Research Article

Construction of Prediction Model of Deep Vein Thrombosis Risk after Total Knee Arthroplasty Based on XGBoost Algorithm

Yuhuan Chen ![ORCID] **and Yingqing Jiang** ![ORCID]

Department of Orthopedics, The First Affiliated Hospital of Soochow University, Suzhou 215006, China

Correspondence should be addressed to Yingqing Jiang; jiangyingqing2021@163.com

Objective. Based on the XGBoost algorithm, the prediction model of the risk of deep vein thrombosis (DVT) in patients after total knee arthroplasty (TKA) was established, and the prediction performance was compared. *Methods*. A total of 100 patients with TKA from January 2019 to December 2020 were retrospectively selected as the study subjects and randomly divided into a training set ($n = 60$) and a test set ($n = 40$). The training set data was used to construct the XGBoost algorithm prediction model and to screen the predictive factors of postoperative DVT in TKA patients. The prediction effect of the model was evaluated by using the test set data. An independent sample T-test was used for comparison between groups, and the χ^2 test was used for comparison between counting data groups. *Results*. The top five items were combined with multiple injuries (35 points), time from injury to operation (28 points), age (24 points), combined with coronary heart disease (21 points), and D-dimer 1 day after operation (16 points). In the training set, the area under the curve of the XGBoost algorithm model was 0.832 (95% CI: 0.748-0.916). *Conclusion*. The model based on the XGBoost algorithm can predict the incidence of DVT in patients after TKA with good performance.

1. Introduction

Total knee arthroplasty (TKA) is an effective method for the treatment of end-stage knee osteoarthritis, which can significantly relieve the pain in patients and promote functional recovery [1]. Deep vein thrombosis (DVT) is one of the common complications of limb fractures in the perioperative period, which can lead to pulmonary embolism and postthrombotic syndrome. The rate of sudden death caused by pulmonary embolism is as high as 34%, which seriously affects the prognosis and life of patients [2]. DVT is a venous reflux disease caused by blood coagulation in the deep vein due to various reasons. There are three factors in patients with DVT: slow blood flow, hypercoagulability, and venous wall injury [3]. Previous studies showed that the overall incidence of DVT was 69.9% per patient postoperatively [4]. Therefore, the early diagnosis and treatment of DVT are extremely important.

In recent years, the extreme Gradient Boosting (XGBoost) model has become one of the common tools for solving classification problems due to its high efficiency, flexibility, and accuracy [5]. The XGBoost model is widely used in statistics, data mining, machine learning, and artificial intelligence, because of its low input data requirements, automatic variable selection, and low computational complexity. Especially in the medical field, the XGBoost model can make use of historical case data to build models for classification and predictions [6]. Direct prediction of postoperative indicators with clinical data and specific algorithms is helpful for doctors to make reasonable and effective judgments and further guide the follow-up medical work.

In recent years, a large number of studies have reported high-risk factors related to DVT [7], but there are few studies on predicting DVT after TKA by the XGBoost model. Therefore, the purpose of this study is to explore the risk factors for DVT in patients after TKA. The prediction model
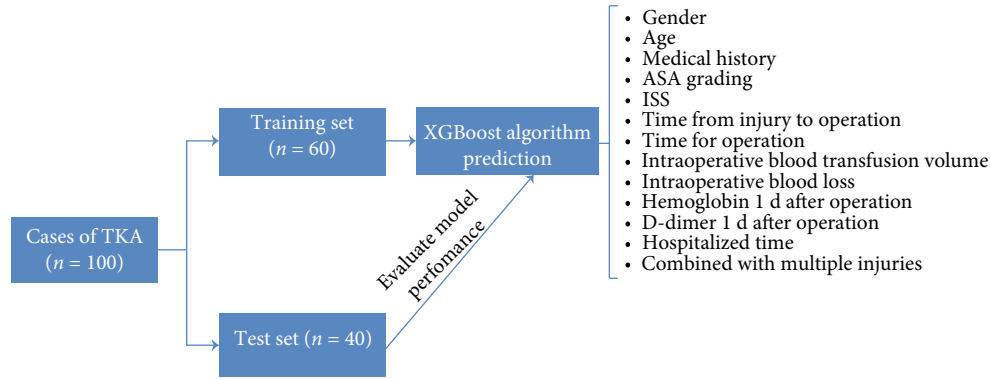
FIGURE 1: Modeling flow chart.

based on the XGBoost algorithm is aimed at providing reliable theoretical guidance for the early intervention of DVT after TKA.

