SINFLARNI MULTISINFLARGA AJRATISH VA PRETSEDENT OB'EKTLAR ORQALI IFODALASH

Ro'zibayev Ortiq Baxtiyorovich[1], Baydullayev Ruslan Tuylibayevich[1]

[1]Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti
E-mail:o.ruzibaev@tuit.uz

K E Y W O R D S

Belgi, F1 o'lchovi (F1-score), Davies–Bouldin indeksi (DBI), ROC-AUC, PR-AUC, fenotipik, sezgirlik (sensitivity), o'ziga xoslik (specificity), muvozanatlangan aniqlik (balanced accuracy), kross-validatsiya, pretsedent, multisinf, ichki jipslik, fenotip

A B S T R A C T

Ushbu maqolada tibbiy tashxislash masalalarida multisinf–pretsedent yondashuvi asosida bemorlarni tasniflashning nazariy asoslari ishlab chiqilgan. Tibbiy tashxis sinflarining ichki geterogenligini hisobga olish maqsadida har bir tashxis sinfini nisbatan bir jinsli multisinflarga ajratish hamda har bir multisinf uchun tipik pretsedent obyektni tanlash usuli taklif etilgan. Pretsedentlarni tanlashning matematik modeli qamrov, ichki jipslik va multisinflararo ajraluvchanlik mezonlari asosida formal ifodalangan. Shuningdek, pretsedentning klinik muvofiqligi va ma'lumotlar sifati hisobga olinadigan integrallashgan baholash mezoni ishlab chiqilgan. Yangi bemorlarni pretsedent obyektlar bilan o'xshashligi asosida tasniflash, qaror ishonchliligini baholash hamda informativ belgilarni tanlash uchun multisinf–pretsedent modelidan foydalanish imkoniyatlari tahlil qilingan.

KIRISH

Tibbiy ma'lumotlar bilan ishlashda bemorlar to'g'risidagi ma'lumotlarni to'g'ri klasterlash va tasniflash, kasalliklarni erta aniqlash hamda samarali qaror qabul qilish jarayonlari uchun hal qiluvchi ahamiyatga ega. Klassik klasterlash usullari ko'pincha ob'ektlarni faqat bitta klasterga mansub deb qarasa, zamonaviy yondashuvlar, jumladan, multiklasterlash, bir ob'ektning bir nechta sinf bilan bog'liq bo'lishi mumkinligini inobatga oladi [1,2].

Multiklasterlash jarayonida hal qiluvchi bosqichlardan biri, bu president ob'ektlarni tanlashdir. President ob'ektlar, klasterlarning namunali vakillari bo'lib, ular asosida boshqa ob'ektlarning klasterga mansubligi baholanadi. Ushbu yondashuv ob'ektlarning noaniq va ko'p o'lchamli fazoda aniqroq ajratilishiga xizmat qiladi [3–5].

Tibbiy tashxis sinfi ko'pincha ichki jihatdan bir jinsli emas. Bir xil nozologik tashxisga ega bemorlarda kasallikning erta va kech bosqichlari, yengil va og'ir kechishi, turli yosh guruhlari, hamroh kasalliklar, laborator fenotiplar yoki tasviriy morfotiplar kuzatilishi mumkin. Agar bunday sinf yagona markaz bilan ifodalansa, markaz real bemorga mos kelmasligi, ayrim lokal guruhlar esa noto'g'ri ajratilishi mumkin. Prototip yoki pretsedent tanlash usullari sinfni real ob'ektlar orqali ifodalash, yaqin qo'shnilar usullarining hisoblash hajmini qisqartirish va qarorni tushunarli qilish uchun qo'llanadi [5, 7, 14].

