

А. І. Поворознюк<sup>1</sup>  
 О. А. Поворознюк<sup>1</sup>  
 Г. Є. Філатова<sup>1</sup>

## ПОБУДОВИ СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ МЕДИЧНИХ РІШЕНЬ НА ОСНОВІ ГЕТЕРОГЕННОГО АНСАМБЛЕВОГО КЛАСИФІКАТОРА

<sup>1</sup> Національний технічний університет «Харківський політехнічний інститут»

*На основі аналізу методів діагностики, які використовуються для постановки діагнозу лікарем згідно вимог доказової медицини, а також методів побудови діагностичних вирішальних правил у системах підтримки прийняття медичних рішень, виділено два підходи до формулювання діагностичного висновку. Перший – традиційний підхід лікарів-практиків на основі виявлення симптомів захворювання та аналізу структури симптомокомплексів з врахуванням досвіду клініцистів в даній предметній галузі медицини, який висвітлений в медичних довідниках, і його можна вважати експертною оцінкою кожного захворювання. Другий – формалізований підхід на основі синтезу ймовірного класифікатора, який навчається на навчальній вибірці. Так як вказані підходи розглядають одну і ту ж проблему з різних точок зору, то їх спільне використання є перспективним. Реалізація різнотипних моделей класифікаторів та агрегація їх результатів можлива при реалізації гетерогенного ансамблевого класифікатора. Мета дослідження – синтез гетерогенного ансамблевого класифікатора, який враховує як експертну, так і ймовірнісну складову процесу постановки діагнозу. В роботі розроблена математична модель гетерогенного ансамблевого класифікатора і обґрунтовано вибір його складових. В якості ймовірнісної складової вибрані широко вживані методи класифікації: метод порівняння з прототипом, метод К-найближчих сусідів та метод потенційних функцій. Формалізована експертна інформація щодо структури симптомокомплексів шляхом представлення симптомокомплексів захворювань числовими інтервалами значень «нижче норми», «норма», «вище норми» лінгвістичних змінних. Розглянуті варіанти врахування експертних оцінок щодо структури симптомокомплексів в ансамблевому класифікаторі. Виконано проектування системи підтримки прийняття рішень та комплексна перевірка розробленої системи на реальних медичних даних, яка підтвердила ефективність роботи системи.*

**Ключові слова:** медична діагностика; симптомокомплекс; експертна інформація; ймовірнісний класифікатор; гетерогенний ансамблевий класифікатор; система підтримки прийняття рішень

### Вступ

Розвиток високих технологій безумовно впливає на медичну діагностику, дозволяючи інтегрувати новітні досягнення в цю важливу сферу. Зараз, коли медичні інформаційні системи (МІС) охоплюють цілий спектр технологій [1]: від електронних медичних карт до інтелектуальних систем підтримки прийняття рішень, які використовують штучний інтелект [2] і можливості телемедицини [3], здатні аналізувати великі обсяги даних, виконувати обробку біомедичних сигналів та зображень, виявляючи діагностичні ознаки (маркери захворювань), що допомагають лікарям більш ефективно виконувати діагностику [4, 5]. Проте, незважаючи на цей прогрес, проблема синтезу діагностичних вирішальних правил залишається актуальною, тому що надійність і точність діагностичних висновків багато в чому залежать від якості і структури цих правил.

Одним з шляхів вирішення цієї проблеми є інтеграція медичної інформації з різних джерел і розробка комплексних діагностичних моделей з метою забезпечення більш глибокого розуміння стану пацієнта і, як результат, покращення якості медичних рішень [5]. Важливими є також етичні аспекти впровадження таких систем, зокрема захист персональних даних пацієнтів та забезпечення прозорості прийняття рішень.

Діагноз, що походить від грецького слова *διάγνωσις*, означає розпізнавання або ідентифікацію. Він представляє собою медичний висновок, що характеризує функції здоров'я, морфологічний стан пацієнта, а також наявні захворювання чи травми. Діагноз може включати зміни в організмі та пояснення причин, що призвели до летального результату. Відповідно, діагноз – це структурований висновок, оформлений за прийнятими класифікаціями захворювань.

