Osteoarthritis and Cartilage Open 2 (2020) 100069

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131

Review

Machine learning in knee osteoarthritis: A review

C. Kokkotis [a,b,*], S. Moustakidis [c,d], E. Papageorgiou [a,c], G. Giakas [b], D.E. Tsaopoulos [a]

[a] Institute for Bio-Economy & Agri-Technology, Center for Research and Technology Hellas, Volos, Greece
[b] Department of Physical Education & Sport Science, University of Thessaly, Trikala, Greece
[c] Electrical Engineering Dept., University of Applied Sciences of Thessaly, Larisa, Greece
[d] AIDEAS OÜ, Narva mnt 5, Tallinn, Harju maakond, Estonia

ARTICLE INFO

Keywords:
Knee osteoarthritis
Feature engineering
Machine learning
Prediction
Classification
Segmentation

SUMMARY

Objective: The purpose of present review paper is to introduce the reader to key directions of Machine Learning techniques on the diagnosis and predictions of knee osteoarthritis.
Design: This survey was based on research articles published between 2006 and 2019. The articles were divided into four categories, namely (i) predictions/regression, (ii) classification, (iii) optimum post-treatment planning techniques and (iv) segmentation. The grouping was based on the application domain of each study.
Results: The survey findings are reported outlining the main characteristics of the proposed learning algorithms, the application domains, the data sources investigated and the quality of the results.
Conclusions: Knee osteoarthritis is a big data problem in terms of data complexity, heterogeneity and size as it has been commonly considered in the literature. Machine Learning has attracted significant interest from the scientific community to cope with the aforementioned challenges and thus lead to new automated pre- or post-treatment solutions that utilize data from the greatest possible variety of sources.

1. Introduction

Knee Osteoarthritis (KOA) is a degenerative disease of the knee joint and the most common form of arthritis causing pain, mobility limitation, affecting independence and quality of life in millions of people [1]. There is no known cure for KOA, but there are several medical, biological and environmental risk factors, both modifiable and non-modifiable, that are known to be involved in the development and progression of the disease [2]. The aforementioned data characterizing KOA are high-dimensional, heterogeneous and the limited number of simple logistic regression models are not capable of handling large numbers of risk factors and most importantly, any interactions between environmental and other medical and biological factors. Furthermore, they cannot identify the tendency of a healthy subject to show signs of the disease and its progression based on patient outcomes. Despite that, the power and importance of correct study design should not be underestimated. In the well-designed study even “simple” analysis can give trustful results. These significant shortfalls in OA risk prediction models require a completely different modelling and computational approach to the problem. Advanced machine learning techniques such as fuzzy-logic theory, discrimination metrics (e.g. mutual information gain indexes and Fisher discrimination ratios) and advanced classification models combined with novel and efficient feature selection methods suitable for very large data sets could significantly contribute to the problem of high dimensionality compared to the existing statistical techniques applied to the OA risk prediction problem.

Machine Learning (ML) is the study of how computer algorithms (i.e., machines) can “learn” complex relationships or patterns from empirical data and hence, produce (mathematical) models linking an even large number of covariates to some target variable of interest [3]. As mentioned before, the ability to analyze complex cases with a huge volume of data and the maximum possible results it renders ML a valuable tool against KOA. It is worth noting that ML has been applied in areas such as robotics [4], medicine [5], biochemistry [6], bioinformatics [7], meteorology [8], agriculture [9] and the economic sciences [10]. The importance of applying ML techniques to KOA has been documented by Jamshidi et al. [11] and Kluzek and Mattei [12] in 2019.

