaptive High-Degree Unscented Kalman Filter with Unknown Process Noise / [D. Xu, B. Wang, L. Zhang et al.] // Electronics. – 2022. – Vol. 11, № 12 – P. 1863–1874. DOI: https://doi.org/10.3390/electronics11121863.
11. A practitioner's guide to generalized linear models / [D. Anderson, S. Feldblum, C. Modlin et al.] // Casualty Actuarial Society Discussion Paper Program. – 2004. – Vol. 11, Issue 3. – P. 1–116.
12. Omerašević A. Risk factors selection with data mining methods for insurance premium ratemaking / A. Omerašević, J. Selimović // Zbornik Radova Ekonomski Fakultet u Rijeka. – 2020. – Vol. 38, № 2. – P. 667–696. DOI: https://doi.org/10.18045/zbefri.2020.2.667.
13. Campo B. D. C. Insurance pricing with hierarchically structured data an illustration with a workers' compensation insurance portfolio / B. D. C. Campo, K. Antonio // Scandinavian Actuarial Journal. – 2023. – Vol. 2023, Issue 9 – P. 853–884. DOI: https://doi.org/10.1080/03461238.2022.2161413.
14. McCullagh P. Generalized Linear Models. Second edition / P. McCullagh, J. Nelder. – London : Chapman & Hall, 1989. – 532 p.
15. Akrout M. On a Conjecture Regarding the Adam Optimizer, 2022. [Electronic resource] / M. Akrout, D. Tweed. – Access mode: https://arxiv.org/pdf/2111.08162.pdf.
16. Roy V. MCMC for GLMMs, 2022. [Electronic resource] / V. Roy. – Access mode: https://arxiv.org/pdf/2204.01866.pdf.


DOI 10.15588/1607-3274-2024-2-5

# НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

# NEUROINFORMATICS AND INTELLIGENT SYSTEMS

UDC 004.93

## CONVOLUTIONAL NEURAL NETWORK SCALING METHODS IN SEMANTIC SEGMENTATION

**Hmyria I. O.** – Post-graduate student of the Department of Software Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

**Kravets N. S.** – PhD, Associate Professor, Associate Professor of the Department of Software Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine.

### ABSTRACT

**Context.** Designing a new architecture is difficult and time-consuming process, that in some cases can be replaced by scaling existing model. In this paper we examine convolutional neural network scaling methods and aiming on the development of the method that allows to scale original network that solves segmentation task into more accurate network.

**Objective.** The goal of the work is to develop a method of scaling a convolutional neural network, that achieve or outperform existing scaling methods, and to verify its effectiveness in solving semantic segmentation task.

**Method.** The proposed asymmetric method combines advantages of other methods and provides same high accuracy network in the result as combined method and even outperform other methods. The method is developed to be appliable for convolutional neural networks which follows encoder-decoder architecture designed to solve semantic segmentation task. The method is enhancing feature extraction potential of the encoder part, meanwhile preserving decoder part of architecture. Because of its asymmetric nature, proposed method more efficient, since it results in smaller increase of parameters amount.

**Results.** The proposed method was implemented on U-net architecture that was applied to solve semantic segmentation task. The evaluation of the method as well as other methods was performed on the semantic dataset. The asymmetric scaling method showed its efficiency outperformed or achieved other scaling methods results, meanwhile it has fewer parameters.

**Conclusions.** Scaling techniques could be beneficial in cases where some extra computational resources are available. The proposed method was evaluated on the solving semantic segmentation task, on which method showed its efficiency. Even though scaling methods improves original network accuracy they highly increase network requirements, which proposed asymmetric method dedicated to decrease. The prospects for further research may include the optimization process and investigation of tradeoff between accuracy gain and resources requirements, as well as a conducting experiment that includes several different architectures.

**KEYWORDS:** convolutional neural network, scaling method, asymmetric scaling, semantic segmentation, encoder-decoder, image.

### ABBREVIATIONS

CNN is a convolutional neural network;

ELU is a Exponential Linear Unit;

ReLU is a Rectified linear unit;

D-Unet is a U-net scaled in depth;

W-Unet is a U-net scaled in width;

R-Unet is a U-net with scaled input image resolution;

WDR-Unet is a U-net scaled in depth, width and with increased image resolution;

AWDR-Unet is a U-net scaled asymmetrically in depth, width and with increased image resolution.

### NOMENCLATURE

$F(X)$is a convolutional neural network;

$X$ is an input space(images);

$Y$ is an output space;

$x$ is an input image;

$H,W,C$ is an input image heigh, width and number of channels respectively;

$R^{H\times W\times C}$ is an three dimensional tensor;

*softmax* is an output layer;

$Dec_i$ is a deconvolutional block, that can contain several layers;

$Conv_i$ is a convolutional block, that contain convolutional and pooling layers;

$W',H',C'$ is a scaled input image heigh, width, channels;

$M_a$ is a a-th scaling method;

$F_o$ is an origin network;

$F_a$ is a CNN scaled with a-th method;

*accuracy*($F(X)$) is a accuracy of the network;

*Params*($F(X)$) is a parameters amount;


DOI 10.15588/1607-3274-2024-2-6

$accur_b$ is a network accuracy for the model scaled with method b;

$\max(accur_b)$ is a maximum of acquired model accuracies;

$e^x$ is an exponential function of x;

$a$ is an hyperparameter;

$p_i$ is a is the predicted probability for the true class;

$\gamma$ is a focusing parameter;

$\alpha$ is a weighing factor;

$\sigma$ is an activation function;

$b_i'$ is the $i$-th adjusted bias;

$X_{scaled}$ is a scaled input space;

$conv_{scaled}$ is a scaled convolutional layer;

$F'$ is the adjusted number of filters;

$W_i'$ is the $i$-th adjusted filter;

$F_{scaled}$ is a scaled amount of layers;

$Conv_s^i$ is a $i$-th scaled convolutional block.

$S$ is a scaling factor for image parameters;

$F_s$ is a scaled network;

$Deconv$ is a deconvolutional part of network;

$X_s$ is a scaled input images;

$P_{origin}$ is an origin amount of parameters;

$P_{scaled}$ is an scaled amount of parameters.

## INTRODUCTION

Thankfully to the rapid development of technologies and the equally rapid growth of computing capabilities of computers and their memory capacity, the wide development and use of approaches based on artificial intelligence and digital image processing became possible.

Many researchers have dedicated their work to developing computer vision and image processing systems to solve complex tasks in life scenarios.

Vision systems are widely used in many aspects of real life. In medicine, computer vision plays an important role in imaging and healthcare applications. In autonomous vehicles – identifying and understanding objects in the environment, helping autonomous vehicles navigate safely. Satellite image analysis for land cover classification, monitoring changes in vegetation, urban areas, and more. In the robotic area such systems help to identify and manipulate objects in their environment.

Often creating a new architecture is not available due to different limitations, but in cases where we have some base model and some extra resources available, we can use a scaling technique.

Our goal in this paper is to examine possible network scaling methods and apply them on the convolutional neural network. Study how different approaches impact network accuracy on solving semantic segmentation tasks. Propose and experimentally verify if there is a reason to scale not the whole network symmetrically, but to scale only part of the network.

**The object of study** is the process of scaling a convolutional neural network for semantic segmentation.

Creating a new convolutional neural network is difficult, iterative, and time-consuming process. The scaling of existing model could be beneficial in cases when we need to achieve better accuracy and don't strictly limit to computational resources. Scaling refers to the practice of increasing the size and complexity of neural networks to improve their performance. It's reasonable because such techniques provide a pathway to building more powerful and expressive models which can solve complex tasks, leverage vast amounts of data, and push the boundaries of performance.

**The subject of study** is the scaling methods for convolutional neural network model.

**The purpose of the work** is to increase the accuracy of the convolutional neural network model by scaling its base architecture.

## 1 PROBLEM STATEMENT

Formally convolutional neural network can be represented as $F(X) = Y$. An input image $x \in X$ is a three-dimensional tensor, $x \in R^{H \times W \times C}$, where dimensions it's image parameters, such as size and channels.

From the architecture perspective model can be represented as:

$$F(X) = soft\max(Dec_i(...(Conv_2(Conv_1(X))))) .$$

Deconvolutional blocks placed in deconvolutional part of the network. Such blocks can consist of different layers, but commonly it consists of convolutional layer, concatenation or skip connection and deconvolutional layer. A convolutional (encode) block is consist of convolutional and pooling layers.

Scaling is the process of increasing the size and complexity of neural network by adjusting number of layers and filters, or other parameters, such as image size (W, H). Let's represent scaling method as M so $M_a(F(X^{W \times H \times C}))) = F_a(X^{W' \times H' \times C'})$, where $F_a$ – model with $i'$ amount of layers, received after applying $M_a$ method.

We have a restriction that $accuracy(F_a(X)) > accuracy(F(X))$, that means that method should positively impact on received accuracy.

Another limitation we have its number of parameters, in this work we put that limitation as $Params(F_a) \le 2 * Params(F_o)$. Modeling in such way computational restrictions.

The task is to find such $M_b$ which results in better network accuracy:


DOI 10.15588/1607-3274-2024-2-6

$$\max(accur_b) > \max(accur_a, ..., accur_{a+n}),$$

and using less or equal number of parameters:

$$Params(F_b) \le Params(F_a).$$

## 2 REVIEW OF THE LITERATURE

Convolutional neural networks allow to solve various types of computer vision tasks, such as recognition [1], classification and segmentations. They have huge potential in real world task solving from simple image processing to complex image search engine [2]. Segmentation of an image is one of the indispensable tasks in computer vision. This task is comparatively more complicated than other vision tasks as it needs low-level spatial information. Basically, image segmentation can be of two types: semantic segmentation [3] and instance segmentation [4]. In this article we work only with semantic segmentation.

Image segmentation problems have been approached using several classic pre-deep learning techniques, such as sparsity-based methods [5], k-means clustering [6], Support vector machines [7], Random forests [8], ect. The situation has changed radically with the growth of computing power and the development of machine learning methods. The number of neural networks designed for segmentation increased notably [9] [10]. Methods based on encoder-decoder architectures have become a popular approach to semantic segmentation, particularly U-net [11] found wide usage in different studies [12–14].

There are different techniques to increase CNN's accuracy, such as data-centric and network-based. When data-centric methods propose operations on data, to benefit in result efficiency, network-based, such as scaling, offers to modify the network. Historically the most common way is scaling in depth. We can scale networks in different ways. We can scale up or down the depth of the network [15] which means increase or decrease the number of layers. Usually, it results in a more accurate but heavyweight network.

Also, we can scale networks in another dimension – width [16]. In this case we are not increasing the number of layers, but we are multiplying the amount of filters on each layer.

Besides the methods mentioned above there is image resolution scaling method [17] [18]. We are increasing the image's height and width, allowing the network to learn more features increasing its accuracy.

All those techniques have their own disadvantages, but mostly researchers must balance between accuracy profit and resources requirements. Increasing network depth or width inevitably leads to an increasing number of parameters and computations required to train the network. Meanwhile increasing image resolution leads to lower batch size since increasing image size increases memory usage.

The problem is to find which scaling method brings more profit with the same or about the same computational requirements.

We are going to examine all three methods to inspect its influence on the U-net and its benefits when applying to solve semantic segmentation problems. The main criteria is the result network accuracy, but we also will compare its training speed.

## 3 MATERIALS AND METHODS

U-net received its name because of its architecture that resembles the letter U. This architecture has an innovative design, containing contracting path, or in other words – encoder, which has the purpose of extracting features from the input image. Following this, an expansive path, or decoder, expanding image to initial size to enable accurate pixel-wise segmentation. In this symmetric architecture information flows seamlessly, preserving spatial information, which is beneficial for accurate segmentation. Base U-net architecture is displayed on Fig. 1.

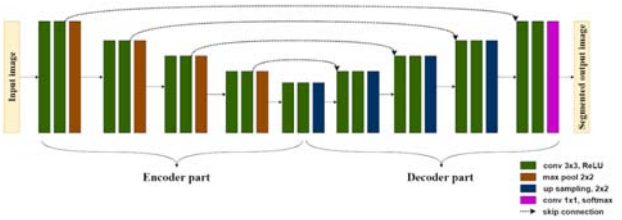



Figure 1 – U-net architecture

The encoder part of the U-Net contains convolutional [19] and pooling layers to systematically reduce the spatial dimensions of the input image. This reduction allows for the extraction of high-level features, creating feature maps. As it moves through consecutive layers, the encoder captures complicated patterns and essential features. During this down sampling process, the U-Net progressively reduces the dimensions of the input image. This down sampling operation involves the use of max-pooling layers, which reduces the information, retaining the most relevant features essential for accurate segmentation.

Unlike the encoder, the decoder section of the U-Net aims to reconstruct the segmented image by expanding the condensed features into the original dimensions. This process is essential to ensure the precise localization of objects within the image and is called up sampling. In this stage, the U-Net uses transposed convolutions, also known as deconvolutions, to reconstruct the segmented image. This method recovers the spatial information lost during the down sampling phase, enabling the network to generate detailed and accurate segmentations.

The basic U-net consists of five blocks in the encoder part and five in decoder. Each encoder block contains consecutive convolutional layers, followed by max-pooling.


DOI 10.15588/1607-3274-2024-2-6

In this article we used U-net as baseline, but made several changes.

In order to reduce the vanishing gradient problem, the activation function was changed from Rectified linear unit (ReLU) to Exponential Linear Unit (ELU). Since ELU mitigates the 'dying ReLU' problem by allowing negative values, which prevents the vanishing gradient issue. And it helps to avoid dead neurons, enhancing the overall robustness of the model during training. ELU activation function could be described as:

$$ELU(x) = f(x) = \begin{cases} x, & x < 0 \\ a(e^x - 1), & x \geq 0. \end{cases}$$

Weight initialization technique was used to receive a more robust and stable learning process. As the technique was chosen He Normalization. The advantage of He Normalization lies in its ability to maintain the stability of gradients, allowing the network to train more effectively. By avoiding the vanishing or exploding gradients, it provides a smoother and more consistent learning process. Consequently, this stability leads to enhanced convergence, enabling the model to reach its optimal state efficiently. As a result, the network requires fewer iterations to reach a desired level of accuracy optimizing the training time.

Also we used another loss function called Focal loss [20], a variation of Binary Cross-Entropy, that serves to lower the impact of simpler to learn instances, thereby encouraging the model to concentrate its learning efforts on more complex examples. This specialized loss function demonstrates decent efficiency in scenarios with significant class imbalances. When some classes appear often and some are rarely seen. The desire to use that loss function was mostly derived from the used dataset, as it's highly imbalanced, so in order to increase focus on other classes this loss function was chosen. Focal Loss proposes to focus on hard training examples, downweighing easy to learn examples, using a modular factor, as shown in formula below:

$$FL = \sum_{i=1}^{i=n} \alpha(i - p_i)^{\gamma} \log(p_i).$$

Here, $\gamma > 0$, but when $\gamma = 1$ this function starts to behave like CrossEntropy loss function. Parameter $\alpha$ usually should be in range [0,1], it can be treated as a hyperparameter, but in our case in order to make this function more data aware modification of inverse class frequency values was used.

In order to receive faster convergence, stability in the learning process, and improved generalization, batch normalization [21] layers were added to the network. They apply a transformation that maintains the mean output close to zero and the output standard deviation close to one, transforming each input in the current mini-batch by subtracting the input mean in the current mini-batch and dividing it by the standard deviation.

Width-wise scaling or expanding Convolutional Neural Networks in width can be beneficial for several reasons. A wider CNN allows the network to capture different features and patterns within the data, improving its accuracy. Besides that, a wider network can better discern finer details in the data due to an enhanced variety of feature maps and activations, which potentially leading to performance improvements, especially in tasks which require specific features extracting. Width-wise scaling refers to adjusting the number of channels increasing number of filters, and can be described as:

$$conv_{scaled}(X) = \sigma(\sum_{i=1}^{F'}(X * W_i' + b_i')).$$

This method can be beneficial, but at the same time wider networks increase computational requirements and need additional memory and processing potential. Besides that, it can lead to overfitting, especially with smaller datasets. As it increases potential for the network to grow overly specialized and less flexible with new or varied data and might compromise its generalization abilities.

The scaling of a CNN depth-wise amplifies its ability to extract complex features and patterns from data, thereby enhancing its representational power. Depth-wise method refers to adjusting number of layers, and can be described as:

$$F_{scaled}(X) = conv_i(..., conv_1(X)),$$

where $i$ – is adjusted amount of layer numbers.

By employing deeper architecture, convolutional neural networks (CNNs) efficiently learn hierarchical features, enabling the network to detect compound patterns and characteristics spanning different tiers, thereby amplifying network effectiveness in complex tasks.

This approach also has its disadvantages, like widthwise scaling, its increasing network's depth leading to enlarging complexity, longer time of training and demanding more computational resources. As networks become deeper it increases probability of overfitting, especially on smaller datasets, as they might memorize patterns instead of generalizing from them, decreasing the model's accuracy performing on new data.

Unlike previous methods which are model-based, image resolution scaling is data-based method. The image resolution refers to the amount of detail that an image holds, typically measured in pixels. It is commonly expressed as the width and height of the image in pixels. The resolution determines the clarity and sharpness of the image, with higher resolutions generally providing more detail. Resolution scaling can be described as:

$$X_{scaled} = resize(X^{W \times H \times C}),$$

where $X_{scaled} \in R^{W*S \times H*S \times C}$, and $S$ – scaling factor.


DOI 10.15588/1607-3274-2024-2-6

Increasing image resolution can potentially increase network accuracy, since higher image accuracy allows models to learn fine-grained patterns. Commonly in order to increase image resolution we would have to use techniques like interpolation. But since our dataset contains high resolution images it allows us just squeeze images to desired size, which is higher than the original one.

To receive all benefits from the previously described methods we applied all those three methods simultaneously on the same baseline network. But instead of raw multiplying of the baseline network's layer and filters, we construct a modified version of the network displayed on Figure 2.

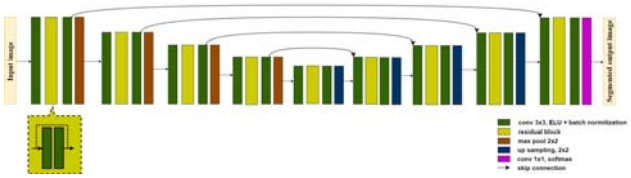


Figure 2 – WDR-Unet architecture

The methods described before applied to whole network, significantly increasing the number of parameters. To use a smaller number of parameters we decided to apply scaling method on U-net architecture asymmetrically, that means applying scaling only on the encoder part of the network. The proposed method can be described as:

$$F_S(X) = Deconv(...,(Conv_S^i(...,Conv_S^1(X_S)))) .$$

Schematic AWDR architecture is displayed on the Figure 3.

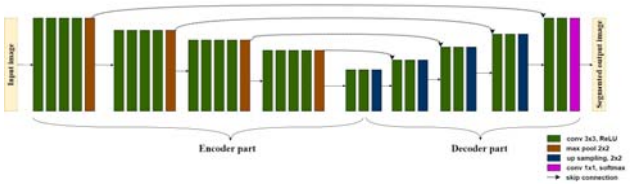



Figure 3 – AWDR-Unet architecture

Encoder part of the network should be scaled using depth-wise and width-wise, besides that input image resolution should be increased. Increasing depth should be performed by adding new convolutional layers in the first half of the network. After that scaling in width should be applied to the encoder part, including new layers.

This concept was inspired by the desire to achieve scaling benefits with lower requirements. Since encoder is responsible for capturing and encoding hierarchical features from the input image, we decided to scale this part. We took a network scaled in all dimensions and modified the decoder part to be the same as in the original baseline. Received network have a smaller number of parameters meanwhile preserves accuracy gain received with combined scaling.

## 4 EXPERIMENTS

For the training and validation processes was used Cityscapes [22] dataset. It's dataset for semantic understanding of urban street scenes. It provides semantic, instance-wise, and dense pixel annotations for 30 classes grouped into 8 categories: vehicles, humans, constructions, flat surfaces, objects, nature, sky, and void. But for this experiment only semantic information was considered. The dataset consists of around 3475 fine annotated images. Data was captured in 50 cities during several months, daytimes, and good weather conditions. Images were thoroughly selected to have the following features: large number of dynamic objects, varying scene layout, and varying background.

Each image has a size of 2048 x 1024 so we performed image resizing using the nearest neighbor algorithm which provides a sharper result image. To decrease overfitting, we performed data processing which includes random cropping and random flipping.

We applied all five methods described in previous section to baseline. All received networks were trained on the same dataset with the same number of epochs (50).

As the result of width-wise scaling we received architecture that almost did not differ from baseline except for the number of filters in each convolutional layer. The origin amount was multiplied by a defined scale factor. That approach increases the width of the whole network symmetrically. Though the last convolutional layer which has classification purpose remained the same, since the number of classes in the dataset wasn't changed. In the experiment we aim to limit scaling in the way that the number of parameters is increased twice, following established limitation. The received scaling factor was 1.4, since it doubled the parameters amount of the original network.

In depth-wise scaling, to receive deeper network, we extended the baseline architecture with additional layers which were copies of existing layers. Each layer was doubled, so depth of the entire network was doubled. The scaling parameter in this case is two, selected in such a way that the number of parameters in received architecture would not break limitation.

Combined scaling method including all three methods follows previous methods principles, except for different scaling parameters, since in the result of the scaling we need to meet limitations. To increase depth of the network we added several layers, but instead of adding a plain sequence of convolutional layers, we decided to use residual blocks, since the model is quite deep and residual blocks facilitate the stable backpropagation of gradients, reducing the possibility of vanishing or exploding gradients.

## 5 RESULTS

In the result of the experiment, we received five different networks, which were received after applying scaling methods to the base model.

For each of the method received networks characteristics are shown in Table 1.


DOI 10.15588/1607-3274-2024-2-6

Table 1 – Scaled networks characteristics

| Network | Accuracy, % | Params, M | Speed, ms/step |
|---|---|---|---|
| AWDR-Unet | 84.8 | 55.2 | 1134 |
| WDR-Unet | 84.7 | 62.8 | 1585 |
| D-Unet | 83.6 | 62.4 | 786 |
| W-Unet | 82.5 | 61.3 | 786 |
| R-Unet | 82.9 | 31.0 | 910 |
| Baseline | 80 | 31.0 | 430 |

As we can see from the table, the best accuracy received AWDR network is 84.8%, approximately the same accuracy as WDR received (84.7%) but has 7.6 M parameters less and faster training speed. Making it a reasonable method of scaling which needs further investigations. In general, methods which combine several methods show better accuracy results (about 1–1.5% better than others), meanwhile have much higher training time.

Increasing model in depth increases the number of parameters by the equation $P_{scaled} = d * P_{origin}$, where $d$ is scaling factor. So, for scaling factor 2 increases number of parameters in twice. For the scaling in width number of parameters can be counted as sum of each layer parameters. The parameters grow is highly depends on the architecture, and in our case increasing each layer's filters number by 1.4 increases overall parameters in twice. Resolution scaling is not increasing parameters amount since it's not affecting architecture. But it has impact on memory usage, we iteratively found that increasing image resolution more than by 1.4 leads to memory overfitting. Using all three methods simultaneously required adjusting scaling factors to meet limitations, so we changed w scaling factor to 1.16, and d to 1.8. Asymmetric method got same scaling factors, but they were applied only to the part of the network.

The computed loss of the five models at each training epoch is shown in Fig. 4.

