Original Article
Clinics in Orthopedic Surgery 2023;15:894-901 • https://doi.org/10.4055/cios22181

Predicting Prolonged Wound Drainage after Hemiarthroplasty for Hip Fractures: A Stacked Machine Learning Study

Sultan Turhan, PhD Student, Umut Canbek, Assoc Dr*, Tugba Dubektas-Canbek, MD†, Eralp Dogu, Assoc Dr

Department of Statistics, Mugla Sitki Kocman University, Mugla,
**Department of Orthopedics and Traumatology, Mugla Sitki Kocman University College of Medicine, Mugla,*
†Department of Internal Medicine, Mugla Sitki Kocman University College of Medicine, Mugla, Türkiye

Background: Prolonged wound drainage (PWD) is one of the most important reasons that increase the risk of early periprosthetic joint infection after arthroplasty. It is very important to evaluate the risk factors for PWD in the surgical field after arthroplasty surgery. This can be accomplished using machine learning or artificial intelligence methods. Our aim in this study was to compare machine learning methods in predicting possible PWD.

Methods: The study was carried out on clinical, laboratory, and radiological data of 313 patients who underwent hemiarthroplasty (HA) for proximal femur fractures. We preprocessed the dataset and trained and tested machine learning methods using cross validation. We compared various machine learning algorithms (linear discriminant analysis, decision tree, k-nearest neighbors, gradient boosting machine, and logistic regression [LR]) based on performance measures. We also combined the most successful algorithms with a metaclassifier. To help understand the relationship between risk factors, we provided a risk factor severity ranking.

Results: To estimate the risk of PWD, classification was performed with first-level classifiers and then integrated as a LR-based meta-learner stacking method. More performance improvements were achieved with the stacking method.

Conclusions: We found that the stacking method was superior to other methods in PWD classification. We determined that the volume of fluid collected from the drain, morbid obesity class, blood transfusion, and body mass index score were the four most important risk factors according to stacking.

Keywords: *Drainage fluid, Hip fracture, Hemiarthroplasty, Stacking, Machine learning*

Today, hip fracture surgery represents a large quota of orthopedic surgeon activities and has significant clinical and social cost implications associated with it.[1,2] For hip fracture treatment, which has high clinical importance and cost, it is very important to evaluate and predict the presence of prolonged wound drainage (PWD) after hip hemiarthroplasty (HA) and to assess risk factors.[3] PWD after HA highly affects the recovery process and results in longer length of stay. Statistical and machine learning (ML) models such as logistic regression (LR) have been widely used to assess underlying complex patterns of risk factors in the hip.[3] However, their use for predictive analysis is limited.

Various studies have been conducted in the literature on ML methods related to hip fractures and femoral neck fractures. Yoo et al.[4] developed ML models to more accurately determine the risk of osteoporosis of the femoral neck in postmenopausal women and compared them with the osteoporosis self-assessment tool. Karnuta et al.[5] used naive Bayes ML method, which provides high ac-

Received June 8, 2022; Revised September 24, 2022;
Accepted September 27, 2022
Correspondence to: Eralp Dogu, Assoc Dr
Department of Statistics, Mugla Sitki Kocman University, Mugla, Türkiye
Tel: +90-25-2211-5446
E-mail: eralp.dogu@mu.edu.tr

curacy and responsiveness in estimating the length of stay and cost of hip fracture care using preoperative variables. Shtar et al.[6] compared ML methods in designing an estimated rehabilitation outcomes model for patients after acute hip fractures. Cary et al.[7] evaluated an ML model designed to predict mortality for those treated for the hip. All these studies were carried out with conventional classification methods.[7] Various studies have been conducted in the literature on ML methods related to hip fractures and femoral neck fractures. All these studies were carried out with conventional classification methods.[4-7] However, no study has been found in the literature on the use of ML method for the analysis of risk factors for PWD after arthroplasty. Moreover, most of the research focus on the best performing models, but the findings are summaries of comparison of models and require high-level coding skills. This results in limited use of ML models in the clinical practice. In other words, clinicians cannot use the models for daily decisions due to technical difficulties.