2. Materials and Methods

2.1. General Information. From January 2019 to December 2020, 100 patients with TKA who met the inclusion and exclusion criteria were selected as research objects.

Inclusive criteria are patients younger than 80 years old who received unilateral TKA for the first time and without cognitive impairment and able to cooperate with the researcher and patients with stable condition at discharge without postoperative complications and who informed and agreed to participate in the study.

Exclusion criteria are patients with severe cardiovascular and cerebrovascular diseases and malignant tumors, patients with mental disorders who are unable to cooperate, patients with hip and ankle joint deformities that affect knee joint function, patients who have undergone TKA surgery or bilateral TKA surgery at the same time, and recent participants in other researches.

2.2. Research Queue. All patients with TKA were randomly divided into a training set ($n = 60$) and a test set ($n = 40$). The training set was used to construct the XGBoost algorithm prediction model, and the test set data was used to evaluate the prediction effect of the XGBoost model. According to the ultrasound results of both lower limbs after the operation, the patients in the training set were divided into a DVT group and a non-DVT group.

2.3. Indicators and Standards of Observation. Collecting baseline data include age, gender, previous medical history (hypertension and coronary heart disease), American Society of Anesthesiologists (ASA) classification, injury severity score (ISS), time from injury to operation, operation time, intraoperative blood transfusion volume, intraoperative blood loss, hemoglobin 1 day after operation, D-dimer 1 day after operation, hospital time, and combined with multiple injuries.

ISS<16 is a minor injury, $16 < ISS \leq 25$ is a moderate injury, and $ISS > 25$ is a severe injury [8]. According to the ASA classification, patients' condition was divided into five grades: Grade I: healthy; Grade II: suffering from systemic disease, with good compensatory function; Grade III: severe systemic disease, daily functions are limited within the scope of compensation; Grade IV: compensatory function is not complete, and daily function is lost; Grade V: the condition is serious and life-threatening [9].

2.4. Treatment Method. The risk assessment of thromboembolism was carried out regularly for all patients after operation. To prevent DVT, patients without contraindications of anticoagulation were injected subcutaneously with low molecular weight heparin sodium 0.4 mL every day. At the same time, the plantar vein pump was performed twice a day for 20 minutes each time. One day after the operation, blood samples were collected from all patients, and the D-dimer, blood routine, and coagulation were monitored. Ultrasound examination of veins of lower limbs was performed 3-5 days after the operation. Positive patients were given low molecular heparin calcium 0.4 mL twice a day, and the plantar venous pump was stopped for DVT treatment. All patients stopped using anticoagulants 12 hours before operation and started using anticoagulants 12 hours after operation.

2.5. Research Outcomes and Predictors. The outcome of the study was that DVT occurred in TKA patients 3-5 days after the operation. The inner diameter of the deep vein of lower limbs was observed from bottom to top along with the operation site, and DVT was diagnosed if the following phenomena occur: The lumen below the embolic site is enlarged, and the venous lumen cannot be closed. There is a solid echo with unequal intensity in the lumen. When the occlusion is complete, pulse and color Doppler show a small amount or no blood flow signal.

2.6. XGBoost Algorithm Model. The XGBoost algorithm is an optimized distributed gradient computing integration algorithm. The idea of this algorithm comes from the gradient lifting iterative decision tree algorithm. Based on the gradient lifting iterative decision tree algorithm, a second-order Taylor function is added to classify the data, which is specifically completed by the following:

Goal function:

$$\overset{\wedge}{y_i} = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F, \tag{1}$$

where $i = 1, 2, \cdots, n$, n is the number of samples, F is the set corresponding to all regression trees, and f_k is the function in F.

