Infection Management after Arthroplasty: Machine Learning and Deep Learning as Tools

Zandi R. MD[*], Ahmadi Abdashti A. R. MD[**], Ahmadi Abdashti Z. M.Sc[***]
Amouzadeh F. MD[****], Vahidi Sh. PhD[*****]

Abstract:

Background and Objective: Nowadays, with the expansion of human knowledge, the use of artificial intelligence (AI), machine learning (ML), and deep learning (DL) has found applications in the field of medicine. In the context of infections following hip and knee arthroplasties in orthopedics, machine learning and deep learning have been utilized for assessing the significance of risk factors, evaluating the importance of diagnostic and prognostic criteria, and analyzing the sensitivity of pathogens to antibiotics. Additionally, AI tools can be employed in patient education and in answering common patient questions regarding periprosthetic joint infection (PJI).

Materials & Methods: In this narrative study, 22 articles from the past decade, extracted from the PubMed database, were reviewed, focusing on the use of artificial intelligence, machine learning, and deep learning in PJI. We highlighted the aspects analyzed by AI, including risk factors, diagnostic criteria, prognostic factors, and effective antibiotics against bacteria. Furthermore, we recalled common machine learning models along with their respective sensitivity, specificity, and accuracy. We also examined a study regarding the use of ChatGPT-4 for educating and responding to patients with PJI.

Results: Given that complex and nonlinear data are effectively and rapidly analyzed by machine learning, significant improvements in the accuracy and speed of PJI management have been achieved through machine learning. Additionally, the use of deep learning has further increased accuracy compared to traditional machine learning methods. Common limitations shared across studies primarily included their retrospective nature and the small sample sizes of the studied populations.

Conclusion: Considering the capabilities of AI, machine learning and deep learning to analyze data more quickly and accurately compared to traditional methods, utilizing these technologies in various aspects of medicine, including PJI management, is reasonable.

Keywords: Machine Learning, Infection, Hip Arthroplasty, Knee Arthroplasty

[*] Associate Professor of Orthopedic Surgery, Musculoskeletal Injuries Research Center, Shahid Beheshti University of Medical Sciences

[**] Orthopedic Surgeon and Researcher at Musculoskeletal Injuries Research Center, Shahid Beheshti University of Medical Sciences

[***] Department of Electrical Engineering, Electronic, Micro and Nano Electronic Devices, Faculty of Mechanic, Electrical and Computer, Science and Research branch, Islamic Azad University

[****] Associate Professor of Orthopedic Surgery, Shahid Beheshti University of Medical Sciences

[*****] PhD Student of Health Education and Promotion, Shahid Beheshti University of Medical Sciences

Received: 09/11/2024
Accepted: 29/03/2025

Corresponding Author: Dr. Ahmadreza Ahmadi Abdashti
Tel: 02122439784

E-mail: ahmadrezaahmadi@sbmu.ac.ir

Background and Objective

Arthroplasty is a widely performed surgical intervention for the hip and knee joints. A recent study conducted in Germany indicated that among individuals aged 70 years and older, the incidence of primary hip arthroplasties was 1.1% from 2007 to 2014, while the rates for primary knee arthroplasties were 0.7% in 2007 and 0.6% in 2014. In 2014, the prevalence of hip surgery in relation to the overall population was 0.26%, compared to 0.19% for knee arthroplasty. Among the cohort aged 70 years and older in 2014, the rates of revision arthroplasties (including corrections without replacement) were 0.19% for the hip and 0.10% for the knee.[1] Several key factors influence the decision to pursue total hip arthroplasty (THA) or total knee arthroplasty (TKA), including pain severity, functional outcomes, radiological changes, and the failure of conservative treatments.[2] The most prevalent in-hospital complications for patients undergoing THA include fractures (0.6%) and deep vein thrombosis (DVT) (0.6%). In patients undergoing TKA, the incidence of DVT increases to 1.4%, while cardiac events are observed at a rate of 0.8%. Postoperative complications following hospital discharge for THA and TKA patients frequently necessitate reoperations due to issues such as bleeding, wound necrosis, wound infection, or various other reasons, with infection presenting one of the most significant challenges following both procedures.[3] The overall incidence of surgical site infection (SSI) following primary total hip arthroplasty varies from 0.2% to 2.2%. Additionally, the incidence of deep joint infection, or peri-prosthetic joint infection (PJI), among patients undergoing THA and TKA ranges from 0.3% to 1.9%, potentially escalating to as high as 10% in revision cases.[4] Numerous risk factors have been identified that correlate with an increased likelihood of PJI. These include male gender, younger age, Type II diabetes, class II obesity, hypertension, hypoalbuminemia, preoperative nutritional status, body mass index, rheumatoid arthritis, post-traumatic osteoarthritis, prior intra-articular injections before TKA, previous multi-ligament knee surgeries, prior steroid treatment, current tobacco use, bilateral cases, hospital stays exceeding 35 days, patellar resurfacing, prolonged surgical duration, blood transfusion use, and significant variability in glucose levels following surgery. Less significant factors have been omitted for brevity.[5] In 2018, a consensus was established regarding the diagnosis of PJI based on criteria derived from synovial fluid analysis, the presence of a fistula at the surgical site, levels of C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), alpha-defensin, D-dimer, polymorphonuclear cells (PMN), histological findings, and presence of pus at the surgical site. Nevertheless, certain cases continue to pose diagnostic challenges, making definitive diagnoses elusive.[6] The primary interventions for managing PJI include debridement, antibiotic therapy, and irrigation with implant retention (DAIR). In specific instances, revisions may be performed as either one-stage or two-stage procedures.[7-9] Successful management of PJI hinges on the careful selection of an appropriate treatment strategy, consideration of the patient's unique characteristics, and acknowledgment of immutable infection-related factors. Therefore, interdisciplinary decision-making is paramount in the effective management of PJI.[10]