Our aim in this study was to propose advanced ML algorithms to detect high risk of developing PWD after arthroplasty. Even developed models might be used for assessment of risk factors, but they may have lower predictive performance. Canbek et al.[3] studied risk factors of PWD after arthroplasty using a logistic model but the study did not cover predictive performance of this method. Here, using the same data, we expanded the use of logistic model by comparing it with advanced ML models such as decision trees, k-nearest neighbors (KNN) algorithm, decision tree, boosted machines and stacking. Our hypothesis was that predictive models could be improved through stacking of trained models to predict PWD after arthroplasty. Our findings support this idea by providing higher accuracy and precision when compared to previously developed logistic models[3] and widely used ML algorithms. Our findings also support that these models do not conflict in terms of the important risk factors such as surgery type, volume of fluid drained from the drain, blood transfusion, and morbid obesity while they perform better. The comparisons showed that stacking models achieved the best performance. Although high-performing ML models were developed, they were not available for clinical practice in most cases. We developed a web-based interface to increase the practical use of these models where the clinician could enter the measurements for the aforementioned risk factors to predict the risk of having PWD after arthroplasty. The web interface is available at https://biodatalab.shinyapps.io/PWDPredictoR/.

METHODS

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This study was approved by the Ethics Committee of Mugla Sıtkı Kocman University (No. 5/01), and informed consent was obtained from all individual participants included in the study.

Patient Population and Sample

The data used in the study were obtained from patients who visited the orthopedic department of Mugla Sıtkı Koçman University Training and Research Hospital between January 1, 2017, and January 1, 2020, and who underwent HA after proximal femur fractures. The dataset was created during the study as explained in detail in the study of Canbek et al.[3] Here, we used the same dataset for further investigation of model development. In the data set, there were 21 features and 313 examples. There were two separate groups, 28 with fluid present and 285 with fluid absent. In all patients, negative pressure closed suction drains were removed under sterile conditions at the 48th hour after surgery. Following the removal of the Hemovac drain, a sterile pediatric urine collection bag with a capacity of 100 mL was placed at the drain outlet. Patients with a daily fluid collection > 2 mL for three consecutive days after placement of the collection bag were considered as disabled cases and labeled as fluid present. Further details of data generation are available in the study of Canbek et al.[3] Table 1 shows the descriptive statistics of the original data set and processed data set.

ML Methods Used in This Study

Linear discriminant analysis

lda is an approach that maximizes the ratio of interclass variance to in-class variance. Here, the goal is to assign units to their actual class with minimal error.[8] The linear discriminant analysis (LDA) method is robust, easy to use, and has high prognostic accuracy.[9]

Decision tree (rpart)

rpart is a tree-based algorithm consisting of a series of decision tests that work with the divide and conquer method. However, it offers a process that is generally less time-consuming than other classification techniques.[10] Decision trees are straightforward and simple.[11]

Table 1. Descriptive Statistics for Wound Drainage Dataset

Variable	Fluid		Total
	Absent (n = 285)	Present (n = 28)	
Age (yr)	81 ± 8	80 ± 8	80.91
Sex			313
Female	183	16	
Male	102	12	
BMI (kg/m^2)	31.41 ± 4.01	35.72 ± 5.74	31.79
DM			
Present	78	7	85
Absent	207	21	228
Fasting glucose (mg/dL)	110.49 ± 45.23	102.92 ± 23.85	109.81
HbA1c (%)	1.82 ± 3.10	1.53 ± 2.73	1.79
ASA score			
ASA 2	154	12	166
ASA 3	117	15	132
ASA 4	14	1	15
Comorbidity score			
2	60	1	61
3	134	12	146
4	71	13	84
5	11	1	12
6	5	1	6
7	4	0	4
Antibiotic type			
Cefazolin	272	25	297
Clindamisin	13	3	16
Preoperative anticoagulant status			
Not using	173	15	188
Aspirin	80	13	93
Low-molecular-weight heparin	6	0	6
Dabigatran	7	0	7
Coumadin	15	0	15
Plavix	4	0	4
Morbid obesity			
BMI < 40 kg/m^2	277	15	292
BMI > 40 kg/m^2	8	13	21
Fracture type			
Femur neck	104	6	110
Trokanterik	181	22	203
Surgical approach			
Posterolateral	237	10	247
Anterolateral	48	18	66
Cerclage presence			
Present	161	16	136
Absent	124	12	177
Bone cement presence			
Present	108	6	114
Absent	177	22	199
Time before operation (day)	23.31 ± 16.96	24.43 ± 11.79	23.41
Time before surgery (day)	0.67 ± 0.82	0.75 ± 0.58	0.67
Transfused blood (U: Units)	1.63 ± 1.47	4.50 ± 2.18	1.88
Input hemoglobin (g/dL)	11.26 ± 1.55	11.02 ± 1.53	11.23
Postoperative day 1 hemoglobin level (g/dL)	9.80 ± 1.48	8.26 ± 1.46	9.66
Hemoglobin level at discharge (g/dL)	10.00 ± 1.01	9.88 ± 0.73	9.98
Volume of fluid drained from the drain (mL)	294.49 ± 105.31	562.50 ± 173.00	318.46