The regular term is the damage function of the model:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2, \quad l(y_{i3}\hat{y}_i). \tag{2}$$

The second-order Taylor function is introduced to approximately expand the loss function, and the objective function is optimized to be closer to the actual value to improve the prediction accuracy:

$$o = l(y, y_i) + \sum_{k=1}^{k} \beta(f_k), \tag{3}$$

where $l(y, y_i)$ is a model for integrating data standardization, O represents the difference between the predicted value of the last formula and the recorded value of the actual data, and β represents the normalized processing coefficient, which is used to calculate the positive value after data weighting to prevent data confusion. Formulas (1) and (2) are fused by superposition calculation of multiple data, and the predicted values of iterative samples of data are substituted into the loss function, and the calculation results are multiplied by the normalization coefficient. The final simplified formula of the XGBoost algorithm is as follows:

$$o^t = \sum_{i=1}^{N} \left(y_i, y^{t-1} + g_i f_i(x_i) + \frac{1}{2} h_i y_i(x_i) \right) + \beta(f_i). \tag{4}$$

2.7. Statistical Methods. Single-factor analysis of variance was used to train the clinical data of the DVT group and the non-DVT group. SPSS20.0 software was used for the statistical analysis of the data. The independent sample T-test was used for comparison between groups, and the χ^2 test was used for comparison between counting data groups.

3. Results

3.1. Process of Model Establishment. 100 patients with TKA were randomly divided into the training set ($n = 60$) and test set ($n = 40$). The training set was used to construct the XGBoost algorithm prediction model, and the test set data was used to evaluate the prediction effect of the model. According to the ultrasound results of both lower limbs after operation, the patients in the training set were divided into the DVT group ($n = 24$) and the non-DVT group ($n = 36$). The predictive factors included all the items of observed indicators. Figure 1 shows the specific modeling process.

3.2. Clinical Data of Patients in Training Set and Test Set. There was no significant difference between the training set and the test set in terms of gender, age, complications, ASA grade, ISS, time from injury to operation, operation time, intraoperative blood transfusion volume, intraoperative blood loss, hemoglobin 1 day after operation, D-dimer 1 day after operation, hospitalized time, and combined with multiple injuries. In addition, there was no significant difference in diabetes mellitus and hyperlipidemia, ASA grade, ISS, operation time, intraoperative blood transfusion volume, intraoperative blood loss, and hospitalized time between the DVT group and the non-DVT group. There were statistically significant differences between the two groups in terms of gender, age, hypertension, coronary heart disease, time from injury to operation, hemoglobin, D-dimer at 1 day after operation, and combined with multiple injuries (Figure 2, Table 1).

3.3. Establishment of XGBoost Algorithm Model. By integrating the data in the training set into the XGBoost algorithm model, we have obtained important feature scoring results. The top five items were combined with multiple injuries (35 points), time from injury to operation (28 points), age (24 points), combined with coronary heart disease (21 points), and D-dimer 1 day after operation (16 points). In addition, the other three items were combined with hypertension (14 points), hemoglobin 1 day after operation (11 points), and gender (9 points), as shown in Figure 3.

3.4. Calibration Degree of XGBoost Algorithm Model. As shown in Figure 4, in the training set, the area under the curve of the XGBoost algorithm model was 0.832 (95% CI: 0.748-0.916).

4. Discussion

With the change of population structure, the number of elderly patients with TKA is increasing [10]. DVT is a common complication after orthopedic surgery. Previous studies showed that the incidence of postoperative DVT was as high as 57%-62%, and the incidence of DVT was 0.5%-44% even with routine prevention [11]. Bawa et al. showed that patients with hypercoagulability diagnoses were at a higher risk of postoperative DVT [12]. Tateiwa et al. found that the total incidence of DVT was 62.5% in 88 patients with TKA. Among the 55 patients with DVT, the incidence of distal DVT was 96.4%, and that of proximal DVT was 3.6% [13]. These results were similar to those of our study. The reason may be that TKA has caused great damage to tissues and blood vessel walls and caused more hemorrhage. In addition, patients with postoperative pain often stay in bed for braking, which further aggravates blood stasis and increases the risk of DVT. Therefore, for TKA patients, it is necessary to attach great importance to the occurrence of DVT and strengthen the screening and prevention of thrombosis.

