Artificial Intelligence, Deep Learning, and Machine Learning Applications in Total Hip Arthroplasty

Abstract Artificial intelligence (AI) recently gained popularity in total hip arthroplasty (THA) applications due to several reasons including technological improvements such as availability of data storage, processor capabilities, AI technique developments, and surgery-related improvements including presurgical analysis techniques developed and data collected for input to algorithms (Mont, et al. J Arthroplast. 34(10):2199–200, 2019). In this work the focus will be on the research literature covering AI, deep learning (DL), and machine learning (ML) techniques that relate to only THA. This coverage excludes the combined results for total knee arthroplasty (TKA) and THA unless THA is analyzed independently from TKA. Applications determined include THA-related economic analysis and payment models, patients' well-being, risk of blood transfusion, hip fracture detection (Kim and MacKinnon. Clin Radiol. 73:439–45, 2018). Biomechanical considerations, optimal implant design, post-THA implant brand detection, hip disability upon THA, inpatient and outpatient THA surgery detection, automating and improving angle of acetabular component, text-based database search for THA-related factors, mechanical loosening detection of the transplant, patient comfort after THA, and implant failure detection. Many more applications are possible using AI, DL, and ML with few of them suggested in the conclusion section.

1 Introduction

Development of algorithms allowing to make informed decisions based on patterns learned from data and mimicking human behavior by using technology has been one of the goals of researchers for real-life applications. Impact of AI, DL, and ML applications recently (within the last 5 years) started to gain popularity even though research on deep learning applications on THA can be seen as early as 1997 [1]. One of the key aspects of THA is to be one of the most successful orthopedic procedures developed in the twentieth century, a feature that can allow to make informed judgments by using algorithms for classification and prediction noting the ability to clearly distinguish many aspects of the operational procedures. For instance, there

E. Tokgöz, *Total Hip Arthroplasty*, https://doi.org/10.1007/978-3-031-08927-5_11

are certain aspects that are clinically observed such as cement versus cementless THA procedures and reasons for implant failures that are known to train algorithms by using the corresponding data sets that can help with accurate results for testing algorithms on testing sets. Large number of features (i.e., input variables) and slowness of manual processes encourage researchers to investigate the use of AI/DL/ML algorithms to determine models that allow predictions by incorporating all the features simultaneously. Some of the challenges that we can list here with these algorithms in applications can include the small size of the sample set used for modeling that would not necessarily allow generalization (depending on conditions) and a team of interdisciplinary researchers with a broad knowledge of concepts. One other challenge is applicability of the developed models on data sets. Throughout this article, we will cover the research literature on AI, DL, and ML techniques that relate to only THA. This coverage excludes the combined results for total knee arthroplasty (TKA) and THA unless THA is analyzed independently from TKA. Applications determined include THA-related economic analysis and payment models, patients' well-being, major complication analysis, sensor-based gait analysis of THA patients, risk of blood transfusion, hip fracture detection, biomechanical considerations, optimal implant design, post-THA implant brand detection, hip disability upon THA, inpatient and outpatient THA detection, automating and improving angle of acetabular component, text-based database search for THA-related factors, mechanical loosening detection of the transplant, patient comfort after THA, and implant failure detection. Even though the following three sections are categorized into AI, DL, and ML, some of the articles have mixes of these methods. The last section is devoted to discussion and potential future research directions by using AI, DL, and ML.

2 Machine Learning

A machine learning algorithm is designed in [2] to propose a risk-adjusted patient-specific payment model (PSPM) that considers patient comorbidity used on preoperative big data to predict length of stay (LOS) and patient-specific inpatient payments after primary THA. The eight variables used are age group, ethnicity, gender, Charlson Comorbidity Index (based on comorbidities such as congestive heart failure renal disease and cancer documented from the 12 months before the hospitalization to 3 days after discharge), discharge disposition, type of admission, all patient refined (APR) risk of mortality, and APR severity of illness (minor, moderate, major, and extreme comorbidities). Data collected from 122,334 patients between 2012 and 2016 undergoing primary THA for osteoarthritis is used to train a naïve Bayesian model. Performance of the machine learning model is evaluated by using percentage of accuracy and area under the curve calculations. Age, race, gender, and comorbidity scores are determined to be the most important characteristics for the generated model to demonstrate validity, reliability, and responsiveness for receiver operating characteristic curve values of 87% for LOS and 71% for LOS payment. The patient complexity and error for predicting payment are determined to be correlated with 3% for moderate, 12% for major, and 32% for extreme comorbidities. The ML algorithm is determined to be good for predicting LOS and payment prior to primary THA.

Noting the financial challenges faced by the patients, authors of [3] developed logistic regression, artificial neural networks and random forest model. The database use consisted of 63,859 recorded patients from 2017 to 2018. No overnight stay in the hospital is compared to 1–3 days of stay for the models developed. Among the 40 candidate variables chosen for modeling, top 10 important features/variables included ethnicity, anesthesia type, race, BMI, age, blood urea nitrogen, year, albumin, sodium, and white blood cell count for the developed models by using artificial neural network (ANN), random forest, and multivariable regression in predicting same-day discharge patients after primary THA. Area under the curve and accuracy values are 71.5% and 65% for logistic regression, 76.2% and 73% for ANN, and 80.4% and 81% for random forest. Therefore, ANN and random forest are determined to be the outstanding classifiers for utilization in the future. These models demonstrated reliability for their future use in ambulance utilization and patient discharge. We refer to [33] for an elastic-net penalized logistic regression model developed for prediction of prolonged postoperative opioid prescriptions of THA patients.