ASOSIY QISM

Mazkur tadqiqotda har bir X_p sinfni uning ichki tuzilishiga mos ravishda r_p ta multisinfga ajratish qabul qilinadi:

$$X_p = \bigcup_{q=1}^{r_p} X_{pq}, \qquad X_{pq} \cap X_{pt} = \varnothing \ (q \neq t), \ (1)$$

bu yerda X_{pq} – p-sinfning q-multisinfi; r_p – p-sinf tarkibida aniqlangan multisinflar soni. Har bir multisinf

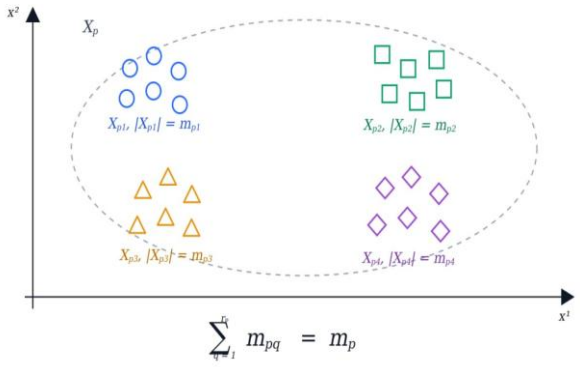
$$X_{pq} = \{x_{pq1}, x_{pq2}, \dots, x_{pqm_{pq}}\}, \qquad q = \overline{1, r_p}, \quad (2)$$

ko'rinishida ifodalanadi va

$$\sum_{q=1}^{r_p} m_{pq} = m_p \qquad (3)$$

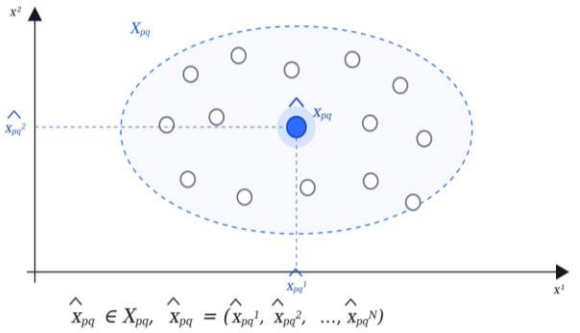
(cc) BY

tenglik bajariladi.



1-rasm. X_p sinfning multisinflarga ajralishi.

«Multisinf» atamasi bu yerda yangi tashxisiy sinfni emas, mavjud p tashxis sinfining ichki, nisbatan bir jinsli qismini anglatadi. Masalan, qandli diabet sinfida insulin rezistentligi yuqori, semirish bilan bog‘liq fenotip; normal vaznli, lekin sekresiya yetishmovchiligi ustun fenotip; asoratlari rivojlangan fenotip alohida multisinflar sifatida shakllanishi mumkin. Jarrohlik stomatologiyasida esa bir tashxis sinfi yallig‘lanishning joylashuvi, tarqalish darajasi, yosh, laborator faollik yoki tasviriy belgilarga ko‘ra ichki guruhlarga ajralishi mumkin. Bunday ifodalash tashxisni o‘zgartirmaydi, balki uning ma’lumotlar fazosidagi ko‘p modalli tuzilishini ochib beradi.



2-rasm. Multisinfning tipik ob’ekti ko‘rinishi

Har bir X_{pq} multisinf ob’ektlari orasidan ushbu guruhning tipik holatini ifodalovchi pretsedent ob’ekt tanlanadi:

$$\hat{x}_{pq} \in X_{pq}, \qquad \hat{x}_{pq} = \left(\hat{x}_{pq}^1, \hat{x}_{pq}^2, \ldots, \hat{x}_{pq}^N\right). \quad (4)$$

Pretsedent ob’ektning real o‘quv ob’ekti bo‘lishi muhim.

O‘rtacha yoki sintetik markazdan farqli ravishda, u klinik jihatdan mavjud bemor holatiga mos keladi, uning belgilar kombinatsiyasi fiziologik va mantiqiy jihatdan tekshirilishi mumkin. Shu sababli pretsedent ob’ekt «etalon bemor» emas, balki muayyan multisinfning eng tipik vakillaridan biri sifatida talqin qilinadi.