Clinical studies have found that DVT after TKA is a risk factor for poor prognosis and death [14]. However, the risk factors leading to DVT are still inconclusive. According to the analysis of hip fracture patients, it is found that the average time interval between operation and hospitalization for patients with DVT after the operation was 9.45 days, while

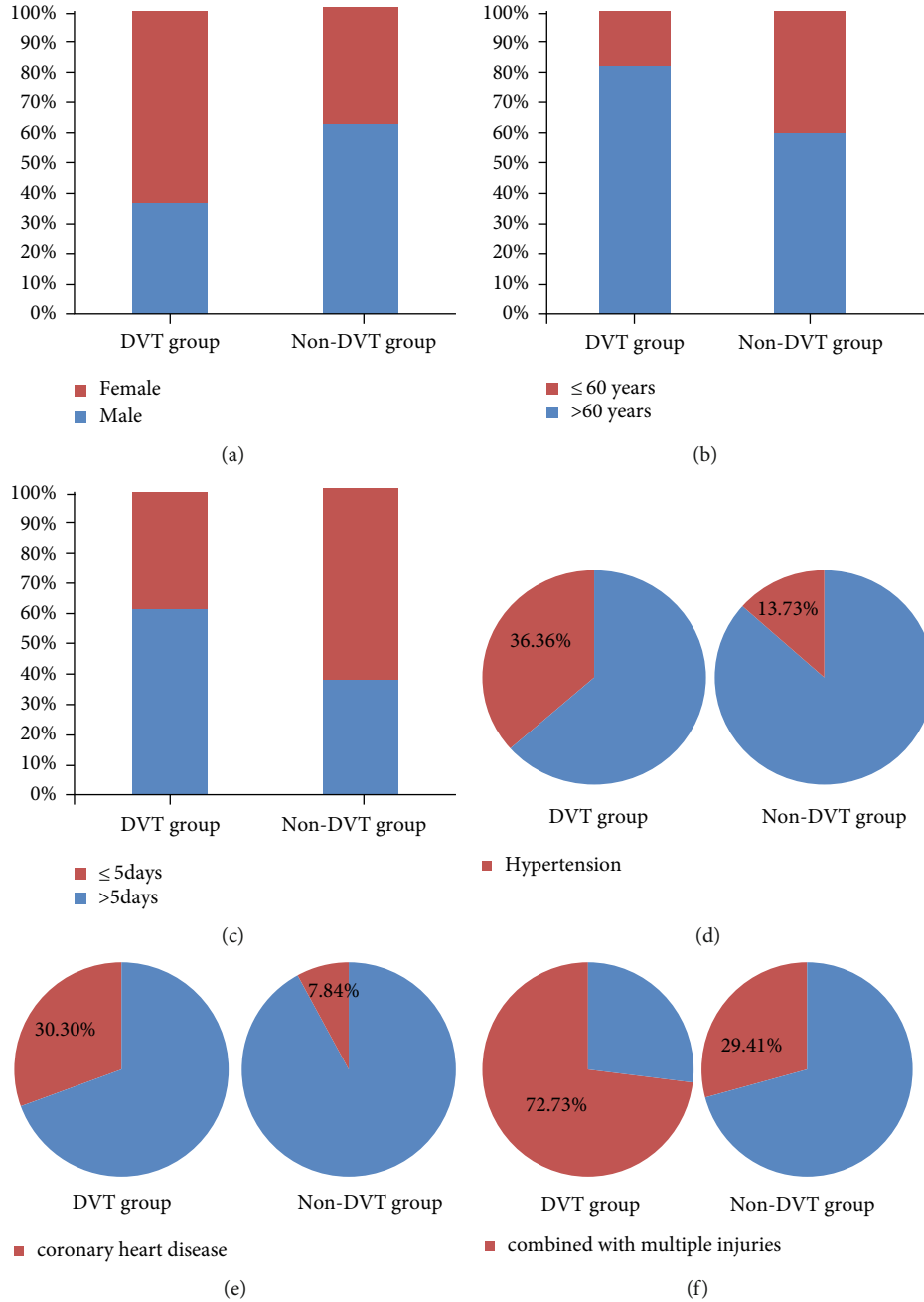(a)

(b)

(c)

(d)

(e)

(f)

Figure 2: Comparison of clinical data between the DVT group and the non-DVT group in the training set. There were significant differences ($P < 0.05$) between the DVT and non-DVT groups in gender (a), age (b), time from injury to operation (c), hemoglobin (d), coronary heart disease (e), and combined with multiple injuries (f).

Table 1: Comparison of clinical data between the DVT group and the non-DVT group in the training set 1 day after operation (mean ± SD).

