

Козак Л.М.<https://orcid.org/0000-0002-7412-3041>Інститут інформаційних технологій та систем
Національної академії наук України**Білявенко Л.В.**<https://orcid.org/0009-0003-0602-1072>Інститут інформаційних технологій та систем
Національної академії наук України

МЕТОД АНАЛІЗУВАННЯ, ОЦІНЮВАННЯ ТА ПРОГНОЗУВАННЯ ДИНАМІКИ СТАНУ РУХОВИХ ФУНКЦІЙ ВНАСЛІДОК ЛІКУВАЛЬНО-РЕАБІЛІТАЦІЙНИХ ЗАХОДІВ

У статті запропоновано метод аналізування, оцінювання та прогнозування динаміки лікувально-реабілітаційних заходів на основі поєднання методів статистичного аналізу, кластерного аналізу та сучасних методів *Data Mining* і машинного навчання. Актуальність дослідження зумовлена необхідністю підвищення обґрунтованості та об'єктивності клінічних рішень у процесі реабілітації пацієнтів із порушеннями функцій опорно-рухового апарату, зокрема у військовослужбовців після поранень і травматичних ушкоджень, в умовах обмежених даних і високої міжіндивідуальної варіабельності клінічних показників. Запропонований метод орієнтовано на роботу з гетерогенними та незбалансованими клінічними даними і передбачає поетапне оброблення інформації: аналіз обсягу та якості первинних медичних даних, формування інформативного вектора ознак, кластеризацію пацієнтів за ступенем тяжкості стану рухових функцій, балансування вибірок і побудову прогностичних моделей. Для кластерного аналізу застосовано алгоритм *HDBSCAN*, який не потребує попереднього задання кількості кластерів і забезпечує автоматичне виділення стійких груп пацієнтів з урахуванням неоднорідності щільності даних та наявності шумових спостережень. З метою підвищення якості навчання моделей у разі дисбалансу класів використано комбінацію методів *SMOTE* та *Random OverSampling*, що дало змогу суттєво покращити показники точності, зокрема *F1-score*, для міноритарних груп пацієнтів. Для прогнозування динаміки стану та кінцевих результатів реабілітації застосовано ансамблеві моделі *Random Forest* та *XGBoost*, які забезпечують високу прогностичну точність, стійкість до перенавчання та можливість інтерпретації внеску окремих клінічних показників. Метод уможливує не лише класифікування пацієнтів за поточним рівнем тяжкості та ризику, а й прогнозування очікуваної динаміки змін стану залежно від вибраних лікувально-реабілітаційних комплексів, що створює підґрунтя для підтримки прийняття лікарських рішень. Апробацію методу проведено в клінічних умовах на вибірці військовослужбовців з діагнозом паралегія, які проходили реабілітацію, що підтвердило його практичну придатність, інформативність та потенціал інтеграції в системи цифрової медицини та клінічні системи підтримки прийняття рішень.

Ключові слова: цифрова медицина, *Data Mining*, машинне навчання, кластеризація, моделювання, прогнозування, *XGBoost*, реабілітація, підтримка прийняття рішень лікарем.

Постановка проблеми. Сучасна медицина дедалі більше покладається на цифрові технології для підвищення ефективності діагностики, лікування та керування даними пацієнтів. Актуальною є потреба інформаційної підтримки надання реабілітаційної допомоги військовим з пораненнями в клінічних умовах, яка забезпечує можливість

отримання, перегляду та аналізу цифрових медичних даних (ЦМД) лікарями на робочому місці, що уможливує удосконалення та пришвидшення процесів діагностування стану пацієнтів та їхньої реабілітації.

Аналіз останніх досліджень і публікацій. Для розроблення класифікаційних діагностичних