The rapid advancements in artificial intelligence (AI) and machine learning (ML) have opened new avenues for evaluating the risk, diagnosis, treatment, and prognosis of patients with peri-prosthetic joint infection (PJI). The primary aim of AI is to replicate machine behaviors that closely resemble human cognitive processes.[11] Fundamentally, machine learning serves as a simulation of the human mind and can be categorized into four principal types: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. By utilizing extensive datasets, machine learning generates algorithms capable of identifying novel features and assigning varying weights to enhance accuracy. In supervised learning, the machine is provided with both input data and corresponding outcomes, which are divided into two groups: the training set and the test set. This phase involves the development of an algorithm that enables the computer to discern relationships within the data, thereby facilitating predictions of specific outcomes based on its training.

Conversely, unsupervised learning seeks to uncover patterns within datasets that lack predefined outcomes. Semi-supervised learning integrates aspects of both supervised and unsupervised learning, functioning particularly well with datasets that comprise a mixture of labeled and unlabeled data. This is frequently observed in medical imaging, where a physician labels a limited subset of images, which the model subsequently utilizes to classify the remaining unlabeled images. The labeled dataset contributes to a more effective operational model in comparison to unsupervised approaches. Reinforcement learning trains algorithms to perform specific tasks without a singular correct answer, focusing on achieving an overall desired outcome through trial and error, employing rewards and penalties throughout the learning process. Although this method is less frequently employed in the medical field, it closely parallels human learning processes.[12] Deep learning, a subset of machine learning,[13,14] enables computers to analyze vast amounts of unlabeled data through artificial neural networks.[15] The key distinction between these approaches lies in the degree of human intervention; while machine learning models necessitate human oversight for enhanced accuracy, deep learning models possess the capability to self-correct errors autonomously.[14] Machine learning aspires to make informed predictions in scenarios devoid of precise solutions, whereas deep learning encapsulates knowledge within a hierarchical framework of interconnected concepts.[13] Contrary to popular belief, machine learning can operate efficiently on smaller datasets,[16] while deep learning excels in analyzing large volumes of data, presenting significant advantages over traditional methods.[17]

This article endeavors to explore the common machine learning and deep learning models relevant to research on PJI and their clinical applications.

Materials and Methods

This article addresses two principal inquiries: What role can machine learning play in the prevention and management of peri-prosthetic joint infection (PJI), and what are the most prevalent models and algorithms employed in this domain?

To investigate these questions, we conducted a comprehensive search within the PubMed database utilizing keywords such as "Total Knee Arthroplasty," "Total Hip Arthroplasty," "Joint Replacement," "Infection," "Machine Learning," and "Artificial Intelligence," focusing on studies published within the last decade. A specialist in orthopedics meticulously screened the titles and abstracts to select the most pertinent articles for detailed review.

Findings

1- Prediction

One notable advantage of machine learning (ML) is its enhanced accuracy in predicting diseases and complications through the analysis of risk factor patterns. In contrast to traditional statistical models that primarily establish relationships and assess variable importance, ML can predict outcomes based on previously unseen data.[18]

Within the context of PJI prediction, the AUTO-PROGNOS model identifies key risk factors, including malnutrition, dementia, malignancy, chronic obstructive pulmonary disease (COPD), and insurance status.[19]

In cases of recurrent PJI following revision procedures, a deep learning model known as the Artificial Neural Network (ANN) has demonstrated effectiveness, revealing that prior irrigation and debridement (I&D)—regardless of whether accompanied by component exchange—serves as a significant predictor. Moreover, microbiological factors such as Enterococcus and Methicillin-resistant Staphylococcus aureus (MRSA) are associated with recurrence, with MRSA recognized as a critical risk factor within the ANN model. Additional risk factors for recurrent PJI include more than four prior surgeries, metastasis, HIV/AIDS, substance abuse, and obesity.[20] In scenarios where PJI develops after revision due to non-infectious causes, the ANN model can also predict outcomes, incorporating factors such as prior surgeries, substance use, and diabetes.[21] For the prediction of treatment failure following I&D for PJI, the Random Forest (RF) algorithm has shown superior performance compared to traditional scoring systems like KLIC and CRIME80, which assess the likelihood of success in washout and debridement procedures.[18] Furthermore, the ANN model has been utilized to evaluate risk factors for surgical site infections (SSI) following arthroplasty, with the Canadian Classification of Health Interventions (CCI), obesity, and smoking emerging as significant predictors. Notably, deep learning has underscored the impact of obesity on infection risk to a greater extent than previously recognized, attributed to its ability to analyze complex and nonlinear data relationships.[22] Concerning the risk of effusion post-hemarthroplasty, the Stacking model has identified four primary risk factors: accumulated fluid volume from the drain, morbid obesity, blood class, and elevated levels.[23] A summary of the findings from the prediction section is presented in Table 1.