Index	DVT group	Non-DVT group	t value	P value
Hemoglobin (g/L)	107.25 ± 16.33	118.39 ± 17.52	2.922	<0.05
D-dimer (mg/L)	6.24 ± 4.87	2.66 ± 3.21	4.065	<0.001

that for patients without DVT was 1.06 days [15]. In this study, we found that the time from injury to operation over 5 days was an independent risk factor for DVT after TKA. This result is similar to previous studies [16]. This may be due to the long-term braking, traction, or pain after the operation, which leads to prolonged stay in bed, muscle atrophy, and hypercoagulability which then leads to the formation of DVT. Therefore, patients should be admitted to the hospital as soon as possible after being injured. Surgery should be performed as soon as possible when conditions permit to reduce the risk of DVT.
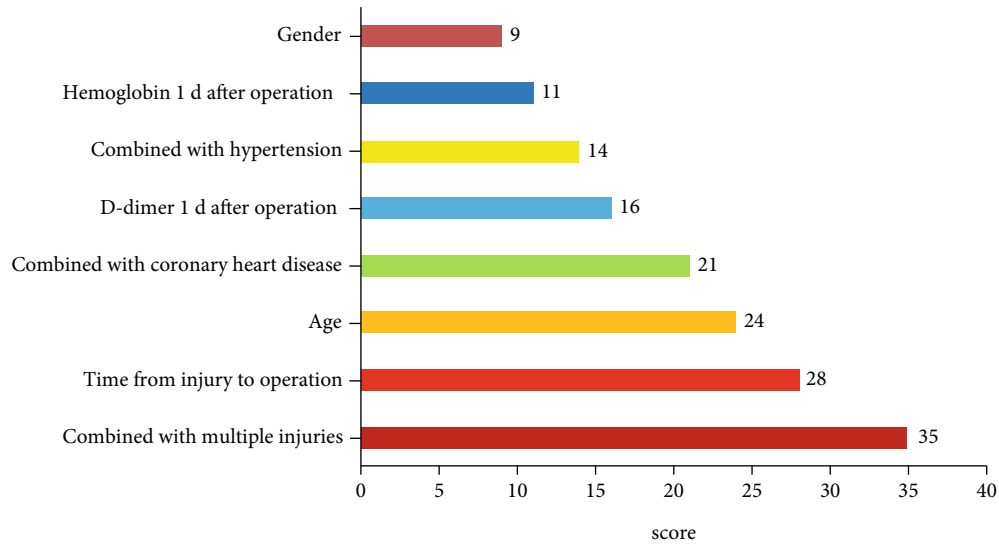
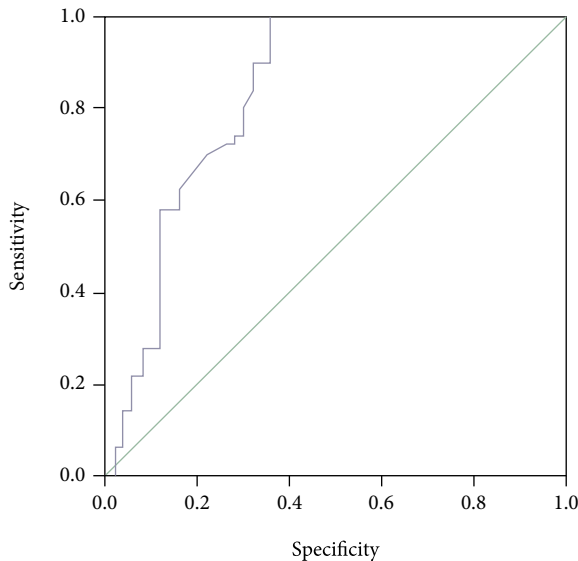FIGURE 3: Score of important features in XGBoost algorithm model.

FIGURE 4: The receiver operation characteristic (ROC) curves of XGBoost algorithm model.