Clinically significant outcome (CSO) for the patient-reported health state (PRHS) is modeled in [4] by using stochastic gradient boosting, random forest, support vector machine, ANN, and logistic regression. Variables used included preoperative PRHS, BMI, age, drug allergies, preoperative opioid use, smoking history, prior ipsilateral hip surgery excluding a THA, and diabetes. Data collected between 2014 and 2017 on a total of 407 patients are analyzed based on discrimination, calibration, Brier score, and decision curve analysis. Stratified splitting of 80–20 on training-testing is conducted on the data. The minimal clinically important difference (MCID) is calculated for the PRHS by using a distribution. Feature selection with random forest algorithms was used recursively to determine the subset of variables to be employed for final algorithm development. Discrimination, calibration, Brier score, and decision curve analysis indicated the random forest algorithm to perform better on predicting patient's achievement of clinically meaningful improvements for the PRHS. It is also observed that preoperative PRHS score, BMI, age, and preoperative opioid use are the most important features. Clinically meaningful improvement for the PRHS after THA is determined for 69.2% of patients.

Machine learning methods are utilized in [5] for modeling major complications of patients after THA. Approximately 90,000 THA patients of a California hospital are included in the data set with 545 patients that had major complications. Variables included in the analysis included age, gender, race, ethnicity, insurance, and medical comorbidities that are used as the variables of the developed models. AutoPrognosis, logistic regression, random forest, gradient boosting, XGBoost, and AdaBoost are compared for their accuracies. AutoPrognosis model demonstrated higher accuracy (73.2%) when compared to logistic regression that had 64.4% and other machine learning algorithms. The outcomes of the modeling resulted in classification attributes to differ for AutoPrognosis and logistic regression: Five features that appeared to be the most important in risk prediction for using AutoPrognosis are chronic obstructive pulmonary disease (COPD), dementia, malnutrition, malignancy, and Medicare coverage, while logistic regression indicated the importance of variables such as chronic atherosclerosis, renal failure, and chronic obstructive pulmonary disease. The success and discriminative ability of AutoPrognosis is due to analyzing complex nonlinear relationships and be able to capture variables that logistic regression and other machine learning algorithms could not capture. It is concluded by the authors that providing more accurate prognostic information by using AutoPrognosis can help facilitating well-versed preoperative shared decision-making.

Falling impacts the THA patients' well-being and increases the chance of post-surgical procedures due to issues that may arise. Wearable sensors can be integrated into fall risk assessment tools to collect data on patients' functional ability. Support vector machine (SVM) and linear discriminant analysis classifier are developed and tested in [6] to predict the risk of THA patients' falling by using the sensor-collected data. Research data is collected at three different stages: preoperatively, 2-week THA follow-up, and 6-week THA follow-up. Feature variables consisted of preoperational and operative trajectory data. Preoperation set consisted of sensor-derived metrics collected preoperatively, while operative trajectory set combined sensor-derived metrics from preoperative and 2-week postoperative appointments. A total of 96 patients initiated the research, and this number is reduced to 72 at the end of the data collection period. SVM demonstrated success based on the measured 87% accuracy, 97% sensitivity, 46% specificity, and 82% area under the curve (AUC) for the preoperative appointment. Upon adding 2-week postoperative data to the preoperative data, an overall improved performance of 90% accuracy, 93% sensitivity, 59% specificity, and 88% AUC is achieved by using the linear discriminant analysis classifier. The importance of the high accuracy of the fall risk prediction models is emphasized for THA patients.

Logistic regression is compared to six machine learning algorithms in [7] for predicting the risk of blood transfusion in both THA and TKA by using long short-term memory networks (LSTM), RF, decision tree (DT), k-nearest neighbors (KNN), SVM, and naïve Bayes classifier. Here we report only the results attained for THA; the postoperative transfusion rate of 22.79% for THA of the 12,642 patients is observed. The variables considered included age, sex, BMI, hemoglobin, type 2 diabetes, operation time, tranexamic acid use, interoperative blood loss, and hypertension. A tenfold cross-validation strategy is used to quantify the predictive ability of each model defined as the AUC of the receiver operating characteristic. Both LSTM and RF models had significantly better accuracies than LR, Naïve Bayes, KNN, SVM, and DT. Hypertension is determined to be a risk factor for transfusion.

24 statistical models are designed in [8] for prediction of hip fractures over time in 4722 women and 717 men with 5 years of follow-up. AUC values of 92% by using the bootstrap aggregated flexible discriminant analysis and 89% by using Extreme Gradient Boosting (GB) are determined to be the best "female model" and

best “male model,” respectively. Identifying features of the model included bone mineral density, glucose measurements, and osteoarthritis diagnosis. ML demonstrated improvement on hip fracture prediction beyond logistic regression.

Length of stay and cost of THA patients’ predictive modeling are conducted in [9] by using naïve Bayes machine learning algorithm. Feature selection included age, sex, ethnicity, race, type of admission, risk of mortality, and severity of illness. Accuracies of 76.5% for length of stay and 79% for cost are attained with performances of 88% and 89% for length of stay and cost, respectively. Model error and risk of mortality are determined to be positively correlated indicating validity of increase in risk-adjusted payment for each risk of mortality. Due to the cost of delivery of hip fracture care depending on non-modifiable patient-specific factors, the bundled care is concluded to be an inconvenient payment model for hip fractures in [9].

