

**МЕДИЧНА ІНФОРМАТИКА
ТА ІНЖЕНЕРІЯ**

FAK 0--16 4 6-6.+

**ВИКОРИСТАННЯ ФОРМУЛИ БАЙЄСА ДЛЯ ПРОГНОЗУВАННЯ ВИНИКНЕННЯ
ЦУКРОВОГО ДІАБЕТУ**

С. М. Злепко⁷, Н. М. Сулова⁷, П. Г. Прудюс², С. В. Павлов⁷

Вінницький національний технічний університет@

Вінницький обласний клінічний ендокринологічний диспансер²

Стаття присвячена проблемі прогнозування цукрового діабету, зокрема діагностиці переддіабетичного стану. Розроблена математична модель прогнозування розвитку у пацієнта цукрового діабету I або II типу на основі діагностичного опитувальника.

Ключові слова: цукровий діабет, прогнозування, математичне моделювання, %ормула Байєса.

**ИСПОЛЬЗОВАНИЕ ФОРМУЛЫ БАЙЕСА ДЛЯ ПРОГНОЗИРОВАНИЯ
ВОЗНИКНОВЕНИЯ САХАРНОГО ДИАБЕТА**

С. М. Злепко⁷, Н. Н. Сулова⁷, Ф. Г. Прудюс², С. В. Павлов⁷

Вінницький національний технічний університет@

Вінницький обласний клінічний ендокринологічний диспансер²

Статья посвящена проблеме прогнозирования сахарного диабета, в частности диагностике преддиабетического состояния. Разработана математическая модель прогнозирования развития у пациента сахарного диабета I или II типа на основе диагностического опросника.

Ключевые слова: сахарный диабет, прогнозирование, математическое моделирование, %ормула Байеса.

**USING OF BAYES FORMULA FOR PROGNOSTICATION OF DIABETES
MELLITUS ORIGIN**

S. M. Zlepko⁷, N. M. Surova⁷, P. H. Prudyus², S. V. Pavlov⁷

Vinnitsia National Technical University@

Vinnitsia Regional Clinical Endocrinology Dispensary²

The article is devoted to the problem of diabetes mellitus prognostication, in particular to diagnostics of the pre-diabetic state. The mathematical model of diabetes mellitus of I or II type development prognostication is developed on the basis of diagnostic questionnaire.

Key words: difbetes mellitus, prognostication, mathematical model, Bayes formula.

Вступ. Останні роки розвитку суспільства характеризуються стрімким зростанням захворюваності на цукровий діабет в усьому світі, а в Україні — особливо. Питання профілактики та лікування цієї важкої патології набувають все більшої актуальності і потребують комплексного підходу, участі клініцистів різного профілю, а також представників неклінічних дисциплін.

Якщо розглянути рівні захворюваності на діабет на різних континентах, то можна зауважити, що в Європі, кількість населення якої приблизно дорівнює кількості

населення Африки, хворих на цукровий діабет значно більше, ніж в Африці.

Вважається, що в Україні близько одного мільйона хворих на цукровий діабет. Можна з впевненістю говорити про набагато більшу кількість хворих у дійсності. Що ж стосується розподілу захворювання по різних регіонах України, то найвища захворюваність реєструється в Києві та Київській області, а також у Донецькій області. З одного боку, це результат кращої діагностики хвороби в цих регіонах, з другого — впливу несприятливих факторів довкілля [1].

© С. М. Злепко, Н. М. Сулова, П. Г. Прудюс², С. В. Павлов

Багато людей не знають, що вони хворі на цукровий діабет або мають переддіабет, тобто порушення глікемії натще (НІН) або порушення толерантності до глюкози (НТГ); у частини таких хворих уже розвинулися характерні мікрovasкулярні зміни, що виникають як наслідок діабету. В усьому світі кількість людей з переддіабетом складає приблизно 314 млн., і, згідно з прогнозами, до 2025 року ця кількість зросте до 418 млн. Оскільки поширеність діабету і потреба у його прогнозуванні й надалі зростають, пов'язані з діабетом захворюваність і смертність стали основними проблемами системи охорони здоров'я.

Постановка проблеми. Переддіабет в 3—10 разів підвищує короткотривалий абсолютний ризик виникнення цукрового діабету, при чому у деяких популяціях цей ризик вищий, ніж в інших. Особи з цукровим діабетом уразливі до виникнення множинних і комплексних медичних ускладнень. До цих ускладнень належать як серцево-судинні захворювання (ССЗ) (ішемічна хвороба серця, інсульт, захворювання периферійних судин), так і мікросудинні ушкодження (а саме ретинопатія, нейропатія).