Patients over 60 years have decreased physical function, easily damaged intima of blood vessels, and decreased elasticity of blood vessels and blood circulation, which increases the risk of DVT. Nybo and Hvas provided that age is a risk factor for DVT [17]. Gómez-Jabalera et al. demonstrated that age over 60 years was an independent risk factor for DVT in patients after fracture operation, and it was highly specific for predicting DVT [18]. These were also consistent with our study. However, some scholars believe that the incidence of DVT has nothing to do with age. This may be due to the timely screening of DVT after injury [11]. In addition, we also found that coronary heart disease was an independent risk factor for DVT after TKA. Our results are consistent with those of previous scholars [19]. The reason may be that coronary atherosclerosis in patients with coronary heart disease leads to vascular endothelial damage, promotes

platelet adhesion and aggregation, and accelerates the coagulation process, thus leading to the increased risk of postoperative DVT. D-dimer is the simplest degradation product of fibrin, and the increase of mass concentration can reflect the existence of hypercoagulability in the body, which is a nonspecific sign of thrombosis. It has been widely applied to the diagnosis and prognosis of clinical thrombotic diseases. Our study found that D-dimer was an independent risk factor for DVT after TKA, which was consistent with previous studies [20].

Davagdorj et al. built a model of noncommunicable diseases caused by smoking based on the XGBoost framework, which can realize a more accurate prediction [5]. In addition, in the diagnosis of chronic kidney disease, Ogunleye and Wang proved that the XGBoost model has a great advantage in prediction accuracy [21]. The XGBoost algorithm model established in this study has high sensitivity and specificity in predicting DVT in patients after TKA operation. Various algorithms have their own advantages and disadvantages. We will try to use new algorithms to evaluate the prognosis of the disease in the future [22, 23].

There are some limitations to our study. This study was a retrospective study of a small single-center sample, which needs further verification in the prospective study of a large multicenter sample. Secondly, there are not enough influencing factors in the analysis, and there was no hierarchical analysis of these factors. In addition, although the XGBoost algorithm model has unique advantages in dealing with high-dimensional variables, complex interactions among variables, and nonlinear relationships, the effectiveness of the prediction model will also be limited by the nature, number, type, and sample of variables.

5. Conclusion

The model based on the XGBoost algorithm can predict the occurrence of DVT in patients after TKA with good prediction performance. The items including combined with

multiple injuries, time from injury to operation, age, coronary heart disease, and D-dimer at 1 day after operation can be used as predictive indicators of DVT after TKA.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

All authors declare no conflicts of interest in this paper.

References

[1] A. M. Schwartz, K. X. Farley, G. N. Guild, and T. L. Bradbury Jr., "Projections and epidemiology of revision hip and knee arthroplasty in the United States to 2030," *The Journal of Arthroplasty*, vol. 35, no. 6, pp. S79–s85, 2020.

[2] H. Meng, Y. Zhu, J. Zhang et al., "Incidence and risk factor for preoperative deep vein thrombosis (DVT) in isolated calcaneal fracture, a prospective cohort study," *Foot and Ankle Surgery*, vol. 27, no. 5, pp. 510–514, 2021.

[3] J. M. Lurie, C. Y. M. Png, S. Subramaniam et al., "Virchow's triad in "silent" deep vein thrombosis," *Journal of Vascular Surgery. Venous and Lymphatic Disorders*, vol. 7, no. 5, pp. 640–645, 2019.

[4] M. J. Chang, M. K. Song, M. G. Kyung, J. H. Shin, C. B. Chang, and S. B. Kang, "Incidence of deep vein thrombosis before and after total knee arthroplasty without pharmacologic prophylaxis: a 128-row multidetector ct indirect venography study," *BMC Musculoskeletal Disorders*, vol. 19, no. 1, p. 274, 2018.

[5] K. Davagdorj, V. H. Pham, N. Theera-Umpon, and K. H. Ryu, "XGBoost-based framework for smoking-induced noncommunicable disease prediction," *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, pp. 6513–6522, 2020.

[6] S. K. Devana, A. A. Shah, C. Lee, A. R. Roney, M. van der Schaar, and N. F. SooHoo, "A novel, potentially universal machine learning algorithm to predict complications in total knee arthroplasty," *Arthroplast Today*, vol. 10, no. 1, pp. 135–143, 2021.

[7] M. A. Alsheef, A. M. Alabbad, R. A. Albassam et al., "Pregnancy and venous thromboembolism: risk factors, trends, management, and mortality," *BioMed Research International*, vol. 2020, Article ID 4071892, 6 pages, 2020.