Biomechanical and bone quality data attained from CT, electromyography, and gait analysis are used in [10] for making a THA surgical decision prosthesis adaptation to the bone by using the BMD of the proximal and the distal region of the femur and cementation. Feature selection for RF included base of support, BMD of the proximal region of femur, and start and stop of the electromyographic signals. Feature selection for GB included base of support, toe in/out operated, velocity, healthy leg BMD, and start and stop of the electromyographic signals. Random forests (RF) and gradient boosted tree are performed as classifiers on 51 patients’ data based on the splitting of the data into 75% training and 25% testing sets. RF method had the best results utilizing the training set, while GB on the test set demonstrated good results including 92.9% accuracy, 100% specificity, and 85.7% value of under the curve of receiver operator characteristic. Features playing key roles in the choice of cemented or uncemented prosthesis selection are determined to be the skeletal muscle parameters such as the start and stop of muscle contraction from EMG signals and temporal and spatial gait parameters. The usefulness of the regression analysis for predicting the BMD of the distal and proximal parts of the operated femur after 1 year from the surgery is also demonstrated to be useful by the authors as a part of the patient follow-up.

Optimal implant design parameter characteristics are structured in [11] by integrating biomechanical analysis into machine learning techniques. 3D finite element analysis is integrated into ANN and SVM with the selected implant geometric features including stem length, lateral thickness, medial thickness, and the distance between the implant neck and the central stem surface. The output is designed to be the strain reduced by the presence of the hip implant. A pattern-search minimization algorithm is used to identify the optimal geometry of the implant by exploring new values of the input parameters in an iterative fashion. The optimization algorithm explored unseen values of the selected parameters of the hip implant geometry to minimize the function. Four geometrical ranges are explored for the dimensions of the bone by considering a clinically admissible shape. ANN and SVM techniques had similar pattern to the pattern-search minimization algorithm; optimizing parameters of the SVM had better prediction of the lower random errors; therefore, it had better results than ANN. An optimized implant that had reduced stress shielding is observed to need a decreased stem length and a reduced implant surface contact with the bone. In the case of thinner stems, the two radiuses associated with the stem width at the distal cross section in contact with the bone played a role for better stress shielding results.

3 Deep Learning

A deep learning application by using ANN on a network that learns and predicts LOS, inpatient charges, and discharge disposition by using 78,335 primary THA is implemented in [12]. The 15 preoperative attributes included age, gender, ethnicity, race, type of admission, location of admission (emergency department or not), patient code, risk of mortality (minor, moderate, major, severe), patient's severity of illness, number of associated chronic conditions and diagnoses, comorbidity status, weekend or weekday admission, hospital type, patient's income quartile, and internal or external (i.e., transfer) patient. All patient refined risk (i.e., minor, moderate, major, severe) is a composite disease-specific (i.e., minor 25% uncomplicated diabetes, moderate 25% diabetes with kidney disease, major 25% prior ketoacidosis, extreme 25% prior diabetic coma) measure accounting for the number and severity of underlying comorbidities. These attributes are used for generation of four hidden layers with 112, 56, 28, and 14 nodes from the input to the final layer that are heuristically chosen. Glorot normalization algorithm is used for initialization of each hidden layer node, and rectified linear activation function is applied by using a kernel constraint. Softmax activation function is used for the output layer consisting of the number of classes to determine the probabilities. Metrics used for validity included accuracy and area under the receiver operating characteristic curve. ANN learning in the first 30 training rounds resulted area under the curve values of 82% for LOS, 83.4% for charges, and 79.4% for disposition. Patient-specific payment model introduced established a risk increase of 2.5% for moderate, 8.9% for major, and 17.3% for severe comorbidities. These results are found to be reliable and valid for using the tier-based patient-specific payment model for future purposes.

A hip implant recognition algorithm is designed in [13] to detect implantation on 170 postoperative hip anteroposterior x-rays collected from 5 hospitals that incorporated 29 implant brands. Images are manually labeled, and they are successfully trained for the stem detection model. A six-layered convolutional neural network (CNN) in Keras deep learning platform is developed. 224 × 224 grayscale image inputs are used that had two layers of convolution and one max pooling layer to generate a feature map that is fed into two fully connected layers that generated 29 class outputs. Validation on 25% of training set is conducted based on the recognition model that had detection and clustering. 99% area under the curve value is attained from the receiver operating characteristic curve generated from a test set containing 25% of all stem-cropped images. The generated CNN showed usefulness in predicting stem detection in THA applications.

Classification of the quality (e.g., the staying length in hospital) after THA procedure in Taiwan is modeled in [14]. The proposed approach incorporated expert knowledge, global discretization, imbalanced bootstrap technique, reduct and core methods, rough sets, rule induction, and rule filter. Logistic regression, SVM, and multilayer perceptron (MLP) are utilized for modeling. The second version of Learning from examples module (LEM2) algorithm is applied for symbolic attributes in their work. The LEM2 algorithm calculates a single local covering for each concept from a decision table to generate decision rules. Calculation of each rule's quality index is based on a specific rule quality function that depends on the measure of support, consistency, and coverage to determine the strength of the rules [35]. Another application used in [14] is rough set theory (RST) approach that is introduced for AI applications. RST is a soft computing technique first proposed by Pawlak [15] that uses mathematical modeling to address class data classification problems and identified to be a very useful tool for decision support systems, especially in cases in which hybrid data, vague concepts, and uncertain data are involved in the decision process [16]. In conclusion, RST is found to be the best model among all considerations as a feasible choice for classification learning of imbalanced class data and combination of core attributes. Comparison of accuracy of different methods for both options of all 17 attributes and 7 core attributes in the THA data set had strong outcomes with a minimum of 85% accuracy calculation.