моделей застосовують різні типи методів машинного навчання: логістичні регресії, методи дерев рішень та випадкового лісу, метод опорних векторів, ансамблі класифікаційних моделей, нейронні мережі та глибоке навчання. Ефективність таких моделей безпосередньо залежить від повноти, якості та інтеперабельності цифрових медичних даних, що використовуються для їх навчання та валідації. Усі процеси цифрової медицини реалізуються з використанням цифрових медичних даних про стан здоров'я людини, тобто персональних медичних даних, що зумовлює необхідність визначення вимог до їх захищеного обміну, стандартизованого подання та надійного зберігання з можливістю оперативного доступу без втрати інформації [1-4]. Розробки за напрямом клінічного спрямування прогностичних моделей, побудованих внаслідок поєднання різних методів Data Mining, дають змогу виявити клінічно значущі закономірності із застосуванням даних моніторингу та лікувальних втручань [5-6]. Зокрема, у роботі [7] продемонстровано ефективність використання ансамблів методів Data Mining для класифікації станів пацієнтів, що дає змогу підвищити точність стратифікації стану та обґрунтованість клінічних рішень. У подальших дослідженнях автори показали можливість застосування нейронних мереж для прогнозування керувальних параметрів хірургічного втручання в кардіології, що підтверджує доцільність використання інтелектуальних моделей для підтримки лікувально-діагностичних процесів [8]. У XX сторіччі розроблено та наразі застосовують класичні інструменти (системи) оцінювання і прогнозування стану пацієнтів: шкали ступеня тяжкості стану, різні спеціалізовані тести, зокрема для оцінювання стану м'язових функцій (індекс Бартела, індекс болю WAS тощо), які класифікують пацієнта в конкретні групи ризику на основі аналізу відхилів анатомічних, фізіологічних, біохімічних показників [9]. Розроблення методів та технологій дослідження, спрямованих на розв'язання конкретних завдань практичної медицини, базується на аналізі поточного стану як цілісного організму, так і його окремих фізіологічних систем. Для виявлення динаміки стану пацієнтів (зокрема військовослужбовців, які проходять реабілітацію після лікування уражених кінцівок внаслідок поранення чи травмування) і визначення ефективності комплексу задіяних реабілітаційних інтервенцій було сформовано ансамбль методів та розроблено схему дослідження, в якій поєднано методи статистичного аналізу та методи Data Mining для аналізування

показників рухових функцій. Зазвичай проводять попередній відбір ознак до початку застосування алгоритмів машинного навчання. Застосовують статистичні критерії кореляції кожного первинного показника з цільовою ознакою та критерії впорядкування (наприклад, за величиною та значимістю Chi-Square Пірсона, F Фішера). Подальший відбір комплексу (комбінацій) інформативних ознак виконують за допомогою алгоритмів класифікації для отримання більшої точності [10].

Кластеризація. Одним з ефективних методів оброблення даних є їх сегментація за допомогою методів кластерного аналізу (неконтрольоване навчання) [11]. Розбиття на кластери (підгрупи) дає змогу застосовувати всю доступну інформацію для побудови кількох моделей, а потім виконувати точніші прогнози за найкращою моделлю. Найпоширеніші методи: *k*-means для чіткої кластеризації та HDBSCAN, які реалізовано в багатьох модулях, в тому числі і в бібліотеках Python [12]. Кластеризація за алгоритмом HDBSCAN у цьому дослідженні первинний розподіл пацієнтів на групи виконано із застосуванням алгоритму HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), який належить до класу методів щільнісної кластеризації та не потребує попереднього задання кількості кластерів. Алгоритм HDBSCAN дає змогу виявляти кластери довільної форми, автоматично ідентифікувати шумові спостереження та забезпечувати стабільний розподіл об'єктів у разі неоднорідної щільності даних.

На відміну від алгоритмів типу *k*-means, у HDBSCAN кластеризація ґрунтується не на мінімізації відстаней до фіксованих центрів, а на аналізі щільності розподілу об'єктів у просторі ознак. Базовим елементом методу є *core-distance* – мінімальна відстань від об'єкта до його *k*-го найближчого сусіда, що визначає локальну щільність даних. Для побудови ієрархічної структури кластерів використовується взаємна досяжна відстань (*mutual reachability distance*), яка визначається як

$$d_{\{mr\}}(x_i, x_j) = \max(\text{core}_k(x_i), \text{core}_k(x_j), d(x_i, x_j)).$$

На основі взаємної досяжної відстані формується мінімальне основне дерево, з якого будується ієрархія кластерів. Подальший відбір оптимальних кластерів здійснюється шляхом максимізації їх стабільності, яка характеризує тривалість існування кластера в ієрархічному дереві при зміні порогів щільності.

Отже, у HDBSCAN кількість кластерів визна-

частіше автоматично на основі їх стабільності. Для оцінювання якості отриманої кластеризації доцільно використовувати індекси валідності кластерів, зокрема показник DBCV (Density-Based Cluster Validation), а також додатково індекс Silhouette для аналізу внутрішньої однорідності та міжкластерної відокремленості [12-14].