2- Diagnosis

Machine learning techniques are proving to be invaluable in determining optimal diagnostic methodologies for a range of diseases. In the context of infections following arthroplasty, algorithms such as XGBoost have significantly enhanced the detection of surgical site infections (SSI).[24] Furthermore, models including Random Forest (RF), Uniform Manifold Approximation and Projection (UMAP), and Markov chains have identified C-reactive protein (CRP) as the most critical diagnostic marker for peri-prosthetic joint infection (PJI). Additional notable factors influencing the diagnosis of PJI comprise decreased blood urea nitrogen (BUN), reduced hemoglobin (Hb) levels, elevated total protein, and increased alanine aminotransferase (ALT). A synergistic combination of biomarkers, particularly high levels of CRP and low levels of hemoglobin, presents an excellent diagnostic combination for PJI. Markov chain analysis has revealed that patients exhibiting PJI post-surgery have an increased risk of transmitting the infection to higher-risk groups, indicating a significant risk of reinfection.[25] Techniques such as K-means clustering and Principal Component Analysis (PCA), in conjunction with decision trees, have further improved the accuracy of ambiguous PJI diagnoses.[26] The application of Support Vector Machines (SVM) alongside the TRACE4 platform in evaluating the significance of MRI findings has emphasized that the most specific MRI features for diagnosing PJI include synovitis, lamellated synovitis, and extracapsular edema. Notably, the presence of bone edema has emerged as the most sensitive finding, boasting a high Negative Predictive Value (NPV). Conversely, indicators such as bone destruction, soft tissue masses surrounding the joint, and fibrotic membranes were not associated with infectious failure in total hip arthroplasty (THA), suggesting that they are more reflective of non-infectious treatment failures.[27] Recent advancements in deep learning methodologies have further enhanced the analysis of dynamic bone scans (Dynamic Bone Scan - DBS). The DBS-eNet (Dynamic Bone Scan Efficient Neural Network) model has demonstrated superior sensitivity and specificity compared to traditional bone scintigraphy methods for diagnosing PJI. It is noteworthy that specialists in nuclear medicine have exhibited improved accuracy in diagnosing infections associated with knee prostheses relative to those associated with hip prostheses.[28]

Table 1- Use of ML in evaluating predictor factors

Author/Year of Publication	*Machine Learning Model (ML)*	*Interpretability*	*Type of Arthroplasty*	*Description*	*Results*
Yeo I et al. (2023)[17]	• *ANN* • *SGB* • *SVM* • *RF* • *Elastic-Net Penalized Logestic Regression*	• *Accuracy > 94.0%* • *AUC = 0.84* • *Calibration Intercept = 0.09* • *Calibration Score = 1.06* • *Brier: 0.054* • *All 5 models have excellent performance in terms of Discrimination, Calibration and Decision Curve Analysis, but ANN is better than the rest.*	*Primary TKA*	*To prevent overfitting, a Five-Fold Validation as well as 80-20 Validation and Recursive Feature Elimination were used.*	• *SSI predictors included CCI, obesity and cigarettes.* • *CCI is most important.Obesity according to ML, became more important than in the past.*
Akash A Shah (2021)[14]	*Autoprognosis: Includes 5 MLBenchmark models* • *Logistic regression (A Liner Classifier)* • *Random Forest (A Tree Based Ensemble Classifier)* • *AdaBoost* • *Gradient Boosting Machine* • *XGBoost (Boosting Ensemble Classifier)*	• *discrimination (AUROC: 0.732 ± 0.01)* • *Brier score of 0.007 ± 0.002* • *The calibration plot demonstrates the enhanced efficacy of the Autoprognosis model.*	*THA*	• *Hyperparents of each model were selected by Grid Search. L2 regularisation was used for logistics regression.* • *Utilization of Five-Fold Cross Validation and 80-20 Validation for validation.*	• *The autoprogenosis model was better than other checked models.* • *5 important features of whistle, dementia, malignancy, COPD, autoprogenosis insurance were important for PJI forecast.*
Christian Klemt et al. (2021)[15]	• *ANN* • *SGB* • *ENP*	• *Discrimination in the ANN model had AUC = 0.84 in the test group.* • *The model was also approved by Calibration and Decision Curve Analysis.* • *Other components in the test group* • *Intercept:0.06* • *Slope:1.09* • *Brier:0.052*	*Revision TKA*	• *A Five-Fold Validation was used to prevent overfitting as well as 80-20 Validation and Recursive Feature Elementation.* • *Coarse Grained Grided Grine Search and Repeated Random Sub-Sampling were used to tune.*	• *The strongest predictor for the PJI Richarnt is the previous I&D with or without modular component replacement.* • *After that,> 4 previous surgery and microbiology (Enterococcus, MRSA) are also strong.* • *MRSA was identified as a risk factor.* • *Metastasis, AIDS, substance abuse, and obesity factors were confirmed, consistent with previous research.*