Prediction of the dependent variable hip disability and osteoarthritis outcome score (HOOS) is the primary outcome of [17] by utilizing THA results. A total of 160 patients with 44% female population is included in the study. The authors used the least absolute shrinkage selection operator (LASSO) [18] as the machine learning algorithm for predictive analysis. LASSO can reduce overfitting through penalization of the regression coefficients by sometimes reducing to zero resulting in excluding a predictor entirely so that the out-of-sample prediction accuracy is maximized. The main objective of LASSO is to minimize the mean squared error by reducing the coefficients. Post-surgery and 3-month follow-up data for analysis of HOOS is collected. In total a 23-item rating scale is designed with 25 coefficients utilized in the model. Independent variables included the following:

- Clinical and demographic variables such as such as age, gender, race, Hispanic ethnicity, marital status, level of education categorized into less than a college degree, college degree, or advanced degree, employment status, number of hours worked per week, planned legal action, and worker's compensation status.
- Patient-reported health and health habits, smoking status (smoked vs. never smoked), BMI, and exercise of number of days per week of mild, moderate, and strenuous.
- Cognitive appraisal processes using Brief Appraisal Inventory© [19].
- Surgical approaches including direct lateral, anterolateral, and direct anterior methods.

LASSO is determined to be a weak predictor that failed to include several important variables that are often considered important in predictive modeling in surgical outcomes such as smoking, age, level of education, and frequency of exercise. Diagnostic plots revealed at most moderate difficulties with the final model that utilized the 2-month postsurgical collected data. The most predictive independent variables of postoperative HOOS are determined to be cognitive appraisal processes. Variables predicting a worse HOOS are anterior surgical approach, increased BMI, thoughts of work, frequent comparison to healthier peers, and increased medical comorbidities. Variables that predicted a better HOOS consisted of thoughts related to family interaction, trying not to complain, employment at the time of surgery, and helping others. In conclusion, authors pointed out the need of an accurate predictive model need due to limited ability to identify patients at risk of having a mismatch in outcome following THA based on the models generated.

THA patient designation using machine learning for inpatient and outpatient classification is implemented in [20]. Of the 1409 medicare patients included in the study by using the data between 2017 and 2019, 77.4% of the patients experienced THA. 80% of the data is used for training and 20% for testing. Extreme Gradient Boosting (XGBoost) is a machine learning tool building predictive models utilizing gradient boosting framework. Inpatient/outpatient are predicted target variables used for the XGBoost method as the training data. Input variables used in the model included the following:

- Patient demographics such as age, gender, and BMI.
- Diagnosis leading to joint pain such as rheumatoid arthritis, osteoarthritis, and avascular necrosis.
- Past medical history such as cardiac history, history of a venous thromboembolic event [VTE], diabetes mellitus [DM], and other rheumatologic disease.
- Charlson Comorbidity Index (CCI).
- American Society of Anesthesiologists' Physical Status Classification (ASA).
- Revised Cardiac Risk Index.
- Modified Frailty Index (mFI).
- Preoperative functional scores.
- Hip disability and osteoarthritis outcome score (HOOSJR).
- VR12 physical component.
- VR12 mental component (mcs) scores.

The XGBoost model demonstrated 78.7% accuracy for predicting an inpatient or outpatient stay with 81.5% that is observed to be the area under the receiver operating characteristic curve. The most influential features in the predictive model included BMI, age, functional scores, and ASA Physical Status Classification.

Angular position of the acetabular component is observed to be a risk factor in implant dislocation following THA. A deep learning approach is undertaken in [21] to automate the angle measurement with the goal of increasing accuracy in measurements, reducing human error, and speeding up the measurement process. The data consisted of 600 anteroposterior (AP) radiographs taken from equal number of male and female THA patients from 2000 to 2017 with 300 of the cases ultimately dislocated and 300 cases without dislocation. Among these cases, 200 had osteoarthritis, 200 had rheumatoid arthritis, and 200 had other indications. Manual annotation, augmentation, and random splitting for 80% training, 10% validation, and 10% testing data sets are applied. Training of the models based on sex, underlying pathology, and ultimate dislocation status are critical considerations in the models generated. Two U-Net CNN models are formed to segment AP pelvis and cross-table lateral hip images independently. The encoders of both models had the VGG-16 architecture, and initial weights were pooled from a model pretrained on the ImageNet database. Well-known Adam optimizer is used after training the network's decoder layers for 50 epochs with a batch size of 8. Model performance is evaluated on independent test data sets that were not used for training and validation. The inclination angle model had performance values of 91.3% for acetabular component and 84.3% for ischial tuberosity. The anteversion angle model had performance value of 90.3% only for acetabular component. Less than 2.5% of the cases had differences of 5° or more when human and deep learning measurements are compared. The high accuracy of the CNN models showed their effectiveness in automating the measurement of angular position of acetabular components.

Deep learning and machine learning models are developed in [22] as a part of natural language processing for efficient and accurate hip dislocation detection following primary THA by using standard (radiology notes) and non-standard (follow-up telephone notes) free-text medical narratives. After preprocessing, 105 out of 1890 patients had a dislocation sustained that resulted in a total of 380 radiology and 174 telephone notes. No indication of a dislocation is found in 2634 radiology and 609 telephone notes. Traditional machine learning models used included generalized linear model, KNN, random forest, SVM, and shallow neural network. The deep learning models included long short-term memory (LSTM) model and a CNN model. The classification of both deep and machine learning models is tuned to detect radiology notes that relate to three categories: (1) current dislocation, (2) evidence of previous dislocation, and (3) no dislocation. The proposed CNN model achieved the best overall performance for classification of both the radiology and telephone notes into the above-mentioned three categories. Therefore, the developed CNN model in [22] can be used for accurate and efficient hip dislocation detection from free-text medical narratives.

