

УДК 004.8.032.26; 57.089

INTELLIGENT PROTEIN BALANCE ANALYZER BASED ON AN ARTIFICIAL NEURAL NETWORK WITH THE FORCED LEARNING ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗАТОР БІЛКОВОГО БАЛАНСУ У БАЗИСІ ШТУЧНОЇ НЕЙРОННОЇ МЕРЕЖІ З ПРИМУСОВИМ НАВЧАННЯМ

Alyoshin S.P. / Альошин С.П.

c.t.s., as. prof. / к.т.н., доц.

Haitan O.M. / Гайтан О.М.

ORCID: 0000-0002-7228-9937

National University «Yuri Kondratyuk Poltava Polytechnic»,

Poltava, Pershotravnevuy Ave. 24, 36011

Національний університет «Полтавська політехніка імені Юрія Кондратюка»,

Полтава, Першотравневий проспект 24, 36011

Анотація. Білок – основний будівельний матеріал організму, норма якого має бути на постійному контролі для недопущення його дефіциту. Одним із найдоступніших і найпоширеніших досліджень рівня білків є біохімічний аналіз крові, що дозволяє обчислити загальну кількість білків, а також окремих його фракцій. Однак для отримання достовірних результатів аналізу необхідно дотримуватися деяких обов'язкових правил, крім того, на результати аналізу можуть суттєво впливати ліки, що приймаються пацієнтом. Для випадків, коли немає можливості провести стандартний білковий аналіз або не вдається забезпечити обмеження та виконання обов'язкових умов, пропонується технологія інтелектуального розпізнавання рівня білкового балансу. Мета дослідження – синтез нейромережевої моделі розпізнавання рівня білкового балансу досліджуваного суб'єкта на основі отримання знань з репрезентативної вибірки прикладів з бази історій хвороб, пов'язаних з дефіцитом білкових фрагментів. Завдання знаходження функціональної залежності білкового балансу від значень симптомів вирішено застосуванням технології примусового навчання нейромереж та реалізовано моделями з різною архітектурою як завдання розпізнавання образів. Практична значимість результатів – можливість отримання додаткового незалежного каналу аналізу білкового балансу без необхідності використання традиційного методу біохімічного аналізу крові з його обмеженнями та протипоказаннями. Синтезований ансамбль моделей, алгоритмічний та програмний інструментарій дозволяють розширити застосування та знизити вимоги до обмежень, що мають місце при традиційних методах аналізу.

Ключові слова: білковий аналіз, інтелектуальний аналізатор, нейронна мережа, розпізнавання, ознаки-симптоми

Вступ.

Білок – основний будівельний матеріал нашого організму. Існує безліч різних типів білків, кожен з яких має своє призначення та виконує певні функції. Будь-яке відхилення від норми призводить до ослаблення організму тією чи іншою мірою [1]. Отже, норма білка в організмі повинна бути на постійному контролі для недопущення його дефіциту, процедура оцінювання – безболісна, оперативна, мало витратна, постійна та ефективна. Матеріалом для досліджень білка найчастіше стає кров. Біохімічний аналіз крові – одне з найдоступніших і найпоширеніших досліджень виявлення рівня білків [1]. Цей аналіз дозволяє обчислити загальну кількість білків та окремих його фракцій.

Однак для того, щоб аналіз дав достовірний результат, необхідно дотримуватись деяких обов'язкових правил. Кров на аналіз загального рівня



білка здають вранці до їжі. При цьому бажано утримуватись від будь-якої їжі протягом 10-14 годин до здачі аналізу. За добу до проведення дослідження необхідно виключити будь-які фізичні навантаження та бажано емоційні, вживання алкоголю навіть у малих дозах. Крім того, слід відмовитися від будь-яких солодких напоїв – соків, газованої води, чаю та кави з цукром.