Pretsedent tanlashda ikki qarama-qarshi talab bir vaqtda bajarilishi kerak. Birinchi talab - **qamrov va ichki jipslik**: $\hat{x}_{pq}$ o‘z multisinfidagi boshqa ob’ektlarga yetarlicha o‘xshash bo‘lishi lozim. Ikkinchi talab – **ajraluvchanlik**: bir X_p sinfning turli multisinflari pretsedentlari bir-biriga haddan tashqari o‘xshash bo‘lmasligi kerak. Ushbu talablar umumiy ko‘rinishda

$$\nu\left(\hat{x}_{pq}, x_{pqk}\right) \geq \delta, \qquad k = \overline{1, m_{pq}}, \quad (5)$$

va

$$\nu\left(\hat{x}_{pi}, \hat{x}_{pt}\right) \leq \frac{\delta}{3}, \qquad i \neq t, \quad i, t = \overline{1, r_p}, \quad (6)$$

shartlari orqali ifodalanadi. Bu yerda $\nu(\cdot,\cdot)$ – ob’ektlar o‘xshashlik darajasi; $\delta \in (0{,}100]$ – multisinf ichida qabul qilinadigan minimal o‘xshashlik foizi. $\delta/3$ chegarasi turli pretsedentlar o‘rtasidagi o‘xshashlik uchun qat’iyroq cheklov bo‘lib, ichki jipslik bilan tashqi ajralish orasida aniq marja hosil qiladi.

$\delta/3$ qiymatini mutlaq universal konstanta sifatida emas, bazaviy nazariy nisbat sifatida qarash lozim. Amaliy ma’lumotda u klinik ekspert bahosi, ichki validatsiya yoki qabul qiluvchi operatsion xarakteristika asosida moslashtirilishi mumkin. Masalan, $\gamma \in (0{,}1)$ ajralish koeffitsienti kiritilib,

$$\nu\left(\hat{x}_{pi}, \hat{x}_{pt}\right) \leq \gamma\delta \qquad (7)$$

deb olinadi; bazaviy holda $\gamma = 1/3$. Agar multisinflar juda yaqin fenotiplarni ifodalasa, γ biroz kattaroq; klinik jihatdan keskin farqli fenotiplarda esa kichikroq tanlanishi mumkin. Biroq dissertatsiyada asosiy belgilashlar izchilligini saqlash uchun keyingi formulalarda $\delta/3$ shart qabul qilinadi.

Pretsedent tanlash masalasini to‘liq formallashtirish uchun nomzod $z \in X_{pq}$ ning qamrovini quyidagicha belgilaymiz:

(cc) BY

$$Cov_{pq}(z;\delta) = \frac{1}{m_{pq}} \sum_{k=1}^{m_{pq}} \mathbf{1}\{v(z, x_{pqk}) \geq \delta\}. \quad (8)$$

Agar $\mathrm{Cov_{pq}}(z;\delta) = 1$ bo‘lsa, z multisinfning barcha ob’ektlarini belgilangan darajada qoplaydi. Real tibbiy ma’lumotlarda shovqin, noto‘g‘ri o‘lchov yoki chegaraviy bemorlar sababli to‘liq qamrov doim ham mumkin emas. Shu bois yumshatilgan shart

$$Cov_{pq}(z;\delta) \geq \eta, \qquad 0 < \eta \leq 1, \quad (9)$$

kiritilishi mumkin. Masalan, $\eta = 0.9$ bo‘lsa, pretsedent multisinf ob’ektlarining kamida 90 foizini qoplashi talab etiladi. Bu yondashuv yakka shovqinli ob’ektlar tufayli butun guruh uchun pretsedent topilmay qolishining oldini oladi.