Епідеміологічні дані підтверджують, що ускладнення діабету починаються рано, прогресуючи від нормальної толерантності до глюкози до вираженого діабету. Рання ідентифікація і лікування осіб з переддіабетом може потенційно зменшити або затримати прогресування діабету і відстрочити ССЗ та мікросудинні ураження.

Діагноз переддіабет найчастіше робиться на підставі лише глікемічних критеріїв, а саме НГН і / або НТГ, причому адекватність порогових точок для встановлення цього діагнозу ще й досі є предметом для обговорення.

Особи з переддіабетичним станом мають три можливі варіанти розвитку подій за довготривалими спостереженнями: у одній третині розвивається діабет другого типу, у одній третині залишається переддіабетичний стан, у одній третині відновлюється нормоглікемія.

На сьогодні не існує затвердженої схеми лікування переддіабету та чіткої системи визначення даного діагнозу, зокрема прогнозування перебігу хвороби на майбутнє.

Мета роботи. Метою роботи є побудова математичної моделі прогнозування розвитку цукрового діабету, зокрема, визначення переддіабетичного стану на основі діагностичних тестів.

Для знаходження ймовірності переддіабетичного стану пацієнта розробляються математичні моделі

пацієнтів, хворих на цукровий діабет першого та другого типу, а також здорових людей на основі анкетних даних (діагностичних тестів) та формули Байєса.

Вирішення проблеми. Для створення системи прогнозування цукрового діабету будемо використовувати анкету, яка являє собою перелік питань і можливих варіантів відповідей. Діагностичні тести, за аналогією з [2], повинні бути побудовані за таким принципом:

1. Перша група питань — це питання, які дозволяють отримати інформацію про стан здоров'я пацієнта, на який не може вплинути лікування.

2. Друга група питань — це питання, які дозволяють отримати інформацію про стан здоров'я пацієнта, який підлягає частковому лікуванню (наприклад, алкоголізм).

3. Третя група питань — це питання, які дозволяють отримати інформацію про стан здоров'я пацієнта, який підлягає лікуванню.

Це пов'язано з тим, що у випадку прогнозування виникнення цукрового діабету, корегуючи питання другої та третьої груп математичної моделі пацієнтів, які хворіють на цукровий діабет, можна отримати оптимальний спосіб профілактичного лікування.

Серед цих питань повинна бути певна мінімальна, але достатня кількість таких, за наявності яких може працювати система прогнозування. Вони обов'язково повинні чітко визначати появу цукрового діабету першого та другого типу або ж його відсутність, а відповіді на них пацієнт повинен давати в першу чергу. Решта питань повинна містити додаткову інформацію, яка допомагає точніше визначити ймовірність переддіабетичного стану.

Анкету, заповнену пацієнтом, можна представити у вигляді матриці, наприклад:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}$$

де a_{ij} — j -тий варіант відповіді на i -те питання. Тобто першому питанню відповідають варіанти відповідей $a_{11}, a_{12}, \dots, a_{1m}$, другому — $a_{21}, a_{22}, \dots, a_{2m}$, n -му — $a_{n1}, a_{n2}, \dots, a_{nm}$.

На кожне питання може бути лише один варіант відповіді, тобто якщо пацієнт вибрав відповідь $a_{..}$ на i -те питання, то цьому елементу матриці відповідатиме одиниця, а всім іншим варіантам автоматично присвоюється значення нуля.

Заповнена анкета матиме вигляд, наприклад, такий:

$$\begin{pmatrix} 0 & 1 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

тобто на перше питання пацієнт обрав другий варіант відповіді, на друге - перший варіант відповіді, на n -те - t -тий варіант відповіді.

Нехай для побудови системи прогнозування цукрового діабету було опитано наступних пацієнтів:

- n_1 - кількість досліджуваних хворих на цукровий діабет 1-го типу;
- n_2 - кількість досліджуваних хворих на цукровий діабет 2-го типу;
- $n_3 = n_1 + n_2$ - кількість досліджуваних хворих на цукровий діабет;
- n_4 - кількість досліджуваних здорових пацієнтів.