На результати аналізу можуть суттєво вплинути ліки, що приймаються пацієнтом. Одні (анаболіки, імунодепресанти, інсулін тощо) можуть підвищити рівень білка, інші – знизити, що спотворює реальну картину білкового балансу.

Мета дослідження.

Для випадків, коли немає можливості провести стандартний білковий аналіз або не вдається виконання перерахованих вище обмежень та обов'язкових умов, пропонується технологія інтелектуального аналізу ретроспективних прецедентних даних на основі репрезентативної вибірки прикладів з бази однотипних історій хвороб, пов'язаних з дефіцитом білкових фрагментів. З урахуванням потенційних можливостей існуючих пакетів технічного аналізу даних, бази прецедентів [2], доцільно вирішити це завдання на основі примусового навчання ансамблю нейромережевих моделей, вибрати найкращу та використовувати її як самостійний канал оперативного, не критичного до ряду побутових обмежень, продуктивного розпізнавання рівня білкового балансу [3-4]. Отже, метою дослідження є синтез нейромережевої моделі розпізнавання рівня білкового балансу досліджуваного суб'єкта на основі отримання знань з репрезентативної вибірки прикладів зі зниженням рівня обмежувальних умов та збереженням необхідної ефективності.

Математична постановка задачі.

Необхідно визначити вид аналітичного зв'язку (F) симптомів організму суб'єкта (X) з рівнем білкового балансу (Y) (норма, дефіцит, надлишок) як відображення простору факторів (ознак) на простір станів (класів) із заданими показниками якості [4]:

$$F : X \rightarrow Y_{opt}, X \subset \mathfrak{R}^m, Y_{opt} \subset \mathfrak{R}, \quad (1)$$

де X – вектор зафіксованих ознак стану білкового балансу суб'єкта; Y_{opt} – вихідне значення класу (рівня білкового балансу).

Аналітичний вид цього зв'язку можна визначити, виходячи з теореми Колмогорова-Арнольда про представлення функції кількох аргументів через суму композицій функцій однієї змінної та її конвертації в нейромережевий формат Хехт-Нільсена [2-4]:

$$y(x) = \alpha \sum_{i=1}^M v_i (w_{i1}x_1 + w_{i2}x_2 + \dots + w_{in}x_n + u_i), \quad (2)$$

де M – потужність навчальної вибірки, α, v – параметри нейромережі, n – кількість нейронів, $w_{i1}, w_{i2}, \dots, w_{in}$ – вагові коефіцієнти нейронів.

Тоді пошук вагових коефіцієнтів шуканої функції (2) зводиться до примусового навчання нейронної мережі (або ансамблю моделей різної архітектури) на прикладах з бази даних, яке завершиться після досягнення заданої продуктивності (ймовірності розпізнавання класу) при припустимих помилках на навчальній та тестовій множинах прикладів. Формально цей



процес можна подати у символічному вигляді [4]:

$$\max P(S, X) \text{ при } \delta \leq \delta_0, \quad (3)$$

де $P(S, X)$ – вирішальне правило розпізнавання класів S у просторі ознак X ; $s \in S$, S – множина рівнів білкового балансу станів суб'єкта дослідження, що розпізнаються; $x \in X$, X – множина вхідних факторів-симптомів суб'єкта; δ – рівень помилок на навчальній та тестовій множинах; δ_0 – допустимі помилки навчання моделей на навчальному та тестовому множинах.

Коли процес навчання буде завершено, мережа стає здатною реалізовувати оптимальний алгоритм розпізнавання образів за правилом перевірки гіпотез (правило Байєса) при подачі на її вхід вектора ознак (симптомів) досліджуваного суб'єкта. Формалізація цього процесу представлена в [4-6].

