

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

Кваліфікаційна наукова праця  
на правах рукопису

СЕМЕНЮК РОМАН СЕРГІЙОВИЧ

УДК 681.142

**ДИСЕРТАЦІЯ**

**МЕТОДИ ОПРАЦЮВАННЯ ВИМІРЮВАЛЬНОЇ ТА ЕКСПЕРТНОЇ ІНФОРМАЦІЇ**  
**З ЗАСТОСУВАННЯМ ШКАЛ КЛАСИФІКАЦІЇ**

05.01.02 – стандартизація, сертифікація та метрологічне забезпечення

15 – Автоматизація та приладобудування

Подається на здобуття наукового ступеня кандидата технічних наук

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

\_\_\_\_\_ Р.С. Семенюк

Науковий керівник – Ніна ЯРЕМЧУК, кандидат технічних наук, доцент

Київ – 2021

## АНОТАЦІЯ

*Семенюк Р.С.* Методи опрацювання вимірювальної та експертної інформації з застосуванням шкал класифікації – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук) за спеціальністю 05.01.02 – стандартизація, сертифікація та метрологічне забезпечення (15 – Автоматизація та приладобудування). – КПІ ім. Ігоря Сікорського, Київ, 2021.

Дисертаційну роботу присвячено вирішенню важливої науково-практичної проблеми – розробці методів опрацювання експериментальних даних (числових і вербальних) із застосуванням класифікації.

В роботі проведено огляд наукових робіт вчених, що працювали в області теорії класифікації. Визначено основні ознаки класифікації, серед яких однією з основних є наявність шкали класифікації, що розробляється заздалегідь або в процесі класифікації, виходячи з умови розрізнення класів еквівалентності. При складних випадках класифікації, коли рішення подається за кількома градаціями, характеристикою якості класифікації є умовна ймовірність віднесення до певного класу еквівалентності. Досвід роботи вчених свідчить про те, що статистичні феномени, з якими зустрічається дослідник в галузі класифікації, складні і не завжди передбачувані, тому дуже важливі методи і технології для різних випадків застосування процедури класифікації. Цим питанням присвячена дисертаційна робота автора, в якій представлені технології використання класифікації при опрацюванні числових даних, при переході від числових даних до вербальних з використанням результатів вимірювання і експертної інформації та при опрацюванні вербальних даних.

Серед основних положень наукової новизни та практичного значення варто відзначити наступне:

– набули подальшого розвитку методи непараметричної ідентифікації форми розподілу малих вибірок, на основі чого розроблено метод непараметричної

класифікації видів розподілу, який дозволяє вибрати ефективну оцінку центру розподілу вибірки і тим самим зменшити невизначеність результату вимірювань;

- вперше застосовано метод непараметричної класифікації виду розподілу при побудові контрольних карт технологічного процесу, що дозволяє збільшити чутливість до контрольованих змін центральної тенденції і визначити клас сумарного розподілу вибірок при нестабільності технологічного процесу;

- набули подальшого розвитку методи переходу від числових до вербальних даних в ІВС із застосуванням шкали метричної класифікації, розроблене теоретичне обґрунтування для об'єднання нечіткості експертної інформації і невизначеності вимірювань, що дозволило провести адекватне врахування невизначеності вимірювальної інформації при формуванні функцій приналежності терм-множини шкали і отримати характеристику якості встановленої шкали класифікації у вигляді матриці відповідності;

- вперше розроблено метод перевірки узгодженості вибірок вербальних даних, отриманих за шкалою метричної класифікації, що дозволяє проводити перевірку узгодженості при нерівномірному розташуванні класів еквівалентності в області їх визначення;

- на основі застосування декількох операторів для роботи з вибірками вербальних даних розроблено технологію класифікації стану об'єктів за експертною інформацією з подальшим використанням отриманої шкали класифікації для побудови контрольних карт технологічного процесу за якісними ознаками.

При опрацюванні вимірювальної інформації класифікація використана як попередня операція, що дозволяє удосконалити процедуру вимірювання. Для цієї мети розроблено метод непараметричної класифікації розподілів вибірок малого об'єму. Розглянуто можливості удосконалення методів опрацювання багаторазових вимірювань за допомогою попередньої непараметричної класифікації розподілів вибірок. В першу чергу це стосується формування результату багаторазових вимірювань, який за попередньою класифікацією є ефективною статистичною

оцінкою. Підвищення точності результату вимірювання продемонстровано моделюванням процедури вимірювання за декількома вибірками з генеральної сукупності випадкових чисел з рівномірним законом розподілу, де за результатами попередньої класифікації отримана ефективна оцінка центру розподілу – середина розмаху, а оцінки стандартної і розширеної невизначеності зменшуються у порівнянні з ситуацією, коли розподіл невідомий і його приймають нормальним.

Розроблено технологію застосування попередньої непараметричної класифікації при побудові контрольних карт технологічних процесів для вибору ефективних статистичних характеристик сумарного розподілу вибірок, за якими будується контрольна карта. Зазвичай цей розподіл вважають нормальним з вибором центральної тенденції за середнім арифметичним. З застосуванням попередньої класифікації центральну тенденцію визначають (за однією чи декількома вибірками) або за середнім арифметичним, або за медіаною вибірки, або за серединою розмаху. Моделювання було проведено для контрольних карт стабільності технологічних процесів, за яким видно, що попередня класифікація дозволяє обрати ефективну для даного класу розподілу оцінку центру вибірки. Тому попереджувальні і контрольні межі стають вузькими, що забезпечує більшу чутливість саме до змін параметрів технологічного процесу. При зміні центральної тенденції всі порядкові статистики зміщуються в один бік і їх порядок слідування не змінюється. Завдяки цьому непараметричну класифікацію можна застосовувати при нестабільності технологічного процесу.

Розроблено технології застосування класифікації при переході від числових даних до вербальних в інформаційно-вимірювальних системах, за яким використовується вимірювальна та експертна інформація. Результати вимірювання, які отримують на виході вимірювальних каналів системи, перетворюють в форму, що може бути застосована надалі при нечітких обчисленнях, прийнятті рішень, отриманні діагнозу про стан об'єктів, а також при необхідності регулювання параметрів об'єктів. Для отримання кінцевого результату використовується метрична класифікація, за якої на метричному носіїв будується шкала з нечіткою лінгвістичною змінною з

використанням експертної інформації. Запропонована наступна послідовність етапів встановлення або відтворення шкали класифікації з нечіткою лінгвістичною змінною: орієнтовне визначення кількості термів множини лінгвістичної змінної і їх границь за експертною інформацією; аналіз експертної інформації щодо нечіткості семантичного правила і складових невизначеності вимірювання; вибір загального критерію, що дозволяє об'єднати складові нечіткості і невизначеності; вибір алгоритму об'єднання складових нечіткості і невизначеності; перевірка максимальної кількості термів множини лінгвістичної змінної за умовою розрізненості; вибір функцій приналежності шкали; моделювання нечіткого виведення з застосуванням встановленої лінгвістичної шкали. При аналізі наведених вище етапів особливу увагу приділено чинникам, які впливають на розмитість функції приналежності окремих термів лінгвістичної змінної і які повинні враховуватись при побудові лінгвістичної шкали, проведено дослідження способів об'єднання чинників, що характеризують нечіткість експертної інформації і невизначеність вимірювання, а також способів побудови функцій приналежності термножини шкали. Мірою нечіткості термножини шкали було обрано індекс нечіткості, носій окремих термів був визначений як вкладений інтервал за заданим рівнем довіри. Як характеристику якості встановленої шкали класифікації запропоновано використати матрицю відповідності шкали, за якою може бути визначений функціонал, що відповідає інтуїтивно поняттю «якості класифікації», а саме норма Фробеніуса.

В роботі запропоновано метод оцінки узгодженості вибірок класифікованих даних за шкалою класифікації та за непараметричними оцінками центру вибірок.

Практичне застосування результатів теоретичних досліджень з визначення термножини лінгвістичної змінної в системі класифікації ґрунтів за вмістом рухомого фосфору для оптимізації норми внесення добрив отримане при вирішенні таких задач як тривалий моніторинг земельних ресурсів, поліпшення якості ґрунтів при внесенні добрив, визначення придатності земель при оптимізації різних способів використання,

тощо. Як критерій вибору оптимальної програми відбору проб ґрунту може бути використана запропонована міра узгодженості.

В роботі наведено результати розробки інтелектуальної системи для визначення стану заповнення сміттям підземних урн, в якій було використано метричну класифікацію та розроблено алгоритм опрацювання вербальних даних для своєчасного виклику сміттєзбиральної техніки.

При розробці технологій застосування класифікації для опрацювання вербальних даних, отриманих за експертною інформацією, було визначено оператори, які можуть бути використані саме при роботі з вербальними даними (без призначення рангів). Запропоновано одночасне використання двох операторів (медіани та емулятора середнього арифметичного для вербальних даних) для визначення центру вибірки. В тих випадках, коли результати визначення центральної тенденції не співпадають, а знаходяться в сусідніх класах еквівалентності, запропоновано використовувати проміжні класи еквівалентності.

Розглянуто можливості побудови контрольних карт поточних станів технологічних процесів за сукупністю вербальних вибірок. Встановлено, що важливими характеристиками при класифікації станів є загальна кількість можливих сполучень вербальних категорій в вибірках і кількість розрізнених класів еквівалентності. При невеликій кількості вербальних категорій і застосуванні декількох операторів і критеріїв всі вибірки можуть бути розрізненими, тобто шкала класифікації може бути побудована заздалегідь. При збільшенні кількості вербальних категорій або їх сполучень велика кількість класів еквівалентності приводить до недоцільності побудови загальної шкали класифікації. Тому вербальні вибірки класифікуються в процесі побудови контрольної карти. Розроблено технологію побудови контрольних карт технологічних процесів за якісними характеристиками.

Теоретичні і практичні результати дисертації пов'язані з науковою тематикою та планами кафедри інформаційно-вимірювальних технологій НТУУ «КПІ ім. Ігоря Сікорського», а також НДР держбюджетної теми «Дослідження та розробка

метрологічного та технічного забезпечення для реалізації завдань екологічного моніторингу в Україні» (№ ДР 0113U002159) та НДР «Теоретичні та практичні засади застосування м'яких вимірювань в системах визначення якості продукції» (№ ДР 0118U001478).

Ключові слова: непараметрична класифікація розподілів, шкала метричної класифікації, нечіткість експертної інформації, невизначеність вимірювання, опрацювання вербальних даних, контрольні карти.

Список публікацій здобувача:

Статті у періодичних наукових виданнях інших держав, які входять до ОЕСР:

1. Semeniuk R.S., Yaremchuk N.A. The methods of obtaining and elaborating classified data in intellectual measurement systems : *Central European Researchers Journal (CERES Journal)*, Volume 6, Issue 1, 2020. С.27-36 (видання держави, що входить до ЄС (Словаччина); посилання на збірник: <http://ceres-journal.eu/iss200601>).

Статті у фахових виданнях:

2. Бугреєв С.С., Семенюк Р.С., Яремчук Н.А. Вимірювання рівня сміття в підземних урнах лазерними рівнемірами : *Український метрологічний журнал*. № 3/2017. С. 36-39.

3. Яремчук Н.А., Семенюк Р.С. Опрацювання вербальних даних за побудови контрольних карт : науково-виробничий журнал *Метрологія і прилади*, № 5-1 (67), 2017. С. 58-61.

4. Семенюк Р.С., Павлишин М.М. Використання методів нечіткої класифікації для оптимізації кількості проб ґрунту в задачах агромоніторингу. *Техніко-технологічні аспекти розвитку та випробування нової техніки і технологій для сільського господарства України* : зб. наук. праць, 2018 р., випуск 22(36). С. 264-267.

5. Семенюк Р., Яремчук Н., Гусар І. Спосіб перевірки узгодженості класифікованих даних під час оцінювання показників якості ґрунту. *Техніко-*

*технологічні аспекти розвитку та випробування нової техніки і технологій для сільського господарства України* : зб. наук. праць, 2019 р., випуск 25(39). С.151-159.

6. Семенюк Р.С. Яремчук Н.А. Використання нечіткої класифікації виду розподілу для вибірок малого об'єму. *Інформаційні системи, механіка та керування* : зб. наук. праць, 2017 р., випуск 17. С. 40-50.

Матеріали коференцій:

7. Яремчук Н.А., Семенюк Р.С. Використання нечіткої класифікації при оцінюванні точності і стабільності технологічних процесів. *Гіротехнології, навігація, керування рухом і конструювання авіаційно-космічної техніки* : зб. тез XI міжнар. наук.-техн. конф., м. Київ, 13-14 квітня 2017 р. С. 91-95.

8. Семенюк Р.С. Опрацювання результатів вимірювання з використанням попередньої нечіткої класифікації. *Метрологія, інформаційно-вимірювальні технології та системи* : зб. тез VI міжнар. наук.-техн. конф., м.Київ, 24-25 жовтня 2017 р. С. 129-131.

9. Семенюк Р.С., Яремчук Н.А. Застосування класифікації розподілів вибірки при побудові контрольних карт. *Приладобудування та метрологія: сучасні проблеми, тенденції розвитку* : зб. тез III всеукр. наук.-практ. конф., м.Луцьк, 11-12 жовтня 2018 р. С. 55-57.

10. Семенюк Р.С. Удосконалення побудови контрольних карт технологічних процесів з застосуванням попередньої класифікації. *Мехатронні системи: інновації та інжиніринг* : зб. тез II міжнар. наук.-практ. конф., м.Київ, 15 червня 2018 р. С. 46-47.

11. Семенюк Р.С. Способы определения терм-множества лингвистической переменной с неопределенностью измерений. *Неопределенность измерений: научные, нормативные и методические аспекты УМ-2018* : зб. тез 15-ї міжнар. наук.-техн. сем., г.Созополь, 10 сентября 2018 р. С. 206-207.

12. Семенюк Р.С., Яремчук Н.А. Основні етапи встановлення лінгвістичних шкал при вимірюваннях і діагностиці. *Метрологія, інформаційно-вимірювальні*



*технології та системи* : зб. тез VII міжнар. наук.-техн. конф., 18-19 лютого 2020 р. С. 131-133.

13. Яремчук Н.А., Семенюк Р.С. Способи урахування невизначеності при побудові лінгвістичних шкал. *Метрологія та вимірювальна техніка (Метрологія–2020)* : зб. доп. XII міжнар. наук.–техн. конф. 6–8 жовтня 2020 р. – Х.: ННЦ Інститут метрології, 2020. – С. 77-81.

14. Ступак С.С., Яремчук Н.А., Семенюк Р.С. Способи урахування впливу невизначеності на результати «м'яких» вимірювань. *Актуальні проблеми автоматики та приладобудування* : матеріали 2-ї міжнар. наук.-техн. конф., м.Харків, 06-07 грудня 2018 р. С. 210-211.

## ANNOTATION

*Semeniuk R.S.* Methods of processing measurement and expert information using classification scales – qualification scientific work on the rights of a manuscript.

Dissertation to obtain a scientific degree of Candidate of Technical Sciences for the specialty 05.01.02 – standardization, certification and metrological support (15 – Automation and Instrumentation). – Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, 2021.

The dissertation work is devoted to solving an important scientific and practical problem – the development of methods for processing experimental data (numerical and verbal) using classification.

The paper reviews the scientific works of scientists who worked in the field of classification theory. The main features of classification are defined, among which one of the main ones is the presence of a classification scale developed in advance or in the classification process, based on the condition of distinguishing equivalence classes. In complex classification cases, when the solution is presented in several gradations, the classification quality characteristic is the conditional probability of being assigned to a certain equivalence class. Scientists' experience shows that the statistical phenomena a researcher

encounters in the field of classification are complex and not always predictable, therefore methods and techniques for different applications of the classification procedure are very important. This issue is the focus of the author's dissertation, which presents techniques for using classification to work through numerical data, transitioning from numerical data to verbal data using measurement and expert information, and working through verbal data.

Among the main points of scientific novelty and practical importance it is worth noting the following:

- methods for the non-parametric identification of the shape of the distribution of small samples have been further developed, based on which a method for the non-parametric classification of distribution types has been developed that makes it possible to select an effective estimate of the sample distribution center and thereby reduce the uncertainty of the measurement result;

- the method of non-parametric classification of the distribution type was applied for the first time in the construction of process control charts, thus increasing the sensitivity to controllable changes in central tendency and determining the class of the total sampling distribution for process instability;

- methods for transition from numerical to verbal data in information and measurement system using a metric classification scale have been further developed, and a theoretical rationale for combining uncertainty in expert information and measurement uncertainty has been developed, that allowed to adequately take into account the uncertainty of the measurement information in the formation of the membership functions of the term set of the scale and to obtain a characteristic of the quality of the established classification scale in the form of a correspondence matrix;

- a method for checking the consistency of samples of verbal data obtained behind a metric classification scale has been developed for the first time, allowing consistency checking when equivalence classes are not evenly spaced in their domain of definition;

- based on the application of several operators for working with samples of verbal data, the technology for classifying the state of objects based on expert information has been

developed, followed by the use of the resulting classification scale to build process control charts for qualitative attributes.

During the elaboration of measurement information, classification is used as a previous operation, which allows to improve the measurement procedure. For this purpose, a method for nonparametric classification of small sample distributions has been developed. Possibilities of improvement of methods of elaboration of repeated measurements by means of the previous nonparametric classification of distributions of samples are considered. Primarily, this concerns the formation of a multiple measurement result, which behind the previous classification is an effective statistical estimate. An increase in the accuracy of the measurement result is demonstrated by simulating the measurement procedure over several samples from a population of random numbers with a uniform distribution model, where the previous classification results in an effective mid-distribution centre estimate, and the standard and expanded uncertainty estimates are reduced in comparison with the situation where the distribution is unknown and assumed to be normal.

Technology has been developed to apply the previous nonparametric classification in the construction of control maps of technological processes to select the effective statistical characteristics of the total distribution of samples, for which the control map is constructed. This distribution is usually assumed to be normal with the central tendency chosen as the arithmetic mean. Using the previous classification, the central tendency is defined (over one or more samples) as either the arithmetic mean, the sample median or the mid-point of the spread. Modelling has been carried out for process stability control charts, for which it can be seen that the previous classification allows the selection of an effective sample center estimate for a given class of distribution. Therefore, the warning and control limits become narrower, which ensures greater sensitivity specifically to changes in process parameters. When the central tendency changes, all ordinal statistics are shifted to one side and their order does not change. This means that non-parametric classification can be used when the process is unstable.

Techniques have been developed for the application of classification in the transition from numerical data to verbal data in information and measurement systems, behind which measurement and expert information is used. The measurement results, which are obtained at the output of the measuring channels of the system, are transformed into a form that can be further applied in fuzzy calculations, decision-making, obtaining diagnosis, about the state of objects, as well as in the need to regulate the parameters of objects. For the final result is used a metric classification in which a scale with a fuzzy linguistic variable is constructed on a metric medium using expert information. The following sequence of steps to establish or recreate a classification scale with a fuzzy linguistic variable is proposed: tentative determination of the number of linguistic variable plural terms and their boundaries by expert information; analysis of expert information regarding the fuzzy semantic rule and the components of measurement uncertainty; selection of a common criteria that allows combining the fuzzy and uncertainty components; selection of an algorithm for combining the fuzzy and uncertainty components; verification of the maximum number of linguistic variable plural terms beyond the disjointness condition; selection of scale membership functions; simulation of fuzzy output using the established linguistic scale. In the analysis of the above steps, special attention is paid to the factors that influence the fuzziness of the membership function of the individual terms of the linguistic variable and that must be taken into account in the construction of the linguistic scale, the methods for combining the factors that characterise the fuzziness of expert information and measurement uncertainty, and the methods for constructing the membership functions of the term set of the scale are investigated. A fuzzy index was chosen as a measure of the term set fuzzyness of the scale, the carrier of the individual terms was defined as a nested interval beyond a given confidence level. As a characteristic of the quality of the established classification scale, it was proposed to use a scale correspondence matrix, behind which there can be a certain functional that corresponds intuitively to the notion of "classification quality", namely the Frobenius norm.

In this work, it is proposed a method for assessing the consistency of samples of classified data behind the classification scale and behind the non-parametric estimates of the centre of samples.

The practical application of the results of theoretical research from the determination of the multiple term linguistic variable in the soil classification system for the content of mobile phosphorus to optimise the rate of fertilisation is obtained in such tasks as long-term monitoring of land resources, improving soil quality, when applying fertilisers, determining land suitability, when optimising different uses, and so on. A proposed measure of consistency can be used as a criteria for selecting the best suitable soil sampling program. In this work are presented the results of the development of an intelligent system for determining the state of filling of underground rubbish bins , in which a metric classification was used and a verbal algorithm was developed for the timely recall of waste collection equipment.

In developing techniques for applying classification to the elaboration of verbal data obtained from expert information, operators that can be used specifically when dealing with verbal data (without assigning ranks) have been identified. The simultaneous use of two operators (median and arithmetic mean emulator for verbal data) to determine the sampling center was suggested. In cases when the results of the central tendency determination do not coincide but are in neighboring equivalence classes, has been suggested the use of intermediate equivalence classes.

It was considered the possibility of constructing control maps of the current states of technological processes over a set of verbal samples. It has been established that the important characteristics in classifying the states are the total number of possible connections of verbal categories in the samples and the number of disparate equivalence classes. For small numbers of verbal categories and the application of multiple operators and criteria, all samples can be disjointed, i.e., the classification scale can be constructed preliminarily. When the number of verbal categories or their compounds increases, a large number of equivalence classes makes it impractical to construct a general classification scale. Therefore, the verbal

samples are classified in the process of constructing a control map. It has developed a technology for constructing process control charts for quality characteristics.

Theoretical and practical results of the thesis are connected with the scientific themes and plans of the department of information and measuring technologies of Igor Sikorsky Kyiv Polytechnic Institute, as well as research paper of the state budget / fiscal theme «Research and development of metrological and technical support for the implementation of environmental monitoring tasks in Ukraine» (№ RP 0113U002159) and research paper «Theoretical and practical principles of the application of soft measurements in product quality systems» (№ RP 0118U001478).

*Keywords:* non-parametric classification of distributions, metric classification scale, expert information uncertainty, measurement uncertainty, verbal data elaboration, control maps.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ .....	18
ВСТУП .....	19
РОЗДІЛ 1. ЗАСТОСУВАННЯ КЛАСИФІКАЦІЇ ПРИ ОПРАЦЮВАННІ ВИМІРЮВАЛЬНОЇ ІНФОРМАЦІЇ. КЛАСИФІКАЦІЯ РОЗПОДІЛІВ ВИБІРОК МАЛОГО ОБ'ЄМУ.....	25
1.1. Визначення процедури класифікації, її основні ознаки.....	25
1.2. Визначення правила класифікації.....	27
1.3. Характеристики якості класифікації .....	29
1.4. Огляд методів ідентифікації та класифікації розподілів вибірок малого об'єму.....	32
1.5. Розширення можливостей непараметричної нечіткої ідентифікації для різних об'ємів вибірки.....	37
1.6. Дослідження нечіткої ідентифікації вибірок для різних розподілів з метою визначення класифікаційних ознак.....	40
1.6.1. Отримання навчальних (еталонних) вибірок для різних розподілів, визначення нечітких оцінок і лінгвістичних кодів.....	41
1.6.2. Формування класів еквівалентності для непараметричної класифікації розподілів вибірок .....	45
1.7. Достовірність класифікації.....	48
1.7.1. Теоретична оцінка достовірності.....	49
1.7.2. Експериментальна оцінка достовірності при моделюванні процедури класифікації.....	52
1.8. Визначення необхідної кількості вибірок за заданою ймовірністю правильної класифікації.....	54
Висновки до розділу 1.....	57
РОЗДІЛ 2. ОПРАЦЮВАННЯ ВИМІРЮВАЛЬНОЇ ІНФОРМАЦІЇ ЗА ПОПЕРЕДНЬОЮ КЛАСИФІКАЦІЄЮ РОЗПОДІЛІВ ВИБІРКИ.....	59

2.1. Ефективні оцінки центру для різних законів розподілу.....	59
2.2. Застосування попередньої класифікації розподілу вибірки для підвищення точності при опрацюванні результатів вимірювань.....	61
2.3. Застосування непараметричної класифікації при побудові контрольних карт.....	64
2.3.1. Застосування непараметричної класифікації при визначенні стабільності технологічних процесів.....	65
2.3.2. Моделювання процедури застосування непараметричної класифікації при побудові контрольних карт.....	67
Висновки до розділу 2.....	74
<b>РОЗДІЛ 3. ОСОБЛИВОСТІ ОПРАЦЮВАННЯ ВИМІРЮВАЛЬНОЇ ТА ЕКСПЕРТНОЇ ІНФОРМАЦІЇ ПРИ ПЕРЕХОДІ ВІД ЧИСЛОВИХ ДАНИХ ДО ВЕРБАЛЬНИХ ЗА МЕТРИЧНОЮ КЛАСИФІКАЦІЄЮ.....</b>	<b>75</b>
3.1. Загальна характеристика метричної класифікації.....	75
3.2. Основні етапи встановлення шкал класифікації при вимірюваннях і діагностиці.....	77
3.3. Визначення кількості класів еквівалентності шкали.....	81
3.4. Встановлення форми функцій приналежності окремих градацій шкали класифікації.....	83
3.5. Визначення міри узгодженості між класифікованими даними за умов метричної класифікації.....	91
3.5.1. Використання міри узгодженості для оцінки розсіювання класифікованих даних.....	91
3.5.2. Застосування шкали класифікації при визначенні міри узгодженості.....	92
3.6. Оцінювання якості шкали класифікації за матрицею відповідності.....	94
3.7. Визначення терм-множини лінгвістичної змінної при класифікації ґрунтів за вмістом рухомого фосфору для оптимізації норми внесення добрив.....	95



3.7.1. Побудова шкали класифікації показників якості ґрунту з урахуванням невизначеності вимірювання.....	97
3.7.2. Застосування критерію перевірки узгодженості (збіжності) класифікованих даних вмісту рухомого фосфору.....	104
3.8. Застосування нечіткої класифікації при визначенні стану заповнення сміттям підземної урни.....	107
3.8.1. Вимірювач рівня сміття.....	108
3.8.2. Застосування шкали класифікації при визначенні стану заповнення сміттям підземної урни.....	109
Висновки до розділу 3.....	113
РОЗДІЛ 4. ЗАСТОСУВАННЯ КЛАСИФІКАЦІЇ ДЛЯ ОПРАЦЮВАННЯ ЕКСПЕРТНОЇ ІНФОРМАЦІЇ, ПОДАНОЇ ЗА ВИБІРКАМИ ВЕРБАЛЬНИХ ДАНИХ .....	115
4.1. Відношення та оператори, які використовуються при роботі з вербальними даними.....	115
4.2. Технології застосування процедури класифікації при визначенні центральної тенденції вербальної вибірки.....	117
4.3. Класифікація поточних станів технологічних процесів при побудові контрольних карт.....	119
4.4. Побудова контрольної карти і знаходження центральної тенденції груп підземних урн.....	121
4.5. Метод класифікації стану об'єкта або технологічного процесу за вибірками вербальних даних при побудові контрольних карт.....	125
Висновки до розділу 4.....	127
ВИСНОВКИ .....	129
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	131
ДОДАТКИ .....	136

## **ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ**

ІЗВ – інтелектуальні засоби вимірювання;

ІВС – інтелектуальні вимірювальні системи;

КК – контрольна карта;

НО – нечіткі оцінки;

ЛК – лінгвістичний код;

СКВ – середнє квадратичне відхилення;

ЛЗ – лінгвістична змінна;

ФП – функція приналежності;

НК – нечіткий класифікатор.

## ВСТУП

**Актуальність теми.** В сучасних засобах вимірювальної техніки, особливо в інтелектуальних засобах вимірювання (ІЗВ), застосовується класифікація, як додаткова операція, що дозволяє удосконалити процедуру вимірювання. Тому в представленій роботі проведено дослідження і розробку методів застосування класифікації при опрацюванні експериментальних даних, як числових (кількісних), так і вербальних (якісних). У розробленні та вирішенні питань застосування класифікації внесли вагомий внесок вітчизняні та зарубіжні вчені: Орнатський П.П., Айвазян С.А., Кендал М., Стьюарт А., Розенберг В.Я., Bashkansky E., Franceschini F. В області опрацювання числових даних актуальним питанням є оптимізація вибору алгоритмів функціонування і опрацювання даних за умов активного використання апріорної і поточної вимірювальної інформації. До таких задач відноситься визначення виду розподілу або класу розподілу вибірок малого об'єму, тобто їх попередня класифікація з метою визначення виду розподілу. Наявність такої інформації дозволяє вибрати найбільш ефективну оцінку вимірюваної величини, що має найменшу стандартну невизначеність і, відповідно, коефіцієнт охоплення при обчисленні розширеної невизначеності. В ІЗВ ідентифікація виду розподілу включається в операцію класифікації і повинна виконуватися без участі оператора. Проте, на даний момент залишаються недостатньо розробленими задачі визначення розподілу малих вибірок, тому в дисертації набули подальшого розвитку питання непараметричної класифікації виду розподілу, яка нечутлива до помірних змін центральної тенденції вибірок, що є значною перевагою для її застосування в контрольних картах (КК) технологічних процесів. Про необхідність визначення класу розподілу при побудові КК зазначено в стандарті ISO 8258 – 2001.

В інтелектуальних вимірювальних системах (ІВС) результати вимірювань використовуються при нечітких виведеннях і нечітких (м'яких) обчисленнях. Тому в таких системах має місце поліморфізм, тобто відображення вимірюваних властивостей декількома шкалами, а саме метричною шкалою і шкалою класифікації. За

М. Кендалом і А. Стьюартом класифікація, що забезпечує перехід від числових даних, поданих за метричною шкалою, до вербальних (класифікованих) даних, називається метричною. Основною складністю цього переходу, на думку спеціалістів, є необхідність розробки теоретичного обґрунтування побудови шкали класифікації з урахуванням складових невизначеності вимірювальної інформації та неоднозначності або нечіткості експертної інформації. Технологія об'єднання вимірювальної та експертної інформації повинна супроводжуватись характеристиками якості метричної класифікації. Розв'язання цих задач набуває особливого значення для ІВС, що використовуються в промисловості і в народному господарстві, при медичних дослідженнях, тобто там, де невизначеність вимірювальної інформації суміжна з нечіткістю експертної інформації.

Опрацювання вербальних даних, отриманих за метричною класифікацією, може бути проведено з урахуванням відстані між класами еквівалентності шкали класифікації. Це забезпечує можливість розробки технологій опрацювання технології, яка орієнтована на рівномірне розміщення класів еквівалентності в області їх визначення.

Актуальною задачею є розробка методів опрацювання експертної інформації у вигляді вибірок вербальних даних з метою визначення (класифікації) стану об'єкта за центральною тенденцією вибірок експертних даних, що відображають властивості об'єкта, та ранжування вибірок для побудови контрольних карт за вербальними даними.

Розробка технології для опрацювання експертної інформації є актуальною задачею через обмеженість кількості встановлених операторів для її вирішення. Цілий ряд добре відомих операторів непридатний для експертних даних, тому що такі операції як додавання і віднімання несправедливі для ординальних даних. Тому задача комбінування операторів для класифікації вибірок вербальних даних відповідає сучасним потребам розвитку систем з опрацювання науково-технічної інформації.

**Зв'язок роботи з науковими програмами, планами, темами.** Дослідження, результати яких викладені в дисертації, пов'язані з науковою тематикою та планами кафедри інформаційно-вимірювальних технологій НТУУ «КПІ ім. Ігоря Сікорського», а також НДР держбюджетної теми «Дослідження та розробка метрологічного та технічного забезпечення для реалізації завдань екологічного моніторингу в Україні» (№ ДР 0113U002159) та НДР «Теоретичні та практичні засади застосування м'яких вимірювань в системах визначення якості продукції» (№ ДР 0118U001478).

**Мета і задачі дослідження.** *Метою дисертаційної роботи є розробка методів опрацювання вимірювальної та експертної інформації (числової та вербальної) з використанням шкал класифікації.*

Для досягнення поставленої мети вирішено наступні задачі дослідження:

1. Розробити непараметричну класифікацію виду розподілу малих вибірок числових даних, що може бути використана для оптимізації методів опрацювання даних і при побудові контрольних карт технологічних процесів.

2. Визначити основні етапи переходу від числових до вербальних даних при метричній класифікації і розробити питання врахування нечіткості експертної інформації і невизначеності числових даних при побудові шкали класифікації, розробити критерії оцінювання якості класифікації.

3. Удосконалити методи опрацювання вербальних (експертних) даних для їх подальшої класифікації за центром малої вибірки для характеристики стану об'єктів та за домінуванням при побудові КК за якісними характеристиками контрольованих властивостей.

4. Розробити технології використання шкали класифікації при вирішенні практичних задач господарської діяльності, а саме для оптимізації норми внесення добрив в ґрунти сільськогосподарського призначення, для перевірки методик отримання проб ґрунту, для визначення рівня заповнення підземних урн сміттям, для ІВС медичного призначення, тощо.

*Об'єктом дослідження* є процес опрацювання числових та вербальних даних в вибірках малого об'єму.

*Предметом дослідження* є методи опрацювання вимірювальної та експертної інформації зі встановленням шкал класифікації як для удосконалення процедур опрацювання даних, так і для переходу від числових даних до вербальних.

**Методи дослідження.** Розв'язання поставлених задач виконано з використанням основних положень системного аналізу та теорії вимірювань; теорії ймовірностей, математичної статистики, теорії нечітких множин та методів експертного оцінювання, теорії шкал.

**Наукова новизна одержаних результатів.** На основі проведених досліджень отримано наступні результати:

- вперше розроблено метод непараметричної класифікації видів розподілу малих вибірок, який відрізняється можливістю вибрати ефективну оцінку центру розподілу вибірки та визначити клас сумарного розподілу вибірок за нестабільності технологічного процесу при побудові контрольних карт;
- набули подальшого розвитку методи переходу від числових до вербальних даних в інформаційних вимірювальних системах із застосуванням шкали метричної класифікації шляхом об'єднання нечіткої експертної інформації і невизначеності вимірювань, що дозволило отримати характеристику якості встановленої шкали класифікації у вигляді матриці відповідності;
- вперше розроблено метод перевірки узгодженості між вербальними даними в вибірках, отриманих за шкалою метричної класифікації, який відрізняється можливістю врахування нерівномірності розташування класів еквівалентності в області їх визначення;
- вперше розроблено метод класифікації стану об'єкту за експертною інформацією на основі застосування методів математичної статистики і нечіткої логіки для роботи з вибірками вербальних даних, який відрізняється можливістю

використання отриманих шкал класифікації для побудови контрольних карт за якісними ознаками.

**Практичне значення одержаних результатів** полягає у тому, що розроблені методи опрацювання даних з застосуванням класифікації можуть бути використані при побудові ІВС для удосконалення алгоритмів опрацювання даних багаторазових вимірювань, при побудові контрольних карт технологічних процесів на основі числових і вербальних даних, при отриманні результатів «м'яких» вимірювань. Результати дослідження метричної класифікації були використані в методиках визначення нерівномірності розподілу добрив при агромоніторингу, при побудові ІВС для визначення стану заповнення сміттям підземних урн. Способи класифікації вибірок вербальних даних можуть бути широко застосовані при візуальному і тактильному контролі якості продукції.

**Особистий внесок здобувача.** Всі результати теоретичного аналізу та експериментальних досліджень дисертаційної роботи, висновки та рекомендації, які винесені до захисту, одержані автором особисто. В спільних публікаціях автором зроблено наступне: [1, 2] – дослідження використання непараметричної класифікації виду розподілу для вибірок малого об'єму, [6] - застосування методів непараметричної класифікації для оптимізації кількості проб ґрунту в задачах агромоніторингу, [4, 7] – дослідження використання контрольних карт технологічних процесів при опрацюванні даних, [12] – застосування лінгвістичних шкал при вимірюваннях і діагностиці, [3] – дослідження використання вербальної шкали при розробці вимірювача сміття в підземній урні, [13] – дослідження способів урахування інструментальної невизначеності при формуванні результатів м'яких вимірювань, [14] - дослідження способів побудови шкали класифікації, за якої отримують ступені приналежності результатів вимірювання окремих ознак встановленим класам еквівалентності.

**Апробація результатів дисертації.** Основні положення та результати дослідження доповідалися та обговорювалися на: XI Міжнародній науково-технічній конференції «Гіротехнології, навігація, керування рухом і конструювання авіаційно-

космічної техніки» (Київ, 13-14 квітня 2017), VI Міжнародній науково-технічній конференції «Метрологія, інформаційно-вимірювальні технології та системи», (Київ, 24-25 жовтня 2017), III Всеукраїнській науково-практичній конференції «Приладобудування та метрологія: сучасні проблеми, тенденції розвитку» (Київ, 11-12 жовтня 2018), II Міжнародній науково-практичній конференції «Мехатронні системи: інновації та інжиніринг» (Київ, 15 червня 2018), XV Міжнародному науково-технічному семінарі «Неопределенность измерений: научные, нормативные и методические аспекты» UM-2018, (Созополь, 10 сентября 2018), VII Міжнародній науково-технічній конференції «Метрологія, інформаційно-вимірювальні технології та системи», (Харків, 18-19 лютого 2020), XII Міжнародній науково-технічній конференції «Метрологія та вимірювальна техніка» та на семінарі «Невизначеність вимірювань: наукові, прикладні, нормативні та методичні аспекти» UM-2020, (Харків 7 жовтня 2020).