Mechanical loosening detection of THA implants is analyzed in [23] by using a deep learning algorithm and two different methods that utilize saliency maps and activation maximization [8]. Saliency map identifies the pixels most significantly affect the CNN classification output by ranking all the pixels of an input image based on their relative influence on a specific class score. An input image is generated by activation maximization for each filter that maximizes that filter's output [8]. 40 patients' image-specific saliency maps are used in [23] for training a CNN with 17 mechanically loose and 23 with well-fixed THA for detecting mechanical loosening of THA implants by classifying the input x-rays into categories of "loose" and "well-fixed." The first layer of CNN that looks directly at the x-ray image learns to detect very simple patterns such as horizontal and vertical lines in the image, while deeper layers that consist of middle and last convolutional layers learned more complex filters. The usefulness of combining saliency maps and activation maximization is shown for accurate mechanical loosening detection that can be used by decision-makers for revision surgeries.

AI, DL, and ML are also used for research on hip fractures that relate to THA; we cover only one research article in this area of interest as an example of an application; however, this area of interest is not a direct application of THA; therefore, it is not covered here extensively. Detection of hip fractures by using a deep convolutional neural network (DCNN) on plain pelvic radiographs upon THA is designed in [24]; 25,505 limb radiographs collected between the beginning of 2012 and end of 2017 are used with the retraining of 3605 frontal pelvic radiographs. Some of the deep learning research evaluating medical images use cropped images to avoid "black box" mechanisms such as [25] and enhance the accuracy of final validation, while authors of [24] reduced the image matrix size to 512 × 512 pixels instead. DenseNet-121 is used as the architecture of the designed neural network by using pixel values from the digital images as inputs using convolution and pooling techniques on each layer and to adjust the weights in the neural network according to the difference between the output and true label. Designed DCNN yield to strong results including 91% accuracy, 98% sensitivity, 2% false-negative rate, and 98% area under the receiver operating characteristic curve (AUC) when tested on 100 additional images collected during 2017. Gradient-weighted class activation mapping (Grad-CAM) is used by the authors to confirm the validity of the model, and 95.9% accuracy is attained by using the visualization algorithm for lesion identification (Figs. 1 and 2).

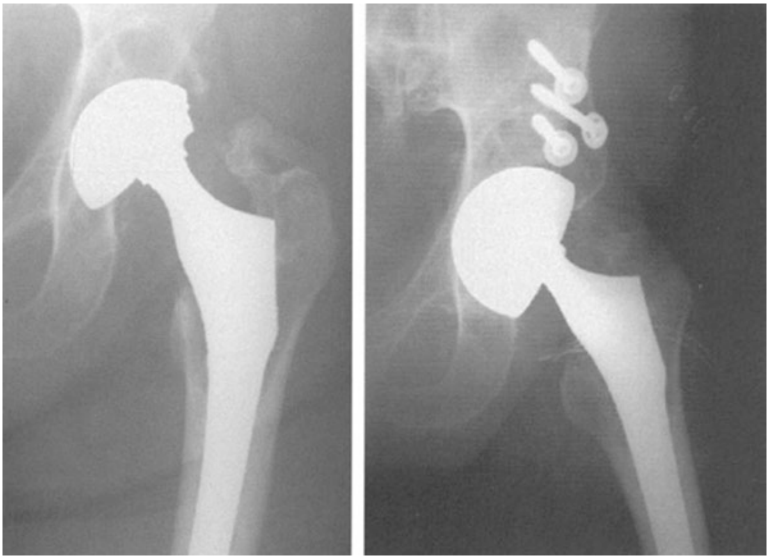

Fig. 1 An image that can be detected easily using DL of polyethylene wear on a radiograph [26]

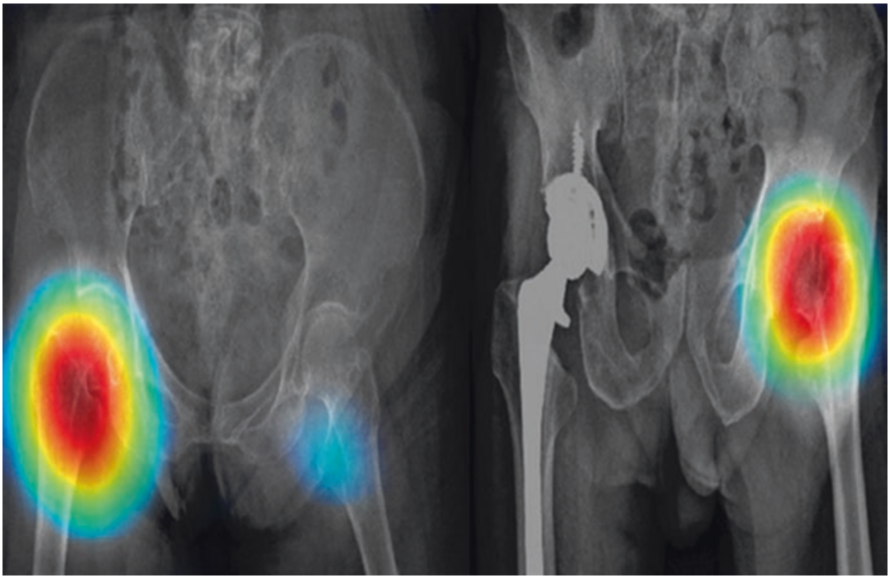

Fig. 2 Two images of gradient-weighted class activation mapping used for visualizing the class of discriminative regions for DL applications [24]