Сукупність вхідних чинників $X^n = \{x_1, x_2, \dots, x_n\} \subset X$ разом із алфавітом класів забезпечують реалізацію правила розпізнавання образів [3]:

$$\omega_g \in \Omega_k, \text{ якщо } L(\omega, \{\omega_g\}) = \sup_i L(\omega, \{\omega_i\}) \quad (4)$$

$$L(\omega, \{\omega_g\}) \rightarrow \omega_g \in \Omega_k,$$

де $X^n = \{x_1, x_2, \dots, x_n\} \subset X$ – сукупність вхідних факторів; $L(\omega, \{\omega_g\})$ – правило віднесення суб'єкта з набором відповідних ознак (ω_g) до відповідного класу $\{\omega\}$; $\{\omega\}$ – множина класів (рівнів білкового балансу) (p, g) у просторі ознак (k, l) при всіх їх можливих поєднаннях (ω_{pk}, ω_{gl}).

Відповідно до теорії та правила навчання мережі слід очікувати, що існує такий набір чисел H, n, α, v_i, u_i , при яких функція у апроксимується рядом (2) по всій області її визначення і може бути реалізована за допомогою тришарової нейронної мережі із заданою похибкою. Це фундаментальне положення, що дозволяє встановити детермінований аналітичний зв'язок усієї сукупності симптомів-ознак зі ступенем білкового балансу суб'єкта.

Формування навчальної вибірки.

Інформативні ознаки ступеня балансу білка (норма, дефіцит) в організмі представлені вектором значень у форматі репрезентативної вибірки прикладів у вигляді [1]: постійні набряки, різке і безпричинне зниження ваги, сонливість та занепад сил, слабкість, тривала діарея, біль у суглобах та кістках, пожовтіння шкірних покривів та білків очей, зниження імунітету, часті застуди. Функцію стану білкового балансу можна описати класами: 1 (норма), 2 (дефіцит).

При цьому наявність або відсутність симптомів дефіциту білка фіксується в номінальному вираженні (символами) і представлена таблично, як показано на фрагменті вибірки прикладів з 9-ма ознаками та 2-ма класами (рис.1).

Вхідні установки режимів роботи, методів навчання, архітектур та складності моделей вибираються на розсуд дослідника та можливостей пакета технічного аналізу.

Підлаштування вагових коефіцієнтів нейронів базується на переборі всіх можливих варіантів поєднань ваг синапсів за алгоритмом зворотного поширення помилки з різними модифікаціями [2-5].



	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1	1	0	1	1	1	0	1	0	0	1
2	0	0	0	1	1	0	1	0	0	0
3	0	0	0	0	0	1	0	0	0	0
4	0	0	0	0	0	0	0	1	1	0
5	1	0	0	1	1	0	0	0	0	0
6	1	1	0	1	1	1	1	1	0	1
7	0	1	0	1	1	1	1	1	1	1
8	1	0	0	0	0	0	0	0	0	0
9	1	1	1	0	0	0	0	0	1	0

Рис.1. Фрагмент навчальної вибірки

Таким чином, у результаті синтезу ансамблю моделей з примусовим навчанням на репрезентативній вибірці прикладів з ретроспективної множини бази даних представлені результати досліджень з оцінкою якості результатів моделювання. Експериментальні дослідження показали стійку збіжність процесу навчання до мінімальних помилок при домінуванні числа симптомів у рядках спостережень кожного класу станів суб'єктів (рис. 2).