**Публікації.** Основні положення дисертаційної роботи опубліковано у 14 наукових працях, у тому числі 6 статей у наукових фахових виданнях (з них 1 стаття у періодичних наукових виданнях інших держав, які входять до ОЕСР та/або Європейського Союзу), 8 тез доповідей в збірниках матеріалів конференцій.

**Структура та обсяг дисертаційної роботи.** Дисертаційна робота складається з переліку умовних скорочень, вступу, 4 розділів, висновків, списку використаних найменувань і 2 додатків. Загальний обсяг дисертації складає 149 сторінок друкованого тексту, у тому числі 137 основного тексту, робота містить 19 рисунків, 46 таблиць.



## **РОЗДІЛ 1. ЗАСТОСУВАННЯ КЛАСИФІКАЦІЇ ПРИ ОПРАЦЮВАННІ ВИМІРЮВАЛЬНОЇ ІНФОРМАЦІЇ. КЛАСИФІКАЦІЯ РОЗПОДІЛІВ ВИБІРОК МАЛОГО ОБ'ЄМУ.**

Розвиток і широке використання комп'ютерних і інтелектуальних вимірювальних систем обумовлює необхідність доповнення переліку традиційних вимірювальних операцій в процедурі вимірювання, а саме операцією класифікації. При опрацюванні числових експериментальних даних операція класифікації дозволяє отримати додаткову інформацію про об'єкт вимірювання, яка може бути потім використана для удосконалення алгоритму опрацювання даних. В даній роботі це інформація про класи розподілу ймовірностей вибіркового даних, що може бути використана для удосконалення процедури опрацювання даних багаторазових вимірювань і при побудові контрольних карт технологічних процесів.

### **1.1. Визначення процедури класифікації, її основні ознаки.**

Визначення і систематизація видів класифікації наведені в роботах М.Дж. Кендала і А. Стьюарта [1], С.А. Айвазяна [2] і В.Я. Розенберга [3].

За [1] процедура класифікації відноситься до процедур диференціації генеральних сукупностей, що отримують в результаті вимірювань. При цьому виникають три класи задач, що вирішуються процедурами дискримінації, класифікації і розбивки.

За [1] класифікація – процедура, що складається з розподілення елементів вибірки чи їх сукупності на більш або менш однорідні групи або класи, кожний з яких характеризується певними ознаками, і віднесенням певного елемента до того чи іншого класу.

За [2] класифікацією називають розділення сукупності об'єктів та явищ на однорідні (в певному сенсі) групування. При цьому термін «класифікація» використовують в залежності від контексту як для позначення самого процесу розділення, так і для його результату. Це поняття тісно пов'язане з такими термінами

як групування, типологізація, систематизація, дискримінація, кластеризація і є одним з основних в практичній і науковій діяльності людини.

Основними ознаками процедури класифікації є [2, 4]:

- наявність шкали класифікації, що охоплює всю генеральну сукупність і яка розробляється заздалегідь або в процесі класифікації;
- наявність правил, що виробляються в процесі класифікації виходячи з умови розрізнення;
- наявність емпіричних відношень, що характеризують об'єкти класифікації (відношення порядку, еквівалентності або толерантності);
- можливість формування навчальних вибірок;
- наявність інформації про класи еквівалентності.

Метою процедури класифікації є виявлення всіх класів і віднесення об'єкта до одного з виявлених класів.

В залежності від мети і ознак, за якими об'єкти класифікуються, виділяють наступні різновиди класифікації (рис.1.1).



Рисунок 1.1 – Різновиди класифікації

Визначення чіткої класифікації наведено в [3]: класифікацією множини  $M$  називається повна сім'я  $\{M_1, \dots, M_L\}$  непустих, попарно неперерізаючих підмножин  $M_i \neq \emptyset, M_i \subset M, i = \overline{1, L}$  множини  $M$ , що в поєднанні дають все  $M$ :

$$\bigcup_{i=1}^L M_i = M, M_i \cap M_j = \emptyset, i \neq j, \forall i, j \in \{\overline{1, L}\}$$

В цьому випадку на  $M$  визначено відношення еквівалентності.

Підмножини  $M_i$  називають класами еквівалентності. При чіткій класифікації рішення про віднесення до класу має вигляд

$$M_i = \{m : m \text{ має властивість } i\}, i = \overline{1, L}$$

При нечіткій класифікації остаточне рішення може бути подано з зазначенням розсіювання за декількома класами еквівалентності.

Якщо класи встановлюються у відповідності з ознаками, тоді елементи  $m$ , що входять до класу  $M_i$ , будуть відрізнятися значеннями параметрів  $a_i$ , які відповідають цьому класу, а класи будуть представляти собою параметричні сім'ї (множини) функцій [3]:

$$M_i = \{m : m = f_i(x, a_i), a_i \in A_i\},$$

де  $f_i$  – функція аргументу  $x$ , що залежить від параметру  $a_i$ , причому  $f, x, a_i$  можуть бути як скалярними, так і векторними величинами. Якщо кожний елемент характеризується лише однією ознакою, тоді

$$M_i = \{m : m = f_i(x, a_i), i \in \{\overline{1, L}\}, a_i \in A_i\}.$$

Впорядкований набір ознак і визначає «список» класів, що створює простий вид класифікації. Прикладом може бути класифікація імпульсів за формою, класифікація одновимірних розподілів вибірки.

Класифікацію за однією вимірюваною ознакою і впорядкованими класами еквівалентності в роботі [1] називають метричною класифікацією.

## 1.2. Визначення правил класифікації.

Матеріали з визначення правила класифікації наведено в [2]. Розглянуто класифікацію для випадку, коли розподіли двох класів визначено повністю. За [2] вхідними даними  $x'$  процедури класифікації є впорядкований набір з  $p$ -ознак або результатів вимірювань:

$$x' = \{x^{(1)}, \dots, x^{(p)}\}.$$

Подія, що  $x'$  отримано з  $j$ -го класу, а також відповідну гіпотезу в [2] позначено  $H_j$ , розподіл вектору  $x$ , що належить  $j$ -му класу ( $j=1, \dots, k$ ) позначено  $F_j(\dots) = F(\dots | H_j)$ , щільність ймовірності відповідно  $f_j(\dots) = f(\dots | H_j)$ .

Задачу побудови класифікаційного правила розглянуто при двох способах подання  $x$ : аналітичному (формула) і вибірковому. Ця задача тісно пов'язана з класичною задачею перевірки гіпотези  $H_1: x \in F_1$  проти гіпотези  $H_2: x \in F_2$ . З помилкою першого роду  $\alpha$ , найменшій помилці другого роду  $\beta$  відповідає відношення правдоподібності, засноване на статистиці:

$$\gamma(x) = \frac{L(x | H_2)}{L(x | H_1)} = \frac{f_2(x)}{f_1(x)},$$

де  $L$  – функція правдоподібності,  $\gamma(x)$  – критерій відношення правдоподібності.

При цьому при  $\gamma(x) \leq c_\alpha$  приймається гіпотеза  $H_1$ , а при  $\gamma(x) > c_\alpha$  приймається гіпотеза  $H_2$ . Таким чином,  $R$  – простір можливих значень  $x$  за допомогою  $\gamma(x)$  розбивається на дві області, що не перерізаються:

- $K_1 = \{x: \gamma(x) \leq c_\alpha\}$ ,  $x \in R$ , що є областю прийняття  $H_1$ ;
- $K_2 = \{x: \gamma(x) > c_\alpha\}$ ,  $x \in R$ , що є областю прийняття  $H_2$ , або критичною областю для прийняття  $H_1$ .

Якщо позначити  $\pi_j = P(H_j)$  як апіорну ймовірність гіпотези, тоді правило класифікації з [2] має вигляд:

$$\frac{\pi_2 f_2(x)}{\pi_1 f_1(x)} \neq 1 \Rightarrow \begin{cases} H_2 \\ H_1 \end{cases}.$$

Таке правило є байєсівським і є частковим випадком відношення правдоподібності. Тоді за формулою повної ймовірності, ймовірність помилкового рішення:

$$P(\text{помилка}) = \pi_1 p\{\text{помилка} | H_1\} + \pi_2 p\{\text{помилка} | H_2\}.$$

### 1.3. Характеристики якості класифікації.

Якщо розглядати класифікацію як перевірку гіпотези "*x* належить  $F_1$ " або " $x \in F_1$ " проти альтернативи "*x* належить  $F_2$ ", тоді якість рішення визначається помилками першого та другого роду. Але на практиці виникають більш складні форми висновків, наприклад [2], рішення, що складається з 3-х градацій: " $x \in F_1$ ", відмова від класифікації, " $x \in F_2$ " або висновок за умовною ймовірністю  $P\{x \in F_1 | x\}$ . Відповідно до цього змінюються показники якості класифікації.

В загальному випадку статистичний критерій класифікації може бути поданий як  $\gamma(x) \neq c$ , де  $\gamma(x)$  – відома функція  $x$ ,  $c$  – поріг критерію.

В залежності від специфічних прикладів класифікації і критерій, і характеристики якості можуть змінюватись. Для простих правил класифікації в [2] запропоновано 2 види характеристик якості класифікації: чутливість і специфічність. Пояснення цих характеристик наведено на прикладі класифікації на класи  $H_1$  і  $H_2$  (табл.1.1)

Таблиця 1.1 – Приклад застосування простого правила класифікації [2]

Результат застосування критерія	Статус об'єкта		Всього
	$H_1$	$H_2$	
Прийнято гіпотезу $H_1$ $\gamma(x) < c$	$b$	$s$	$b + s$
Відкинуто гіпотезу $H_1$ $\gamma(x) \geq c$	$g$	$l$	$g + l$
Всього	$b + g$	$s + l$	$n$

Чутливість критерія, тобто доля випадків, для яких  $\gamma(x) < c$  пов'язана з помилкою першого роду, тобто чутливість дорівнює  $1 - \alpha$ .

Специфічність критерія за [2] – це відношення  $l / (s + l)$ , тобто доля випадків, для яких  $\gamma(x) \geq c$ , що дорівнює  $1 - \beta$ .

Відносний ризик відношення ймовірності бути випадком, за умов, коли гіпотезу «випадок» прийнято, до ймовірності бути випадком, коли цю гіпотезу відкинута, що є відношенням відносної долі прийнятих до відносної долі неприйнятих:

$$R = \frac{b}{b+s} : \frac{g}{g+l}.$$

Характеристики якості класифікації можуть бути змінені виходячи зі зміни порогу критерію, тому виникає необхідність визначення якості класифікації за допомогою заданої функції  $\gamma(x)$  при різних значеннях  $c$ . Для цієї мети розглядають залежність між чутливістю та специфічністю. В роботі [2] зазначено, що досвід роботи в галузі класифікації свідчить про те, що статистичні феномени, з якими зустрічається дослідник в галузі класифікації, складні і не завжди передбачувані. Тому закінчених теоретичних результатів, які могли б виконувати роль орієнтирів для практики, у порівнянні з іншими областями прикладної статистики, значно менше, у зв'язку з чим дуже важливі напрацювання рекомендацій для різних випадків застосування процедури класифікації.

В [1] зазначено, що наслідки помилкової класифікації можуть суттєво відрізнятися. Тому в таких випадках може бути застосована функція витрат. Нехай ймовірності приналежності двом генеральним сукупностям позначено  $\pi_1$  і  $\pi_2$ , а помилкам прийняття рішення приписані певні числа, що характеризують втрати при помилковій класифікації. Наприклад, помилки першого та другого роду дорівнюють відповідно  $c_1$  і  $c_2$ . Тоді мінімізація проводиться за співвідношенням:

$$c_2 \int_R \pi_2 f_2 dx + c_1 \int_{1-R} \pi_1 f_1 dx = c_1 \pi_1 + \int_R (c_2 \pi_2 f_2 - c_1 \pi_1 f_1) dx.$$

Це співвідношення досягає мінімуму, якщо область  $R$  визначено за формулою:

$$\frac{c_2 \pi_2 f_2}{c_1 \pi_1 f_1} < 1.$$

Якщо між введеними класами можна ввести транзитивне відношення порядку, то такі класи називають впорядкованими. Упорядкування може бути пов'язане зі змістовним тлумаченням класів і з їх геометричним розташуванням вздовж певної прямої у вибірковому просторі. У випадках, коли класи відповідають послідовним стадіям певного процесу, їх змістовне і геометричне впорядкування часто співпадають. При роботі з впорядкованими класами використовують інтуїтивне уявлення дослідника про «відстань» між класами, що використовується у подальшому при класифікації і при визначенні якості класифікації.

Таким чином, основними ознаками процедури класифікації є наявність шкали класифікації, що охоплює всю генеральну сукупність і яка розробляється заздалегідь або в процесі класифікації, виходячи з умови розрізнення; наявність емпіричних відношень, що характеризують об'єкти класифікації, можливість формування навчальних вибірок; наявність інформації про класи еквівалентності.

Метою класифікації є виявлення всіх класів і віднесення об'єкта до одного з виявлених класів.

При проведенні класифікації визначається правило класифікації або критерій класифікації, за яким проводять віднесення до того чи іншого класу.

Характеристики якості класифікації обирають виходячи з правила або критерію класифікації і способів подання рішення, що приймається при класифікації. В залежності від специфіки прикладів класифікації характеристики якості можуть бути різними. Це можуть бути ймовірності помилок першого і другого роду, похідні від цих характеристик: чутливість і специфічність, відносний ризик, функція витрат. При складних випадках класифікації, коли рішення подається за кількома градаціями, це висновок за умовною ймовірністю віднесення до певного класу еквівалентності.

Досвід роботи вчених, що працюють в галузі класифікації, свідчить про те, що статистичні феномени, з якими зустрічається дослідник в галузі класифікації, складні і

не завжди передбачувані, тому дуже важливі напрацювання і рекомендації для різних випадків застосування процедури класифікації. Цим питанням присвячена представлена в дисертації робота автора.

#### 1.4. Огляд методів ідентифікації та класифікації розподілів вибірок малого об'єму.

Методи ідентифікації та класифікації розподілів вибірки можна поділити на такі основні групи.

В першу групу методів входять методи ідентифікації, засновані на критеріях узгодження. Ці критерії застосовують для оцінки близькості розподілу вибірки експериментальних даних до прийнятої аналітичної моделі закону розподілу. Є цілий ряд критеріїв, запропонований різними авторами, такі як: критерій Колмогорова, критерій  $\chi^2$ , критерій  $\omega^2$ , тощо (табл.1.2).

До другої групи належать методи класифікації розподілів вибірок за контрексесом і коефіцієнтом ентропії [5].

Таблиця 1.2 – Можливості критеріїв, що покладені в основу способу перевірки узгодження емпіричного розподілу з теоретичною моделлю

Критерій, що покладений в основу способу перевірки	Вид розподілу	Кількість елементів вибірки
Критерій Колмогорова	Будь який неперервний	>100
Критерій $\chi^2$	–	>200
Критерій $\omega^2$	Будь який неперервний	>50
Критерій $W$	Нормальний, лог-нормальний	3 ... 50
Складений критерій	Нормальний	15 ... 50

Інтегральні ознаки класифікації нормуються як лінійні функціонали від функції щільності ймовірності  $p(x)$ . Зокрема ними можуть служити нормовані моменти розподілів. При використанні четвертого моменту форма розподілу характеризується



значенням ексцесу  $\xi$ . Але так як ексцес різних розподілів коливається в нескінченних межах, то цей параметр незручний для цілей ідентифікації. Тому проводиться його нелінійне перетворення в значення контрексцесу:

$$\chi = \frac{1}{\sqrt{\xi}}, \quad (1.1)$$

яке для будь-яких розподілів укладено в межах від 0 до 1. Тобто в якості першої ознаки використовують значення контрексцесу. Для класифікації розподілів одного контрексцесу недостатньо. Тому необхідна друга ознака, в якості якої приймають значення ентропійного коефіцієнта:

$$k_e = \frac{\Delta_e}{\sigma}, \quad (1.2)$$

де  $\Delta_e$  – ентропійне значення.

При використанні цих двох ознак зображується точка з координатами  $k$  і  $\chi$  буде завжди знаходитися в межах прямокутника, обмеженого значеннями  $k$  від 0 до 2.066 і  $\chi$  від 0 до 1. Цим можна скористатися для винесення судження про вид досліджуваного розподілу за допомогою топографічної класифікації математичних моделей розподілів в координатах  $k$  і  $\chi$  [5].

Метод класифікації з використанням контрексцесу і ентропійного коефіцієнта дозволяє класифікувати велику кількість розподілів, але топографічна класифікація не дозволяє здійснити автоматичну класифікацію і тому в такому вигляді є неприйнятною для ІЗВ. Крім того, розглянутий вище метод розрахований на великий об'єм вибірки (30-200), що обмежує область його застосування.

До третьої групи відноситься метод непараметричної нечіткої ідентифікації виду розподілу вибірки малого об'єму (починаючи з числа вимірювань, рівного 9), що може бути використаний для великого числа розподілів і легко реалізується програмним шляхом [6].

Алгоритм нечіткої ідентифікації наведено на рис.1.2.

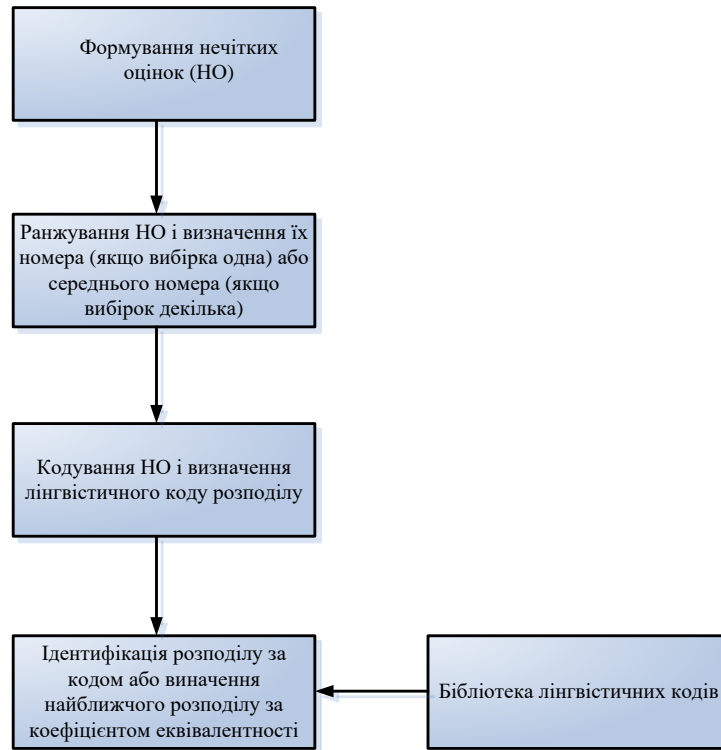


Рисунок 1.2 – Алгоритм нечіткої ідентифікації

При ідентифікації результати вимірювань випадкової величини  $q(i)$  розміщують в порядку зростання, тобто отримують впорядковану вибірку. Оскільки побудова нечітких оцінок заснована на використанні порядкових статистик з певними номерами, обсяг вибірки фіксується. В роботі [6] він становить 9, 17, 25, ... Нечіткі оцінки (НО) являють собою ознаки, що залежать від форми розподілу. В [6] для ідентифікації обрані шість оцінок, які для обсягу вибірки  $n=9$  рівні:

$$HO1 = HO(168) = \frac{1}{3}[q(1) + q(6) + q(8)] \quad (1.3)$$

$$HO2 = HO(249) = \frac{1}{3}[q(2) + q(4) + q(9)] \quad (1.4)$$

$$HO3 = HO(267) = \frac{1}{3}[q(2) + q(6) + q(7)] \quad (1.5)$$

$$HO4 = HO(348) = \frac{1}{3}[q(3) + q(4) + q(8)] \quad (1.6)$$

$$HO5 = HO(1478) = \frac{1}{4}[q(1) + q(4) + q(7) + q(8)] \quad (1.7)$$

$$HO6 = HO(2369) = \frac{1}{4}[q(2) + q(3) + q(6) + q(9)] \quad (1.8)$$

де  $q(i)$  – порядкова статистика з номером  $i$ .

Для обсягу вибірки  $n = 17$  значення нечітких оцінок обчислюються за формулами:

$$HO1 = \frac{1}{6}[q(1) + q(2) + q(10) + q(11) + q(14) + q(15)] \quad (1.9)$$

$$HO2 = \frac{1}{6}[q(3) + q(4) + q(7) + q(8) + q(16) + q(17)] \quad (1.10)$$

$$HO3 = \frac{1}{6}[q(3) + q(4) + q(10) + q(11) + q(12) + q(13)] \quad (1.11)$$

$$HO4 = \frac{1}{6}[q(5) + q(6) + q(7) + q(8) + q(14) + q(15)] \quad (1.12)$$

$$HO5 = \frac{1}{8}[q(1) + q(2) + q(7) + q(8) + q(12) + q(13) + q(14) + q(15)] \quad (1.13)$$

$$HO6 = \frac{1}{8}[q(3) + q(4) + q(5) + q(6) + q(10) + q(11) + q(16) + q(17)] \quad (1.14)$$

Отримані нечіткі оцінки ранжуються за зростанням і визначаються порядкові номери імен нечітких оцінок.

Кодування здійснюється за наступним принципом: якщо номер нечіткої оцінки менше номера медіани ряду оцінок, то цьому номеру присвоюється код 1, якщо більше, то присвоюється код 0.

Отриманий таким чином лінгвістичний код використовується для нечіткої ідентифікації розподілу досліджуваної вибірки за кодами, наведеними в табл.1.3.

Для прийняття рішення про належність вибірки до того чи іншого розподілу порозрядно порівнюють лінгвістичний код вибірки і еталонна. Відносна кількість елементів лінгвістичного коду, що збіглися за належністю в досліджуваному і «еталонному» термах, використовується в якості міри еквівалентності і позначається у вигляді

$$E_j = \frac{F_{xj}}{F} \quad (1.15)$$

де  $E_j$  – коефіцієнт еквівалентності,

$j$  – число еталонних термів (число моделей розподілів при ідентифікації),

$E_{xj}$  – число елементів аналізованого терму, що співпали з елементами еталонного терму,

$F$  – число елементів еталонного терму або число нечітких оцінок.

Якщо  $E_j = 1$ , то  $F_{xj} = F$ , при  $0 < E_j < 1$   $F_{xj}$  нечітко дорівнює  $F$ .

Таблиця 1.3 – Двійкові лінгвістичні коди для різних розподілів [6]

Види розподілу	Розряди вертикального і горизонтального лінгвістичних кодів					
	1	2	3	4	5	6
Коши	1	1	1	0	0	0
	1	1	1	0	0	0
АРК	0	0	0	1	1	1
	0	0	0	1	1	1
ЛЕКС	1	1	0	0	0	1
	1	1	0	0	1	0
ЕКСП	0	1	1	1	0	0
	1	0	1	1	0	0
2МОД	0	1	0	1	0	1
	0	0	1	0	1	1

Але метод нечіткої ідентифікації розроблений тільки для вибірок фіксованого об'єму, тому потрібно розширити його функції для різного об'єму вибірок. Також процедура ідентифікації розроблена для обмеженої кількості розподілів [6]. Тому в дисертаційній роботі виникає необхідність проведення дослідження для розподілів вибірок, що застосовуються в вимірювальній практиці. Оскільки немає відомостей про якість ідентифікації, треба провести аналіз достовірності ідентифікації. Тому подальше дослідження особливостей і удосконалення нечіткої ідентифікації у вибірках малого об'єму, спрямоване на розширення області застосування нечіткої ідентифікації і дослідження вірогідності розрізнення розподілів.

## 1.5. Розширення можливостей непараметричної нечіткої ідентифікації для різних об'ємів вибірки

Нечітка ідентифікація, що продемонстрована в роботі [6], розроблена тільки для фіксованого об'єму вибірки ( $n=9$ ,  $n=17$ ). Тому в даній роботі проведене подальше дослідження алгоритму нечіткої ідентифікації, розширення його можливостей дослідження.

При різних об'ємах вибірки буде змінюватись і кількість нечітких оцінок. Тому постає питання формування алгоритму знаходження НО.

При підборі нечітких оцінок необхідно дотримуватися кількох умов:

- оцінки повинні бути лінгвістично еквівалентними, тобто середнє значення їх індексів має дорівнювати медіані;
- оцінки повинні бути незалежними, тобто такими, які не можна було б отримати лінійною комбінацією інших НО;
- оцінки повинні бути асиметричними, тобто такими, в яких індекси елементів розташовуються несиметрично щодо медіани.
- оцінки повинні бути попарно симетричними, тобто в наборі оцінок кожна оцінка повинна мати собі пару, що дзеркально відображає її щодо медіани.

Оцінка, що складається з двох значень, завжди буде симетричною, тому мінімальний розмір асиметричної нечіткої оцінки – три індекси. Тоді сума їх індексів повинна дорівнювати потроєному індексу медіани ( $3M$ ). Знайдемо три значення, які будуть відповідати умові лінгвістичної еквівалентності. Для цього використовуємо вибірку, що складається з 11 елементів, її медіана  $M = 6$ , а  $3M = 18$ . Першим візьмемо максимальне значення 11. Можна помітити, що більше ми не можемо брати значення по праву сторону від медіани (більше медіани), так як навіть мінімальне число  $7 + 11 = 18 = 3M$ .

$$\begin{array}{ccccccccccc} \circ & \circ & \circ & \circ & \circ & \bullet & \bullet & \circ & \circ & \circ & \bullet \\ & & & & & 7 & + & 11 & = & 18 & \end{array}$$

Тому візьмемо максимальний індекс з вибірки, який менше медіани, тобто 5.  $18 - 11 - 5 = 2$ . Отримаємо першу нечітку оцінку з ім'ям 2-5-11.



Тепер якщо ми до лівого значення додамо одиницю, а від правого віднімемо одиницю, то отримаємо наступну еквівалентну оцінку:  $2+1=3$ ,  $11-1=10$ .



Вона називається 3-5-10, сума індексів цієї оцінки  $3+5+10=18=3M$ . Аналогічно знаходимо ще одну оцінку 4-5-9.



Як ми можемо бачити лівий і центральний індекси зійшлися, і при наступному повторенні візьмуть однакове значення.



Тепер зменшимо значення середньої оцінки на 1, а праву знову візьмемо 11.



$18 - 4 - 11 = 3$ . Нова оцінка 3-4-11. Інших комбінацій, при яких два індекси менше медіани, немає.



Якщо розглянути асиметричні оцінки для 9-ти вимірювань, то можна помітити, що вони дзеркально відображають одна одну щодо медіани. Наприклад, для оцінки 2-5-11 дзеркально симетричною буде 1-7-10.



Знайдемо дзеркальні оцінки для вже виведених. Для 3-5-10 - це 2-7-9, для 4-5-9 – це 3-7-8 і для 3-4-11 симетричною буде 1-8-9. В результаті отримаємо весь набір трьохелементних оцінок для вибірки в 11 значень.

1	○ ● ○ ○ ○ ● ○ ○ ○ ○ ○ ●	2 5 11
2	● ○ ○ ○ ○ ○ ● ○ ○ ○ ● ○	1 7 10
3	○ ○ ○ ● ○ ● ○ ○ ○ ○ ● ○	3 5 10
4	○ ● ○ ○ ○ ○ ● ○ ● ○ ○ ○	2 7 9
5	○ ○ ○ ● ● ○ ○ ○ ○ ● ○ ○	4 5 9
6	○ ○ ● ○ ○ ○ ● ● ● ○ ○ ○	3 7 8
7	○ ○ ○ ● ● ○ ○ ○ ○ ○ ○ ●	3 4 11
8	● ○ ○ ○ ○ ○ ● ○ ● ● ○ ○	1 8 9

Таким чином методика підбору нечітких оцінок для трьох індексів наступна: беремо максимально близьке значення до медіани з будь-якої сторони, і найдалші від нього можливі значення. Тоді перша оцінка буде 2 M-1 N. Далі «зрушуємо» крайні індекси на одиницю до медіани, отримуємо 3 M-1 N-1. Так зрушуємо до тих пір, поки ліва оцінка не стане дорівнювати M-2. Потім центральний індекс зменшуємо на одиницю і знову беремо найдалші від нього можливі значення. Отримуємо оцінку 3 M-2 N і, якщо можливо, повторюємо операцію. Таким перебором знаходимо всі можливі асиметричні оцінки для одного боку вибірки, після чого дзеркально відображаємо їх щодо медіани і отримуємо всі можливі оцінки для іншої сторони вибірки.

На основі даних міркувань можна вивести формули шести оцінок, необхідних для знаходження розподілу. За цими формулами можна обчислити НО для будь-якого числа вимірів.

Таким чином, отримуємо рекомендації для імен нечітких оцінок при різних об'ємах вибірок.

Отож, за допомогою сформованого алгоритму знаходження НО були отримані формули для оцінок, які можна використати для будь якого об'єму вибірки (табл. 1.4).

Таблиця 1.4 – Рекомендації для імен НО при різних об'ємах вибірок

Імена оцінок при різних об'ємах вибірок N				
N=9	N=11	N=13	N=15	N=25
1-6-8	1-8-9	2-6-13	2-7-15	4-16-24
2-4-9	3-4-11	3-6-12	3-7-14	5-13-25
2-6-7	1-7-10	3-5-13	3-6-15	4-19-22
3-4-8	2-5-11	1-9-11	1-10-13	6-12-23
1-4-7-8	2-5-7-9	2-5-8-11	2-6-9-13	3-10-21-24
2-3-6-9	3-5-8-10	1-4-8-12	1-6-9-14	4-9-18-23

### **1.6. Дослідження нечіткої ідентифікації для різних розподілів вибірок з метою визначення класифікаційних ознак.**

Для побудови класифікації на основі нечіткої ідентифікації були використані «навчальні» (еталонні) вибірки, що були «входом» задачі класифікації. На «виході» задачі повинен бути клас або перелік об'єктів, що створює цей клас. Якщо дослідник має і «входи» і «виходи» задачі класифікації, то таку інформацію називають навчальною і метою дослідження є опис процедур, за допомогою яких при наявності тільки вхідних даних нового об'єкту, його можна було б віднести з найбільшою (в певному сенсі) точністю до одного з класів. Однією з головних задач в цьому випадку є відбір ознак, за якими будуть формуватись класи.

Для дослідження нечіткої ідентифікації були проаналізовані лінгвістичні коди видів розподілу. Автором даної роботи були отримані еталонні вибірки для розподілів Коші, Лапласа, експоненціального та арксинусного. Ці розподіли відрізнялися найбільшою стабільністю середніх номерів нечітких оцінок.

При дослідженні якості нечіткої ідентифікації, проведеної автором роботи, лінгвістичні коди («вихід») були отримані теоретично за навчальною (еталонною) вибіркою («вхід»), яка, в свою чергу, була отримана за допомогою моделюючого співвідношення  $\xi = \varphi(\alpha)$ , де  $\varphi(\alpha)$  – строго монотонна і неперервна функція на інтервалі



(0,1); для  $\xi$  задана щільність розподілу  $f(x)$  ( $a < x < b$ ), причому межі  $a$  і  $b$  можуть бути і нескінченними.

При припущенні монотонного зростання  $\varphi(\alpha)$  отримують єдину моделюючу формулу:

$$\xi = F^{-1}(\alpha) \quad (1.16)$$

де  $F(\alpha)$  – функція розподілу ймовірностей. У разі монотонного спадання  $\varphi(\alpha)$  є також єдина моделююча формула

$$\xi = F^{-1}(1 - \alpha) \quad (1.17)$$



Рисунок 1.3 – Алгоритм отримання лінгвістичного коду

### 1.6.1. Отримання навчальних (еталонних) вибірок, нечітких оцінок і лінгвістичних кодів для різних розподілів.

Отримання навчальної (еталонної) вибірки, НО і ЛК для двостороннього розподілу Лапласа проводиться наступним чином.

Функція щільності ймовірності розподілу Лапласа [7, 8]:

$$f(x) = \frac{\lambda}{2} \cdot e^{-\lambda|x|} \quad (1.18)$$

де  $\lambda$  – параметр розподілу.

Функція розподілу ймовірностей для монотонного зростання  $\varphi(\alpha)$  дорівнює

$$F(x) = \frac{1}{2} \cdot e^{\lambda x} \quad (1.19)$$

Використовуючи формулу (1.16), отримуємо

$$\xi = F^{-1}(\alpha)$$

$$\alpha = \frac{1}{2} \cdot e^{\lambda \xi}$$

$$\xi = \frac{1}{\lambda} \ln 2\alpha \quad (1.20)$$

Формула (1.20) являє собою моделююче співвідношення, що дозволяє отримати випадкову величину з розподілом Лапласа

$$F(x) = \frac{1}{2} \cdot e^{-\lambda x} \quad (1.21)$$

Використовуючи формулу (1.17), отримаємо

$$\xi = -\frac{1}{\lambda} \ln(1-\alpha) \quad (1.22)$$

Якщо об'єм вибірки  $n$ , то значення  $\alpha$  відповідно дорівнюють:

$$\alpha = \left\{ \frac{1}{n+1}; \frac{2}{n+1}; \dots; \frac{n}{n+1} \right\}. \text{ Тобто при } n=9 \quad \alpha = \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9\}.$$

Для значень  $\alpha = \{0,1; 0,2; 0,3; 0,4; 0,5\}$  використовуємо формулу (1.20), для  $\alpha = \{0,5; 0,6; 0,7; 0,8; 0,9\}$  використовуємо формулу (1.22).

Тоді «еталонна» вибірка розподілу Лапласа має вигляд:

$$\xi = \frac{1}{\lambda} (-1,61; -0,92; -0,51; -0,22; 0; 0,22; 0,51; 0,92; 1,61)$$

Значення нечітких оцінок, отриманні по «еталонній» вибірці, наступні:

$$HO1 = -0,138 \cdot \frac{1}{\lambda}, \quad HO2 = 0,138 \cdot \frac{1}{\lambda}, \quad HO3 = -0,062 \cdot \frac{1}{\lambda}, \quad HO4 = 0,062 \cdot \frac{1}{\lambda}, \quad HO5 = -0,101 \cdot \frac{1}{\lambda},$$

$$HO6 = 0,101 \cdot \frac{1}{\lambda}.$$

Номери імен нечітких оцінок:

$$N(HO1) = 1, \quad N(HO2) = 6, \quad N(HO3) = 3, \quad N(HO4) = 4, \quad N(HO5) = 2, \quad N(HO6) = 5.$$

Лінгвістичний код розподілу Лапласа «101010».

Отримання еталонних вибірок, НО і ЛК для розподілів Коши проводиться наступним чином.

Функція щільності ймовірності розподілу Коши [7, 8]:

$$f(x) = \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + x^2},$$

$\lambda$  - параметр розподілу Коши.

Для розподілу Коши у випадку монотонного зростання щільності ймовірності, отримаємо:

$$F(x) = \int_{-\infty}^x \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + u^2} du = \frac{1}{\pi} \operatorname{arctg} \frac{u}{\lambda} \Big|_{-\infty}^x = \frac{1}{\pi} \operatorname{arctg} \frac{x}{\lambda} + \frac{1}{2}.$$

У випадку монотонного спадання щільності ймовірності, отримаємо:

$$F(x) = \frac{1}{2} + \int_0^x \frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + u^2} du = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{x}{\lambda}.$$

Таким чином, вся область значень  $\xi$  може бути отримана з допомогою виразу:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} \frac{x}{\lambda} = \alpha \quad (1.26)$$

$$\xi = \lambda \cdot \operatorname{tg} \left[ \pi \left( \alpha - \frac{1}{2} \right) \right] \quad (1.27)$$

Тоді «еталонна» вибірка розподілу Коши має вигляд:

$$\xi = \lambda \cdot (-2, 1; -1, 13; -0, 63; -0, 29; 0; 0, 29; 0, 63; 1, 13; 2, 1)$$

Значення нечітких оцінок, отримані по «еталонній» вибірці, наступні:

$$HO1 = -0,23 \cdot \lambda, \quad HO2 = 0,23 \cdot \lambda, \quad HO3 = -0,07 \cdot \lambda, \quad HO4 = 0,07 \cdot \lambda, \quad HO5 = -0,16 \cdot \lambda,$$

$$HO6 = 0,16 \cdot \lambda.$$

Номери імен нечітких оцінок:

$$N(HO1) = 1, \quad N(HO2) = 6, \quad N(HO3) = 3, \quad N(HO4) = 4, \quad N(HO5) = 2, \quad N(HO6) = 5.$$

Лінгвістичний код розподілу Коши «101010».

Отримання еталонної вибірки, НО і ЛК для арксинусного розподілу наступним чином.