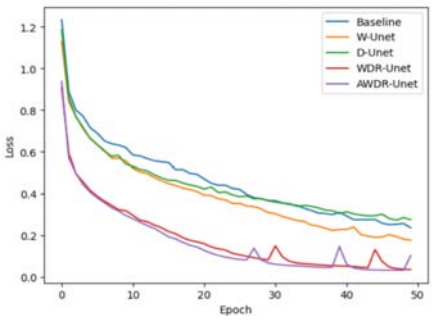



Figure 4 – Training losses of each model

We can observe that WDR and AWDR have lower and approximately equal loss values. From this observation we can confirm that those networks are training better. On the other hand, wider and deeper networks have higher loss values. And decreasing near baseline rate.

On the Fig. 5 we displayed validation loss graphics.

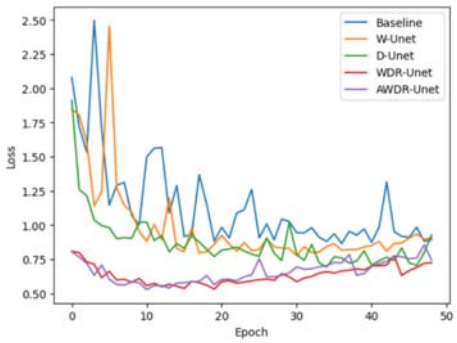



Figure 5 – Validation losses

Here we can see about the same situation as we have with training loss. WDR and AWDR networks have lowest loss values, but here we can see an interesting situation, after the 20-th epoch validation loss value starts to increase. It could be the signal that the model started to learn to perform well on the training data but fails to generalize to new, unseen data. Since those networks are much deeper and complex, they are more vulnerable to overfitting. To fix that in further research we are going to apply more advanced data augmentation techniques to significantly extend the dataset

On Fig. 6 we show accuracy graphic, its display how each model is becoming more accurate with each epoch.

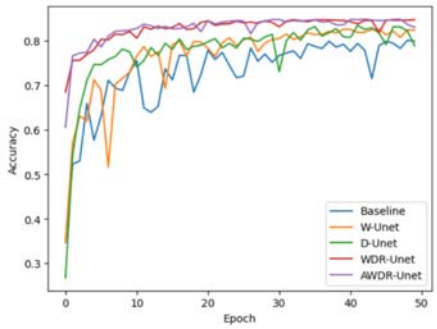



Figure 6 – Accuracy of each model

We can observe from this figure how WDR and AWDR models are converging faster than others, and results in the highest accuracy value. Meanwhile, wider, and deeper networks have second highest values, and have approximately the same values, but deeper networks have slightly better accuracy. Baseline has the lowest accuracy among all networks. Also, it's reasonable to consider using early stopping techniques to stop earlier not perform redundant training and train on lower epochs amount.

## 6 DISCUSSION

In this study we investigated methods of scaling convolutional neural network and its usage in solving semantic segmentation task. The proposed method allows us to combine benefits from other methods, meanwhile using less parameters, which makes network more accurate.

We conducted a review of literature and explored existing approaches in semantic segmentation task, and existing scaling techniques, which can be applied. Existing


DOI 10.15588/1607-3274-2024-2-6

scaling methods includes model-based (depth-wise, width-wise) and data-based (resolution scaling). The scaling in depth and width [23] in general improves accuracy, but received accuracy gain with our network and parameters limit reaches about 3%. Image resolution scaling method also allow to achieve accuracy gain, it can reach from 1% to 13% [24] depending on the architecture and dataset, but for our model and parameters limitation its accuracy gain reached 2.9%. Hybrid scaling [25] produces about 4.7% accuracy improvement. The proposed asymmetric method achieves 4.8% accuracy increase, meanwhile produces network with about 12% less parameters.

After it we conducted an experiment with different scaling methods, including model-wise and data-wise scaling. In a result we acquired five architecture modifications with different characteristics. We empirically verified that scaling a CNN is a beneficial approach in cases where computation capabilities are not strictly limited, and some extra resources are available.

Even though scaling in one dimension potentially can lead to accuracy improvements, it's more efficient to use combined scaling, increasing network architecture in both depth-wise and width-wise ways. Not least important is scaling an input image size. The higher image resolution is the more sophisticated patterns are available for the network to learn. Using proposed method, we received a network architecture that has less parameters while preserving the approximately same high accuracy as existing methods.

The results of the experiment showed that proposed method helps to obtain high-accuracy network, which has high accuracy, but using less parameters than existing methods. However, with benefits from other methods, its also took over the problem of increasing computational requirements and training time, which leaves an open question for further research about tradeoffs between accuracy gain and resources requirements.

## CONCLUSIONS

The paper analyzes scaling method for convolutional neural network designed to solve semantic segmentation task.

**The scientific novelty** of obtained results is the proposed method of asymmetric scaling. This method is appliable to the semantic segmentation models which follows encoder-decoder architecture pattern. It allows to obtain accuracy gain similar to existing methods, but at the same time it uses a smaller number of parameters, that makes it more recourse efficient.

**The practical significance** of obtained results is that the neural network is trained and validated, that allow method to be used in software development. The experimental results allow to recommend the proposed method for use in practice where semantic segmentation task needs to be solved, it can have potential in the safety area, autonomous driving, and traffic systems.

**Prospects for further research** are to study the effectiveness of proposed method on other types of networks with different architectures.


DOI 10.15588/1607-3274-2024-2-6

## ACKNOWLEDGEMENTS

We thank the management of Kharkiv National University of Radioelectronics for the opportunity to conduct scientific research.

## REFERENCES

1. Smelyakov K., Chupryna A., Bohomolov O. et al. The Neural Network Models Effectiveness for Face Detection and Face Recognition, *2021 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Lithuania, 22 April 2021 : proceedings*. Vilnius, IEEE, 2021, pp. 1–7. DOI: 10.1109/estream53087.2021.9431476.
2. Smelyakov K. , Sandrkin D., Ruban I. et al. Search by Image. New Search Engine Service Model, *Problems of Infocommunications. Science and Technology (PIC S&T) : 2018 International Scientific-Practical Conference, Ukraine, 9–12 October 2018 : proceedings*. Kharkiv, IEEE, 2018, pp. 181–186. DOI: 10.1109/infocommst.2018.8632117.
3. Hao S., Zhou Y., Guo Y. A Brief Survey on Semantic Segmentation with Deep Learning, *Neurocomputing*, 2020, Vol. 406, pp. 302–321. DOI: 10.1016/j.neucom.2019.11.118 2020.
4. Hafiz A. M., Bhat G. M. A survey on instance segmentation: state of the art, *International Journal of Multimedia Information Retrieval*, 2020, Vol. 9, No. 3, pp. 171–189. DOI: 10.1007/s13735-020-00195-x.
5. Minaee S., Wang Y. An ADMM Approach to Masked Signal Decomposition Using Subspace Representation, *IEEE Transactions on Image Processing*, 2019, Vol. 28, No. 7, pp. 3192–3204. DOI: 10.1109/tip.2019.2894966.
6. Dhanachandra N., Manglem Khumanthem, Chanu Y. J. Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm, *Procedia Computer Science*, 2015, Vol. 54, pp. 764–771. DOI: 10.1016/j.procs.2015.06.090.
7. Yu Z., Wong H.-S., Wen G. A modified support vector machine and its application to image segmentation, *Image and Vision Computing*, 2011, Vol. 29, No. 1, pp. 29–40. DOI: 10.1016/j.imavis.2010.08.003.
8. Hatami T., Hamghalam M., Reyhani-Galangashi O. et al. A Machine Learning Approach to Brain Tumors Segmentation Using Adaptive Random Forest Algorithm, *2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), Iran, 28 February – 1 March 2019 : proceedings.* Tehran, IEEE, 2019, pp. 76–82. DOI: 10.1109/kbei.2019.8735072.
9. Minaee S. , Boykov Y., Porikli F. et al. Image Segmentation Using Deep Learning: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, P. 1. DOI: 10.1109/tpami.2021.3059968.
10. Ulku I., Akagündüz E. A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images, Applied Artificial Intelligence, 2022, pp. 1–45. DOI: 10.1080/08839514.2022.2032924.
11. Ronneberger O., Philipp F., Thomas B. U-Net: Convolutional Networks for Biomedical Image Segmentation, *Lecture Notes in Computer Science*. Cham, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
12. Huang H., Lin L., Tong R. et al. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation, *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Spain, 4–8 May 2020 : proceedings*. Barcelona, IEEE, 2020, pp. 1055–1059. DOI: 10.1109/icassp40776.2020.9053405.

13. Cao H., Wang Y., Chen J., et al. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation, *Lecture Notes in Computer Science*. Cham, 2023, pp. 205–218. DOI: 10.1007/978-3-031-25066-8_9.
14. Zhang S., Zhang C. Modified U-Net for plant diseased leaf image segmentation, *Computers and Electronics in Agriculture*, 2023, Vol. 204, P. 107511. DOI: 10.1016/j.compag.2022.107511.
15. Kozal J., Wozniak M. Increasing depth of neural networks for life-long learning, *Information Fusion*, 2023, P. 101829. DOI: 10.1016/j.inffus.2023.101829.
16. Yang G., Hu E. Tensor programs IV: Feature learning in infinite-width neural networks, *International Conference on Machine Learning : 38th International Conference, 18–24 July 2021 : proceedings*. San Diego, PMLR, 2021, pp. 11727–11737.
17. Sabottke C. F., Spieler B. M. The Effect of Image Resolution on Deep Learning in Radiography, *Radiology: Artificial Intelligence*, 2020, Vol. 2, No. 1, P. e190015. DOI: 10.1148/ryai.2019190015.
18. Thambawita V., Strümke I., Hicks S. et al. Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images, *Diagnostics*, 2021, Vol. 11, No. 12, P. 2183. DOI: 10.3390/diagnostics11122183.
19. Smelyakov K., Shupyliuk M., Martovytskyi V. et al. Efficiency of image convolution, *Advanced Optoelectronics and Lasers (CAOL) : 8th International Conference, Bulgaria, 6–8 September 2019 : proceedings*. Sozopol, IEEE, 2019, pp. 578–583. DOI: 10.1109/caol46282.2019.9019450.
20. Jadon S. A survey of loss functions for semantic segmentation, *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Chile, 27–29 October 2020 : proceedings.* Viña del Mar, IEEE, 2020, pp. 1–7. DOI: 10.1109/cibcb48159.2020.9277638.
21. Furusho Y., Ikeda K. Theoretical analysis of skip connections and batch normalization from generalization and optimization perspectives, *APSIPA Transactions on Signal and Information Processing*, 2020, Vol. 9. DOI: 10.1017/atsip.2020.7.
22. Cordts M., Omran M., Ramos S. et al. The Cityscapes Dataset for Semantic Urban Scene Understanding, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA, 27–30 June 2016 : proceedings*. Las Vegas, IEEE, 2016, pp. 3213–3223. DOI: 10.1109/cvpr.2016.350.
23. Lingjiao C., Wang H., Zhao J., et al. The Effect of Network Width on the Performance of Large-batch Training, *Advances in Neural Information Processing Systems 31*, 2018.
24. Jerubbaal J., Rajkumar J., Mahesh B. Impact of image size on accuracy and generalization of convolutional neural networks, *IJRAR*, 2019, Vol. 6, No. 1, pp. 70–80.
25. Guocheng Q., Li Y., Peng H. et al. PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies, *Advances in Neural Information Processing Systems 35*, 2022, pp. 23192–23204.

Received 19.02.2024.
Accepted 16.04.2024.

УДК 004.93

## МЕТОДИ МАСШТАБУВАННЯ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ

**Гмиря І. О.** – аспірант кафедри програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

**Кравець Н. С.** – канд. техн. наук, доцент, доцент кафедри програмної інженерії, Харківський національний університет радіоелектроніки, Харків, Україна.

### АНОТАЦІЯ

**Актуальність.** Розробка нової архітектури нейронної мережі є складним і трудомістким процесом, який у деяких випадках може бути замінений масштабуванням існуючої моделі. У цій статті ми розглядаємо методи масштабування згорткової нейронної мережі та прагнемо розробити метод, який дозволяє масштабувати оригінальну мережу, яка вирішує завдання сегментації, у більш точну мережу.

**Мета роботи.** Метою роботи є розробка методу масштабування згорткової нейронної мережі, який досягає або перевершує існуючі методи масштабування, і перевірити його ефективність у вирішенні задачі семантичної сегментації.

**Метод.** Запропонований асиметричний метод поєднує в собі переваги інших методів і забезпечує таку ж високу точність мережі в результаті, як і комбінований метод, і навіть перевершує інші методи. Метод розроблено для застосування до згорткових нейронних мереж, які слідують архітектурі кодера-декодера, призначеної для вирішення завдання семантичної сегментації. Метод посилює потенціал виділення ознак що відбувається в частині кодера, водночас зберігає початкову архітектуру частини декодера. Через свою асиметричність запропонований метод більш ефективний, оскільки призводить до меншого приросту кількості параметрів.

**Результати.** Запропонований метод реалізовано на архітектурі U-net, яка застосовувалася для вирішення задачі семантичної сегментації. Оцінка методу, а також інших методів була виконана на семантичному наборі даних. Метод асиметричного масштабування показав, що його ефективність перевершує або досягає результатів інших методів масштабування, при цьому він є більш ефективний за кількістю параметрів.

**Висновки.** Методи масштабування можуть бути корисними у випадках, коли доступні додаткові обчислювальні ресурси. Запропонований метод був застосований до згорткової нейронної мережі та оцінювався при вирішенні завдання семантичної сегментації, на якому метод показав свою ефективність. Незважаючи на те, що методи масштабування покращують початкову точність мережі, вони значно підвищують вимоги до мережі, для зменшення яких пропонується асиметричний метод. Перспективи подальших досліджень можуть включати процес оптимізації та дослідження оптимального компромісу між підвищенням точності та вимогами до ресурсів, а також проведення експерименту, який включає кілька різних архітектур.

**КЛЮЧОВІ СЛОВА:** згорточна нейронна мережа, метод масштабування, асиметричне масштабування, семантична сегментація, кодер-декодер, зображення.


DOI 10.15588/1607-3274-2024-2-6

**ЛІТЕРАТУРА**

1. The Neural Network Models Effectiveness for Face Detection and Face Recognition / [K. Smelyakov, A. Chupryna, O. Bohomolov et al.] // 2021 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Lithuania, 22 April 2021 : proceedings. – Vilnius : IEEE, 2021. – P. 1–7. DOI: 10.1109/estream53087.2021.9431476.
2. Search by Image. New Search Engine Service Model / [K. Smelyakov, D. Sandrkin, I. Ruban et al.] // Problems of Infocommunications. Science and Technology (PIC S&T) : 2018 International Scientific-Practical Conference, Ukraine, 9–12 October 2018 : proceedings. – Kharkiv : IEEE, 2018. – P. 181–186. DOI: 10.1109/infocommst.2018.8632117.
3. Hao S. A Brief Survey on Semantic Segmentation with Deep Learning / Shijie Hao, Yuan Zhou, Yanrong Guo // Neurocomputing. – 2020. – Vol. 406. – P. 302–321. DOI: 10.1016/j.neucom.2019.11.118 2020.
4. Hafiz A. M. A survey on instance segmentation: state of the art / Abdul Mueed Hafiz, Ghulam Mohiuddin Bhat // International Journal of Multimedia Information Retrieval. – 2020. – Vol. 9, No. 3. – P. 171–189. DOI: 10.1007/s13735-020-00195-x.
5. Minaee S. An ADMM Approach to Masked Signal Decomposition Using Subspace Representation / Shervin Minaee, Yao Wang // IEEE Transactions on Image Processing. – 2019. – Vol. 28, No. 7. – P. 3192–3204. DOI: 10.1109/tip.2019.2894966.
6. Dhanachandra N. Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm / Nameirakpam Dhanachandra, Khumanthem Manglem, Yambem Jina Chanu // Procedia Computer Science. – 2015. – Vol. 54. – P. 764–771. DOI: 10.1016/j.procs.2015.06.090.
7. Yu Z. A modified support vector machine and its application to image segmentation / Zhiwen Yu, Hau-San Wong, Guihua Wen // Image and Vision Computing. – 2011. – Vol. 29, No. 1. – P. 29–40. DOI: 10.1016/j.imavis.2010.08.003.
8. A Machine Learning Approach to Brain Tumors Segmentation Using Adaptive Random Forest Algorithm / [T. Hatami, M. Hamghalam, O. Reyhani-Galangashi et al.] // 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), Iran, 28 February – 1 March 2019 : proceedings. – Tehran : IEEE, 2019. – P. 76–82. DOI: 10.1109/kbei.2019.8735072.
9. Image Segmentation Using Deep Learning: A Survey / [S. Minaee, Y. Boykov, F. Porikli et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2021. – P. 1. DOI: 10.1109/tpami.2021.3059968.
10. Ulku I. A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D Images / Irem Ulku, Erdem Akagündüz // Applied Artificial Intelligence. – 2022. – P. 1–45. DOI: 10.1080/08839514.2022.2032924.
11. Ronneberger O. U-Net: Convolutional Networks for Biomedical Image Segmentation / Olaf Ronneberger, Philipp Fischer, Thomas Brox // Lecture Notes in Computer Science. – Cham, 2015. – P. 234–241. DOI: 10.1007/978-3-319-24574-4_28.
12. UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation / [H. Huang, L. Lin, R. Tong et al.] // ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Spain, 4–8 May 2020 : proceedings. – Barcelona: IEEE, 2020. – P. 1055–1059. DOI: 10.1109/icassp40776.2020.9053405.
13. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation / [H. Cao, Y. Wang, J. Chen et al.] // Lecture Notes in Computer Science. – Cham, 2023. – P. 205–218. DOI: 10.1007/978-3-031-25066-8_9.
14. Zhang S. Modified U-Net for plant diseased leaf image segmentation / Shanwen Zhang, Chuanlei Zhang // Computers and Electronics in Agriculture. – 2023. – Vol. 204. – P. 107511. DOI: 10.1016/j.compag.2022.107511.
15. Kozal J. Increasing depth of neural networks for life-long learning / Jedrzej Kozal, Michal Wozniak // Information Fusion. – 2023. – P. 101829. DOI: 10.1016/j.inffus.2023.101829.
16. Yang G. Tensor programs IV: Feature learning in infinite-width neural networks / G. Yang, E. Hu // International Conference on Machine Learning : 38th International Conference, 18–24 July 2021 : proceedings. – San Diego : PMLR, 2021. – P. 11727–11737.
17. Sabottke C. F. The Effect of Image Resolution on Deep Learning in Radiography / Carl F. Sabottke, Bradley M. Spieler // Radiology: Artificial Intelligence. – 2020. – Vol. 2, No. 1. – P. e190015. DOI: 10.1148/ryai.2019190015.
18. Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images / [V. Thambawita, I. Strümke, S. Hicks et al.] // Diagnostics. – 2021. – Vol. 11, No. 12. – P. 2183. DOI: 10.3390/diagnostics11122183.
19. Efficiency of image convolution / [K. Smelyakov, M. Shupyliuk, V. Martovytskyi et al.] // Advanced Optoelectronics and Lasers (CAOL) : 8th International Conference, Bulgaria, 6–8 September 2019 : proceedings. – Sozopol : IEEE, 2019. – P. 578–583. DOI: 10.1109/caol46282.2019.9019450.
20. Jadon S. A survey of loss functions for semantic segmentation / Shruti Jadon // 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), Chile, 27–29 October 2020 : proceedings. – Viña del Mar : IEEE, 2020. – P. 1–7. DOI: 10.1109/cibcb48159.2020.9277638.
21. Furusho Y. Theoretical analysis of skip connections and batch normalization from generalization and optimization perspectives / Yasutaka Furusho, Kazushi Ikeda // APSIPA Transactions on Signal and Information Processing. – 2020. – Vol. 9. DOI: 10.1017/atsip.2020.7.
22. The Cityscapes Dataset for Semantic Urban Scene Understanding / [M. Cordts, M. Omran, S. Ramos et al.] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA, 27–30 June 2016 : proceedings. – Las Vegas : IEEE, 2016. – P. 3213–3223. DOI: 10.1109/cvpr.2016.350.
23. The Effect of Network Width on the Performance of Large-batch Training / [C. Lingjiao, H. Wang, J. Zhao et al.] // Advances in Neural Information Processing Systems 31. – 2018.
24. Jerubbaal J. Impact of image size on accuracy and generalization of convolutional neural networks / John Jerubbaal, Joseph Rajkumar, Balaji Mahesh // IJRAR. – 2019. – Vol. 6, No. 1. – P. 70–80.
25. PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies / [Q. Guocheng, Y. Li, H. Peng et al.] // Advances in Neural Information Processing Systems 35. – 2022. – P. 23192–23204.


DOI 10.15588/1607-3274-2024-2-6

UDC 004.023, 519.237, 616.1-07

# FUZZY MODEL FOR INTELLECTUALIZING MEDICAL KNOWLEDGE

**Malyar M. M. –** Dr. Sc., Professor, Dean of the Faculty of Mathematics and Digital Technologies, Uzhhorod National University, Uzhhorod, Ukraine.

**Malyar-Gazda N. M. –** PhD, Associate Professor, Doctor anaesthetist, Borsod-Abauj-Zemplen Central Hospital and University Teaching Hospital, Miskolc, Hungary.

**Sharkadi M. M. –** PhD, Associate Professor, Associate Professor of the Department of Cybernetics and Applied Mathematics, Uzhhorod National University, Uzhhorod, Ukraine.

## ABSTRACT

**Context.** The research is devoted to the development of a flexible mathematical apparatus for the intellectualisation of knowledge in the medical field. As a rule, human thinking is based on inaccurate, approximate data, the analysis of which allows us to formulate clear decisions. In cases where there is no exact mathematical model of an object, or the model is difficult to implement, it is advisable to use a fuzzy logic apparatus. The article is aimed at expanding the range of knowledge of researchers working in the field of medical diagnostics.

**Objective.** The aim of the study is to improve the quality of reflection of the subject area of the medical sphere on the basis of building type-2 fuzzy knowledge bases with interval membership functions.

**Method.** The article describes an approach to formalising the knowledge of a medical specialist using second-order fuzzy sets, which allows taking into account the uncertainty and vagueness inherent in medical data and solving the problem of interpreting the results obtained.

**Results.** The developed approach is implemented on a specific problem faced by an anaesthetist when admitting a patient to elective (planned) surgery.

**Conclusions.** Experimental studies have shown that the presented type-2 fuzzy model with interval membership functions allows to adequately reflect the input medical variables of a qualitative nature and take into account both the knowledge of a specialist in medical practice and research medical and biological data. The acquired results hold substantial practical importance for medical practitioners, especially anesthetists, as they lead to enhanced patient assessments, error reduction, and tailored recommendations. This research fosters the advancement of intelligent systems capable of positively influencing clinical practices and improving patient outcomes within the realm of medical diagnostics.

**KEYWORDS:** fuzzy clustering, medical diagnostics, membership functions of the second kind.

## ABBREVIATIONS

ASA is the American Society of Anaesthesiologists;
FOU is a footprint of uncertainty;
FS-1 is a first-order fuzzy set;
FS-2 is a second-order fuzzy set;
LMF is a lower membership function;
MET is a metabolic equivalent;
MF is a membership function;
UMF is an upper membership function.

## NOMENCLATURE

$A$ is a fuzzy set;
$A_{ij}$ is a linguistic term;
$X$ is the set of features;
$x_i$ is an $i$-th element of the set $X$;
$K$ is the set of classes;
$k_j$ is a class label, j-th element of the set $K$;
$F(X)$ is a fuzzy classifier;
$J_{x'}$ is a primary membership function;
$\mu$ is a membership function;
$\mu(x^*)$ is the value of the membership function at a particular point;
$\mu_{\tilde{A}}(x')$ is the secondary membership function;
$n$, $m$ – number of elements in the set;
$P$ is a number of rules;
$R_{pj}$ is a product rule;
$x$ is the primary variable;
$x^*$ is specific value of the indicator;
$u$ is the secondary variable.