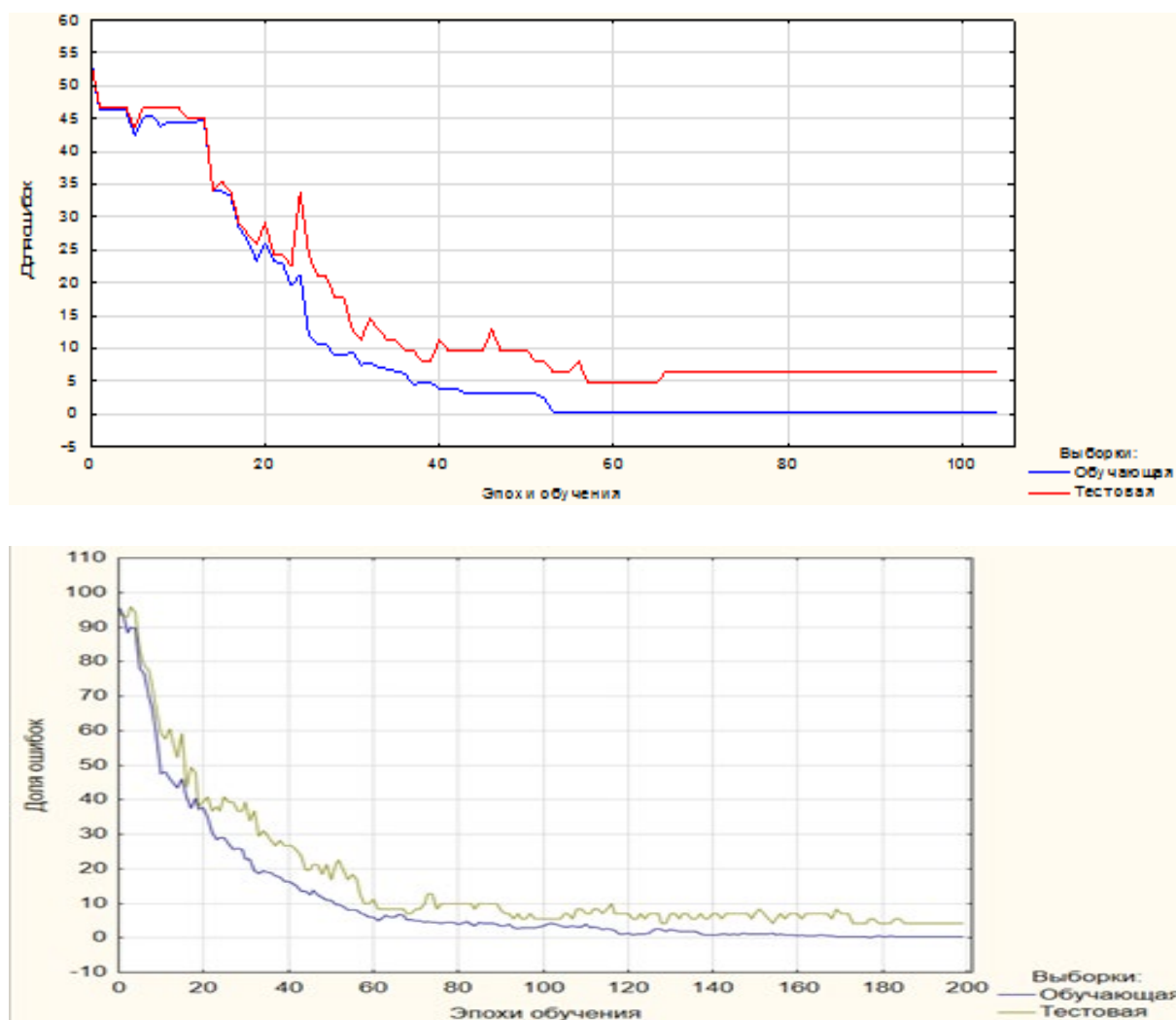


Рис 2. Графіки навчання нейромережевих моделей з різними архітектурами та складністю



У цьому випадку експериментально досягалися прийнятні для практики продуктивність та точність, що дозволяє говорити про спроможність результатів моделювання та доцільність практичного використання отриманих результатів для розпізнавання рівнів білкового балансу суб'єкта, що досліджується.

В результаті моделювання вдалося домогтися стійкої тенденції у збіжності ітераційного процесу навчання до мінімальних помилок на навчальних та тестових множинах. Продуктивність синтезованих моделей виявилася прийнятною для практичного застосування. Інструментально завдання вирішено на платформі пакета технічного аналізу даних.