Традиційна первинна діагностика базується на комплексному огляді пацієнта, аналізі анамнезу,

скарг, об'єктивних симптомів, які виявляються під час фізичного обстеження, а також при проведенні інструментальних та лабораторних досліджень. Лікар групує виявлені симптоми в синдроми та формує остаточний діагноз на основі принципів доказової медицини. Важливими факторами у цьому процесі є професійна майстерність лікаря, а також досвід клініцистів в даній предметній галузі медицини, який висвітлений а медичних довідниках, і його можна вважати експертною оцінкою кожного захворювання.

В сучасних МІС діагностика формалізується як класична задача класифікації, де модель об'єкта діагностики сприймається як "чорний ящик" і вивчається зв'язок між формалізованими станами об'єкта діагностики  $Y$  та вектором вхідних ознак  $X$ , тобто  $Y=f(X)$  [1]. Цей ймовірнісний зв'язок встановлюється на етапі навчання моделі класифікатора за допомогою аналізу навчальної вибірки з пацієнтів, у яких діагноз уже підтверджено. Однак існують суттєві виклики. По-перше, не завжди вдається сформувати репрезентативну вибірку, особливо для рідкісних захворювань. По-друге, лікарі не завжди довіряють результатам комп'ютерної діагностики, ставлячи під сумнів їх інтерпретацію. Так як вказані підходи розглядають одну і ту ж задачу з різних точок зору, то перспективним напрямом є поєднання традиційних та формалізованих методів діагностики. Комбінація різнотипних моделей класифікаторів та агрегація їх результатів можлива при розробці гетерогенного ансамблевого класифікатора, який підвищує точність і надійність діагностики окремих класифікаторів, а також дозволить лікарям більш впевнено користуватися результатами комп'ютерних систем у своїй практиці.

*Метою роботи* є розробка гетерогенного ансамблевого класифікатора, який враховує як експертну, так і ймовірнісну складову процесу постановки діагнозу.

### Розробка моделі ансамблевого гетерогенного класифікатора

Поєднання окремих класифікаторів через ансамблеві методи машинного навчання є ефективним способом зменшення впливу випадкових помилок або недоліків окремих класифікаторів і підвищення точності діагностики [6, 7]. Основні методи агрегації, такі як бегінг, (Bootstrap Aggregating), бустинг (Boosting) і стекінг (Stacking) [8], дозволяють досягти кращих результатів, ніж використання базових моделей окремо.

Бегінг передбачає створення кількох випадкових підвбірок даних шляхом вибірки з повтором. На цих підвбірках навчаються різні моделі, а результати їх прогнозів об'єднуються. Цей метод знижує варіативність і робить модель більш стійкою до випадкових помилок.

Бустинг базується на ідеї послідовного навчання моделей, де кожна нова модель фокусується на помилках попередніх. Цей підхід дозволяє підсилити важливість правильного прогнозування, особливо для екстремальних випадків, і забезпечує високу точність діагностики.

Стекінг поєднує прогнози різних моделей, використовуючи мета-моделі, які навчаються на виходах базових моделей. Ця стратегія дозволяє ефективно об'єднувати результати різних класифікаторів, але потребує значних обчислювальних ресурсів і складного налаштування параметрів.

Гомогенні ансамблі складаються з однотипних класифікаторів, які навчаються незалежно на різних навчальних вибірках. Їх результати комбінуються для отримання остаточного прогнозу. Цей підхід забезпечує певну простоту у реалізації, але може обмежити різноманітність моделей. Гетерогенні ансамблі, навпаки, поєднують різні методи машинного навчання, що дозволяє вловити різні аспекти даних. Цей підхід відкриває можливості для більш детального аналізу, оскільки різні алгоритми можуть мати різні сильні сторони.

Для розробки гетерогенного ансамблевого класифікатора в складі комп'ютеризованої системи підтримки прийняття діагностичних рішень в медицині, розглянемо кілька базових моделей, які можуть бути використані в якості ймовірнісних складових.