Values are presented as number or mean ± standard deviation.
BMI: body mass index, DM: diabetes mellitus, HbA1c: hemoglobin A1c, ASA: American Society of Anesthesiologists physical status classification.

KNN

KNN is an example-based learning method. The logic of KNN was to put forward with the idea that similar examples belonging to the same class have a high probability. The basic idea of KNN is to first select KNN for each test sample, then use the learned neighbors to estimate this test sample.[12] One of the biggest advantages of the KNN algorithm is that it is easy to implement and sensitive to the local structure of the data.[13]

Gradient boosting machine

The general premise of the gradient boosting machine (GBM) approach is that a single decision tree is not strong enough. A decision tree algorithm is applied in GBM and a new decision tree is designed with the resulting error. This process is continued until the error is minimized.[14,15] Since GBM is an ensemble method, it is an algorithm that helps reduce variance and bias.[16]

Logistic regression

It is a traditional statistical learning method that uses a logistic link function to model a bilateral outcome based on patient-level risk factors.[17,18]

Stacking

Stacking method presents the general architecture in which multiple algorithms (first-level learners) are trained to be combined on a second-level learner (meta learner).[19] The basic idea is to educate first-level learners using the original training data set and then create a meta learner (here, we used an LR model) with a new dataset in which the outputs of the first-level learners are considered input characteristics.[19,20] First-level learners are produced either by applying algorithms from different types of classifiers or by applying a uniform classifier, and, therefore, stacked models are often hybrid of various classifiers. Stacking method offers a better generalization ability by reducing the standard deviation and provides strong flexibility. However, as many algorithms are combined, the interpretation of these models is not easy. In this study, LR method was used as the stacking method.

Implementation

We used R (R Foundation for Statistical Computing), an open-source software and caret, caretEnsemble, ggplot2, packages for implementation. The data set is divided into two as 75% training and 25% test. In the preprocessing step, subtracting the average of the predictor's data from the predictor's value (center) and scaling the average of the predictor's data by the standard deviation were applied. To increase the accuracy of the model, a random search was set as tunelenght = 5 and 10 fold cross-validation was applied. After all these data preprocessing steps, the data set was integrated in GBM, KNN, LDA, and rpart methods. The predictions obtained are combined for stacking (LR) with equal weight. Recursive feature elimination is a good selection way to select key features. Recursive feature elimination, the ML algorithm, can determine which features are important to predict the response variable. For the current data set, four important properties were selected with 10-fold cross validation. These features are fluid discharged from the drain, morbid obesity, transfused blood, and surgical approach. Rpart tuning parameter "cp" was held constant at a value of 0. KNN final value used for the model was k = 7. GBM final values used for the model were n.trees = 644, interaction.depth = 8, shrinkage = 0.018, and n.minobsinnode = 17. There are no hyperparameters for LDA.

A web-based interface that allows prediction of PWD after arthroplasty per patient by using the R programming language on Shiny, an open-source R package, was created. In the "Calculate PWD risk" section, the existence of morbid obesity, type of surgical procedure, number of units of blood transfused, and volume of fluid drained of a new patient were entered and the probability of PWD risk after arthroplasty was given. The decision was made using the stacking model that was developed in this study.