Балансування обсягів кластерів. Серед найбільших викликів для ML треба виокремити таку характеристику даних, як збалансованість. Під збалансованістю мають на увазі рівномірну кількість елементів (пацієнтів, подій, елементів тощо) в одному певному класі або кластері порівняно з іншими. Через дисбаланс класи з одиничними елементами (конкретними показниками) погано відслідковувати та класифікувати. Це призводить до проблеми, що з потенційним збільшенням пацієнтів загальна точність та ефективність буде зменшуватись внаслідок збільшення частки класу, який недостатньо було заповнено в початкових навчальних даних. Також зазначимо, що дисбаланс класів призводить до перенавчання на мажоритарних (великих за обсягом) класах та недостатнє врахування міноритарних (малих за обсягом) класів. [15]

Прогностичні моделі в клінічних задачах. Прогностичні моделі машинного навчання широко застосовуються в сучасній медицині для передбачення перебігу захворювань, оцінювання ризиків ускладнень і прогнозування результатів лікувальних та реабілітаційних заходів. Використання ансамблевих методів дає змогу врахувати складні нелінійні залежності між клінічними показниками, даними моніторингу та параметрами лікування, що підвищує точність і надійність прогнозів. Серед найефективніших підходів для побудови клінічно орієнтованих прогностичних моделей виділяють методи випадкового лісу (Random Forest) та екстремального градієнтного підсилення (Extreme Gradient Boosting, XGBoost). Випадковий ліс є ансамблевим методом машинного навчання, що базується на побудові великої кількості дерев рішень з використанням випадкових підвибірок даних та підмножин ознак. Остаточний прогноз формується шляхом мажоритарного голосування (для задач класифікації) або усереднення (для задач регресії). У клінічних завданнях Random Forest добре зарекомендував себе завдяки стійкості до шуму, здатності працювати з гетерогенними медичними даними та можливості оцінювання важливості ознак, що є критично важливим для інтерпретації результатів і підтримки прийняття лікарських рішень [16].

Екстремальний градієнтний бустінг (XGBoost) є високоефективною реалізацією методу градієнтного бустінгу дерев рішень, у якому дерева додаються послідовно, а кожне наступне дерево коригує помилки попередніх моделей. XGBoost використовує регуляризацію, оптимізовані алгоритми обчислень і механізми контролю перенавчання, що забезпечує високу точність прогнозування навіть на складних і незбалансованих клінічних вибірках. У медицині XGBoost широко застосовується для прогнозування клінічних результатів, стратифікації пацієнтів за рівнем ризику та оцінювання ефективності лікувальних заходів, демонструючи переваги над класичними моделями за показниками точності та стабільності [17].

Постановка завдання. Метою роботи є розроблення та апробація методу аналізування, оцінювання та прогнозування динаміки лікувально-реабілітаційних заходів на основі методів Data Mining і машинного навчання для підтримки прийняття рішень лікарем у процесі реабілітації пацієнтів з порушеннями рухових функцій.

Виклад основного матеріалу. Метод аналізування, оцінювання та прогнозування ефективності і динаміки стану рухових функцій внаслідок лікувально-реабілітаційних заходів. Розроблення методу для аналізу зміни стану уражених функцій військовослужбовців об'єднувало чотири узагальнених етапів:

- аналіз обсягу та якості первинних даних про пацієнтів (розмір дата фрейму, кількість потенційних ознак, пропущені значення);
- визначення кластерів за ступенем тяжкості стану рухових функцій (кластерний аналіз із використанням HDBSCAN);
- балансування кластерів – застосовано комбінацію SMOTE та ROS;
- застосування прогностичних моделей та методів Data Mining, аналіз результатів та метрик ефективності моделей.

Запропонований метод оцінювання та прогнозування стану і динаміки пацієнта базується на попередньо відібраних інформативних ознаках і реалізується як послідовність трьох етапів (Рис. 1):

- формування вхідного вектора ознак (на основі визначених важливих показників);
- класифікація пацієнта за поточним станом/рівнем ризику (або ступенем тяжкості);
- прогнозування динаміки та кінцевого результату реабілітації (очікуваний тренд змін і прогноз фінального стану).