4 Artificial Intelligence

One of the earliest applications of ANN on THA is focused on patient comfort after THA based on bodily pain reduction in [1]. A total of 221 patients' survey data on 14 variables included gender, race, income, education, age at surgery, BMI, marital status, availability of help at home, preoperative effect of pain on physical function, preoperative support requirements, preoperative reported change in health over the year prior to surgery, pain-limiting activities' frequency, effect of pain on work, and preoperative SF-36 pain score. The ANN designed is trained by using 26 input nodes to predict the relative success of THA surgery using the presurgical patient survey information and a backpropagation feedforward neural network training to predict the output variable using the jackknife method. The best ANN achieved 83% of total percentage correctness and 62% of weighted percentage correctness. Area under the receiver operating characteristic curve is determined to be 79%. In conclusion, authors pointed out the success of neural networks to predict the success of THA accurately. Such an approach found to be feasible for predicting patients at greatest risk of poor outcomes based on their reported surveys.

The usefulness of ANN for failed implant identification is investigated in [27]. A total of 2116 AP hip radiographs capturing femoral stem implantation following THA from 2002 to 2019 are analyzed. Training is conducted on 1410 AP hip radiographs with an additional 706 used for validation and a unique consecutive series of 324 radiographs used for testing accuracy. The neural network architecture performance is trained, validated, and tested by using AlexNet, DenseNet, GoogLeNet, Inception-ResNet-v2, Inception-v3, ResNet-101, ResNet-50, ResNet-18, SqueezeNet, VGG-19, and VGG-16. Among all the options, Dense-Net 201 architecture attained 100% accuracy in training data, 95.15% accuracy on validation data, and 91.16% accuracy that outperformed the other options. The ANN utilization in iPhone 6 cellular phone application resulted in approximately 1-second runtime. Therefore, the ANN designed is determined to be a strong predictor for failed implant identification.

It is important to determine the manufacturer and the model of the hip implant upon hip arthroplasty. Radiographs are used for implant classification by experts specialized in the subject matter. Delays in care, increased morbidity, and additional economic burden are consequences of unidentifiable hip implants. A CNN algorithm is designed in [28] for differentiating and detecting 18 different hip implants by Zimmer, DePuy, Stryker, and Smith & Nephew manufacturers based on plain radiographs. 1972 AP plain radiographs from 4 sites are collected with 1559 used for training, 207 used for validation, and 206 used for external testing of the CNN. Input images are rescaled to 299 × 299 pixels. After preprocessing, inception V3 network is utilized with pixel normalization to the range of −1 to 1. The network is trained by using all training images for a total of 1000 epochs. Accuracy, sensitivity, specificity, and area under the receiver operating characteristics curve of the model are calculated for determining model performance in predicting the correct implant during both validation and external testing sets. Designed CNN demonstrated progressive "learning" through the 1000 epochs by improving validation accuracy and decreasing validation loss function values. CNN achieved 99.6% accuracy, 94.3% sensitivity, 99.8% specificity, and a value of 99.9% for area under the receiver operating characteristics curve as the average of all 18 manufacturers' implant identification. Implant stem designs for all of accuracy, sensitivity, and specificity included the following:

- 100% for Zimmer Biomet Arcos, Zimmer Biomet Taperloc, DePuy Corail, DePuy SROM, Smith & Nephew Birmingham, Smith & Nephew Synergy, Stryker ABG, and Stryker Exeter
- At least 99.5% for DePuy AML, DePuy Summit, Stryker PCA, and Stryker Restoration Modular.

The other six brands also had strong results with a minimum of 98.1% accuracy and a minimum value of 98.3% specificity, except with two minimum values of 66.7% sensitivity attained for two brands. Hence the CNN generated in [28] for differentiating the 18 hip arthroplasty implant models from four industry leading manufacturers demonstrated its effectiveness (Fig. 3).

An ANN non-parametric metamodel is used as a tool for sensitivity analysis in a cost-effectiveness model in [29]. The decision analytical model used is developed in [30] to investigate the effectiveness and cost-effectiveness of alternative hip prostheses. The metamodels are developed in two stages with the first screening phase emphasizing a nonlinear factor screening for importance analysis to reduce the number of variables attained from the simulation and second phase employing an

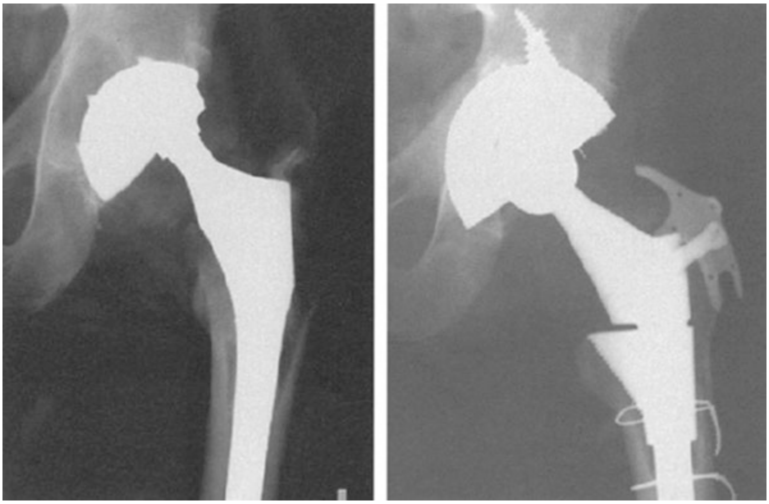

Fig. 3 An image of DePuy anatomic medullary locking cup with acetabular cup system liner [26]