## INTRODUCTION

Modern information technologies are widely used in many areas of human activity, in particular, they are being intensively implemented in the daily practice of healthcare. Fuzzy modelling is one of the most popular and rapidly developing areas in the field of modern methods of managing poorly formalized objects.

The massive introduction of digital technologies in the modern world demonstrates successful results, in particular in the diagnosis of socially significant diseases. The use of computer systems by a practitioner requires continuous improvement of both the doctor's knowledge and the systems as a whole. Both subjective and objective factors should be taken into account. Subjective factors include physician errors in examination and diagnosis. Objective factors include, first of all, the constant increase in the number of criteria that must be taken into account and analysed by a doctor when making a medical decision, which significantly exaggerate the cognitive capabilities of a person, as well as the lack of time for decision-making. To solve these problems, it is advisable to introduce intelligent systems based on artificial intelligence technologies into clinical medicine. Such systems are able to provide clinicians with personalized assess-


DOI 10.15588/1607-3274-2024-2-7

ments and/or recommendations to assist them in making medical decisions.

Intelligent decision-making systems should be focused on the knowledge of a specialist doctor and be based on a set of explicit, understandable, easily interpretable linguistic "IF-THEN" rules. As a rule, in the process of assessing a patient's condition, medical practice uses linguistic terms such as "high fever", "old age", "low blood pressure", which are intuitive and well aligned with the doctor's thinking and judgement, and can be used to describe uncertain and inaccurate information. These properties encourage researchers to use fuzzy logic theory in intelligent decision support systems.

When admitting a patient to surgery, the anaesthetist must comprehensively examine the patient, i.e. objectively examine the patient's condition, analyse laboratory, *X*-ray and other types of examinations, and take into account the presence of other diseases of the patient. Processing a large amount of information about a patient is a complex task that an intelligent decision support system can help solve.

**The object of study** is the development of a flexible mathematical apparatus for intellectualizing knowledge in the medical field, with a specific focus on medical diagnostics.

**The subject of study** is the intellectualization of knowledge in the medical field.

**This paper proposes** the use of a mathematical apparatus using second-order fuzzy sets to determine the class of a patient's readiness for surgery.

## 1 PROBLEM STATEMENT

The problems faced by medical professionals are related to the situation when all parameters cannot be taken into account, and the required number of parameters cannot be determined, i.e. there are only significant parameters, while the final solution is considered, which may be approximate.

Such problems include the tasks of diagnosing and predicting the consequences of a disease, choosing a treatment strategy and tactics, determining the level of patient readiness for surgery, etc.

To solve such problems in conditions where the description of medical data has a qualitative form, it will be appropriate to build fuzzy logic systems that have the ability to describe existing statements of medical professionals.

As a rule, the formation of fuzzy knowledge bases, the choice of the type of membership functions and the number of input parameters will significantly depend on the degree of participation of medical specialists and laboratory information.

To solve this problem, it is proposed to use type-2 fuzzy models with interval membership functions[1], which are adequately applied in the medical field and are capable of taking into account both the knowledge of a medical expert and laboratory biomedical data in the context of descriptive, qualitative data.

The problem of medical diagnosis can be formulated as a classification problem. Such a problem can be solved by finding an appropriate classifier, i.e. a mathematical function $F$, that corresponds to a set of symptoms $X = \{x_1, x_2, ..., x_n\} = \{x_i, i = \overline{1,n}\}$ with a certain class label $k_j$:

$$F(X): X \to k_j .$$

It is proposed to use a fuzzy modelling approach based on observation data as a classifier. Such an approach allows for a compromise between classification accuracy and interpretation of the result.

The task of classification is to predict the class of an object by its feature vector.

Let the set of features $X = \{x_1, x_2, ..., x_n\} = \{x_i, i = \overline{1,n}\}$ and the set of classes $K = \{k_1, k_2, ..., k_m\} = \{k_j, j = \overline{1,m}\}$.

A fuzzy classifier is represented as a function that assigns a class label to a point in the input feature space with a calculated degree of confidence:

$$F(X): X_1 \times X_2 \times ... \times X_n \to [0,1]^n .$$

The basis of a fuzzy classifier is a product rule of the following form:

$$R_{pj}: \text{IF } x_1 = A_{1j} \text{ and } x_2 = A_{2j} \text{ and...and } x_n = A_{nj} \text{ THEN } class = k_j .$$

The class is defined by the rule for which the IF-part maximally corresponds to the description given by the input vector *X*:

$$class = k_{j^*},\ j^* = \arg \max_{1 \le j \le m} \prod_{i=1}^{n} \mu_{A_{ij}}(x_i) .$$

The construction of fuzzy classifiers requires solving the following tasks: selection of informative features, formalization of knowledge, formulation of a fuzzy rule base, and optimization of the parameters of the membership function.

The problem of selecting informative features is to find such input attributes from the dataset that most realistically reflect the patient's condition and the doctors' understanding of the result. These can be statistical, theoretical, information and metaheuristic methods.

The formalization of knowledge is a problem whose solution is to build a model that adequately reflects the information of the subject area.

## 2 REVIEW OF THE LITERATURE

Today, computer-based clinical decision support systems for disease detection and patient monitoring are being developed and implemented[2–5], which are able to provide clinicians with personalised assessments and/or


DOI 10.15588/1607-3274-2024-2-7

recommendations to assist in medical decision-making. As a rule, such systems are based on easily interpretable linguistic "If-Then" rules. Such decision-making systems are human-oriented and based on the apparatus of fuzzy set theory [5–13], which is the basis for the development of computer diagnostic systems [14–18].

In the development of computer diagnostic systems, artificial intelligence methods are often used, such as neural networks [19, 20], the nearest neighbours method [21, 22], genetic algorithms [23, 24], support vector machines [25, 26].

## 3 MATERIALS AND METHODS

The theory of fuzzy sets and fuzzy logic systems makes it possible to build decision-making models for tasks that are poorly formalized and operate with expert information[27].

To formalize expert knowledge through fuzzy sets, appropriate procedures for creating membership functions are required. These procedures are a key stage of decision-making, as the quality of the decision depends on the adequacy of the membership function that represents the expert knowledge. The choice of the type of fuzzy set for the construction of membership functions and the corresponding fuzzy model poses the researcher with the task of optimal choice [30].

The issue of constructing membership functions is one of the key issues in fuzzy logic, and many scientists, starting with L. Zadeh, have devoted their research to it. In the theory of fuzzy sets, the membership function is the main characteristic of a fuzzy object, and all operations with fuzzy objects are performed through their membership functions. Defining a membership function is the first and very important step for further work with fuzzy sets. The membership function can be built on the basis of statistical data or with the participation of an expert or a group of experts. Depending on this, we get a frequency or expert interpretation.

Depending on the degree of fuzziness of fuzzy sets, which is taken into account when building a fuzzy model, we distinguish between type 1 fuzzy models, general type 2 models and interval type 2 models [28].

Type 1 fuzzy models, which are based on first-order fuzzy sets, use membership functions with clear values of membership degrees and produce only a point (clear) value at the output.

Lotfi Zadeh proposed a convenient interpretation of type 1 fuzzy sets in solving practical problems:

$$A = \{(x, \mu_A(x)) \mid x \in X, 0 \le \mu_A(x) \le 1\}.$$

Based on first-order fuzzy sets, various models and algorithms have been developed to address uncertainty, such as the model for assessing the effectiveness of investment projects [4]. However, the analysis of these methods and models shows that they often do not provide complete solutions due to insufficiently justified choice of modelling parameters, requiring multiple implementations to select the optimal parameters.

There are general and interval second-order fuzzy sets. General and interval type-2 fuzzy models are based on second-order fuzzy sets.

The concept of fuzzy sets of the second order (fuzzy sets type-2) was given by the founder of fuzzy logic L. Zadeh in 1975. Second-order fuzzy sets were understood as "fuzzy" sets in which the degree of membership is a fuzzy set of the first order.

The main reason for considering fuzzy sets with fuzzy membership functions is the close relationship that exists between the concept of linguistic truth, for example, with such values on the one hand as true, completely true, very true, more or less true, and so on, and fuzzy sets, the degree of membership of which is described by such linguistic terms as low, medium, high, very low, not low and not high, etc. on the other hand.

In [6], second-order fuzzy sets are described using the lower and upper membership functions (MF). Each of these functions can be represented as a first-order fuzzy set. The interval between these two functions is the footprint of uncertainty (FOU) [5], which is the main characteristic of a second-order fuzzy set (SF-2). The footprint of uncertainty describes the blurring of the first-order membership function, which is fully represented by its two limiting functions: lower (LMF) and upper (UMF), which are first-order fuzzy sets (FS-1) (Fig. 1).

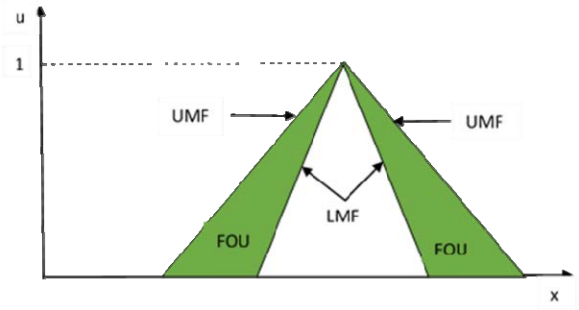



Figure 1 – Type-2 membership function

The introduction of fuzziness in the membership function makes it possible to bring the fuzzy model closer to human thinking and perception. Reality is interpreted differently by different people, and the same word can have different meanings for different individuals, especially when it comes to evaluation statements. Therefore, it is important to avoid a strict correspondence between the values of the degree of membership by expanding the range of uncertainty for each value of the interval. This helps to reduce the risk of errors arising from the lack of consideration of questionable points located near the boundaries of the function.

One of the main tasks is to determine the size of the uncertainty trace, as it affects the accuracy of the model and the time required for computations in computer systems. Obviously, the size of the uncertainty depends on the type of membership function used.


DOI 10.15588/1607-3274-2024-2-7

The membership function of a general second-order fuzzy set is depicted in a three-dimensional model in Figure 2, where the third dimension of the membership function at each point in the two-dimensional domain represents the so-called “footprint of uncertainty” (FOU).

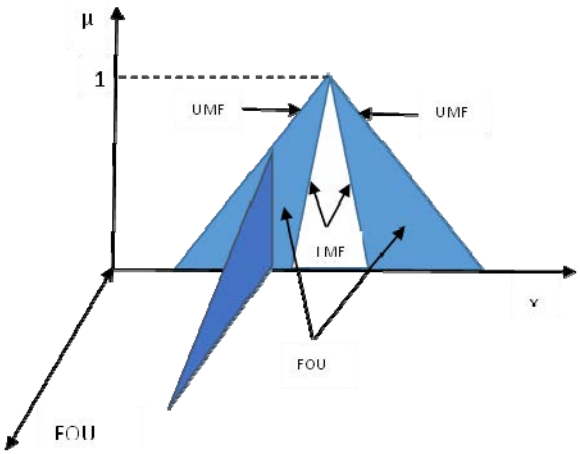



Figure 2 – Uncertainty trace of a type 2 fuzzy set

In theory, you can choose any type of membership function, it is unlimited. However, for type 2 fuzzy sets, the most common are Gaussian, triangular, trapezoidal, and bell-shaped membership functions.

Second-order fuzzy sets are characterised by the blurred boundaries of the membership function (MF) and the way the degrees of membership are distributed to the values of the arguments. Blurring the boundaries is the first step in the transition from type 1 to type 2 fuzzy sets. At the next stage, it is important to choose the type of membership function, just like for type 1 fuzzy sets.

There are two types of FS-2. If for any value of the argument from the universe over the entire interval, from the lower degree of membership to the upper, the value of FS-2 is unchanged, then this type of FS-2 is unified (homogeneous). A fuzzy set with such a type of FS-2 is called an interval fuzzy set of the second order (IFS-2).

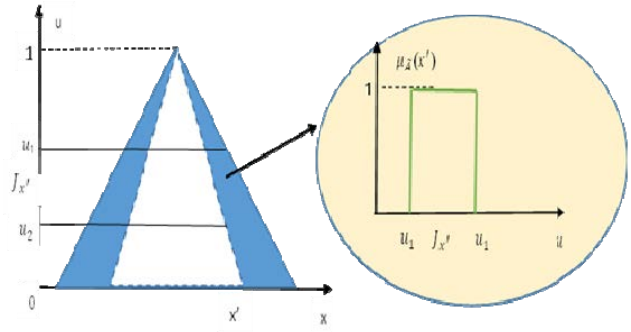



Figure 3 – Interval fuzzy set of the second order

Interval fuzzy models of type 2 use membership functions built on the basis of fuzzy sets with interval values of membership degrees (Fig. 3).

These models, unlike type 1 fuzzy models, provide point and interval values at the output. They cope well with various types of uncertainties and require significantly less computational resources than general type 2 fuzzy models. For example, studies [6–8] provide examples of the use of interval membership functions to solve practical problems.

If, for any value of the argument from the universe in the specified interval, the value of the FN-2 changes, then a fuzzy set with this type of FN-2 is called a second-order fuzzy set of general form.

The second-order membership function in the general (heterogeneous) form can be defined by:

– type 1 characteristics (primary variable and membership function);

– type 2 characteristics (secondary variable and membership function).

Type 2 characteristics define the parameters of the vertical section of the second-order membership function (Fig. 4).

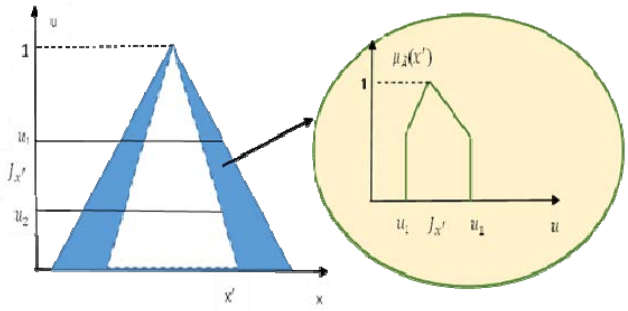



Figure 4 – General second-order fuzzy set

Fig. 4 shows the main characteristics of the second-order membership function in general (heterogeneous) form.

The heterogeneous type of second-order membership function is not used very often due to the high cost of computation, although it has a large number of degrees of freedom. Therefore, expert systems are mostly based on the interval type of second-order fuzzy sets. They allow you to use all the features of interval computing and have a wide range of practical applications.

It is also important to distinguish between different types of uncertainty when constructing membership functions in fuzzy systems, such as intra-uncertainty and inter-uncertainty. Intra-uncertainty arises due to insufficient knowledge or fuzzy expert judgement. Inter-expert uncertainty results from different estimates by several experts. Intra-uncertainty can be described by a second-order fuzzy set, and inter-uncertainty by combining several FS-2 [7].

## 4 EXPERIMENTS

In most practical tasks of medical diagnostics, the synthesis of fuzzy knowledge bases and the construction of fuzzy models in the conditions of qualitative data depend on the ability of a medical specialist to formalise his knowledge and understand the importance of the parameters provided to the developer for the further design of a fuzzy logic system.


DOI 10.15588/1607-3274-2024-2-7

The construction of type-2 fuzzy knowledge bases with interval membership functions for solving the problem of medical diagnosis and prognosis of disease states consists of two main stages. The first stage is designed to generate a fuzzy model based on sample *X*, which is actually verified data from medical practice of the results of a patient's disease examination.

Let's demonstrate the use of this approach when making a decision by an anaesthetist. Here is an example of signs that are used to assess the patient's physical condition during elective (planned) surgical interventions:

1) Physical status. In the provision of anaesthetic care, the classification proposed by the American Society of Anaesthesiologists (ASA) is most often used for assessment [29]. The physical status of patients according to the ASA classification is an assessment of the patient's condition before surgery, endoscopy or other manipulation. There are 5 classes of physical status: from a healthy patient to a patient in an extremely serious condition.

ASA I – Healthy patient (non-smoker, low alcohol drinker).

ASA II – Patient with mild systemic disease (mild disease only without significant functional limitations). Examples include, but are not limited to: smoker, social drinker, pregnant, obese (<30 BMI <40), compensated diabetes mellitus, controlled hypertension, mild respiratory disease.

ASA III – Patient with a severe systemic disease (significant limitations of functional activity). Examples include (but are not limited to): poorly controlled hypertension or subcompensated diabetes mellitus, COPD, morbid obesity (BMI ≥40), active hepatitis, alcohol dependence or abuse, implanted pacemaker, moderate reduction in cardiac output, chronic renal failure requiring regular scheduled haemodialysis. A history (more than 3 months) of myocardial infarction, stroke, transient ischaemic attack, coronary heart disease or stenting.

ASA IV – A patient with a severe systemic disease that poses a permanent threat to life. Examples include, but are not limited to: myocardial infarction, stroke, transient ischaemic attack, coronary artery disease or stenting, ongoing myocardial ischaemia or severe heart valve dysfunction, severe ejection fraction reduction, sepsis, DIC, acute or chronic renal failure, with irregular haemodialysis.

ASA V – Dying patient. Surgery for vital indications. Examples include (but are not limited to): ruptured aortic aneurysm, severe polytrauma, intracranial haemorrhage, acute intestinal ischaemia with concomitant severe cardiac disease or multiple organ failure.

ASA VI – Brain death has been established, organs are removed for donor purposes.

The addition of the letter "E" indicates the urgency of surgical intervention. An emergency situation is defined as existing when a delay in treating the patient will lead to a significant increase in the threat to life. For example: ASA I E, II E, III E or IV E. ASA V class is usually always ASA V E. There is no ASA VI E class. The Borg Scale is used to assess a patient's exercise tolerance.

2) Metabolic equivalent. A metabolic equivalent (MET) is a unit of measurement of the body's energy needs that is used during a treadmill test to assess a person's functional abilities (i.e., their tolerance to physical activity).

The baseline value (1 MET) is the metabolic rate at complete rest (under basal metabolic conditions), which is 1 kcal/kg/h. Thus, 2 MET corresponds to a load that causes a 2-fold increase in the body's energy requirement compared to the resting state. Activity requiring energy expenditure of 2–4 MET (slow walk, doing routine housework, etc.) is considered light, while running and climbing uphill can be accompanied by an increase in energy demand of up to 10 or more MET.

The functional capacity of a person unable to perform a load of more than 5 MET during a treadmill test is considered to be reduced, which is associated with a more severe prognosis, while the ability to perform a load above this level indicates a favourable outcome.

The Borg Scale is used to assess the perceived exertion of physical activity, i.e. how much a person feels their body is working. Perceived exertion is based on the physical sensations that a person experiences during physical activity, including an increase in heart rate, increased respiratory rate, increased sweating, and muscle fatigue. In its original form, the scale started at 6 points and went up to 20, which corresponded to a heart rate of 60 to 200. In recent years, an updated scale from 1 to 10b, the so-called Newman scale, has been used.

3) Laboratory parameters:

Complete blood count:

Haemoglobin 120 – 160 g/l (90–180),

Red blood cells 4.1–5.2 T/l (2.5–7.0)

Leukocytes 4.4–11.3 G/l (2–20.0)

Platelets 140– 400 G/l (75–600)

Kidneys and their functioning:

Creatinine 38–85 μmol/l (30–200)

GFR (glomerular filtration rate) – more than 90 ml/min/1.73 m2 (less than 25)

Homeostasis:

Potassium 3.5–5.1 mmol/l (3–6)

Sodium 137–150 mmol/l (130–155)

Sugar 3.7–6.2 mmol/l (3.2–9.0)

4) Echocardioscopy, if available (LVEF%, normal 55–75%, poor if less than 55%).

5) Coagulogram (blood coagulation).

INR 0.80–1.20 (0.6–1.4).

Fibrinogen 1.8–3.5 g/l (1.2–5.0).

In addition to the above permissible values of indicators for signs of 3–5 groups, there are normative values, which, like personal indicators of patient's symptoms, depend on gender, age, clinical, biochemical and other parameters of comorbidities, which complicates decision-making depending on different stages of the life cycle. Such a relationship of parameters and symptom indicators can be set not only analytically, but also algorithmically by building various models of functioning systems.


DOI 10.15588/1607-3274-2024-2-7

Information on the first and second signs is obtained by interviewing the patient. The scores should correspond to the following levels:

– ASA (class I, II, if class III, the indicators will need to be assessed more carefully);

– MET (4 or more is ideal, less than 4 should be considered whether to allow elective intervention, if so, to set a higher risk of anaesthesia).

The analysis of the information regarding the indicators of groups 3–5 showed that it is advisable to use the apparatus of fuzzy set theory to interpret them, since, as you can see, the values of these indicators are blurred. Among the piecewise linear functions, the most commonly used are the so-called "triangular" and "trapezoidal" membership functions. The first type of function is usually used to formalise uncertainties such as "approximately equal" and "average", and the second type is used to represent uncertainties such as "located in an interval".

To describe the values of these indicators, we will use second-order trapezoidal membership functions (Fig. 5):

$$\mu = \begin{cases} 0, & x \le a_1; \\ \dfrac{x-a_1}{b_1-a_1}, & a_1 \le x \le a_2 \ (UMF); \\ \dfrac{x-a_2}{b_1-a_2}, & a_2 \le x \le b_1 \ (LMF); \\ 1, & b_1 \le x \le b_2 \ (Norm); \\ \dfrac{c_1-x}{c_1-b_2}, & b_2 \le x \le c_1 \ (LMF); \\ \dfrac{c_2-x}{c_2-b_2}, & c_1 \le x \le c_2 \ (UMF); \\ 0, & x \ge c_2. \end{cases}$$

## 5 RESULTS

As an example, consider the application of this approach in determining the degree of fuzziness for the indicator "sugar" of group 3. Homeostasis using second-order fuzzy sets.

For example, a normal blood sugar level in men aged two to 60 years is considered to be in the range of 74–106 mg/dl (4.1–5.9 mmol/l). According to the permissible values, this indicator "sugar" can be in the range of 3.7–6.2 mmol/l (3.2–9.0 mmol/l). Let's describe the value of this indicator using a second-order trapezoidal membership function:

$$\mu = \begin{cases} 0, & x \le 3.2; \\ \dfrac{x-3{,}2}{0{,}9}, & 3.2 \le x \le 3.7; \\ \dfrac{x-3.7}{0.4}, & 3.7 \le x \le 4.1; \\ 1, & 4.1 \le x \le 5.9; \\ \dfrac{6.2-x}{0.3}, & 5.9 \le x \le 6.2; \\ \dfrac{9-x}{3.1}, & 6.2 \le x \le 9.0; \\ 0, & x \ge 9.0. \end{cases}$$

The result of the graphical representation is shown in Figure 6.