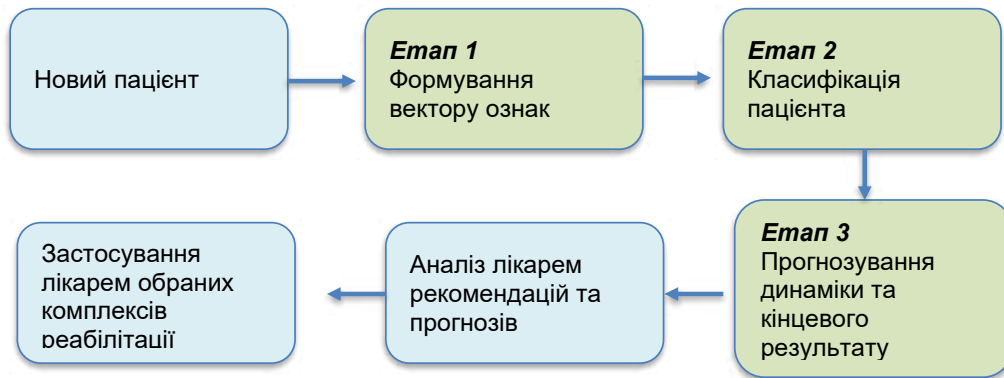


Рис. 1. Структура методу аналізування, оцінювання та прогнозування динаміки стану рухових функцій внаслідок лікувально-реабілітаційних заходів

Етап 1 – Формування вхідного набору ознак на основі раніше визначених важливих показників

На першому етапі метод використовує не весь масив даних, а підмножину ознак, попередньо підтверджених, як найбільш інформативні для завдання оцінювання та прогнозування. Цей набір містить: клінічні показники та результати тестів рухових функцій (Berg Balance Scale, 10MWT, Four Square Step Test, VAS, Barthel Index, Rivermead тощо); параметри, що відображають базовий стан та функціональні обмеження запланованих; характеристики інтервенцій (типи та інтенсивність реабілітаційних заходів). Ознаки подаються у стандартизованому вигляді (кодування категоріальних значень тощо), щоб забезпечити коректний ввід у модель.

Вихід етапу 1: уніфікований числовий вектор ознак пацієнта

Етап 2 – Класифікація пацієнта за поточним станом або рівнем ризику

На другому етапі на основі сформованого вектора ознак виконується класифікація пацієнта, віднесення до класу тяжкості, визначення групи ризику.

Класифікація реалізована на основі методів Data Mining, а саме за допомогою моделей Random Forest, XGBoost та кластерного аналізу, які забезпечують не лише клас, а і ймовірності належності до класів, що підвищує цінність для прийняття клінічного рішення (Рис. 2).

Вихід етапу 2: клас пацієнта

Етап 3 – Прогнозування динаміки та кінцевих результатів лікувально реабілітаційних заходів

3.1 Аналіз ймовірних прогнозів в залежності від застосованих лікувально реабілітаційних заходів

3.2 Візуалізація результатів та представлення лікарю для прийняття рішення

На третьому етапі за допомогою метода виконується прогноз динаміки стану (очікувана зміна ключових функціональних показників) та/або прогноз кінцевого стану в результаті застосування конкретних лікувально-реабілітаційних комплексів. Прогноз включає:

- очікуваний кінцевий ступінь тяжкості стану;
- оцінку впливу комплексу інтервенцій на прогноз;
- ймовірність досягнення функціонально значущого покращення;



Рис. 2. Приклад складника Дерево рішень у моделі XGBoost

прогнозовану траєкторію зміни стану (початковий → проміжний → кінцевий етап);

Цей етап є основою для підтримки рішення лікаря: етап дає змогу порівнювати очікувані результати для різних сценаріїв реабілітації та визначати найбільш перспективну тактику.

Вихід етапу 3: прогноз кінцевого стану та динаміки (подальше прийняття рішення лікарем)

Апробацію методу проведено під час аналізу змін стану функцій опорно-рухового апарата у військових (з діагнозом параплегія), які проходили реабілітацію в ДП «Санаторій Конча-Заспа» у 2025 році. Для побудови моделі використано обсяг даних – 53 пацієнта, для апробації методу – 26 пацієнтів. Визначено 3 кластери за значеннями тяжкості стану рухових функцій (Рис. 3).

За допомогою методу було класифіковано та зроблено прогноз стосовно клінічно значущих покращень стану пацієнтів для трьох виділених кластерів (Табл. 1).

Таблиця 1

Прогнозований та фактичний відсоток клінічно значущих покращень в певному кластері тяжкості

Тяжкість	Кількість пацієнтів	Прогноз	Факт
низька	6	83,3%	83,3%
середня	9	77,8%	88,9%
висока	11	90,9%	72,7%

Отже, результати апробації запропонованого методу свідчать про його високу прогностичну спроможність та практичну придатність для використання у разі аналізу змін стану опорно-рухового апарата після проведення реабілітації.