In this context this review has been carried out to allow each researcher to refer to the appropriate ML method in relation to KOA. To achieve this aim, the structure of the review is as follows. Section 2 *Machine Learning in a nutshell* presents the terminology and definitions, the types, tasks and models, which are used in the studies on which this review was based. Section 3 *Review of studies* presents the steps of the methodology that were followed for the collection and classification of the studies concerning ML techniques in KOA. In addition, it presents a

* Corresponding author. Institute for Bio-Economy & Agri-Technology, Center for Research and Technology Hellas, Volos, Greece.
E-mail address: c.kokkotis@certh.gr (C. Kokkotis).

https://doi.org/10.1016/j.ocarto.2020.100069
Received 21 December 2019; Received in revised form 15 April 2020; Accepted 17 April 2020

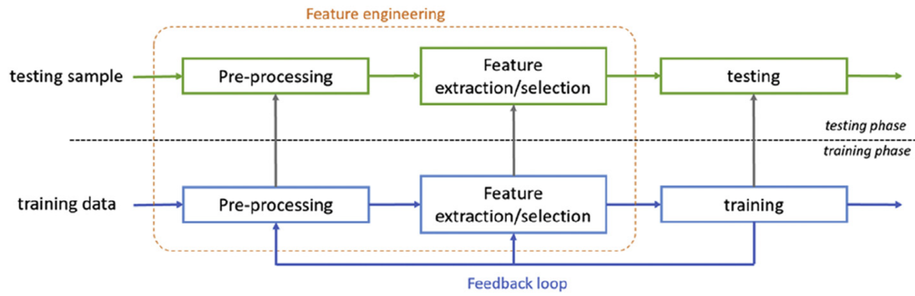

Fig. 1. A typical machine learning system.

summary of the studied literature, highlighting the main characteristics of proposed ML approached divided into four categories. The review ends with Section 4 *Discussion and Conclusions*, which mentions the future expectations and advantages that exist through the usage of machine learning in knee osteoarthritis.

1.1. Machine learning in a nutshell

In ML, a sample (e.g. a patient) is represented by a number of features which come in various forms and formats including patient's characteristics, risk factors, shape/texture characteristics in medical images or clinical history data. To facilitate the learning process, there features are typically concatenated forming a multidimensional feature vector. ML systems (Fig. 1) operate in two phases: the learning phase (training) and testing one. Indicatively, the role of the pre-processing unit can be broadly categorised into the following: (i) data cleaning aiming to remove noise, missing and inconsistent examples (ii) data integration in cases where multiple data sources are available and (iii) data transformation including discretisation and normalisation. The feature extraction/selection unit (also referred as feature engineering unit) attempts to generate and/or identify the most informative feature subset in which the learning model will be subsequently applied during the training phase [13]. The feedback loop allows adjustments of the pre-processing and feature extraction/selection units that will further improve the performance of the learning model. During the testing phase, the trained model is shown previously unseen samples (represented as images or feature vectors) that need to be classified. The model makes an appropriate decision (classification or regression) based on the features that are present in each sample. Deep learning [14], that is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain, sets an alternative architecture by shifting the burden of feature engineering (the process of transforming raw data into features) to the underlying learning system. From this perspective, feature extraction or selection are omitted leading to a fully trainable

Table 1
Presentation of indicative ML models along their characteristics.

Category	Models	Description	Advantages	Disadvantages
Bayesian	Naive Bayes, Gaussian Naive Bayes, Multinomial Naive Bayes, Bayesian Belief Network [29–32]	Probabilistic graphical models in which the analysis is undertaken within the context of Bayesian inference	They model uncertainty; easy to handle missing and hidden data	Increased computational cost in high-dimensional spaces; they require subjective definition of prior probabilities
Linear	Linear regression [17,18]	The best fit line through all data points	Easy to understand and implement; models can be easily interpreted	Too simple to capture complex associations between variables: prone to overfitting
	Logistic regression [17]	The adaptation of linear regression in classification problems		
Tree-based [33–36]	Decision trees (DT) [37–39]	A decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility		Not powerful enough in problems of high complexity
	Random forest (RF) [33]	Ensemble model that produces multiple decision trees, using a randomly selected subset of training samples and variables.	Fast to train and powerful	Not so interpretable; slower than other techniques
	Gradient boosting [40]	Uses weak decision trees as base models. Predictive results are obtained through increasingly refined approximations.	Fast and high performing	Interpretability issues; sensitive to small changes
Neural networks	Neural networks [41–50]	Information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information.	Can handle complex problems	Not interpretable; Slow
	Deep Neural networks (DNN) [51] such as CNN [52], deep belief network [53], and auto-encoders [54].		Can handle extremely complex problems	Require a lot of power; not interpretable; Slow
Instance based models	K-Nearest Neighbor [55], Locally Weighted Learning [56], Learning Vector Quantization algorithm [57], Self-Organising Maps [58]	Memory-based techniques that learn by comparing new examples with instances in the training database	Simple and fast to implement	Complexity grows with data (up to O(n) where n is the number of the training examples), prone to overfitting
Support vector machines (SVMs)	SVM [59,60] Least Squares SVM [61]	Finds a solution (linear or non-linear) that maximizes the margin between classes	SoA performance; generalized solutions; robust to high dimensionality	Tuning hyperparameters is crucial; time consuming and difficult to interpret