Nomzodning ichki tipiklik ko‘rsatkichi o‘rtacha o‘xshashlik orqali

$$T_{pq}(z) = \frac{1}{m_{pq}-1} \sum_{\substack{k=1 \\ x_{pqk} \neq z}}^{m_{pq}} v\left(z, x_{pqk}\right) \quad (10)$$

aniqlanadi. Turli multisinf pretsedentlariga nisbatan ajralish esa

$$D_{pq}(z) = 1 - \max_{t \neq q} \frac{v(z, \hat{x}_{pt})}{100} \quad (11)$$

ko‘rinishida baholanadi. Shunda pretsedent tanlashning birlashgan mezoni

$$\hat{x}_{pq} = arg\max_{z \in X_{pq}} \left[\alpha \frac{T_{pq}(z)}{100} + (1 - \alpha) D_{pq}(z)\right], \qquad 0 \leq \alpha \leq 1, \quad (12)$$

bo‘lishi mumkin. Agar birinchi navbatda multisinfning tipik vakili talab etilsa α katta, pretsedentlar orasidagi farq muhimroq bo‘lsa α kichik tanlanadi. Biroq nomzod (5) va (6) qattiq shartlarni ham qanoatlantirishi kerak.

Tibbiy jihatdan pretsedent ob’ektga quyidagi mazmuniy talablar qo‘yiladi:

- uning tashxisi klinik va laborator yoki instrumental dalillar bilan tasdiqlangan bo‘lishi;
- asosiy belgilarda yetishmaydigan qiymatlar minimal bo‘lishi;
- davolashdan keyingi yoki natija haqida axborot beruvchi belgilarni o‘z ichiga olmasligi;
- o‘lchov xatosi, favqulodda artefakt yoki ma’lumot kiritish xatosiga ega bo‘lmasligi;
- guruh ichidagi ko‘pchilik ob’ektlarga yaqin va boshqa multisinflarga nisbatan ajraluvchan bo‘lishi;
- klinik ekspert tomonidan «tipik holat» sifatida asoslanishi mumkin bo‘lishi.

Ushbu talablar matematik mezonni klinik nazorat bilan birlashtiradi. Faqat masofaga asoslangan tanlash ba’zan ma’lumotlar markazida joylashgan, ammo klinik jihatdan kam uchraydigan kombinatsiyaga ega ob’ektni tanlashi mumkin. Shuning uchun pretsedentning matematik reytingi bilan birga ma’lumot sifati va klinik muvofiqlik indeksi □(□) ∈ [*0,1*] kiritilib,

$$Q_{pq}(z) = e(z)\left[\alpha \frac{T_{pq}(z)}{100} + (1-\alpha) D_{pq}(z)\right] \quad (13)$$

mezoni qo‘llanishi mumkin.

Multisinflarni shakllantirishning o‘zi alohida masala hisoblanadi. Agar $\mathrm{r_p}$ oldindan ma’lum bo‘lsa, ob’ektlar klinik qoidalar, ierarxik klasterlash, k-medoids, aralash ma’lumotlar uchun k-prototypes yoki graf usullari asosidagi guruhlash orqali ajratilishi mumkin [7, 15]. Agar $\mathrm{r_p}$ noma’lum bo‘lsa, uni Silhouette, Davies–Bouldin indeksi (DBI), gap statistic kabi ichki mezonlar bilan birga pretsedent shartlari asosida tanlash maqsadga muvofiq. Bunda eng kichik $\mathrm{r_p}$ qiymat shunday tanlanadiki, har bir hosil bo‘lgan $\mathrm{X_{pq}}$ uchun (5) shartni qanoatlantiruvchi pretsedent mavjud bo‘lsin va turli pretsedentlar (6) bo‘yicha ajralsin.

Formal ravishda p-sinf uchun maqbul multisinflar soni

$$r_p^* = min\left\{r_p: \forall q \ \exists \hat{x}_{pq} \in X_{pq}, \ \min_k v(\hat{x}_{pq}, x_{pqk}) \geq \delta, \ \max_{i \neq t} v(\hat{x}_{pi}, \hat{x}_{pt}) \leq \frac{\delta}{3}\right\} \quad (14)$$

sifatida aniqlanishi mumkin. Ushbu mezon ortiqcha mayda klasterlar hosil qilishni cheklaydi. Agar har bir ob’ekt alohida multisinf qilinsa, ichki o‘xshashlik avtomatik yuqori bo‘ladi, ammo model umumlashtirish xususiyatini yo‘qotadi.