*Метод порівняння з еталоном* є ефективним підходом для аналізу кількісних характеристик, коли класи  $\Omega_m$  ( $m = \overline{1, M}$ ) формують компактні множини об'єктів, що мають сферичну форму у функціональному просторі ознак. Цей метод особливо корисний у ситуаціях, коли дані можна адекватно описати за допомогою геометричних характеристик.

У цьому методі кожен клас  $\Omega_m$  ( $m = \overline{1, M}$ ) представлений моделлю його еталона  $\omega^{mr}$ , в якості якого обирається геометричний центр (центроїд) класу. На етапі навчання координати кожного еталона розраховуються на основі навчальної вибірки об'єктів, що належать до відповідного класу, за формулою:

$$x_i^{mr} = \frac{1}{n_m} \sum_{j=1}^{n_m} x_i^j, \quad (i = \overline{1, p}), \quad (m = \overline{1, M}) \quad (1)$$

де  $x_i^{mr}$  –  $i$ -та координата еталона класу  $m$ ;  $n_m$  – кількість об'єктів класу  $m$  у навчальній вибірці;  $x_i^j$  –  $i$ -та координата  $j$ -го об'єкта класу  $m$  у навчальній вибірці;  $p$  – розмір координатного простору (кількість діагностичних ознак);  $M$  – кількість класів, на які виконується класифікація невідомих об'єктів (кількість захворювань, що діагностуються у цій галузі медицини).

Ця формула дозволяє знайти середнє значення кожної ознаки для всіх об'єктів у класі, що визначає позицію еталона у просторі ознак.

На етапі класифікації новий об'єкт  $\omega$  порівнюється з еталонами  $\omega^{mr}$  всіх класів  $\Omega_m$ . Для цього можуть використовуватися різні метрики, наприклад: евклідова чи манхеттенська відстань, косинусна подібність, тощо. Для евклідової відстані між об'єктом  $\omega$  та кожним з еталонів  $\omega^{mr}$  обчислюється міра близькості  $R(\omega, \omega^{mr})$  за формулою

$$R(\omega, \omega^{mr}) = \sqrt{\sum_{i=1}^p (x_i^\omega - x_i^{\omega^{mr}})^2}, \quad (m = \overline{1, M}), \quad (2)$$

де  $x_i^{\omega^{mr}}$  –  $i$ -та координата еталонного об'єкта класу  $m$ ;  $x_i^\omega$  –  $i$ -та координати об'єкта, який підлягає класифікації;  $p$  – розмір координатного простору (кількість діагностичних ознак);  $M$  – кількість класів, на які виконується класифікація.

На основі вибраної метрики визначається найближчий еталон, і новий об'єкт класифікується відповідно до класу  $\Omega_t$ , до якого належить цей еталон

$$R(\omega, \omega^{tr}) = \min_{m=1, M} R(\omega, \omega^{mr}), \quad (3)$$

де  $R(\omega, \omega^{mr})$  відстань між об'єктом  $\omega$  та еталонем  $\omega^{mr}$  класу  $\Omega_t$ , розраховується за виразом (2).

Метод Фікса–Ходжеса, відомий як метод "к-найближчих сусідів", є популярним алгоритмом класифікації, який застосовується в ситуаціях, коли структура класів є складною і далекою від сферичної. Цей метод базується на гіпотезі про безперервність багатовимірної щільності розподілу класів в кожній локальній області простору ознак. Суть даного методу полягає у визначенні деякого заданого числа  $k$  найближчих до невідомого об'єкта (в обраній метриці) об'єктів навчальної вибірки ("найближчих сусідів"). Невідомий об'єкт  $\omega$  належить до того класу, число представників якого переважає серед цих  $k$  сусідів.  $K$  – ціле додатне число, як правило, невелике. Якщо  $k = 1$ , то об'єкт просто приписується до класу цього єдиного найближчого сусіда.