Функція щільності ймовірності арксинусного розподілу [7, 8]:

$$f(x) = \frac{1}{\pi \sqrt{a^2 - x^2}} \quad (1.28)$$

У випадку монотонного зростання і спадання функції щільності ймовірності в діапазоні  $\pm a$ , отримаємо:

$$F(x) = \frac{1}{\pi} \arcsin \frac{x}{a} + \frac{1}{2} \quad (1.29)$$

$$\xi = a \cdot \sin \pi \left( \alpha - \frac{1}{2} \right) \quad (1.30)$$

Тоді «еталонна» вибірка арксинусного розподілу має вигляд (з врахуванням того, що арксинусний розподіл обмежений – граничні члени вибірки дорівнюють  $\pm a$ )

$$\xi = a \cdot (-1; -0,87; -0,65; -0,34; 0; 0,34; 0,65; 0,87; 1)$$

Значення нечітких оцінок, отримані за «еталонною» вибіркою, наступні:

$$HO1 = 0,07 \cdot a; \quad HO2 = -0,07 \cdot a; \quad HO3 = 0,04 \cdot a; \quad HO4 = -0,04 \cdot a; \quad HO5 = 0,045 \cdot a; \quad HO6 = -0,045 \cdot a$$

Номери імен нечітких оцінок:

$$N(HO1) = 6; \quad N(HO2) = 1; \quad N(HO3) = 4; \quad N(HO4) = 3; \quad N(HO5) = 5; \quad N(HO6) = 2$$

Лінгвістичний код арксинусного розподілу «010101».

Отримання еталонної вибірки, НО і ЛК для одностороннього експоненціального розподілу наступним чином.

Функція щільності ймовірності експоненціального розподілу [7, 8]:

$$f(x) = \lambda \cdot e^{-\lambda x}, \quad x \geq 0 \quad (1.31)$$

Функція розподілу ймовірностей

$$F(x) = \int_0^x \lambda \cdot e^{-\lambda u} = 1 - e^{-\lambda x} \quad (1.32)$$

Тоді «еталонну» вибірку знаходять із співвідношення:

$$\xi = \frac{1}{\lambda} \ln(1 - \alpha) \quad (1.33)$$

Для значень  $\alpha = \{0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9\}$ , отримуємо:

$$\xi = \frac{1}{\lambda} (0,1; 0,22; 0,36; 0,51; 0,69; 0,92; 1,2; 1,61; 2,3)$$

Значення нечітких оцінок, отримані за «еталонною» вибіркою, наступні:

$$HO1 = 0,88 \cdot \frac{1}{\lambda}; \quad HO2 = 1,01 \cdot \frac{1}{\lambda}; \quad HO3 = 0,78 \cdot \frac{1}{\lambda}; \quad HO4 = 0,83 \cdot \frac{1}{\lambda}; \quad HO5 = 0,86 \cdot \frac{1}{\lambda}; \quad HO6 = 0,95 \cdot \frac{1}{\lambda}$$

Номери імен нечітких оцінок:

$$N(HO1) = 4; N(HO2) = 6; N(HO3) = 1; N(HO4) = 2; N(HO5) = 3; N(HO6) = 5.$$

Лінгвістичний код експоненціального розподілу «001110».

Отримання еталонної вибірки, НО і ЛК для одностороннього розподілу Лапласа наступним чином.

Аналогічний аналіз можна провести для розподілу з щільністю

$$f(x) = \lambda \cdot e^{\lambda x}, \quad x \leq 0 \quad (1.34)$$

В [2] цей розподіл називають розподілом Лапласа.

"Еталонна" вибірка для такого розподілу має вигляд:

$$\xi = \frac{1}{\lambda} (-2,3; -1,61; -1,2; -0,92; -0,69; -0,51; -0,36; -0,22; -0,1)$$

Значення нечітких оцінок, отримані за "еталонною" вибіркою, наступні:

$$HO1 = -1,01 \cdot \frac{1}{\lambda}; \quad HO2 = -0,88 \cdot \frac{1}{\lambda}; \quad HO3 = -0,83 \cdot \frac{1}{\lambda}; \quad HO4 = -0,78 \cdot \frac{1}{\lambda}; \quad HO5 = -0,95 \cdot \frac{1}{\lambda};$$

$$HO6 = -0,86 \cdot \frac{1}{\lambda}.$$

Номера імен нечітких оцінок:

$$N(HO1) = 1; N(HO2) = 3; N(HO3) = 5; N(HO4) = 6; N(HO5) = 2; N(HO6) = 4.$$

Лінгвістичний код розподілу «110010».

На основі отриманих оцінок отримано бібліотеку всіх лінгвістичних кодів для розподілів, яка значно спростить процедуру класифікації форми розподілу.

### **1.6.2. Формування класів еквівалентності для непараметричної класифікації розподілів вибірок.**

Отримані результати за розділом 1.6.1 занесені в таблицю 1.5.

Як видно із результатів, наведених в таблиці, нечітка ідентифікація не дозволяє отримати однозначний результат за лінгвістичним кодом при розподілах одного класу (наочно це видно по ЛК нормального розподілу і розподілу Коши).

Тому клас розподілів визначається як одномодальний симетричний.

Таблиця 1.5 – Лінгвістичні коди для різних видів розподілів  
при об'ємі вибірки  $N=9$

Найменування розподілу	Лінгвістичні коди	Клас еквівалентності
Коши	101010	Одноmodalьний симетричний
Нормальний	101010	
Двосторонній Лапласа	101010	
Сімпсона	101010	
Арксинусний	010101	Антимодальний
Рівномірний	110001	Розподіл типу Шапо
Односторонній Лапласа	110010	Правосторонній
Експоненціальний	001110	Лівосторонній

Проведений аналіз на опорних вибірках показав, що в дійсності нечітка ідентифікація дозволяє розрізнити розподіли з різних класів еквівалентності.

Остаточний розподіл на класи, що наведений в табл.1.5, складається з наступних класів еквівалентності: одноmodalьний симетричний, антимодальний, розподіл типу Шапо, правосторонній одноmodalьний, лівосторонній одноmodalьний.

Вибір ознак, за якими проведено класифікацію, наступний [6]: симетричність розподілу (якщо ліва половина кривої щільності ймовірності є дзеркальним відображенням її правої половини), кількість максимумів кривої щільності ймовірності, що відповідають модам розподілу (табл.1.6).

За другою ознакою, як і в роботі [5], розподіли розділено на безmodalьні, одноmodalьні і двохmodalьні.

В свою чергу одноmodalьні розподіли розділено за розміщенням моди: мода в центрі розподілу, мода в лівій половині розподілу, мода в правій половині розподілу.

Таблиця 1.6 – Сукупність ознак класифікації розподілів

Ознаки класифікації розподілів	Шифр ознаки	
Симетричність розподілу	Симетричний (С)	Мода в центрі розподілу (ОЦ)
	Несиметричний (Н)	
Кількість максимумів кривої щільності ймовірності	Безмодальні (Б)	Мода в правій стороні розподілу (ОП)
	Одномодальні (О)	Мода в лівій стороні розподілу (ОЛ)
	Двохмодальні (Д)	

Загальна матриця «ознаки» – «класи» наведена в табл.1.7. З таблиці видно, що вектор ознак у кожного класу різний, тобто вони не перерізаються. В свою чергу це відповідає наступним властивостям класифікації:  $\bigcup_{i=1}^L M_i = M$ ;  $M_i \cap M_j = \emptyset$  для  $\forall i, j \in \{1, \overline{L}\}$ , де множина  $M$  об'єднує родину  $\{M_1, \dots, M_L\}$  непустих підмножин  $M_i \neq \emptyset$ ,  $M_i \subset M$ , що попарно не перерізаються. У цьому випадку на множині  $M$  визначено відношення еквівалентності, а підмножини  $M_i$  є класами еквівалентності.

Таблиця 1.7 – Матриця «ознаки» – «класи еквівалентності»

Шифр ознак	Наявність ознаки за класами еквівалентності				
	Одномодальний симетричний	Антимодальний	Розподіл типу Шапо	Одномодальний лівосторонній	Одномодальний правосторонній
С	+	+	+	-	-
Н	-	-	-	+	+
Б	-	-	+	-	-
ОЦ	+	-	-	-	-
ОП	-	-	-	-	+
ОЛ	-	-	-	+	-
Д	-	+	-	-	-

У даному випадку прийнятий чіткий вектор ознак. А вихідною інформацією для класифікації є нечіткі оцінки і функція належності, що визначається за місцями нечітких оцінок при їх ранжуванні. Тому класифікація за наведеними вище класами

еквівалентності є нечіткою, і для оцінки її якості повинна бути визначена достовірність класифікації.

### 1.7. Достовірність класифікації.

Для аналізу достовірності класифікації використані результати попереднього теоретичного дослідження і подальшого експериментального дослідження.

Аналіз ЛК, що відповідають певним класам розподілів, показав, що неправильна класифікація можлива при зміні 0 в 1 (чи навпаки) третьої і четвертої позицій ЛК вибірки. Розгляд цих змін показав, що новий ЛК не співпадає ні з яким ЛК для всіх класів. Тому результатом чіткої класифікації буде віднесення до певного класу еквівалентності або нечітка класифікація за частковим співпаданням ЛК або «класифікація неможлива». Нечіткий висновок про близькість до певного класу еквівалентності був не завжди можливим, оскільки найбільший коефіцієнт еквівалентності (схожості) спостерігався до більше одного класів.

Наприклад, при зміні ЛК (початковий клас одномодальний симетричний) на 100110 отримано:

- клас одномодальний симетричний,  $K_{екв} = 0,66$ ;
- клас одномодальний,  $K_{екв} = 0,33$ ;
- клас типу Шапо,  $K_{екв} = 0,33$ ;
- клас правосторонній, несиметричний,  $K_{екв} = 0,66$ ;
- клас лівосторонній, несиметричний,  $K_{екв} = 0,66$ .

Таким чином, отримуємо результат «класифікація неможлива».

Зважаючи на вищенаведене для характеристики достовірності класифікації, обрано ймовірність правильного віднесення до певного класу еквівалентності.

При одному переході  $0 \rightarrow 1$  можливе отримання нечіткого результату за коефіцієнтом еквівалентності, що свідчить про близькість розподілу вибірки до певного класу еквівалентності.



Теоретично достовірність класифікації визначається як ймовірність правильної класифікації  $P_H$ , або як  $1 - P_H$ , де  $P_H$  - ймовірність неправильної класифікації.

### 1.7.1. Теоретична оцінка достовірності.

Задачею теоретичного аналізу достовірності є оцінка ймовірності переходу  $0 \rightarrow 1$  в ЛК вибірки при аналізі однієї вибірки. Це відповідає рис.1.4, на якому наведено нечіткі оцінки, що розташовані в районі медіани, за якою приймається рішення 0 або 1.

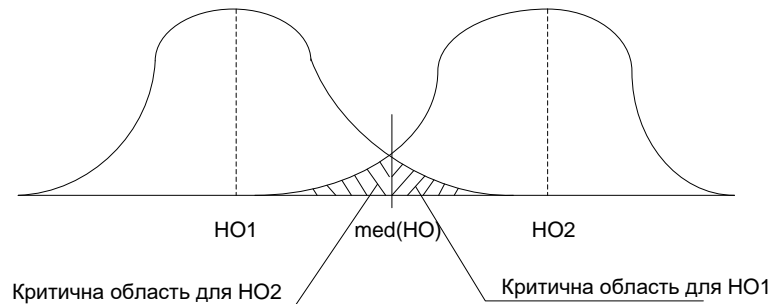


Рисунок 1.4 – Ілюстрація одночасного знаходження нечітких оцінок в критичних областях

де  $НО1$  – оцінка  $НО_{2369}$ ,  $НО2$  – оцінка  $НО_{1478}$ .

НО апроксимуються нормальним розподілом, дисперсія якого визначається за дисперсією порядкових статистик, що входять до складу НО.

Значення порядкових статистик підлягають статистичним змінам. Математичною моделлю цих змін може бути бета-розподіл, що відповідає розподілу порядкових статистик [7, 8].

Якщо  $\xi_1, \xi_2, \dots, \xi_n$  незалежні і рівномірно розподілені на відрізку  $[0, 1]$  і  $\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(n)}$  - впорядковані за зростанням величини, де  $\xi_{(k)}$  ( $k = 1, \dots, n$ ) -  $k$ -та порядкова статистика, то щільність ймовірності  $f_{(k)}(x)$   $k$ -ї порядкової статистики має  $\beta$ -розподіл з  $\alpha = k$ ,  $\beta = n - k + 1$ .

$$D_{(\xi)} = \frac{\alpha \cdot \beta}{(\alpha + \beta)^2 \cdot (\alpha + \beta + 1)}.$$

Тому дисперсія порядкових статистик для нечіткої класифікації розраховується за формулою:

$$D_k = \frac{k \cdot (n - k + 1)}{(n + 1)^2 \cdot (n + 2)}. \quad (1.35)$$

Достовірність класифікації визначається за можливими змінами значень порядкових статистик, що приводить до змін НО.

Дисперсія нечітких оцінок з урахуванням (1.35) розраховується за сумою порядкових статистик. Наприклад, для імені НО(1478) вибірки з рівномірним розподілом

$$D_{1478} = \frac{1}{16} [D_1 + D_4 + D_7 + D_8], \text{ де за формулою (1.35):}$$

$$D_1 = 0,82 \cdot 10^{-2}, \quad D_4 = 2,18 \cdot 10^{-2}, \quad D_7 = 1,91 \cdot 10^{-2}, \quad D_8 = 1,45 \cdot 10^{-2}.$$

$$\text{Тоді } D_{1478} = 0,398 \cdot 10^{-2}$$

Отримані результати для інших НО приведені в таблиці 1.8.

Таблиця 1.8 – НО і їх дисперсії для вибірки з рівномірним розподілом

Імена	168	267	2369	<i>med</i>	1478	348	249
Оцінки	0,399	0,403	0,4745	0,499	0,523	0,595	0,598
Дисперсії	$0,494 \cdot 10^{-2}$	$0,616 \cdot 10^{-2}$	$0,398 \cdot 10^{-2}$	–	$0,398 \cdot 10^{-2}$	$0,616 \cdot 10^{-2}$	$0,494 \cdot 10^{-2}$

Для оцінки ймовірності правильної класифікації визначається критична відстань, тобто відстань  $h_{kp}$  між НО, найближчими до медіани.

Для отриманої таблиці:

$$h_{kp} = NO(1478) - med = med - NO(2369).$$

Так як НО складається з 3-х або 4-х порядкових статистик, то розподіл нечіткої оцінки можна прийняти приблизно нормальним з дисперсією, що наведена в таблиці 1.7.

Тоді ймовірність зміни лінгвістичного коду визначатиметься за умов одночасного знаходження НО в критичних областях:

$$P = P_{11} + P_{12} + P_{13},$$

де  $P_{11}$  – ймовірність одночасного знаходження НО в критичних областях,  $P_{12}$  та  $P_{13}$  – добуток ймовірності знаходження однієї оцінки в критичній відстані на умовну ймовірність, що одна оцінка менша за другу.

$$P_{11} = P_1 \cdot P_2,$$

де  $P_1$  – ймовірність знаходження  $HO_{1478}$  за областю  $HO_{1478} - med$  від  $HO_{1478}$ , а  $P_2$  – ймовірність знаходження  $HO_{2369}$  за областю  $med - HO_{2369}$  від  $HO_{2369}$ ,  $P_{11}$  – ймовірність одночасного знаходження оцінок в наведених областях.

Ймовірність знаходження однієї оцінки в критичній області:  $1 - P_H$ , де за умов наближеності розподілу НО до нормального.

$$P_1 = 0,5 - \Phi\left(\frac{h_{кр}}{\sqrt{D_{1478}}}\right).$$

За умов симетрії оцінок (рис.1.4):

$$P_2 = P_1.$$

За наведеними на рис.1.4 оцінками, отримаємо:

$$P_1 = 0,308, \quad P_{11} = P_1 \cdot P_2 = 0,095 \cdot 100\% = 9,5\%$$

Також потрібно врахувати, коли оцінки знаходяться в наведеній області не одночасно. Відповідно:

$$P_{12} = P(x_1 > x) \cdot P(x_2 < x), \quad \text{де } P(x_1 > x) - \text{знаходження } HO_{1478} \text{ в критичній області,}$$

$$P_{13} = P(x_1 < x) \cdot P(x_2 > x), \quad \text{де } P(x_2 > x) - \text{знаходження } HO_{2369} \text{ в критичній області.}$$

За допомогою числового інтегрування, отримаємо:

$$P_{12} = 0,07, \quad P_{13} = 0,07$$

$$P = P_{11} + P_{12} + P_{13} = 0,24.$$

Ймовірність правильної класифікації:  $(1 - P) \cdot 100\% = (1 - 0,24) \cdot 100\% = 76\%$ .

Результати оцінки достовірності класифікації, що були проведені для інших розподілів за однією вибіркою, приведені в таблиці 1.9.

Таблиця 1.9 – Ймовірність правильної класифікації розподілу за теоретичним дослідженням

Найменування розподілу	Ймовірність зміни лінгвістичного коду	Ймовірність правильної класифікації
Нормальний	20%	80%
Коши	22%	78%
Рівномірний	24%	76%
Лапласа	18%	82%
Арксинусний	16%	84%

### 1.7.2. Експериментальна оцінка достовірності при моделюванні процедури класифікації.

Експериментальне дослідження достовірності були проведені за наступним алгоритмом моделювання для нормального розподілу вибірок (рис.1.5):

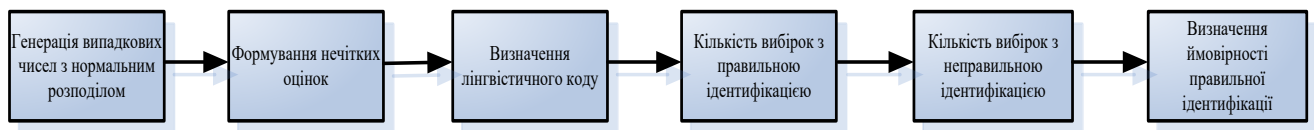


Рисунок 1.5 – Алгоритм визначення ймовірності правильної класифікації

Послідовно генерувались  $n$  вибірок випадкових чисел з нормальним розподілом, об'єм кожної вибірки дорівнював 9 (табл.1 додатку).

Об'єм вибірки при моделюванні визначався наступним чином.

В [5] для статистичної перевірки гіпотези відносно ймовірності наведено наступну формулу для середнього квадратичного відхилення оцінки ймовірності  $\tilde{P}$  за вибіркою об'ємом  $n$ :

$$\sigma(\tilde{P}) = \frac{\sqrt{n\tilde{P}q}}{n\tilde{P}} = \sqrt{\frac{q}{n\tilde{P}}},$$

де  $q = 1 - \tilde{P}$ .

Якщо в середньому ймовірність правильної класифікації за однією вибіркою становить  $P = 0,8$ , то для вибірки  $n = 100$

$$\sigma(\tilde{P}) = \sqrt{\frac{0,2}{100 \cdot 0,8}}, \text{ а довірчий інтервал } \pm 0,07.$$

Для  $n = 200$ ,  $\sigma(\tilde{P}) = 0,027$ , довірчий інтервал становить  $\pm 0,054$ , що гарантує знаходження ймовірності в інтервалі  $0,75 \div 0,85$ , що і було використано для перевірки збіжності теоретичних результатів і моделювання.

Моделювання процедури класифікації було проведено за 200 вибірками (додаток А). Отримані вибірки ранжувались (табл.2 додатку) і за цими даними визначались НО і лівістичний код (табл.3 додатку). Якщо отриманий ЛК співпадав з ЛК нормального розподілу, тоді знак "+" означав факт правильної класифікації.

Оцінка ймовірності правильної класифікації визначалась як  $\tilde{P} = \frac{n(+)}{n}$ , де  $n$  – загальна кількість вибірок,  $n_+$  – кількість результатів правильної класифікації.

Фрагмент моделювання класифікації для перших десяти вибірок наведено в таблиці 1.10.

Таблиця 1.10 – Результати моделювання класифікації

Нечіткі оцінки	«168»	«267»	«249»	«348»	«1478»	«2369	Med	ЛК	Правильна класифікація
№1	-0,1465	0,2053	0,6865	0,8280	0,3840	0,4695	0,4267	110010	+
№2	-0,5768	-0,7155	0,1000	0,0988	-0,3575	-0,3568	-0,3846	110010	+
№3	-0,6883	-0,5800	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№4	-0,3168	-0,2803	0,8638	0,3950	0,1253	0,2508	0,188	110010	+
№5	-0,5123	-0,5390	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№6	-1,0560	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,193	110010	+
№7	-0,2823	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№8	-0,7983	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№9	-0,9075	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№10	-0,5455	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+

Ймовірність правильної класифікації за моделюванням становить 0,79, що збігається з результатами, отриманими за таблицею 1.9 для нормального розподілу.

Розрахунок дисперсій НО показав, що розсіяння НО при об'ємі вибірки  $n=9$  доволі велике і зменшується при збільшенні об'єму вибірки. Тому зменшення об'єму вибірки від 9 недоцільне, тому що достовірність класифікації різко зменшується. При  $n=17$  ймовірність неправильної класифікації зменшується вдвоє, при  $n=25$  вона незначна.

Для зменшення дисперсії НО рекомендується проводити непараметричну класифікацію за декількома вибірками малого об'єму.

Тому подальше дослідження присвячене визначенню необхідної кількості вибірок за заданою ймовірністю правильної класифікації.

### **1.8. Визначення необхідної кількості вибірок за заданою ймовірністю правильної класифікації (при малому об'ємі вибірки).**

Проведений автором аналіз достовірності класифікації дозволяє визначити необхідну кількість вибірок за умов заданої вірогідності класифікації.

Якщо задати, наприклад, ймовірність правильної класифікації  $P_k = 95\%$ , то можна визначити необхідну кількість вибірок.

Задана ймовірність зміни лінгвістичного коду визначається як:  $P = 1 - P_k$

Тоді за заданою ймовірністю  $P$  можна отримати вимоги до ймовірності знаходження однієї НО в критичній області навколо медіани, використовуючи апроксимацію:

$$\sqrt{1 - P_k} = 0,5 - \Phi\left(\frac{h_{kp}}{\sqrt{D_{HO}}}\right),$$

При використанні декількох вибірок отримуємо:

$$\sqrt{1 - P_k} = 0,5 - \Phi\left(\frac{h_{kp} \sqrt{n}}{\sqrt{D_{HO}}}\right)$$

$$\Phi\left(\frac{h_{kp} \sqrt{n}}{\sqrt{D_{HO}}}\right) = 0,5 - \sqrt{1 - P_k}$$

$$\frac{h_{кр} \sqrt{n}}{\sqrt{D_{HO}}} = \Phi^{-1}(0,5 - \sqrt{1 - P_k})$$

звідси

$$n = \frac{D_{HO} [\Phi^{-1}(0,5 - \sqrt{1 - P_k})]^2}{h_{кр}^2}$$

Для умов  $D_{HO} = D_{1478}$ :

$$h_{кр} = 0,523 - 0,499 = 0,024,$$

$$P_k = 0,98$$

$$\Phi^{-1}(0,5 - \sqrt{1 - 0,98}) = \Phi^{-1}(0,36) = 1,1,$$

$$n = \frac{D_{HO} [\Phi^{-1}(0,5 - \sqrt{1 - P_k})]^2}{h_{кр}^2} \approx 10.$$

Результати розрахунків кількості вибірок, що були проведені для нормального і інших розподілів, приведені в таблиці 1.11.

Таблиця 1.11 – Рекомендована кількість вибірок для різних розподілів за заданою вірогідністю класифікації

Назва розподілу	Кількість вибірок за ймовірністю правильної класифікації при $n = 9$	
	$P_k = 95\%$	$P_k = 99\%$
Нормальний	10	18
Коши	9	17
Рівномірний	10	18
Лапласа	9	17
Арксинусний	8	16

Була визначена достовірність віднесення вибірок до певного класу за допомогою генерації вибірок з різними розподілами. Дослідження достовірності за однією вибіркою  $n = 9$  показало, що ймовірність зміни лінгвістичного коду досить велика. Так, якщо об'єм вибірки менший 9, то із-за великої дисперсії порядкових статистик

отримуємо низьку достовірність класифікації. При збільшенні об'єму вибірки достовірність класифікації значно підвищується.

Алгоритм опрацювання декількох вибірок для класифікації розподілу набуває наступного вигляду:

- ранжування вибірок;
- обчислення нечітких оцінок за кожною вибіркою;
- ранжування НО з обчисленням номеру НО;
- обчислення середніх номерів НО за всіма вибірками;
- обчислення медіани середніх номерів;
- формування ЛК за всіма вибірками;
- визначення класу розподілу за ЛК.

Якщо згідно з рекомендаціями табл.1.11 обмежитись 10-ма вибірками для нормального розподілу, то згідно з наведеним вище алгоритмом за даними таблиці 1.10 отримуємо наступні результати (табл.1.12).

Таблиця 1.12 – Номери нечітких оцінок і результат класифікації за всіма вибірками

Імена НО № вибірки	168	267	249	348	1478	2369
1	1	2	5	6	3	4
2	2	1	6	5	4	3
3	1	2	6	5	4	3
4	1	2	6	5	3	4
5	2	1	6	5	3	4
6	1	3	6	5	2	4
7	1	2	6	4	3	5
8	2	1	6	5	4	3
9	1	2	5	6	3	4



Продовження таблиці 1.12						
Імена НО № вибірки	168	267	249	348	1478	2369
10	1	2	6	5	3	4
Середній номер вибірки	1.3	1.8	5.8	5.1	3.2	3.8
Медіана середніх номерів	3.5					
ЛК за 10-ма вибітками	1	1	0	0	1	0

Була визначена достовірність віднесення вибірок до певного класу за допомогою генерації вибірок з різними розподілами. Для підвищення ймовірності правильної класифікації рекомендується усереднення декількох вибірок. Кількість вибірок може бути визначена за заданою вірогідністю.

### **Висновки за розділом 1**

В першому розділі розглянуто визначення процедури класифікації та її основні ознаки, що наведені в роботах вчених, які працювали у цій галузі. Визначено, що однією з головних ознак є наявність шкали класифікації, яка або розробляється заздалегідь, або в процесі класифікації.

При опрацюванні вимірювальної інформації класифікація використовується як попередня операція, результати якої використовуються у подальшому для удосконалення процедури опрацювання вимірювальної інформації. Найчастіше це інформація про клас розподілу вибірок числових даних.

За розділом був проведений огляд існуючих способів класифікації та ідентифікації розподілів вибірок малого об'єму. В результаті був обраний метод нечіткої ідентифікації, запропонований Клікушиним. Даний метод має як ряд переваг,

так і ряд недоліків. Тому було проведене подальше дослідження особливостей і удосконалення нечіткої ідентифікації у вибірках малого об'єму, розширення області застосування нечіткої ідентифікації (для будь-якої кількості елементів вибірки, більше 9) і дослідження вірогідності розрізнення розподілів.

В процесі роботи було показано, що в дійсності нечітка ідентифікація дозволяє розрізняти розподіли з різних класів еквівалентності. Таким чином у подальшому введено поняття непараметричної класифікації, сформовані класи еквівалентності за матрицею «ознаки – класи».

За розробленим методом непараметричної класифікації видів розподілів виділено наступні класи еквівалентності: одномодальний симетричний, антимодальний, безмодальний симетричний, правосторонній одномодальний, лівосторонній одномодальний.

В процесі дослідження достовірності непараметричної класифікації виду розподілу за однією вибіркою була визначена ймовірність правильної класифікації за ймовірністю зміни лінгвістичного коду. Також було проведене експериментальне дослідження достовірності, яке підтвердило результати теоретичного.

Для підвищення ймовірності правильної класифікації рекомендується усереднення декількох вибірок. Мінімальна кількість вибірок може бути визначена за заданою вірогідністю. В роботі був проведений розрахунок необхідної кількості вибірок для різних розподілів за заданою вірогідністю класифікації.

## РОЗДІЛ 2. ОПРАЦЮВАННЯ ВИМІРЮВАЛЬНОЇ ІНФОРМАЦІЇ ЗА ПОПЕРЕДНЬОЮ КЛАСИФІКАЦІЄЮ РОЗПОДІЛІВ ВИБІРКИ.

### 2.1. Ефективні оцінки центру для різних законів розподілу.

На практиці часто виникають задачі опрацювання результатів вимірювань при малих об'ємах вибірок. При цьому невизначеність результату вимірювання доволі велика. В зв'язку з цим постає задача підвищення точності результату вимірювання. Тому пошук нових методів, що дозволяють підвищити точність при опрацюванні результатів вимірювань є актуальним.

Як показали дослідження, проведені в роботах [9], [10], знання виду розподілу має важливе значення при опрацюванні результатів вимірювань.

Знання виду розподілу дозволяє:

- в значній мірі підвищити точність оцінок, що є результатами вимірювань за рахунок вибору найбільш ефективних [4].
- значно точніше визначати коефіцієнт охоплення для розширеної невизначеності [9], [10] замість вибору оцінки зверху. Таким чином знання розподілу дозволяє зменшити і стандартну, і розширену невизначеність.

Так в роботах [9, 11] наведено ефективні оцінки (медіана, середина розмаху), знайдені статистичним опрацюванням вибірок, розподілених за іншими, ніж нормальний, розподілами. Якщо розподіл генеральної сукупності, з якої отримують практичну вибірку скінченного обсягу, є інший ніж нормальний, то середнє значення не обов'язково буде найкращою оцінкою результату вимірювання.

Зокрема, у разі розподілів з обмеженими граничними значеннями, наприклад, рівномірного, особливо арксинусного розподілів генеральної сукупності за обсягів вибірки значно ефективнішою оцінкою результату (з меншою дисперсією за заданого обсягу вибірки) є середина розмаху вибірки.

У таблиці 2.1 наведені залежності від обсягу вибірки відносної ефективності  $E_n$  середини розмаху відносно середнього значення як відношення стандартної

невизначеності  $u_A(\bar{x})$  середнього значення вибірки до стандартної невизначеності  $u_A(x_c)$  середини розмаху [10]:

$$E_n = \frac{u_A(\bar{x})}{u_A(x_c)}$$

Таблиця 2.1 – Відносна ефективність середини розмаху та медіани з вибірок деяких розподілів

Розподіл/оцінка	$E_n$	$n = 9$	$n = 25$
Рівномірний / середина розмаху	$\sqrt{\frac{(n-1) \cdot (n-2)}{6n}}$	1,02	1,92
Арксинусний / середина розмаху	$\frac{n}{\pi^2} \sqrt{\frac{n}{5}}$	1,22	5,66
Лапласа / медіана	$\sqrt{2}$	1,41	1,41

Під час оцінювання якості вимірювання важливою є не тільки стандартна невизначеність, але також і розширена. Оскільки форма розподілу Лапласа і особливо рівномірного та арксинусного розподілів істотно відрізняється від нормального, то і зв'язок між розширеною і стандартною невизначеністю середини розмаху вибірок є іншим ніж у разі нормального розподілу вибірки. Цей зв'язок, як відомо, кількісно відображається коефіцієнтом охоплення, а також ефективною кількістю ступенів свободи. Якщо розподіл не відомий, то потрібно брати оцінку зверху, тобто максимальне значення.

Тому при опрацюванні експериментальних даних попереднім етапом є класифікація розподілу вибірки, на основі якого вибирають ефективну оцінку центру вибірки.

## 2.2. Застосування попередньої класифікації розподілу вибірки для підвищення точності при опрацюванні результатів вимірювань.

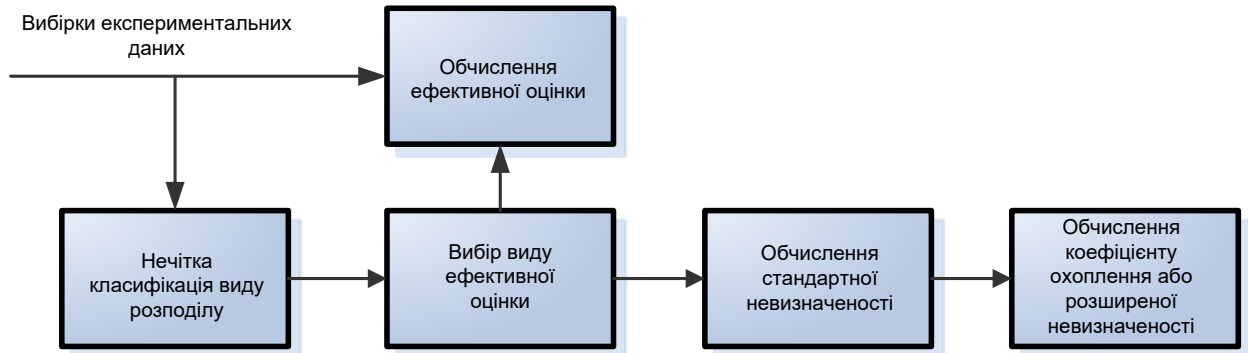


Рисунок 2.1 – Алгоритм процедури опрацювання результатів вимірювання з використанням нечіткої класифікації

На рис.2.1 наведено послідовність опрацювання результатів вимірювання з використанням попередньої класифікації. При апробації наведеного алгоритму вибірки було отримано від генератора випадкових чисел з рівномірним розподілом. Кількість вибірок 6, об'єм вибірки 9.

Таблиця 2.2 – Ранжовані результати вимірювання окремих вибірок

№1	9,58	9,6	9,62	9,81	9,87	10,13	10,16	10,35	10,49
№2	9,61	10,15	10,23	10,24	10,3	10,33	10,38	10,48	10,49
№3	9,59	9,62	9,65	9,94	10,11	10,13	10,19	10,3	10,41
№4	9,5	9,54	9,73	9,82	9,85	9,92	9,96	10,09	10,44
№5	9,68	9,69	9,83	9,95	10,19	10,29	10,3	10,44	10,48
№6	9,59	9,61	9,69	9,98	10,04	10,14	10,19	10,24	10,26

Таблиця 2.3 – Нечітка класифікація сумарного розподілу вибірок

Місце НО \ Ім'я НО	1	2	3	4	5	6	Середнє місце	ЛК
«168»	1	5	0	0	0	0	1,83	1
«267»	5	0	1	0	0	0	1,33	1

Продовження таблиці 2.3								
«249»	0	0	0	0	3	3	5,5	0
«348»	0	0	0	0	3	3	5,5	0
«1478»	0	1	1	4	0	0	3,5	0
«2369»	0	0	4	2	0	0	3,33	1

Висновок про вид розподілу – безмодальний симетричний (що відповідає рівномірному розподілу).

Ефективна оцінка центру – середина розмаху  $x_p$ .

$$x_p = \frac{1}{2}(x_n + x_1). \quad (2.1)$$

Якщо використати інші оцінки центру розподілу, вони будуть менш ефективні, що продемонстровано на прикладі з моделюванням.

Непараметрична оцінка СКВ середини розмаху (стандартна невизначеність):

$$S(x_p) = \frac{R}{\sqrt{2(n+1)(n+2)}} = \sigma(x_p) \quad (2.2)$$

Непараметрична оцінка розширеної невизначеності середини розмаху [11]:

$$U(x_p) = R \left[ (1-P)^{\frac{-1}{n-1}} - 1 \right] / 2 \quad (2.3)$$

Якщо попередньої класифікації не проведено і обрано нормальний закон розподілу, то оцінка центру – середнє арифметичне, а параметрична оцінка СКВ середнього арифметичного (стандартна невизначеність):

$$\sigma(\bar{x}) = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n(n-1)}}$$

Оцінка розширеної невизначеності:

$$U(\bar{x}) = t \cdot u(\bar{x}), \text{ де } t = 2,31 \text{ при } n = 9 \text{ и } P = 0,95.$$

Результати опрацювання наведених в табл.2.2 вибірок за умов попередньої класифікації і використання ефективної оцінки (2.1) та формул 2.2 і 2.3, а також

результати опрацювання при невідомому розподілі і заміні його нормальним, наведені в табл.2.4.

Таблиця 2.4 – Результати обчислення ефективної оцінки та стандартної і розширеної невизначеності

	1	2	3	4	5	6	7	8	9	$\bar{x}$	$\sigma(\bar{x})$	$U(\bar{x})$	$x_p$	$\sigma(x_p)$	$U(x_p)$
№1	9,58	9,6	9,62	9,81	9,87	10,13	10,16	10,35	10,49	9,96	0,113	0,261	10,035	0,061	0,205
№2	9,61	10,15	10,23	10,24	10,3	10,33	10,38	10,48	10,49	10,25	0,088	0,203	10,05	0,059	0,198
№3	9,59	9,62	9,65	9,94	10,11	10,13	10,19	10,3	10,41	9,99	0,103	0,237	10	0,055	0,185
№4	9,5	9,54	9,73	9,82	9,85	9,92	9,96	10,09	10,44	9,87	0,095	0,219	9,97	0,063	0,212
№5	9,68	9,69	9,83	9,95	10,19	10,29	10,3	10,44	10,48	10,09	0,104	0,241	10,08	0,054	0,180
№6	9,59	9,61	9,69	9,98	10,04	10,14	10,19	10,24	10,26	9,97	0,091	0,209	9,925	0,045	0,151

Проведені розрахунки показали, що при збільшенні об'єму вибірки стандартна і розширена невизначеності при виборі середини розмаху як ефективної оцінки для рівномірного закону розподілу, значно зменшились.

Для вибірок більшого об'єму, наприклад, 25, можна провести класифікацію виду розподілу за однією вибіркою.