Performance Evaluation

The correct evaluation of these methods is very important in terms of concluding the study.[21] The accuracy criterion among the evaluation performances can be explained as the ratio of correctly predicted answers in the model to all answers. Kappa is a statistical method that measures the reliability of agreement between two raters.[22] Sensitivity is the ratio of predicted positive class values to all positive class values. Specificity is the ratio of correctly predicted negative class values to all negative class values. Precision is the ratio of the correctly predicted positive class values to all positively predicted class values. F1-Measure was developed because evaluating precision and sensitivity criteria together will give more accurate results. The F1-measure is the harmonic mean of precision and sensitivity.[23]

RESULTS

Stacking Models Perform Better Than Traditional ML Methods

Model 1: stacked learning with linear discriminant analysis, KNN, decision tree, and gradient boosted machine

We trained, validated, and tested each model mentioned in the methods. We also developed a stacked model of LDA, KNN, rpart, and GBM on a logistic meta learner. The performance evaluation based on accuracy, F1, kappa, precision, sensitivity, and specificity for LR, LDA, KNN, rpart, GBM, and stacking models are given in Fig. 1. Overall, the stacking (accuracy = 0.89, f1 = 0.84, kappa = 0.81, precision = 0.88, sensitivity = 0.80, specificity = 0.98) algorithm

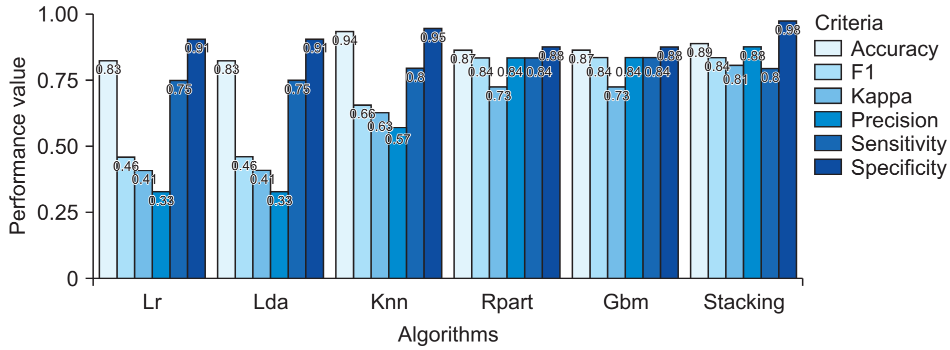

Fig. 1. Performance comparisons for model 1. Accuracy, F1, kappa, precision, sensitivity, and specificity results for logistic regression, linear discriminant analysis, k-nearest neighbors, decision tree, gradient boosted machines, and stacking model 1. Stacking model uses a logistic meta learner to stack linear discriminant analysis, k-nearest neighbors, decision tree, and gradient boosted machines. Results show that decision tree, gradient boosted machine, and stacking model 1 achieved similar performance and outperformed the others.

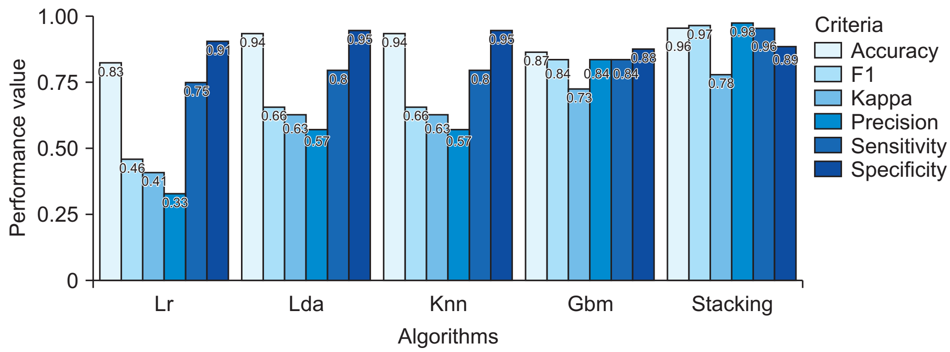

Fig. 2. Performance comparisons for model 2. Accuracy, F1, kappa, precision, sensitivity, and specificity results for logistic regression, linear discriminant analysis, k-nearest neighbors, gradient boosted machines, and stacking model 2. Stacking model uses a logistic meta learner to stack linear discriminant analysis, k-nearest neighbors, and gradient boosted machines. Results show that stacking model 2 outperformed the others in terms of all metrics.