Таким чином, інтелектуальний аналізатор білкового балансу в базисі штучних нейронних мереж практично реалізуємий, дозволяє досягти необхідних характеристик (точність та продуктивність), забезпечує оперативність аналізу та не вимагає додаткових матеріальних та фінансових витрат.

Висновки.

Для розпізнавання ступеня білкового балансу суб'єкта потрібно визначити функціональну залежність її від значень симптомів. Ця задача вирішена застосуванням технології примусового навчання нейромереж та реалізована моделями з різною архітектурою та складністю як задача розпізнавання образів.

Практична значимість результатів моделювання полягає у синтезі продуктивних моделей оцінки стану білкового балансу та можливості отримання додаткового незалежного каналу аналізу білкового балансу без необхідності використання традиційного методу біохімічного аналізу крові з його обмеженнями та протипоказаннями. Синтезований ансамбль моделей, алгоритмічний та програмний інструментарій дозволяють розширити застосування та знизити вимоги до обмежень, що мають місце при традиційних методах аналізу.

Функціонально програма синтезованих моделей може бути реалізована як самостійна програма в основному програмному коді пакету технічного аналізу, що розширює можливості практичної реалізації штучного інтелекту, розширює коло користувачів та доступність до більш ефективного механізму самоконтролю здоров'я громадян.

Література:

1. Остерман Л. А. Методы исследования белков и нуклеиновых кислот: электрофорез и ультрацентрифугирование (практическое пособие). – М.: Наука, 1981. – 288 с.
2. Хайкин, С. Нейронные сети: Полный курс. 2-е издание / Хайкин С. – М.: "Вильямс", 2006. – 1104 с.
3. Боровиков В.П. STATISTICA NN - Техническое описание. – М.: Мир, 1999. – 239 с.
4. Алёшин С.П. Нейросетевое моделирование процесса неинвазивной сатурации субъекта с обеспечением инвариантности к воздействию маскирующих факторов // С.П. Алёшин, О.М. Гайтан // Modern engineering and



innovative technologies. – Karlsruhe, Germany – 2021. – Issue 16. Part 2. – P. 83–89.

5. Алёшин С.П. Covid-19 Coronavirus Screening Analysis Neural Network Technology / С.П. Алёшин, И.В. Хоменко, Н.А. Фурсова // Системи управління, навігації та зв'язку. – 2021. – № 2 (64). – С. 53–57.

6. Алёшин С.П. Нейросетевой базис поддержки решений в пространстве факторов и состояний высокой размерности: [монография] / С.П. Алёшин. – Полтава: Издание «Скайтек», 2013. – 208 с.

Abstract. Protein is the main building material of the body, the norm of which must be constantly monitored to prevent its deficiency. One of the most accessible and common analyses of protein level is a biochemical blood test, which allows you to calculate the total amount of proteins, as well as its individual fractions. However, to obtain reliable results of the analysis, it is necessary to follow some mandatory rules. The medicine taken by the patient can also significantly affect the results of the analysis. For cases where it is impossible to conduct a standard protein analysis or to provide restrictions and fulfillment of rules, a technology for intelligent recognition of the protein balance level is proposed. The purpose of the study is to synthesize a neural network model for recognizing the protein balance level based on extracting knowledge from a representative sample of examples in the database of medical histories associated with a deficiency of protein fragments. The task of finding the functional dependence of the protein balance on the values of the symptoms was solved using the technology of forced learning of neural networks and implemented by models with different architectures as a problem of pattern recognition. The practical significance of the results is the possibility of obtaining an additional independent channel for the analysis of protein balance without need to use the traditional method of biochemical blood analysis with its limitations and contraindications. The synthesized ensemble of models, algorithmic and software tools make it possible to expand the application and reduce the requirements for the restrictions of traditional methods of analysis.

Keywords: protein analysis, intelligent analyzer, neural network, recognition, signs-symptoms.

Стаття відправлена: 20.08.2022 г.

© Альошин С.П., Гайтан О.М.