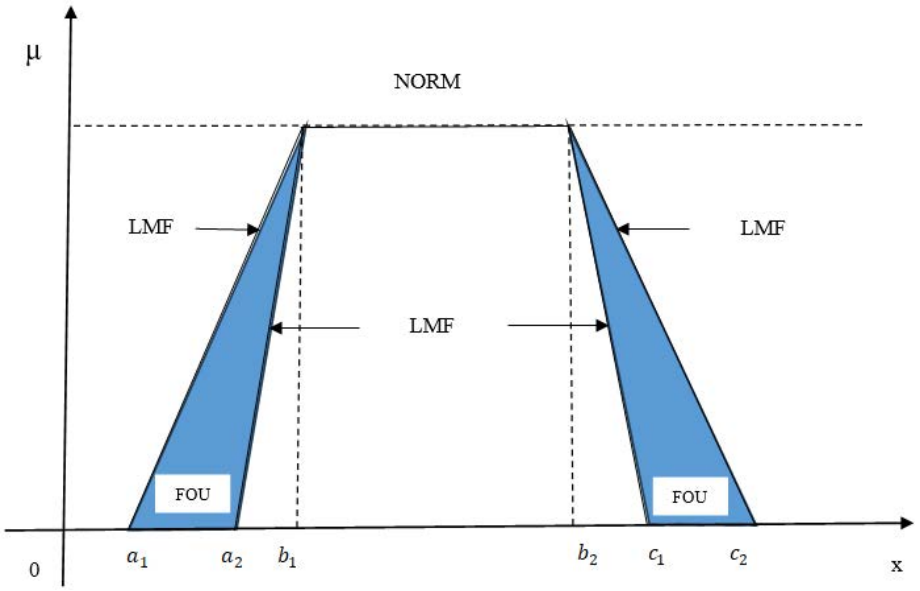



Figure 5 – General interval fuzzy set of type-2


DOI 10.15588/1607-3274-2024-2-7

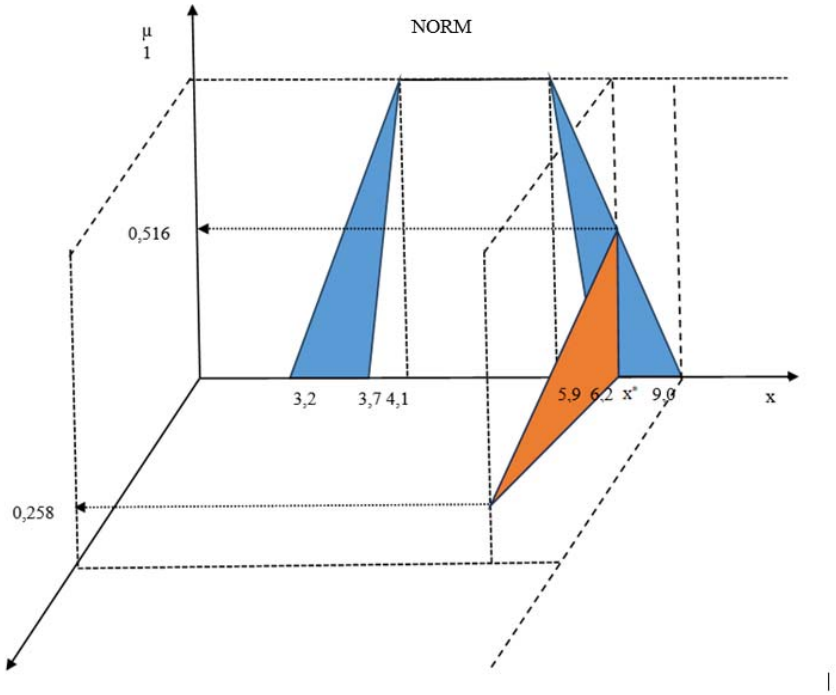



Figure 6 – Interval fuzzy set of type-2 for the indicator “sugar”

Let the value of the indicator “sugar”, for a particular patient, is $x^* = 7.4$ mmol/l, then the value of the primary membership function of the second order $\mu^1(x^*) \approx 0.516$, and, accordingly, the value of the secondary membership function (triangular) $\mu^2(x^*) \approx 0.258$ (Fig. 6).

Similarly, for this indicator, it is possible to build interval fuzzy sets taking into account the age and gender of the patient.

## 6 DISCUSSION

The construction of second-order interval fuzzy sets for indicators of 3–5 groups will allow to formalise knowledge and generate, on the basis of research medical data, an adequate fuzzy knowledge base model for an intelligent decision support system, which will allow doctors to access not only a clinical decision adapted to a particular patient, but also a set of clinical rules on the basis of which this decision was obtained.

Experimental studies have shown the convenience and effectiveness of the proposed approach of using fuzzy interval sets of the second order to intellectualise the knowledge of an anaesthetist.

## CONCLUSIONS

The knowledge obtained from experts usually contains various types of uncertainties, and the knowledge resulting from the processing of experimental data contains a significant amount of noise.

Therefore, the methods of fuzzy set theory are the most suitable for processing incomplete and contradictory information.

For a qualitative representation of the subject area under conditions of uncertainty, it is advisable to use fuzzy interval sets of the second order.

In conclusion, it should be noted that the construction of IF-THEN rules based on biomedical data will be considered in the following publications, which will be devoted to the development of fuzzy logic systems.

**The scientific novelty** of obtained results lies in the pioneering application of type-2 fuzzy sets with interval membership functions to address uncertainty in medical data, particularly in the domain of anesthesiology, while integrating expert knowledge and biomedical data, thereby advancing the capabilities of intelligent medical decision support systems.

**The practical significance** of obtained results lies in the tangible benefits for medical professionals, particularly anesthetists, in improving the quality of patient assessments, reducing errors, and providing personalized recommendations. The research contributes to the development of intelligent systems that can positively impact clinical practice and patient outcomes in the field of medical diagnostics.

**Prospects for further research** involve refining the proposed type-2 fuzzy model, exploring its application across diverse healthcare domains, integrating advanced technologies for real-time decision support, collaborating with medical practitioners to enhance practical utility, and investigating ethical implications, aiming to advance the field of intelligent medical decision support systems.


DOI 10.15588/1607-3274-2024-2-7

## ACKNOWLEDGEMENTS

The work was performed within the framework of the state budget research topic of Uzhhorod National University "Methods of computational intelligence for data processing and analysis" (state registration number 0121U109279).

## REFERENCES

1. Mendel J. M., John R. I., Liang Q. Interval Type-2 fuzzy logic systems: theory and design, *IEEE Transactions on Fuzzy Systems*, 2000, No. 8, pp. 535–550. http://dx.doi.org/10.1109/91.873577
2. Fernandes M., Vieira S. M., Leite F. et al. Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: a Review, *Artif. Intell. Med.*, 2020, No. 102, P. 101762. https://doi/org/10.1016/j.artmed.2019.101762
3. Mustaqeem A., Anwar S. M., Majid M. A modular cluster based collaborative recommender system for cardiac patients, *Artif. Intell. Med.*, 2020, No. 102. P. 101761. https://doi.org/10.1016/j.artmed.2019.101761
4. Souza-Pereira L., Pombo N., Ouhbi S. et al. Clinical decision support systems for chronic diseases: A systematic literature review, *Comput. Methods Program Biomed*, 2020, No. 195, P. 105565. https://doi.org/10.1016/j.cmpb.2020.105565
5. Olakotan O. O., Yusof M. M. Evaluating the alert appropriateness of clinical decision support systems in supporting clinical workflow, *Journal Biomedical Informatics*, 2020, No. 106, P. 103453. https://doi.org/10.1016/j.jbi.2020.103453
6. MsRae M. P., Bozkurt B., Ballantyne C. M. et al. Cardiac ScoreCard: A diagnostic multivariate index assay system for predicting a spectrum of cardiovascular disease, *Expert Systems with Applications: An International Journal*, 2016, No. 54, pp. 136–147. https://doi.org/10.1016/j.eswa.2016.01.029
7. Thukral S., Rana V. Versatility of fuzzy logic in chronic diseases: A review, *Medical Hypotheses*, 2019, No. 122, pp. 150–156. https://doi.org/ 10.1016/j.mehy.2018.11.017
8. Gadaras I., Mikhailov L. An interpretable fuzzy rule-based classification methodology for medical diagnosis, *Artif. Intell. Med.*, 2009, No. 47(1), pp. 25–41. https://doi.org/10.1016/j.artmed.2009.05.003
9. Mokeddem S. A. A fuzzy classification model for myocardial infarction risk assessment, *Applied Intelligens*, 2018, No. 48, pp. 1233–1250. https://doi.org/ 10.1007/s10489-017-1102-1
10. Nauck D., Kruse R. Obtaining interpretable fuzzy classification rules from medical data, *Artif. Intell. Med.*, 1999, No. 16(2), pp. 149–169. https://doi.org/ 10.1016/s0933-3657(98)00070-0
11. Kalantari A., Kamsin A., Shamshirband S. et al. Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions, *Neurocomputing*, 2018, No. 276, pp. 2–22. https://doi.org/ 10.1016/j.neucom.2017.01.126
12. Jemal H., Kechaou Z., Ayed M.B. Enhanced decision support systems in intensive care unit based on intuitionistic fuzzy sets, *Advances in Fuzzy Systems*, 2017, 5b, pp. 1–8. https://doi.org/ 10.1155/2017/7371634
13. Pota M., Esposito M., Pietro G. Designing rule-based fuzzy systems for classification in medicine, *Knowl-Based Systems*, 2017, 124(C), pp. 105–132. https://doi.org/10.1016/j.knosys.2017.03.006
14. Minutolo A., Esposito M., Pietro G. A fuzzy framework for encoding uncertainty in clinical decision-making, *Knowl-Based Systems*, 2016, No. 98, pp. 95–116. https://doi.org/10.1016/j.knosys.2016.01.020
15. Ahmadi H., Gholamzadeh M., Shahmoradi L. et al. Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review, *Comput. Methods Program Biomed*, 2018, No. 161, pp. 145–172. https://doi.org/10.1016/j.cmpb.2018.04.013
16. Kour H., Manhas J., Sharma V. Usage and implementation of neuro-fuzzy systems for classification and prediction in the diagnosis of different types of medical disorders: a decade review, *Artif. Intell. Rev*., 2020, No. 53, pp. 4651–4706. https://doi.org/10.1007/s10462-020-09804-x
17. Sajadi N. A., Borzouei S., Mahjub H. et al. Diagnosis of hypothyroidism using a fuzzy rule-based expert system, *Clinical Epidemiology and Global Health*, 2019, No. 7(4), pp. 519–524. https://doi.org/10.1016/j.cegh.2018.11.007
18. Arji G., Ahmadi H., Nilashi M. et al. Fuzzy logic approach for infectious disease diagnosis: A methodical evaluation, literature and classification, *Biocybernetics and Biomedical Engineering*, 2019, No. 39(4), pp. 937–955. https://doi.org/10.1016/j.bbe.2019.09.004
19. Amato F., Lopez A., Pena-Mendez E. M. et al. Artificial neural networks in medical diagnosis, *J. Appl. Biomed*, 2013, No. 11(2), pp. 47–58. https://doi.org/10.2478/v10136-012-0031-x
20. Jiang J., Wang H., Xie J. et al. Medical knowledge embedding based on recursive neural network for multi-disease diagnosis, *Artif. Intell. Med*, 2020, No. 103, P. 101772. https://doi.org/ 10.1016/j.artmed.2019.101772
21. Alizadehsani R. Machine learning-based coronary artery disease diagnosis: A comprehensive review, *Computers in Biology and Medicine*, 2019, No. 111, P. 103346. https://doi.org/10.1016/j.compbiomed.2019.103346
22. Acharya U. R., Fujita H., Sudarshan V. K. et al. Automated characterization of coronary artery disease, myocardial infarction, and congestive heart failure using contourlet and shearlet transforms of electrocardiogram signal, *Knowl.-Based Syst.*, 2017, No. 132(15), pp. 156–166. https://doi.org/10.1016/j.knosys.2017.06.026
23. [Tan K. C., Yu Q., Heng C. M. et al. Evolutionary computing for knowledge discovery in medical diagnosis, *Artif. Intell. Med*., 2003, No. 27(2), pp. 129–154. https://doi.org/10.1016/S0933-3657(03)00002-2
24. Park Y.-J., Chun S.-H., Kim B.-C. Cost-sensitive case-based reasoning using a genetic algorithm: Application to medical diagnosis, A*rtif. Intell. Med.*, 2011, No. 51(2), pp. 133–145. https://doi.org/ 10.1016/j.artmed.2010.12.001
25. Wang M., Chen H. Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis, *Applied Soft Computing*, 2020, No. 88, P. 105946. https://doi.org/10.1016/j.asoc.2019.105946
26. Chen H. L., Yang B., Wang G. et al. Support vector machine based diagnostic system for breast Cancer using swarm intelligence, *J. Med. Syst.*, 2012, No. 36(4), pp. 2505–2519. https://doi.org/10.1007/s10916-011-9723-0
27. Zadeh L.A. Fuzzy sets as a basis for theory of possibility, *Fuzzy Sets and Systems 100 Suplements*,1999, pp. 9–34. https://doi.org/10.1016/S0165-0114(99)80004-9
28. Mendel J. M., John R. I., Liu F. Interval type-2 fuzzy logic systems made simple, *IEEE Trans. Fuzzy Syst*,2006, № 6,


DOI 10.15588/1607-3274-2024-2-7

pp. 808–821. http://dx.doi.org/10.1109/TFUZZ.2006.879986
29. ASA Physical Status Classification System [Electronic resourse]. Access mode: https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system/.
30. Sharkadi M.bM. Nechitki mnozhyny druhoho rodu, Naukovyi visnyk Uzhhorodskoho universytetu.Serija “Matematyka i informatyka, 2022, No. 2 (41), pp. 163–170. https://doi.org/10.24144/2616-7700.2022.41(2).163-170

Received 20.02.2024.
Accepted 25.04.2024.

УДК 004.023, 519.237, 616.1-07

# НЕЧІТКА МОДЕЛЬ ІНТЕЛЕКТУАЛІЗАЦІЇ ЗНАНЬ МЕДИЧНОЇ СФЕРИ

**Маляр М. М. –** д-р техн. наук, професор, декан факультету математики та цифрових технологій, Ужгородський національний університет, Ужгород, Україна.
**Маляр-Газда Н. М. –** канд. мед. наук, доцент, лікар-анестезіолог, Боршод-Обуй-Земплинська центральна регіональна лікарня та університетська навчальна лікарня, Мішкольц, Угорщина.
**Шаркаді М. М. –** канд. економ. наук, доцент, Ужгородський національний університет, Ужгород, Україна.

## АНОТАЦІЯ

**Актуальність.** Дослідження присвячено розробці гнучкого математичного апарату для інтелектуалізації знань у медичній сфері. Як правило, людське мислення базується на неточних, наближених даних, аналіз яких дозволяє формувати чіткі рішення. У випадках коли не існує точної математичної моделі об’єкта, або модель складна для реалізації доцільно використовувати апарат нечіткої логіки. Стаття направлена на розширення діапазону знань дослідників, які працюють в області медичної діагностики.

**Мета роботи** – підвищення якості відображення предметної області медичної сфери на основі побудови нечітких баз знань типу-2 з інтервальними функціями належності.

**Метод.** Описано підхід формалізації знань фахівця медичної галузі за допомогою нечітких множин другого порядку, який дозволяє враховувати невизначеність і нечіткість, яка властива медичним даним, а також вирішувати проблему інтерпретації отриманих результатів.

**Результати.** Розроблений підхід реалізовано на конкретній проблемі з якою стикається лікар-анестезіолог при допуску пацієнта до елективного (планового) оперативного втручання.

**Висновки.** Проведені експериментальні дослідження показали, що представлена нечітка модель типу-2 з інтервальними функціями належності дозволяє адекватно відображати вхідні медичні змінні якісного характеру та враховувати, як знання фахівця з медичної практики, так і дослідні медико-біологічні дані. Отримані результати мають важливе практичне значення для лікарів-практиків, особливо анестезіологів, оскільки дозволяють покращити оцінку стану пацієнта, зменшити кількість помилок та надати індивідуальні рекомендації. Це дослідження сприяє розвитку інтелектуальних систем, здатних позитивно впливати на клінічну практику та покращувати результати лікування пацієнтів у сфері медичної діагностики.

**КЛЮЧОВІ СЛОВА:** нечітка кластеризація, медична діагностика, функції належності другого роду.

## ЛІТЕРАТУРА

1. Mendel J. M. Interval Type-2 fuzzy logic systems: theory and design/ J. M. Mendel, R. I. John, Q. Liang // IEEE Transactions on Fuzzy Systems. – 2000. – № 8. – P. 535–550. http://dx.doi.org/10.1109/91.873577
2. Clinical Decision Support Systems for Triage in the Emergency Department using Intelligent Systems: a Review/ [Fernandes M., Vieira S. M., Leite F. et al.] // Artif. Intell. Med. – 2020. – No. 102. – P. 101762. https://doi/org/10.1016/j.artmed.2019.101762
3. Mustaqeem A. A modular cluster based collaborative recommender system for cardiac patients / A. Mustaqeem, S. M. Anwar, M. Majid // Artif. Intell. Med. – 2020. – No. 102.– P. 101761. https://doi.org/10.1016/j.artmed.2019.101761
4. Clinical decision support systems for chronic diseases: A systematic literature review / [L. Souza-Pereira, N. Pombo, S. Ouhbi et al.] // Comput. Methods Program Biomed. – 2020. – 195. – 105565. https://doi.org/10.1016/j.cmpb.2020.105565
5. Olakotan O. O. Evaluating the alert appropriateness of clinical decision support systems in supporting clinical workflow / O. O. Olakotan, M. M. Yusof // Journal Biomedical Informatics. – 2020. – No. 106. – P. 103453. https://doi.org/10.1016/j.jbi.2020.103453
6. Cardiac ScoreCard: A diagnostic multivariate index assay system for predicting a spectrum of cardiovascular disease / [M. P. MsRae, B. Bozkurt, C. M. Ballantyne et al.] // Expert Systems with Applications: An International Journal. – 2016. – No. 54. – P. 136–147. https://doi.org/10.1016/j.eswa.2016.01.029
7. Thukral S. Versatility of fuzzy logic in chronic diseases: A review / S. Thukral, V. Rana // Medical Hypotheses. – 2019. – No. 122. – P. 150–156. https://doi.org/10.1016/j.mehy.2018.11.017
8. Gadaras I. An interpretable fuzzy rule-based classification methodology for medical diagnosis / I. Gadaras, L. Mikhailov // Artif. Intell. Med. – 2009. – No. 47(1). – P. 25–41. https://doi.org/10.1016/j.artmed.2009.05.003
9. Mokeddem S. A. A fuzzy classification model for myocardial infarction risk assessment / S. A. Mokeddem // Applied Intelligens. – 2018. – No. 48. – P. 1233–1250. https://doi.org/ 10.1007/s10489-017-1102-1
10. Nauck D. Obtaining interpretable fuzzy classification rules from medical data / D. Nauck, R. Kruse // Artif. Intell. Med. – 1999. – No. 16(2). – P. 149–169. https://doi.org/10.1016/s0933-3657(98)00070-0
11. Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions / [A. Kalantari, A. Kamsin, S. Shamshirband et al.] // Neurocomputing. – 2018. – No. 276. – P. 2–22. https://doi.org/ 10.1016/j.neucom.2017.01.126
12. Jemal H. Enhanced decision support systems in intensive care unit based on intuitionistic fuzzy sets / H. Jemal, Z. Ke-


DOI 10.15588/1607-3274-2024-2-7

chaou, M. B. Ayed // Advances in Fuzzy Systems. – 2017. – 5 b. –P. 1–8. https://doi.org/ 10.1155/2017/7371634
13. Pota M. Designing rule-based fuzzy systems for classification in medicine / M. Pota, M. Esposito, G. Pietro // Knowl-Based Systems. – 2017. – 124(C). – P. 105–132. https://doi.org/10.1016/j.knosys.2017.03.006
14. Minutolo A. A fuzzy framework for encoding uncertainty in clinical decision-making / A. Minutolo, M. Esposito, G. Pietro // Knowl-Based Systems. – 2016. – No. 98. – P. 95–116. https://doi.org/10.1016/j.knosys.2016.01.020
15. Diseases diagnosis using fuzzy logic methods: A systematic and meta-analysis review / [H. Ahmadi, M. Gholamzadeh, L. Shahmoradi et al.] // Comput. Methods Program Biomed. – 2018. – No. 161. – P. 145–172. https://doi.org/10.1016/j.cmpb.2018.04.013
16. Kour H. Usage and implementation of neuro-fuzzy systems for classification and prediction in the diagnosis of different types of medical disorders: a decade review / H. Kour, J. Manhas, V. Sharma // Artif. Intell. Rev. – 2020. – No. 53. – P. 4651–4706. https://doi.org/10.1007/s10462-020-09804-x
17. Diagnosis of hypothyroidism using a fuzzy rule-based expert system/ [N. A. Sajadi, S. Borzouei, H. Mahjub et al.] // Clinical Epidemiology and Global Health. – 2019. – No. 7(4). – P. 519–524. https://doi.org/10.1016/j.cegh.2018.11.007
18. Fuzzy logic approach for infectious disease diagnosis: A methodical evaluation, literature and classification / [G. Arji, H. Ahmadi, M. Nilashi et al.] // Biocybernetics and Biomedical Engineering. – 2019. – No. 39(4). – P. 937–955. https://doi.org/10.1016/j.bbe.2019.09.004
19. Artificial neural networks in medical diagnosis / [F. Amato, A. Lopez, E. M. Pena-Mendez et al.] // J. Appl. Biomed. – 2013. – No. 11(2). – P. 47–58. https://doi.org/10.2478/v10136-012-0031-x
20. Medical knowledge embedding based on recursive neural network for multi-disease diagnosis / [J. Jiang, H. Wang, J. Xie et al.] // Artif. Intell. Med. – 2020. – No. 103. – 101772. https://doi.org/ 10.1016/j.artmed.2019.101772
21. Alizadehsani R. Machine learning-based coronary artery disease diagnosis: A comprehensive review / R. Alizadehsani // Computers in Biology and Medicine. – 2019. – No. 111. – P. 103346. https://doi.org/10.1016/j.compbiomed.2019.103346
22. Automated characterization of coronary artery disease, myocardial infarction, and congestive heart failure using contourlet and shearlet transforms of electrocardiogram signal / [U. R. Acharya, H. Fujita, V. K. Sudarshan et al.] // Knowl.-Based Syst. – 2017. – No. 132(15). – P. 156–166.https://doi.org/10.1016/j.knosys.2017.06.026
23. Evolutionary computing for knowledge discovery in medical diagnosis / [K. C. Tan, Q. Yu, C. M. Heng et al.] // Artif. Intell. Med. – 2003. – No. 27(2). – P. 129–154. https://doi.org/10.1016/S0933-3657(03)00002-2
24. Park Y.-J. Cost-sensitive case-based reasoning using a genetic algorithm: Application to medical diagnosis / Y.-J. Park, S.-H. Chun, B.-C. Kim // Artif. Intell. Med. – 2011. – No. 51(2). – P. 133–145. https://doi.org/ 10.1016/j.artmed.2010.12.001
25. Wang M. Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis / M. Wang, H. Chen // Applied Soft Computing. – 2020. – 88. – 105946. https://doi.org/10.1016/j.asoc.2019.105946
26. Support vector machine based diagnostic system for breast Cancer using swarm intelligence / [H. L. Chen, B. Yang, G. Wang et al.] // J. Med. Syst. – 2012. – No. 36(4). – P. 2505–2519. https://doi.org/10.1007/s10916-011-9723-0
27. Zadeh L. A. Fuzzy sets as a basis for theory of possibility / L. A. Zadeh // Fuzzy Sets and Systems 100 Suplements. – 1999. – P. 9–34. https://doi.org/10.1016/S0165-0114(99)80004-9
28. Mendel J. M. Interval type-2 fuzzy logic systems made simple / J. M. Mendel, R. I. John, F. Liu // IEEE Trans. Fuzzy Syst. – 2006. – № 6. – P. 808–821. http://dx.doi.org/10.1109/TFUZZ.2006.879986
29. ASA Physical Status Classification System[Електронний ресурс]. – Код доступу: https://www.asahq.org/standards-and-guidelines/asa-physical-status-classification-system/.
30. Шаркаді М. М. Нечіткі множини другого роду / М. М. Шаркаді // Науковий вісник Ужгородського університету. Серія «Математика і інформатика». – 2022. – № 2 (41). – С. 163–170. https://doi.org/10.24144/2616-7700.2022.41(2).163-170


DOI 10.15588/1607-3274-2024-2-7

UDC 004.658

# USING MODULAR NEURAL NETWORKS AND MACHINE LEARNING WITH REINFORCEMENT LEARNING TO SOLVE CLASSIFICATION PROBLEMS

**Leoshchenko S. D. –** PhD, Senior Lecturer of the Department of Software Tools, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine.