Отримано відповідність між прогнозованими та фактичними показниками клінічно значущих покращень. Апробація підтверджує адекватність побудованих моделей і коректність вибраної методології аналізу даних. Метод демонструє здатність не лише до якісної класифікації пацієнтів за ступенем тяжкості стану рухових функцій, а й до прогнозування клінічно значущих змін цього стану внаслідок застосування відповідного комплексу лікувально-реабілітаційних заходів. Це надає змогу розглядати його як ефективний інструмент підтримки прийняття рішень лікарем, орієнтованих на індивідуалізацію лікувально-реабілітаційних заходів.

Висновки. Розроблено інтегрований метод аналізування, оцінювання та прогнозування динаміки стану рухових функцій внаслідок лікувально-реабілітаційних заходів, який поєднує кластерний аналіз, балансування даних та сучасні ансамблеві моделі машинного навчання, що забезпечує системний і відтворюваний підхід до аналізу складних клінічних процесів. Запропонований метод реалізує прикладний сценарій підтримки прийняття рішень лікарем, оскільки дає змогу не лише класифікувати пацієнта за поточним станом, а й прогнозувати очікувану динаміку та кінцевий результат реабілітації залежно від вибраного комплексу лікувально-реабілітаційних заходів. Результати клінічної апробації підтвердили практичну придатність методу для використання в реабілітаційних закладах, зокрема під час надання реабілітаційних заходів військовослужбовцям з ураженнями опорно-рухового апарата, а також можливість інтеграції запропонованого методу в інформаційні системи цифрової медицини та системи підтримки прийняття рішень лікарем.

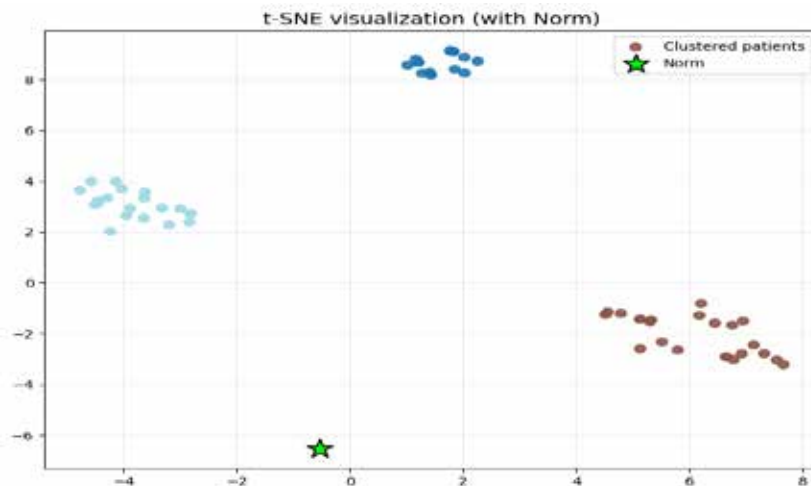


Рис. 3. Візуалізація кластерного аналізу із застосуванням алгоритму HDBSCAN

Список літератури:

1. Dwarampudi T. M., Alugolu A. Review of machine learning classification algorithms. *International Journal of Progressive Research in Engineering Management and Science*. 2024. DOI: <https://doi.org/10.58257/ijprems36877>
2. Saberioon M. et al. Comparative Performance Analysis of Support Vector Machine, Random Forest, Logistic Regression and k-Nearest Neighbours in Rainbow Trout (*Oncorhynchus Mykiss*) Classification Using Image-Based Features. *Sensors*. 2018. Vol. 18, no. 4. P. 1027. DOI: <https://doi.org/10.3390/s18041027>
3. Degadwala D. S. D., Degadwala D. V. A Review on Machine Learning and Deep Learning Methods on Medical Image Classification. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2024. Vol. 10, no. 3. P. 546–555. DOI: <https://doi.org/10.32628/cseit24103205>
4. Romaniuk O. O., Kozak L. M., Kovalenko O. S. Formation of Interoperable Digital Medicine Information Environment: Personal Medical Data. *Sci. innov.* 2021. V. 17, no. 5. P. 50–62.
5. Georga E. I., Tachos N. S., Sakellarios A.I., Kigka V. I., Exarchos T. P., Pelosi G. Artificial intelligence and data mining methods for cardiovascular risk prediction Cardiovascular Computing. *Methodologies and Clinical Applications*. 2019, N 3, PP. 279–301.
6. Amin M., Chiam Y. Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*. 2019. V. 36, N 1, PP. 82–93.
7. Kryvova O.A., Kozak L.M. Information Technology for Classification of Donosological and Pathological States Using the Ensemble of Data Mining Methods. *Cybernetics and Computer Engineering*. 2021, 1(203), pp 77–96.
8. Kryvova O.A., Kozak L.M., Kovalenko O.S., Nenasheva L.V. Prediction of Surgery Control Parameters in Cardiology to Optimize the Emission Fraction Values with the Help of Neural Networks. *Cybernetics and Computer Engineering*. 2021, 4(206), pp 54–72.
9. Kaieski N., da Costa C. A., da Rosa Righi R., Lora P. S. Application of artificial intelligence methods in vital signs analysis of hospitalized patients: A systematic literature review. *Applied Soft Computing*. 2020. V. 96, N 6, PP. 56–89.
10. Richman J. S., Randall M. J. et al. Physiological time-series analysis using approximate entropy and sample entropy. *Heart Circ. Physiol.* 2000. V. 278, N 6. PP. H22039–H2049.
11. İşler Y., Kuntalp M. Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Computers in Biology and Medicine*. 2007. Vol. 37. N. 10. PP. 1502–1510.
12. Campello R. J.G. B., Moulavi D., Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds) *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science*, vol 7819. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37456-2_14
13. Campello R. J. G. B. et al. Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Transactions on Knowledge Discovery from Data*. 2015. Vol. 10, no. 1. P. 1–51. DOI: <https://doi.org/10.1145/2733381>
14. McInnes et al. HDBSCAN: Hierarchical density based clustering, *Journal of Open Source Software*, 2017, 2(11), 205. DOI: <https://doi.org/10.21105/joss.00205>
15. Mohan S., Thirumalai C., Srivastava G. Effective heart disease prediction using hybrid machine learning techniques. *IEEE Access*, 2019. 7:81542–81554.
16. Couronné R., Probst P., Boulesteix A.-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*. 2018. Vol. 19, no. 1. DOI: <https://doi.org/10.1186/s12859-018-2264-5>
17. Zhang Z. et al. Predictive analytics with gradient boosting in clinical medicine. *Annals of Translational Medicine*. 2019. Vol. 7, no. 7. P. 152. URL: <https://doi.org/10.21037/atm.2019.03.29>

Kozak L.M., Biliavenko L.V. METHOD FOR ANALYZING, ASSESSING AND PREDICTING THE DYNAMICS OF MOTOR FUNCTIONS STATE AS A RESULT OF TREATMENT AND REHABILITATION INTERVENTIONS

The article proposes a method for analyzing, evaluating and predicting the effectiveness and dynamics of medical and rehabilitation measures based on a combination of statistical analysis methods, cluster analysis and modern methods of Data Mining and machine learning. The relevance of the study of the conditioned need increases the validity and objectivity of clinical decisions in the process of rehabilitation with disorders of the musculoskeletal system, in particular, military personnel after injuries and traumatic injuries, in conditions of limited data and high interindividual variability of clinical indicators. The proposed method, focused on working with heterogeneous and unbalanced clinical data, provides step-by-step information processing:

analysis of the volume and quality of primary medical data, formation of an informative feature vector; clustering created by the severity of motor functions, balancing the choice and building predictive models. For cluster analysis, the HDBSCAN algorithm was used, which does not require a preliminary setting of the number of clusters and provides automatic recovery of stable groups taking into account the heterogeneity of data density and the presence of noisy observations. In order to improve the quality of model training in the event of class imbalance, a combination of SMOTE and Random OverSampling methods is used, which allowed us to significantly increase the accuracy indicators, in particular the F1-score, for minority groups. To predict the dynamics of the state and the final results of rehabilitation, the ensemble models Random Forest and XGBoost were used, which provide high predictive accuracy, resistance to overtraining and the possibility of interpreting out-of-the-ordinary clinical indicators. The method allows not only to classify diseases by the current level of severity and risk, but also to predict the expected dynamics of changes in the state depending on the selected treatment and rehabilitation complexes, which is confirmed by confirmation for supporting medical decision-making. The testing method was conducted in clinical conditions on a sample of military personnel diagnosed with paraplegia who were undergoing rehabilitation, which confirmed its practical applicability, informativeness, and potential for integration into digital medicine systems and clinical decision support systems.

Keywords: *digital medicine, Data Mining, machine learning, clustering, modeling, forecasting, XGBoost, rehabilitation, physician decision support.*

Дата першого надходження статті до видання: 26.01.2026

Дата прийняття статті до друку після рецензування: 20.02.2026

Дата публікації (оприлюднення) статті: 08.04.2026