(cc) BY

Minimal r_p^* prinsipi esa sinf tuzilishini imkon qadar sodda saqlaydi.

Pretsedentlar asosida yangi ob'ektni tasniflash qoidasi ham tabiiy ravishda hosil bo'ladi. Noma'lum

$$x = (x^1, x^2, \dots, x^N)$$

ob'ekt uchun barcha pretsedentlarga o'xshashlik hisoblanadi va eng yaqin multisinf

$$(p^*, q^*) = \underset{p,q}{argmax} v\left(\hat{x}_{pq}, x\right) \qquad (15)$$

orqali tanlanadi. Agar

$$v\left(\hat{x}_{p^*q^*}, x\right) \geq \delta, \qquad (16)$$

bo'lsa, ob'ekt p^* tashxis sinfiga va q^* ichki fenotipga tegishli deb qabul qilinadi. Agar maksimal o'xshashlik δ dan past bo'lsa, «noaniq», «qo'shimcha tekshiruv talab qilinadi» yoki «o'quv tanlanmada yetarli analog mavjud emas» degan qaror chiqariladi. Tibbiy qarorni majburiy sinflashtirishdan ko'ra, ishonchsiz holatlarni alohida belgilash xavfsizroqdir.

Bir nechta sinf pretsedentlariga o'xshashlik yuqori bo'lgan holatda qaror marjasi hisoblanadi. Eng katta va ikkinchi katta o'xshashliklar $\nu_{(1)}(x)$ va $\nu_{(2)}(x)$ bo'lsa,

$$M(x) = \nu_{(1)}(x) - \nu_{(2)}(x) \qquad (17)$$

marja sifatida olinadi. $M(x)$ kichik bo'lsa, bemor ikki fenotip yoki tashxis chegarasida joylashgan bo'lishi mumkin. Bunday holatda qo'shimcha klinik belgi yoki tekshiruvni tavsiya qilish mumkin. Demak, pretsedent modeli faqat sinf raqamini emas, qarorning tushunarli asosini – qaysi tipik holatga qanchalik o'xshashligini ham ko'rsatadi.

Multisinf–pretsedent ifodalashning informativ belgilarni tanlash uchun asosiy afzalligi shundaki, belgi qiymati ikki xil funksiyani bajaradi. Birinchidan, u bir multisinf ichida pretsedent va ob'ektlarni birlashtirishi kerak; ikkinchidan, turli multisinf pretsedentlarini farqlashi lozim. Masalan, bir laborator marker barcha X_{pq} ob'ektlarida o'xshash bo'lsa, ichki jipslikka hissa qo'shadi. Agar shu marker boshqa X_{pt} pretsedentida ham xuddi shunday bo'lsa, ajratishga hissa qo'shmaydi. Aksincha, guruh ichida beqaror, ammo pretsedentlar orasida farqli belgi ham yetarli emas, chunki u yangi ob'ektni o'z pretsedentiga ishonchli bog'lamaydi. Informativ belgi ikki shartni bir vaqtda ta'minlashi kerak.

Shu nuqtai nazardan har bir j-belgi uchun ichki muvofiqlik ko'rsatkichi

$$a_j = \frac{1}{\sum_{p=1}^{r}\sum_{q=1}^{r_p}(m_{pq}-1)} \sum_{p=1}^{r}\sum_{q=1}^{r_p}\sum_{\substack{k=1 \\ x_{pqk}\neq\hat{x}_{pq}}}^{m_{pq}} \rho_{pq}^{j}\left(\hat{x}_{pq}, x_{pqk}\right) \qquad (18)$$

va pretsedentlararo ajratish ko'rsatkichi

$$b_j = 1 - \frac{1}{\sum_{p=1}^{r}\binom{r_p}{2}} \sum_{p=1}^{r}\sum_{1\leq i<t\leq r_p} \rho_{p,it}^{j}\left(\hat{x}_{pi}, \hat{x}_{pt}\right) \qquad (19)$$

aniqlanadi. Bu yerda ρ^j – j-belgi bo'yicha binar o'xshashlik indikatori. a_j katta bo'lsa, belgi multisinf ichida barqaror; b_j katta bo'lsa, turli pretsedentlarni ajratadi. Keyingi paragrafda ρ^j nominal va sonli belgilar uchun aniq formulalar bilan quriladi.