performed better than other algorithms. Stacking method is a two-stage classification method. In the event that the basic classifiers used in the first step learned a certain region of the feature space incorrectly, the meta-stage classifier detected this undesirable situation and corrected this wrong training, outstripping other methods in terms of evaluation criteria from other methods. The logistic model widely used for risk factor assessment has the lowest performance in terms of all metrics (accuracy = 0.83, f1 = 0.46, kappa = 0.41, precision = 0.33, sensitivity = 0.75, specificity = 0.91). Although rpart (accuracy = 0.87, f1 = 0.84, kappa = 0.73, precision = 0.84, sensitivity = 0.84, specificity = 0.88) and GBM (accuracy = 0.87, f1 = 0.84, kappa = 0.73, precision = 0.84, sensitivity = 0.84, specificity = 0.88) provided promising results, stacking of these algorithms overperform them in terms of accuracy, kappa, precision, and specificity. The stacking model 1 did not achieve better results in terms of F1 and sensitivity. In order to increase the detection ability of Model 1, we developed a several models that combine set of these algorithms. The results of Model 2 given in the next section are the highest performing model among those where we combine GBM, KNN, and LDA on a logistic meta learner.

Model 2: stacked learning with linear discriminant analysis, KNN, and gradient boosted machine

As a result of removing the rpart from the stacking model, we achieved the best performance (accuracy = 0.96, f1 = 0.97, kappa = 0.78, precision = 0.98, sensitivity = 0.96, specificity = 0.89) when compared to high-performing GBM (accuracy = 0.87, f1 = 0.84, kappa = 0.73, precision = 0.84, sensitivity = 0.84, specificity = 0.88) and model 1 (accuracy = 0.89, f1 = 0.84, kappa = 0.81, precision = 0.88, sensitivity = 0.80, specificity = 0.98). The results of GBM, KNN, LDA, and stacking algorithms are given in Fig. 2.

Model 2 Can Be Used for Detecting High-Risk Patients

The risk assessment for PWD after arthroplasty can guide

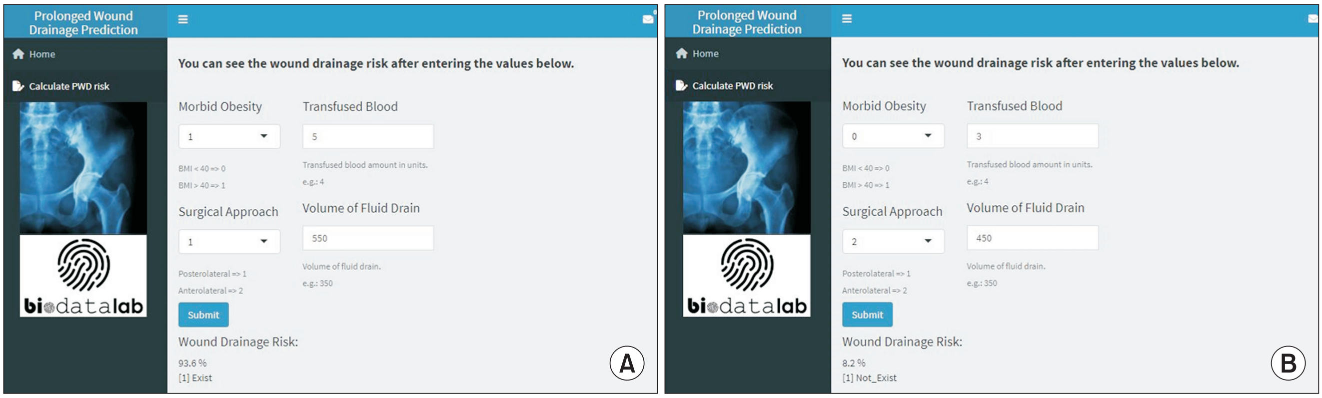

Fig. 3. Illustrative examples of prediction of prolonged wound drainage (PWD) after arthroplasty for patient A (A) and patient B (B). Patient A has higher risk of developing PWD when compared to patient B due to increased body mass index, surgery type, increased amount of blood transfused, and increased amount of liquid from the drain.

clinicians to take preventive actions for the patients with high risk. We developed a web-based interface for patient-by-patient evaluation in the clinical setting. The result of the tool is shown in Fig. 3A for a hypothetical patient A for illustrative purposes. The risk of PWD for this patient was tested on the best ML model (model 2). The patient was morbidly obese (body mass index > 40 kg/m^2), underwent posterolateral surgery, and was transfused with five units of blood, and the liquid discharged from the drain was 550 mL. The tool predicted 93.6% risk for this particular patient. Fig. 3B shows another hypothetical patient B for illustrative purposes. The risk of PWD for this patient was tested on the best ML model (model 2). The patient was not morbidly obese (body mass index < 40 kg/m^2), underwent anterolateral surgery, and was transfused with three units of blood, and the liquid from the drain was 450 mL. The tool predicted 8.2% risk for patient B.