[8] S. M. Galvagno Jr., M. Massey, P. Bouzat et al., "Correlation between the revised trauma score and injury severity score: implications for prehospital trauma triage," *Prehospital Emergency Care*, vol. 23, no. 2, pp. 263–270, 2019.

[9] K. M. Knuf, C. V. Maani, and A. K. Cummings, "Clinical agreement in the American Society of Anesthesiologists physical status classification," *Perioperative Medicine*, vol. 7, no. 1, pp. 1–6, 2018.

[10] M. Weber, F. Völlner, A. Benditz et al., "Total knee arthroplasty in the elderly," *Orthopade*, vol. 46, no. 1, pp. 34–39, 2017.

[11] F. Bengoa, G. Vicencio, D. Schweitzer, M. J. Lira, T. Zamora, and I. Klaber, "High prevalence of deep vein thrombosis in elderly hip fracture patients with delayed hospital admission," *European Journal of Trauma and Emergency Surgery*, vol. 46, no. 4, pp. 913–917, 2020.

[12] H. Bawa, J. W. Weick, D. R. Dirschl, and H. H. Luu, "Trends in deep vein thrombosis prophylaxis and deep vein thrombosis rates after total hip and knee arthroplasty," *The Journal of the American Academy of Orthopaedic Surgeons*, vol. 26, no. 19, pp. 698–705, 2018.

[13] T. Tateiwa, T. Ishida, T. Masaoka et al., "Clinical course of asymptomatic deep vein thrombosis after total knee arthroplasty in Japanese patients," *Journal of Orthopaedic Surgery*, vol. 27, no. 2, 2019.

[14] M. Sloan, N. Sheth, and G. C. Lee, "Is obesity associated with increased risk of deep vein thrombosis or pulmonary embolism after hip and knee arthroplasty? A large database study," *Clinical Orthopaedics and Related Research*, vol. 477, no. 3, pp. 523–532, 2019.

[15] Y. H. Cho, Y. S. Byun, D. G. Jeong, I. H. Han, and Y. B. Park, "Preoperative incidence of deep vein thrombosis after hip fractures in Korean," *Clinics in Orthopedic Surgery*, vol. 7, no. 3, pp. 298–302, 2015.

[16] B. F. Zhang, X. Wei, H. Huang et al., "Deep vein thrombosis in bilateral lower extremities after hip fracture: a retrospective study of 463 patients," *Clinical Interventions in Aging*, vol. - Volume 13, pp. 681–689, 2018.

[17] M. Nybo and A. M. Hvas, "Age-adjusted D-dimer cut-off in the diagnostic strategy for deep vein thrombosis: a systematic review," *Scandinavian Journal of Clinical and Laboratory Investigation*, vol. 77, no. 8, pp. 568–573, 2017.

[18] E. Gómez-Jabalera, S. Bellmunt Montoya, E. Fuentes-Camps, and J. R. Escudero Rodríguez, "Age-adjusted D-dimer for the diagnosis of deep vein thrombosis," *Phlebology*, vol. 33, no. 7, pp. 458–463, 2018.

[19] K. Kaikita, K. Hosokawa, J. R. Dahlen, and K. Tsujita, "Total thrombus-formation analysis system (t-tas): clinical application of quantitative analysis of thrombus formation in cardiovascular disease," *Thrombosis and Haemostasis*, vol. 119, no. 10, pp. 1554–1562, 2019.

[20] J. Gregson, S. Kaptoge, T. Bolton et al., "Cardiovascular risk factors associated with venous thromboembolism," *JAMA Cardiology*, vol. 4, no. 2, pp. 163–173, 2019.

[21] A. Ogunleye and Q. G. Wang, "XGBoost model for chronic kidney disease diagnosis," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 17, no. 6, pp. 2131–2140, 2020.

[22] Z. Tang, Y. Li, X. Chai, H. Zhang, and S. Cao, "Adaptive nonlinear model predictive control of nox emissions under load constraints in power plant boilers," *Journal of Chemical Engineering of Japan*, vol. 53, no. 1, pp. 36–44, 2020.

[23] N. Li, P. Luo, C. Li, Y. Hong, M. Zhang, and Z. Chen, "Analysis of related factors of radiation pneumonia caused by precise radiotherapy of esophageal cancer based on random forest algorithm," *Mathematical Biosciences and Engineering*, vol. 18, no. 4, pp. 4477–4490, 2021.