Реалізація методу. Визначають відстані  $R(\omega, \omega_i)$  ( $i = \overline{1, N}$ ), де  $N$  – об'єм навчальної вибірки, між невідомим об'єктом  $\omega$  та усіма об'єктами навчальної вибірки, які являють собою сукупність всіх класів  $\Omega_m$ , ( $m = \overline{1, M}$ ). Після цього знайдені відстані ранжуються за зростанням, вибираються перші  $k$  елементів – "найближчі сусіди", серед яких визначається клас  $\Omega_t$ , число представників якого переважає серед обраних  $k$  "найближчих сусідів". До класу  $\Omega_t$  і належить невідомий об'єкт  $\omega$ .

Оскільки метод використовує функцію відстані для класифікації, то у випадку, коли ознаки представляють різні фізичні одиниці або мають дуже різні масштаби, використовується нормалізація навчальних даних. Крім того, можна призначати вагові значень внеску сусідів, щоб внесок у більш близьких сусідів був більше, ніж у віддалених, наприклад, надати кожному сусіду вагу  $1/d$ , де  $d$  – відстань до сусіда

Метод потенційних функцій. В якості потенційної функції, що характеризує належність об'єкта відповідному класові, використовується усюди позитивна і монотонно спадна функція відстані, аналогічна за формою електричному потенціалові  $\phi$ . Прикладами таких функцій можуть бути

$$\varphi(R) = \left| \frac{\sin(\alpha R^2)}{\alpha R^2} \right|, \quad \varphi(R) = e^{-\alpha R^2} \quad \text{або} \quad \varphi(R) = \frac{1}{1 + \alpha R^2}, \quad (4)$$

де  $R$  – відстань між точкою-джерелом і точкою-приймачем, у якій обчислюється потенціал;  $\alpha > 0$  – ваговий коефіцієнт, який характеризує швидкість убуття потенціалу  $\phi$ .

Точками-джерелами потенціалу виступають об'єкти класів  $\Omega_m$ , ( $m = \overline{1, M}$ ), а точкою-приймачем – об'єкт  $\omega$ , який підлягає класифікації. Об'єкт  $\omega$  належить до класу  $\Omega_i$ , сумарний потенціал якого буде максимальним. Метод Фікса–Ходжеса та метод потенційних функцій є потужними інструментами для класифікації об'єктів у складних умовах. Вони дозволяють враховувати локальну структуру даних і забезпечують ефективний підхід до класифікації в багатовимірному просторі ознак.

### Врахування експертної інформації щодо структури симптомокомплексу

Структура симптомокомплексу є відображенням думки експертів про конкретний діагноз, вироблену кількома поколіннями лікарів. Інформація про симптоми захворювань в неформалізованому виді наведена в різних медичних довідниках [9, 10].

Симптомокомплекс складається з різних типів симптомів, серед яких виділяються: патогномонічні симптоми – чітко вказують на наявність конкретного захворювання та слугують маркерами; специфічні симптоми – проявляються лише при певних захворюваннях, але не є однозначними показниками; неспецифічні симптоми – можуть бути характерними для кількох захворювань, не вказуючи точно на конкретну хворобу.

Структура симптомокомплексу відображає думки експертів щодо специфічного діагнозу. Вплив кожного симптому розраховується на основі лінгвістичних змінних, які описують відповідні симптоми (наприклад, «висока температура» або «високий артеріальний тиск»).

Кожному симптому  $x_j$  присвоюється експертний рейтинг, який вказує на його значущість у симптомокомплексі. Ці рейтинги приймають значення:  $e_0$  – для патогномонічних симптомів;  $e_1$  – для специфічних симптомів;  $e_2$  – для неспецифічних симптомів;  $e_3$  – для показників, яке не належать до синдрому цього захворювання. При цьому справедлива нерівність  $e_0 \geq e_1 \geq e_2 \geq e_3$ ,  $\sum_{i=0}^3 e_i = 1$ .