На основі даних анкет досліджуваних пацієнтів побудуємо:

H_1 - математичну модель пацієнта, за допомогою якої можна отримати прогноз, що у нього на протязі 5 років може виникнути цукровий діабет 1-го типу і яка сформована на основі інформації про n_1 ;

H_2 - математичну модель пацієнта, за допомогою якої можна отримати прогноз, що у нього на протязі 5 років може виникнути цукровий діабет 2-го типу і яка сформована на основі інформації про n_2 ;

H_3 - математичну модель пацієнта, за допомогою якої можна отримати прогноз, що у нього на

$$\begin{pmatrix} \frac{c_{11}}{n_2} & \frac{c_{12}}{n_2} & \dots & \frac{c_{1m}}{n_2} \\ \frac{c_{21}}{n_2} & \frac{c_{22}}{n_2} & \dots & \frac{c_{2m}}{n_2} \\ \dots & \dots & \dots & \dots \\ \frac{c_{n1}}{n_2} & \frac{c_{n2}}{n_2} & \dots & \frac{c_{nm}}{n_2} \end{pmatrix} \begin{pmatrix} \frac{b_{11} + c_{11}}{n_1 + n_2} & \frac{b_{12} + c_{12}}{n_1 + n_2} & \dots & \frac{b_{1m} + c_{1m}}{n_1 + n_2} \\ \frac{b_{21} + c_{21}}{n_1 + n_2} & \frac{b_{22} + c_{22}}{n_1 + n_2} & \dots & \frac{b_{2m} + c_{2m}}{n_1 + n_2} \\ \dots & \dots & \dots & \dots \\ \frac{b_{n1} + c_{n1}}{n_1 + n_2} & \frac{b_{n2} + c_{n2}}{n_1 + n_2} & \dots & \frac{b_{nm} + c_{nm}}{n_1 + n_2} \end{pmatrix} \begin{pmatrix} \frac{d_{11}}{n_4} & \frac{d_{12}}{n_4} & \dots & \frac{d_{1m}}{n_4} \\ \frac{d_{21}}{n_4} & \frac{d_{22}}{n_4} & \dots & \frac{d_{2m}}{n_4} \\ \dots & \dots & \dots & \dots \\ \frac{d_{n1}}{n_4} & \frac{d_{n2}}{n_4} & \dots & \frac{d_{nm}}{n_4} \end{pmatrix}$$

Формула Байєса має вигляд [3]:

$$P(H_i / A) = \frac{P(H_i) \cdot P(A / H_i)}{\sum_{k=1}^n P(H_k) \cdot P(A / H_k)}$$

$H_i, i = 1, 4$ - визначені вище, а подією A означимо те, що пацієнт взяв анкету і заповнив її, а в результаті

протязі 5 років може виникнути цукровий діабет якого-небудь типу і яка сформована на основі інформації про n_3 ;

H_4 - математичну модель пацієнта, за допомогою якої можна отримати прогноз, що він на протязі 5 років не захворіє на цукровий діабет і яка сформована на основі інформації про n_4 .

Математичні моделі $H_i, i = 1, 4$ являють собою матриці, елементами яких є імовірності, що знайдені на основі інформації про хворих або здорових людей відповідно. Так, математична модель H_1 матиме вигляд:

$$\begin{pmatrix} \frac{b_{11}}{n_1} & \frac{b_{12}}{n_1} & \dots & \frac{b_{1m}}{n_1} \\ \frac{b_{21}}{n_1} & \frac{b_{22}}{n_1} & \dots & \frac{b_{2m}}{n_1} \\ \dots & \dots & \dots & \dots \\ \frac{b_{n1}}{n_1} & \frac{b_{n2}}{n_1} & \dots & \frac{b_{nm}}{n_1} \end{pmatrix}$$

тобто B - кількість пацієнтів, які дали позитивну відповідь на 1-ий варіант 1-го питання, b - кількість пацієнтів, які дали позитивну відповідь на 2-ий варіант 1-го питання, $B_{пт}$ - кількість пацієнтів, які дали позитивну відповідь на m -ий варіант p -го питання.

Аналогічно будуються й моделі H_2, H_3, H_4 відповідно:

ми отримали матрицю діагностичного тесту, з якою будемо працювати далі.