Christian Klemt et al. (2024)[16]	- *ANN* - *SVM* - *KNN*	• *Discrimination in the ANN model had Ro-Auc = 0.78.* • *The model was also approved by Calibration and Decision Curve Analysis. Based on the above, ANN was better than the other two models.* • *Other components in the test group:* • *Intercept:0.10* • *Slope:1.09* • *Brier:0.055*	*Aseptic Revision TKA*	• *Using Five-Fold Cross Validation and 80-20 Validation and Recursive Feature Elementation to prevent overfitting.* • *Using Coarse Grained Grid Search and Repeated Random Sub-Sampling to tune each Hyper Parameter*	• *The strongest predictor for PJI Richarnt was previous open surgery before regeneration surgery.* • *Supplies, obesity and diabetes are also strong predictors*
Noam Shohat (2020)[13]	• *RF* • *Logistic Regression with all variables* • *Stepwise (Forward) Logistic Regression* • *Stepwise (Forward) Logistic Regression including all integration with the variable*	*According to AUC = 0.74, the RF model was the most reliable.*	*THA* *TKA*	*Using 8-Fold Cross Validation for Validation.*	*The predictor of treatment failure for I & D after PJI is stronger than KLIC and CRIME80.*
Sultan Turhan (2023)[18]	• *linear discriminant analysis* • *decision tree* • *KNN* • *GBM* • *LR as Meta Model for Stacking*	• *The following are related to the Stacking model:* • *Accuracy = 0.89* • *f1 = 0.84* • *kappa = 0.81* • *precision= 0.88* • *Sensitivity = 0.80* • *Specificity = 0.98*	Hemiarthroplasty	• *Enhancing the generalization capability of the model by training the results of other models (first-level learner) as input to the LR model (meta-learner) in the Stacking method.* • *Utilizing F1 measure to examine accuracy and sensitivity.*	• *More accurate results in predicting discharge after hemiarthroplasty using model 2 (meta).* • *Volume of fluid collected from drain, morbid obesity, blood type, and high BMI were four significant risk factors for discharge based on the Stacking model.*

In a 2021 study, Kuo and colleagues employed a Stacking model with SVM as the meta-classifier to identify numerous significant features aligned with the ICM2018 (International Consensus Meeting) criteria for diagnosing PJI, while also incorporating additional important variables such as hemoglobin and prothrombin time, thereby establishing varied baseline values for these parameters.[29] In 2022, Tao and colleagues further refined diagnostic criteria for PJI by applying deep learning techniques to pathology slide images, as summarized in Table 2.[30]

In a groundbreaking study, Florez Balado et al. (2023), introduced a novel tool named AI-HPRO (Artificial Intelligence - Hip Prosthesis), which integrates natural language processing (NLP) techniques with the multivariate predictive algorithm XGBoost. This innovative AI-driven approach significantly enhances the monitoring of surgical site infections (SSI) in patients undergoing hip arthroplasty. By reducing the manual monitoring workload by over 88%, the time required for record reviews decreased dramatically from 975 hours to just 63.75 hours, thereby enabling infection prevention teams to oversee additional surgical procedures effectively. The automation of SSI monitoring is increasingly recognized for its numerous benefits, including reduced monitoring time, improved isolation practices, enhanced credibility for inter-center comparisons, and increased professional satisfaction. The rate of automated SSI diagnoses has risen remarkably from 32% of hospitals in 2009 to approximately 83% in 2021, with university-affiliated hospitals demonstrating even higher levels of automation. This shift allows staff to focus on higher-value tasks, such as developing infection control strategies based on surveillance data. The study indicates that the AI-HPRO model demonstrates nearly 100% Negative Predictive Value (NPV) and an impressive sensitivity of 99.7%, thereby outperforming previous models that relied solely on NLP or that combined it with logistic regression. This innovative integration facilitates instantaneous and effective SSI monitoring, bolstering proactive strategies and creating new avenues for disease outbreak prevention.[31]

3- Antibiotics and Pathogenic Microorganisms

- Staphylococcus aureus

Deuis et al.[32] developed a machine learning classification model employing Adaptive Boosting to predict methicillin resistance in Staphylococcus aureus. This study analyzed a total of 606 bacterial genomes sourced from the Pathosystems Resource Integration Center database. By utilizing k-mer counting in the DNA to represent antimicrobial resistance regions within the bacterial genome, the algorithm attained exceptional performance metrics, with an Area Under the Curve (AUC) of 0.991 and an accuracy rate of 99.5%. This model parallels the work conducted by Druen et al.,[33] who also achieved over 98.7% accuracy in predicting the resistance profiles of Staphylococcus aureus.[34]