Table 2
Studies with Predictions/Regression techniques.

Author	Year	Data	Feature engineering	Learning Algorithm	Validation	Results
Abedin, J. [70]	2019	Questionnaire data /X-ray	LASSO	Elastic Net (EN), Random Forests (RF) and a convolution neural network (CNN)	70% training/30% testing	Root Mean Square Error (RMSE) for the CNN, EN, and RF models was 0.77, 0.97 and 0.94 respectively
Ashinsky, B. G. [68]	2017	MRI	–	Weighted neighbor distance using compound hierarchy of algorithms representing morphology WN(D-CHRM)	LOOCV	75% acc
Donoghue, C. [65]	2011	MRI	Laplacian Eigenmap Embedding	Multiple linear regression	270 knees as external validation group	Up to $R^2 = 0.75$
Du, Y. [69]	2018	MRI	PCA	ANN, SVM, Random forest, Naïve Bayes	10-fold cross validation (10F-CV)	ANN with AUC = 0.761 for KL grade Random forest with area under the curve (AUC) = 0.785 for JSM
Du, Y. [67]	2017	MRI	PCA	ANN, SVM, Random forest, Naïve Bayes	10F-CV	receiver operating characteristic (ROC) AUC of 0.761 (ANN)
Halilaj, E. [75]	2018	X-rays and pain scores	–	LASSO regression	10F-CV for model selection and 10% for model evaluation	AUC of 0.86 for Radiographic progression
Lazzarini, N. [77]	2017	Clinical variables, food and pain questionnaires, biochemical markers (BM) and imaging-based information	Ranked Guided Iterative Feature Elimination, PCA	Random Forest	10F-CV	AUC of 0.823 by using only 5 variables
Marques, J. [66]	2013	MRI	Texture Analysis for extraction and Partial least squares (PLS) regression for selection	Fisher linear discriminant analysis	10F-CV for model selection. 10% for evaluation	ROC AUC of 0.92
Nelson, A.E. [73]	2019	Demographic, MRI and biochemical variables	Distance weighted discrimination (DWD), PCA	K- means, t-SNE	Validation on 597 participants-	z = 10.1 (z-scores)
Pedoia, V. [71]	2018	MRI and biomechanics multidimensional data	Topological Data Analysis	Logistic Regression	–	AUC 83.8%
Tiulpin, A. [74]	2019	X-ray, Clinical data	CNN	Logistic Regression (LR) and Gradient Boosting Machine (GBM)	OAI dataset for training and MOST dataset for testing, 5F-CV	AUC of 0.79
Widera, P. [72]	2019	Clinical and X-ray image assessment metrics	Recursive feature elimination	Logistic regression, KNN, SVC (linear kernel), SVC (RBF kernel) and RF	Standard 10-fold stratified cross-validation protocol	F1 score 0.573–0.689
Yoo, T. K. [76]	2013	Kinematic data	–	SVM	Leave-one-out cross-validation (LOOCV)	97.4% acc

system that begins from raw or pre-processed input (e.g. image pixels or time-series) and ends with the final output of recognized objects or predicted values.