**Oliinyk A. O. –** Dr. Sc., Professor, Professor of the Department of Software Tools, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine.

**Subbotin S. A.** – Dr. Sc., Professor, Head of the Department of Software Tools, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine.

**Kolpakova T. O. –** PhD, Associate Professor, Associate Professor of the Department of Software Tools, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine.

## ABSTRACT

**Context.** The solution of the classification problem (including graphical data) based on the use of modular neural networks and modified machine learning methods with reinforcement for the synthesis of neuromodels that are characterized by a high level of accuracy is considered. The object of research is the process of synthesizing modular neural networks based on machine learning methods with reinforcement.

**Objective** is to develop a method for synthesizing modular neural networks based on machine learning methods with reinforcement, for constructing high-precision neuromodels for solving classification problems.

**Method.** A method for synthesizing modular neural networks based on a reinforcement machine learning approach is proposed. At the beginning, after initializing a system of modular neural networks built on the bottom-up principle, input data is provided – a training set of data from the sample and a hyperparameter to select the size of each module. The result of this method is a trained system of modular neural networks. The process starts with a single supergroup that contains all the categories of the data set. Then the network size is selected. The output matrix is softmax, similar to the trained network. After that, the average probability of softmax is used as a similarity indicator for group categories. If new child supergroups are formed, the module learns to classify between new supergroups. The training cycle of modular neural network modules is repeated until the training modules of all supergroups are completed. This method allows you to improve the accuracy of the resulting model.

**Results.** The developed method is implemented and investigated on the example of neuromodel synthesis based on a modular neural network for image classification, which can later be used as a model for technical diagnostics. Using the developed method significantly reduces the resource intensity of setting up hyperparameters.

**Conclusions.** The conducted experiments confirmed the operability of the proposed method of neuromodel synthesis for image classification and allow us to recommend it for use in practice in the synthesis of modular neural networks as a basis for classification models for further automation of tasks of technical diagnostics and image recognition using big data. Prospects for further research may lie in using the parallel capacities of GPU-based computing systems to organize directly modular neural networks based on them.

**KEYWORDS:** modular neural networks, image classification, synthesis, diagnostics, topology, artificial intelligence, reinforcement learning.

## ABBREVIATIONS

ANN is an artificial neural network;
BP is a backpropagation method;
CNN is a convolutional neural network;
GAN is a generative adversarial network;
GPU is a graphics processing unit;
MNN is a modular neural network;
PCA is a principal component analysis;
VAE is a variational autoencoders;
ViT is a vision transformer.

## NOMENCLATURE

$\Delta AD$ is a density of accuracy changes;
$DataSet$ is a data set;
$E$ is a relative error;
$Error_{ANN}$ is a neuromodel error;
$F1\,score$ is a F1 score;
$FN$ represents false-negative results: the number of samples that the model predicted as not belonging to a class, while they actually belong to that particular class;
$FP$ represents false positive results: the number of samples that were predicted to belong to the class, while they do not belong to the class at all;
$i, j$ is a reference to the number of layers in two network modules;
$Model$ is a synthesized neuromodel;
$MS$ is a block network size;
$n$ is a number of input features that characterize sample instances;
$Precision$ is a precision of work of neuromodel;
$Recall$ is a recall of work of neuromodel;
$TP$ represents true positive results: the number of samples that were predicted to belong to the class and correctly belong to the class;


DOI 10.15588/1607-3274-2024-2-8

*VA* is a verification accuracy obtained during image classification;

$x_n$ is an independent attribute of the sample instance.

## INTRODUCTION

Artificial intelligence technologies such as machine learning, deep learning, and computer vision can help use automation to structure and organize almost all data [1].

Images, including images and video glasses, provide the bulk of global data [2]. To interpret and systematize this data, developers of modern computing systems and software turn to image classification based on artificial intelligence [2].

Image classification analyzes digital images (photos, drawings, scanned copies, etc.) using analytical models based on artificial intelligence, which can identify and identify a wide range of criteria-from the content of the image to the time of day [1, 2].

Image classification is the task of classifying and assigning labels to groups of images or vectors within an image based on certain criteria [3]. A label can be assigned based on one or more criteria.

Image classification is a serious resource-intensive task for comparing modern architectures and methodologies in the field of Computer Vision [1, 3].

In a single-attribute classification, each image has only one label or annotation, as the name implies. As a result, for each image that the model sees, it analyzes and classifies it according to only one criterion [4].

On the other hand, when classified by multiple attributes, images can have multiple labels, and some images contain all the labels that you use at the same time [3].

The whole process begins with preprocessing – data preparation [5, 6].

This step improves image data by eliminating unwanted deformations and improving certain key aspects of the image so that computer vision models can work with these advanced data. In fact, data is cleaned up and prepared for processing by the artificial intelligence model [7, 8].

Data cleaning, sometimes called preprocessing, is an important step in preparing data for training your model, as data inaccuracies lead to inaccuracies in the image classification model [1–4]. As the results of clearing the data, you can see what happened:

– removing duplicates: duplicating data frees up the learning process and may cause the model to add more weight to duplicated data unnecessarily. [7, 8];

– deleting irrelevant data: including irrelevant data will not help train the model to achieve the desired goal [7, 8];

– filtering unwanted plague data (information outliers): some data, although technically relevant, do not help in training the artificial intelligence model [7, 8]. Data that goes beyond the norm can distort your model's predictions, so it's best to just delete it [3];

– missing data detection: missing data can cause problems in the learning process-during data cleaning, missing data can be identified and updated accordingly [7, 8];

– correction of structural errors: most machine learning methods are not able to identify errors as a human would, which means that each piece of data must be precisely organized [7, 8].

Well-organized data guarantees greater success when it comes to training an image classification model or any artificial intelligence model [7, 8].

Object detection: determines the location of objects in a set of images. This is the process of determining the location of an object, which includes image segmentation and determining the location of the object [4].

Object recognition and training: marking arranged images.

Deep learning methods reveal laws in the image and characteristics that may be unique to a particular feature.

The model is trained on this data set and becomes more accurate in the future [7, 8].

After the data has been identified, you need to learn your artificial intelligence model [7, 8]. This involves uploading a large amount of data to each of the attributes to give the artificial intelligence model the ability to learn something. The more training data is loaded, the more accurately the model will determine the content of each image [6].

Object classification: the model is ready to classify images. This is the final step in the development of an artificial intelligence model that classifies images according to several different criteria [7, 8].

The method uses an appropriate classification approach to classify the observed elements into predefined classes [1].

This is done by comparing the sample patterns with the desired patterns.

Now elements that were added as tags in the next step will be recognized by the algorithm in real images [7, 8].

However, most ANN training methods are still the same modifications of the BP method with certain adaptations to the more complex ANN architecture [11]. Moreover, such methods do not in any way touch the synthesis stage of the network topology. That is why the task of developing new methods based on the evolutionary approach for full-fledged ANN synthesis is urgent.

In modern computer vision, the model architecture is most often used as the most optimal standard by CNN [9].

The convolution operation introduces a matrix called a kernel into the input, performing pairwise multiplication of the Matrix and summing the result. Note that the filter and kernel are the same thing [10].

To apply the filter, imagine that it is applied to the input matrix, starting from the upper-right edge, multiply each value by the corresponding kernel value and sum them [9, 10]. Then the core is shifted by a certain amount, which is called a step, usually by one, and the process is repeated.

To apply the filter, imagine that it is applied to the input matrix, starting from the upper-right edge, each value


DOI 10.15588/1607-3274-2024-2-8

is multiplied by the corresponding kernel value and summed [9, 10]. Then the core is shifted by a certain amount, which is called a step, in this case by one, and the process is repeated again [9, 10].

The number of cores used simultaneously is called functions. For example, if you apply four cores $3\times3$ to an image, the resulting matrix will have four elements [9, 10].

In practice, core numbers are usually randomly initiated, put together, and updated during training using gradient descent [9, 10].

One of the key aspects of the convolution operation is the distribution of weights: most often, the same core, the gray mat, is reused at each step, covering the entire image [9, 10]. This significantly reduces the number of weights required, reducing calculation costs. Moreover, this leads to a shift in position, as the network learns to rely on the input matrix, combining pixels closely with each other. This helps because adjacent pixels are usually semantically related [9, 10].

Typically, in neural networks, several convolutional operation blocks are stacked together to form a layer. Several layers are then stacked together like constructors to form the final model. The number of layers is called depth; the more layers, the deeper the network [9, 10].

To further reduce the overhead of counting, different types of pools are applied between layers. Pool management is a complex process. It takes the Matrix and reduces it, summing up its values. The most common is the maximum Union, where the researcher determines the size of the window and takes the maximum value inside it [9, 10].

The final vector obtained by feeding an image packet to this series of convolution layers is then passed to a fully connected layer, which produces a vector with the classes [9, 10].

We present a comparison of the methods discussed above based on artificial neural networks in the following Table 1. we will use the following comparison criteria:

– structural complexity: the structural complexity of the network under consideration;

– using the GPU during training: it is possible to use the GPU during training in order to speed up this process;

– complexity of activation functions: assessment of the complexity and resource intensity of the activation function;

– working on large data sets: general assessment of the network's performance with large data samples.

Studding at the results of the comparison, we can conclude that today the general trend is the use of various modifications of CNN. At the same time, a not always more complex topology guarantees greater accuracy, as in the case of comparing the results of AlexNet and ResNet, because ResNet has a simpler and clearer architecture, but demonstrates higher accuracy. On the other hand, the use of a combination of layers of different topologies of artificial neural networks, as in the case of ViT, demonstrates the highest accuracy, but this imposes significant limitations in the use of modern computing capabilities (performing parallel calculations on the GPU). This is due to the fact that when performing calculations on parallel threads, constant synchronization and data forwarding between threads sometimes require significantly more resources than parallel calculations themselves.

Table 1 –Comparison of neural network models for image classification

| | AlexNet | ResNet | ViT |
|---|---|---|---|
| Structural complexity | CNN-based, more simplified architecture (fewer layers, but more neurons) | CNN-based, more simplified architecture (fewer layers) | More complex, composite structure (layer Norm + MSP+ MLP) |
| Using the GPU during training | Possible | Possible | Possible (complicated by data forwarding) |
| Complexity of activation functions | Advanced function for improved accuracy (ReLU) | Advanced function for improved accuracy (ReLU) | Composite function for different blocks (GELU + softmax layers) |
| Working on large data sets | Possible (average accuracy and speed of operation) | Possible (medium accuracy and high speed of operation) | Possible (high accuracy and low speed) |

That is why the task of finding the most optimal neuromodel architecture for image classification remains relevant, which would be characterized by topological simplicity and the possibility of full use of the capabilities of parallel calculation systems.

**The object of study** the process of MNN synthesis, based on reinforcement learning.

Existing solutions based on artificial neural networks are highly accurate, but they are characterized by excessive topological complexity and lack of the ability to fully use the capabilities of parallel computing systems.

**The subject of the study** MNN synthesis method using a machine learning approach with reinforcement learning.

Today, the general trend is the use of various modifications of deep ANNs. However, a more complex topology does not always guarantee greater accuracy. On the other hand, the use of a combination of layers of different ANN topologies demonstrates the highest accuracy, but complicates their implementation. Therefore, the paper proposes an approach based on the use of machine learning with reinforcement for MNN training, which allows combining different topologies and simplifying their synthesis.

**The purpose of the work** is to development of a method for MNN synthesis based on machine learning methods with reinforcement, for constructing high-precision neuromodels for solving classification problems.


DOI 10.15588/1607-3274-2024-2-8

## 1 PROBLEM STATEMENT

Let us define a sample of data consisting of images that are characterized by a number of features: $\{x_1, x_2, ..., x_n\}$. Then, using a model based on an artificial neural network: $Model = \{ANN\}$, you need to pass the entire sample of data $DataSet = \{x_1, x_2, ..., x_n, y\}$ through it, determine the class of each of the images, determine the dependent feature $\{y\}$. The conditions for regulating the learning process are to determine the reward/penalty during training. $Model = \{ANN\}$ in order to obtain a solution characterized by the greatest accuracy of work: $Error_{ANN} \to 0$.

## 2 REVIEW OF THE LITERATURE

To teach a model to classify images, huge amounts of data are required [13]. There are most often two approaches to machine learning models:

– supervised learning;

– unsupervised learning.

The more common of the two is learning with a teacher. This is when the expert advisor provides the system with data samples that already have a class designation attribute (or rule responses). This teaches the model to recognize correlations and apply procedures to new data [13].

Learning without a teacher is not as common as learning with a teacher. Learning without a teacher is characterized by sloppy, raw data without human intervention. This approach does not use basic data [14].

However, learning without a teacher can lead to solving problems when a person is an expert who has not yet been identified [13].

Supervised machine learning is a type of machine learning in which a model is trained on a set of data with markers, that is, the model is provided with input – output pairs. The goal is for the model to examine the mapping between inputs and outputs so that it can make predictions or make decisions when new, previously unmarked data arrives.

Supervised learning usually has clear stages.

Data collection. The work begins with collecting a data set consisting of the values of input object attributes and the corresponding class labels. For example, signs may include data such as color, size, age, gender, clinical indicators, and so on.

Preprocessing of data. Once the tag is created, preprocessing usually takes place to clean up and convert the data to a format suitable for training. This may include processing missing values, scaling functions, or encoding categorical variables.

Data separation. The entire sample must be divided into two or more subsets: usually this is a time for training the model and, accordingly, a part for testing. The training set is used to train the model, while the test set is used to evaluate its performance on invisible data.

Selecting a model. In the future, the structure of Delhi and the specific method of training are selected. The choice of model depends on various factors, such as the nature of the data, the complexity of the task, and the available computing resources.

Training the model. The model is trained based on training data that processes input objects along with the corresponding labels. During training, the model adjusts its internal parameters to minimize the difference between its forecasts and actual labels.

Evaluating the model. After training, you need to evaluate the performance of the model using a set of tests. This gives you an idea of how well model summarizes new, invisible data. Common evaluation measures include accuracy, precision, responsiveness, F1 score, and so on.

Setting up hyperparameters. In many cases, you may need to fine-tune the model's hyperparameters to further improve its performance. Hyperparameters are config parameters that are not studied during training, but affect the learning process.

Deployment. Once a satisfactory level of model performance is reached, you can deploy it to make predictions based on new, unknown data in real-world applications.

Throughout this process, it is important to repeat and improve your approach based on the performance of the model and the conclusions obtained from the analysis of results. Supervised learning is widely used in various fields, including, but not limited to, regression, classification, and ranking tasks.

Unsupervised machine learning is a type of machine learning in which you have input data without the appropriate labels. The goal is to identify hidden laws, structures, or relationships within them. Unlike learning with a teacher, there are no pre-determined results or correct answers that guide the learning process. The method attempts to find internal structures or representations in the data.

Unsupervised learning is usually similar in stages to learning with teachers.

Data collection. As with learning with a teacher, the robot starts by collecting a set of data. However, when teaching without a teacher, the data set consists only of input objects without any corresponding labels or target variables.

Preprocessing of data. Preprocessing is similar to supervised learning, processing missing values, scaling functions, encoding categorical variables, and so on.

Selecting a model. Choose a teaching method without a teacher or a model that matches the task. Advanced unsupervised learning methods include clustering, dimensionality reduction, and generative modeling.

Training the model. When teaching without a teacher, Model learns exclusively from input data without explicit control. The method examines the structure of data and examines patterns or representations that reflect their main characteristics.

Clustering. Clustering methods group similar data points based on some degree of similarity. K-means clustering and hierarchical clustering are examples of popular clustering methods.


DOI 10.15588/1607-3274-2024-2-8

Dimensionality reduction: dimensionality reduction methods are aimed at reducing the number of objects while maintaining as much relevant information as possible. PCA and t-SNE are commonly used to reduce dimensionality.

Generative modeling: generative models learn the basic probability distribution of data and can generate new patterns similar to training data. Examples are autoencoders, VAE, and GAN.

Evaluating the model. Unlike learning with a teacher, there may not always be clear assessment indicators for learning tasks without a teacher. Evaluation often involves qualitative evaluation by checking learned representations, visualizing clusters, or evaluating the quality of non-selected samples.

Interpretation and application: once the model is trained, it is possible to interpret the learned patterns or representations and apply them to solve specific problems or obtain information about the data.

Unsupervised learning is used in a variety of applications, such as customer segmentation, anomaly detection, data compression, and character learning. This can be especially useful when tagging data is scarce or expensive, or when learning and understanding the data structure is the main goal.

Reinforcement learning is an area of machine learning that considers taking appropriate steps to maximize reward in a particular situation [13–16]. This approach is used by various software and machines to find the best possible behavior or path to choose in a particular situation. Reinforcement learning differs from training with a teacher in that when training with a teacher, the training yes-no contains a key to the answer, so the model learns with the correct answer by itself, whereas when training with reinforcement, there is no answer, but the reinforcement agent decides what to do to complete the task. In the absence of a training data set, he must learn from his own experience [13–16].

A well-known example shows a robot, a diamond, and a fire. The goal of the robot is to get a reward in the form of a diamond and avoid obstacles that are triggered [13–16]. The robot learns, tries all possible paths, and then chooses the one that gives it the reward with the fewest cross-codes. Each correct step will give the robot a reward, and each wrong step will take the reward from the robot. The total reward will be calculated when it reaches the final reward-a diamond. [13–16]

Basic principles of reinforcement learning [13–16]:

– input data: the input data must be the initial state from which the model will start;

– result: there are many possible outcomes, as there are many solutions to a particular problem;

– training is based on input data, the model will return the state, and the user will decide to reward or punish the model based on its results;

– the model continues to learn.

– the best decision is made on the basis of a mock reward.

Let's compare machine learning methods for classifying images by the following criteria:

– input requirements: overall assessment of the machine learning approach and appropriate methods for organizing and presenting input data;

– requirements for training metaparameters: requirements for parameters and settings necessary for using the training process;

– complexity of the organization: assessment of the complexity of organizing the model training process, including in terms of computing resources;

– adaptability of solutions (models): assessment of the adaptability and universality of the obtained Solutions, their ability to work with updated data about the task (or environment).

The results of the analysis are presented in the form of a Table 2.

Table 2 – Comparison of machine learning approaches

| | Supervising learning | Unsupervised learning | Reinforcement learning |
|---|---|---|---|
| Input requirements | The highest | Minimum | Average |
| Requirements for training weather parameters | High | High | Average |
| Complexity of the organization | Average | High | High |
| Adaptability of solutions (models) | Average | High | The highest |

From the results of the comparison, it can be concluded that in general, learning with and without a teacher, as the main approaches to machine learning, already take into account a large number of developed and studied advantages and disadvantages. Thus, the very concept of learning with a teacher requires a detailed preprocessing of input data with their clear classification in the educational part. Learning without a teacher does not require this from the data and can even work with them without much verification, but most methods of learning without a teacher require prior determination of many metaparameters to start the learning process.

Reinforcement learning is a more modern approach that attempts to solve the known problems of previous strategies. Yes, it has requirements for input data about the environment, but they are not excessive, and metaparameters are usually limited to defining criteria for stopping and determining the success/failure of training. Also, Solutions (mods) obtained on the basis of reinforcement learning are most adaptable and versatile, because they are immediately synthesized taking into account the variability of the environment in which the agent operates.

## 3 MATERIALS AND METHODS

The main idea of MNN is based on modularity, and is that: the best approach to solving a given complex problem is not to perform one giant task, but a system of separate and mostly independent subtasks working together to


DOI 10.15588/1607-3274-2024-2-8

achieve one big goal [17–20]. This concept has a biological basis. There are functionally specialized areas in the brain that are specific to various cognitive processes. In a part of the brain called the thalamus, there is a lateral cranial nucleus, which is divided into layers that separately process color and color: both the main components of vision [17–20]. MNN use this idea to solve complex artificial intelligence problems. Several independent neural networks are simultaneously trained for a specific subtask, and their results are eventually combined to perform a single task. Advantages of MNN include [17–20]:

– simplicity of structural structure and organization;
– a combination of teaching techniques and methods;
– scalability of solutions;
– efficient use of computing systems.

As you know, artificial intelligence does not inherently speak language, so the main task is to teach an analytical model to reason over a given text. Reasoning refers to tasks such as arithmetic, sorting, comparison, and counting [17–20].

For example, you need to create a system of responses to questions Fig. 1 illustrates how the quality control model works [17–20].

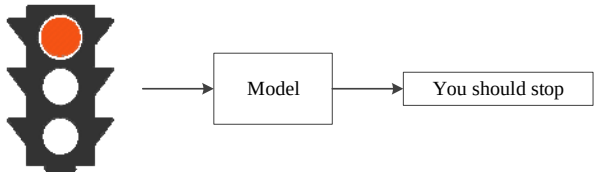



Figure 1 – Visualization of the answer to a question-the task is to answer a question about an image to show that the system understands the image

So, in the problem with determining what the sign indicates: the color mark of the traffic light in Fig. 1. To answer this question, you need to perform several reasoning steps: find a traffic light, determine the color mark, find out the color, and then provide a decision on further actions [17–20].

MnMs are able to provide such reasoning. Using this question, the model builds a specific network architecture, and then performs the assembled sequence of neural modules to output the answer, as shown in Fig. 2 [17–20].

MNNs refer to artificial neural networks that consist of several different neural networks connected together in combination with an intermediary. To illustrate this point further, consider a consumer who owns several smart devices, such as a smartphone, smartwatch, and tablet, such as an iPad, in addition to a laptop or desktop computer [17–20]. Despite the different capabilities of these respective devices, all of them will be connected to a modem or traffic jam route, which will allow users of these devices to quickly and efficiently access online and mobile services. In addition, this online connection also allows users to combine the functionality of their various devices to achieve a specific goal, such as streaming a popular television program or calling a friend or family member, among other things [17–20].

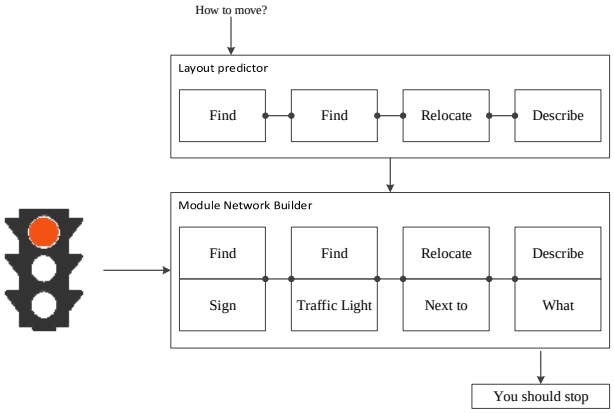



Figure 2 – For each instance, the model first assumes a layout, then, using the image functions, it builds the assembled network of neural modules to output the response

Taking all of the above into account, MNNs enable software developers to use the capabilities of individual neural networks in a more consistent and efficient way. To do this, each neural network within a larger MNN will be used to solve part of a specific problem [17–20]. At this point, an intermediary known as an integrator will be used to organize and analyze these many modules to create the final result of the neural network. Thanks to this configuration, simple neural networks can be implemented in a more complex way, and some common applications of these networks include image processing, high-level input compression, and stock market forecasting software [17–20].

Ensemble training. The concepts and ideas that formed the basis for creating modular neural networks were first theorized in the 1980s and led to the development of a machine learning method called ensemble learning [17–20]. This method is based on the idea that weaker machine learning models can be combined together to create a single stronger Model [17–20]. Moreover, this collective approach can be used to produce more significant results than those that could be obtained using a single deep learning model. Another way to conceptualize this process is the divide-and-conquer approach, in which a large problem is solved by breaking it down into smaller parts that can be solved in a simpler or more viable way [17–20].