Table 2 - Use of ML in evaluating factors related to diagnosis

Author/ Year of Publication	***Model (ML)***	***Interpretability***	***Type of Arthroplasty***	***Description***	***Results***
Guosong Wu et al. (2023)[19]	*XGBoost*	*• ROC AUC of 0.906 (95% CI 0.835-0.978)* *• PR AUC of 0.637 (95% CI 0.528-0.746)* *• F1 Score of 0.79 (0.67-0.90)* *The use of nurses' notes significantly increased F1 and PRAUC measures.*	*THA* *TKA*	*Grid Search was used for optimization.*	*• The model uses structured (e.g. ICD10 codes) and unstructured data (nurses' writings modeled with Bag of Words and TF-IDF techniques) to identify SSI.* *• Benefits: More accurate SSI identification, reduced workload especially for monitoring teams, time and cost savings, optimal use of unstructured data that is often not taken seriously.*
Alvaro Flores-Balado (2023)[26]	*AI-HPRO was designed using Natural Language Processing (NLP) and Extreme Gradient Boosting.*	*• Sensitivity (99.18%)* *• Specificity (91.01%)* *• F1-score of 0.32* *• AUC of 0.989* *• accuracy of 91.27%* *• Negative predictive value of 99.98%.*	*THA*		*• Reduction of more than 88% in manual monitoring workload and the time required to review records from 975 hours to only 63.75 hours using this AI-based method.* *• Rapid SSI monitoring, enhancement of preventive strategies, and providing new opportunities for disease outbreak prevention through this innovative combination.*
Eryo Kawakami et al. (2023)[20]	*RF was used to differentiate PJI from NonPJI, k-Mean, and UMAP were used to subclassify into high-risk and low-risk groups, given the overlapping characteristics in the groups. Of course, in this study, Markov chains were also used to examine the probability of remaining in the current group over time.*	*RO-AUC:0.778*	*THA*	*For comparison, the logistic regression model was also used, which had an AUC of 0.743.*	*• CRP was identified as the most important diagnostic factor for PJI.* *• Decreased BUN and Hb and increased total protein and ALT were the other most important factors for diagnosing PJI.* *• The combination of biomarkers is also important. For example, high CRP and low hemoglobin was a unique combination for diagnosing PJI.* *• People with PJI have a higher probability of transitioning to higher risk groups after surgery (PJI is a risk for reinfection).*
Pearl R Paranjap (2023)[21]	*• Unsupervised: K-means clustering & PCA.* *• Supervised: The decision tree algorithm*	*Area under the curve (AUC) of 0.993, with high sensitivity (99.1%) and specificity (97.1%).* *SHAP was used for AD and MID Sum markers.*	*TKA ,THA*	*Decision tree overfitting was corrected using Cost Complexity Pruning.*	*Using synovial fluid biomarkers and patient age, this model significantly reduced the rate of indeterminate PJI diagnoses from 7.4% to 1.3%.*

Domenico Albano et al. (2023)[22]	*SVM TRACE4 platorm*	*a-All MRI features:* *1) Internal group:* *Accuracy82%,AUC81%* *92% sensitivity* *62% specificity* *79% PPV,83% NPV* *2) External group:* *Accuracy88%,AUC89%* *92% sensitivity* *79% specificity* *89% PPV,83% NPV* *b-Bone marrow edema* *1) Internal group:* *Accuracy80%,AUC74%* *81% sensitivity* *76% specificity* *66% PPV,88% NPV* *2) External group:* *Accuracy75%,AUC79%* *68% sensitivity* *89% specificity* *93% PPV,60% NPV*	*THA*	*Nested five-fold cross validation*	*• The most specific MRI findings for diagnosing PJI according to machine learning statistics include synovitis, stratified synovitis, and extracapsular edema.* *• Bone edema is also the most sensitive with a high NPV.* *• Bone destruction and periarticular soft tissue mass and fibrotic membrane were not associated with infectious THA failure and are likely a sign of non-infectious treatment failure.*
Liangbing Nie (2023)[23]	*DBS-eNet, which is an effective neural network and part of deep learning.*	*•Diagnostic accuracies:* *Knee: 87.74%* *Hip: 86.36%* *• AUC:* *Knee: 0.957* *Hip: 0.906*	*THA* *TKA*	*Preprocessing hip data into DOI and knee data into standard* *• Identification of data with greater impact by Grad-CAM* *• Use of DCA to evaluate the clinical utility of the model*	*• It provided higher sensitivity and specificity for the diagnosis of PJI than the original classification models.* *• The diagnosis was better than that of nuclear medicine specialists, especially in peri-knee prosthesis infection.*
Kuo F.C. (2021)[24]	*Two-level stacked generalization architecture:* *•Meta-classifier: SVM* *•Base classifiers: RF, EGB, LR, naïve bayesian*	*• AUC: 0.988* *• Accuracy: 96.4%* *• Recall: 0.981* *• F1 score: 0.97 Matthews correlation coefficient: 0.926* *• Precision: 0.96*	*THA* *TKA*	*5-fold cross-validation was performed.*	*• The model's performance in diagnosing chronic PJI was compared with the 2018 European Society of Osteoarthritis criteria.* *• Identification of many common features listed in the 2018 ICM criteria for diagnosing PJI.* *• Consideration of other important features such as hemoglobin and prothrombin time and setting different baseline values for these features.*
Tao Y. (2022)[25]	*ResNet34 deep learning convolutional network*	*• AUC: 0.8136* *• Average accuracy: 93.3%* *• Average recall rate: 0.9739* *• F1 score: 0.9482*	*Revision* *TKA* *THA*	*• Converting frozen sections of 20 patients using UN scanner and McUddy scanner into electronic images.* *• Artificially segmenting the entire image under high magnification (400×) and viewing using IViewer software* *• Saving segmented electronic images in TIFF format.* *• Studying images and pathological diagnosis based on ICM 2018 diagnostic criteria by two pathologists.* *• If the number of white blood cells at high magnification is more than 10, it is considered positive, otherwise it is considered negative.* *• Using Resnet deep learning convolutional network model for training.* *• Reading and recognizing positive and negative images according to ICM 2018 diagnostic indicators by a self-made intelligent image reader system using two sets.*	*• The aim was to identify high-magnification images of pathological soft tissue around joints against diagnostic criteria for acute infection.* *• A convolutional neural network was used for image recognition and analysis. This technique shows that better diagnostic accuracy can be achieved by comparing the new method with standard strategies.* *• This study demonstrates the high potential of deep learning in improving the diagnosis of acute infections through pathological image analysis, although continuous improvement of the training data is necessary to achieve optimal accuracy.*