Таблиця 2.5 – Ранжовані результати вимірювання

Номер елемента	1	2	3	4	5	6	7	8	9	10	11	12
Значення	9,566	9,6	9,632	9,816	9,828	9,873	10,035	10,074	10,154	10,175	10,351	10,352
13	14	15	16	17	18	19	20	21	22	23	24	25
10,368	10,43	10,464	10,499	10,521	10,531	10,539	10,548	10,59	10,592	10,683	10,705	10,709

Таблиця 2.6 – Результати обчислення ефективної оцінки та стандартної і розширеної невизначеності

$\bar{x}$	$\sigma(\bar{x})$	$U(\bar{x})$	$x_p$	$\sigma(x_p)$	$U(x_p)$
10,07	0,11	0,26	10,14	0,06	0,14

При збільшенні об'єму вибірки стандартна і розширена невизначеності за умов вибору ефективної оцінки для конкретного розподілу значно зменшуються.

Також є можливість використання при калібруванні при наявності декількох вимірювань в одній точці калібрувальної характеристики і наявності декількох точок.

Класифікація дозволяє удосконалити процедуру опрацювання результатів вимірювань.

Класифікація емпіричних розподілів може бути першим етапом при використанні узагальнених розподілів (наприклад,  $gh$ -розподілу Тьюки [12]), в якому при  $g = 0$  мають сім'ю симетричних розподілів) для оцінки розширеної невизначеності, тому що такі узагальнені розподіли можуть застосовуватись для спрощення об'єднання складових невизначеності вимірювання.

### **2.3 Застосування непараметричної класифікації при побудові контрольних карт.**

Контрольні карти (КК) використовуються для контролю якості процесу, що є сукупністю взаємопов'язаних і взаємодіючих видів діяльності [12]. Загальний метод отримання КК включає [13, 14]:

1. Отримання вибірки;
2. Розрахунок необхідної статистики, такої як вибіркоче середнє, середина розмаху, медіана, розмах чи накоплена сума;
3. Графічну побудову цієї статистики на карті у вигляді функції від вибіркової послідовності чи часу.

На карті необхідно певним способом відобразити правила прийняття рішень про те чи знаходиться дана змінна «під контролем», тобто на КК вказуються верхня і нижня контрольні границі. До тих пір, поки статистика, що відкладається на КК, потрапляє між цими двома межами, процес вважається під контролем.

Результатом процесу є продукт. Продукт теж може бути проконтрольованим. Для проведення контролю визначають характеристики процесу і продукту, які можуть бути кількісними і якісними. При контролі визначають верхню і нижню межі поля допуску, які визначають верхню і нижню межу допустимих значень характеристики, при яких характеристика якості все ще відповідає вимогам.



Правила прийняття рішень, що використовуються для фіксування цих ліній, засновані на виді розподілу (зазвичай без інформації його приймають нормальним).

### **2.3.1. Застосування непараметричної класифікації при визначенні стабільності технологічних процесів.**

Основною метою виробників певної продукції є постійне підвищення продуктивності праці і якості продукції при мінімальних витратах на виробництво. Вирішення даної задачі можливе тільки при наявності відповідної інформації про хід процесу виробництва на всіх рівнях, тобто поточного контролю практично всіх етапів виробництва. В свою чергу це обумовлює необхідність використання відповідних систем контролю і забезпечення якості (як окремих деталей і виробів, так і технологічних процесів).

Найбільш розповсюджений спосіб вирішення даної задачі обумовлює застосування управління процесами, що, в свою чергу, вимагає використання спеціальних контрольних карт параметрів, які характеризують якість процесу. За цими картами фіксується як випадкове розсіювання цих параметрів, так і можливі систематичні відхилення в усіх характерних точках процесу виробництва.

Стабільний процес, процес у стані статистичної керованості – процес з постійним середнім, мінливість якого викликана тільки випадковими причинами [12].

Вибірки з стабільного процесу в будь-який момент часу являють собою прості випадкові вибірки з однієї і тієї ж сукупності.

Стабільність процесу не означає, що випадкова мінливість більша, менша або відповідає вимогам, а лише означає, що мінливість може бути спрогнозована з допомогою статистичних методів.

Кarti статистичного контролю якості з окремими значеннями параметрів легко отримати і використати. Але управління виробництвом за допомогою таких карт не є ефективним. Більш інформативними є карти з вибічковими статистичними характеристиками, що дають можливість більш ефективного і правильного управління технологічним процесом. В цьому випадку на карті статистичного контролю

позначають час або номер вибірки, а на вісі ординат контрольований параметр. В залежності від типу карти на вісі ординат вказують:

- окремі значення параметру  $x_i$ ;
- центральну тенденцію або за середнім арифметичним  $\bar{x}$ , або за медіаною  $medx$ , або за серединою розмаху  $R_c$ ;
- середнє квадратичне відхилення  $S$  або розмах  $R$ ;
- контрольні границі;
- попереджувальні границі.

Контрольні границі відповідають границям  $\bar{x} \pm 3S$  для нормального розподілу вибірки і ймовірності 99,7%. Попереджувальні границі відповідають інтервалу  $\bar{x} \pm 2S$  для нормального розподілу вибірки і можуть бути тільки умовно перевищені.

Існуючі методи опрацювання вибірок контрольних карт орієнтовані на нормальний розподіл, хоча в багатьох джерелах відмічають, що розподіл може відрізнитися від нормального. Результати вимірювань представляють собою ряд вибірок невеликого об'єму. При цьому допускається, що є сталою лише форма розподілу без допущень щодо рівності центральних тенденцій і дисперсій. В такому випадку для визначення (а в основному для перевірки нормальності) використовуються непараметричні критерії.

Так в [8] наведено непараметричний критерій перевірки гіпотези нормальності за сукупністю малих вибірок, де перевіряється нормальність перших або останніх порядкових статистик сукупності ранжованих вибірок.

Метою роботи є застосування непараметричної класифікації для побудови контрольних карт технологічного процесу, що дозволяє визначити клас розподілу або сам розподіл результатів вимірювань окремих вибірок малого об'єму без дотримання вимог щодо рівності центрів розподілів окремих вибірок.

### **2.3.2. Моделювання процедури застосування непараметричної класифікації при побудові контрольних карт.**

На КК певним способом відображаються правила прийняття рішень про те, чи знаходиться характеристика процесу «під контролем». Тому на КК вказують верхню і нижню контрольну межу. До тих пір, поки статистика, що відкладається між цими границями знаходиться в межах, процес вважається «під контролем». Правила прийняття рішення, що використовуються для фіксації цих ліній, можуть бути засновані на припущенні про вид розподілу. Звичайно припускають, що розподіл нормальний, або використовуються непараметричні оцінки, такі як медіана і розмах (абсолютне значення різниці між найбільшим і найменшим значеннями у вибірці). Але в літературних джерелах з КК ([13, 14]) зазначається, що будь-яка інформація про вид або клас розподілу вибірки корисна при побудові КК. Клас розподілу за [14] – група розподілів, що мають спільні параметри, що повністю визначають дану групу розподілів.

Модель розподілу характеристики продукції – це конкретний розподіл або вид розподілу, наприклад – унімодальний симетричний розподіл.

Тому класифікацію виду розподілу можна вважати як попередню операцію побудови КК, що дозволяє визначити клас розподілу і вибрати оцінку, яка буде використовуватись як центральна тенденція при побудові контрольних карт технологічного процесу [14, 15], а саме середнє арифметичне, медіану, середину розмаху, тощо, а потім у відповідності з обраною центральною тенденцією провести розрахунок розсіювання і побудувати обмежувальні границі для центральної тенденції технологічного процесу.

У відповідності з цим алгоритм побудови контрольних карт наступний, враховуючи, що вихідною інформацією є декілька вибірок з малою кількістю вимірювань в кожній, причому центри розподілів вибірок можуть відрізнитись внаслідок нестабільності технологічного процесу.

Вибірки ранжуються, формуються НО, НО ранжуються, визначається середній номер нечіткої оцінки за декількома вибірками, за розробленим алгоритмом непараметричної класифікації формується код, що порівнюється з кодами класів розподілів для визначення відповідного класу розподілу або самого розподілу.

З метою визначення алгоритму опрацювання даних і порівнянням з традиційним з припущенням нормального розподілу проведено моделювання процесу побудови КК. З використанням генератора випадкових чисел отримано 6 вибірок по 9 результатів вимірювання в кожній (табл.2.7) розмаху вибірок.

Таблиця 2.7 – Ранжовані результати вимірювання окремих вибірок, отриманих при контролі технологічного процесу

№1	9,58	9,6	9,62	9,81	9,87	10,13	10,16	10,35	10,49
№2	9,61	10,15	10,23	10,24	10,3	10,33	10,38	10,48	10,49
№3	9,59	9,62	9,65	9,94	10,11	10,13	10,19	10,3	10,41
№4	9,5	9,54	9,73	9,82	9,85	9,92	9,96	10,09	10,44
№5	9,68	9,69	9,83	9,95	10,19	10,29	10,3	10,44	10,48
№6	9,59	9,61	9,69	9,98	10,04	10,14	10,19	10,24	10,26

Проведено нечітку класифікацію сумарного розподілу вибірок (табл.2.8)

Таблиця 2.8 – Нечітка класифікація сумарного розподілу вибірок

Місце Ім'я НО	1	2	3	4	5	6	Середнє місце	ЛК
«168»	1	5	0	0	0	0	1,83	1
«267»	5	0	1	0	0	0	1,33	1
«249»	0	0	0	0	3	3	5,5	0
«348»	0	0	0	0	3	3	5,5	0
«1478»	0	1	1	4	0	0	3,5	0
«2369»	0	0	4	2	0	0	3,33	1

За лінгвістичним кодом визначено клас розподілу як безмодальний, наближений до рівномірного. Тоді за центральну тенденцію приймають середину розмаху, а СКВ визначають за формулою [10]:

$$S(x_p) = \frac{R}{\sqrt{2(n+1)(n+2)}},$$

де  $R = x'_n - x'_1$  (вибірковий розмах)

Дані для побудови контрольної карти з урахуванням ефективної оцінки центру розподілу наведено в табл. 2.9.

Таблиця 2.9 – Дані для побудови контрольної карти

Номер вибірки	Середина розмаху $x_p$	Відхилення окремих значень від середини розмаху								
		№1	10,04	-0,46	-0,44	-0,42	-0,22	-0,17	0,09	0,12
№2	10,05	-0,44	0,1	0,18	0,19	0,25	0,28	0,33	0,43	0,44
№3	10,0	-0,41	-0,38	0,35	-0,06	0,11	0,13	0,19	0,30	- 0,41
№4	9,97	-0,47	-0,43	-0,24	-0,15	-0,12	-0,05	-0,01	0,12	0,47
№5	10,08	-0,40	-0,39	-0,25	-0,13	0,11	0,21	0,22	0,36	0,40
№6	9,93	-0,34	-0,32	-0,24	0,05	0,11	0,21	0,26	0,31	0,33

Якщо без проведення класифікації припускаємо, що розподіл нормальний, то результати опрацювання даних табл.2.7 мають наступний вигляд (2.10)

Таблиця 2.10 – Дані для побудови контрольної карти

Номер вибірки	Середнє арифметичне $\bar{x}$	Відхилення від заданого значення технологічного процесу								
		№1	9,96	-0,38	-0,36	-0,34	-0,14	-0,09	0,17	0,20
№2	10,25	-0,64	-0,1	-0,02	-0,01	0,05	0,08	0,13	0,23	0,24
№3	9,99	-0,40	-0,37	-0,34	-0,05	0,12	0,14	0,20	0,31	0,42
№4	9,87	-0,37	-0,33	-0,14	-0,05	-0,02	0,05	0,09	0,22	0,57
№5	10,09	-0,41	-0,40	-0,26	-0,14	0,1	0,2	0,21	0,35	0,39
№6	9,97	-0,38	-0,36	-0,28	0,01	0,07	0,17	0,22	0,27	-0,29

Для побудови КК з використанням середнього значення є багато способів і допоміжних таблиць. Якщо скористатись класичним підходом, тоді центром сумарного розподілу буде  $\bar{x} = \frac{1}{L} \sum_{i=1}^L \bar{x}_i$ , що для даних табл.2.10 становить  $\bar{x} = 10.02$ .

При об'єднанні  $L$  однорідних груп, що вміщують по  $n$  результатів вимірювання, СКВ середнього  $\bar{x}$  обчислюють за формулою [10]:

$$\sigma(\bar{x}) = \sqrt{\frac{n-1}{N(N-1)} \sum_{i=1}^L S_i^2 + \frac{1}{L(N-1)} \sum_{i=1}^L (\bar{x}_i - \bar{x})^2}, \quad (2.5)$$

де  $S_i$  - вибіркоче СКВ.

Так як  $S_i^2 = n\sigma^2[\bar{x}_i]$ , то формула (2.5) приймає вигляд:

$$\sigma(\bar{x}) = \sqrt{\frac{n-1}{N(N-1)} \sum_{i=1}^L \sigma^2[\bar{x}_i] + \frac{1}{L(N-1)} \sum_{i=1}^L (\bar{x}_i - \bar{x})^2},$$

Результати обчислень даних таблиці 2.10 наведено в табл.2.11.

Таблиця 2.11 – Границі знаходження процесу «під контролем»

Середнє арифметичне $\bar{x}$	$\sigma[\bar{x}_i]$	Попереджувальні границі	Контрольні границі
9,96	0,113		
10,25	0,088		
9,99	0,103		
9,87	0,095		
10,09	0,104		
9,97	0,09		
$\bar{x} = 10.02$	$\sigma[\bar{x}_i] = 0.0392$	$\bar{x} \pm 0.0784$	$\bar{x} \pm 0.1176$

Контрольна карта, побудована за даними таблиці 2.11, наведена на рис.2.2

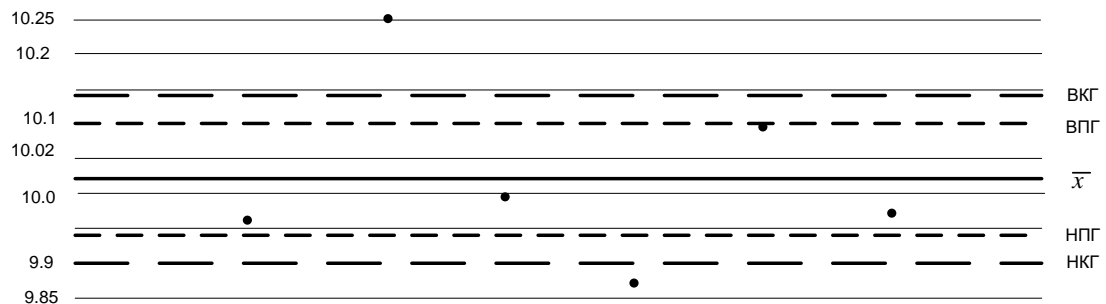


Рисунок 2.2 – Контрольна карта технологічного процесу при прийнятті нормального розподілу

При роботі з попередньою класифікацією розподілу для побудови КК скористаємось характеристикою центру вибірки – серединою розмаху і даними таблиці 2.9.

Таблиця 2.12 – Границі знаходження процесу «під контролем»

Середина розмаху $x_p$	Розмах $R$	$\sigma[\bar{x}_{pi}]$	Контрольні границі	Попереджувальні границі
10,04	0,91	0,0614	-	-
10,05	0,88	0,0593	-	-
10,0	0,82	0,0553	-	-
9,97	0,94	0,0634	-	-
10,08	0,8	0,0539	-	-
9,93	0,67	0,0452	-	-
$\bar{x}_p = 10,01$	$\bar{R} = 0,837$	$\sigma[\bar{x}_p] = 0,0231$	$\bar{x}_p \pm 2\sigma[\bar{x}_p] = 10,01 \pm 0,0463$	$\bar{x}_p \pm 3\sigma[\bar{x}_p] = 10,01 \pm 0,0693$

Загальне СКВ отримуємо за формулою 2.6:

$$\sigma(x_p) = \sqrt{\frac{n-1}{N(N-1)} \sum_{i=1}^L \sigma^2[x_p] + \frac{1}{L(N-1)} \sum_{i=1}^L (x_{pi} - \bar{x}_p)^2}, \quad (2.6)$$

Контрольна карта, побудована за даними табл.2.12, наведена на рис.2.3

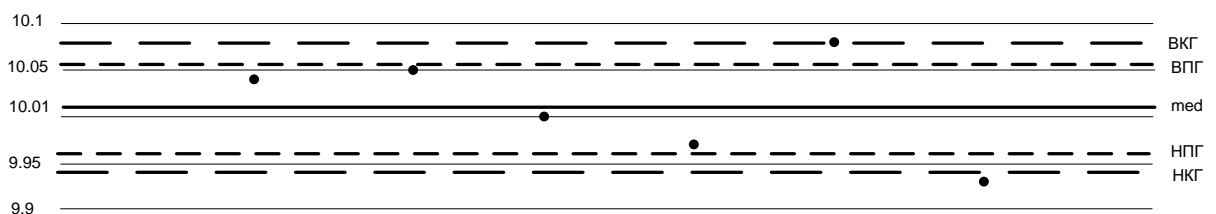


Рисунок 2.3 – Контрольна карта технологічного процесу при проведенні попередньої класифікації

З проведеного дослідження і моделювання отримані наступні висновки. Попередня класифікація сумарного розподілу дозволяє обрати вид контрольної карти за центральною тенденцією, а саме:

- КК середніх арифметичних (average control chart), за якою кількісною ознакою мінливості процесу є середні арифметичні в підгрупах;
- КК медіан (median control chart), за якою кількісною ознакою мінливості процесу є значення медіан в підгрупах;
- КК розмахів (range control chart), за якою кількісною ознакою мінливості процесу є значення розмаху в підгрупах.

Завдяки вибору КК за ефективною для даного класу розподілу оцінкою центру вибірки зменшується дисперсія і СКВ за даною вибіркою. Тому попереджувальні і контрольні межі стають вузькими, що забезпечує більшу чутливість саме до змін параметрів технологічного процесу. Якщо ж зміни відбуваються в контрольних межах, то неможливо вирішити чим вони обумовлені: нестабільністю технологічного процесу чи обмеженістю статистичних даних. Рішення про клас розподілу декількох вибірок завдяки непараметричному оцінюванню приймається при певному порядку слідування нечітких оцінок (що є лінійними комбінаціями порядкових статистик) і не залежить від зміни центральної тенденції. При зміні центральної тенденції всі порядкові статистики зміщуються в один бік і їх порядок слідування не змінюється. Завдяки цьому непараметричну класифікацію можна застосувати при нестабільності технологічного процесу. Цього не можна сказати при використанні нормального розподілу для окремих вибірок, тому що поєднання вибірок при визначенні розподілу можливе лише при статистичній рівності середніх значень. Якщо повернутись до таблиці 2.11, то вибірки з номерами 2 і 4 не відповідають вимозі про рівність середніх [16]:

$$|\bar{x}_1 - \bar{x}_2| < t_p \sqrt{\sigma^2[\bar{x}_1] + \sigma^2[\bar{x}_2]}, \text{ а саме}$$

$$|10,25 - 9,87| = 0,38, \quad t_p = 2,31, \quad (\text{для } P = 0,95, \quad n = 9)$$

$$\sqrt{\sigma^2[\bar{x}_2] + \sigma^2[\bar{x}_p]} = \sqrt{0,095 + 0,088} = 0,129, \quad \text{тобто } |\bar{x}_2 - \bar{x}_1| > 0,298.$$

Можливість використання класифікації була перевірена при значній нестабільності технологічного процесу. Для моделювання були використанні дані табл.3.7 з додаванням лінійного дрейфу. Аналогічне моделювання нестабільності



технологічного процесу було проведено Робертсом [13, 17]. Він порівнював декілька типів контрольних карт, використовуючи для кожної карти один чи ряд критеріїв. Спостереження імітувались за допомогою таблиці випадкових чисел, розподілених за нормальним законом з математичним очікуванням, що дорівнює нулю і дисперсією, що дорівнює 1. Після того, як було отримано 100 чисел, до всіх чисел, починаючи з наступного, додавалась 1, щоб відтворити зсув середнього рівня процесу.

Вихідні дані з врахуванням нестабільності технологічного процесу наведено в табл.2.13

Таблиця 2.13 – Ранжовані результати вимірювання окремих вибірок, отриманих при контролі технологічного процесу

№1	9,566	9,600	9,632	9,828	9,873	10,154	10,175	10,351	10,499
№2	9,816	10,368	10,430	10,464	10,521	10,539	10,592	10,683	10,705
№3	10,035	10,074	10,352	10,531	10,548	10,590	10,709	10,816	10,993
№4	10,112	10,149	10,336	10,444	10,465	10,523	10,581	10,708	11,040
№5	10,493	10,504	10,645	10,756	10,990	11,111	11,118	11,249	11,292
№6	10,614	10,616	10,711	10,989	11,043	11,158	11,202	11,240	11,275

В таблиці 2.14 наведено результати визначення класу сумарного розподілу з використанням даних табл.2.10.

Таблиця 2.14 – Непараметрична класифікація виду розподілу

НО	1	2	3	4	5	6	Середнє місце	ЛК
«168»	1	5	0	0	0	0	1,83	1
«267»	5	0	1	0	0	0	1,33	1
«249»	0	0	0	0	3	3	5,5	0
348»	0	0	0	0	3	3	5,5	0
«1478»	0	1	1	4	0	0	3,5	0
«2369»	0	0	4	2	0	0	3,33	1

З таблиці 2.14 видно, що порядок слідування нечітких оцінок не змінився при зміні центральної тенденції між вибірками (якщо за час отримання вибірки суттєвих зміщень центру не було). Це ще раз свідчить про доцільність застосування непараметричної класифікації розподілів вибірки при побудові КК.

### **Висновки за розділом 2**

Розглянуто впровадження непараметричної класифікації виду розподілу вибірок числових даних як попередню операцію при опрацюванні вимірювальної інформації.

Показано, що непараметрична класифікація виду розподілу дозволяє обрати ефективну оцінку центру розподілу і тим самим підвищити точність результату багаторазового вимірювання.

При застосуванні попередньої непараметричної класифікації забезпечується вибір виду контрольної карти за центральною тенденцією, що забезпечує більшу чутливість саме до змін параметрів технологічного процесу.

Показано, що при зміні центральної тенденції всі порядкові статистики зміщуються в один бік і їх порядок слідування не змінюється. Завдяки цьому непараметричну класифікацію можна застосувати для побудови контрольних карт при нестабільності технологічного процесу.

## **РОЗДІЛ 3. ОСОБЛИВОСТІ ОПРАЦЮВАННЯ ВИМІРЮВАЛЬНОЇ ТА ЕКСПЕРТНОЇ ІНФОРМАЦІЇ ПРИ ПЕРЕХОДІ ВІД ЧИСЛОВИХ ДАНИХ ДО ВЕРБАЛЬНИХ ЗА МЕТРИЧНОЮ КЛАСИФІКАЦІЄЮ.**

### **3.1. Загальна характеристика метричної класифікації.**

Сучасні інтелектуальні вимірювальні системи використовуються разом з системами прийняття рішень при виконанні різних виробничих задач. В такому випадку за традиційних операцій, що входять до складу процедури вимірювання, додається операція класифікації, функціонування якої забезпечує шкала класифікації. Тому вважають, що особливістю інтелектуальних вимірювальних систем є поліморфізм, який полягає в відображенні вимірювальної величини декількома шкалами.

Найчастіше використовується шкала метричної класифікації з впорядкованими класами еквівалентності, що є шкалою квазіпорядку. В загальному випадку, це шкала з нечіткою лінгвістичною змінною (ЛЗ), яка забезпечує можливість отримання даних, що використовуються в обчисленнях на основі нечіткої логіки [18].

При встановленні шкали з нечіткою ЛЗ використовується семантичне правило, яке ставить у відповідність кожному елементу терм-множини  $T = \{T_1, T_2, \dots, T_n\}$  нечітку підмножину з області визначення вимірюваної величини [19, 20]. Семантичне правило за експертною інформацією може бути операцією, що встановлює функцію приналежності (ФП) або включати оператори, що змінюють індекс нечіткості встановленої ФП. Так як результати вимірювання супроводжуються невизначеністю, є необхідність в аналізі способів урахування невизначеності вимірювання при встановленні ФП терм-множини лінгвістичної змінної [21]. В роботах зі встановлення шкал з нечіткою ЛЗ враховується тільки нечіткість семантичного правила або експертної інформації. Тому однією з задач представленої роботи є аналіз і розробка способів побудови шкали класифікації, тобто визначення ФП окремих класів з урахуванням невизначеності результатів вимірювання.

Для отриманих класифікованих даних використовується нечіткий класифікатор [22, 23]. У загальному випадку задача класифікації полягає в визначенні ступеня приналежності ознак об'єкта окремим класам або самого об'єкту за сукупністю ознак. У випадку наявності однієї ознаки використовують одновимірні ФП. Одновимірні ФП є такими, що найбільш часто використовуються в нечіткій логіці. В вимірювальних системах на основі результату вимірювання ознаки  $x_i$  класифікатор обчислює ступінь її приналежності терм-множинам  $T_1, T_2, \dots, T_n$  (рис. 3.1). Найбільше значення ступеня приналежності (що може приймати значення в інтервалі від 0 до 1) вказує, до якої терм-множини належить ознака об'єкта, тобто яким чином його потрібно класифікувати.

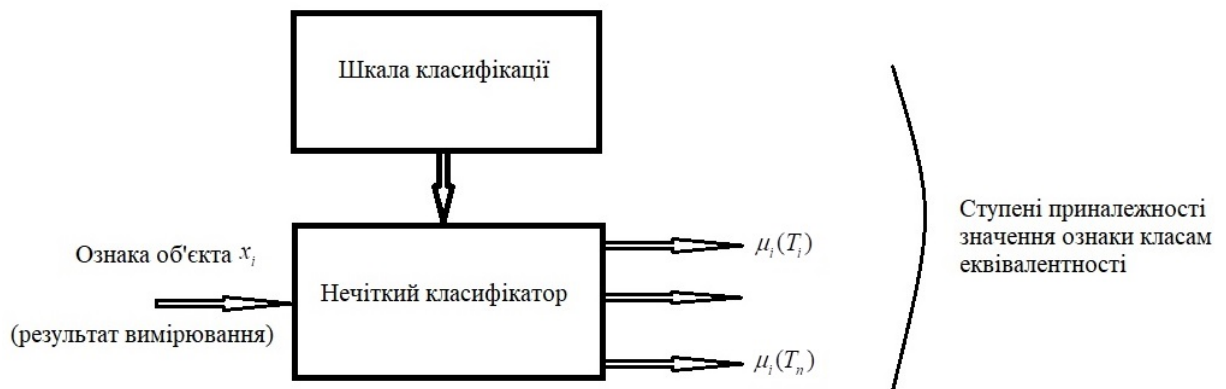


Рисунок 3.1 – Ілюстрація отримання класифікованих даних

Якщо результат класифікації є чітким, тобто обрано один клас еквівалентності або один з термів множини, то характеристикою якості є ймовірність правильної класифікації. Якщо результат класифікації подається як нечітка підмножина, тобто враховується розсіювання ступенів приналежності за класами еквівалентності, то подання результату класифікації набуває вигляду:

$$T_1 | \mu_i(T_1); T_2 | \mu_i(T_2); \dots T_n | \mu_i(T_n)$$

Отримані таким чином класифіковані дані можуть бути використані далі при нечітких обчисленнях і виведеннях з метою прийняття певних рішень. Але в роботі розглянуто способи опрацювання класифікованих даних для двох напрямків, які частіше всього зустрічаються в інтелектуальних вимірювальних системах, а саме:

- багаторазове вимірювання ознаки об'єкта з подальшою класифікацією поточного стану об'єкта;
- класифікація (визначення) узагальненої характеристики об'єкта за вектором результатів вимірювання окремих ознак.

### **3.2. Основні етапи встановлення шкал класифікації при вимірюваннях і діагностиці.**

Розглянемо метричну класифікацію, за якої на метричному носіївбудується шкала класифікації як шкала за нечіткою лінгвістичною змінною.

Якщо звернутись до визначення шкали, що наведено в [24], то можна визначити лінгвістичну шкалу з нечіткою ЛЗ як впорядкований ряд імен термів (нечітких підмножин), носії яких відповідають певним частинам діапазону вимірюваної величини.

Вимірювання і діагностування відносять до процедур експериментальної інформатики. Загальною рисою цих процедур є відображення певних властивостей об'єкта на основі експериментальних даних. При вимірюваннях має місце відображення розміру величини, при діагностуванні відображення загального стану об'єкта і причин цього стану діагнозом [24]. Зіставлення процедур вимірювання і діагностування свідчить про наявність загальних етапів при виконанні цих процедур. При діагностуванні стану об'єкта на основі результатів вимірювання діагностичних ознак теж може використовуватись шкала з нечіткою ЛЗ, певні рівні або градації якої відображають характеристики стану об'єкта діагностування. Таким чином, при побудові шкали з нечіткою ЛЗ при вимірюваннях і діагностуванні використовується операція фазифікації, що дозволяє приписати відповідні нечіткі множини градаціям шкали, що стають елементами терм-множини лінгвістичної змінної. При побудові шкали повинна враховуватись загальна невизначеність вимірювання.

На основі досліджень, проведених автором роботи, можна запропонувати наступну послідовність етапів встановлення або відтворення лінгвістичної шкали при вимірюваннях і діагностуванні:

- визначення кількості термів множини і їх границь;
- аналіз нечіткості семантичного правила і складових невизначеності вимірювання;
- врахування впливу невизначеності вимірювання на індекс нечіткості нечітких підмножин, що є елементами терм-множини ЛЗ;
- апробація (при наявності еталонних елементів) або моделювання нечіткого виведення з застосуванням встановленої лінгвістичної шкали.

При відтворенні шкали з ЛЗ спочатку визначається кількість термів і їх чіткі границі у відповідності з існуючими рекомендаціями, настановами, нормативними документами, тощо. Якщо такої інформації немає, використовуються інтуїтивні функції приналежності, які на підсвідомому рівні використовує людина [25]. Для того, щоб експертним шляхом визначити ширину терма, можна скористатись поняттям критичної точки функції приналежності терма, під якою розуміють точку зі ступенем приналежності, що дорівнює 0,5.

Лінгвістичні шкали складаються з якісних оцінок фізичної величини, стану об'єктів та систем. Формалізація якісних оцінок ускладнена по-перше лінгвістичною невизначеністю понять (наприклад, «малий», «невеликий»), а по-друге суб'єктивністю сприйняття цих понять різними експертами. При створенні математичного апарату, що може забезпечити адекватний опис і формалізацію невизначеності такого роду якраз і використовуються нечіткі ЛЗ, індекс нечіткості функцій приналежності яких залежить від нечіткості семантичного правила [26, 27, 28].

Якщо при встановленні границь термів використовуються певні узгоджені норми, то границі є чіткими, а подальше врахування нечіткості пов'язане зі складовими невизначеності, що можуть впливати на зміну границь. І першою складовою такої невизначеності є невизначеність від неповної ідентифікації об'єкта при встановленні шкали. Тобто, є можливість уточнення шкали за повною ідентифікацією об'єкта, але цією можливістю користуються не завжди, тому що не хочуть збільшувати кількість правил бази знань.

Другий етап при встановленні шкали з нечіткою ЛЗ полягає в аналізі нечіткості семантичного правила і складових невизначеності вимірювання.

При діагностуванні теж виділяють множину  $S$  рівнів певної діагностичної ознаки і визначають відношення порядку між цими рівнями. Найменування термів, що відповідають цим рівням, теж знаходяться у відношенні порядку або у відношенні домінування. На основі результату вимірювання діагностичної ознаки визначається рівень шкали з нечіткою ЛЗ або елемент терм-множини, якому відповідає результат вимірювання. Фазифікований результат використовується у подальшому для визначення стану об'єкта за однією або декількома діагностичними ознаками.

Тобто при урахуванні деталей для повної ідентифікації об'єкта збільшується кількість правил. При неповній ідентифікації об'єкта збільшується індекс нечіткості терм-множини ЛЗ. Зменшити індекс нечіткості до певної межі можна за рахунок збільшення бази правил. Тому при встановленні шкали необхідно знайти компроміс між критеріями допустимої точності шкали і критеріями компактності бази правил. Невизначеність від неповної ідентифікації об'єкта об'єднується з іншими складовими невизначеності.

Якщо звернутись до [29], то невизначеність від неповної ідентифікації об'єкта відповідає дефініційній невизначеності вимірювання, що виникає внаслідок обмеженої кількості деталей у визначенні вимірювальної величини.

Після цього проводиться аналіз складових невизначеності, що супроводжують отримання результату вимірювання за метричною шкалою. Загальний перелік складових невизначеності, що впливають на індекс нечіткості терм-множини ЛЗ наведено нижче.



Рисунок 3.2 – Складові невизначеності

Спосіб об'єднання чинників залежить від алгоритму роботи нечіткого класифікатора, за яким результат вимірювання подають як нечітке число (fuzzy number), або чітке число (singleton). Останній вид алгоритму використовується в нечіткому додатку до пакету MatLab, за яким визначають саме активовані ФП класів еквівалентності. Ці ФП модифікують у відповідності з максимумом і мінімумом перерізу і за центром ваги визначають клас еквівалентності, якому відповідає результат вимірювання. Тому був обраний спосіб об'єднання чинників, що характеризують нечіткість експертної інформації і невизначеність вимірювання. Для об'єднання чинників необхідно було визначити критерій з дотриманням вимог обумовленості, практичності і реалістичності. Цим вимогам відповідає індекс нечіткості, що є мірою нечіткості множини і визначається як відстань цієї множини до найближчої до неї звичайної множини. Ця метрика може бути застосована і для дискретної і для неперервної множини, тобто при всіх способах подання нечіткості семантичного правила.

При визначенні типу функцій приналежності окремих термінів ЛЗ виходять з рекомендації з їх практичного застосування [30, 31].

Доволі часто використовуються функції приналежності, що складаються з прямолінійних ділянок, що обумовлені їх простотою. Суттєвою перевагою



багатокутових функцій приналежності є те, що для їх визначення необхідний найменший (у порівнянні з іншими) об'єм інформації, що обмежується даними з кутових точок. Найбільш розповсюдженою формою є трапецієподібна або трикутна. В роботі [21] проведено аналіз впливу невизначеності вимірювання на форму трапецієподібної і трикутної функції приналежності і показано, що нахил бічних сторін визначається за сумарною невизначеністю вимірювання.

Останнім етапом є апробація нечіткого виведення або перевірка отриманого діагнозу за встановленою шкалою. Апробація може бути проведена за наявності еталонних зразків, що відповідають певним градаціям лінгвістичної шкали. Якщо такої можливості немає можна скористатись моделюванням. При наявності недоліків встановлена шкала підлягає корегуванню і удосконаленню.

### **3.3. Визначення кількості класів еквівалентності шкали.**

Першим етапом встановлення шкали класифікації є визначення кількості класів еквівалентності або термів лінгвістичної змінної. Другим етапом є вибір форми ФП терм-множини лінгвістичної змінної.

Розрахунок кількості класів еквівалентності запропоновано проводити виходячи з невизначеності вимірювання. Для цього використовується інформаційний підхід, заснований на ентропійному інтервалі невизначеності [5, 32, 33, 34, 35]. За цим підходом кількість розрізняваних градацій шкали визначається за границями діапазону вимірювання  $x_1, x_2$  (нижньою і верхньою відповідно) та ентропійним інтервалом невизначеності  $\alpha$ , як 
$$N = \frac{x_2 - x_1}{\alpha}.$$

Ентропійний інтервал невизначеності знаходять з урахуванням розподілу похибки. Значення  $N$  є верхньою границею кількості класів еквівалентності. Тому при встановленні шкали класифікації їх кількість може бути меншою виходячи з розрізняльної здатності якісної оцінки різних класів еквівалентності. Після цього встановлюються назви або коди чи умовні позначення класів еквівалентності і їх границі, що на першому етапі є чіткими. Але наявність похибки вимірювання є

причиною появи зони невизначеності на границях між окремими класами еквівалентності.

За інформаційним підходом [5], кількість інформації, отримана в результаті вимірювань, дорівнює різниці вихідних ентропій, що залишилися, тобто

$$I = H(X) - H(X / X_n) = \ln(X_2 - X_1) = \ln(2\Delta) = \ln \frac{X_2 - X_1}{2\Delta} = \ln N.$$

Число  $N = (X_2 - X_1) / (2\Delta)$  показує, скільки інтервалів невизначеності завдовжки  $X$   $d = 2\Delta$  укладається у всьому діапазоні  $X_2 - X_1$ , тобто яке число помітних градацій вимірюваної величини дозволяє отримати даний прилад або метод вимірювання.

Співвідношення  $I = \ln Na$  і  $N = (X_2 - X_1) / d$  справедливі при будь-якому законі розподілу похибки, якщо лише інтервал невизначеності  $d$  буде знайдений через ентропію. Автори роботи [5] запропонували називати величину  $N$  числом різних градацій вимірюваної величини, а  $d$  – інтервалом ентропії невизначеності результату виміру.

Основні переваги інформаційного підходу щодо питання математичного опису випадкових похибок полягає в тому, що розмір ентропійного інтервалу невизначеності може бути визначений суто математично для будь-якого закону розподілу похибки, як величина, що стоїть під знаком логарифма у формулі для ентропії  $H(X / X_n)$ .