MNN structure. Consistent with the idea that neural networks are based on multiple functions and capabilities of the human brain, the human brain consists of a hierarchy of networks consisting of millions of neurons, with each network specialized to perform specific work related to the functioning of the human body [17–20]. To give an example of this, let’s look at 3 different direct-link neural networks that have been trained to solve problems related to pattern recognition, such as license plate recognition. For many reasons, these 3 networks may have difficulty analyzing and identifying license plates in videos or images on an ongoing basis [17–20]. Thus, the software developer could combine the source data of these three models to create a single output signal in order to create a


DOI 10.15588/1607-3274-2024-2-8

single model that is more accurate than the previous 3 models [17–20].

In addition, software developers can also use MNN to break down the learning problem itself into smaller, more manageable parts [17–20]. Going back to the example of license plate recognition, this problem can be divided into three subtasks, such as identifying license plates in images, identifying license plates in videos, and identifying license plates in images that depict bad weather, such as heavy rain or snowfall. Thanks to these combined conclusions, the developer could create a deep learning algorithm that would be able to identify the appearance of license plates in both images and videos depicting various weather conditions that a driver may encounter when passing a busy intersection or street [17–20].

Using MNN offers a wide range of benefits. For example, combining multiple neural networks that have been trained to perform a specific task can significantly reduce the training time that is often required to train neural networks, as well as the associated costs [17–20]. Conversely, this approach also allows software developers to combine different types of machine learning models and approaches that would not be possible with traditional methods. Finally, as already mentioned, the combination of several neural networks can lead to the development of a single neural network that can work much more accurately and efficiently than any of the previous networks were capable of independently [17–20].

Since artificial neural networks have generally become a hot topic for both research and software development over the past few years, it is appropriate that MNN is an approach that is also becoming increasingly popular [17–20]. Moreover, since the cost of training machine learning models in practice can be exorbitant, MNNs allow software developers to create models much cheaper and more sustainable. At the same time, MNNs will continue to be used in the coming years, since approaches that were considered inadequate in previous scenarios can be combined to create a single viable solution [17–20].

In the usual formulation of the problem of image classification as a computer vision problem, it is reduced to maximizing the confidence function from a set of hypothetical target locations, where reliability can be studied in a fully controlled or poorly controlled setting [17–20]. In the formulation of a sliding window, the hypothesis set consists of a large set of rectangular windows, and the maximization problem is solved by exhaustive search. Since this process, as a rule, is too expensive in practice, many methods have been proposed to speed it up, from methods that use the properties of a confidence function, to sentence methods or cascade methods. All these methods retain the property of exhaustive search in the hypothesis space, aimed either at reducing the number of hypotheses to begin with, or at effectively searching for them [17–20].

In general, the advantages of using subcripping training in image classification are as follows:

– reinforcement learning can be used to solve very complex problems that cannot be solved by conventional methods when the input volumes are too large [17–20];

– this approach is preferred for achieving high rates of adaptability and versatility of solutions [17–20];

– provides an opportunity to correct mistakes made in the learning process [17–20];

– as soon as the error is corrected by the model, the probability of occurrence of the same error is very small;

– in the absence of a training data set, the sub-course must learn from its own experience [17–20];

– reinforcement learning models can translate people into many tasks [17–20];

– reinforcement learning is designed to achieve the ideal behavior of a model in a particular text in order to maximize its performance [17–20];

– reinforcement learning methods maintain a balance between research and exploitation. Research is the process of testing different things to see if they are better than what has been done before. Exploitation is the process of testing what has worked best in the past. Other learning algorithms do not provide such a balance [17–20].

To build an MNN system, you need to select the appropriate network size for each module, and then find the average softmax probability for playing categories. The MNN system is built on the bottom-up principle. The input data for the method is a training dataset from the sample and a hyperparameter for selecting the size of each module. The result of the method is a trained MNN system [17–20]. First, we start with a single supergroup that contains all the categories of the data set. Then the network size is determined. The output matrix is softmax, similar to the trained network. After that, the average probability of softmax is used as a similarity indicator for group categories. If new child supergroups are formed, the module is trained to classify between new supergroups. Training of MNN modules is described below. The cycle is repeated until the modules of all superg-Rup [17–20] are trained.

Selecting neural network sizes and hyperparames. You must select the number of layers and the size (configuration) of each module. You should use a new metric called the accuracy change density [17–20] $(\Delta AD)$. It measures the accuracy gain per unit increase in model size between two network models:

$$\Delta AD_{ij} = \frac{VA_i - VA_j}{MS_i - MS_j}. \quad (1)$$

If we consider the efficiency of use $\Delta AD$ as an indicator, it should be noted that when increasing the size of the model, there is no noticeable increase in accuracy. The accuracy is calculated for each MNN module before other networks are grouped into supergroups [17–20]. For example, $\Delta AD$ for a root module, MNN is calculated based on the accuracy obtained for classification between all categories of the data set, and for any other module in MNN, it is calculated by classification between all its


DOI 10.15588/1607-3274-2024-2-8

child categories [17–20]. Using $\Delta AD$, you can define effective network configurations and distinguish categories with high accuracy. This approach can lead to a slight loss of accuracy compared to large monolithic neural networks, since small networks with MNN need to classify only several groups of visually similar categories, rather than all categories of the data set.

Fig. 3 shows a common variant of communication between MNN modules during operation. This is how the main module is allocated, which is responsible for distributing operations and further synchronizing their results. Fig. 4 shows a schematic example of simulated models.

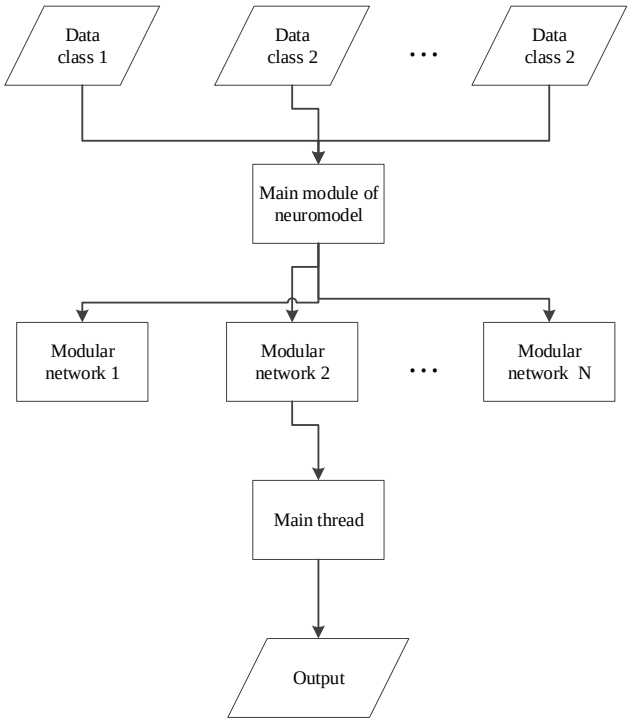



Figure 3 – General representation of communication between MNN modules during operation

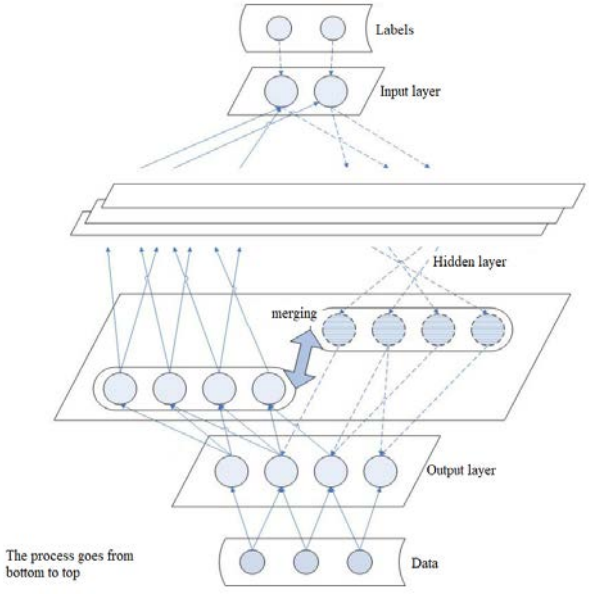



Figure 4 – Schematic example of synthesized networks

## 4 EXPERIMENTS

Several sets of image data were used in the experiments. CEFAR [27–29], SVHN [30], and EMNIST [31], which contain centered and fixed-size images with only one object in each image. Images in large data sets such as ImageNet 2012 [32] and Caltech-256 [33] have different sizes and display real images more accurately.

The CEFAR datasets [27–29] consist of color images of $32\times32$ different categories in size. The training and test kits contain 50 thousand and 10 thousand images, respectively. During the experiments, the generally accepted practice of using 5,000 images from the training set to form a test set was observed [27–29]. The SVHN dataset contains 73,257 pixel-sized color images in the training set $32\times32$ and 531,131 images for additional training [30]. When reporting the results of the SVHN data set, it was generally accepted to use all training data without any increase in data. A set of 6000 images is used to check learning outcomes. EMNIST is an extension of the popular MNIST dataset [31]. There are six configurations of the EMNIST dataset, and an EMNIST-balanced configuration was used. It contains 131,600 pixel – sized grayscale images belonging to 47 categories. The ImageNet training set contains 1000 categories of approximately 1000 images each [33] $28\times28$. ImageNet also includes a verification kit and a testing kit. A subset of the ImageNet dataset with 20 categories was also used to easily visualize MNN and fully understand the details and properties of the hierarchy. The Caltech dataset is also used in experiments [32]. As suggested in [32], a subset of 11 categories from the Caltech dataset was used to provide a fair comparison with existing work. Each category of the Caltech dataset contains approximately 100 training images and 20 test images. The parameters of the data sets are described in Table 3.

Table 3 – Detailed information about sampling data for testing

| Data set | Size of the images | Number of images to study | Number of images to test | Number of image classes |
|---|---|---|---|---|
| CIFAR 10 | $32\times32\times3$ | 50000 | 10000 | 10 |
| CIFAR 100 | $32\times32\times3$ | 50000. | 10000 | 100 |
| SVHN | $32\times32\times3$ | 604388 | 26032 | 10 |
| EMNIST | $28\times28\times1$ | 112800 | 18800 | 47 |
| ImageNet 2012 | Different | 1200000 | 75000 | 1000 |
| ImageNet 2012 | Different | 26000 | 2000 | 20 |
| Caltech | Different | 2000 | 400 | 11 |

Image classification models should be evaluated to determine how well they perform compared to other models [7, 8]. Here are some well-known indicators used in image classification [7, 8].

Accuracy. Accuracy is an indicator that is defined for each class. Class accuracy tells us how much of the data provided by the machine learning model for class membership was actually part of the class in the validation data. A simple formula can demonstrate this [7, 8]:


DOI 10.15588/1607-3274-2024-2-8

$$Precision = \frac{TP}{TP + FP}. \quad (2)$$

Completeness. Completeness, similar to accuracy, is defined for each class [7, 8].

Completeness tells us what proportion of data from the validation set belonging to the class was correctly defined (as belonging to the Class) [7, 8].

Completeness can be represented as [7, 8]:

$$Recall = \frac{TP}{TP + FN}. \quad (3)$$

F1 Rating. The F1 score helps to achieve a balance between accuracy and completeness in order to get an average idea of how the Model Works [7, 8]. The F1 score as an indicator is calculated as follows:

$$F1\,Score = \frac{2 \cdot (Precision \times Recall)}{Precision + Recall}. \quad (4)$$

Relative error. The relative error in this case will be calculated as the ratio of the classification error to the total sample size (the number of its instances).

$$E = \frac{error_{class}}{Number_{sampl}} \cdot 100\%. \quad (5)$$

## 5 RESULTS

A comparison of MNN test results with other neural networks when classifying selected data samples is shown in Table 4.

Table 4 – General test results

| Dataset | ANN | Model size in memory (KB) | Model error on test data |
|---|---|---|---|
| CIFAR-10 | VGG-16 | 78.410 | 0.067 |
| | VGG-Pruned | 28.200 | 0.066 |
| | **MNN** | 806 | 0.079 |
| CIFAR-100 | VGG-16 | 78.590 | 0.295 |
| | VGG-Pruned | 28.910 | 0.252 |
| | ResNet | 141.100 | 0.192 |
| | **MNN** | 832 | 0.209 |
| SVHN | ResNet | 11.000 | 0.016 |
| | MNN | 522 | 0.018 |
| | EDEN | — | 0.117 |
| EMNIST | **MNN** | 363 | 0.078 |
| | VGG-16 | 528.000 | 0.076 |
| ImageNet 2012 (subset) | ResNet | 84.000 | 0.081 |
| | **MNN** | 1.872 | 0.124 |
| | VGG-16 | 528.120 | 0.295 |
| ImageNet 2012 | ResNet-34 | 84.100 | 0.276 |
| | **MNN** | 2.515 | 0.313 |

## 6 DISCUSSION

The MNN architecture was compared with several existing architectures.

MNN has been shown to have the lowest memory requirements and the number of iterations. This performance boost is achieved with a small test error cost. ResNet and VGG contain 54 and 16 layers, respectively. In some tests, it was also recommended to use a shortened and quantum version of the VGG architecture: VGG-Pruned. MNN also equaled ResNet. For the EMNIST dataset, we use EDEN for comparison.

These architectures contain large deep neural networks with inverted bottleneck filters to reduce the number of operations.

Table 4 shows that the MNN tree has the smallest model size. Compared to VGG-Pruned on the CIFAR – 100, MNN requires a 97.12% smaller model. Smaller models require fewer memory accesses, achieve faster convergence, and reduce power consumption. The table shows the number of floating-point multiplications and additions performed during the output of a single image. The specified model size and the number of MNN operations are the sum of module values along the longest run path.

There is a slight difference in memory requirements and number of operations when comparing the MNN for ImageNet 2012 (subset) and the entire ImageNet 2012 dataset. this shows the scalability of the proposed method when building using the methods presented in Section 1. the table does not report the model size and number of operations for EDEN because the data and source code are not publicly available.

Table 4 shows that MNN achieves the lowest error of 7.8% for the EMNIST dataset. The accuracy of MNN is comparable to the state of the art for the CIFAR-10 and SVHN datasets. It is worth noting that the test error achieved in modern monolithic deep neural network architectures is achieved after significant efforts to configure hyperparameters.

## CONCLUSIONS

The actual scientific and applied problem of developing a method for synthesizing MNN based on machine learning methods with reinforcement, for constructing high-precision neuromodels for solving classification problems.

**The scientific novelty** lies in the fact that a method has been developed implemented and investigated on the example of neuromodel synthesis based on a modular neural network for image classification, which can later be used as a model for technical diagnostics.

**The practical significance** lies in the fact that the using the developed method significantly reduces the resource intensity of setting up hyperparameters.

**Prospects for further research** may lie in using the parallel capacities of GPU-based computing systems to organize directly modular neural networks based on them.


DOI 10.15588/1607-3274-2024-2-8

## ACKNOWLEDGEMENTS

The work was carried out with the support of the state budget research project of the state budget of the National University "Zaporozhzhia Polytechnic" "Intelligent information processing methods and tools for decision-making in the military and civilian industries" (state registration number 0121U107499).

## REFERENCES

1. Kuo C. Transfer Learning for Image Classification: With Python Examples. New York, Kindle, 2022, 127 p.
2. Borra S., Thanki R., Dey N. Satellite Image Analysis: Clustering and Classification (SpringerBriefs in Applied Sciences and Technology). Cham, Springer, 2019, 113 p.
3. Siahaan V. Step by Step Tutorial image classification Using Scikit-Learn, Keras, And TensorFlow with Python. New York, Kindle, 2021, 141 p.
4. Canty M. J. Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for Python. New York, CRC Press, 2019, 530 p.
5. Image Classification [Electronic resource]. Mode of access: https://www.sciencedirect.com/topics/engineering/image-classification
6. Image Classification Explained [+V7 Tutorial] [Electronic resource]. Mode of access: https://www.v7labs.com/blog/image-classification-guide#h4
7. Image Classification in AI: How it works [Electronic resource]. Mode of access: https://levity.ai/blog/image-classification-in-ai-how-it-works
8. Image Classification [Electronic resource]. Mode of access: https://huggingface.co/tasks/image-classification
9. Dayhoff J. Neural Network Architectures: An Introduction. New York, Van Nostrand Reinhold Company, 1990, 259 p.
10. Aggarwal C.C. Neural Networks and Deep Learning: A Textbook. Cham, Springer, 2018, 520 p.
11. Rashid T. Make Your Own Neural Network. CreateSpace Independent Publishing Platform, 2016, 222 p.
12. Jones H. Neural Networks: An Essential Beginners Guide to Artificial Neural Networks and their Role in Machine Learning and Artificial Intelligence. CreateSpace Independent Publishing Platform, 2018, 76 p.
13. Theobald O. Machine Learning for Absolute Beginners: A Plain English Introduction (Machine Learning with Python for Beginners). London, Scatterplot Press, 2020, 181 p.
14. Sebastian C. Machine Learning for Beginners: Absolute Beginners Guide, Learn Machine Learning and Artificial Intelligence from Scratch. New York, Kindle, 2018, 102 p.
15. Julian D. Designing Machine Learning Systems with Python. Birmingham, Packt Publishing, 2016, 232 p.
16. Burkov A. The Hundred-Page Machine Learning Book. New York, Kindle, 2022, 160 p.
17. Modular Neural Networks for Low-Power Image Classification on Embedded Devices [Electronic resource]. Mode of access: https://dl.acm.org/doi/10.1145/3408062
18. Reasoning Using Modular Neural Networks [Electronic resource]. Mode of access: https://towardsdatascience.com/reasoning-using-modular-neural-networks-f003cb6109a2
19. What is a Modular Neural Network? Weak vs. Strong Learners [Electronic resource]. Mode of access: https://caseguard.com/articles/what-is-a-modular-neural-network-weak-vs-strong-learners/
20. Modular neural networks using multiple gradients [Electronic resource]. Mode of access: http://essay.utwente.nl/82225/
21. The CIFAR-10 dataset [Electronic resource]. Mode of access: https://www.cs.toronto.edu/~kriz/cifar.html
22. The CIFAR-100 dataset [Electronic resource]. Mode of access: https://www.cs.toronto.edu/~kriz/cifar.html
23. CIFAR-100 [Electronic resource]. Mode of access: https://paperswithcode.com/dataset/cifar-100
24. The Street View House Numbers (SVHN) Dataset [Electronic resource]. Mode of access: http://ufldl.stanford.edu/housenumbers/
25. The EMNIST Dataset [Electronic resource]. Mode of access: https://www.nist.gov/itl/products-and-services/emnist-dataset
26. ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [Electronic resource]. Mode of access: https://www.image-net.org/challenges/LSVRC/2012/
27. Caltech 256 Image Dataset [Electronic resource]. Mode of access: https://www.kaggle.com/datasets/jessicali9530/caltech256

Received 18.03.2024.
Accepted 01.05.2024.

УДК 004.896

# ВИКОРИСТАННЯ МОДУЛЯРНИХ НЕЙРОННИХ МЕРЕЖ ТА МАШИННОГО НАВЧАННЯ З ПІД-КРІПЛЕННЯМ ДЛЯ ВИРІШЕННЯ ЗАДАЧ КЛАСИФІКАЦІЇ

**Леощенко С. Д. –** д-р філософії, ст. викладач кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя Україна.

**Олійник А. О.** – д-р техн. наук, професор, професор кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

**Субботін С. О. –** д-р техн. наук, професор, завідувач кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

**Колпакова Т.О. –** канд. техн. наук, доцент кафедри програмних засобів Національного університету «Запорізька політехніка», Запоріжжя, Україна.

## АНОТАЦІЯ

**Актуальність.** Розглянуто вирішення задачі класифікації (в тому числі графічних даних) на основі використання модулярних нейронних мереж та модифікованих методів машинного навчання з підкріпленням для синтезу нейромоделей, які відрізняються високим рівнем точності роботи. Об'єктом дослідження є процес синтезу модулярних нейронних мереж на основі методів машинного навчання з підкріпленням.


DOI 10.15588/1607-3274-2024-2-8

**Мета роботи** полягає у розробці методу синтезу модулярних нейронних мереж на основі методів машинного навчання з підкріпленням, для побудови нейромоделей високої точності для розв'язання задач класифікації.

**Метод.** Запропоновано метод синтезу імпульсних нейронних мереж на основі еволюційного підходу. На початку, після ініціалізації системи модулярних нейронних мереж, що побудована за принципом знизу вгору, подаються вхідні дані – навчальний набір даних з вибірки і гіперпараметр, для вибору розміру кожного модуля. Результатом роботи методу є навчена система модулярних нейронних мереж. Процес починають з однієї супергрупи, яка містить усі категорії набору даних. Потім обирається розмір мережі. Вихідна матриця softmax, подібна для навченої мережі. Після чого середня ймовірність softmax використовується як показник подібності для групових категорій. Якщо формуються нові дочірні супергрупи, модуль навчається класифікації між новими супергрупами. Цикл навчання модулів модулярних нейронних мереж повторюється до тих пір, поки не будуть навчені модулі всіх супергруп. Метод дозволяє підвищити точність результуючої моделі.

**Результати.** Розроблений метод реалізовано та досліджено на прикладі синтезу нейромоделі на основі модулярної нейронної мережі для класифікації зображень, яка в подальшому зможе використовуватися у якості моделі для технічного діагностування. Використання розробленого методу значно знижує ресурсоємність налаштування гіперпараметрів.

**Висновки.** Проведені експерименти підтвердили працездатність запропонованого методу синтезу нейромоделі для класифікації зображень та дозволяють рекомендувати його для використання на практиці при синтезі модулярних нейронних мереж у якості основи класифікаційних моделей для подальшої автоматизації задач технічного діагностування та розпізнавання образів з використанням великих даних. Перспективи подальших досліджень можуть полягати у використанні паралельних потужностей обчислювальних систем на базі GPU для організації на їх базі безпосередньо модульних нейромереж.

**КЛЮЧОВІ СЛОВА:** модулярні нейронні мережі, класифікація зображень, синтез, діагностування, топологія, штучний інтелект, навчання з підкріпленням.