Будемо обчислювати наступні ймовірності:

$P(H_i / A)$ - ймовірність того, що матриця, заповнена пацієнтом, збігається з математичною моделлю H_i тобто існує ймовірність, що пацієнт захворіє на цукровий діабет першого типу;

$P(H_3/A)$ - ймовірність того, що матриця, заповнена пацієнтом, збігається з математичною моделлю H_2 , тобто існує ймовірність, що пацієнт захворіє на цукровий діабет другого типу;

$P(H_3/A)$ - ймовірність того, що матриця, заповнена пацієнтом, збігається з математичною моделлю H тобто існує ймовірність того, що пацієнт захворіє на цукровий діабет будь-якого типу;

$P(H_4/A)$ - ймовірність того, що матриця, заповнена пацієнтом, збігається з математичною моделлю H_4 ,

тобто існує ймовірність того, що пацієнт не захворіє на цукрових діабет.

Для формування прогнозу спочатку обчислюють ймовірності $P(H/A)$ та $P(H_4/A)$ для визначення того, чи потрапляє пацієнт в групу ризику захворіти на цукровий діабет, і, якщо потрапляє, то знаходять ймовірності $P(H_3/A)$ та $P(H_4/A)$, що дозволяє визначити тип цукрового діабету.

Розглянемо перший етап прогнозування:

$$P(H_4 / A) = \frac{P(H_4) \cdot P(A / H_4)}{P(H_3) \cdot P(A / H_3) + P(H_4) \cdot P(A / H_4)}, \quad P(H_3 / A) = 1 - P(H_4 / A)$$

де $P(H_3) = \frac{n_3}{n_3 + n_4}$, $P(H_4) = \frac{n_4}{n_3 + n_4}$, $P(H_3/A)$ та $P(H_4/A)$ визначаються наступним чином: якщо матри-

ця H_3 має вигляд

$$\begin{pmatrix} \frac{b_{11} + c_{11}}{n_1 + n_2} & \frac{b_{12} + c_{12}}{n_1 + n_2} & \dots & \frac{b_{1m} + c_{1m}}{n_1 + n_2} \\ \frac{b_{21} + c_{21}}{n_1 + n_2} & \frac{b_{22} + c_{22}}{n_1 + n_2} & \dots & \frac{b_{2m} + c_{2m}}{n_1 + n_2} \\ \dots & \dots & \dots & \dots \\ \frac{b_{n1} + c_{n1}}{n_1 + n_2} & \frac{b_{n2} + c_{n2}}{n_1 + n_2} & \dots & \frac{b_{nm} + c_{nm}}{n_1 + n_2} \end{pmatrix}, \text{ а матриця, заповнена пацієнтом - } \begin{pmatrix} 0 & 1 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

$$P(A / H_3) = \frac{b_{12} + c_{12}}{n_1 + n_2} \cdot \frac{b_{21} + c_{21}}{n_1 + n_2} \cdot \dots \cdot \frac{b_{nm} + c_{nm}}{n_1 + n_2} \quad \text{Аналогічно визначається і} \quad P(A / H_4) = \frac{d_{12}}{n_4} \cdot \frac{d_{21}}{n_4} \cdot \dots \cdot \frac{d_{nm}}{n_4}$$

Отже, врахувавши, що $n_3 = n_1 + n_2$, отримаємо наступну формулу для обчислення ймовірності того,

що пацієнт на протязі наступних 5-ти років не захворіє на цукровий діабет:

$$P(H_4 / A) = \frac{\frac{n_4}{n_1 + n_2 + n_4} \cdot \left(\frac{d_{12}}{n_4} \cdot \frac{d_{21}}{n_4} \cdot \dots \cdot \frac{d_{nm}}{n_4} \right)}{\frac{n_1 + n_2}{n_1 + n_2 + n_4} \cdot \left(\frac{b_{12} + c_{12}}{n_1 + n_2} \cdot \frac{b_{21} + c_{21}}{n_1 + n_2} \cdot \dots \cdot \frac{b_{nm} + c_{nm}}{n_1 + n_2} \right) + \frac{n_4}{n_1 + n_2 + n_4} \cdot \left(\frac{d_{12}}{n_4} \cdot \frac{d_{21}}{n_4} \cdot \dots \cdot \frac{d_{nm}}{n_4} \right)}, \quad (1)$$