ANN to structure an input-output relationship of the cost-effectiveness model [29]. The performance of the resulting ANN is compared with multiple linear regression and Gaussian process based on Charnley and Spectron prosthesis. 12 of the 31 features are selected from the simulation. Mean square error of prediction and mean absolute percentage deviation of the ANN meta models displayed the best performance measures for predicting both costs and quality-adjusted life years for the two prostheses. Overall, both ANN and linear regression models predicted the quality-adjusted life years highly accurately while ANN showing the best predictive capability for costs in THA model. ANN model is determined to be a good predictive modeling technique for health economic simulations.

Automated record search for text detection by using ANN in comparison to classic record search by two manual reviewers is investigated in [31]. Manual patient record analysis included hospitalization report, surgery report, and postoperative outpatient clinical report and excluded radiographic, laboratory, and pathology reports that were not reviewed. Surgery and implant characteristics such as implant size and implant articulation were extracted with any reported adverse events, and their respective treatments were recorded. The purpose of ANN development is to establish ease of access and increasing quality of accurate monitoring of the THA patients' records. A text mining engine utilizing a natural language processing technology and machine learning for extracting key concepts from electronic medical records are the two key components of the algorithm. Recall, precision, accuracy, and F-values are used as the statistical measures for the data collected from 532 patients and 613 hips. As a result, the comparison of manual and ANN search for implant characteristics resulted in significantly higher accuracy of the algorithm with 94.8% than the accuracy of the reviewer with 93.4%. ANN algorithm demonstrated better results than the manual process even in the case of existing clear pattern for implant sizes with low-level training. Overall performance of the algorithm is measured as 96% for recall, 88% for precision, and an F-value of 0.89 for all adverse events. The automated ANN search algorithm is determined to be capable of analyzing and interpreting large quantities of electronic medical records faster than the manual search with a performance level equivalent of comparable or slightly better than a human reviewer.

5 Conclusion and Possible Improvements

In this work applications of AI, DL, and ML that relate to THA in the research literature are covered. A variety of research results are covered throughout this article with implant design and failure, post-THA patient satisfaction, database search and text detection, and biomechanical considerations. Deep learning results attained for the THA applications covered in this work particularly have strong results for the most part. Current applications of the algorithms have limited scope; however, more advanced results can be attained. Such applications can include parallel computing, integration of DL directly into hardware applications used in THA, and integrating optimization algorithms into AI/DL/ML algorithms. Supervised learning methods can be particularly helpful in applications. Adaptive learning approaches can also be included based on multiple surgeries on same patient types. There are many more AI/DL/ML applications that can be integrated into other advanced technologies that can guide surgeons during THA. We must note the results of the reviewed articles in this work are particular instances of applications of the AI theory; therefore, they may not be able to yield good results in other collected data sets necessarily; there are many factors that play in such research results.

To the best of our knowledge, utilization of AI, DL, and ML on psychological treatment of patients to prevent them go through THAs has not been investigated in the research literature. Such research requires specific data collection from THA candidates who go through psychological treatment; after such a therapy, patients' decision to pursue or not pursue with THA treatment can be determined. The current practice in elective orthopedics does not routinely include psychological interventions despite evidence that psychological factors such as personality, anxiety, depression, and negative thinking styles can influence outcomes and recovery from surgery [32]. In fact, there is very limited research and investment on impact of psychological treatment on patients to prevent going through THA, and the majority of the literature focuses on the impact of psychological treatment based on pre- and post-THA outcomes. The application of AI theory with the corresponding feature (i.e., variable) selection during psychological treatment and analyzed along with the success of the treatment for declining occurrence of THA appears as a brand-new research area. Noting that the average age of THA patients is getting younger over the years, effectiveness of psychological treatment can be investigated for declining the increase in THA over the years. This idea leaves us with a brand-new THA research area application from a psychological standpoint that can also be applied in other surgical procedures: Can we use AI, DL, and ML effectively to determine features that help THA candidates prevent going through THA after psychological treatments and help them to heal naturally? If the answer is yes, then these features can help to decline the increase in THA procedures by the help of psychologists focusing on helping the patients.