## ЛІТЕРАТУРА

1. Kuo C. Transfer Learning for Image Classification: With Python Examples / C. Kuo. – New York : Kindle, 2022. – 127 p.
2. Borra S. Satellite Image Analysis: Clustering and Classification (SpringerBriefs in Applied Sciences and Technology) / S. Borra, R. Thanki, N. Dey. – Cham : Springer, 2019. – 113 p.
3. Siahaan V. Step by Step Tutorial image classification Using Scikit-Learn, Keras, And TensorFlow with Python GUI / V. Siahaan. – New York : Kindle, 2021. – 141 p.
4. Canty M. J. Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for Python / M. J. Canty. – – New York : CRC Press, 2019. – 530 p.
5. Image Classification [Electronic resource]. – Access mode: https://www.sciencedirect.com/topics/engineering/image-classification
6. Image Classification Explained [+V7 Tutorial] [Electronic resource]. – Access mode: https://www.v7labs.com/blog/image-classification-guide#h4
7. Image Classification in AI: How it works [Electronic resource]. – Access mode: https://levity.ai/blog/image-classification-in-ai-how-it-works
8. Image Classification [Electronic resource]. – Access mode: https://huggingface.co/tasks/image-classification
9. Dayhoff J. Neural Network Architectures: An Introduction / J. Dayhoff. – New York : Van Nostrand Reinhold Company, 1990. – 259 p.
10. Aggarwal C. C. Neural Networks and Deep Learning: A Textbook / C. C. Aggarwal. – Cham : Springer, 2018. – 520 p.
11. Rashid T. Make Your Own Neural Network / T. Rashid. – Scotts Valley : CreateSpace Independent Publishing Platform, 2016. – 222 p.
12. Jones H. Neural Networks: An Essential Beginners Guide to Artificial Neural Networks and their Role in Machine Learning and Artificial Intelligence / H. Jones. – Scotts Valley : CreateSpace Independent Publishing Platform, 2018. – 76 p.
13. Theobald O. Machine Learning for Absolute Beginners: A Plain English Introduction (Machine Learning with Python for Beginners) / O. Theobald. – London : Scatterplot Press, 2020. – 181 p.
14. Sebastian C. Machine Learning for Beginners: Absolute Beginners Guide, Learn Machine Learning and Artificial Intelligence from Scratch / C. Sebastian. – New York : Kindle, 2018. – 102 p.
15. Julian, D. Designing Machine Learning Systems with Python / D. Julian. – Birmingham : Packt Publishing, 2016. – 232 p.
16. Burkov, A. The Hundred-Page Machine Learning Book / A. Burkov. – New York : Kindle, 2022. – 160 p.
17. Modular Neural Networks for Low-Power Image Classification on Embedded Devices [Electronic resource]. – Access mode: https://dl.acm.org/doi/10.1145/3408062
18. Reasoning Using Modular Neural Networks [Electronic resource]. – Access mode: https://towardsdatascience.com/reasoning-using-modular-neural-networks-f003cb6109a2
19. What is a Modular Neural Network? Weak vs. Strong Learners [Electronic resource]. – Access mode: https://caseguard.com/articles/what-is-a-modular-neural-network-weak-vs-strong-learners/
20. Modular neural networks using multiple gradients [Electronic resource]. – Access mode: http://essay.utwente.nl/82225/
21. The CIFAR-10 dataset [Electronic resource]. – Access mode: https://www.cs.toronto.edu/~kriz/cifar.html
22. The CIFAR-100 dataset [Electronic resource]. – Access mode: https://www.cs.toronto.edu/~kriz/cifar.html
23. CIFAR-100 [Electronic resource]. – Access mode: https://paperswithcode.com/dataset/cifar-100
24. The Street View House Numbers (SVHN) Dataset [Electronic resource]. – Access mode: http://ufldl.stanford.edu/housenumbers/
25. The EMNIST Dataset [Electronic resource]. – Access mode: https://www.nist.gov/itl/products-and-services/emnist-dataset
26. ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) [Electronic resource]. – Access mode: https://www.image-net.org/challenges/LSVRC/2012/
27. Caltech 256 Image Dataset [Electronic resource]. – Access mode: https://www.kaggle.com/datasets/jessicali9530/caltech256


DOI 10.15588/1607-3274-2024-2-8

UDC 004.93

# FACE RECOGNITION USING THE TEN-VARIATE PREDICTION ELLIPSOID FOR NORMALIZED DATA BASED ON THE BOX-COX TRANSFORMATION

**Prykhodko S. B. –** Dr. Sc., Professor, Head of the Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

**Trukhov A. S.** – Post-graduate student of the Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine.

## ABSTRACT

**Context.** Face recognition, which is one of the tasks of pattern recognition, plays an important role in the modern information world and is widely used in various fields, including security systems, access control, etc. This makes it an important tool for security and personalization. However, the low probability of identifying a person by face can have negative consequences, so there is a need for the development and improvement of face recognition methods. The object of research is the face recognition process. The subject of the research is a mathematical model for face recognition.

One of the frequently used methods of pattern recognition is the construction of decision rules based on the prediction ellipsoid. An important limitation of its application is the need to fulfill the assumption of a multivariate normal distribution of data. However, in many cases, the multivariate distribution of real data may deviate from normal, which leads to a decrease in the probability of recognition. Therefore, there is a need to improve mathematical models that would take into account the specified deviation.

**The objective** of the work is to increase the probability of face recognition by constructing a ten-variate prediction ellipsoid for data normalized by the Box-Cox transformation.

**Method.** Application of the Mardia test to test the deviation of a multivariate distribution of data from normality. Building decision rules for face recognition using a ten-variate prediction ellipsoid for data normalized based on the Box-Cox transformation. Obtaining estimates of the parameters of the univariate and ten-variate Box-Cox transformations using the maximum likelihood method.

**Results.** A comparison of the results of face recognition using decision rules, which were built using a ten-variate ellipsoid of prediction for data normalized by various transformations, was carried out. In comparison with the use of univariate normalizing transformations (decimal logarithm and Box-Cox) and the absence of normalization, the use of the ten-variate Box-Cox transformation leads to an increase in the probability of face recognition.

**Conclusions.** For face recognition, a mathematical model in the form of a ten-variate prediction ellipsoid for data normalized using the multivariate Box-Cox transformation has been improved, which allows to increase in the probability of recognition in comparison with the use of corresponding models that are built either without normalization or with the use of univariate normalizing transformations. It was found that a mathematical model built for normalized data using a multivariate Box-Cox transformation has a higher probability of recognition since univariate transformations neglect the correlation between geometric features of the face.

**KEYWORDS:** face recognition, prediction ellipsoid, multivariate Box-Cox transformation, normalizing transformation.

## ABBREVIATIONS

BCT is the Box-Cox transformation;

SMD is the squared Mahalanobis distance;

PRFP is the probability of recognizing the first person.

## NOMENCLATURE

$k$ is a number of variables (geometrical facial features);

$m$ is a number of degrees of freedom;

$N$ is a number of data points;

$\mathbf{S_Z}$ is a sample covariance matrix for normalized data;

$\mathbf{X}$ is a non-Gaussian random vector;

$\overline{\mathbf{X}}$ is a vector of sample means of the $X_j$ variables;

$X_j$ is a $j$-th non-Gaussian variable;

$\overline{X}_j$ is a sample mean of the $X_j$ values;

$\mathbf{Z}$ is a Gaussian random vector;

$\overline{\mathbf{Z}}$ is a vector of sample means of the $Z_j$ variables;

$Z_j$ is a $j$-th Gaussian variable that is obtained by transforming the variable;

$\overline{Z}_j$ is a sample mean of the $Z_j$ values;

$\alpha$ is a significance level;

$\beta_1$ is a multivariate skewness;

$\beta_2$ is a multivariate kurtosis;

$\chi^2_{m,\alpha}$ is the Chi-Square distribution quantile with $m$ degrees of freedom and significance level $\alpha$;

$\boldsymbol{\psi}$ is a vector of multivariate normalizing transformation;

$\boldsymbol{\Theta}$ is a $k$-variate vector of normalizing transformation parameters.

## INTRODUCTION

Facial recognition is becoming increasingly popular due to its wide range of applications in fields like computer vision, security systems, and others. The process of facial recognition involves automatically identifying individuals based on distinct facial features, including the shape of the eyes, nose, mouth, and other characteristics. The technology behind facial recognition is continually advancing leading to higher accuracy and opening up a multitude of possibilities for its use in various aspects of life.


DOI 10.15588/1607-3274-2024-2-9

The accuracy and effectiveness of facial recognition systems are heavily reliant on the specific decision rule chosen for the identification process. This decision rule essentially dictates how an individual's facial features are classified into predefined categories or classes within the system.

In modern methods used for face recognition, an important limitation is the assumption of a multivariate normal distribution of the data [1]. However, real data often have a non-Gaussian distribution. As a result, such deviations can be the cause of errors in the face recognition process. Therefore, there is a need to improve mathematical models that can take into account deviations from the normal distribution of data.

**The object of study** is the process of face recognition.

The facial recognition process involves a series of key steps, it initiates with image preprocessing, involving face detection, and alignment for optimal analysis. Feature extraction follows, identifying key facial elements such as the position of the eyes, nose, mouth, and other distinctive attributes. These extracted features serve as the foundation for generating a feature vector, a mathematical representation encapsulating the unique facial characteristics, which is pivotal in the recognition process. Pattern recognition, through the application of mathematical models, determines which individual the feature vector corresponds to. This process involves comparing the feature vector to a database of known individuals [2].

**The subject of study** is a mathematical model for face recognition. One of the frequently employed methods in pattern recognition involves building decision rules based on prediction ellipsoids.

**The purpose of the work** is to increase the probability of face recognition by constructing a ten-variate prediction ellipsoid for normalized data using Box-Cox transformation.

## 1 PROBLEM STATEMENT

Suppose given the original data sample set of the ten geometrical facial features the multivariate distribution for which is not Gaussian. Suppose that there are bijective ten-variate normalizing transformation $\boldsymbol{\psi} = \{\psi_Y, \psi_1, \psi_2, \ldots, \psi_{10}\}^T$ of non-Gaussian random vector $\mathbf{X} = \{X_1, X_2, \ldots, X_{10}\}^T$ to Gaussian random vector $\mathbf{Z} = \{Z_1, Z_2, \ldots, Z_{10}\}^T$ is given by:

$$\mathbf{Z} = \boldsymbol{\psi}(\mathbf{X}) \tag{1}$$

and the inverse transformation for (1)

$$\mathbf{X} = \boldsymbol{\psi}^{-1}(\mathbf{Z}). \tag{2}$$

It is required to build the prediction ellipsoid for normalized data in the form:

$$(\mathbf{Z} - \overline{\mathbf{Z}})^T \mathbf{S}_{\mathbf{Z}}^{-1} (\mathbf{Z} - \overline{\mathbf{Z}}) = \chi^2_{m,\alpha}, \tag{3}$$

where

$$\mathbf{S}_{\mathbf{Z}} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{Z}_i - \overline{\mathbf{Z}})(\mathbf{Z}_i - \overline{\mathbf{Z}})^T .$$

Also, it is required to develop the decision rule for face recognition based on equation (3) and the transformations (1) and (2).

## 2 REVIEW OF THE LITERATURE

Mahalanobis distance is a way of measuring how far a point is from a distribution of points, taking into account the shape and orientation of the distribution. It is based on the idea that the distance between two points should be scaled by the variance and covariance of the variables involved. The squared Mahalanobis distance (SMD) is widely used in statistics and multivariate data analysis for pattern recognition [3], classification [4, 5], and outlier detection [6–8]. This metric provides a valuable measure for assessing the relationships between data points in various applications.

The value of the SMD is approximately equal to the value of the Chi-squared distribution with *k* degrees of freedom, which is equal to the number of characteristics [9, 10]. Recognition takes place with the help of a prediction ellipsoid, which defines the space of allowed values such that all elements within the same class are inside the ellipsoid, while others are outside it.

Many statistical procedures assume that the variables are normally distributed, and an assumption of homoscedasticity or homogeneity of variance. Significant violations of either assumption can increase the chances of committing either a type I or II error. Rectifying these issues through data transformations can significantly improve analysis accuracy [11].

The construction of the ellipsoid relies on the assumption that the data follows a multivariate normal distribution [12]. However, real data may have a non-normal multivariate distribution, which leads to a lower recognition probability. To solve certain practical problems, which are based on the use of the Mahalanobis distance, in the case of non-Gaussian data, normalization is used [13, 14]. Its application allows solving the corresponding problems for data whose multivariate distribution deviates from normal.

In [15], a decision rule for pattern recognition was improved based on the application of the SMD for normalized data from 10 characteristics using a decimal logarithm transformation. This allowed to increase in the probability of recognition, but when using the decimal logarithm, the probability of recognition is not always satisfactory, so it is necessary to apply other normalizing transformations, such as the Box-Cox transformation (BCT).


DOI 10.15588/1607-3274-2024-2-9

## 3 MATERIALS AND METHODS

To create feature vectors, a special program was developed in Python using the Dlib computer vision library. After detecting a face in the input image, the program performs several image processing steps, including cropping and aligning the face so that the eyes are at the same level. Such processing helps to remove some of the distortions caused by the position of the face in the input image [16]. In the last step, the program obtains a set of characteristics from the aligned image. Each feature is the pixel distance between facial landmarks defined by the Dlib library.

After analyzing the studies [17–19], 17 key landmarks of the face were identified. Using the pixel distances between these landmarks, a vector consisting of 10 features was constructed. The symmetrical distances were averaged, resulting in the following features: $X_1$ – the average distance from the eyes to the middle of the nose, $X_2$ – the average distance from the eyes to the center of the mouth, $X_3$ – the average distance from the eyes to the center of the eyebrows, $X_4$ – the average distance from the eyebrows to the top of the nose, $X_5$ – the average distance from the corners of the eyes to the top of the nose, $X_6$ – the distance between the eyebrows, $X_7$ – the distance between the nose and the middle of the mouth, $X_8$ – the distance between the corners of the mouth, $X_9$ – the distance between the edges of the nose, $X_{10}$ – the distance from the mouth to the chin.

To account for variations in the position of the face in the image and different distances to the camera, a normalization process is used by dividing each feature by the distance between the eyes [20].

In the final version, the vector takes the form: $\mathbf{X}$ ={(v1+v2)/2d, (v3+v4)/2d, (v5+v6)/2d, (v7+v8)/2d, (v9+v10)/2d, v11/d, v12/d, v13/d, v14/d, v15/d}.

A dataset from work [15] containing 200 photos of two people was chosen, where 100 photos are used to build a prediction ellipsoid for recognizing the first person, and 300 are used for testing.

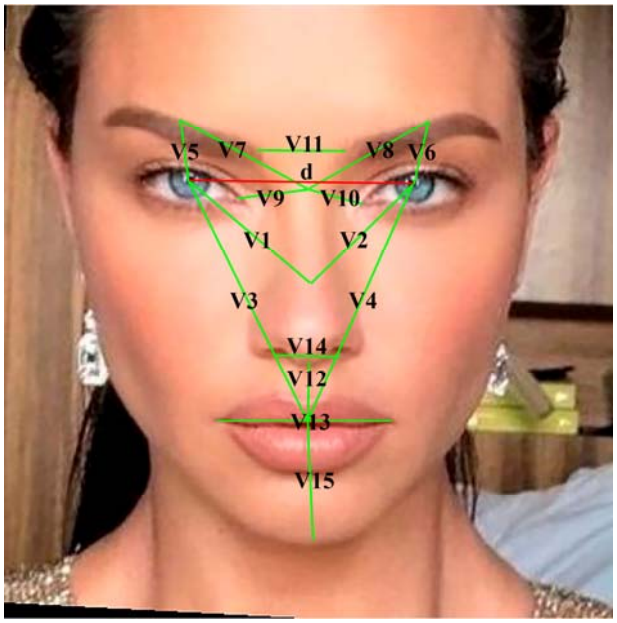



Figure 1 – Distances between face key points used for recognition

As a result, 400 feature vectors were obtained, one per photo, consisting of 10 elements.

A vector of first-person sample means $\overline{\mathbf{X}} = \{\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_{10}\}^T$ to construct a prediction ellipsoid: $\overline{\mathbf{X}}$ = {0.6330; 1.1629; 0.3010; 0.6259; 0.2994; 0.3670; 0.3015; 0.8011; 0.3584; 0.6290}, the covariance matrix is shown in Table 1, and the characteristic ranges in Table 2.

Table 1 – Covariance matrix of the initial sample

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.00101 | 0.00089 | –0.00060 | –0.00006 | –0.00012 | –0.00011 | –0.00052 | 0.00001 | 0.00034 | –0.00113 |
| 0.00089 | 0.00200 | 0.00000 | 0.00026 | –0.00016 | 0.00006 | 0.00052 | –0.00004 | 0.00046 | 0.00024 |
| –0.00060 | 0.00000 | 0.00111 | 0.00047 | 0.00001 | 0.00019 | 0.00071 | –0.00009 | –0.00009 | 0.00121 |
| –0.00006 | 0.00026 | 0.00047 | 0.00039 | –0.00003 | 0.00017 | 0.00031 | –0.00027 | –0.00006 | 0.00034 |
| –0.00012 | –0.00016 | 0.00001 | –0.00003 | 0.00008 | 0.00002 | 0.00003 | 0.00000 | –0.00010 | 0.00008 |
| –0.00011 | 0.00006 | 0.00019 | 0.00017 | 0.00002 | 0.00073 | 0.00027 | 0.00003 | –0.00002 | 0.00002 |
| –0.00052 | 0.00052 | 0.00071 | 0.00031 | 0.00003 | 0.00027 | 0.00119 | –0.00015 | –0.00007 | 0.00149 |
| 0.00001 | –0.00004 | –0.00009 | –0.00027 | 0.00000 | 0.00003 | –0.00015 | 0.00276 | 0.00100 | 0.00129 |
| 0.00034 | 0.00046 | –0.00009 | –0.00006 | –0.00010 | –0.00002 | –0.00007 | 0.00100 | 0.00104 | 0.00016 |
| –0.00113 | 0.00024 | 0.00121 | 0.00034 | 0.00008 | 0.00002 | 0.00149 | 0.00129 | 0.00016 | 0.00416 |

Table 2 – Ranges of characteristics of the initial sample

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Min | 0.55734 | 1.07032 | 0.23998 | 0.57780 | 0.27292 | 0.31088 | 0.22610 | 0.69870 | 0.29332 | 0.46811 |
| Max | 0.70846 | 1.28726 | 0.39580 | 0.66983 | 0.32167 | 0.43156 | 0.39956 | 1.00075 | 0.46041 | 0.80951 |


DOI 10.15588/1607-3274-2024-2-9

The Mardia test was used to assess the deviation of the multivariate data distribution from normality. It is based on the analysis of the multivariate skewness $\beta_1$ and kurtosis $\beta_2$ of the data, which are indicators of how much the data deviate from the normal distribution and are calculated according to the following formulas:

$$\beta_1 = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)^T \mathbf{S}_{\mathbf{X}}^{-1}\left(\mathbf{X}_j - \overline{\mathbf{X}}\right)\right\}^3, \quad (4)$$

$$\beta_2 = \frac{1}{N}\sum_{j=1}^{N}\left\{\left(\mathbf{X}_j - \overline{\mathbf{X}}\right)^T \mathbf{S}_{\mathbf{X}}^{-1}\left(\mathbf{X}_j - \overline{\mathbf{X}}\right)\right\}^2, \quad (5)$$

where

$$\mathbf{S}_{\mathbf{X}} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)\left(\mathbf{X}_i - \overline{\mathbf{X}}\right)^T.$$

According to the Mardia test, the multivariate distribution of the received sample is not Gaussian since the test statistic for multivariate skewness $N\beta_1/6$ of the data, which equals 289.20, is greater than the quantile of the Chi-Square distribution, which is 277.77 for 220 degrees of freedom and 0.005 significance level. In contrast, the test statistic for multivariate kurtosis $\beta_2$, which equals 122.35, does not exceed the value of the Gaussian distribution quantile, which is 127.97 for the mean of 120, the variance of 9.6, and a significance level of 0.005. That is why, there is a need to apply a normalizing transformation (1).

The original BCT is a univariate transformation with one parameter $\lambda$ and is applied element-wise to a vector. For multivariate data, it is usually applied $k$ times as univariate mapping to each column with different values for $\lambda$. Therefore, the overall transformation is specified by a $k$-variate vector $\boldsymbol{\Theta} = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$ [21].

As in [22], normalization by the BCT is given by:

$$Z_j = x(\lambda_j) = \begin{cases} \left(X_j^{\lambda_j} - 1\right)\Big/\lambda_j, \lambda_j \neq 0; \\ \ln(X_j), \quad \lambda_j = 0. \end{cases} \quad (6)$$

The main task when using the method is to find the optimal value of the input parameter in such a way that, as a result of the transformation, the distribution of the output value is as close as possible to the normal one. The most popular method of finding the optimal value of the lambda parameter is the maximum likelihood estimation:

$$l(\lambda) = C - \frac{N}{2}\ln\sum_{i=1}^{N}\frac{\left(x(\lambda)_i - \overline{x(\lambda)}\right)^2}{N} + (\lambda - 1)\sum_{i=1}^{N}\ln(x_i). \quad (7)$$

The multivariate Box-Cox method uses a separate transformation parameter for each variable. When variables are transformed to joint normality, they become approximately linearly related, constant in conditional variance, and marginally normal in distribution. In the case of using the ten-dimensional BCT, the components of the vector **T** are defined as (6).

For the ten-variate BCT, the log-likelihood function can be written as:

$$l(X, \theta) = \sum_{j=1}^{k}(\lambda_j - 1)\sum_{i=1}^{N}\ln(x_{ji}) - \frac{N}{2}\ln\left[\det(\mathbf{S}_{\mathbf{Z}})\right]. \quad (8)$$

After applying normalizing transformations, a ten-variate prediction ellipsoid is built based on (3):

$$\left(\mathbf{Z} - \overline{\mathbf{Z}}\right)^T \mathbf{S}_{\mathbf{Z}}^{-1}\left(\mathbf{Z} - \overline{\mathbf{Z}}\right) = \chi^2_{10,\,0.005}. \quad (9)$$

The value of the quantile of the Chi-square distribution is 25.19 for 10 degrees of freedom and a significance level of 0.005. The decision rule is based on a prediction ellipsoid (9), which describes the space of admissible values for each class such that all objects of one class must lie within the bounds of this ellipsoid, and of another class – outside the bounds.

## 4 EXPERIMENTS

For comparison, two prediction ellipsoids are built based on the data from [14] and two normalization transformations: the univariate BCT and the ten-variate BCT.

A univariate BCT is applied to the initial sample. As a result of solving the task using the maximum likelihood method of the logarithmic function (7), the following parameter estimates were obtained: $\hat{\lambda}_1 = 1.7451$, $\hat{\lambda}_2 = -4.5493$, $\hat{\lambda}_3 = -0.7145$, $\hat{\lambda}_4 = 0.3643$, $\hat{\lambda}_5 = 5.1055$, $\hat{\lambda}_6 = -0.8785$, $\hat{\lambda}_7 = -0.4222$, $\hat{\lambda}_8 = -2.7221$, $\hat{\lambda}_9 = -1.8611$, $\hat{\lambda}_{10} = 1.0102$.

As a result of the application of the univariate BCT with components (6), where each element of the vector **T** is calculated independently of the others, a sample with the following vector of means $\overline{\mathbf{Z}} = \left\{\overline{Z}_1, \overline{Z}_2, \ldots, \overline{Z}_{10}\right\}^T$ was obtained: $\overline{\mathbf{Z}}$ = {–0.31461; 0.10718; –1.92555; –0.43103; –0.19545; –1.62004; –1.57614; –0.31836; –3.16345; –0.37015}. The covariance matrix of the sample in Table 3, ranges of characteristics in Table 4.

The normalized sample obtained as a result of applying the univariate BCT does not deviate from the multivariate normal distribution, because the test statistic for multivariate skewness $N\beta_1/6$, which equals 276.51, does not exceed the critical value 277.77; the test statistic for multivariate kurtosis $\beta_2$, which equals 120.4, is less than the critical value of 127.97.

Applying the ten-variate BCT to normalize the initial sample. As a result of solving the task using the maximum likelihood method of the logarithmic function (8), the following parameter estimates were obtained:


DOI 10.15588/1607-3274-2024-2-9

$\hat{\lambda}_1 = -0.5799$, $\hat{\lambda}_2 = -0.9732$, $\hat{\lambda}_3 = 0.4871$, $\hat{\lambda}_4 = 2.888$, $\hat{\lambda}_5 = 4.0714$, $\hat{\lambda}_6 = -0.231$, $\hat{\lambda}_7 = 0.087$, $\hat{\lambda}_8 = -1.2973$, $\hat{\lambda}_9 = -1.1866$, $\hat{\lambda}_{10} = 1.3321$.

As a result of the application of the ten-variate BCT with components (6), a sample with the following vector of means $\overline{\mathbf{Z}} = \{\overline{Z}_1, \overline{Z}_2, \ldots, \overline{Z}_{10}\}^T$ was obtained: $\overline{\mathbf{Z}}$ = {–0.52627; 0.13911; –0.91083; –0.25654; –0.24379; –1.13224; –1.14383; –0.26329; –2.03304; –0.34494}. The covariance matrix of the sample in Table 5, ranges of characteristics in Table 6.