а ймовірність того, що пацієнт на протязі 5-ти наступних років захворіє на цукровий діабет, визначається за формулою:

$$P(H_3 / A) = 1 - P(H_4 / A) \quad (2)$$

На другому етапі прогнозування виконують розра-

хунки, аналогічні розрахункам на першому етапі, використовуючи дані матриць H_1 , H і матрицю, заповнену пацієнтом, в результаті чого отримуємо формулу для обчислення ймовірності того, що пацієнт на протязі 5 -ти наступних років захворіє на цукровий діабет 1-го типу:

$$P(H_1 / A) = \frac{\frac{n_1}{n_1 + n_2} \cdot \left(\frac{b_{12}}{n_1} \cdot \frac{b_{21}}{n_1} \cdot \dots \cdot \frac{b_{nm}}{n_1} \right)}{\frac{n_2}{n_1 + n_2} \cdot \left(\frac{c_{12}}{n_2} \cdot \frac{c_{21}}{n_2} \cdot \dots \cdot \frac{c_{nm}}{n_2} \right) + \frac{n_1}{n_1 + n_2} \cdot \left(\frac{b_{12}}{n_1} \cdot \frac{b_{21}}{n_1} \cdot \dots \cdot \frac{b_{nm}}{n_1} \right)}, \quad (3)$$

тоді ймовірності того, що пацієнт на протязі 5-ти наступних років захворіє на цукровий діабет 2-го типу визначається за формулою:

$$P(H_2 / A) = 1 - P(H_1 / A) \quad (4)$$

Дані ймовірності обчислюються на основі даних досліджуваної групи людей, в результаті чого формуються проміжки (α_1, α_2) , (α_2, α_3) , (α_3, α_4) , (β_1, β_2) , (β_2, β_3) , (β_3, β_4) : (α_1, α_2) - інтервал, до якого належать пацієнти, які на протязі 5-ти років не захворіють на цукровий діабет; (α_2, α_3) - інтервал, до якого належать пацієнти, які можуть як захворіти, так і не захворіти на цукровий діабет; (α_3, α_4) - інтервал, до якого належать пацієнти, які на протязі 5-ти років можуть захворіють на цукровий діабет; (β_1, β_2) - інтервал, до якого належать пацієнти, які на протязі 5-ти років можуть захворіти на цукровий діабет першого типу; (β_2, β_3) - інтервал, до якого належать пацієнти, які на протязі 5-ти років можуть захворіти на цукровий діабет як першого, так і другого типу; (β_3, β_4) - інтервал, до якого належать пацієнти, які на протязі 5-ти років можуть зах-

Література

1. Діагностика і лікування переддіабету в континуумі гіперглікемії: коли починається ризик розвитку цукрового діабету? / Погоджена думка Американського коледжу ендокринології і Американської асоціації клінічних ендокринологів // Ліки України: науково-практичний журнал для лікарів та провізорів. - 2009. - № 1. - С. 8-15.
2. Мисюк И. С. Система математического прогнозирова-

воріти на цукровий діабет другого типу. Таким чином, інтервали ймовірностей (α_2, α_3) та (α_3, α_4) визначають переддіабетичний стан.

Висновки. На основі даних діагностичного тесту запропонованої структури спочатку обчислюється ймовірність того, що пацієнт не захворіє на протязі 5-ти років (за допомогою формули (1)); ймовірність того, що пацієнт захворіє на протязі 5-ти років (формула (2)); якщо ж пацієнт потрапляє до групи ризику захворіти на цукровий діабет (у нього діагностується переддіабетичний стан), то обчислюються ймовірності виникнення 1-го та 2-го типів цукрового діабету (формули (3) і (4)). На основі отриманих ймовірностей та встановлених інтервальних проміжків робиться висновок про здоров'я пацієнта.

Для визначення правильного профілактичного лікування використовуються питання анкети другого та третього типів. Для цього в формулах (3) та (4) послідовно замінюються значення ймовірностей математичних моделей H_1 та H_2 на значення ймовірностей математичної моделі H_4 , після чого знову визначають ймовірність захворювання на цукровий діабет. Спочатку вказану вище процедуру виконують з відповідями на питання третього типу, оскільки вони містять інформацію про здоров'я пацієнтів, які найкраще піддаються профілактичному лікуванню, а потім - з відповідями на питання другого типу.

3. Мисюк И. С. Система математического прогнозирования мозговых инсультов. Метод рекомбинации / И. С. Мисюк, А. Е. Семак, В. И. Адамович. - Минск: Изд-во БГУ им. В. И. Ленина, 1982. - 21 с.
3. Сеньо П. С. Теорія ймовірності та математична статистика: підручник для вищих навч. закладів / П. С. Сеньо. - 2-ге вид., перероб. і доп. - К.: Знання, 2007. - 556 с.