Для нормально розподіленої похибки, тобто при

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right);$$

$$\ln p(x) = -\ln(\sigma\sqrt{2\pi}) - x^2 / (2\sigma^2).$$

Звідси ентропія похибки

$$H(X / X_n) = - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx = \int_{-\infty}^{+\infty} p(x) \left( \ln \sigma\sqrt{2\pi} + \frac{x^2}{2\sigma^2} \right) dx = \ln \sigma\sqrt{2\pi} \int_{-\infty}^{+\infty} p(x) dx + \frac{1}{2\sigma^2} \int_{-\infty}^{+\infty} x^2 p(x) dx$$

Враховуючи, що  $\int_{-\infty}^{+\infty} p(x) dx = 1$  і за визначенням дисперсії  $\int_{-\infty}^{+\infty} x^2 p(x) dx = \sigma^2$ , отримуємо

$$H(X / X_n) = - \int_{-\infty}^{+\infty} p(x) \ln p(x) dx = \ln \sigma \sqrt{2\pi} + 1/2 = \ln \sigma \sqrt{2\pi} + \ln \sqrt{e} = \ln \sqrt{2\pi e},$$

Тобто інтервал невизначеності  $d$  результату виміювання, знайдений через ентропію відповідно до теорії інформації, однозначно (без яких-небудь припущень про вибір рівня довірчої вірогідності) дорівнює  $d = \sqrt{2\pi e} \sigma \approx 4,133\sigma$ , а число різних градацій результату виміру при рівномірному розподілі вірогідності різних значень вимірюваної величини

$$N = (X_2 - X_1) / d = (X_2 - X_1) / (4,133\sigma).$$

Розрахуємо число градацій для конкретного прикладу, зокрема для вимірювача рівня сміття. Вимірюванню підлягає відстань від верхньої точки урни до рівня насипу сміття. Максимальна довжина складає 2400 мм. Оскільки у нас мультиплікативна похибка, то розрахунок проводимо для середини діапазону.

Відповідно:

$$N = (X_2 - X_1) / d = (2400 - 10) / (4,133\sigma),$$

$$\sigma = \frac{\Delta \cdot X_{cp}}{\sqrt{3}},$$

$$N = 2400 / (4,133 \cdot 131,6) = 4,4 \approx 5.$$

За результатами розрахунків ми отримали 5 термів. Діапазон значень для нечітких змінних ЛПП для нашого прикладу наступний: «мінімальна довжина», «менше половини довжини», «половина довжини», «більше половини довжини», «максимальна довжина».

### **3.4. Встановлення форми функцій приналежності окремих градацій шкали класифікації.**

За думками спеціалістів, що працюють в області нечіткої логіки, найбільш уразливим для критики питанням в теорії нечітких множин є питання про методи побудови головної характеристики нечіткої множини – функції приналежності [19]. Основною складністю, що заважає інтенсивному застосуванню теорії нечітких множин при вирішенні практичних задач є те, що функція належності може бути задана без

теоретичного обґрунтування і не може бути таким чином перевірена на адекватність засобами цієї теорії. В кожному відомому методі побудови функції приналежності формулюються свої вимоги і обґрунтування до вибору саме такої побудови [22]. Якщо для побудови шкали лінгвістичної змінної визначена кількість термів множини [19, 22], то виникає питання про вибір форми функції приналежності за адекватністю індексу нечіткості і відносною розширеною невизначеністю вимірювання.

Другим етапом формування шкали класифікації є встановлення форми ФП окремих градацій шкали класифікації або визначених раніше класів еквівалентності. Дослідження проведено для шкали однієї ознаки, тобто ФП класів еквівалентності створюють шкалу нечіткої класифікації, а індекс нечіткості окремих класів як нечітких множин залежить від невизначеності вимірювання. Дослідження зв'язку індексу нечіткості класу еквівалентності і похибки вимірювання було проведено для трапецієподібної форми ФП, і для прикладу було обрано клас еквівалентності «малий», тобто початковий терм множини класів еквівалентності з аналітичним поданням ФП:

$$\mu_{T_1}(x) = \left\{ \begin{array}{l} 1, \text{ якщо } 0 \leq x \leq a_1 \\ \frac{a_2 - x}{a_2 - a_1}, \text{ якщо } a_1 \leq x \leq a_2 \\ 0, \text{ якщо } a_2 < x \end{array} \right\} \quad (3.1)$$

Чітка верхня границя класу еквівалентності була встановлена на рівні  $\frac{a_1 + a_2}{2}$ . Від наявності похибки вимірювання  $\Delta$ , що дорівнює  $\frac{a_2 - a_1}{2}$  форма ФП від прямокутної стає трапецієподібною (рис. 3.4).

Якщо використати розширену невизначеність в носії ФП як  $a_1 = a - \Delta = a(1 - \delta)$ ;  $a_2 = a + \Delta = a(1 + \delta)$ , де  $\delta$  – відносна розширена невизначеність, тоді 3.1 набуває вигляду

$$\mu_{T_1}(x) = \begin{cases} 1, & \text{якщо } 0 \leq x \leq a(1-\delta) \\ \frac{a(1+\delta)-x}{2a\delta}, & \text{якщо } a(1-\delta) \leq x \leq a(1+\delta) \\ 0, & \text{якщо } x > a(1+\delta) \end{cases} \quad (3.2)$$

Лінійний індекс нечіткості неперервної множини з функцією приналежності  $\mu_A(x)$  визначається за формулою [1]:

$$I_A^L = \frac{2}{b-a} \int_a^b |\mu_{\underline{A}}(x) - \mu_{\bar{A}}(x)| dx \quad (3.3)$$

де  $\mu_{\underline{A}}(x)$  – звичайна множина, найближча до нечіткої.

У відповідності з (3.1) і (3.2) індекс нечіткості терма множини «значення величини мале» з урахуванням відносної невизначеності  $\delta$  набуває вигляду:

$$I_A^L = \frac{2}{a(1+\delta)} \left[ \int_0^a |\mu_{T_1}(x) - 1| dx + \int_a^{a(1+\delta)} |\mu_{T_1}(x)| dx \right] = \frac{\delta}{1+\delta} \quad (3.4)$$

Якщо для оцінки ступеня нечіткості обрати індекс нечіткості, заснований на відносній евклідовій відстані [16] для неперервної множини  $U = [a, b]$ :

$$I_A^E = \frac{2}{\sqrt{b-a}} \sqrt{\int_a^b (\mu_{\bar{A}}(x) - \mu_{\underline{A}}(x))^2 dx}, \quad (3.5)$$

а  $\mu_{\bar{A}}(x)$  – ФП нечіткої множини  $\bar{A}$ ,  $\mu_{\underline{A}}(x)$  – ФП чіткої множини, найближчої до  $\bar{A}$ .

У відповідності до формули (3.4) з урахуванням (3.1) отримуємо індекс нечіткості для терма  $T_1 = \text{«малій»}$ :

$$\begin{aligned} I_{T_1}^E &= \frac{2}{\sqrt{a_2}} \sqrt{\int_0^{\frac{a_1+a_2}{2}} (\mu_{T_1}(x) - 1)^2 dx + \int_{\frac{a_1+a_2}{2}}^{a_2} (\mu_{T_1}(x))^2 dx} = \frac{2}{\sqrt{a_2}} \sqrt{\int_{a_1}^{a_2} (\mu_{T_1}(x))^2 dx + 2 \int_{a_1}^{\frac{a_1+a_2}{2}} \mu_{T_1}(x) dx + \int_{a_1}^{\frac{a_1+a_2}{2}} dx} = \\ &= \frac{2}{\sqrt{a_2}} \cdot \sqrt{\frac{a_2 - a_1}{12}}. \end{aligned}$$

Якщо ввести позначення  $\frac{a_1 + a_2}{2} = x_r$ ,  $\frac{a_2 - a_1}{2} = \Delta$ ,  $\frac{\Delta}{x_r} = \delta$ , де  $x_r$  – результат вимірювання, що знаходиться на верхній границі терма «малий»,  $\Delta$ ,  $\delta$  – абсолютна і відносна похибка значення  $x_r$ , то формула (3.5) для індексу нечіткості приймає вигляд:

$$I_{T_1}^E = \sqrt{\frac{2\Delta}{(x_r + \Delta) \cdot 3}} = \sqrt{\frac{2\delta}{(1 + \delta) \cdot 3}} \quad (3.6)$$

Таким чином отримано залежність між індексом нечіткості ФП терма «малий» і похибкою вимірювання граничного значення терма, форма якого встановлюється з урахуванням похибки вимірювання. Зі збільшенням похибки  $\Delta$  і зменшенням  $a_1$  індекс нечіткості збільшується і при  $a_1 = 0$  форма ФП стає трикутною (рис. 3.4),  $\delta = 100\%$ ,  $I_{T_1}^E = 0,57$ .

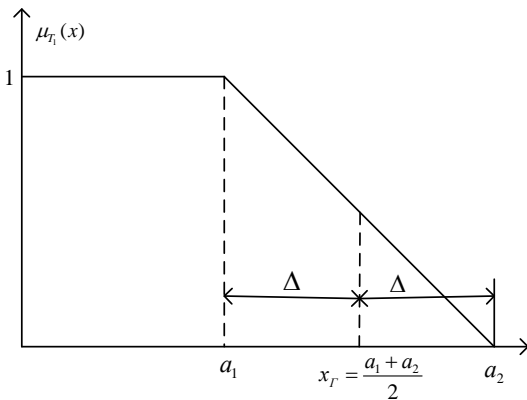


Рисунок 3.3 – ФП терма «малий»

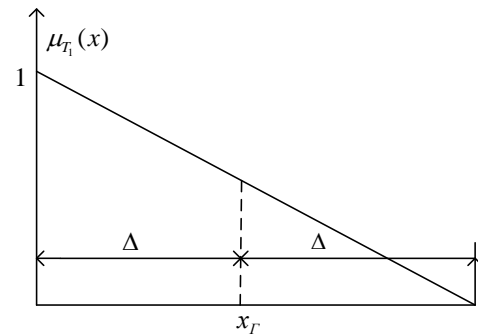


Рисунок 3.4 – ФП терма «малий» при  $\delta = 100\%$

Виходячи з (3.4) і (3.6) при об'єднанні чинників, що впливають на індекс нечіткості, треба визначити загальну область розмитості  $\Delta$ , як носія нечіткого числа, симетричного відносно границі терму або класу еквівалентності  $(a_1 + a_2)/2$ . З застосуванням теорії нечітких множин при об'єднанні результатів вимірювання, отриманих за допомогою «фізичних» систем і результатів оцінювання з «суб'єктивною» неточністю, носій визначено як вкладений інтервал (nested interval) за заданим рівнем довіри [36, 37]. Як приклад при встановленні шкали артеріального

тиску в ІВС медичного призначення терм-множина ЛЗ визначається за експертними даними. Розбіжність даних, отриманих з різних джерел для певного віку і статі особи, ідентифікується як нечіткість семантичного правила. Наприклад, для терму «нормальний» вона складала  $\pm 3 \text{ мм рт. ст.}$ . Для повної ідентифікації об'єкта треба було розробити шкали з урахуванням статі і віку (по 3 роки в кожній групі). Це збільшувало об'єм бази правил і, відповідно, кількість антецедентів в ній. Тому треба було знайти компроміс між критеріями точності при побудові шкали і критеріями компактності бази правил. Виходячи з подальших чинників, що впливають на індекс нечіткості, а саме інструментальної невизначеності  $\pm 3 \text{ мм рт. ст.}$  і нестабільності латентного параметру  $\pm 5 \text{ мм рт. ст.}$  (від зміни тиску протягом доби), залишкову складову невизначеності від неповної ідентифікації об'єкту було встановлено на рівні  $\pm 2,5 \text{ мм рт. ст.}$  з організацією шкал для трьох вікових груп: «молодий», «середній», «похилий». У відповідності з рекомендаціями було визначено загальний покривний інтервал  $\pm 10 \text{ мм рт. ст.}$  за покривною ймовірністю  $P = 0,95$ .

Дослідження показали, що використання нелінійних форм ФП мало впливає на значення індексу нечіткості, але при цьому порушується правило розділення одиниці, що справедливе при лінійних ФП. А це, в свою чергу, впливає на рішення, які приймаються при нечіткій класифікації. Для ілюстрації отриманих результатів наведемо ФП двох класів еквівалентності  $\mu_{T_1}(x)$  («малий») і  $\mu_{T_2}(x)$  («середній») з використанням трапецієподібної і степеневої форми (увігнутої  $\mu_{T_1}'(x)$  та випуклої  $\mu_{T_1}''(x)$ ) (рис. 3.5).

Функції приналежності  $\mu_{T_1}'(x)$  і  $\mu_{T_1}''(x)$  для  $U = R^+ \cup \{0\}$  становлять:

$$\mu_{T_1}''(x) = \begin{cases} 1, & \text{якщо } 0 \leq x \leq a_1; \\ 1 - \left(\frac{x - a_1}{a_2 - a_1}\right)^2, & \text{якщо } a_1 \leq x \leq a_2; \\ 0, & \text{якщо } a_2 < x. \end{cases}$$

$$\mu_{T_1}(x) = \begin{cases} 1, & \text{якщо } 0 \leq x \leq a_1; \\ 1 - \left(\frac{x-a_1}{a_2-a_1}\right)^2, & \text{якщо } a_1 \leq x \leq a_2; \\ 0, & \text{якщо } a_2 < x. \end{cases}$$

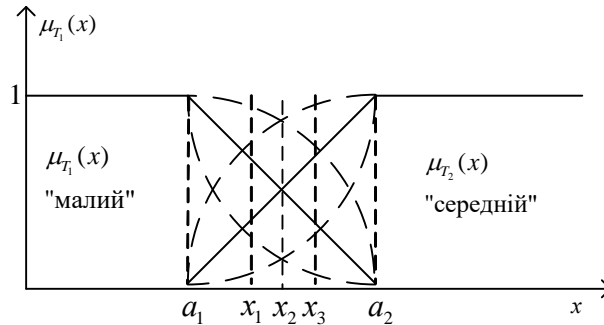


Рисунок 3.5 – Ілюстрація класифікації при різних формах ФП

Функції належності  $\mu_{T_2}(x)$  («середній») на перетині двох класів еквівалентності побудовано за зворотнім перетворенням ФП  $\mu_{T_1}(x)$  (рис. 3.5).

Якщо властивість об'єкта  $x$  класифікується за результатами вимірювання  $x_1, x_2, x_3$ , то результати нечіткої класифікації отримують за перерізом ординат, що відповідають значенням  $x_1, x_2, x_3$  з ФП двох класів еквівалентності (табл. 3.1).

Таблиця 3.1 – Результати класифікації за ступенем належності класам «малий» і «середній»

Форма ФП	Результати вимірювань		
	$x_1 = a_1 + 0,25 \cdot (a_2 - a_1)$	$x_2 = a_1 + 0,5 \cdot (a_2 - a_1)$	$x_3 = a_1 + 0,75 \cdot (a_2 - a_1)$
Трапецієподібна	$T_1   0,75; T_2   0,25$	$T_1   0,5; T_2   0,5$	$T_1   0,25; T_2   0,75$
Увігнута	$T_1   0,5; T_2   0,06$	$T_1   0,3; T_2   0,25$	$T_1   0,13; T_2   0,56$
	$T_1   0,89; T_2   0,11$	$T_1   0,55; T_2   0,45$	$T_1   0,19; T_2   0,81$
Випукла	$T_1   0,94; T_2   0,5$	$T_1   0,75; T_2   0,7$	$T_1   0,44; T_2   0,87$
	$T_1   0,65; T_2   0,35$	$T_1   0,52; T_2   0,48$	$T_1   0,34; T_2   0,66$

За результатами розрахунків (табл. 3.1) можна зробити наступні висновки:



– тільки для трапецієподібної ФП результати класифікації відповідають умові розділення одиниці. Тому другий рядок результатів класифікації для увігнутої та опуклої ФП отримано з перетворенням для виконання умови розділення одиниці;

– за всіма формами ФП результат віднесення до класу еквівалентності за максимумом ступеня приналежності однаковий;

– нелінійні ФП дозволяють провести класифікацію на лінії розділення двох класів еквівалентності;

– якщо в області розмежування двох класів еквівалентності застосувати дві форми ФП – випуклу та увігнуту, тоді пріоритет при віднесенні до класу еквівалентності буде у класу з випуклою ФП (в цьому випадку можна заздалегідь передбачити пріоритетне віднесення до вищого або нижчого за порядком класу еквівалентності).

Якщо вихідна терм-множина ЛЗ є нечіткою, то збільшення її нечіткості з урахуванням невизначеності вимірювання можна отримати з використанням операторів  $DIL(A)$  (для неперервної множини) і оператора нечіткості  $K$  (для дискретної множини) [19].

При урахуванні невизначеності вимірювання індекс нечіткості або нечіткого числа (що є результатом вимірювання), або термів нечіткої ЛЗ (тобто шкали квазіпорядку) повинен збільшуватись. Для збільшення індексу нечіткості неперервної множини може бути використаний оператор розтягування  $DIL(A) = A^{1/N}$ , де  $N > 1$ .

Говорять, за певних умов, що « $A^{1/N}$  більш нечітке, ніж  $A$ » [19]. При урахуванні невизначеності вимірювання обирають  $N$  на основі адекватного збільшення індексу нечіткості від внеску невизначеності в загальний індекс нечіткості. Якщо індекс нечіткості ФП раніше встановленого терма ЛЗ значний, тоді при малій невизначеності вимірювання її внеском можна знехтувати. Критерієм служить співвідношення між індексом нечіткості ФП терма ЛЗ і його можливим збільшенням від урахування невизначеності вимірювання.

Якщо терм ЛЗ задано дискретною чіткою або нечіткою множиною, тоді для врахування внеску невизначеності можна використати оператор нечіткості  $K$  [19].

Оператор нечіткості  $K$  використовується для перетворення звичайних множин в нечіткі і для збільшення нечіткості нечітких множин. Для дискретної універсальної множини оператор представляють як матрицю, що взаємодіє з кожним елементом множини. Як приклад, для чотирьох членів множини оператор  $K$  має вигляд:

$$K = \begin{vmatrix} K_{11} & K_{12} & K_{13} & K_{14} \\ K_{21} & K_{22} & K_{23} & K_{24} \\ K_{31} & K_{32} & K_{33} & K_{34} \\ K_{41} & K_{42} & K_{43} & K_{44} \end{vmatrix}$$

Взаємодія оператора  $K$  з множиною  $A$ , що складається з чотирьох членів

$$A = \sum_{i=1}^4 \mu_A(i) / i,$$

де сума означає тільки формальний перелік, відображається наступним рівнянням [19]:

$$KA = \sum_{i=1}^4 \sum_{j=1}^4 (\mu_A(j) \cdot K_{ij}) / i,$$

де друга сума – це об'єднання нечітких множин, в даному випадку за правилом логічної конорми.

Якщо вихідна нечітка множина, що відповідає терму ЛЗ становить

$$A = 1/a_1 + 0,66/a_2 + 0,33/a_3 + 0/a_4$$

з лінійним індексом  $I_A^L = \frac{2}{n} \sum_{i=1}^n |\mu_{\bar{A}}(x_i) - \mu_{\underline{A}}(x_i)| = 0,33$ , то після взаємодії з оператором

нечіткості  $K$ , що дорівнює

$$K = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0,9 & 1 & 0 & 0 \\ 0 & 0,9 & 1 & 0 \\ 0 & 0 & 0,8 & 0 \end{vmatrix}$$

і означає «більше або менше» [1], отримаємо

$$KA = 1/a_1 + 0,9/a_2 + 0,59/a_3 + 0,26/a_4 \text{ з індексом нечіткості } I_{KA}^L = 0,39.$$

Таким чином, збільшення індексу нечіткості становить

$$I_{KA}^L - I_A^L = 0,06.$$

З огляду на адекватність збільшення індексу нечіткості і значенням розширеної невизначеності, вплив якої враховується, за моделлю (3.2) отримуємо  $\delta = 6,4\%$ .

Але спеціалісти з використання операторів нечіткої логіки застерігають щодо адекватності дій оператора нечіткості, тому що не завжди дія оператора  $K$  на нечіткі множини приводить до збільшення її нечіткості. Тому завжди потрібна перевірка індексу нечіткості множин  $A$  і  $KA$ .

### **3.5. Визначення міри узгодженості між класифікованими даними за умов метричної класифікації**

#### **3.5.1. Використання міри узгодженості для оцінки розсіювання класифікованих даних**

Для аналізу вибірок класифікованих даних важлива оцінка їх розсіювання відносно центральної тенденції, тому що прийняття рішення за вибіркою з меншим розсіюванням більш інформативне, ніж тоді, коли розсіювання велике. В роботі [33] запропоновано використання інформаційних критеріїв, що є мірою узгодженості між ординальними даними ( $Cns$  – consensus measure) і мірою розбіжності ( $Dnt$  – dissention measure), пов'язаних між собою співвідношенням:

$$Cns(x) = 1 - Dnt(x).$$

Недоліком запропонованої раніше формули

$$Cns = 1 - \frac{\sum p_i \cdot \ln(p_i)}{\ln\left(\frac{1}{n}\right)} \quad (3.7)$$

де  $n$  – кількість категорій або класів еквівалентності, автори [33] вважають відсутність врахування розподілу вибірки або порядку групування членів вибірки. І дійсно, для унімодального і антимодального розподілу за формулою (3.6) можуть бути отримані дані і ті ж значення узгодженості, тому що відсутнє значення центральної тенденції. Тому для категорій або класів шкали було запропоновано рангову арифметизацію і отримане співвідношення:

$$Cns(x) = 1 + \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|x_i - \mu_x|}{d_x} \right), \quad (3.8)$$

де  $x_i$  – ранг класу еквівалентності,  $p_i$  – ймовірність або частота, поєднана з  $x_i$ ,  $d_x$  – область визначення, що дорівнює  $x_{\max} - x_{\min}$ ,  $\mu_x$  – середнє значення, що дорівнює

$$\mu_x = \sum_{i=1}^n p_i x_i.$$

Опоненти штучного введення відстані за допомогою рангів [38, 39] вважають, що це протирічить встановленому обмеженню операцій з ординальними даними. Це дійсно так. Але коли ми маємо побудовану раніше шкалу класифікації, де визначена відстань між класами еквівалентності, формулу (3.8) можна пристосовувати для використання міри узгодженості між вибірковими класифікованими даними.

### 3.5.2. Застосування шкали класифікації при визначенні міри узгодженості

Недоліками способів обчислення міри узгодженості вибірки класифікованих даних, заснованому на використанні формули (3.8), є, по-перше, застосування операцій з рангами, які проявляються тільки у відношенні порядку (а не пропорційності), а, по-друге, на визначенні центральної тенденції як середнього арифметичного рангів. Тому запропоновано визначати центральну тенденцію за операторами *med* і *OWA*. Якщо визначено клас еквівалентності  $T_u$ , що відповідає центру вибірки, тоді з застосуванням шкали класифікації може бути отримане значення  $x(T_u)$ , що відповідає центру (тобто середині) класу еквівалентності  $T_u$ . У відповідності з цим може бути запропоноване наступне співвідношення для визначення міри узгодженості:

$$Cns(x) = 1 + \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{x(T_i) - x(T_u)}{x_{\max} - x_{\min}} \right), \quad (3.9)$$

де  $x(T_i)$  – середина класу еквівалентності  $T_i$  за областю визначення шкали класифікації.

Такий спосіб оцінювання міри узгодженості дозволяє її застосовувати при наявності нерівномірності шкали класифікації в області її визначення.

При рівномірному розміщенні класів еквівалентності формула (3.9) може бути подана як:

$$Cns(x) = 1 + \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{n(T_i | T_y)}{n} \right), \quad (3.10)$$

де  $n(T_i | T_y)$  – кількість класів еквівалентності між  $T_i$  і  $T_y$ ,  $n$  – загальна кількість класів еквівалентності.

Але при рівномірному розміщенні класів еквівалентності може бути використана і ймовірнісна оцінка дисперсії, запропонована в роботі [40]

$$VAR = \frac{\sum_{i=1}^{n-1} F_i(1-F_i)}{\max_{F_i} [\sum_{i=1}^{k-1} F_i(1-F_i)]}, \quad (3.11)$$

де  $F_i$  – кумулятивна відносна частота для  $i$ -го класу еквівалентності. Знаменник відповідає максимуму, якщо вибіркові дані поляризовані на два крайні класи еквівалентності. У подальшому формула (3.11) використовується у наступному вигляді:

$$D = \frac{\sum_{k=1}^n F_k(1-F_k)}{(n-1)/4} \quad (3.12)$$

При цьому  $F_n$  завжди дорівнює одиниці. Міра дисперсії для ординальної шкали досягає максимуму, що дорівнює 1, коли розподіл даних поляризований на два екстремальних рівня, та нулю, коли один клас еквівалентності вміщує всі дані. Таким чином крайні значення діапазону міри дисперсії  $D$  і міри розбіжності даних  $Dnt$  співпадають. Але як показав проведений аналіз в інших точках діапазону значення  $D$  завжди перевищувало значення  $Dnt$ . Тобто при порівнянні розсіювання вибірових даних треба обирати одну з вищезначених мір.

Далі в роботі наведено приклади практичного застосування результатів теоретичного дослідження.

### 3.6. Оцінювання якості шкали класифікації за матрицею відповідності

Як характеристику якості встановленої шкали класифікації автором роботи запропоновано використати матрицю відповідності шкали, рядки якої відповідають носіям окремих термів або градаціям шкали, а стовпчики – класам еквівалентності, до яких віднесено розмір величини за класифікацією. Якщо позначити терм-множину  $T_i$  ( $i=1,..N$ ), де  $N$  – кількість класів еквівалентності, що відповідає вимірюваній величині, а  $T_j'$  ( $j=1,..N$ ) – множину результатів класифікації, то на перерізі рядків і стовпчиків отримаємо ймовірність  $P_{ij}$  віднесення розміру величини з  $i$ -того терму до  $j$ -го класу еквівалентності. При цьому якщо  $i=j$ , це відповідає ймовірності правильної класифікації, значення якої розташовані в діагональних елементах матриці відповідності.

Таблиця 3.2 – Матриця відповідності шкали класифікації

$T_i' \backslash T_j'$	$T_1'$	$T_2'$	...	$T_N'$
$T_1(0 \div x_{k1})$	$P_{11}$	$P_{12}$	...	$P_{1N}$
$T_2(x_{k1} \div x_{k2})$	$P_{21}$	$P_{22}$	...	$P_{2N}$
...	...	...	...	...
$T_N(x_{k(N-1)} \div x_{\max})$	$P_{N1}$	$P_{N2}$	...	$P_{NN}$

За матрицею відповідності може бути визначений функціонал, що відповідає інтуїтивно поняттю «якості класифікації». Таким функціоналом або загальним показником якості встановленої шкали обрана норма Фробеніуса:

$$G = \left( \sum_{i=1}^N \sum_{j=1}^N (P_{ij} - I_{ij})^2 \right) / 2N, \text{ де } I_{ij} \text{ – ідеальна матриця відповідності. Показник якості}$$

знаходиться в межах  $0 \leq G \leq 1$ , при цьому 0 відповідає правильній класифікації.

Матриця відповідності характеризує ситуацію, коли апріорної інформації немає і ймовірність знаходження вимірюваної величини однакова за усіма класами еквівалентності. При наявності результату вимірювання з інформацією про його

класифікацією за певними класами еквівалентності результат класифікації може бути скоригований за формулою Байеса. Необхідність корекції визначається за нормою Фробеніуса.

### **3.7. Визначення терм-множини лінгвістичної змінної при класифікації ґрунтів за вмістом рухомого фосфору для оптимізації норми внесення добрив**

Визначення та контроль стану родючості ґрунтів або якості земельних ділянок потрібні при вирішенні таких задач, як тривалий моніторинг земельних ресурсів, поліпшення якості ґрунтів при внесенні добрив, визначення придатності земель при оптимізації різних способів використання, тощо [41, 42]. При вирішенні цих задач можна виділити наступні етапи:

- отримання проб ґрунту за детально визначеною програмою;
- визначення показників родючості ґрунтів за стандартизованими методиками;
- прийняття рішень щодо стану якості ґрунту або способів поліпшення стану ґрунту за внесенням необхідної кількості добрив.

Для отримання необхідних даних використовується шкала класифікації, яка дозволяє представити дані, отримані за метричною шкалою, як категоризовані дані, що є певними рівнями лінгвістичної змінної «показник якості ґрунту» (наприклад, «низький», «середній», «високий»).

Групування ґрунтів за вмістом рухомого фосфору за методом Мачигіна [42] наведено в табл.3.3.

У відповідності з ДСТУ 4114-2002 значення відносного відхилення результатів аналізу для двохсторонньої довірчої ймовірності  $P=0,95$  складають у відсотках:

30 – при вмісті  $P_2O_5$  до 15 мг/кг;

20 – більше 15 мг/кг.

Тоді у відповідності з [5] кількість розрізнявальних градацій при рівномірному розподілі ймовірності значень вимірювальної величини і ентропійному інтервалу невизначеності, знайденому для нормального закону розподілу похибки вимірювання, становить

$$N = \frac{x_2 - x_1}{d} = \frac{x_2 - x_1}{4,133 \cdot \sigma}$$

де  $x_2, x_1$  – верхня і нижня границі інтервала вимірювання,  $d$  – ентропійний інтервал невизначеності,  $\sigma$  – середнє квадратичне відхилення похибки вимірювання:

$$N = \frac{80 \cdot 100}{4,133 \cdot 10 \cdot 40} = 4,85 \approx 5$$

Тобто 5 термів множини відповідають інтервалам групування ґрунтів і умові розрізнення термів.

Таблиця 3.3 – Класифікація ґрунтів

<b>Лінгвістична характеристика</b>	<b>Вміст рухомого фосфору в пробах ґрунту за методом Мачигіна мг/кг</b>
Низький	менше 15
Середній	16-30
Підвищений	31-45
Високий	46-60
Дуже високий	більше 60

Варто зазначити, що в рекламних проспектах зустрічаються рекомендації щодо групування ґрунтів з більшою кількістю градацій, наприклад AgriLab (табл. 3.4).

За теперішнього рівня точності вимірювання вмісту фосфору в пробах ґрунту ці градації є нерозрізненими.

Для визначення функції приналежності терм-множини лінгвістичної змінної «рівень вмісту рухомого фосфору в пробах ґрунту» використовуються трапецієподібні та трикутні форми ФП. Як зазначалось вище, такі функції застосовуються на практиці доволі часто, для їх визначення потрібний найменший у порівнянні з іншими об'єм інформації, для них виконується умова розбиття одиниці [22].



Таблиця 3.4 – Класифікація ґрунтів

Лінгвістична характеристика	Вміст рухомого фосфору в пробах ґрунту за методом Мачигіна мг/кг
Дуже високий	Більше 51 мг/кг
Високий (б)	34-50
Високий (а)	26-33
Середній (б)	21-26
Середній (а)	13-20
Низький (б)	10-12
Низький (а)	6-9
Дуже низький	Менше 5

### 3.7.1. Побудова шкали класифікації показників якості ґрунту з урахуванням невизначеності вимірювання.

Якщо звернутись до першого етапу, що стосується програми відбору проб, то вона, в певній мірі, пов'язана зі способом внесення добрив на досліджуваній ділянці ґрунту. При внесенні добрив за їх розпорошенням на ділянці може бути застосований рандомізований відбір проб без застосування певної системи. При використанні рідких форм добрив, які вважають на процедурою, еталонним матеріалом або їх комбінацією. У відповідності з цим шкали показників якості ґрунту пов'язані з процедурою, що використовується при їх вимірюванні. Так, наприклад, визначення рухомих сполук фосфору в ґрунті може виконуватись за методом Кірсанова (ДСТУ 4405:2005) і за методом Мачигіна (ДСТУ 4114:2002), і відповідні цим методам шкали будуть відрізнятись. Тому ілюстрацією побудови шкали в даній роботі буде шкала, заснована на визначенні рухомих сполук фосфору за методом Мачигіна. За цим методом лінгвістичні характеристики терм-множини шкали «вміст рухомого фосфору в пробах ґрунту за методом Мачигіна в мг/кг» наступні:  $T_1$  – «низький» (менше 15),  $T_2$  – «середній» (16-30),  $T_3$  – «підвищений» (31-45),  $T_4$  – «високий» (46-60),  $T_5$  – «дуже

високий» (більше 60). Зважаючи на відносну похибку вимірювання вмісту фосфору, встановлену як границю довірчого інтервалу з ймовірністю 0.95, що становить 30% для вмісту  $P_2O_5$  до 15 мг/кг і 20% для вмісту більше 15 мг/кг [42], було побудовано шкалу для нечіткої лінгвістичної змінної з використанням трапецієподібних і трикутних функцій приналежності окремих термів [22]:

$$\mu_{T_1}(x) = \begin{cases} 1, & \text{для } 0 \leq x \leq 12,5, \\ (18,5 - x) / 6 & \text{для } 12,5 < x \leq 18,5, \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_2}(x) = \begin{cases} (x - 12,5) / 6 & \text{для } 12,5 < x \leq 18,5, \\ 1 & \text{для } 18,5 < x \leq 24,5, \\ (36,5 - x) / 12 & \text{для } 24,5 < x \leq 36,5, \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_3}(x) = \begin{cases} (x - 24,5) / 12 & \text{для } 24,5 < x \leq 36,5, \\ 1 & \text{для } 36,5 < x \leq 38, \\ (53 - x) / 15 & \text{для } 38 < x \leq 53 \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_4}(x) = \begin{cases} (x - 38) / 15 & \text{для } 38 < x \leq 53, \\ (68 - x) / 15 & \text{для } 53 < x \leq 68 \\ 0 & \text{в інших випадках} \end{cases}$$

$$\mu_{T_5}(x) = \begin{cases} (x - 53) / 15 & \text{для } 53 < x \leq 68, \\ 1 & \text{для } 68 < x \leq 75 \\ 0 & \text{в інших випадках} \end{cases}$$

Терм-множина лінгвістичної змінної «рівень вмісту рухомого фосфору в пробах ґрунту» наведена на рисунку 3.6.

Форма ФП окремих термів змінюється від трапецієподібної до трикутної в залежності від значення похибки. Якщо ентропійний інтервал невизначеності займає практично весь носій ФП терма, форма ФП відповідає трикутній.

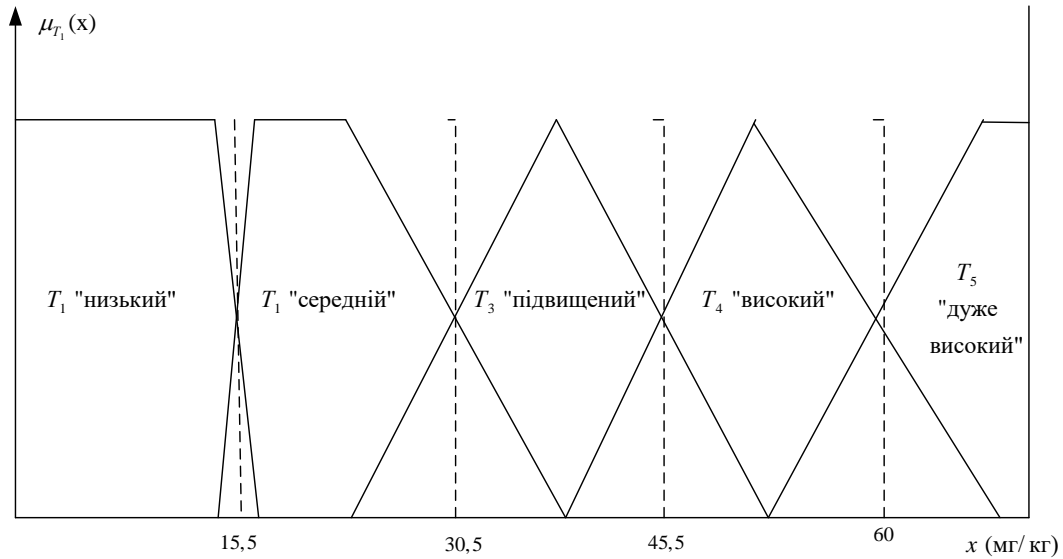


Рисунок 3.6 – Терм-множина лінгвістичної змінної «рівень вмісту рухомого фосфору в пробах ґрунту», побудована з урахуванням невизначеності вимірювання

Враховуючи переріз функцій приналежності сусідніх термів побудовано матрицю відповідності встановленої шкали, де в комірках наведені значення, що є ймовірностями віднесення до певних класів еквівалентності  $P_{ij}$ . Діагональні значення відповідають ймовірностям правильного віднесення ( $i = j$ ), а сусідні відповідають ймовірностям неправильного віднесення ( $i \neq j$ ), що спричинені наявністю невизначеності вимірювання. Наприклад, для першого класу еквівалентності

$P_{11} = (\int_0^{x_{кр}} \mu_{T_1}(x) dx) / (\int_0^{x_{мак}} \mu_{T_1}(x) dx)$ , де  $x_{кр}$  – критична точка функції приналежності, під якою розуміють точку зі ступенем приналежності 0,5 [22]. Для першого класу еквівалентності  $x_{кр} = 15,5$  мг / кг,  $P_{11} = 0,952$ ,  $P_{12} = 0,048$ .