a- Enterococcus faecium

In research conducted by Druen et al.,[33] a set-cover model was implemented to predict antibiotic susceptibility across twelve bacterial pathogens, including Staphylococcus aureus, Enterococcus faecium, Escherichia coli, Klebsiella pneumoniae, and Pseudomonas aeruginosa. This study encompassed a total of 56 different antibiotics and utilized k-mer counting of DNA sourced from the Pathosystems Resource Integration Center database. The resulting model achieved an impressive accuracy rate of 100% for Enterococcus faecium. The model not only exhibited high interpretability but also benefited from the application of sample compression theory, facilitating a brief computational time and thereby making it accessible for individuals without expertise in machine learning (ML).[34]

b- Escherichia coli

In a study conducted by Moradi-Garavand et al.,[35] a gradient boosted decision tree model was employed to forecast the resistance of Escherichia coli to eleven commonly utilized antibiotics. The model was developed using a comprehensive dataset that encompassed genetic content, year of isolation, demographic structures, and polymorphism information derived from 1,936 complete genome sequences of Escherichia coli. The performance of this machine learning model was evaluated against a rule-based methodology developed within the same research framework, yielding an impressive average accuracy of 91%. This result significantly surpassed both the rule-based method and earlier endeavors by Druen and colleagues, whose model achieved an accuracy of 81.8% for Escherichia coli.[34]

c- Klebsiella pneumoniae

Druen et al.[33] constructed a model utilizing binary phenotypic classification (sensitive or resistant), achieving a remarkable accuracy of 95% for Klebsiella pneumoniae. In another noteworthy study by Nguyen et al.,[36] the Minimum Inhibitory Concentration (MIC) for Klebsiella pneumoniae was predicted employing a gradient boosted tree model to assess antibiotic resistance levels. This model was trained on data extracted from 1,668 whole genome sequences, resulting in an overall accuracy of 92%.[34]

d- Pseudomonas aeruginosa

Using genomic sequences and transcriptomic data from 414 samples of Pseudomonas aeruginosa, Khaledi et al.,[37] developed a support vector machine (SVM) classifier to predict the bacterium's sensitivity to four commonly prescribed anti-Pseudomonas antibiotics, specifically ceftazidime, meropenem, ciprofloxacin, and tobramycin. This model demonstrated high sensitivity and predictive values ranging from 0.8 to over 0.9, comparable to the model developed by Druen et al.,[33] which achieved an accuracy of 93.9% for Pseudomonas aeruginosa. Despite these advancements in predictive modeling, limited efforts have been made to create user-friendly software accessible to the public. In response to this gap, Aoun and colleagues[38] introduced PhenotypeSeeker, a straightforward software application that employs MIC values and binary phenotypes to assess the resistance of Pseudomonas aeruginosa to ciprofloxacin through two regression models. This approach achieved an accuracy of 88%, utilizing k-mers from 200 genomes to inform the model. The model can be constructed using assembled genomes in under five hours for each phenotype, delivering phenotype predictions in merely one second.[34] A summary of the findings related to bacterial species and antibiotic sensitivity is provided in Table 3.

Responses to PJI Frequently Asked Questions by ChatGPT

In 2024, Zhong and colleagues[39] evaluated the performance of ChatGPT in addressing frequently asked questions regarding Periprosthetic Joint Infection (PJI) following Total Knee Arthroplasty (TKA) and Total Hip Arthroplasty (THA). They identified twelve categories of frequently asked questions sourced from the International Clinical Skills Center's website. The accuracy of ChatGPT's responses was assessed by a multidisciplinary team. Overall, while a small fraction of the responses required minor clarifications, the answers were predominantly unbiased and evidence-based, even in instances where discrepancies were noted. Thus, the potential utilization of ChatGPT-4 as an educational tool for patients presents significant benefits.

Table 3 - Using ML to assess antibiotic susceptibility

Author/Year of Publication	*ML Model*	*Interpretability*	*Type of Use*
Davis J.J. (2016)[27]	*Adaptive boosting*	*Area under the receiver operating characteristic curve: 0.991* *Accuracy: 99.5%, F1 score: 0.995*	*Using the ML-based pAST method and NGS-pAST to assess the susceptibility of Staphylococcus aureus to the antibiotic methicillin*
Drouin A. (2019)[28]	*Set Covering Machine*	*Accuracy:* • *Staphylococcus aureus: 98.7%* • *Enterococcus faecium: 100.0%* • *Escherichia coli: 81.8%* • *Klebsiella pneumoniae: 95.0%* • *Pseudomonas aeruginosa: 93.9%*	*Using the pAST method based on ML and NGS - pAST to assess the susceptibility of the following bacteria to the antibiotics listed against them:* • *Staphylococcus aureus: methicillin* • *Enterococcus faecium: vancomycin* • *Escherichia coli: co-amoxiclav* • *Klebsiella pneumoniae: gentamicin* • *Pseudomonas aeruginosa: levofloxacin*
Moradi garavand D. (2018)[30]	*Gradient-boosting decision tree*	*Average accuracy: 91.0%*	*Using the ML-based pAST and NGS-pAST method to assess the susceptibility of Escherichia coli to 11 antibiotics*
Nguyen, M. (2018)[31]	*Gradient-boosting tree*	*Average accuracy: 92.0%*	*Using the ML-based pAST and NGS-pAST method to evaluate MICs for Klebsiella pneumoniae to determine the degree of resistance to 20 antibiotics*
Khaledi, A. (2020)[32]	*Support vector machine*	• *Using single-nucleotide polymorphism's information or gene expression data alone: Sensitivity and Predictive value:0.8–0.9* • *In combination: Sensitivity and Predictive value:>0.9*	*Using the ML-based pAST and NGS-pAST method to assess the susceptibility of Pseudomonas aeruginosa to the following four antibiotics:* • *Ceftazidime* • *Meropenem* • *Ciprofloxacin* • *Tobramycin*
Aun, E. (2018)[33]	• *Logistic regression for binary phenotypes (susceptible or resistant)* • *Linear regression for continuous phenotypes (MIC value)*	• *Logistic regression: Accuracy:88.0%, F1-measure:0.88 Sensitivity: 0.90, Specificity: 0.87* • *Linear regression: Coefficient of determination* (R^2)*:0.42* *Pearson correlation coefficient:0.68* *Spearman correlation coefficient:0.84*	*Using the ML-based pAST and NGS-pAST method to assess the susceptibility of Pseudomonas aeruginosa to ciprofloxacin*