Виконаємо формалізацію думок фахівців щодо структури симптомокомплексу при створенні ансамблевого методу класифікації. При постановці діагнозу лікарі часто спираються на поняття норми для кожного симптому. Динамічний діапазон значень симптомів поділяється на три категорії: «нижче норми», «норма» та «вище норми». Так само визначаються бінарні ознаки, де лінгвістичні змінні позначаються термінами «ознака проявляється» або «ознака не проявляється».

Отже, кожен патологічний синдром (еталон для кожного класу) визначається діагностичними симптомами  $x_j$ , кожен з яких приймає значення відповідних лінгвістичних змінних. Опис типових наборів симптомів разом із варіантами лінгвістичних змінних формує формалізовану експертну оцінку еталонів для кожного класу захворювання. Якщо відомі значення динамічного діапазону для всіх діагностичних ознак і порогові значення «норми», центри цих діапазонів («нижче норми», «норма», «вище норми») слугують кількісним представленням еталона  $\omega^{ml}$  класу  $\Omega_m$ , базуючись на висновках експертів щодо структури симптомокомплексу, а не на навчальних даних. Ці еталони можуть використовуватися в якості базових моделей в гетерогенному ансамблевому класифікаторі. Розглянемо два сценарії спільного використання методів на основі навчальної вибірки та експертної оцінки.

*Сценарій 1:* Агрегація результатів базових моделей. У класичній діаграмі стекінг-ансамблевого класифікатора результати класифікації різних моделей об'єднуються через мета-модель. Для кожного класу  $\Omega_m$ , ( $m = \overline{1, M}$ ) створюються основні моделі класифікації, які ґрунтуються на аналізі навчальної вибірки та на експертних оцінках структури симптомокомплексів. Класифікація нових об'єктів здійснюється кожною базовою моделлю відповідно алгоритмів, описаних вище. Якщо результати класифікації різних моделей відрізняються, результати агрегуються, і остаточний діагноз  $D_k$  формується шляхом зваженого голосування.

*Сценарій 2:* Агрегація моделей в методі порівняння з еталоном при визначенні координат еталонів класів. В даному варіанті ансамблевого класифікатора виконується агрегацію результатів розрахунку координат еталонів класів різними моделями. Для кожного класу  $\Omega_m$ , ( $m = \overline{1, M}$ ) створюються окремі еталони ( $\omega^{mr}$  на основі навчальної вибірки та  $\omega^{ml}$  на основі експертних оцінок). Координати еталону мета-моделі розраховуються за формулою:

$$x_i^m = k_1 x_i^{mr} + k_2 x_i^{ml}, \quad (i = \overline{1, p}), \quad (m = \overline{1, M}), \quad (5)$$

де  $k_i > 0$ ,  $\sum k_i = 1$ ,  $i = \overline{1, 2}$  – вагові коефіцієнти, що відповідають довірі до навчальної вибірки  $k_1$  і

до експертної оцінки  $k_2$  відповідно.

Після навчання мета-моделі нові об'єкти класифікуються за стандартними алгоритмами методу порівняння з еталоном (2) і (3). Остаточний діагноз, визначений ансамблевим класифікатором, є підтримкою в прийнятті діагностичного рішення, яке остаточно приймає лікар. Врахування експертної інформації щодо структури симптомокомплексу та використання різних методів класифікації дозволяє створити більш надійний і точний гетерогенний ансамблевий класифікатор. Це поєднання експертної думки та статистичних методів відкриває нові можливості для покращення процесу діагностики в медичній практиці.