References

1. Schwartz MH, et al. Using neural networks to identify patients unlikely to achieve a reduction in bodily pain after total hip replacement surgery. Med Care. 1997;35(10):1020.
2. Ramkumar PN, et al. Development and validation of a machine learning algorithm after primary total hip arthroplasty: applications to length of stay and payment models. J Arthroplast. 2019;34(4):632–7. https://doi.org/10.1016/j.arth.2018.12.030. Epub 2018 Dec 27
3. Zhong H, et al. Machine learning approaches in predicting ambulatory same day discharge patients after total hip arthroplasty. Reg Anesth Pain Med. 2021;46(9):779–83.
4. Kunze KN, et al. Development of machine learning algorithms to predict clinically meaningful improvement for the patient-reported health state after total hip arthroplast. J Arthroplast. 2020;35(8):2119–23.
5. Shah AA, et al. Development of a novel, potentially universal machine learning algorithm for prediction of complications after total hip arthroplasty. J Arthroplast. 2021;36(5):1655–62.
6. Polus JS, et al. Machine learning predicts the fall risk of total hip arthroplasty patients based on wearable sensor instrumented performance tests. J Arthroplast. 2021;36(2):573–8.
7. Huang ZY, et al. Predicting postoperative transfusion in elective total HIP and knee arthroplasty: Comparison of different machine learning models of a case-control study. Int J Surg. 2021;96:106183.
8. Huang G, Liu Z, Pleiss G, et al. Convolutional networks with dense connectivity. IEEE Trans Pattern Anal Mach Intell. 2019:1–1. https://doi.org/10.1109/tpami.2019.2918284.
9. Karnuta JM, et al. Bundled care for hip fractures: a machine learning approach to an untenable patient-specific payment model. J Orthop Trauma. 2019;33(7):324–30. https://doi.org/10.1097/BOT.0000000000001454.
10. Ricciardi C, et al. Improving prosthetic selection and predicting BMD from biometric measurements in patients receiving total hip arthroplasty. Diagnostics. 2020;10(10):815.
11. Cilla M, et al. Machine learning techniques for the optimization of joint replacements: application to a short-stem hip implant. PLoS One. 2017;12(9):e0183755.
12. Ramkumar PN, et al. Preoperative prediction of value metrics and a patient-specific payment model for primary total hip arthroplasty: development and validation of a deep learning model. J Arthroplasty. 2019;34(10):2228–2234.e1. https://doi.org/10.1016/j.arth.2019.04.055. Epub 2019 May 2
13. Kang Y-J, et al. Machine learning–based identification of hip arthroplasty designs. J Orthop Translat. 2020;21:13–7.
14. Chen Y-S, Cheng C-H. Identifying the medical practice after total hip arthroplasty using an integrated hybrid approach. Comput Biol Med. 2012;42(8):826–40.
15. Pawlak Z. Rough sets. Inf J Comput Inf Sci. 1982;11:341–56.
16. Greco S, et al. Rough sets theory for multicriteria decision analysis. Eur J Oper Res. 2001;129(1):1–47.
17. Sniderman J, et al. Patient factors that matter in predicting hip arthroplasty outcomes: a machine-learning approach. J Arthroplast. 2021;36(6):2024–32.
18. Hastie T. GLMNET: fit a GLM with Lasso or Elasticnet regularization. Vienna, Austria: R Foundation; 2008.

19. Kingma DP, Ba J. Adam: a method for stochastic optimization. BT – 3rd International Conference on Learning Representations, ICLR 2015. San Diego, CA, USA: Conference Track Proceedings 2015; 2015.
20. Kugelman DN, et al. A novel machine learning predictive tool assessing outpatient or inpatient designation for Medicare patients undergoing Total hip arthroplasty. Arthroplast Today. 2021;8:194–9.
21. Rouzrokh P, et al. A deep learning tool for automated radiographic measurement of acetabular component inclination and version after total hip arthroplasty. J Arthroplast. 2021;36(7):2510–2517.e6.
22. Borjali A, et al. Natural language processing with deep learning for medical adverse event detection from free-text medical narratives: a case study of detecting total hip replacement dislocation. Comput Biol Med. 2021;129:104140.
23. Borjali A, et al. Deep learning in orthopedics: how do we build trust in the machine? Healthcare Transformation (2020).
24. Cheng C-T, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. Eur Radiol. 2019;29(10):5469–77.
25. Gale W, Oakden-Rayner L, Carneiro G, et al (2017) Detecting hip fractures with radiologist-level performance using deep neural networks. arXiv:1711.06504.
26. Bono J, et al. Revision Total hip arthroplasty. New York: Springer; 1999.
27. Murphy M, et al. Artificial intelligence accurately identifies total hip arthroplasty implants: a tool for revision surgery. HIP Int (2021): 1120700020987526.
28. Karnuta JM, et al. Artificial intelligence to identify arthroplasty implants from radiographs of the hip. J Arthroplast. 2021;36(7):S290–4.
29. Alam MF, Briggs A. Artificial neural network metamodel for sensitivity analysis in a total hip replacement health economic model. Expert Rev Pharmacoecon Outcomes Res 2019;1.
30. Briggs A, Sculpher M, Dawson J, et al. The use of probabilistic models in technology assessment: the case of total hip replacement. Appl Health Econ Health Policy. 2004;3:79–89.
31. Van de Meulebroucke C, Beckers J, Corten K. What can we expect following anterior total hip arthroplasty on a regular operating table? A validation study of an artificial intelligence algorithm to monitor adverse events in a high volume, nonacademic setting. J Arthroplast. 2019;34(10):2260.
32. Bay S, Kuster L, McLean N, Byrnes M, Kuster MS. A systematic review of psychological interventions in total hip and knee arthroplasty. BMC Musculoskelet Disord. 2018;19(1):201. https://doi.org/10.1186/s12891-018-2121-8. Published 2018 Jun 21
33. Karhade AV, et al. Development of machine learning algorithms for prediction of sustained postoperative opioid prescriptions after total hip arthroplasty. J Arthroplast. 2019;34(10):2272–7.
34. Mont MA, et al. Artificial intelligence: influencing our lives in joint arthroplasty. J Arthroplast. 2019;34(10):2199–200.
35. Rapkin BD, et al. Development of a practical outcome measure to account for individual differences in quality-of life appraisal: the brief appraisal inventory. Qual Life Res. 2018;27:823e33.
36. Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clin Radiol. 2018;73:439–45. https://doi.org/10.1016/j.crad.2017.11.015.