Limitations

Common limitations identified in machine learning studies include their retrospective design and the relatively small size of patient populations involved.

Discussion and Conclusion

The integration of machine learning and deep learning methodologies into the assessment of Periprosthetic Joint Infection (PJI) risk, the enhancement of diagnostic and prognostic standardization, and the evaluation of bacterial sensitivity to antibiotics marks the advent of a new era in the management of this arthroplasty-related complication. The advantages associated with the use of machine learning in these inquiries encompass improved accuracy and analytical speed; machine learning models possess the capability to simultaneously process multiple features, thereby allowing for the analysis of complex and nonlinear relationships that are challenging for traditional methodologies. Moving forward, it is anticipated that addressing the limitations currently observed in studies will lead to enhanced accuracy in research findings, alongside the development of machine learning-based software that can facilitate more precise and cost-effective decision-making in medical practice. Improved management of complications arising from arthroplasty will ultimately contribute to enhanced treatment processes and improved patient outcomes.

References:

1. Rothbauer F, Zerwes U, Bleß H-H, Kip M. Prevalence of hip and knee arthroplasty. White Paper on Joint Replacement: Status of Hip and Knee Arthroplasty Care in Germany; 15-39; 2018.
2. Gademan MG, Hofstede SN, Vliet Vlieland TP, Nelissen RG, Marang-Van de Mheen PJ. Indication criteria for total hip or knee arthroplasty in osteoarthritis: a state-of-the-science overview. BMC musculoskeletal disorders; 11-17:1; 2016.
3. Cushner F, Agnelli G, FitzGerald G, Warwick D. Complications and functional outcomes after total hip arthroplasty and total knee arthroplasty: results from the Global Orthopaedic Registry (GLORY). 2010.
4. Dobson PF, Reed MR. Prevention of infection in primary THA and TKA. EFORT open reviews; 5-13(10): 604; 2020.
5. Rodriguez-Merchan EC, Delgado-Martinez AD. Risk factors for periprosthetic joint infection after primary total knee arthroplasty. Journal of Clinical Medicine. 2022; 11(20): 6128.
6. Parvizi J, Tan TL, Goswami K, Higuera C, Della Valle C, Chen AF, et al. The 2018 definition of periprosthetic hip and knee infection: an evidence-based and validated criteria. The Journal of arthroplasty; 33(5): 1309.e2; 2018.
7. Sherrell JC, Fehring TK, Odum S, Hansen E, Zmistowski B, Dennos A, et al. The Chitranjan Ranawat Award: fate of two-stage reimplantation after failed irrigation and débridement for periprosthetic knee infection. Clinical Orthopaedics and Related Research®. 2011; 25-469: 18.
8. Gardner J, Gioe TJ, Tatman P. Can this prosthesis be saved?: implant salvage attempts in infected primary TKA. Clinical Orthopaedics and Related Research®. 2011; 6-469: 970.
9. Sousa R, Abreu MA. Treatment of prosthetic joint infection with debridement, antibiotics and irrigation with implant retention-a narrative review. Journal of bone and joint infection; 17-108: (3)3; 2018.
10. Davis JS, Metcalf S, Clark B, Robinson JO, Huggan P, Luey C, et al., editors. Predictors of treatment success after periprosthetic joint infection: 24-month follow up from a multicenter prospective observational cohort study of 653 patients. Open Forum Infectious Diseases; 2022: Oxford University Press US.
11. R. Z, R. AAA, Z. AA, A. OM, R. MSM. Investigating the Role of Artificial Intelligence Technology Development in Medicine and Health Promotion. Iranian Journal of Surgery, 1-23: (2)32; 2024.
12. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. Translational vision science & technology. 2020; 9-14(2).
13. Tiwari T, Tiwari T, Tiwari S. How Artificial Intelligence, Machine Learning and Deep.
14. Sasikala S, Subhashini S, Alli P, Angelina JJR. Deep Learning Applications in Medical Imaging: Artificial Intelligence, Machine Learning, and Deep Learning. DeepLearning Applications in Medical Imaging: IGI Global; p.178-208; 2021.
15. Alaskar H, Saba T. Machine learning and deep learning: a comparative review. Proceedings of Integrated Intelligence Enable Networks and Computing: IIENC: 50-143; 2020-2021.
16. Cui Z .Machine learning and small data. Educational Measurement: Issues and Practice. 8-12:40(4); 2021.
17. Tripathi AN, Sharma B. A Depth of Deep Learning for Big Data and its Applications. International Journal of Engineering Research & Technology (IJERT) ISSN. 78-81; 2020-2021.
18. Shohat N, Goswami K, Tan TL, Yayac M, Soriano A, Sousa R, et al. 2020 Frank Stinchfield Award: identifying who will fail following irrigation and debridement for prosthetic joint infection: a machine learning-based validated tool. The Bone & Joint Journal. 2020; 102 (7 Supple B): 9-11.
19. Shah AA, Devana SK, Lee C, Kianian R, van der Schaar M, SooHoo NF. Development of a novel, potentially universal machine learning algorithm for prediction of complications after total hip arthroplasty. The Journal of arthroplasty. 1655e1(5): 36-62; 2021.
20. Klemt C, Laurencin S, Uzosike AC, Burns JC, Costales TG, Yeo I, et al. Machine learning models accurately predict recurrent infection following revision total knee arthroplasty for periprosthetic joint infection. Knee Surgery, Sports Traumatology, Arthroscopy; 25-30(8): 82-90; 2022.
21. Klemt C, Yeo I, Harvey M, Burns JC, Melnic C, Uzosike AC, et al. The use of artificial intelligence for the prediction of periprosthetic joint infection following aseptic revision total knee arthroplasty. The Journal of Knee Surgery. 37(02): 66-158; 2024.
22. Yeo I, Klemt C, Robinson MG, Esposito JG, Uzosike AC, Kwon Y-M. The use of artificial neural networks for the prediction of surgical site infection following TKA. The Journal of Knee Surgery. 36(06):43-637; 2023.
23. Turhan S, Canbek U, Dubektas-Canbek T, Dogu E. Predicting Prolonged Wound Drainage after Hemiarthroplasty for Hip Fractures: A Stacked Machine Learning Study. Clinics in Orthopedic Surgery. 15(6): 894; 2023.