### Результати тестування

Для тестування створеного класифікатора була сформована навчальна вибірка, що включає 200 пацієнтів. Діагностичними ознаками стали 9 показників клінічних аналізів крові та сечі, за якими були визначені діагностичні стани захворювань молочної залози:  $D_1$  – кіста молочної залози;  $D_2$  – мастопатія;  $D_3$  – ліпома;  $D_4$  – фіброаденома;  $D_5$  – практично здорова. Результати діагностики подані в табл. 1. У табл. 1 використані наступні умовні позначення:  $P(x)$  – результати застосування

Таблиця 1 – Результати діагностики

Код	Кількість пацієнтів	Результати діагностики					
		$P(x)$		$\mu(x)$		$\Omega$	
		$N$	%	$N$	%	$N$	%
D1	40	35	87	34	85	38	95
D2	60	54	90	53	88	56	93
D3	20	17	85	17	85	19	95
D4	30	26	87	24	80	26	87
D0	50	45	90	47	94	49	98
Разом	200	177	88	175	87,5	188	94

методу порівняння з еталоном;  $\mu(x)$  – результат використання детермінованої логіки через інтерпретацію симптомокомплексу в термінах нечіткої логіки;  $\Omega$  – результат роботи розробленого класифікатора;  $N$  – кількість правильно ідентифікованих об'єктів; % – відсоток від загальної кількості правильно визначених об'єктів. У виразі (5) вагові коефіцієнти мають однакову вагу, тобто  $k_1 = k_2 = 0,5$ . Згідно з даними табл. 1, частка правильно класифікованих об'єктів за результатами клінічного аналізу при діагностиці захворювань молочної залози методом порівняння з еталоном становить 88%, за детермінованою логікою – 87,5%, а використання розробленого класифікатора – 94%. Ці результати підтверджують функціональність і

ефективність розробленого методу діагностики.

### Висновки

У даній роботі обґрунтовано необхідність синтезу класифікатора для медичної діагностики, який поєднує два методи формування діагностичних висновків: ймовірнісний підхід, що базується на аналізі навчальних вибірок, та метод, заснований на формалізації експертних знань про структуру симптомокомплексів. Розроблено математичну модель гетерогенного ансамблевого класифікатора, в якому базовими моделями виступають класифікатори, що використовують метод порівняння з еталоном, а також формалізовані експертні оцінки. Запропоновано кілька альтернатив агрегації результатів базових моделей, що дозволяє підвищити точність і надійність діагностичних висновків. Результати тестування, проведеного на основі реальних медичних даних, підтвердили працездатність розробленого підходу та продемонстрували ефективність у процесі діагностики. Подальші дослідження будуть спрямовані на застосування розроблених правил прийняття рішень не лише до кількісних параметрів, але й до різноманітних діагностичних даних, таких як біомедичні сигнали та зображення, що відкриває нові перспективи для вдосконалення медичної діагностики.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] О.Г. Аврунін, Є.В. Бодяньський, М.В. Калашник, В.В. Семенець, В.О. Філатов, *Сучасні інтелектуальні технології функціональної медичної діагностики: монографія*. Харків : ХНУРЕ, 2018. 236 с.
- [2] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q. L. Han, and Y. Tang, A brief overview of ChatGPT: The history, status quo and potential future development, *IEEE/CAA Journal of Automatica Sinica*, 2023, vol. 10, No. 5, pp. 1122-1136.
- [3] Y. T. Yang, U. Iqbal, J. Horn-Yu Ching et al. Trends in the growth of literature of telemedicine: A bibliometric analysis. *Computer Methods and Programs in Biomedicine*. 2015; Vol. 122 (3): 471–479. DOI: 10.1016/j.cmpb.2015.09.008
- [4] T. Biloborodova, L. Scislo, I. Skarga-Bandurova et al. Fetal ECG signal processing and identification of hypoxic pregnancy conditions in-utero *Mathematical Biosciences and Engineering*, 2021, № 4, p. 4919-4942 [https://doi: 10.3934/mbe.2021250](https://doi.org/10.3934/mbe.2021250)
- [5] А.І. Поворознюк, О.А. Поворознюк, Х. Шехна, Синтез комбінованого вирішального правила в медичних системах підтримки прийняття рішень. *Системи управління, навігації та зв'язку*, 2021, випуск 1(63),– С.103-106.
- [6] A. Z. Kiflay, A. Tsokanos, and R. Kirner, A Network Intrusion Detection System Using Ensemble Machine Learning, *International Camahan Conference on Security Technology (ICCST)*, 2021, pp. 1-6.