Pretsedent ifodalashni tibbiy ekspert tizimi nuqtai nazaridan ham talqin qilish mumkin. Har bir $\hat{x}_{pq}$ ma'lum tashxisning klinik portretini, ρ^j esa bemorning shu portret bilan j–ko'rsatkich bo'yicha mos kelishini ko'rsatadi. ν mos kelgan belgilar ulushini ifodalaydi. Shunday qilib, qaror «model shunday dedi» shaklida emas, «bemorning tanlangan ℓ ta belgisidan s tasi muayyan pretsedentga mos keldi, eng yaqin raqib pretsedentga esa faqat t tasi mos keldi» tarzida tushuntirilishi mumkin. Bu xususiyat klinik ishonch, audit va xatolar tahlili uchun muhim.

Biroq pretsedent modeli samarali ishlashi uchun aralash belgilar bo'yicha o'xshashlikni bir xil shkalada ifodalash zarur. Nominal belgilarda qiymatlar tengligi tabiiy mezon bo'lsa, sonli belgilarda mutlaq tenglik deyarli uchramaydi va klinik qabul qilinadigan farq chegarasi kerak. Shuning uchun har bir multisinf va belgi uchun adaptiv bo'sag'aviy qiymat shakllantiriladi. Ushbu mexanizm keyingi paragrafda batafsil asoslanadi.

XULOSA

Ushbu maqolada tibbiy tashxislash masalalarida bemorlar to‘g‘risidagi ma’lumotlarni samarali tasniflash uchun multisinf–pretsedent yondashuvining nazariy asoslari ishlab chiqildi. Tadqiqotda an’anaviy yagona markazga asoslangan ifodalash usullari tibbiy sinflarning ichki geterogenligini to‘liq aks ettira olmasligi asoslab berildi hamda har bir tashxis sinfini nisbatan bir jinsli multisinflarga ajratish orqali ushbu muammoni bartaraf etish taklif qilindi. Har bir multisinf uchun tipik pretsedent obyektni tanlash sinfning real klinik vakili bilan ishlash, qarorlarni tushuntirish va yangi bemorlarni ishonchli tasniflash imkoniyatini yaratishi ko‘rsatildi.

Maqolada pretsedent obyektni tanlashning matematik modeli ishlab chiqilib, uning ichki jipslik, qamrov darajasi va turli multisinflar orasidagi ajraluvchanlik mezonlari formal ifodalandi. Shuningdek, pretsedentlarni tanlashda ma’lumotlarning sifati, klinik muvofiqligi va ekspert bahosini hisobga oluvchi integrallashgan mezon taklif etildi. Ushbu yondashuv shovqinli va noaniq tibbiy ma’lumotlar sharoitida ham tipik klinik holatlarni aniqlash imkonini beradi.

Bundan tashqari, yangi bemorni pretsedentlar bilan o‘xshashligi asosida tasniflash qoidasi hamda qaror ishonchliligini baholovchi marja mezoni kiritildi. Bu esa modelning nafaqat tashxisiy aniqligini, balki uning klinik talqin qilinuvchanligini ham oshiradi. Taklif etilgan yondashuv shifokorga bemorning qaysi tipik klinik holatga qanchalik yaqinligini izohlash imkonini berib, ekspert tizimlarining shaffofligi va ishonchliligini ta’minlaydi.