DISCUSSION

It is not clear why some patients develop PWD after arthroplasty. Although the exact source is unknown, some authors think that PWD develops as a result of hematoma or seroma formation.[24] Despite the similarity in the biochemical content of serum and transudative disability fluid, Canbek et al.,[25] upon further proteomic analysis, found that PWD contained lymph-specific proteins. Especially in the field of orthopedics, increasing postoperative complications and economic cost rates necessitate a prediction of the risk of developing PWD. Today, an orthopedic surgeon needs accurate predictions of the outcome of their patients' disease, so high-performance methods are vital to support prevention of PWD. To this end, we developed several high- performing ML algorithms in this study. Canbek et al.[3] showed the amount of fluid drained from the drain, blood transfusion, surgery type, and morbid obesity are of great importance. We found that the risk factors determined by the new models were compatible with the literature. Moreover, advanced models we developed have higher predictive ability when compared to traditional predictive models such as logistic model. We suggest that the method we developed can be used as a decision support tool for the pre-evaluation of risk factors. We also developed an interface for this purpose. The interface can be used as a decision support tool in the clinical setting.

It has been shown that there is a significant relationship between morbid obesity and fluid volume from the drain in patients with PWD after arthroplasty.[26] Ahmed et al.[27] reported that hypertensive patients were more disabled than normotensive patients in their study, and they attributed this to prolongation of bleeding. Although the exact source of the injury is not known, it is thought that bleeding and hematoma from the vessels in the operation area may cause disability.[28] Finally, PWD fluid analysis has been shown as lymph fluid accumulating in the surgical field after major surgical procedures such as PWD and HA. The results shown as risk factors for PWD with the new modeling we found are consistent with the recent literature. One of the most important risk factors is the use of low-molecular-weight heparin in the literature.[26] However, in our case, all patients were using low-molecular-weight heparin. So, the existence/nonexistence of this risk factor could not be evaluated.

Our study has several limitations. Our models rely on one center dataset. Our model development pipeline can be fed by other datasets such as studies for large cohorts and multicenter studies to cover a wide range of population characteristics. It should be noted that although stacking algorithms can help improve prediction performance, interpreting with these models is not an easy task. We used separate models for risk assessment and recommended the use of the most successful level one model. In addition, linear regression and decision tree regression were applied to predict the volume of mean drainage output (mL), and significant results could not be obtained.

In the study, stacking models were developed to predict PWD after arthroplasty. First, we developed individual linear discriminant analysis, KNN, decision tree, and gradient boosted machines models. Then, a meta learner (logistic) was used to stack linear discriminant analysis, KNN, decision tree, and gradient boosted machines models. We found that model 2 outperformed other methods. Model 2 was integrated into a Shiny web-based interface to help clinicians use the model for their daily practice.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ORCID

Sultan Turhan https://orcid.org/0000-0002-9704-1700
Umut Canbek https://orcid.org/0000-0003-1741-7059
Tugba Dubektas-Canbek https://orcid.org/0000-0002-3730-0029
Eralp Dogu https://orcid.org/0000-0002-8256-7304