- [7] A. Sarkar, H. S. Sharma, and M. M. Singh, A supervised machine learning based solution for efficient network intrusion detection using ensemble learning based on hyperparameter optimization, *International Journal of Information Technology*, 2022, vol. 15, No. 1, pp. 423–434.
- [8] O. Hornostal, V. Chelak, and S. Gavrylenko, Research of Intelligent Data Analysis Methods for Identification of Computer System State, *Proceedings of the 30th International Scientific Symposium Metrology and Metrology Assurance (MMA)*, 2020, pp. 1-5.
- [9] І. І. Мавров *Оптимальна діагностика та лікування захворювань шкіри та венеричних захворювань. Посібник для дерматологів і венерологів* – Київ. ТОВ «Доктор Медіа», 2007 – 344 с.
- [10] К. І. Григор'єва *Педіатрія: посібник для практикуючих лікарів* – Київ. Медпрес-інформ, 2014, – 816 с.

**Поворознюк Анатолій Іванович** – д-р техн. наук, професор, професор кафедри комп'ютерної інженерії та програмування, e-mail: ai.povoroznjuk@gmail.com

**Поворознюк Оксана Анатоліївна** – канд. техн. наук, доцент, доцент кафедри комп'ютерної інженерії та програмування, e-mail: povoks76@gmail.com

**Філатова Ганна Євгенівна** – д-р техн. наук, професор, професор кафедри комп'ютерної інженерії та програмування, e-mail: filatova@gmail.com

**A. I. Povoroznyuk<sup>1</sup>**  
**O. A. Povoroznyuk<sup>1</sup>**  
**A. Ye. Filatova<sup>1</sup>**

## **Building an acceptance support system medical decisions based on heterogeneous ensemble classifier**

<sup>1</sup> National Technical University "Kharkiv Polytechnic Institute"

*Based on the analysis of diagnostic methods used to make a diagnosis by a doctor in accordance with the requirements of evidence-based medicine, as well as methods of constructing diagnostic decision rules in medical decision-making support systems, two approaches to the formulation of a diagnostic conclusion are distinguished. The first is the traditional approach of medical practitioners based on the identification of disease symptoms and analysis of the structure of symptom complexes, taking into account the experience of clinicians in this subject field of medicine, which is covered in medical handbooks, and it can be considered an expert assessment of each disease. The second is a formalized approach based on the synthesis of a probabilistic classifier that is trained on a training sample. Since these approaches consider the same problem from different points of view, their joint use is promising. Implementation of different types of classifier models and aggregation of their results is possible when implementing a heterogeneous ensemble classifier. The goal of the study is the synthesis of a heterogeneous ensemble classifier that takes into account both the expert and probabilistic components of the diagnosis process. In the paper, a mathematical model of a heterogeneous ensemble classifier is developed and the choice of its components is substantiated. As a probabilistic component, widely used classification methods are selected: the method of comparison with a prototype, the method of K-nearest neighbors, and the method of potential functions. Formalized expert information on the structure of symptom complexes by presenting the symptom complexes of diseases with numerical intervals of values "below normal", "normal", "above normal" of linguistic variables. Considered options for taking into account expert assessments regarding the structure of symptom complexes in the ensemble classifier. The decision support system was designed and a comprehensive check of the developed system was carried out on real medical data, which confirmed the effectiveness of the system.*

**Keywords:** medical diagnosis; symptom complex; expert information; probabilistic classifier; heterogeneous ensemble classifier; decision support system

**Povoroznyuk Anatoly Ivanovich** – Dr. Tech. Sciences, professor, professor of the department of computer engineering and programming, e-mail: ai.povoroznjuk@gmail.com

**Povoroznyuk Oksana Anatoliivna** – Candidate of technical Sciences, associate professor, associate professor of the Department of Computer Engineering and Programming, e-mail: povoks76@gmail.com

**Filatova Anna Evgenivna** – Dr. Tech. Sciences, professor, professor of the department of computer engineering and programming, e-mail: filatova@gmail.com