24. Wu G, Cheligeer C, Southern DA, Martin EA, Xu Y, Leal J ,et al. Development of machine learning models for the detection of surgical site infections following total hip and knee arthroplasty: a multicenter cohort study. Antimicrobial Resistance & Infection Control. 12(1): 88; 2023.
25. Kawakami E, Kobayashi N, Ichihara Y, Ishikawa T, Choe H, Tomoyama A, et al. Monitoring of blood biochemical markers for periprosthetic joint infection using ensemble machine learning and UMAP embedding. Archives of Orthopaedic and Trauma Surgery. 6057(10): 67-143; 2023.
26. Paranjape PR, Thai-Paquette V, Miamidian JL, Parr J, Kazin EA, McLaren A, et al. Achieving high accuracy in predicting the probability of periprosthetic joint infection from synovial fluid in patients undergoing hip or knee arthroplasty: the development and validation of a multivariable machine learning algorithm. Cureus. 15(12); 2023.
27. Albano D, Gitto S, Messina C, Serpi F, Salvatore C, Castiglioni I, et al. MRI-based artificial intelligence to predict infection following total hip arthroplasty failure. La radiologia medica. 128(3): 6-340; 2023.
28. Nie L, Sun Z, Shan F, Li C, Ding X, Shen C. An artificial intelligence framework for the diagnosis of prosthetic joint infection based on 99 mTc-MDP dynamic bone scintigraphy. European Radiology. 33(10): 803-6794; 2023.
29. Kuo F-C, Hu W-H, Hu Y-J. Periprosthetic joint infection prediction via machine learning: comprehensible personalized decision support for diagnosis. The Journal of Arthroplasty. 37(1): 41-132; 2022.
30. Tao Y, Hu H, Li J, Li M, Zheng Q, Zhang G, et al .A preliminary study on the application of deep learning methods based on convolutional network to the pathological diagnosis of PJI. Arthroplasty. 4(1): 49; 2022.
31. Flores-Balado Á, Méndez CC, González AH, Gutierrez RM, de las Casas Cámara G, Cordero BV, et al. Using artificial intelligence to reduce orthopedic surgical site infection surveillance. Workload: Algorithm design, validation, and implementation in 4 Spanish hospitals. American Journal of Infection Control. 51(11): 9-1225; 2023.
32. Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, et al. Antimicrobial resistance prediction in PATRIC and RAST. Scientific reports. 6(1): 27930; 2016.
33. Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. Scientific reports. 9(1): 4071; 2019.
34. Chong YY, Chan PK, Chan VWK, Cheung A, Luk MH, Cheung MH, et al. Application of machine learning in the prevention of periprosthetic joint infection following total knee arthroplasty: a systematic review. Arthroplasty. 5(1): 38; 2023.
35. Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLoS computational biology. 14(12): e1006258; 2018.
36. Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, et al. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Scientific reports. 8(1): 421; 2018.
37. Khaledi A, Weimann A, Schniederjans M, Asgari E, Kuo TH, Oliver A, et al. Predicting antimicrobial resistance in Pseudomonas aeruginosa with machine learning-enabled molecular diagnostics. EMBO molecular medicine. 12(3):e10264; 2020.
38. Aun E, Brauer A, Kisand V, Tenson T, Remm M. A k-mer-based method for the identification of phenotype-associated genomic biomarkers and predicting phenotypes of sequenced bacteria. PLoS computational biology. 14(10): e1006434; 2018.
39. Hu X, Niemann M, Kienzle A, Braun K, Back DA, Gwinner C, et al. Evaluating ChatGPT responses to frequently asked patient questions regarding periprosthetic joint infection after total hip and knee arthroplasty. Digital Health. 10: 20552076241272620; 2024.