REFERENCES

1. Aslam-Pervez N, Riaz O, Gopal S, Hossain F. Predictors of intraoperative fractures during hemiarthroplasty for the treatment of fragility hip fractures. Clin Orthop Surg. 2018; 10(1):14-9.
2. Bergstrom U, Jonsson H, Gustafson Y, Pettersson U, Stenlund H, Svensson O. The hip fracture incidence curve is shifting to the right. Acta Orthop. 2009;80(5):520-4.
3. Canbek U, Akgun U, Aydogan NH, Canbek TD, Turgut A, Erciyes OS. Factors associated with prolonged wound drainage after hemiarthroplasty for hip fractures in elderly. Int Orthop. 2020;44(9):1823-31.
4. Yoo TK, Kim SK, Oh E, Kim DW. Risk prediction of femoral neck osteoporosis using machine learning and conventional methods. In: Rojas I, Joya G, Cabestany J, eds. Advances in computational intelligence. IWANN 2013. vol 7903. Lecture Notes in Computer Science. Berlin: Springer; 2013. 181-8.
5. Karnuta JM, Navarro SM, Haeberle HS, Billow DG, Krebs VE, Ramkumar PN. Bundled care for hip fractures: a machine-learning approach to an untenable patient-specific payment model. J Orthop Trauma. 2019;33(7):324-30.
6. Shtar G, Rokach L, Shapira B, Nissan R, Hershkovitz A. Using machine learning to predict rehabilitation outcomes in postacute hip fracture patients. Arch Phys Med Rehabil. 2021;102(3):386-94.
7. Cary MP Jr, Zhuang F, Draelos RL, et al. Machine learning algorithms to predict mortality and allocate palliative care for older patients with hip fracture. J Am Med Dir Assoc. 2021;22(2):291-6.
8. Balakrishnama S, Ganapathiraju A. Linear discriminant analysis: a brief tutorial. Philadelphia: Institute for Signal and information Processing; 1998.
9. Arabameri A, Pourghasemi HR. Spatial modeling in GIS and R for earth and environmental sciences. Amsterdam: Elsevier; 2019. 299-321.
10. Jimenez-Grande D, Farokh Atashzar S, Martinez-Valdes E, Marco De Nunzio A, Falla D. Kinematic biomarkers of chronic neck pain measured during gait: a data-driven classification approach. J Biomech. 2021;118:110190.
11. Wu X, Kumar V, Ross Quinlan J, et al. Top 10 algorithms in data mining. Knowl Inf Syst. 2008;14:1-37.
12. Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf Theory. 1967;13(1):21-7.
13. Soucy P, and Mineau GW. A simple KNN algorithm for text categorization. In: Proceedings of 2001 IEEE International Conference on Data Mining; 2001 San Jose, CA, USA.
14. Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat. 2001;29(5):1189-232.
15. Nisbet R, Elder J, Miner G. Handbook of statistical analysis and data mining applications. Cambridge: Academic Press; 2009. 864.
16. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot. 2013;7:21.
17. Schapire RE, Freund Y, Bartlett P, Lee WS. Boosting the margin: a new explanation for the effectiveness of voting methods. In: Proceedings of the Fourteenth International Conference on Machine Learning; 1997 Jul 8-12; San Francisco, CA, USA.
18. Wright RE. Logistic regression. In: Grimm LG, Yarnold PR, eds. Reading and understanding multivariate statistics.

Washington, DC: American Psychological Association; 1994. 217-44.

19. Wolpert DH. Stacked generalization. Neural Netw. 1992; 5(2):241-59.

20. Wang R. Significantly improving the prediction of molecular atomization energies by an ensemble of machine learning algorithms and rescanning input space: a stacked generalization approach. J Phys Chem C. 2018;122(16):8868-73.

21. Turhan S, Tunc M, Dogu E, Balci Y. Machine learning in forensic science and forensic medicine: research on the literature. Adli Tıp Dergisi. 2022;36(1):1-7.

22. Dwivedi AK. Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Comput Appl. 2018;29:685-93.

23. Li Y, Chen Z. Performance evaluation of machine learning methods for breast cancer prediction. Appl Comput Math. 2018;7(4):212-6.

24. Kilpadi DV, Cunningham MR. Evaluation of closed incision management with negative pressure wound therapy (CIM): hematoma/seroma and involvement of the lymphatic system. Wound Repair Regen. 2011;19(5):588-96.

25. Canbek U, Dibek E, Akgun U, et al. Analysis of the fluid biochemistry in patients with prolonged wound drainage after hip hemiarthroplasty. Injury. 2021;52(4):918-25.

26. Patel VP, Walsh M, Sehgal B, Preston C, DeWal H, Di Cesare PE. Factors associated with prolonged wound drainage after primary total hip and knee arthroplasty. J Bone Joint Surg Am. 2007;89(1):33-8.

27. Ahmed AA, Mooar PA, Kleiner M, Torg JS, Miyamoto CT. Hypertensive patients show delayed wound healing following total hip arthroplasty. PLoS One. 2011;6(8):e23224.

28. Strahovnik A, Fokter SK, Kotnik M. Comparison of drainage techniques on prolonged serous drainage after total hip arthroplasty. J Arthroplasty. 2010;25(2):244-8.