The normalized sample by using the ten-variate BCT does not deviate from the multivariate normal distribution, because the test statistic for multivariate skewness $N\beta_1/6$, which equals 265.59, does not exceed the critical value 277.77; the test statistic for multivariate kurtosis $\beta_2$, which equals 121.52, is less than the critical value of 127.97.

Table 3 – Covariance matrix of the sample normalized by univariate Box-Cox

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.00051 | 0.00027 | –0.00329 | –0.00006 | 0.00000 | –0.00054 | –0.00197 | 0.00006 | 0.00437 | –0.00080 |
| 0.00027 | 0.00035 | –0.00016 | 0.00014 | 0.00000 | 0.00022 | 0.00115 | –0.00006 | 0.00294 | 0.00005 |
| –0.00329 | –0.00016 | 0.06670 | 0.00489 | 0.00000 | 0.00923 | 0.02936 | –0.00234 | –0.01845 | 0.00937 |
| –0.00006 | 0.00014 | 0.00489 | 0.00070 | 0.00000 | 0.00149 | 0.00234 | –0.00085 | –0.00193 | 0.00046 |
| 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | –0.00001 | 0.00000 |
| –0.00054 | 0.00022 | 0.00923 | 0.00149 | 0.00000 | 0.03103 | 0.00950 | 0.00025 | –0.00533 | 0.00025 |
| –0.00197 | 0.00115 | 0.02936 | 0.00234 | 0.00000 | 0.00950 | 0.03558 | –0.00250 | –0.01137 | 0.00810 |
| 0.00006 | –0.00006 | –0.00234 | –0.00085 | 0.00000 | 0.00025 | –0.00250 | 0.01315 | 0.03704 | 0.00273 |
| 0.00437 | 0.00294 | –0.01845 | –0.00193 | –0.00001 | –0.00533 | –0.01137 | 0.03704 | 0.34189 | 0.00232 |
| –0.00080 | 0.00005 | 0.00937 | 0.00046 | 0.00000 | 0.00025 | 0.00810 | 0.00273 | 0.00232 | 0.00412 |

Table 4 – Ranges of characteristics of the sample normalized by univariate Box-Cox

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Min | –0.36642 | 0.05845 | –2.48076 | –0.49720 | –0.19561 | –2.03872 | –2.06852 | –0.60754 | –4.72990 | –0.53009 |
| Max | –0.25900 | 0.15013 | –1.31438 | –0.37285 | –0.19527 | –1.24335 | –1.12041 | 0.00075 | –1.73859 | –0.19029 |

Table 5 – Covariance matrix of the sample normalized by ten-variate Box-Cox

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.00436 | 0.00134 | –0.00233 | –0.00006 | –0.00001 | –0.00080 | –0.00323 | 0.00007 | 0.00616 | –0.00204 |
| 0.00134 | 0.00106 | –0.00002 | 0.00008 | 0.00000 | 0.00017 | 0.00112 | –0.00004 | 0.00281 | 0.00013 |
| –0.00233 | –0.00002 | 0.00374 | 0.00036 | 0.00000 | 0.00115 | 0.00381 | –0.00035 | –0.00198 | 0.00192 |
| –0.00006 | 0.00008 | 0.00036 | 0.00007 | 0.00000 | 0.00023 | 0.00038 | –0.00018 | –0.00027 | 0.00012 |
| –0.00001 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00000 | –0.00002 | 0.00000 |
| –0.00080 | 0.00017 | 0.00115 | 0.00023 | 0.00000 | 0.00846 | 0.00270 | 0.00012 | –0.00122 | 0.00009 |
| –0.00323 | 0.00112 | 0.00381 | 0.00038 | 0.00000 | 0.00270 | 0.01047 | –0.00090 | –0.00277 | 0.00378 |
| 0.00007 | –0.00004 | –0.00035 | –0.00018 | 0.00000 | 0.00012 | –0.00090 | 0.00705 | 0.01390 | 0.00175 |
| 0.00616 | 0.00281 | –0.00198 | –0.00027 | –0.00002 | –0.00122 | –0.00277 | 0.01390 | 0.08567 | 0.00116 |
| –0.00204 | 0.00013 | 0.00192 | 0.00012 | 0.00000 | 0.00009 | 0.00378 | 0.00175 | 0.00116 | 0.00305 |

Table 6 – Ranges of characteristics of the sample normalized by ten-variate Box-Cox

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Min | –0.69589 | 0.06576 | –1.02852 | –0.27523 | –0.24437 | –1.34124 | –1.39454 | –0.45651 | –2.76948 | –0.47758 |
| Max | –0.38152 | 0.22388 | –0.74583 | –0.23742 | –0.24319 | –0.92748 | –0.88171 | 0.00075 | –1.27277 | –0.18419 |

After data normalization by univariate and multivariate BCTs, ten-variate ellipsoids were constructed based on (9). The computer program implementing the constructed models was developed to conduct experiments. The program was written in the Python language.

## 5 RESULTS

The recognition check is based on two criteria, such as the probability of recognizing the first person (PRFP), and the probability of type II errors, which occur when the decision rule mistakenly identifies another person as person 1.

Application of (9) for normalized data with a univariate BCT to recognize 300 test photos allowed obtaining the following results: the PRFP is 94%, with a probability of type II errors being 5.5%. The application of (9) for data normalized using the ten-variate BCT allowed obtaining the PRFP of 97% with type II errors of 2.5%.

Table 7 shows a comparison of the results of using models for non-normalized data (Source); data normalized by the transformation of the decimal logarithm (Lg) [15]; normalized data using univariate BCT; normalized data using the ten-variate BCT.


DOI 10.15588/1607-3274-2024-2-9

Table 7 – Comparison of the results

| | Source | Lg | Univariate BCT | Ten-variate BCT |
|---|---|---|---|---|
| PRFP | 92% | 95% | 94% | 97% |
| Type II errors | 4.5% | 5.5% | 5.5% | 2.5% |

The use of ten-variate prediction ellipsoid (9) for data normalized using a multivariate BCT resulted in the highest probability of recognition.

## 6 DISCUSSION

As is evident from Table 7, the decision rule built for the initial data resulted in the lowest recognition probability. Application of decision rules for normalized data by decimal logarithm and univariate BCT has increased the PRFP and reduced the possibility of type II errors. The most notable enhancement in recognition accuracy was achieved with the ten-variate BCT.

Taking into account that the use of univariate transformations had a lesser impact, it is important to note that this can be explained by the fact that univariate transformations do not account for data correlation. Thus, their limitation lies in their inability to consider the interrelationships between different variables, which affect the analysis results. Unlike univariate transformations, the ten-variate BCT preserves inter-variable relationships crucial in capturing complex facial features.

## CONCLUSIONS

The important problem of increasing the probability of face recognition by constructing a ten-variate prediction ellipsoid for data normalized by the BCT is solved.

**The scientific novelty** of the obtained results is that the ten-variate prediction ellipsoid for normalized data for face recognition is firstly constructed based on the BCTs. The application of univariate BCT resulted in a slight improvement, which is explained by the fact that the method does not take into account the correlation between features. The construction prediction ellipsoid for normalized data based on ten-variate BCT allowed increased PRFP and reduced type II errors.

**The practical significance** of the obtained results is that the software realizing the constructed model is developed in the Python language. The experimental results allow us to recommend the constructed model for use in practice.

**Prospects for further research** may include the use of other multivariate normalizing transformations to construct prediction ellipsoid for face recognition.

## ACKNOWLEDGEMENTS

The work is of an initiative nature and was carried out within the scope of the authors' scientific activities during their regular working hours in their primary positions.

## REFERENCES

1. Sen S., Diawara N., Iftekharuddin K. Statistical pattern recognition using Gaussian copula, Journal of Statistical Theory and Practice, 2015, No. 9, pp. 768–777. DOI: 10.1080/15598608.2015.1008607
2. Barnouti N. H., Al-Dabbagh S. S. M., Matti W. E. Face recognition: A literature review, *International Journal of Applied Information Systems*, 2016, No.11(4), pp. 21–31. DOI: 10.5120/ijais2016451597
3. Kapoor S., Khanna S., Bhatia R. Facial gesture recognition using correlation and Mahalanobis distance, *International Journal of Computer Science and Information Security*, 2010, Vol. 7, pp. 267–272.
4. Xiang S., Nie F., Zhang C. Learning a Mahalanobis distance metric for data clustering and classification, *Pattern recognition*, 2008, No. 41(12), pp. 3600–3612. DOI: 10.1016/j.patcog.2008.05.018
5. Nader P., Honeine P., Beauseroy P. Mahalanobis-based one-class classification, *In 2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2014. pp. 1–6. IEEE. DOI: 10.1109/MLSP.2014.6958934
6. Todeschini R., Ballabio D., Consonni V., Sahigara F., Filzmoser P. Locally centred Mahalanobis distance: a new distance measure with salient features towards outlier detection, *Analytica chimica acta*, 2013, No. 787, pp. 1–9. DOI: 10.1016/j.aca.2013.04.034
7. Dashdondov K., Kim M. H. Mahalanobis distance based multivariate outlier detection to improve performance of hypertension prediction, *Neural Processing Letters*, 2021, P. 113. DOI: 10.1007/s11063-021-10663-y
8. Prykhodko S., Prykhodko N., Makarova L. et al. Application of the Squared Mahalanobis Distance for Detecting Outliers in Multivariate Non-Gaussian Data, *Radioelectronics, Telecommunications and Computer Engineering : 14th International Conference on Advanced Trends (TCSET)*, Lviv-Slavske, Ukraine, February 20–24, 2018, proceedings, P. 962–965. DOI: 10.1109/TCSET.2018.8336353
9. Etherington T. R. Mahalanobis distances for ecological niche modelling and outlier detection: implications of sample size, error, and bias for selecting and parameterising a multivariate location and scatter method, PeerJ, 2021, No. 9, e11436. DOI: 10.7717/peerj.11436
10. Ghorbani H. Mahalanobis distance and its application for detecting multivariate outliers, *Facta Universitatis, Series: Mathematics and Informatics*, 2019, pp. 583–595. DOI: 10.22190/FUMI1903583G
11. Osborne J. Improving your data transformations: Applying the Box-Cox transformation, *Practical Assessment, Research, and Evaluation*, 2010, Vol. 15(1), No. 12. DOI: 10.7275/qbpc-gk17
12. Gallego G., Cuevas C., Mohedano R., Garcia N. On the Mahalanobis distance classification criterion for multidimensional normal distributions. / [G. Gallego,] //IEEE Transactions on Signal Processing, 2013, No. 61(17), pp. 4387–4396. DOI: 10.1109/TSP.2013.2269047
13. Prykhodko S. Makarova L., Prykhodko K., Pukhalevych A. Application of transformed prediction ellipsoids for outlier detection in multivariate non-Gaussian data, *In 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, 2020, pp. 359–362. IEEE. DOI: 10.1109/TCSET49122.2020.235454
14. Meshkani S. A., Mehrabi B., Yaghubpur A., Alghalandis Y. F. The application of geochemical pattern recognition to regional prospecting: A case study of the Sanandaj-Sirjan metallogenic zone, Iran, *Journal of geochemical exploration*, 2011, No. 108(3), pp. 183–195. DOI: 10.1016/j.gexplo.2011.01.006


DOI 10.15588/1607-3274-2024-2-9

15. Prykhodko S., Trukhov A. Pobudova pravyl pryiniattia rishen dlia rozpiznavannia oblych na osnovi kvadrata vidstani Makhalanobisa dlia normalizovanykh danykh, *Information Technology: Computer Science, Software Engineering and Cyber Security*, 2023, No. 2, pp. 50–58. DOI: 10.32782/IT/2023-2-6. (in Ukraine).
16. Haghpanah M. A., Saeedizade E., Masouleh M. T., Kalhor A. Real-time facial expression recognition using facial landmarks and neural networks, *In 2022 International Conference on Machine Vision and Image Processing (MVIP)*, 2022, pp. 1–7. IEEE. DOI: 10.1109/MVIP53647.2022.9738754
17. Li D., Wang Z., Gao Q., Song Y., Yu X., Wang C. Facial expression recognition based on Electroencephalogram and facial landmark localization, *Technology and Health Care*, 2019, No. 27(4), pp. 373–387. DOI: 10.3233/THC-181538
18. Osman Ali A. S., Asirvadam V. S., Malik A. S., Eltoukhy M. M., Aziz A. Age-invariant face recognition using triangle geometric features, *International Journal of Pattern Recognition and Artificial Intelligence*, 2015, No. 29(05), pp. 1556006. DOI: 10.1142/S0218001415560066
19. Bijarnia S., Singh P. Age invariant face recognition using minimal geometrical facial features, *In Advanced Computing and Communication Technologies: Proceedings of the 9th ICACCT*, 2016, pp. 71–77. Springer Singapore. DOI: 10.1007/978-981-10-1023-1_7.
20. Johnston B., Chazal P. D. A review of image-based automatic facial landmark identification techniques, *EURASIP Journal on Image and Video Processing*, 2018, No. 1, pp. 1–23. DOI: 10.1186/s13640-018-0324-4
21. Blum L., Elgendi M., Menon C. Impact of Box-Cox Transformation on Machine-Learning Algorithms, *Frontiers in Artificial Intelligence*, 2022, No. 5, pp. 877569. DOI: 10.3389/frai.2022.877569
22. Prykhodko S. B., Shutko I. S., Prykhodko A. S. A nonlinear regression model to estimate the size of web apps created using the CakePHP framework, *Radio Electronics, Computer Science, Control*, 2021, No. 4 (59), pp. 129–139. DOI: 10.15588/1607-3274-2021-4-12

Received 06.02.2024.
Accepted 21.04.2024.

УДК 004.93

## РОЗПІЗНАВАННЯ ОБЛИЧЧЯ ЗА ДОПОМОГОЮ ДЕСЯТИВИМІРНОГО ЕЛІПСОЇДА ПРОГНОЗУВАННЯ ДЛЯ НОРМАЛІЗОВАНИХ НА ОСНОВІ ПЕРЕТВОРЕННЯ БОКСА-КОКСА ДАНИХ

**Приходько С. Б.** – д-р техн. наук, професор, завідувач кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна.

**Трухов А. С.** – аспірант кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування ім. адмірала Макарова, Миколаїв, Україна.

### АНОТАЦІЯ

**Актуальність.** Розпізнавання обличчя, яке є одним із завдань розпізнавання образів, відіграє важливу роль у сучасному інформаційному світі та знаходить широке застосування в різних галузях, включаючи системи безпеки, управління доступом та ін. Це робить його важливим інструментом для забезпечення безпеки та персоналізації. Однак низька ймовірність ідентифікації особи за обличчям може мати негативні наслідки, тому існує потреба в розробці та вдосконаленні методів розпізнавання обличчя. Об’єктом дослідження є процес розпізнавання обличчя. Предметом дослідження є математична модель для розпізнавання обличчя.

Один з часто використовуваних методів розпізнавання образів полягає в побудові правил прийняття рішень на основі еліпсоїда прогнозування. Важливим обмеженням його застосування є необхідність виконання припущення про багатовимірний нормальний розподіл даних. Однак у багатьох випадках багатовимірний розподіл реальних даних може відхилятися від нормального, що призводить до зниження ймовірності розпізнавання. Тому виникає необхідність удосконалення математичних моделей, які враховували б зазначене відхилення.

**Мета роботи** полягає у підвищенні ймовірності розпізнавання обличчя шляхом побудови десятивимірного еліпсоїду прогнозування для нормалізованих за допомогою перетворення Бокса-Кокса даних.

**Метод.** Застосування тесту Мардіа для перевірки відхилення багатовимірного розподілу даних від нормального. Побудова правил прийняття рішень для розпізнавання обличчя за допомогою десятивимірного еліпсоїду прогнозування для нормалізованих на основі перетворення Бокса-Кокса даних. Отримання оцінок параметрів одновимірного та десятививимірного перетворень Бокса-Кокса за допомогою методу максимальної правдоподібності.

**Результати.** Здійснено порівняння результатів розпізнавання облич за допомогою правил прийняття рішень, які побудовані за допомогою десятивимірного еліпсоїду прогнозування для нормалізованих за різними перетвореннями даних. У порівнянні із застосуванням одновимірних нормалізуючих перетворень (десяткового логарифму та Бокса-Кокса) та у випадку відсутності нормалізації використання десятивимірного перетворення Бокса-Кокса призводить до збільшення ймовірності розпізнавання облич.

**Висновки.** Для розпізнавання обличчя удосконалено математичну модель у вигляді десятивимірного еліпсоїду прогнозування для нормалізованих за допомогою багатовимірного перетворення Бокса-Кокса даних, що дозволяє підвищити ймовірність розпізнавання у порівнянні із застосуванням відповідних моделей, які побудовані або без нормалізації, або із використанням одновимірних нормалізуючих перетворень. Досліджено, що математична модель, побудована для нормалізованих даних за допомогою багатовимірного перетворення Бокса-Кокса, має більшу ймовірність розпізнавання за рахунок того, що одновимірні перетворення нехтують кореляцією між геометричними ознаками обличчя.

**КЛЮЧОВІ СЛОВА:** розпізнавання обличчя, еліпсоїд прогнозування, нормалізуюче перетворення, багатовимірне перетворення Бокса-Кокса, правило прийняття рішення.


DOI 10.15588/1607-3274-2024-2-9

**ЛІТЕРАТУРА**

1. Sen S. Statistical pattern recognition using Gaussian copula. / S. Sen, N. Diawara, K. Iftekharuddin // Journal of Statistical Theory and Practice. – 2015. – No. 9. – P.768–777. DOI: 10.1080/15598608.2015.1008607
2. Barnouti N. H. Face recognition: A literature review. / N. H. Barnouti, S. S. M. Al-Dabbagh, W. E. Matti // International Journal of Applied Information Systems. – 2016. – No. 11(4). – P. 21–31. DOI: 10.5120/ijais2016451597
3. Kapoor S. Facial gesture recognition using correlation and Mahalanobis distance. / S. Kapoor, S. Khanna, R. Bhatia // International Journal of Computer Science and Information Security. – 2010. – Vol. 7. – P. 267–272.
4. Xiang S. Learning a Mahalanobis distance metric for data clustering and classification. / S. Xiang, F. Nie, C. Zhang // Pattern recognition. – 2008. – No. 41(12). – P. 3600–3612. DOI: 10.1016/j.patcog.2008.05.018
5. Nader P. Mahalanobis-based one-class classification. / P. Nader, P. Honeine, P. Beauseroy // In 2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP). – 2014. – P. 1–6. IEEE. DOI: 10.1109/MLSP.2014.6958934
6. Todeschini R. Locally centred Mahalanobis distance: a new distance measure with salient features towards outlier detection. / [R. Todeschini, D. Ballabio, V. Consonni et al.] // Analytica chimica acta. – 2013. – No. 787. – P. 1–9. DOI: 10.1016/j.aca.2013.04.034
7. Dashdondov K. Mahalanobis distance based multivariate outlier detection to improve performance of hypertension prediction. / K. Dashdondov, M. H. Kim // Neural Processing Letters, 2021 113. DOI: 10.1007/s11063-021-10663-y
8. Prykhodko S. Application of the Squared Mahalanobis Distance for Detecting Outliers in Multivariate Non-Gaussian Data / [S. Prykhodko, N. Prykhodko, L. Makarova et al.] // Radioelectronics, Telecommunications and Computer Engineering : 14th International Conference on Advanced Trends (TCSET), Lviv-Slavske, Ukraine, February 20–24, 2018 : proceedings. – P. 962–965. DOI: 10.1109/TCSET.2018.8336353
9. Etherington T. R. Mahalanobis distances for ecological niche modelling and outlier detection: implications of sample size, error, and bias for selecting and parameterising a multivariate location and scatter method. / T. R. Etherington // PeerJ. – 2021. – No. 9, e11436. DOI: 10.7717/peerj.11436
10. Ghorbani H. Mahalanobis distance and its application for detecting multivariate outliers. / H. Ghorbani // Facta Universitatis, Series: Mathematics and Informatics. – 2019. – P. 583–595. DOI: 10.22190/FUMI1903583G
11. Osborne J. Improving your data transformations: Applying the Box-Cox transformation. / J. Osborne // Practical Assessment, Research, and Evaluation. – 2010. – Vol.15(1), No. 12. DOI: 10.7275/qbpc-gk17
12. Gallego G. On the Mahalanobis distance classification criterion for multidimensional normal distributions. / [G. Gallego, C. Cuevas, R. Mohedano, N. Garcia] //IEEE Transactions on Signal Processing. 2013. – No. 61(17). – P. 4387–4396. DOI: 10.1109/TSP.2013.2269047
13. Application of transformed prediction ellipsoids for outlier detection in multivariate non-Gaussian data. / [S. Prykhodko, L. Makarova, K. Prykhodko, A. Pukhalevych] // In 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). – 2020. – P. 359–362. IEEE. DOI: 10.1109/TCSET49122.2020.235454
14. The application of geochemical pattern recognition to regional prospecting: A case study of the Sanandaj-Sirjan metallogenic zone, Iran. / [S. A. Meshkani, B. Mehrabi, A. Yaghubpur, Y. F. Alghalandis] // Journal of geochemical exploration. – 2011. – No. 108(3). – P. 183–195. DOI: 10.1016/j.gexplo.2011.01.006
15. Приходько С. Побудова правил прийняття рішень для розпізнавання облич на основі квадрата відстані Махаланобіса для нормалізованих даних. / С. Приходько, А. Трухов // Information Technology: Computer Science, Software Engineering and Cyber Security. – 2023. – No. 2. – P. 50–58. DOI: 10.32782/IT/2023-2-6
16. Real-time facial expression recognition using facial landmarks and neural networks. / [M. A. Haghpanah, E. Saeedizade, M. T. Masouleh, A. Kalhor] // In 2022 International Conference on Machine Vision and Image Processing (MVIP). – 2022. – P. 1–7. IEEE. DOI: 10.1109/MVIP53647.2022.9738754
17. Facial expression recognition based on Electroencephalogram and facial landmark localization. / [D. Li, Z. Wang, Q. Gao et al.] // Technology and Health Care. – 2019. – No. 27(4). – P. 373–387. DOI: 10.3233/THC-181538
18. Age-invariant face recognition using triangle geometric features. / [A. S. Osman Ali, V. S. Asirvadam, A. S. Malik et al.] // International Journal of Pattern Recognition and Artificial Intelligence. – 2015. – No. 29(05). P. 1556006. DOI: 10.1142/S0218001415560066
19. Bijarnia S. Age invariant face recognition using minimal geometrical facial features. / S. Bijarnia, P. Singh // In Advanced Computing and Communication Technologies: Proceedings of the 9th ICACCT. – 2016. – P. 71–77. Springer Singapore. DOI: 10.1007/978-981-10-1023-1_7.
20. Johnston B. A review of image-based automatic facial landmark identification techniques. / B. Johnston, P. D. Chazal // EURASIP Journal on Image and Video Processing. – 2018. – No.1. – P. 1–23. DOI: 10.1186/s13640-018-0324-4
21. Blum L. Impact of Box-Cox Transformation on Machine-Learning Algorithms. / L. Blum, M. Elgendi, C. Menon // Frontiers in Artificial Intelligence. – 2022. – No. 5, P. 877569. DOI: 10.3389/frai.2022.877569
22. Prykhodko S.B. A nonlinear regression model to estimate the size of web apps created using the CakePHP framework / S. B. Prykhodko, I. S. Shutko, A. S. Prykhodko // Radio Electronics, Computer Science, Control. – 2021. – No. 4 (59). – P. 129–139. DOI: 10.15588/1607-3274-2021-4-12


DOI 10.15588/1607-3274-2024-2-9