Learning can be classified as supervised, unsupervised or reinforcement learning. In *supervised learning*, each data sample is represented by a pair consisting of an input (typically a multi-dimensional feature vector) and a desired output value (e.g. a label having real-world meaning such as Kellgren Lawrence grades in case of KOA). The training phase involves the task of learning a function that maps every input to its associated output. The generated inferred function is used to map unknown inputs during the testing phase. *Unsupervised learning* [15] is a class of ML techniques that operate with unlabeled data with the goal of discovering structures or patterns in the dataset. Novel paradigms for unsupervised learning (the so-called self-supervised learning) have been also proposed exploiting different labelings that are freely available besides or within visual data to learn general-purpose features [16]. In *reinforcement learning*, a model learns through trial and error interactions with its environment using reward and penalty assignments.

In the terminology of ML, *classification* is considered as an instance of supervised learning. In short, it is the task of identifying to which of a set of categories (sub-populations) a new example belongs, on the basis of a training set of data (experience) containing examples whose label is known. *Regression* constitutes another supervised learning task, which aims to provide a prediction of an output variable according to the input variables which are known. The most known regression algorithms are the linear regression [17], as well as, stepwise regression [18]. Also, more complex regression algorithms have been developed, such as ordinary least squares regression [19], multivariate adaptive regression splines [20], multiple linear regression, and locally estimated scatterplot smoothing [21]. Table 1 cites the most well-known state-of-the-art ML models of the literature. *Dimensionality reduction* (DR) is a task that belongs in both families of supervised and unsupervised learning types, with the aim of providing a more compact lower-dimensional representation of a dataset preserving as much information as possible from the original data. It is usually performed prior to applying a classification or regression model in order to avoid the effects of the curse of dimensionality. Some of the most common DR algorithms are the following: (i) principal component analysis (PCA) [22], (ii) partial least squares (PLS) regression [23] and (iii) linear discriminant analysis (LDA) [24]. Finally, *clustering* [25] is an application of unsupervised learning typically used to find natural groupings of data (clusters). Well established clustering techniques are the K-means technique [26] hierarchical clustering [27], and the expectation-maximization technique [28].

Recently, deep learning has attracted wide-spread attention because of its enormous representing power, automated feature learning capability and best-in-class performance in solving complex problems [62]. Deep NNs make use of deeper architectures, extensible hidden units and nonlinear activation functions to model complex data, whereas one of their most attractive aspects is that they automate feature engineering thus alleviating the need for domain expertise and hardcore feature extraction. Currently, DL models have dramatically improved the state-of-the-art in many different sectors and industries including healthcare [63]. DL models can be either supervised, partially supervised, or even unsupervised. Convolutional neural networks (CNN) are among the most famous DL networks where feature maps are extracted by performing convolutions in the image domain. A comprehensive introduction on CNNs is given in Refs. [52]. Other typical DL architectures that belong to the family of probabilistic undirected graphical models are deep Boltzmann machines, and deep belief networks [53]. Auto-encoders [54] are unsupervised DNNs whose main idea is to encode high dimensional data into a low-dimensional latent vector and try to reconstruct the input data as flawlessly as possible by only using its

Table 3
Classification studies employing biomechanical data and/or distinct variables.