Natijada multisinf–pretsedent modeli informativ belgilarni tanlash uchun ham qulay nazariy asos yaratadi. Belgilar multisinf ichidagi o‘xshashlikni saqlashi va turli pretsedentlarni samarali ajrata olishi talab etilishi natijasida diagnostik jihatdan eng muhim belgilarni aniqlash imkoniyati yuzaga keladi. Ushbu yondashuv tibbiy tashxislash tizimlarining aniqligi, barqarorligi va talqin qilinuvchanligini oshirish bilan birga, keyingi tadqiqotlarda aralash turdagi belgilar uchun adaptiv o‘xshashlik mezonlarini ishlab chiqish va amaliy klinik ma’lumotlarda sinovdan o‘tkazish uchun mustahkam nazariy asos bo‘lib xizmat qiladi.

ADABIYOTLAR

1. *Журавлев Ю.И.* Об алгебраическом подходе к решению задач распознавания или классификации.// Проблемы кибернетики, 1978. Вып. 33. С. 5-68.
2. *Неймарка Ю.* Распознавание образов и медицинская диагностика/ Под ред. М.: Наука, 1972.
3. *Енюков И.С.* Методы, алгоритмы, программы многомерного статистического анализа. - М.: Финансы и статистика, 1986.
4. *Guyon I., Elisseeff A.* An Introduction to Variable and Feature Selection // Journal of Machine Learning Research. – 2003. – Vol. 3. – P. 1157–1182.
5. *Varoquaux G., Cheplygina V.* Machine Learning for Medical Imaging: Methodological Failures and Recommendations for the Future // npj Digital Medicine. – 2022. – Vol. 5. – Art. 48. DOI: 10.1038/s41746-022-00592-y.
6. *Collins G.S., Moons K.G.M.,* Dhiman P. et al. TRIPOD+AI Statement: Updated Guidance for Reporting Clinical Prediction Models That Use Regression or Machine Learning Methods // BMJ. – 2024. – Vol. 385. – e078378. DOI: 10.1136/bmj-2023-078378.
7. *Vasey B., Nagendran M., Campbell B.* et al. Reporting Guideline for the Early-Stage Clinical Evaluation of Decision Support Systems Driven by Artificial Intelligence: DECIDE-AI // Nature Medicine. – 2022. – Vol. 28. – P. 924–933. DOI: 10.1038/s41591-022-01772-9.
8. *Liu X., Rivera S.C., Moher D.* et al. Reporting Guidelines for Clinical Trial Reports for Interventions Involving Artificial Intelligence: The CONSORT-AI Extension // Nature Medicine. – 2020. – Vol. 26. – P. 1364–1374. DOI: 10.1038/s41591-020-1034-x.
9. *Chicco D., Jurman G.* The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation // BMC Genomics. – 2020. – Vol. 21. – Art. 6. DOI: 10.1186/s12864-019-6413-7.