Отримана матриця відповідності може бути використана при заданих границях термів множини як показник якості шкали класифікації, а також для оптимізації вибору границь, тому що, як видно з табл. 3.5, в центральній частині області визначення шкали класифікації ймовірність правильного віднесення зменшується.

Таблиця 3.5 – Матриця відповідності шкали класифікації

Границі значень вимірюваного показника (мг/кг)	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>
0 – 15,5	0,952	0,048	0	0	0
15,5 – 30,5	0,050	0,850	0,1	0	0
30,5 – 45,5	0	0,100	0,775	0,125	0
45,5 – 60,5	0	0	0,125	0,75	0,125
60,5 – 75	0	0	0	0,129	0,871

Характеристики якості встановленої шкали класифікації наступні: норма Фробеніуса  $G = 0,204$ ,  $R = G^{1/2} = 0,452$ ; ймовірнісна оцінка дисперсії  $D = 0,713$ . Але якщо невизначеність вимірювання, за якою встановлювалась шкала, була б вдвічі меншою, то характеристики якості класифікації дорівнювали: норма Фробеніуса  $G = 0,058$ ,  $R = 0,241$ ,  $D = 0,411$ .

Якщо результат класифікації отримують за одноразовими вимірюваннями, він може бути скоригований з використанням матриці відповідності і формули Байеса. Так, якщо результат класифікації за перерізом результату вимірювання зі шкалою класифікації (рис.3.6) становить  $T_1 | 0,7; T_2 | 0,3$  тоді з розрахунком апостеріорної ймовірності для класу еквівалентності «низький» отримуємо

$$P(T_1 | A) = \frac{P_{11} \cdot P(A | T_1) + P_{12} \cdot P(A | T_2)}{P_{11} \cdot P(A | T_1) + P_{12} \cdot P(A | T_2) + P_{22} \cdot P(A | T_2) + P_{21} \cdot P(A | T_1)} = 0,701,$$

звідки  $P(T_2 | A) = 0,299$ . Таким чином, якщо значення норми Фробеніуса мале, результати класифікації не потребують корекції. Але якщо розсіяння терм-множини збільшується, це спричиняє зменшення діагональних ймовірностей, що як наслідок приводить до збільшення норми Фробеніуса. Тоді результати класифікації потребують корекції. Так, якщо  $P_{11} = 0,8$ ;  $G = 0,032$ , тоді за шкалою класифікації при попередньому значенні вимірюваної величини отримуємо  $T_1 | 0,8; T_2 | 0,2$ . При проведенні корекції за формулою Байеса отримуємо  $T_1 | 0,74; T_2 | 0,26$ .

Результати нечіткої класифікації використовуються для прийняття рішення про внесення добрив за правилами:

Правило 1:

Якщо вміст рухомого фосфору «низький», то норма внесення амофосу «висока (90 кг/га)»

Правило 2:

Якщо вміст рухомого фосфору «середній», то норма внесення амофосу «середня (71 кг/га)»

Правило 3:

Якщо вміст рухомого фосфору «підвищений», то норма внесення амофосу «нижче середньої (57 кг/га)»

Правило 4:

Якщо вміст рухомого фосфору «високий», то норма внесення амофосу «низька (33 кг/га)»

Правило 5:

Якщо вміст рухомого фосфору «дуже високий», то амофос вносити не потрібно.

Для класифікації вмісту фосфору нечіткий класифікатор визначає переріз ординати, що відповідає результату вимірювання з терм-множиною лінгвістичної змінної. Якщо переріз відповідає такому значенню  $\mu_{T_i}(x)$ , яке дорівнює одиниці, використовується відповідне правило з сукупності Пр1-Пр-5. Якщо переріз включає два терми множини, наприклад  $T_1 | 0,6$ ;  $T_2 | 0,4$  тоді норма внесення добрив  $H$  розраховується за двома правилами з ваговими коефіцієнтами, а саме

$$H = 90 \text{ кг} / \text{га} \cdot 0,6 + 71 \text{ кг} / \text{га} \cdot 0,4 = 82,4 \text{ кг} / \text{га}$$

Але така ситуація з обчисленням необхідної норми внесення добрив має місце коли база правил працює з узгодженими даними. Тому першим етапом в опрацюванні класифікованих даних повинна бути їх перевірка на узгодженість. Як приклад було обрано вибірку порівняно узгоджених даних:

Таблиця 3.6 – Вибірка узгоджених даних

Результати вимірювання вмісту рухомого фосфору	16	14	15	16	17	16	17
Дані після класифікації у відповідності зі шкалою встановленою шкалою	$T_2$	$T_1$	$T_1$	$T_2$	$T_2$	$T_2$	$T_2$

Медіані ранжованої вибірки відповідає клас еквівалентності  $T_2$ , відповідно значення  $x(T_2) = 23 \text{ мг/кг}$ .

Інтегральна характеристика віднесення даних до відповідних класів еквівалентності для бази правил:

$$T_1 | 0,29; T_2 | 0,71.$$

Міра узгодженості отриманих даних за формулою:

$$Cns(x) = 1 + 0,29 \cdot \log_2\left(1 - \frac{|7,5 - 23|}{75}\right) + 0,71 \cdot \log_2\left(1 - \frac{|23 - 23|}{75}\right) = 0,9.$$

Міра розбіжності  $Dnt = 0,1$ .

Так як класи еквівалентності розташовані рівномірно в діапазоні вимірювання (табл. 3.6), для оцінки розсіювання можна скористатись формулою (3.12) і інтегральною характеристикою вибірки, за якою  $Dnt = 0,205$ .

За результатами обчислень можна вважати, що вибірка даних достатньо узгоджена.

Для вибірки з більшим розсіюванням:

Таблиця 3.7 – Вибірка узгоджених даних з більшим розсіюванням

Результати вимірювання вмісту рухомого фосфору	15	20	31	47	18	25	50
Дані після класифікації у відповідності зі шкалою табл.	$T_1$	$T_2$	$T_3$	$T_4$	$T_2$	$T_2$	$T_4$

Отримано  $med = T_2$ ,  $x(T_2) = 23 \text{ мг/кг}$ , інтегральну характеристику віднесення даних до відповідних класів еквівалентності:  $T_1 | 0,14$ ;  $T_2 | 0,43$ ;  $T_3 | 0,14$ ;  $T_4 | 0,29$ .

Міра узгодженості  $Cns(x) = 0,693$ , міра розбіжності  $Dnt = 0,307$ , міра розсіювання  $D = 0,541$ .

За результатами класифікації дані вибірки неузгоджені і потрібний додатковий аналіз методики їх отримання.

Якщо врахувати невизначеність вимірювання і варіацію вмісту рухомого фосфору в стрічці і при рандомізованому виборі точок відбору проб (за даними, отриманими в [43]), то можна провести моделювання з використанням отриманого середньоквадратичного відхилення і прийнятого розподілу значень отриманих проб ґрунту. Для нормального розподілу вибірки, сумарного середнього квадратичного відхилення 10% і об'єму вибірки  $n = 20$  було проведено моделювання в межах третього класу еквівалентності з середнім значенням вмісту рухомого фосфору 38 мг/кг. Вибір цього класу еквівалентності був обумовлений тим, що форма ФП цього класу відповідає трикутній, що в свою чергу свідчить про максимальне розсіювання. Результати класифікації були отримані за функціями приналежності  $\mu_{T_i}(x)$  для двох алгоритмів роботи нечіткого класифікатора: за максимумом ступеня приналежності перерізу (стовпчик 3 табл.3.8) і з урахуванням всіх ступенів приналежності перерізу результату вимірювання сім'єю з ФП  $\mu_{T_i}(x)$  (стовпчик 4 табл.3.8).

Багаторазове повторення моделювання для вибірки  $n = 20$  показало високу збіжність загальних результатів класифікації. При цьому в процедурі моделювання враховано розсіювання реальних даних за рандомізованим відбором (перший стовпчик) і невизначеність вимірювання (результати класифікації за шкалою). Якщо порівняти дані моделювання з даними, що наведені в табл. 3.5, то можна зробити висновок, що дані таблиці відповідності можуть бути використані для визначення максимальної границі міри збіжності даних відбору проб ґрунту, за яких програму відбору проб можна вважати задовільною.

Таблиця 3.8 – Результати класифікації за моделюванням процедури відбору проб за вибіркою  $n = 20$  при нормальному розподілі результатів вимірювання

Ранжовані результати вимірювання (мг/кг)	Кількість результатів	Результат класифікації за алгоритмом 1	Результат класифікації за алгоритмом 2
28	1	$T_2$	$T_2   0,71; T_3   0,29$
35	1	$T_3$	$T_2   0,13; T_3   0,87$
36	3	$T_3$	$T_2   0,04; T_3   0,96$
37	3	$T_3$	$T_3   1$
38	3	$T_3$	$T_3   1$
39	2	$T_3$	$T_3   0,93; T_4   0,07$
40	1	$T_3$	$T_3   0,87; T_4   0,13$
41	1	$T_3$	$T_3   0,8; T_4   0,2$
42	1	$T_3$	$T_3   0,73; T_4   0,27$
43	2	$T_3$	$T_3   0,67; T_4   0,33$
44	1	$T_3$	$T_3   0,6; T_4   0,4$
47	1	$T_4$	$T_3   0,4; T_4   0,6$
51	1	$T_4$	$T_3   0,13; T_4   0,87$
Загальний результат нечіткої класифікації		$T_2   0,05; T_3   0,85; T_4   0,1$	$T_2   0,048; T_3   0,792; T_4   0,16$

### 3.7.2. Застосування критерію перевірки узгодженості (збіжності) класифікованих даних вмісту рухомого фосфору.

Сучасні технології внесення мінеральних добрив є на теперішній час не лише високотехнологічними, але й такими, що можуть спричинити негативні зміни, тому необхідна певна система отримання проб ґрунту. Так в роботі [43] зазначено, що так, як застосування безводного аміаку можливе лише шляхом його локального внесення у



стрічки на певну глибину, зміни, що відбуваються у стрічках його внесення можуть бути доволі контрастними порівняно з показниками ґрунту у міжрядді внаслідок формування осередків з високою концентрацією солей. Тому в роботі [43] було запропоновано і досліджено відбір проб за схемою зигзагоподібного перетинання, що забезпечувало однакову кількість проб у стрічці і у міжрядді. Як критерій вибору оптимальної програми відбору проб в роботі [43] було обрано коефіцієнт парної кореляції між вмістом рухомого фосфору та амонійного азоту у стрічці та у змішаному зразку, що складався з 20 індивідуальних проб, відібраних за схемою зигзагоподібного перетинання. Отримані значення коефіцієнту кореляції були достатньо високими при вибірці в кількості 20-ти проб, але зі зменшенням їхньої кількості до 10 або навпаки їх збільшення до 30-40 не було виявлено чіткого зв'язку між показниками родючості, визначеними у стрічці внесення добрив та у зразках, відібраних за зигзагоподібною схемою. Також не було виявлено кореляції за умов рандомізованого відбору.

З розглянутого вище можна зробити висновки, що по-перше, є необхідність в перевірці узгодженості вихідних даних при прийнятті рішень щодо показників якості ґрунту, а по-друге необхідно мати декілька критеріїв, що дозволяють перевірити узгодженість саме між класифікованими даними. Тому в представленій роботі запропоновано використання інформаційних критеріїв перевірки узгодженості класифікованих даних з використанням прийнятої при класифікації шкали.

В п.3.5.2 автором було запропоновано наступний критерій перевірки узгодженості (збіжності) класифікованих даних за інформаційною мірою узгодженості  $Cns$  (за [33]  $Cns$  – consensus measure) або неузгодженості  $Dnt$  (за [33]  $Dnt$  – dissention measure), де  $Cns(x) = 1 - Dnt(x)$ :

$$Cns(x) = 1 + \sum_{i=1}^n p_i \log_2 \left( 1 - \frac{|x(T_i) - x(T_y)|}{x_{\max} - x_{\min}} \right), \quad (3.13)$$

де  $x(T_i)$  – середина класу еквівалентності  $T_i$  за областю визначення шкали класифікації,  $x(T_y)$  – середина класу еквівалентності, що відповідає центру вибірки,  $p_i$  – оцінка ймовірності знаходження в класі еквівалентності  $T_i$ .

Граничне значення  $Cns$  для заданої невизначеності можна отримати з застосуванням матриці відповідності (табл. 3.5) за центром вибірки, що відповідає певному класу еквівалентності. Наприклад, для  $T_y = T_3$ ,  $x(T_y) = 38 \text{ мг/кг}$ ,

$$\begin{aligned} (Cns)^{sp} &= 1 + P_2 \log_2 \left( 1 - \frac{|x(T_2) - x(T_y)|}{x_{\max} - x_{\min}} \right) + P_4 \log_2 \left( 1 - \frac{|x(T_4) - x(T_y)|}{x_{\max} - x_{\min}} \right) = \\ \text{отримуємо} \quad &= 1 + 0,1 \cdot \log_2 \left( 1 - \frac{|23 - 38|}{75} \right) + 0,125 \cdot \log_2 \left( 1 - \frac{|53 - 38|}{75} \right) = 0,928. \end{aligned}$$

Якщо міра узгодженості даних вибірки більша або дорівнює  $(Cns)^{sp}$ , вибірка вважається узгодженою. Для даних таблиці 3.8  $Cns = 0,933$ . Отже, дана вибірка вважається узгодженою. Якщо програма отримання проб є недосконалою, що є причиною отримання аномальних результатів, міра узгодженості даних зменшується, що є сигналом для удосконалення програми. В таблиці 3.9 наведено результати моделювання процедури класифікації, за якої до нормально розподілених даних додано суміш з більшим розсіюванням.

Таблиця 3.9 – Результати класифікації за моделюванням засміченої вибірки

Ранжовані результати вимірювання (мг/кг)	Кількість результатів	Результат класифікації за алгоритмом 1	Результат класифікації за алгоритмом 2
23	1	$T_2$	$T_2   1$
24	1	$T_2$	$T_2   1$
25	1	$T_2$	$T_2   0,96; T_3   0,04$
26	2	$T_2$	$T_2   0,88; T_3   0,12$
27	2	$T_2$	$T_2   0,79; T_3   0,21$

Продовження таблиці 3.9			
28	1	$T_2$	$T_2   0,71; T_3   0,29$
30	1	$T_2$	$T_2   0,52; T_3   0,48$
36	2	$T_3$	$T_2   0,04; T_3   0,96$
37	2	$T_3$	$T_3   1$
38	2	$T_3$	$T_3   1$
39	2	$T_3$	$T_3   0,93; T_4   0,07$
40	1	$T_3$	$T_3   0,87; T_4   0,13$
41	1	$T_3$	$T_3   0,8; T_4   0,2$
42	1	$T_3$	$T_3   0,73; T_4   0,27$
Загальний результат		$T_2   0,45; T_3   0,55$	$T_2   0,38; T_3   0,583; T_4   0,037$

Дані таблиці 3.9 не можна вважати достатньо узгодженими, про що свідчить значення міри узгодженості:  $Cns = 0,866$ , що менше за граничне значення, яке обчислене з урахуванням невизначеності вимірювання. Перевагою формули 3.13 для міри узгодженості є те, що її можна використовувати при різних областях визначення окремих термів, тобто при нерівномірній шкалі класифікації.

### 3.8. Застосування нечіткої класифікації при визначенні стану заповнення сміттям підземної урни

Однією з актуальних проблем великих міст є збирання і раціональне перероблення сміття. Для збирання сміття в центрі міст, в екскурсійних зонах пропонується використовувати підземні урни [44, 55]. Успішна експлуатація таких сміттєвих урн, що представляють собою великий сміттєзбиральний контейнер, можлива при виконанні цілої низки умов: автономного живлення, можливості вимірювання рівня сміття, можливості виклику сміттєзбиральної техніки при наявності інформації про середнє заповнення сміттям групи урн, тощо.

### 3.8.1. Вимірювач рівня сміття

Структурна схема вимірювача рівня сміття наведена на рис.3.7, де GSM – модуль GSM/GPRS мобільного зв'язку, МК – мікроконтролер, БДР1 – блок датчика рівня 1 (основний датчик рівня), БДР2 – блок датчика рівня 2 (допоміжний датчик рівня), БДТ – блок датчика температури, СМ – сонячний модуль, АКК – акумулятор, СП – система пожежогасіння.

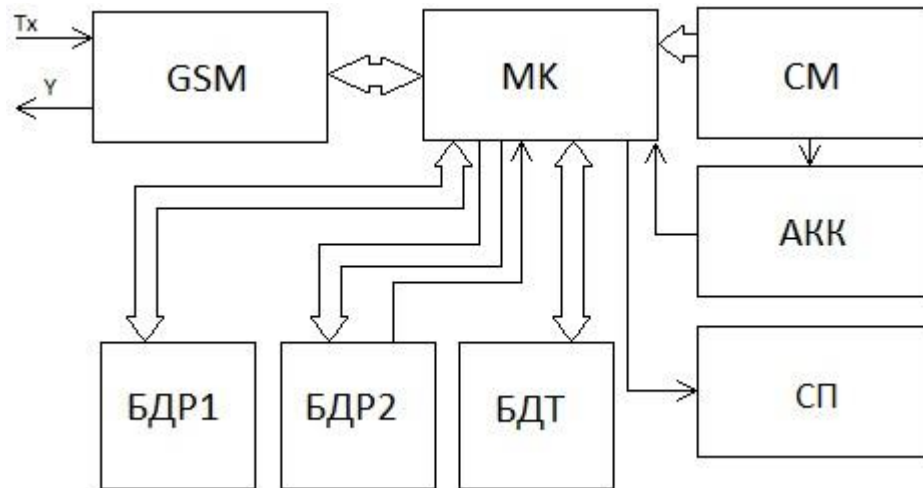


Рисунок 3.7 – Структурна схема вимірювача рівня сміття в підземній урні

Основний принцип роботи схеми наступний: після того як мікроконтролер прокинувся від режиму сну, він одразу подає сигнали для вмикання БДР1, БДР2, БДТ та GSM. Після вимірювання датчиками рівня та температури, дані надходять до мікроконтролера, який у свою чергу передає їх до GSM модуля, де йде пересилання даних до сервера. АКК живить усю периферію та передає дані до МК про стан заряду акумулятора. Дані живлення АКК з МК передаються до GSM, а далі до сервера. СМ – живить акумулятор та передає дані про стан сонячних панелей. При підвищенні температури в урні, що вимірюється за допомогою датчика БДТ, спрацьовує СП – система пожежогасіння. БДР1 включає в себе основний датчик вимірювання рівня сміття в урні, який передає дані до мікроконтролера через інтерфейс. БДР2 – допоміжний резервний датчик вимірювання рівня сміття передає дані в аналоговому вигляді до мікроконтролера, що має вбудований АЦП для перетворення аналогових даних в кодові. Як основний датчик використовувався лазерний датчик відстані

VL53LOX, виконаний в найменшому корпусі для поверхневого монтажу. Він має інтерфейс I2C, що полегшує передачу та програмування даних.

Джерело оптичного сигналу, повністю невидиме для людського ока, в поєднанні з вбудованим інфрачервоним фільтром, забезпечує більшу відстань вимірювання, меншу чутливість до рівня зовнішнього освітлення і більшу стійкість до перехресних перешкод, що можуть бути викликані скляними поверхнями. Розташування датчика буде по центру, на рівні максимальної довжини урни. Для захисту датчика від механічних пошкоджень БДР1 розташовується в колбі з акрилу.

За паспортними даними лазерного далекоміра VL53LOX похибка вимірювання рівня складає  $\pm 7\%$  на білому смітті і  $\pm 12\%$  на сірому смітті. Похибка від зміни форми насипу (плоска, горбик) не перевищує  $2\%$ .

За правилами, розміщеними в базі правил на сервері, визначається терм, до якого відноситься результат вимірювання довжини. Рішення приймається за максимальним значенням ступеня приналежності. За певним термом, до якого відноситься результат вимірювання довжини, визначається вербальна характеристика рівня сміття в певній підземній урни.

### **3.8.2 Застосування шкали класифікації при визначенні стану заповнення сміттям підземної урни.**

Безпосередньому вимірюванню підлягає відстань від верхньою точки урни до рівня насипу сміття. Максимальна висота урни становить 2400 мм. При побудові шкали класифікації визначення кількості термів і формування їх функцій приналежності відбувалось з урахуванням сумарної невизначеності вимірювання відстані, складовими якої були похибка далекоміра, складові від впливу на результати вимірювання щільності та різрорідності сміття, кольору сміття (сірий, білий), форми насипу (наявність «горбика») та інших впливних величин. Загальна похибка вимірювання становить  $\pm 14\%$ . У відповідності з цим встановлено 5 термів множини станів заповнення урни:

- $T_1$  – урна практично порожня ( $0 \div 300$  мм);

- $T_2$  – урна заповнена менше ніж наполовину ( $300 \div 900$  мм);
- $T_3$  – урна заповнена наполовину ( $900 \div 1500$  мм);
- $T_4$  – урна заповнена більше ніж наполовину ( $1500 \div 2100$  мм);
- $T_5$  – урна практично заповнена ( $2100 \div 2400$  мм).

Шкала і матриця класифікації представлені на рис. 3.8 і в таблиці 3.10, де в діагоналі матриці розміщені ймовірності правильного віднесення до певного класу еквівалентності. Значення ймовірності правильного віднесення до певного класу еквівалентності свідчать про те, що кількість класів еквівалентності не можна збільшувати при великій похибці вимірювання.

Таблиця 3.10 – Матриця шкали класифікації з ймовірностями правильного віднесення до певних класів еквівалентності

Стан заповнення урни	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$
Рівень заповнення урни					
$0 \div 300$ мм	$P_{11} = 0,75$	$P_{12} = 0,25$	$P_{13} = 0$	$P_{14} = 0$	$P_{15} = 0$
$300 \div 900$ мм	$P_{21} = 0,12$	$P_{22} = 0,79$	$P_{23} = 0,09$	$P_{24} = 0$	$P_{25} = 0$
$900 \div 1500$ мм	$P_{31} = 0$	$P_{32} = 0,09$	$P_{33} = 0,86$	$P_{34} = 0,05$	$P_{35} = 0$
$1500 \div 2100$ мм	$P_{41} = 0$	$P_{42} = 0$	$P_{43} = 0,05$	$P_{44} = 0,93$	$P_{45} = 0,02$
$2100 \div 2400$ мм	$P_{51} = 0$	$P_{52} = 0$	$P_{53} = 0$	$P_{54} = 0,03$	$P_{55} = 0,97$

При виконанні одноразових вимірювань можна рекомендувати алгоритм роботи нечіткого класифікатора (НК) за перерізами всіх термів шкали класифікації (рис. 3.8). На рис. 3.8 наведено приклад результату класифікації:  $T_1 | 0,17$ ;  $T_2 | 0,83$ . При формуванні цього результату враховано невизначеність вимірювання, яка приводить до розсіювання результату класифікації за класами еквівалентності.

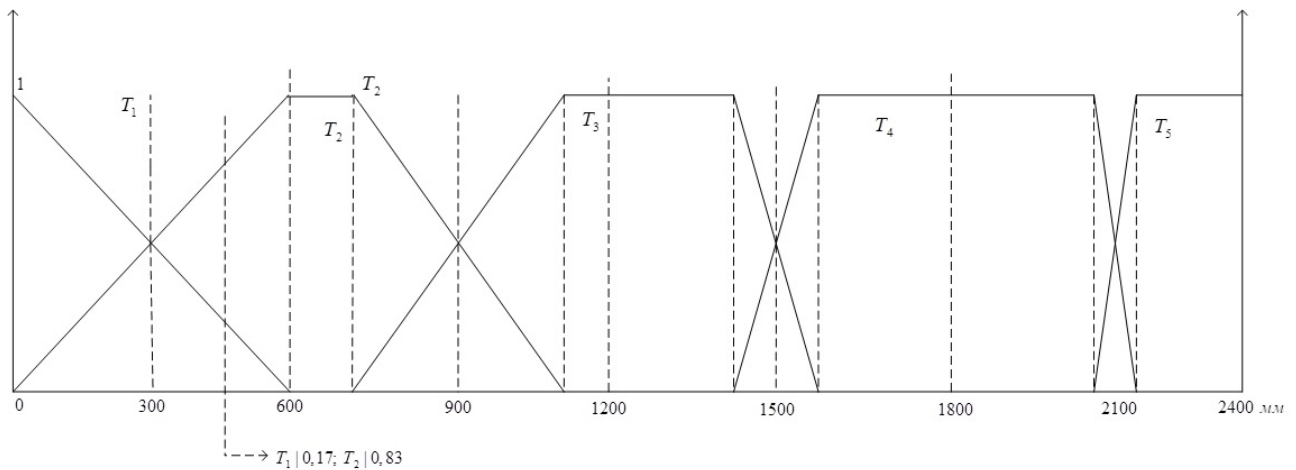


Рисунок 3.8 – Шкала класифікації стану заповнення смітцевої урни з урахуванням похибки вимірювання відстані від рівня насипу сміття

Якщо виконуються багаторазові вимірювання і використовується алгоритм НК за максимумом перерізу, то результати класифікації характеризують тільки розсіювання результатів вимірювання без урахування невизначеності вимірювання. Для врахування невизначеності вимірювань можна використати обернену матрицю шкали (табл. 3.11), і за добутком результатів класифікації за НК (терм множина  $T_1, T_2, \dots, T_n$ ).

Таблиця 3.11 – Обернена матриця шкали класифікації (з нормалізацією рядками)

$\nearrow$	$T'_1$	$T'_2$	$T'_3$	$T'_4$	$T'_5$
$T_1$	0,86	0,14	0	0	0
$T_2$	0,22	0,70	0,08	0	0
$T_3$	0	0,09	0,86	0,05	0
$T_4$	0	0	0,05	0,92	0,03
$T_5$	0	0	0	0,02	0,98

Для одноразового вимірювання застосування оберненої матриці приведе до результату класифікації, що справедливий при довільному розташуванні результату вимірювання в області визначення ФП лінгвістичної змінної. Тому розсіювання результату класифікації буде значно перевищувати реальне. Так при кінцевому

результаті класифікації за максимумом перерізу отримуємо  $T_2$ . Тоді за оберненою матрицею загальне розсіювання, що відповідає довільному розташуванню становить  $T_1 | 0,22; T_2 | 0,70; T_3 | 0,08$ , що перевищує реальне для конкретного випадку  $T_1 | 0,17; T_2 | 0,83$ . Але застосування оберненої матриці корисне для порівняння шкал з різною кількістю і розташуванням термів лінгвістичної змінної.

Якщо при багаторазових вимірюваннях використовується алгоритм роботи НК з урахуванням перерізів всіх термів, то кінцевий результат класифікації можна отримати за загальною ФП.

Приклад застосування цих способів наведено для опрацювання результатів вимірювання рівня заповнення сміттям декількох урн (для зменшення об'єму обчислень і наочності).

Таблиця 3.12 – Ранжовані результати класифікації

№ урни	Результати класифікації стану заповнення урни за перерізами терм-множини	Результати класифікації стану заповнення урни за максимумами перерізу
1	$T_1   0,34; T_2   0,66$	$T_2$
2	$T_1   0,25; T_2   0,75$	$T_2$
3	$T_2   1$	$T_2$
4	$T_2   0,8; T_3   0,2$	$T_3$
5	$T_2   0,05; T_3   0,95$	$T_3$
	$T_1   0,12; T_2   0,65; T_3   0,23$	$T_2   0,8; T_3   0,2$

Результат класифікації наведено в останньому рядку. Для першого стовпчика при класифікації враховано розсіювання результатів вимірювання і невизначеність вимірювання. Для другого стовпчика враховано тільки розсіювання результатів вимірювання. Для врахування невизначеності скористаємось оберненою матрицею (табл. 3.13).



Таблиця 3.13 – Обернена матриця

$T_1'$	$T_2'$	$T_3'$	$T_4'$
0,22·0,8	0,7·0,8	0,08·0,8	
	0,09·0,2	0,86·0,2	0,05·0,2
0,18	0,58	0,23	0,01

Нехтуючи термами з малими значеннями ступеня приналежності отримуємо кінцевий результат:

$$T_1 | 0,18; T_2 | 0,59; T_3 | 0,23.$$

Порівняння отриманих результатів як нечітких множин проведено за відносною відстанню Хемінга  $\rho_{ij}$  (таблиця 3.14).

Таблиця 3.14 – Оцінка збіжності результатів класифікації

1	$T_1   0,12; T_2   0,65; T_3   0,23$	
2	$T_2   0,8; T_3   0,2$	$\rho_{12} = 0,1$
3	$T_1   0,18; T_2   0,59; T_3   0,23$	$\rho_{13} = 0,04$

На цьому прикладі показано, що результати класифікації з використанням алгоритму НК за перерізом всіх термів ЛЗ і узагальненої ФП і з використанням алгоритму НК за максимумом перерізу і оберненою матрицею шкали класифікації збігаються (при чому збіжність збільшується зі збільшенням об'єму вибірки).

### Загальні висновки за розділом 3.

Для того, щоб вимірювальний канал інтелектуальної системи співпрацював з нечіткою базою правил, а результати вимірювань використовувались в нечітких обчисленнях, необхідний блок, що виконує функцію нечіткого класифікатора і реалізує процедуру класифікації. При цьому використовується шкала класифікації і обирається певний алгоритм роботи нечіткого класифікатора.

В роботі проведено аналіз процедури встановлення шкали нечіткої класифікації з урахуванням невизначеності вимірювань як при визначенні кількості термів лінгвістичної змінної, так і при виборі форми функції приналежності окремих термів.

При цьому враховується як нечіткість експертної інформації так і невизначеність вимірювання.

Встановлена послідовність етапів розробки або відтворення шкали метричної класифікації.

Проведено аналіз чинників, які впливають на розмитість функцій приналежності окремих термів лінгвістичної змінної, встановлено критерій і технології їх об'єднання.

Як оцінку якості встановленої шкали класифікації обрано матрицю відповідності і норму Фробеніуса.

Розглянуто оператори та критерії, які застосовуються при визначенні центральної тенденції вибірки класифікованих даних, ранжуванні вибірок, перевірки їх узгодженості.

Запропоновано метод застосування шкали класифікації при визначенні узгодженості або розсіювання класифікованих даних, які використовуються у подальшому для отримання висновку за базою правил.

Наведено приклади практичного застосування способів отримання класифікованих даних в інтелектуальних вимірювальних системах, що стосуються встановлення шкали класифікації з урахуванням невизначеності вимірювання, взаємодії шкали класифікації з алгоритмом нечіткого класифікатора, способів використання класифікованих даних.

## **РОЗДІЛ 4. ЗАСТОСУВАННЯ КЛАСИФІКАЦІЇ ДЛЯ ОПРАЦЮВАННЯ ЕКСПЕРТНОЇ ІНФОРМАЦІЇ, ПОДАНОЇ ЗА ВИБІРКАМИ ВЕРБАЛЬНИХ ДАНИХ.**

Одним з сучасних напрямків розвитку теорії вимірювань є вимірювання ординальних величин, які за [39, 45, 46, 47] визначаються у відповідності з домовленою процедурою вимірювання, для яких може бути встановлено загальне відношення порядку серед інших однорідних величин у відповідності з їх розміром, але для яких не існує алгебраїчних операцій. Ординальна шкала значень величин або впорядкованих класів еквівалентності встановлюється у відповідності з процедурою вимірювань. Серед ординальних шкал виділяють більш сильні і більш слабкі шкали. До більш сильних відносять арифметизовані ординальні шкали, для яких справедливе обмежене відношення комбінування і які за властивостями наближаються до метричних інтервальних шкал. Такі шкали отримали назву «асоціативних» [45]. Серед більш слабких виділяють шкали впорядкованих вербальних класів еквівалентності, або шкали «квазіпорядку», що є впорядкованими шкалами класифікації. За такими шкалами можуть бути отримані дані про якість технологічного процесу.

Шкали квазіпорядку або передпорядку використовують при поданні експертної інформації [48, 49], при поданні результатів спостережень, при проведенні візуального контролю, при проведенні тактильного контролю за зразками, тощо.

Даний розділ присвячений розробці і аналізу методів опрацювання експертної інформації, поданої за вибірками вербальних даних за умов наявності відношення порядку між вербальними категоріями.

### **4.1. Відношення та оператори, які використовуються при роботі з вербальними даними.**

При роботі з вербальними даними можуть бути використані співвідношення і оператори, що побудовані на непараметричних статистиках або на нечіткій логіці. Для визначення центральної тенденції при непарній кількості членів вибірки може бути використана медіана вибірки  $med$ , що є центральним членом ранжованої вибірки, і

незалежно від кількості членів вибірки – оператор  $OWA$ , що може виступати як емулятор середнього арифметичного [50, 51, 52, 53].

$$OWA = \max_{k=1}^n [\min \{Q(k), b_k\}] \quad (4.1)$$

$Q(k) = S(f_k)$ ,  $k = 1, 2, \dots, n$ ,  $n$  – розмір вибірки;  $S(f_k)$  –

$f_k$ -тий рівень лінгвістичної шкали;

$$f_k = \text{Int} \left\{ 1 + \left[ k \frac{t-1}{n} \right] \right\} \quad (4.2)$$

$f_k$  – найближче ціле число в формулі (4.2);  $t$  – кількість градацій вербальної шкали;  $b_k$  – тий елемент вибірки, попередньо ранжованої за порядком спадання градацій вербальної шкали. Кількість розрізняваних станів визначається за кількістю градацій вербальної шкали, що використовується при отриманні відповідних вербальних характеристик кожного члена вибірки. При використанні оператора  $OWA$  існує обмеження: кількість членів вибірки повинна бути більшою за кількість градацій вербальної шкали.

Відношення домінування встановлюється між двома впорядкованими вибірками, що складаються з  $n$  чисел [20]. Якщо є дві впорядковані вибірки:

$$v = (k_1, k_2, \dots, k_n) \quad \text{та} \quad v' = (k'_1, k'_2, \dots, k'_n),$$

тоді вибірка  $v'$  домінує над вибіркою  $v$ , тобто  $v' \geq v$

тоді і тільки тоді, коли

$$k'_1 \geq k_1, \quad k'_2 \geq k_2, \quad k'_n \geq k_n \quad (4.3)$$

Символи  $\geq$  та  $\geq$  використовуються для відношення несуворого порядку. Для відношення суворого порядку  $v$

$$v' > v, \quad (4.4)$$

за яким " $v'$  суворо домінує  $v$ ", потрібно, щоб виконувалась (4.3) і було хоча б одне  $k'_i$  і  $k_i$  між якими існує суворий порядок, тобто  $k'_i > k_i$ .

Згідно [20] відношення домінування встановлює відношення порядку (цілкове або часткове) між вибірками об'ємом  $n$ .

Відношення домінування може бути розповсюджене на вибірки даних, отриманих з застосуванням вербальних шкал квазіпорядку, для яких встановлені відношення порядку між класами еквівалентності. Для цього розповсюджене співвідношення (4.3) має вигляд:

$$k'_1 \geq k_1, k'_2 \geq k_2, k'_n \geq k_n \quad (4.5)$$

#### **4.2. Технології застосування процедури класифікації при визначенні центральної тенденції вербальної вибірки.**

На основі використання перелічених вище операторів в роботі запропоновано наступні технології застосування процедури класифікації за центральною тенденцією вербальної вибірки при визначенні стану об'єкту і при побудові контрольних карт технологічних процесів. При перевірці збіжності оцінок центральної тенденції вибірки за операторами *med* і *OWA* (табл. 4.1), показано, що результати визначення центральної тенденції за медіаною і оператором *OWA* не співпадають. Це пояснюється тим, що медіана нечутлива до змін крайніх членів впорядкованої вибірки, а оцінка *OWA* навпаки – чутлива, але внаслідок ступінчастості оператора *OWA* збільшується кількість вибірок, класифікованих за середнім класом еквівалентності (табл. 4.1). Так як обидві оцінки не є універсальними, то для підвищення точності запропоновано використати дві оцінки: за медіаною і за оператором *OWA*. В тих випадках, коли результати визначення центральної тенденції не співпадають, а знаходяться в сусідніх класах еквівалентності, запропоновано використовувати проміжні класи еквівалентності (для табл.8 це напівгрубий НГ, напівточний НТ).

Таким чином розроблено метод опрацювання вербальних вибірок малого об'єму з метою класифікації стану об'єкту з обчисленням двох оцінок центральної тенденції, за якими при їх нееквівалентності визначають додаткові (проміжні) класи еквівалентності стану об'єкту.

Таблиця 4.1 – Способи оцінювання центральної тенденції вербальних вибірок, отриманих при візуальному контролі рівня шорсткості оброблених поверхонь, де Г – грубий, С – середній, Т – точний

Досліджувані вибірки	<i>med</i>	<i>OWA</i>	<i>med i OWA</i>	Досліджувані вибірки	<i>med</i>	<i>OWA</i>	<i>med i OWA</i>
ГГГГГ	Г	Г	Г	ТСССГ	С	С	С
СГГГГ	Г	Г	Г	ТСССС	С	С	С
ТГГГГ	Г	Г	Г	ТТССГ	С	С	С
ССГГГ	Г	С	НГ	ТТССС	С	С	С
ТСГГГ	Г	С	НГ	ТТТГГ	Т	С	НТ
ТТГГГ	Г	С	НГ	ТТТСГ	Т	С	НТ
СССГГ	С	С	С	ТТТСС	Т	С	НТ
ССССГ	С	С	С	ТТТТГ	Т	Т	Т
ССССС	С	С	С	ТТТТС	Т	Т	Т
ТССГГ	С	С	С	ТТТТТ	Т	Т	Т
ТТСГГ	С	С	С				

При більших об'ємах вербальних вибірок ( $n \geq 10$ ) збіжність оцінок центральної тенденції може бути перевірена за непараметричними критеріями.