(cc) BY

10. *Tibshirani R.* Regression Shrinkage and Selection via the Lasso // Journal of the Royal Statistical Society: Series B. – 1996. – Vol. 58, No. 1. – P. 267–288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.
11. *Cover T.M., Hart P.E.* Nearest Neighbor Pattern Classification // IEEE Transactions on Information Theory. – 1967. – Vol. 13, No. 1. – P. 21–27. DOI: 10.1109/TIT.1967.1053964.
12. *Huang Z.* Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values // Data Mining and Knowledge Discovery. – 1998. – Vol. 2. – P. 283–304. DOI: 10.1023/A:1009769707641.
13. *Kapoor S., Narayanan A.* Leakage and the Reproducibility Crisis in Machine-Learning-Based Science // Patterns. – 2023. – Vol. 4, No. 9. – Art. 100804. DOI: 10.1016/j.patter.2023.100804.
14. *Van Calster B., McLernon D.J., van Smeden M., Wynants L., Steyerberg E.W.* Calibration: The Achilles Heel of Predictive Analytics // BMC Medicine. – 2019. – Vol. 17. – Art. 230. DOI: 10.1186/s12916-019-1466-7.
15. *Lundberg S.M., Lee S.-I.* A Unified Approach to Interpreting Model Predictions // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 4765–4774.
16. *Wolff R.F., Moons K.G.M., Riley R.D.* et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies // Annals of Internal Medicine. – 2019. – Vol. 170, No. 1. – P. 51–58. DOI: 10.7326/M18-1376.
17. *Saito T., Rehmsmeier M.* The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets // PLoS ONE. – 2015. – Vol. 10, No. 3. – e0118432. DOI: 10.1371/journal.pone.0118432.
18. *Nishanov A.Kh., Rakhmanov A.T., Ruzibaev O.B., Allamov O.T., Zaripov F.M.* Алгоритм многоклассовой классификации с использованием шарового покрытия. Высокопроизводительные вычислительные системы и технологии в научных исследованиях, автоматизации управления и производства: XII Междунар. науч.-практ. конф. (ВВСТ-2022), 20–21 мая 2022 г. // Высокопроизводительные вычислительные системы и технологии. — 2022. — Т. 6, № 1. — С. 119–125.
19. *Zhuravlev Y.I., Ryazanov V.V.* Solution of instance-based recognition problems with a large number of classes. Doklady Mathematics. 2017. Vol. 96, No. 2. P. 488–490.
20. *Ollamberganov F.F., Ro‘ziboyev O.B., Akbarova M., Nishanov A.X., Muxsinov Sh.* Multiklasterlashda pretsedent obyektlarni tanlash algoritmi // «Muhammad al-Xorazmiy avlodlari» ilmiy-amaliy va axborot-tahliliy jurnali. – Toshkent, 2025. – № 2(32). – B. 80–84.
21. *Gorelik A.L.* General statement of the problem of object and phenomenon recognition. Cybernetics. 1980. No. 6. P. 72–75.
22. *Nishanov A.Kh., Akbaraliev B.B., Samandarov B.S., Akhmedov O.K., Tajibaev S.K.* An algorithm for classification, localization and selection of informative features in the space of polytypic data. Webology. 2020. Vol. 17, No. 1. P. 341–364.
23. *Rakhmanov A.T., Akbaraliev B.B., Ruzibaev O.B.* On one modified method of solving the classification problem. Chemical Technology. Control and Management. 2014. No. 4. P. 85–91.
24. *Rakhmanov A.T., Ruziboev O.B.* To the solution of the classification problem for three classes. Science and World. 2015. Vol. 2, No. 18. P. 21–25.
25. *Nishanov A.X., Ruzibaev O.B., Tran N.H.* Modification of decision rules “Ball Apolonia” in the problem of classification. International Conference on Information Science and Communications Technologies (ICISCT). 2016. P. 1–3.
26. *Nishanov A.Kh., Ruzibaev O.B., Chedjou J.C., Kyamakya K.* et al. Algorithm for the selection of informative symptoms in the classification of medical data. Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference. 2020. P. 647–658.

(cc) BY

27. *Chandrashekar, G., & Sahin, F.* (2014). A Survey on Feature Selection Methods. Computers & Electrical Engineering, 40(1), 16–28.
28. *Li, J., Cheng, K., Wang, S.*, et al. (2017). Feature Selection: A Data Perspective. ACM Computing Surveys, 50(6), 94.
29. *Saeys, Y., Inza, I., & Larrañaga, P.* (2007). A Review of Feature Selection Techniques in Bioinformatics. Bioinformatics, 23(19), 2507–2517.
30. *Guyon, I., Weston, J., Barnhill, S., & Vapnik, V.* (2002). Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46(1–3), 389–422.
31. *Remeseiro, B., & Bolon-Canedo, V.* (2019). A Review of Feature Selection Methods in Medical Applications. Computers in Biology and Medicine, 112, 103375.
32. *Al Mamun, A., Sohel, F.*, et al. (2021). A Comprehensive Review of Feature Selection Methods for Biomedical Data. IEEE Access, 9, 114125–114148.

(cc) BY