В табл. 4.2 наведені результати класифікації стану окремих урн, за якими треба прийняти рішення про виклик сміттєзбиральної техніки

Таблиця 4.2 – Стан заповнення окремих урн

№ урни	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	МІД	МІД	МІД	МПД	МПД	МПД	МПД	МПД	МПД	ПД	БПД	БПД	БПД	МАД	МАД

Центральна тенденція за *med* – це рівень МПД, за оператором *OWA* – це рівень ПД. Для визначення збіжності оцінок центральних тенденцій був використаний довірчий інтервал медіани з границями [54]  $\Delta_{rl} = x_u - med$ ,  $\Delta_{rh} = x_v - med$ , де  $x_u$ ,  $x_v$  – порядкові статистики вибірки з номерами  $u$ ,  $v$ ;  $u$  – ближнє ціле число, що менше від

$(n+1-z_p\sqrt{n})/2$ ;  $v$  – ближнє ціле число, що більше від  $(n+1+z_p\sqrt{n})/2$ ;  $z_p$  – квантиль нормального розподілу; для  $P=0,95$ ,  $x_u = x_4 = \text{МПД}$ ;  $x_v = x_{12} = \text{БПД}$ .

Так як  $X_{OWA}$  знаходиться в межах МПД і БПД, то результати  $X_{OWA}$  і  $X_{med}$  збігаються. Рішення про виклик сміттезбиральної техніки приймаємо, якщо  $X_{med}$  або  $X_{OWA}$  дорівнює або перевищує контрольний рівень, наприклад МПД.

### 4.3. Класифікація поточних станів технологічних процесів при побудові контрольних карт.

Класифікацію поточних станів технологічних процесів при побудові КК проводять за вибірками невеликого об'єму, розділених певними проміжками часу. Як показали дослідження, для класифікації вибірок недостатньо одного критерію. Це наочно продемонстровано в табл. 4.3, де кількість рівнів  $t=3$  з терм-множиною  $T = \{\text{«низький»}, \text{«середній»}, \text{«високий»}\} = \{H, C, B\}$ , об'єм вибірок  $n=3$ . Важливою характеристикою сукупності вибірок є загальна кількість  $N$  можливих сполучень в вибірках, де  $N = C_t^n = (t+n-1)! / (n!(t-1)!)$  і кількість розрізнених класів еквівалентності  $N_{екв}$  після класифікації. Значення  $N$  характеризує потенційну розрізнявальну здатність при побудові КК, що збільшується при збільшенні  $t$  і  $n$ . Значення  $N_{екв}$  характеризує реальну розрізнявальну здатність для визначення на КК центральної тенденції, попереджувальних і контрольних границь. В табл. 4.3  $N = N_{екв}$ . За рядками табл. 4.3 введені наступні позначення: А – ранжовані вибірки, що відповідають різним станам; В – номер класу еквівалентності за відношенням домінування вибірок; С – номер класу еквівалентності за медіаною та розсіюванням.

Таблиця 4.3 – Класифікація поточних станів за вербальними вибірками

А	ННН	СНН	ВНН	ССН	ВСН	ССС	ВСС	ВВН	ВВС	ВВВ
В	1	2	3		4		5		6	7
С	1	2	3	4	5	6	7	8	9	10

Але при збільшенні  $t$  і  $n$  навіть при використанні всіх операторів частина вибірок залишається нерозрізною і кількість класів еквівалентності стає меншою за кількість можливих сполучень вербальних категорій, тобто  $N_{екв} < N$ .

При збільшенні об'єму вибірок збільшується кількість класів еквівалентності, але частина вибірок залишається нерозрізною одна від одної. Наприклад, при об'ємі вибірки  $n = 4$  і терм-множини  $t = 3$ , тобто  $\{H;C;B\}$  з застосуванням оцінки центральної тенденції за оператором  $OWA$ , отримуємо 3 класи еквівалентності у відповідності з обраними термами (табл. 4.4). А потім класифікуємо їх за відношенням домінування і критерієм розсіювання.

Таблиця 4.4 – Класифікація станів, що відповідають вибіркам при  $n = 4$ ,  $t = 3$

Ранжовані вибірки, що відповідають різним станам	Клас еквівалентності за оператором $OWA$	Номер класу еквівалентності за відношенням домінування і розсіювання	Шкала для визначення рівня якості за вербальними вибірками
НННН	Н	1	Низький
СННН		2	
ВННН		3	
ССНН	С	4	Нижче середнього
ВСНН		5	
СССН		6	
ВВНН		7	Середній
ВССН			
СССС			
ВВСН			
ВССС		8	Вище середнього
ВССС			
ВВСС		10	



Продовження таблиці 4.4			
ВВВН	В	11	Високий
ВВВС		12	
ВВВВ		13	

Але середню частину вибірок складно проранжувати за цими критеріями. Такі вибірки залишаються нерозрізненими. Тобто стан об'єкту за цими вибірками відповідає одному класу еквівалентності. Для розрізнення таких станів в роботі [39] додано ще ранговий критерій, що прийнятий як вирішальний. Тоді кількість класів еквівалентності дорівнює кількості варіантів ранжованих вибірок (для таблиці 4.4 це 15). Але в цьому випадку вже використовується арифметизована рангова шкала при визначенні вихідних терм-множин.

Зважаючи на протиріччя, що складаються при застосуванні багатьох критеріїв, автором роботи було прийнято рішення про обмеження кількості критеріїв при класифікації, а отже про обмеження розрізняваних класів еквівалентності. Тим більше, що така ситуація посилюється при збільшенні об'єму вибірок. Якщо шкала використовується для характеристики рівня якості технологічного процесу, за отриманими вербальними вибірками можуть бути розрізнені наступні класи еквівалентності: «низький», «нижче середнього», «середній», «вище середнього», «високий». Тоді карта технологічного процесу може бути побудована за цими рівнями.

#### **4.4. Побудова контрольної карти і знаходження центральної тенденції груп підземних урн.**

Для визначення стану заповнення декількох урн з метою керування рухом сміттєзбиральної техніки необхідна контрольна карта стану заповнення урн і процедура агрегування даних контрольної карти.

Один із варіантів отриманої вибірки вербальних результатів вимірювання, що використовуються для побудови контрольної карти, наведено в таблицях 4.2, 4.5.

Таблиця 4.5 – Результати оцінювання стану заповнення сміттєвих урн.

№ урни	Стан заповнення	№ урни	Ранжована вибірка стану заповнення
1	ПД	4	МІД
2	МПД	9	МІД
3	МПД	14	МІД
4	МІД	2	МПД
5	БПД	3	МПД
6	БПД	10	МПД
7	МАД	11	МПД
8	БПД	12	МПД
9	МІД	15	МПД
10	МПД	1	ПД
11	МПД	5	БПД
12	МПД	6	БПД
13	МАД	8	БПД
14	МІД	7	МАД
15	МПД	13	МАД

Позначення в табл. 4.5:

- МІД (заповнена) – максимальне значення;
- МПД (більше половини);
- ПД (наполовину заповнена) – середнє значення;
- БПД (менше половини);
- МАД (пуста) – мінімальне значення.

На контрольній карті (рис.4.1) представлені ці результати:

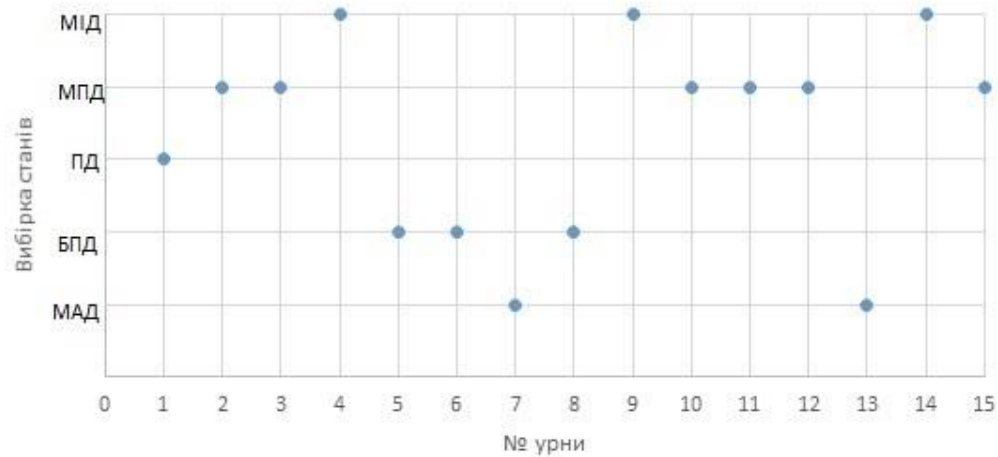


Рисунок 4.1 – Контрольна карта стану заповнення урн.

Для визначення виходу вимірних об'єктів за границю буде використовуватись лише одна межа – нижня контрольна межа, яка буде визначатись замовником аналізу роботи того чи іншого об'єкта. Елементом, який буде виходити за контрольну нижню межу або залишатись в її межах для аналізу об'єкта, є центральна тенденція, яку можна визначити за оператором OWA, або за медіаною.

Оператор OWA, як емулятор середньоарифметичного значення, визначається за наступною формулою:

$$X_{OWA} = \text{Max}_{k=1}^n [\text{Min}\{Q(k), b_k\}] \quad (4.6)$$

де  $b_k$  –  $k$ -й елемент зразка;

$Q(k) = S_{f(k)}$ ,  $k = 1, 2, \dots, n$  – вага оператора OWA з:

$$f(k) = \text{Int}\left[1 + k \frac{t-1}{n}\right]$$

де  $S_{f(k)}$  – це  $f(k)$ -й рівень лінгвістичної шкали,  $\text{Int}(a)$  – функція цілого числа,  $t$  – число рівнів,  $n$  – об'єм вибірки.

Для поточної вибірки ваги операторів OWA наступні:

$$Q(1) = \text{МАД},$$

$$Q(2) = Q(3) = Q(4) = Q(5) = \text{БПД},$$

$$Q(6) = Q(7) = Q(8) = Q(9) = \text{ПД}$$

$$Q(10) = Q(11) = Q(12) = \text{МПД}$$

$$Q(13) = Q(14) = Q(15) = \text{МІД}$$

За формулою  $X_{OWA}$  розраховуємо емулятор середньоарифметичного значення вибірки терм-множин з урахуванням ваг оператора OWA та ранжованих елементів зразка (вибірки) в порядку спадання:

$$\begin{aligned} X_{OWA} &= \text{Max}_{k=1}^n [\text{Min}\{Q(k), b_k\}] = \text{Max}_{k=1}^{15} [\text{Min}\{Q(1), b_1\}, \text{Min}\{Q(2), b_2\}, \text{Min}\{Q(3), b_3\}, \text{Min}\{Q(4), b_4\}] = \\ &= \text{Min}\{Q(5), b_5\}, \text{Min}\{Q(6), b_6\}, \text{Min}\{Q(7), b_7\}, \text{Min}\{Q(8), b_8\}, \text{Min}\{Q(9), b_9\}, \text{Min}\{Q(10), b_{10}\} = \\ &= \text{Min}\{Q(11), b_{11}\}, \text{Min}\{Q(12), b_{12}\}, \text{Min}\{Q(13), b_{13}\}, \text{Min}\{Q(14), b_{14}\}, \text{Min}\{Q(15), b_{15}\} = \text{Min}\{\text{МАД}, \text{МІД}\} = \\ &= \text{Min}\{\text{БПД}, \text{МІД}\}, \text{Min}\{\text{БПД}, \text{МІД}\}, \text{Min}\{\text{БПД}, \text{МПД}\}, \text{Min}\{\text{БПД}, \text{МПД}\}, \text{Min}\{\text{ПД}, \text{МПД}\} = \\ &= \text{Min}\{\text{ПД}, \text{МПД}\}, \text{Min}\{\text{ПД}, \text{МПД}\}, \text{Min}\{\text{ПД}, \text{МПД}\}, \text{Min}\{\text{МПД}, \text{ПД}\}, \text{Min}\{\text{МПД}, \text{БПД}\} = \\ &= \text{Min}\{\text{МПД}, \text{БПД}\}, \text{Min}\{\text{МІД}, \text{БПД}\}, \text{Min}\{\text{МІД}, \text{МАД}\}, \text{Min}\{\text{МІД}, \text{МАД}\} = \text{Max}\{\text{МАД}, \text{БПД}, \text{БПД}\} = \\ &= \text{БПД}, \text{БПД}, \text{ПД}, \text{ПД}, \text{ПД}, \text{ПД}, \text{ПД}, \text{БПД}, \text{БПД}, \text{БПД}, \text{МАД}, \text{МАД}\} = \text{ПД}. \end{aligned}$$

Отримане за наведеною вибіркою (табл.5.3)  $X_{OWA}$  – ПД (наполовину заповнена).

При непарній кількості елементів в вибірці результатів вимірювань можна скористатись медіаною вибірки [5].

$$X_{med} = X_{\left(\frac{n+1}{2}\right)}$$

Скориставшись четвертим стовпцем таблиці 1 визначаємо, що  $X_{med} = \text{МПД}$ .

Позначаємо  $X_{OWA}$  і  $X_{med}$  на контрольній карті (рис.4.2):

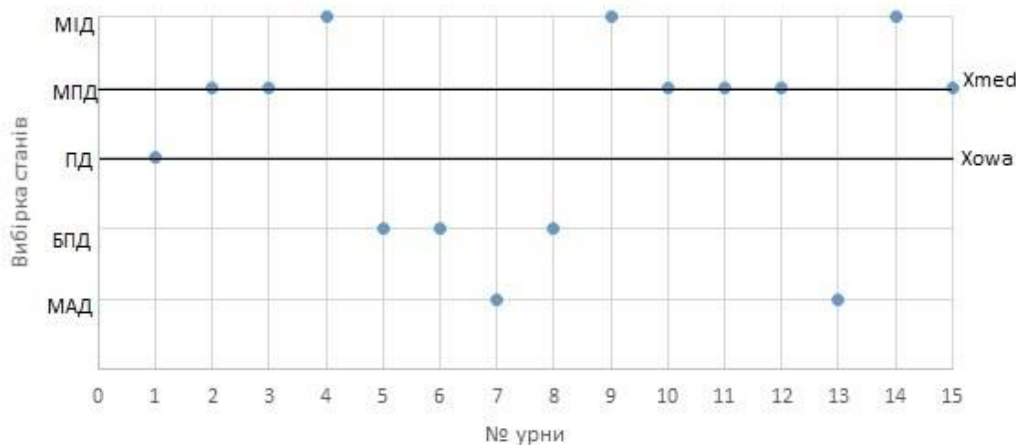


Рисунок 4.2 – Контрольна карта стану заповнення урн.

Визначення збіжності отриманих за формулами оцінок центральної тенденції було проведено в підрозділі 4.2.

Рішення про виклик сміттезбиральної техніки приймаємо, якщо  $X_{med}$  або  $X_{OWA}$  дорівнює або перевищує контрольний рівень, наприклад МПД.

Контрольні карти рис.4.1, 4.2 використовувались для оптимізації розміщення смітєвих урн. Наявність на карті протилежних станів МІД і МАД свідчить про неоптимальність розміщення деяких урн.

#### **4.5. Метод класифікації стану об'єкта або технологічного процесу за вибірками вербальних даних при побудові контрольних карт.**

В попередніх підрозділах розглянуто ситуацію, коли шкала класифікації будується заздалегідь за умов розрізнення окремих вибірок як класів еквівалентності. Але при збільшенні  $t$  і  $n$  кількість можливих сполучень вербальних категорій стає занадто великою.

Велика кількість класів еквівалентності приводить до недоцільності побудови загальної шкали класифікації. Тому вербальні вибірки класифікуються в процесі побудови контрольної карти. При цьому послідовність застосування операторів зворотна по відношенню до табл.4.1, а саме:

- вибірки впорядковуються за спаданням властивості;
- вибірки розподіляються на класи еквівалентності за центральною тенденцією (використовується або медіана, або  $OWA$ , або обидві оцінки з визначенням проміжних класів еквівалентності);
- вибірки в кожному класі еквівалентності ранжуються за відношенням домінування і розсіювання.

Вибірки, що залишились нерозрізненими, належать до одного класу еквівалентності. За отриманими класами еквівалентності може бути побудована контрольна карта.

Результати класифікації 31 вибірки при  $t = 3$ ;  $n = 5$  наведено в табл. 4.6.

Таблиця 4.6 – Класифікація вибірок для побудови контрольної карти

Класифікація за центром вибірки – медіаною	Класифікація за відношенням домінування і розсіюванням	Номер класу еквівалентності	Номери впорядкованих вибірок для непараметричного оцінювання
Клас Н	ССННН	1	1
Клас С	СССНН	2	2,3,4,5
	ССССН	3	6,7,8,9
	ССССС	4	10
	ВСССН		11,12,13,14,15
	ВВССН	5	16,17,18,19,20
Клас В	ВВВНН		21
	ВВВСН	6	22,23,24
	ВВВСС	7	25,26,27,28
	ВВВВС	8	29,30,31

Результати розрахунків за даними таблиці 4.6 для побудови контрольної карти:  $med = x'_{16} = (ВВССН); ; \quad НПГ = x'_{10} = (ССССС); ; \quad ВПГ = x'_{22} = (ВВВСН); \quad НКГ = x'_7 = (ССССН);$   
 $ВКГ = x'_{25} = (ВВВСС)$ . Вибірki на рис.4.3 розташовують за лініями, які відповідають 8-ми класам еквівалентності. Якщо класів еквівалентності недостатньо для розрізнення центральної тенденції, попереджувальних та контрольних границь необхідно збільшити або кількість градацій  $t$  або об'єм вибірки  $n$ .

Таким чином, розроблено технології встановлення відношення порядку між вербальними вибірками малого об'єму для побудови шкали і метод класифікації вербальних вибірок за рівнем якості технологічного процесу на основі використання критерію домінування, оцінок центральної тенденції вибірки і розсіювання, який може

бути використаний при побудові КК технологічних процесів за якісними характеристиками.

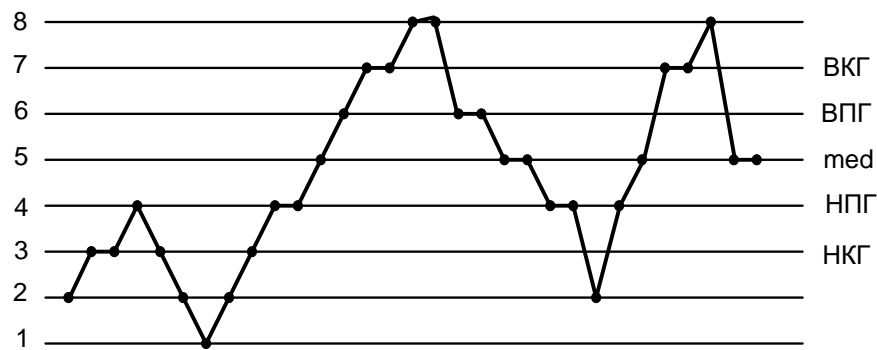


Рисунок 4.3 – Контрольна карта за даними таблиці 4.6

#### Висновки за розділом 4.

В розділі розглянуто питання опрацювання вербальних вибірок з метою визначення шкали класифікації станів об'єкту або рівня якості при побудові контрольних карт. Для визначення класів еквівалентності шкали були застосовані наступні критерії: критерій визначення центральної тенденції (за медіаною або оператором  $OWA$ ), відношення домінування між вербальними вибірками і характеристики розсіювання вибірки. Вибірки, що залишаються нерозрізненими за вищеназваними критеріями і операторами, об'єднуються в один клас еквівалентності, що відповідає певному стану об'єкта або рівню якості. Отримана шкала класифікації може бути використана для побудови контрольних карт за вербальними даними.

З метою класифікації стану об'єкту за центральною тенденцією експертних оцінок його властивостей розроблено метод опрацювання вербальних вибірок з обчисленням двох оцінок центральної тенденції, за яким при їх нееквівалентності визначається додаткові (проміжні) класи еквівалентності стану об'єкта.

Розроблено метод класифікації стану об'єкта або технологічного процесу за вербальними даними, що ґрунтується на технології використання декількох критеріїв і оцінок і дозволяє вирішити задачу класифікації стану з використанням або

встановлених класів еквівалентності або з додатковими (проміжними), який може бути використаний при побудові контрольних карт за якісними ознаками.



## ВИСНОВКИ

Дисертаційна робота містить нові науково обгрунтовані результати застосування шкал класифікації при опрацюванні вимірювальної та експертної інформації. Висновки, що узагальнюють отримані наукові та практичні результати, полягають у наступному:

1. Розроблено метод непараметричної класифікації видів розподілу вибірок малого об'єму, що дозволяє провести попередню класифікацію вимірювальної інформації та вибрати ефективну оцінку центру розподілу і тим самим підвищити точність результату вимірювання.

2. Вперше застосовано метод непараметричної класифікації виду розподілу при побудові контрольних карт технологічно процесу, що дозволяє збільшити чутливість до контрольованих змін центральної тенденції і визначити клас сумарного розподілу вибірок при нестабільності технологічного процесу.

3. Набули подальшого розвитку технології побудови лінгвістичних шкал в інтелектуальних вимірювальних системах для переходу від числових даних до вербальних в напрямку об'єднання вимірювальної та експертної інформації із застосуванням нечіткої метричної класифікації, що дозволило провести адекватне врахування невизначеності при формуванні функцій приналежності терм-множини шкали і отримати характеристику якості встановленої шкали класифікації у вигляді матриці відповідності.

4. Вперше розроблено метод визначення узгодженості вибірок класифікованих даних, заснований на непараметричних оцінках центру вибірки і шкалі метричної класифікації, що дозволяє провести оцінку узгодженості при нерівномірному розташуванні класів еквівалентності в області їх визначення.

5. З метою класифікації стану об'єкту за центральною тенденцією експертних оцінок його властивостей розроблено метод опрацювання вербальних вибірок з обчисленням двох оцінок центральної тенденції, за яким при їх нееквівалентності визначається додаткові (проміжні) класи еквівалентності стану об'єкта.

6. Розроблено метод класифікації стану об'єкта або технологічного процесу за вербальними даними, що ґрунтується на технології використання декількох критеріїв і оцінок і дозволяє вирішити задачу класифікації стану з використанням або встановлених класів еквівалентності або з додатковими (проміжними), це підвищує чутливість контрольних карт за якісними ознаками.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кендалл М., Стюарт А. Многомерный статистический анализ и временные ряды. Москва : Наука, 1976. 736 с.
2. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика : Классификации и снижение размерности. Москва : Финансы и статистика, 1989. 607 с.
3. Розенберг В. Я. Введение в теорию точности измерительных систем. Москва : Советское радио, 1975. 394 с.
4. Орлов А. И. Некоторые вероятностные вопросы теории классификации : зб. Прикладная статистика : Ученые записки по статистике. Москва : Наука, 1983. Т. 45. С. 166-179.
5. Новицкий П. В., Зограф И. А. Оценка погрешностей результатов измерений. Ленинград : Энергоатомиздат, 1985. 248 с.
6. Кликушин Ю. Н. Представление случайных сигналов с помощью принадлежностных спектров. *Журнал радиоэлектроники / Омский государственный технический университет*. 2000. № 2. URL: <http://jre.cplire.ru/jre/feb00/4/text.html>.
7. Королюк В. С., Портенко Н. И., Скороход А. В., Турбин А. Ф. Справочник по теории вероятностей и математической статистике. Москва : Наука, 1985. 640 с.
8. Смирнов Н. В., Дунин-Барковский Н. В. Курс теории вероятностей и математической статистики для технических приложений. Москва : Наука, 1969. 511 с.
9. Дорожовець М. М. Опрацювання результатів вимірювань : навч. посіб. Львів : Вид-во Національного університету «Львівська політехніка», 2007. 624 с.
10. Грановський В. А., Сирая Т. Н. Методы обработки экспериментальных данных при измерениях. Ленинград : Энергоатомиздат, 1990. 288 с.
11. Основи метрології та вимірювальної техніки : у 2 т. : навч. посіб. / В. Д. Ціделко та ін. К. : НТУУ «КПІ», 2013. Т. 1. 236 с.
12. Martinez J., Iglewicz B. Some properties of the Tukey g- and h-family of distributions. *Communications in Statistics – Theory and Methods*. 1984. No. 13. P. 353-359.

13. Химмельблау Д. Анализ процессов статистическими методами. Москва : Мир, 1973. 957 с.
14. ДСТУ ISO 8258:2001. Статистичний контроль. Контрольні карти Шухарта (ISO 8258:1991, IDT) [Чинний від 2003-07-01]. К. : Держспоживстандарт України, 2003. 32 с.
15. ISO 22514-1. Statistical methods in process management – Capability and performance. ISO-2014.
16. Ціделко В. Д., Яремчук Н. А., Невизначеність вимірювання : Обробка даних і подання результату вимірювання : Монографія. К. : ІВЦ Вид-во «Політехніка», 2002. 176 с.
17. Roberts F. S. On the indicator function of the plurality function. *Mathematical Social Sciences*. 1991. No 22. P. 163-174.
18. Яремчук Н. А. Інтелектуальні засоби вимірювальної техніки : навч. пос. : Том 1. Методологія інтелектуальних засобів вимірювальної техніки. К. : Корнійчук, 2017. 208 с.
19. Конышева Л. К., Незаров Д. М. Основы теории нечетких множеств : учеб. пособие. СПб. : Питер, 2011. 192 с.
20. Кофман А. Введение в теорию нечетких множеств / пер. с франц. Москва : Радио и связь, 1982. 432 с.
21. Семенюк Р. С. Способы определения терм-множества лингвистической переменной с неопределенностью измерений. *Неопределенность измерений: научные, нормативные и методические аспекты UM-2018* : зб. тез 15-ї міжнар. наук.-техн. сем., г.Созополь, 10 сентября 2018 р. С. 206-207.
22. Пегат А. Нечеткое моделирование и управление / пер. с англ. Москва : БИНОМ. Лаборатория знаний, 2009. 798 с.
23. Танаки К. Итоги рассмотрения факторов неопределенности и неясности в инженерном искусстве. *Нечеткие множества и теория возможностей. Последние достижения* / пер. с англ. ; под ред. Р.Р. Ягера. Москва : Радио и связь, 1986. 408 с.

24. Орнатський П. П. Вступ до методології науки про вимірювання : навч. пос. К. : ІСЛЮ. 180 с.
25. Гудмэн И. Нечеткие множества как классы эквивалентности случайных множеств. *Нечеткие множества и теория возможностей. Последние достижения* / пер. с англ. ; под ред. Р.Р. Ягера. Москва : Радио и связь, 1986. 408 с.
26. Grabisch M., Nguyen H. T., Walker E. A. Fundamentals of uncertainty calculi with applications to fuzzy inference. Dordrecht, The Netherlands : Kluwer Academic Publishers, 1995.
27. Torra V. Aggregation of linguistic labels when semantics is based on antonyms. *International Journal of Intelligent Systems*. 2001. No. 16. P. 513-524.
28. Torra V. Negation functions based semantics for ordered linguistic labels. *International Journal of Intelligent Systems*. 1996. No. 11. P. 975-988.
29. ICGM 200:2008. International vocabulary of metrology – Basic and general concepts and associated terms (VIM). ICGM 2008.
30. Саати Томас Л. Принятие решений при зависимостях и обратных связях : Аналитические сети / пер. с англ. ; науч. ред. А. В. Андрейчиков, О.Н. Андрейчикова. 3-е изд. Москва : Книжный дом «ЛИБРОКОМ», 2011. 360 с.
31. Федин С. С. Оценка и прогнозирование качества промышленной продукции с использованием адаптивных систем искусственного интеллекта : монография. К. : Интерсервис, 2012. 206 с.
32. Рабинович В. И., Цапенко М. П. О количестве измерительной информации. *Измерительная техника*. 1963. № 4. С. 7-11 ; № 6, С. 1-6 ; № 10, С. 8-13.
33. Tastle W. I., Wierman M. I. An information theoretic measure for the evaluation of ordinal scale data. *Behavior Research Methods*. 2006. No. 38(3). P. 487-494.
34. Bashansky E., Gadrich T. Evaluating Quality Measured on a Ternary Ordinal Scale. *Quality and Reliability Engineering International*. 2008. No. 24. P. 957-971.

35. Flynn, D., van Schaik, P., van Wersch, A. A comparison of multi-item likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment*. 2004. No 20(1). P. 49-58.
36. Солопченко Г. Н. Представление измеряемых величин и погрешностей измерений как нечетких переменных. *Измерительная техника*. 2007. № 2. С. 3-7.
37. Борисов А. Н. Модели принятия решений на основе лингвистической переменной. Рига : Зинатне, 1982. 256 с.
38. Соболев В. И. Информационно-статистическая теория измерений : учебник для вузов. Москва : Машиностроение, 1983. 224 с.
39. Franceschini F., Galetto M., Varetto M. Qualitative ordinal scales : The concept of ordinal range. *Quality Engineering*. 2004. No. 16(4). P. 515-524.
40. Blair I., Lacy M. G. Statistics of ordinal variation. *Sociological Methods and Research*. 2000. No. 28. P. 251-280.
41. ДСТУ 4362:2004. Якість ґрунту. Показники родючості ґрунтів. [Чинний від 2006-01-01]. Київ : Держспоживстандарт України, 2005. 20 с.
42. ДСТУ 4114-2002. Ґрунти. Визначення рухомих сполук форфору і калію за модифікованим методом Мачигіна. [Чинний від 2003-01-01]. Київ : Держстандарт України, 2002. 10 с.
43. Мірошніченко М. М., Гладкіх Є. Ю., Ревтьє А. В. Особливості відбору проб ґрунту за локального внесення мінеральних добрив. *Агроекологічний журнал*. 2016. № 4. С. 75-80.
44. Franceshini F., Galetto M., Varetto M. Orderer samples control charts for ordinal variables. *Quality and reliability engineering international*. 2005. No. 21. P. 177-195.
45. Кнорринг В. Г. Метрология и репрезентационная теория измерений. *Измерительная техника*. 1986. № 1. С. 6-7.
46. Хованов Н. З. Математические основы теории шкал измерения качества. Ленинград, 1982. 169 с.

47. Bashkansky E., Gadrich T. Mathematical and computational aspects of treatment ordinal measurement results. *Advanced Mathematical And Computational Tools In Metrology And Testing IX*. 2012. P. 35-42.
48. Гнатієнко Г. М., Снитюк В. Є. Експертні технології прийняття рішень : монографія. К. : ТОВ «Маклаут», 2008. 444 с.
49. Cecconi P., Franceschini F., Galetto M., Measurements, evaluations and preferences. *A scheme of classification according to the representational theory, Measurement*. 2006. Vol. 39, Issue 1. P.1-11.
50. Yager R. R. Quantifier guided aggregation using OWA operators. *International Journal of Intelligent Systems*. 1996. No. 11. P. 49-73.
51. Yager R. R. On ordered weighted averaging aggregation operators in multi-criteria decisionmaking. *IEEE Trans Syst Man Cybern*. 1988. P. 183-190.
52. Domingo-Ferrer J., Torra V. Median-based aggregation operators for prototype construction in ordinal scales. *International Journal of Intelligent Systems*. No. 18. 2003. P. 633-655.
53. Herrera F., Herrera-Viedma E. Aggregation operators for linguistic weighted information. *IEEE Transactions on Systems, Man, and Cybernetics*. 1997. No. 27. P. 646-656.
54. Сікоза О. М., Яремчук Н. А., Обчислення невизначеності при експертному оцінюванні вагових коефіцієнтів. *Система обробки інформації*. 2011. № 1(91). С. 48.
55. Бугреєв С. С., Семенюк Р. С., Яремчук Н. А. Вимірювання рівня сміття в підземних урнах лазерними рівнемірами. *Український метрологічний журнал*. 2017. № 3. С. 36-39.

## СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

*Статті у періодичних наукових виданнях інших держав, які входять до ОЕСР:*

1. Semeniuk R.S., Yaremchuk N.A. The methods of obtaining and elaborating classified data in intellectual measurement systems : *Central European Researchers Journal (CERES Journal)*, Volume 6, Issue 1, 2020. С.27-36 (видання держави, що входить до ЄС (Словаччина); посилання на збірник: <http://ceres-journal.eu/iss200601>).

*Статті у фахових виданнях:*

2. Бугреєв С.С., Семенюк Р.С., Яремчук Н.А. Вимірювання рівня сміття в підземних урнах лазерними рівнемірами : *Український метрологічний журнал*. № 3/2017. С. 36-39.

3. Яремчук Н.А., Семенюк Р.С. Опрацювання вербальних даних за побудови контрольних карт : науково-виробничий журнал *Метрологія і прилади*, № 5-1 (67), 2017. С. 58-61.

4. Семенюк Р.С., Павлишин М.М. Використання методів нечіткої класифікації для оптимізації кількості проб ґрунту в задачах агромоніторингу. *Техніко-технологічні аспекти розвитку та випробування нової техніки і технологій для сільського господарства України* : зб. наук. праць, 2018 р., випуск 22(36). С. 264-267.

5. Семенюк Р., Яремчук Н., Гусар І. Спосіб перевірки узгодженості класифікованих даних під час оцінювання показників якості ґрунту. *Техніко-технологічні аспекти розвитку та випробування нової техніки і технологій для сільського господарства України* : зб. наук. праць, 2019 р., випуск 25(39). С.151-159.

6. Семенюк Р.С. Яремчук Н.А. Використання нечіткої класифікації виду розподілу для вибірок малого об'єму. *Інформаційні системи, механіка та керування* : зб. наук. праць, 2017 р., випуск 17. С. 40-50.

*Матеріали коференцій:*

7. Яремчук Н.А., Семенюк Р.С. Використання нечіткої класифікації при оцінюванні точності і стабільності технологічних процесів. *Гіротехнології, навігація, керування*



*рухом і конструювання авіаційно-космічної техніки* : зб. тез XI міжнар. наук.-техн. конф., м. Київ, 13-14 квітня 2017 р. С. 91-95.

8. Семенюк Р.С. Опрацювання результатів вимірювання з використанням попередньої нечіткої класифікації. *Метрологія, інформаційно-вимірювальні технології та системи* : зб. тез VI міжнар. наук.-техн. конф., м.Київ, 24-25 жовтня 2017 р. С. 129-131.

9. Семенюк Р.С., Яремчук Н.А. Застосування класифікації розподілів вибірки при побудові контрольних карт. *Приладобудування та метрологія: сучасні проблеми, тенденції розвитку* : зб. тез III всеукр. наук.-практ. конф., м.Луцьк, 11-12 жовтня 2018 р. С. 55-57.

10. Семенюк Р.С. Удосконалення побудови контрольних карт технологічних процесів з застосуванням попередньої класифікації. *Мехатронні системи: інновації та інжиніринг* : зб. тез II міжнар. наук.-практ. конф., м.Київ, 15 червня 2018 р. С. 46-47.

11. Семенюк Р.С. Способы определения терм-множества лингвистической переменной с неопределенностью измерений. *Неопределенность измерений: научные, нормативные и методические аспекты UM-2018* : зб. тез 15-ї міжнар. наук.-техн. сем., г.Созополь, 10 сентября 2018 р. С. 206-207.

12. Семенюк Р.С., Яремчук Н.А. Основні етапи встановлення лінгвістичних шкал при вимірюваннях і діагностиці. *Метрологія, інформаційно-вимірювальні технології та системи* : зб. тез VII міжнар. наук.-техн. конф., 18-19 лютого 2020 р. С. 131-133.

13. Яремчук Н.А., Семенюк Р.С. Способы учета неопределенности при построении лингвистических шкал. *Метрологія та вимірювальна техніка (Метрологія–2020)* : зб. доп. XII міжнар. наук.–техн. конф. 6–8 жовтня 2020 р. – Х.: ННЦ Інститут метрології, 2020. – С. 77-81.

14. Ступак С.С., Яремчук Н.А., Семенюк Р.С. Способы учета влияния неопределенности на результаты «мягких» измерений. *Актуальні проблеми автоматики та приладобудування* : матеріали 2-ї міжнар. наук.-техн. конф., м.Харків, 06-07 грудня 2018 р. С. 210-211.

**Особистий внесок здобувача.** Всі результати теоретичного аналізу та експериментальних досліджень дисертаційної роботи, висновки та рекомендації, які винесені до захисту, одержані автором особисто. В спільних публікаціях автором зроблено наступне: [1, 2] – дослідження використання непараметричної класифікації виду розподілу для вибірок малого об'єму, [6] - застосування методів непараметричної класифікації для оптимізації кількості проб ґрунту в задачах агромоніторингу, [4, 7] – дослідження використання контрольних карт технологічних процесів при опрацюванні даних, [12] – застосування лінгвістичних шкал при вимірюваннях і діагностиці, [3] – дослідження використання вербальної шкали при розробці вимірювача сміття в підземній урні, [13] – дослідження способів урахування інструментальної невизначеності при формуванні результатів м'яких вимірювань, [14] - дослідження способів побудови шкали класифікації, за якої отримують ступені приналежності результатів вимірювання окремих ознак встановленим класам еквівалентності.

## ДОДАТОК Б

## Моделювання процедури оцінювання достовірності непараметричної класифікації

Таблиця 1 – Вихідні вибірки, отриманні при генерації випадкових чисел за нормальним розподілом

Номер вибірки	1	2	3	4	5	6	7	8	9
№1	-1,30427	-1,49086	0,20291	0,628734	0,260707	0,729884	-2,08547	-0,44368	-0,32104
№2	-0,29822	0,43382	-0,16569	0,569396	-0,79652	0,930811	0,673441	0,303248	-0,84697
№3	0,748698	1,768258	0,064575	-0,33807	-0,20905	0,192303	2,702157	-0,30627	-1,26773
№4	-1,7966	-0,09545	0,424673	-0,10187	-0,33052	0,216842	0,652377	-1,60278	-1,14045
№5	1,723654	1,544291	0,827043	1,859537	-1,37554	-1,98582	1,577151	-0,29424	0,359216
№6	-0,52851	2,503668	0,247747	0,836541	-1,38446	-0,36844	0,69184	-0,28435	0,275815
№7	-0,5994	-1,31217	1,460439	-0,62548	0,140457	0,812882	0,883101	1,419493	-0,14382
№8	1,661209	0,349804	0,501577	-1,32389	0,221263	-0,47901	0,619084	-0,41163	-1,34146
№9	-0,20462	-0,48706	-0,51664	-0,14352	-0,36933	1,629867	0,166189	-0,96092	-0,0398
№10	-2,49964	-0,99765	2,410532	-0,25519	0,521995	-0,97176	0,67678	-1,53332	0,459438
№11	-0,44569	0,462329	-0,40026	-0,03402	-1,45864	0,973688	0,745373	1,078184	0,689834
№12	1,015566	2,746462	1,000884	-0,40691	0,679624	2,461567	0,033164	1,152879	2,917201
№13	0,558308	1,07247	1,506879	0,208881	0,496265	0,6414	-0,02672	-0,18496	-0,037
№14	-0,37207	-0,1358	0,779368	1,883159	1,119019	0,379418	-0,4341	-1,95256	-0,12492
№15	0,075518	-0,1691	-1,78054	0,216452	-1,23174	0,280967	-0,34201	-0,38502	0,637457
№16	1,496916	-0,58841	-0,49974	0,473203	-1,0368	1,80357	-0,736	-0,50872	0,068619
№17	-0,28896	0,446203	-0,91484	0,794603	-1,85377	-0,32669	-0,92964	-1,65892	-0,81736
№18	0,906852	0,343665	-1,35627	-1,07916	-0,59573	0,893995	-1,2897	0,105501	-0,0164
№19	-0,20213	2,412723	0,331584	1,662653	0,867514	-1,04876	0,693812	1,053982	0,127998
№20	-1,09344	-1,44322	0,071481	-0,14203	0,90791	-0,71726	2,0456	0,199177	0,578492
№21	0,904314	-0,86342	0,4766	0,613184	1,866085	-0,48573	0,369783	0,134893	0,367289
№22	-1,5006	-1,73989	-1,53731	1,171906	1,583633	0,348574	0,857586	0,569942	-0,55565
№23	0,255133	-0,86019	-1,58541	-0,15405	-0,3131	-1,81364	0,73422	0,475809	0,896157
№24	2,334688	-1,46174	0,008539	-1,12094	-0,31143	0,25361	0,911409	-0,18597	-0,60677
№25	-0,52052	-0,54165	-0,74141	0,074353	0,337359	1,202123	0,690598	0,232119	-0,61148
№26	-0,27343	0,387673	-0,91909	0,602786	-1,40685	-0,85449	1,234954	-0,87287	-1,22176
№27	2,004166	0,091728	0,337604	-0,48128	0,575301	0,441478	-0,70958	-2,21335	-0,78453
№28	0,839556	1,215978	-0,84937	-0,65756	1,281245	-0,47309	-0,30995	-1,02584	-0,6227
№29	1,151044	-1,28777	1,676542	1,878203	0,050657	-0,24123	0,985516	0,210519	1,441019
№30	-0,28593	1,099007	1,128041	-0,41039	-1,09655	-0,17468	0,663684	-0,84142	-0,40929

Продовження табл. 1

Номер вибірки	1	2	3	4	5	6	7	8	9
№31	0,080928	0,669183	1,356735	0,318165	-0,40468	-0,42938	-0,60575	-0,019	-1,04365
№32	-0,55257	-0,37794	1,04969	-1,1744	-1,23917	0,181357	0,928433	0,528332	1,072307
№33	-1,26525	0,08791	-1,11073	-0,96453	0,209037	-0,38001	-1,3631	0,843861	0,267186
№34	-0,31243	0,228166	0,31976	0,529329	0,219745	-0,30615	0,066793	0,533313	1,069083
№35	-0,22118	1,001063	0,371714	0,876722	1,153332	-0,06949	-0,7184	-0,47063	0,277739
№36	0,077741	-2,35078	-0,17511	0,738645	-0,27002	1,8029	0,339643	0,15605	0,785327
№37	-1,58415	-0,30823	-0,20045	-0,28653	0,406431	0,005666	2,048036	-1,26919	-0,34266
№38	0,284994	0,359549	0,394743	-2,13788	0,201026	-0,84378	-1,63559	0,539965	-0,24369
№39	0,939701	-0,20034	-0,0123	-0,3239	0,391017	-0,72957	1,068435	0,317677	-0,96104
№40	1,838417	-0,53246	1,349756	-0,54229	-0,89422	-0,06132	0,070723	-1,90911	1,553387
№41	-1,23436	1,274586	-0,54184	-0,64602	0,351888	-0,86804	1,311865	-0,5264	-0,94583
№42	-0,58774	0,799907	-0,85782	-0,91383	0,862724	-0,52571	1,650441	-1,20593	-0,88801
№43	1,304959	0,580929	0,721667	1,215632	-0,28213	-0,00282	0,737244	-0,88903	0,346307
№44	-0,46409	0,405951	0,416301	-0,42814	0,337541	0,951067	1,065192	-0,79778	0,032067
№45	1,524169	-0,32499	-0,21575	-0,15818	-0,04703	1,364651	0,548536	-1,3159	-0,68211
№46	0,203749	-0,09683	1,680916	0,207478	-1,45215	0,588068	-0,2097	-1,36907	0,902411
№47	-0,26371	0,727511	1,463133	-0,66768	1,029818	0,907839	0,241024	0,916177	0,014277
№48	-0,90091	0,677526	0,39096	-1,01534	2,053233	-0,3764	-0,038	0,626876	-0,97756
№49	0,301369	0,999508	-0,62168	1,038809	-0,27421	-0,0307	0,216528	0,489132	0,042762
№50	-1,48771	-0,70538	0,587945	1,066727	1,663474	1,167907	0,486981	0,148405	-0,09241
№51	-0,47145	0,798071	-0,51549	0,949842	0,294695	-0,1175	0,884651	0,034626	-1,86167
№52	2,055975	-1,93177	1,194094	0,496005	-0,89188	1,200498	0,709579	0,451766	0,801832
№53	-0,36968	0,973666	1,012523	0,564112	-1,29124	0,268203	0,341005	-0,35879	-0,90232
№54	1,661168	0,548888	1,303838	-0,9512	0,446943	0,458985	-0,12314	0,648605	1,066123
№55	1,241149	1,242665	-0,52306	0,176109	-0,57604	-1,65693	0,348929	0,038886	-1,31559
№56	-1,19535	-0,54532	-0,41062	-0,96909	-0,90055	-0,12784	-0,58701	1,325801	-0,72931
№57	-0,92354	1,175748	1,125939	-0,93888	1,231451	-0,40644	0,338302	0,237412	0,023951
№58	-1,25394	-1,22638	1,2509	-1,44478	-1,12977	0,206346	0,470362	0,577459	-1,4663
№59	-1,47258	-1,17512	-0,81479	-0,76086	-2,20585	0,255106	-0,84808	0,619939	0,001359
№60	-0,33759	1,821534	-0,2715	-1,26502	0,047698	0,837155	2,196931	-0,35207	0,4037
№61	0,70808	-0,82455	-1,13744	0,661818	-1,33257	-0,90527	-0,69267	1,400359	-1,16712
№62	-2,01689	-0,83915	-0,64391	0,622161	-0,41384	1,535605	1,085569	1,875383	0,204602
№63	0,85173	-0,43645	-1,6716	0,759665	-0,62467	-0,13988	1,287968	-0,87001	1,231418
№64	-0,59497	0,505948	-0,36977	0,47288	-1,62956	-0,54586	0,493949	0,307914	-0,71968
№65	-0,78769	-1,24758	-0,14255	-0,24376	1,063171	-1,62364	-0,81164	-0,49174	1,985319

Продовження табл. 1

Номер вибірки	1	2	3	4	5	6	7	8	9
№66	-0,58659	-0,33673	0,687548	-0,02083	0,714614	0,831839	0,989987	1,104563	0,179226
№67	-0,56281	0,054924	1,567391	-0,68715	0,440672	0,552338	-0,44111	-0,22801	0,761934
№68	1,608231	0,203024	-0,72625	0,183767	-0,9388	1,007819	-0,57149	-0,87386	-0,69798
№69	1,000403	0,32434	-2,1862	2,106543	-1,15232	-0,28519	2,674733	0,864474	-0,41931
№70	0,706819	0,700919	-0,04415	0,590978	0,268346	-0,56271	0,19672	1,067196	-1,63073
№71	0,749581	-1,27489	0,133269	0,862954	-1,27376	-0,7717	-2,83161	0,14263	2,705987
№72	1,323892	-0,57498	0,74181	-0,32261	1,384316	-1,30966	-1,26115	-0,38862	1,448972
№73	-0,52189	1,077091	0,459124	0,362972	0,551539	-0,65993	0,477002	0,311809	-1,57517
№74	-0,13804	-1,03427	0,542153	-2,50232	-0,05323	0,080534	-1,44022	-1,63422	1,900694
№75	0,200522	0,133961	-0,89472	0,860054	-1,11185	0,662876	-0,45974	-0,11143	-0,56244
№76	0,342171	-2,20964	-2,10835	-0,05002	-1,02481	-0,07534	0,528052	-1,27192	-1,24268
№77	-0,64351	0,913748	-0,80467	0,742245	1,02759	-0,04472	-0,20426	-0,16803	1,148151
№78	-1,51546	1,108225	-0,8547	-1,41641	0,498853	0,565218	-1,6462	0,588339	-0,82965
№79	-0,03669	-1,25717	-1,70421	-0,5312	1,156154	-0,7775	-0,29735	-1,04243	0,232115
№80	-0,2977	1,903248	-2,30256	-0,81978	-0,22126	-0,43441	-0,23348	2,450658	-0,70052
№81	-1,07761	0,15186	-0,44284	0,55345	-1,6033	1,166045	0,757594	0,197408	0,238528
№82	0,615144	-1,4128	0,320147	-0,36685	-0,04364	1,261954	1,79747	-0,10653	0,591693
№83	-1,69818	0,254471	1,388357	-0,61311	-0,06187	-0,4697	-0,12041	1,22017	-0,68824
№84	1,231538	2,487735	-1,29739	1,66724	-0,03224	-1,55376	1,262817	0,161214	0,394035
№85	0,049492	-0,10903	-0,01961	-0,89924	-0,50734	0,979015	0,878504	-0,86613	-3,29277
№86	-0,93943	1,587095	0,995545	-0,07411	-0,17668	0,812995	0,860439	-1,35838	0,648139
№87	-0,97249	-0,64273	-0,10945	-0,05885	-0,25159	0,485933	-0,98957	0,599614	0,65685
№88	-0,45801	0,104316	0,303415	-1,82194	0,989689	0,057195	-0,97047	-0,64065	0,395836
№89	1,054466	1,348432	-0,011	1,636921	-0,47683	-1,57207	-0,33111	-1,27365	-1,54246
№90	-0,64315	-1,22032	1,473169	-0,34177	0,308003	1,328846	2,069919	0,303262	0,480372
№91	-0,94934	0,818808	0,359668	0,611513	0,709245	0,395474	-0,47641	0,042081	-0,23342
№92	-0,13709	1,687859	-0,28873	1,701858	-1,33477	1,563581	0,758647	-0,89181	1,136509
№93	-1,59376	0,304772	-0,17519	1,010675	0,874221	-0,14267	-0,14304	0,669859	-0,09116
№94	0,216539	0,173429	-0,24562	-0,50369	-0,88042	-0,62335	-1,34453	-0,53426	-0,1729
№95	1,21018	0,021583	1,129576	0,190844	0,521576	-0,82022	0,112071	-1,76245	1,190258
№96	-1,69177	-0,21797	-0,14094	0,557041	-0,01705	3,162216	0,568615	1,253433	-4,16873
№97	1,103368	0,858219	0,552009	-0,6095	-1,34277	2,795349	0,756399	1,255537	-0,42656
№98	-0,03625	0,309644	-0,94828	-1,20894	-0,96465	0,229558	2,015325	2,206009	1,090485
№99	0,506098	-0,06277	1,066335	-0,4918	-0,73826	0,191561	0,228625	1,369818	-0,3724
№100	-0,3179	0,191898	0,457306	-3,1622	0,225446	1,091386	-0,08304	1,522047	-1,44822

Таблиця 2 - Ранжовані вибірки

Номер вибірки	1	2	3	4	5	6	7	8	9
№1	-1,30427	-1,49086	0,20291	0,628734	0,260707	0,729884	-2,08547	-0,44368	-0,32104
№2	-0,29822	0,43382	-0,16569	0,569396	-0,79652	0,930811	0,673441	0,303248	-0,84697
№3	0,748698	1,768258	0,064575	-0,33807	-0,20905	0,192303	2,702157	-0,30627	-1,26773
№4	-1,7966	-0,09545	0,424673	-0,10187	-0,33052	0,216842	0,652377	-1,60278	-1,14045
№5	1,723654	1,544291	0,827043	1,859537	-1,37554	-1,98582	1,577151	-0,29424	0,359216
№6	-0,52851	2,503668	0,247747	0,836541	-1,38446	-0,36844	0,69184	-0,28435	0,275815
№7	-0,5994	-1,31217	1,460439	-0,62548	0,140457	0,812882	0,883101	1,419493	-0,14382
№8	1,661209	0,349804	0,501577	-1,32389	0,221263	-0,47901	0,619084	-0,41163	-1,34146
№9	-0,20462	-0,48706	-0,51664	-0,14352	-0,36933	1,629867	0,166189	-0,96092	-0,0398
№10	-2,49964	-0,99765	2,410532	-0,25519	0,521995	-0,97176	0,67678	-1,53332	0,459438
№11	-0,44569	0,462329	-0,40026	-0,03402	-1,45864	0,973688	0,745373	1,078184	0,689834
№12	1,015566	2,746462	1,000884	-0,40691	0,679624	2,461567	0,033164	1,152879	2,917201
№13	0,558308	1,07247	1,506879	0,208881	0,496265	0,6414	-0,02672	-0,18496	-0,037
№14	-0,37207	-0,1358	0,779368	1,883159	1,119019	0,379418	-0,4341	-1,95256	-0,12492
№15	0,075518	-0,1691	-1,78054	0,216452	-1,23174	0,280967	-0,34201	-0,38502	0,637457
№16	1,496916	-0,58841	-0,49974	0,473203	-1,0368	1,80357	-0,736	-0,50872	0,068619
№17	-0,28896	0,446203	-0,91484	0,794603	-1,85377	-0,32669	-0,92964	-1,65892	-0,81736
№18	0,906852	0,343665	-1,35627	-1,07916	-0,59573	0,893995	-1,2897	0,105501	-0,0164
№19	-0,20213	2,412723	0,331584	1,662653	0,867514	-1,04876	0,693812	1,053982	0,127998
№20	-1,09344	-1,44322	0,071481	-0,14203	0,90791	-0,71726	2,0456	0,199177	0,578492
№21	0,904314	-0,86342	0,4766	0,613184	1,866085	-0,48573	0,369783	0,134893	0,367289
№22	-1,5006	-1,73989	-1,53731	1,171906	1,583633	0,348574	0,857586	0,569942	-0,55565
№23	0,255133	-0,86019	-1,58541	-0,15405	-0,3131	-1,81364	0,73422	0,475809	0,896157
№24	2,334688	-1,46174	0,008539	-1,12094	-0,31143	0,25361	0,911409	-0,18597	-0,60677
№25	-0,52052	-0,54165	-0,74141	0,074353	0,337359	1,202123	0,690598	0,232119	-0,61148
№26	-0,27343	0,387673	-0,91909	0,602786	-1,40685	-0,85449	1,234954	-0,87287	-1,22176
№27	2,004166	0,091728	0,337604	-0,48128	0,575301	0,441478	-0,70958	-2,21335	-0,78453
№28	0,839556	1,215978	-0,84937	-0,65756	1,281245	-0,47309	-0,30995	-1,02584	-0,6227
№29	1,151044	-1,28777	1,676542	1,878203	0,050657	-0,24123	0,985516	0,210519	1,441019
№30	-0,28593	1,099007	1,128041	-0,41039	-1,09655	-0,17468	0,663684	-0,84142	-0,40929
№31	0,080928	0,669183	1,356735	0,318165	-0,40468	-0,42938	-0,60575	-0,019	-1,04365
№32	-0,55257	-0,37794	1,04969	-1,1744	-1,23917	0,181357	0,928433	0,528332	1,072307
№33	-1,26525	0,08791	-1,11073	-0,96453	0,209037	-0,38001	-1,3631	0,843861	0,267186
№34	-0,31243	0,228166	0,31976	0,529329	0,219745	-0,30615	0,066793	0,533313	1,069083
№35	-0,22118	1,001063	0,371714	0,876722	1,153332	-0,06949	-0,7184	-0,47063	0,277739

Продовження табл.2									
Номер вибірки	1	2	3	4	5	6	7	8	9
№36	0,077741	-2,35078	-0,17511	0,738645	-0,27002	1,8029	0,339643	0,15605	0,785327
№37	-1,58415	-0,30823	-0,20045	-0,28653	0,406431	0,005666	2,048036	-1,26919	-0,34266
№38	0,284994	0,359549	0,394743	-2,13788	0,201026	-0,84378	-1,63559	0,539965	-0,24369
№39	0,939701	-0,20034	-0,0123	-0,3239	0,391017	-0,72957	1,068435	0,317677	-0,96104
№40	1,838417	-0,53246	1,349756	-0,54229	-0,89422	-0,06132	0,070723	-1,90911	1,553387
№41	-1,23436	1,274586	-0,54184	-0,64602	0,351888	-0,86804	1,311865	-0,5264	-0,94583
№42	-0,58774	0,799907	-0,85782	-0,91383	0,862724	-0,52571	1,650441	-1,20593	-0,88801
№43	1,304959	0,580929	0,721667	1,215632	-0,28213	-0,00282	0,737244	-0,88903	0,346307
№44	-0,46409	0,405951	0,416301	-0,42814	0,337541	0,951067	1,065192	-0,79778	0,032067
№45	1,524169	-0,32499	-0,21575	-0,15818	-0,04703	1,364651	0,548536	-1,3159	-0,68211
№46	0,203749	-0,09683	1,680916	0,207478	-1,45215	0,588068	-0,2097	-1,36907	0,902411
№47	-0,26371	0,727511	1,463133	-0,66768	1,029818	0,907839	0,241024	0,916177	0,014277
№48	-0,90091	0,677526	0,39096	-1,01534	2,053233	-0,3764	-0,038	0,626876	-0,97756
№49	0,301369	0,999508	-0,62168	1,038809	-0,27421	-0,0307	0,216528	0,489132	0,042762
№50	-1,48771	-0,70538	0,587945	1,066727	1,663474	1,167907	0,486981	0,148405	-0,09241
№51	-0,47145	0,798071	-0,51549	0,949842	0,294695	-0,1175	0,884651	0,034626	-1,86167
№52	2,055975	-1,93177	1,194094	0,496005	-0,89188	1,200498	0,709579	0,451766	0,801832
№53	-0,36968	0,973666	1,012523	0,564112	-1,29124	0,268203	0,341005	-0,35879	-0,90232
№54	1,661168	0,548888	1,303838	-0,9512	0,446943	0,458985	-0,12314	0,648605	1,066123
№55	1,241149	1,242665	-0,52306	0,176109	-0,57604	-1,65693	0,348929	0,038886	-1,31559
№56	-1,19535	-0,54532	-0,41062	-0,96909	-0,90055	-0,12784	-0,58701	1,325801	-0,72931
№57	-0,92354	1,175748	1,125939	-0,93888	1,231451	-0,40644	0,338302	0,237412	0,023951
№58	-1,25394	-1,22638	1,2509	-1,44478	-1,12977	0,206346	0,470362	0,577459	-1,4663
№59	-1,47258	-1,17512	-0,81479	-0,76086	-2,20585	0,255106	-0,84808	0,619939	0,001359
№60	-0,33759	1,821534	-0,2715	-1,26502	0,047698	0,837155	2,196931	-0,35207	0,4037
№61	0,70808	-0,82455	-1,13744	0,661818	-1,33257	-0,90527	-0,69267	1,400359	-1,16712
№62	-2,01689	-0,83915	-0,64391	0,622161	-0,41384	1,535605	1,085569	1,875383	0,204602
№63	0,85173	-0,43645	-1,6716	0,759665	-0,62467	-0,13988	1,287968	-0,87001	1,231418
№64	-0,59497	0,505948	-0,36977	0,47288	-1,62956	-0,54586	0,493949	0,307914	-0,71968
№65	-0,78769	-1,24758	-0,14255	-0,24376	1,063171	-1,62364	-0,81164	-0,49174	1,985319
№66	-0,58659	-0,33673	0,687548	-0,02083	0,714614	0,831839	0,989987	1,104563	0,179226
№67	-0,56281	0,054924	1,567391	-0,68715	0,440672	0,552338	-0,44111	-0,22801	0,761934
№68	1,608231	0,203024	-0,72625	0,183767	-0,9388	1,007819	-0,57149	-0,87386	-0,69798
№69	1,000403	0,32434	-2,1862	2,106543	-1,15232	-0,28519	2,674733	0,864474	-0,41931
№70	0,706819	0,700919	-0,04415	0,590978	0,268346	-0,56271	0,19672	1,067196	-1,63073

Продовження табл.2									
Номер вибірки	1	2	3	4	5	6	7	8	9
№71	0,749581	-1,27489	0,133269	0,862954	-1,27376	-0,7717	-2,83161	0,14263	2,705987
№72	1,323892	-0,57498	0,74181	-0,32261	1,384316	-1,30966	-1,26115	-0,38862	1,448972
№73	-0,52189	1,077091	0,459124	0,362972	0,551539	-0,65993	0,477002	0,311809	-1,57517
№74	-0,13804	-1,03427	0,542153	-2,50232	-0,05323	0,080534	-1,44022	-1,63422	1,900694
№75	0,200522	0,133961	-0,89472	0,860054	-1,11185	0,662876	-0,45974	-0,11143	-0,56244
№76	0,342171	-2,20964	-2,10835	-0,05002	-1,02481	-0,07534	0,528052	-1,27192	-1,24268
№77	-0,64351	0,913748	-0,80467	0,742245	1,02759	-0,04472	-0,20426	-0,16803	1,148151
№78	-1,51546	1,108225	-0,8547	-1,41641	0,498853	0,565218	-1,6462	0,588339	-0,82965
№79	-0,03669	-1,25717	-1,70421	-0,5312	1,156154	-0,7775	-0,29735	-1,04243	0,232115
№80	-0,2977	1,903248	-2,30256	-0,81978	-0,22126	-0,43441	-0,23348	2,450658	-0,70052
№81	-1,07761	0,15186	-0,44284	0,55345	-1,6033	1,166045	0,757594	0,197408	0,238528
№82	0,615144	-1,4128	0,320147	-0,36685	-0,04364	1,261954	1,79747	-0,10653	0,591693
№83	-1,69818	0,254471	1,388357	-0,61311	-0,06187	-0,4697	-0,12041	1,22017	-0,68824
№84	1,231538	2,487735	-1,29739	1,66724	-0,03224	-1,55376	1,262817	0,161214	0,394035
№85	0,049492	-0,10903	-0,01961	-0,89924	-0,50734	0,979015	0,878504	-0,86613	-3,29277
№86	-0,93943	1,587095	0,995545	-0,07411	-0,17668	0,812995	0,860439	-1,35838	0,648139
№87	-0,97249	-0,64273	-0,10945	-0,05885	-0,25159	0,485933	-0,98957	0,599614	0,65685
№88	-0,45801	0,104316	0,303415	-1,82194	0,989689	0,057195	-0,97047	-0,64065	0,395836
№89	1,054466	1,348432	-0,011	1,636921	-0,47683	-1,57207	-0,33111	-1,27365	-1,54246
№90	-0,64315	-1,22032	1,473169	-0,34177	0,308003	1,328846	2,069919	0,303262	0,480372
№91	-0,94934	0,818808	0,359668	0,611513	0,709245	0,395474	-0,47641	0,042081	-0,23342
№92	-0,13709	1,687859	-0,28873	1,701858	-1,33477	1,563581	0,758647	-0,89181	1,136509
№93	-1,59376	0,304772	-0,17519	1,010675	0,874221	-0,14267	-0,14304	0,669859	-0,09116
№94	0,216539	0,173429	-0,24562	-0,50369	-0,88042	-0,62335	-1,34453	-0,53426	-0,1729
№95	1,21018	0,021583	1,129576	0,190844	0,521576	-0,82022	0,112071	-1,76245	1,190258
№96	-1,69177	-0,21797	-0,14094	0,557041	-0,01705	3,162216	0,568615	1,253433	-4,16873
№97	1,103368	0,858219	0,552009	-0,6095	-1,34277	2,795349	0,756399	1,255537	-0,42656
№98	-0,03625	0,309644	-0,94828	-1,20894	-0,96465	0,229558	2,015325	2,206009	1,090485
№99	0,506098	-0,06277	1,066335	-0,4918	-0,73826	0,191561	0,228625	1,369818	-0,3724
№100	-0,3179	0,191898	0,457306	-3,1622	0,225446	1,091386	-0,08304	1,522047	-1,44822



Таблиця 3 – Результати обчислення НО і класифікації

Нечіткі оцінки	«168»	«267»	«249»	«348»	«1478»	«2369	Med	ЛК	Правильна класифікація
№1	-0,1465	0,2053	0,6865	0,8280	0,3840	0,4695	0,4267	110010	+
№2	-0,5768	-0,7155	0,1000	0,0988	-0,3575	-0,3568	-0,3846	110010	+
№3	-0,6883	-0,5800	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№4	-0,3168	-0,2803	0,8638	0,3950	0,1253	0,2508	0,188	110010	+
№5	-0,5123	-0,5390	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№6	-1,0560	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,193	110010	+
№7	-0,2823	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№8	-0,7983	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№9	-0,9075	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№10	-0,5455	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+
№11	-0,1358	0,2253	0,6761	0,8006	0,3821	0,4683	0,4198	110010	+
№12	-0,5616	-0,7032	0,1042	0,0975	-0,3465	-0,3382	-0,4035	110010	+
№13	-0,6773	-0,5713	0,0589	-0,0742	-0,3538	-0,3287	-0,3429	110001	-
№14	-0,3164	-0,2813	0,8636	0,3940	0,1243	0,2501	0,1875	110010	+
№15	-0,5119	-0,5388	0,2164	0,1648	-0,1749	-0,0607	-0,1179	110010	+
№16	-0,9986	-0,2727	0,1288	0,0887	-0,4647	-0,0559	-0,1925	110010	+
№17	-0,2828	-0,0713	0,5773	0,1272	-0,0453	0,2079	0,0398	110010	+
№18	-0,7978	-0,9348	0,1857	0,0739	-0,3096	-0,4011	-0,3561	110001	-
№19	-0,9068	-0,4497	-0,0402	0,0451	-0,4245	-0,1583	-0,2921	110010	+
№20	-0,5452	-0,4301	1,1879	0,4223	-0,1329	0,3601	0,1132	110010	+
№21	-0,1502	0,2049	0,6901	0,8325	0,3845	0,4723	0,4362	110010	+
№22	-0,5772	-0,7152	0,0978	0,0955	-0,3572	-0,3565	-0,3844	110010	+
№23	-0,6883	-0,5801	0,0609	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№24	-0,3168	-0,2803	0,8638	0,3954	0,1253	0,2508	0,1881	110010	+
№25	-0,5123	-0,5391	0,2168	0,1661	-0,1755	-0,0610	-0,1183	110010	+
№26	-1,0558	-0,2616	0,1295	0,0902	-0,4651	-0,0569	-0,193	110010	+

Продовження табл.3

Нечіткі оцінки	«168»	«267»	«249»	«348»	«1478»	«2369	Med	ЛК	Правильна класифікація
№27	-0,2823	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,04031	110010	+
№28	-0,7983	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3556	110001	-
№29	-0,6883	-0,5825	0,0610	-0,0773	-0,3268	-0,3363	-0,3314	110001	-
№30	-0,9075	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2914	110010	+
№31	-0,5455	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1133	110010	+
№32	-0,1465	0,2053	0,6865	0,8280	0,3840	0,4695	0,4266	110010	+
№33	-0,5768	-0,7155	0,1011	0,0988	-0,3575	-0,3568	-0,3855	110010	+
№34	-0,3168	-0,2803	0,8638	0,3950	0,1253	0,2508	0,1879	110010	+
№35	-0,5123	-0,5391	0,2168	0,1660	-0,1755	-0,0610	-0,1182	110010	+
№36	-1,0561	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,1931	110010	+
№37	-0,2823	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0411	110010	+
№38	-0,7983	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№39	-0,9075	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№40	-0,5455	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+
№41	-0,1466	0,2054	0,6866	0,8281	0,3839	0,4691	0,4271	110010	+
№42	-0,5768	-0,7155	0,1003	0,0988	-0,3575	-0,3568	-0,3846	110010	+
№43	-0,6883	-0,5831	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№44	-0,3168	-0,2803	0,8638	0,3950	0,1253	0,2508	0,1878	110010	+
№45	-0,5123	-0,5392	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№46	-1,0562	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,1929	110010	+
№47	-0,2823	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№48	-0,7983	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№49	-0,9075	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№50	-0,5455	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+
№51	-0,1465	0,2053	0,6865	0,8280	0,3840	0,4695	0,4267	110010	+
№52	-0,5759	-0,7155	0,0999	0,0988	-0,3575	-0,3568	-0,3846	110010	+

Продовження табл.3

Нечіткі оцінки	«168»	«267»	«249»	«348»	«1478»	«2369	Med	ЛК	Правильна класифікація
№53	-0,6874	-0,5786	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№54	-0,3158	-0,2803	0,8638	0,3950	0,1253	0,2508	0,1864	110010	+
№55	-0,5119	-0,5397	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№56	-1,0559	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,1933	110010	+
№57	-0,2821	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№58	-0,7984	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№59	-0,9073	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№60	-0,5454	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+
№61	-0,1465	0,2053	0,6865	0,8280	0,3840	0,4695	0,4267	110010	+
№62	-0,5768	-0,7155	0,1000	0,0988	-0,3575	-0,3568	-0,3846	110010	+
№63	-0,6885	-0,5822	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№64	-0,3168	-0,2803	0,8638	0,3950	0,1253	0,2508	0,1902	110010	+
№65	-0,5123	-0,5391	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№66	-1,0558	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,2003	110010	+
№67	-0,2823	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№68	-0,7983	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№69	-0,9076	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№70	-0,5454	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+
№71	-0,1462	0,2053	0,6865	0,8280	0,3840	0,4695	0,4267	110010	+
№72	-0,5773	-0,7155	0,1000	0,0988	-0,3575	-0,3568	-0,3846	110010	+
№73	-0,6881	-0,5811	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№74	-0,3166	-0,2803	0,8638	0,3950	0,1253	0,2508	0,1883	110010	+
№75	-0,5123	-0,5395	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№76	-1,0561	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,1933	110010	+
№77	-0,2822	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№78	-0,7982	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-

Продовження табл.3

Нечіткі оцінки	«168»	«267»	«249»	«348»	«1478»	«2369	Med	ЛК	Правильна класифікація
№79	-0,9079	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№80	-0,5458	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+
№81	-0,1464	0,2053	0,6865	0,8280	0,3840	0,4695	0,4267	110010	+
№82	-0,5771	-0,7155	0,1000	0,0988	-0,3575	-0,3568	-0,3846	110010	+
№83	-0,6885	-0,5802	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№84	-0,3169	-0,2803	0,8638	0,3950	0,1253	0,2508	0,1869	110010	+
№85	-0,5124	-0,5392	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№86	-1,0561	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,1927	110010	+
№87	-0,2823	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№88	-0,7983	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№89	-0,9075	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№90	-0,5457	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+
№91	-0,1462	0,2053	0,6865	0,8280	0,3840	0,4695	0,4267	110010	+
№92	-0,5768	-0,7155	0,1000	0,0988	-0,3575	-0,3568	-0,3846	110010	+
№93	-0,6883	-0,5799	0,0610	-0,0773	-0,3268	-0,3363	-0,3315	110001	-
№94	-0,3171	-0,2803	0,8638	0,3950	0,1253	0,2508	0,1877	110010	+
№95	-0,5123	-0,5391	0,2168	0,1660	-0,1755	-0,0610	-0,1183	110010	+
№96	-1,0561	-0,2618	0,1293	0,0900	-0,4658	-0,0570	-0,1935	110010	+
№97	-0,2824	-0,0718	0,5783	0,1278	-0,0473	0,2088	0,0403	110010	+
№98	-0,7982	-0,9355	0,1868	0,0745	-0,3100	-0,4013	-0,3557	110001	-
№99	-0,9077	-0,4505	-0,0398	0,0448	-0,4245	-0,1585	-0,2915	110010	+
№100	-0,5453	-0,4308	1,1875	0,4228	-0,1328	0,3598	0,1135	110010	+

## Акти впровадження дисертаційного дослідження

**ЗАТВЕРДЖУЮ**  
перший проректор НТУУ «КПІ ім. І. Сікорського»  
академік НАН України, проф.  
Ю. І. Якименко  
2018 р.

**АКТ**

впровадження результатів дисертаційного дослідження аспіранта Семенюка Романа Сергійовича на тему «Удосконалення способів опрацювання експериментальних даних з застосуванням класифікації»

Комісія у складі:  
голова – д.т.н., доц. Сременко В.С.  
члени комісії – к.т.н., доц. Павлишин М.М.;  
к.т.н., доц. Яремчук Н.А.

цим Актом засвідчує, що результати дисертаційного дослідження Семенюка Романа Сергійовича на тему «Удосконалення способів опрацювання експериментальних даних з застосуванням класифікації», що виконувалась на кафедрі інформаційно-виміральної техніки факультету авіаційних та космічних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», впроваджені в НДР «Система метрологічного забезпечення реалізації завдань агромоніторингу» (номер держреєстрації 0115U000004).

<b>Голова комісії</b> д.т.н., доц. (науковий ступінь, ім'я та прізвище)	 (підпис)	<u>Сременко В.С.</u> (прізвище та ініціали)
<b>Члени комісії</b> к.т.н., доц. (науковий ступінь, ім'я та прізвище)	 (підпис)	<u>Яремчук Н.А.</u> (прізвище та ініціали)
к.т.н., доц. (науковий ступінь, ім'я та прізвище)	 (підпис)	<u>Павлишин М.М.</u> (прізвище та ініціали)

«11» червня 2018 р.

**ЗАТВЕРДЖУЮ**  
перший проректор НТУУ «КПІ ім. І. Сікорського»  
академік НАН України, проф.  
Ю. І. Якименко  
2021 р.

**АКТ**

впровадження результатів дисертаційного дослідження аспіранта Семенюка Романа Сергійовича на тему «Методи опрацювання виміральної та експертної інформації з застосуванням шкал класифікації»

Комісія у складі:  
голова – д.т.н., доц. Сременко В.С.  
члени комісії – к.т.н., доц. Павлишин М.М.;  
к.т.н., доц. Шведова В.В.

цим Актом засвідчує, що результати дисертаційного дослідження Семенюка Романа Сергійовича на тему «Методи опрацювання виміральної та експертної інформації з застосуванням шкал класифікації», що виконувалась на кафедрі інформаційно-виміральної техніки приладобудівного факультету Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», впроваджені в НДР «Теоретичні та практичні засади застосування м'яких вимірювань в системах визначення якості продукції» (номер держреєстрації 0118U001478).

<b>Голова комісії</b> д.т.н., доц. (науковий ступінь, ім'я та прізвище)	 (підпис)	<u>Сременко В.С.</u> (прізвище та ініціали)
<b>Члени комісії</b> к.т.н., доц. (науковий ступінь, ім'я та прізвище)	 (підпис)	<u>Шведова В.В.</u> (прізвище та ініціали)
к.т.н., доц. (науковий ступінь, ім'я та прізвище)	 (підпис)	<u>Павлишин М.М.</u> (прізвище та ініціали)

«01» березня 2021 р.