Author	Year	Data	Feature engineering	Learning Algorithm	Validation	Results
Aksehirli, Ö [92]	2013	Demographic characteristics and some gene polymorphisms	–	SVM, PNN	152 OA knees for training and 102 healthy for testing	76,77% acc & 90,55% acc
Beynon, M. J. [78]	2006	Biomechanical Data	Simulated annealing (SA) and genetic algorithms (GAs)	Dempster–Shafer theory of evidence (DST) & Linear discriminant analysis (LDA)	LOOCV	96.7% & 93.3% acc
de Dieu Uwisengeyimana, J [89].	2017	Biomechanical Data	Time-domain statistical features	Multilayer perceptron, Quadratic support vector machine, complex tree & deep learning network with k-NN	22 subjects (11 healthy and 11 OA)	99.5%, 99.4% 98.3% & 91.3% acc
Deluzio, K.J. [84]	2007	Biomechanical Data	PCA	Discriminant analysis	CV	Misclassification rate 8%
Jones, L. [85]	2008	Biomechanical Data	PCA	The Dempster–Shafer (DS)-based classifier & ANN	LOOCV	97.62% & 77.82% acc
Kotti, M. [83]	2014	Biomechanical data	PPCA	Bayes classifier	47F-CV	82.62% acc
Kotti, M. [90]	2017	Biomechanical data	–	Random forest	50% training/50% testing, 5F-CV	72.61% acc
Lim J. [86]	2019	Demographic and personal characteristics, lifestyle- and health status-related variables	PCA	DNN	66% training/34% testing	AUC of 76.8%
Long, M. J. [94]	2017	Outcome scores and biomechanical gait parameters	–	KNN	70% training/30% test. 30% of training was left out for validation	AUC of 1.00
McBride, J. [95]	2011	Biomechanical data	–	Neural networks	50% training/50% testing	75.3% acc
Mezghani, N. [79]	2008	Biomechanical data	Discrete wavelet transform (DWT)	Nearest neighbor classification (NNC)	LOOCV	38 of 42 cases acc
Mezghani, N. [80]	2008	Biomechanical data	Discrete wavelet transform (DWT) & Polynomial expansion	Nearest neighbor classifier (NNC)	LOOCV	91% acc 67% acc
Mezghani N. [96]	2017	Biomechanical Data	–	Regression tree	10F-CV for model selection. 10% for model evaluation	ROC AUC of 0.85
Moustakidis, S. [81]	2010	Biomechanical data	Wavelet Packet, FS via SVMFuzCoC	KNN1 SVM (AAA) SVM (1AA) FCT C4.5 FDT-SVM	10F-CV	86.09% acc 89.71% acc 90.18% acc 88.35% acc 91.12% acc 93.44% acc
Moustakidis, S. [88]	2019	Clinical Data	Feature subsets exploration	DNN Adaboost Fuzzy KNN Fuzzy NPC CFKNN	10F-CV	86.95% acc (for age 70+) 78.60% acc 77.39% acc 72.40% acc 73.60% acc
Phinyomark, A. [87]	2016	Biomechanical Data	PCA	SVM	10F-CV	98–100% acc
Şen Köktaş, N. [93]	2006	Biomechanical data	–	MLPs	CV	1.5 of the subjects has been misclassified
Şen Köktaş, N. [82]	2010	Biomechanical data (Also included age, body mass index and pain level)	Mahalanobis Distance algorithm	Decision tree - MLP multi-classifier	10F-CV	80% acc
Yoo, T. K. [91]	2016	Predictors of the scoring system in the Fifth Korea National Health and Nutrition Examination Surveys (KNHANES V-1) data	Logistic regression	ANN	66.7% training/33.3% validation, KNHANES V-1 (internal validation group) and OAI (external validation group)	ROC AUC of 0.66–0.88

Table 4
Medical image-based classification studies of KOA.

Author	Year	Data	Localization of joints	Feature engineering	Learning Algorithm	Validation	Results
Bien, N. [97]	2018	MRI	–	–	CNN (MRNet)	Validation A: 82,9% training, 8.5% tuning and 8,6 validation B: 60%-20%–20% into training, tuning, and validation sets using an external dataset	AUC of 0.937
En, Chuah Zhi [98]	2013	MRI	–	Discrete Wavelet Transform (DWT)	ANN-based	57,1% (200 images) training/ 42,9% testing (150 images)	94.67% acc
Kubkaddi, Sanjeevakumar [100]	2017	MRI	–	GLCM	SVM with RBF kernel, SVM with linear kernel & SVM with polynomial kernel	70% training/30% testing	95.45% acc, 95.45% acc & 87.8% acc
Kumarv, A. [101]	2017	MRI	–	GLCM	SVM	15 images/hold out validation	86.66% acc
Marques, J. [102]	2012	MRI	–	PLS with forward feature selection (PLS-FFS)	Fisher LDA, PLS regression, sparse PLS and sparse LDA	10F CV	ROC AUC of 0.86 (Diagnosis) & 0.63 (Prognosis), ROC AUC of 0.88 & 0.67, ROC AUC of 0.89 & 0.69, ROC AUC of 0.93 & 0.70, ROC AUC of 0.89 & 0.59
Pedoia, V. [99]	2019	MRI (T_2 relaxation time maps), Demographics and KOOS	–	PCA	Densely Connected Convolutional Neural Network (DenseNet) RF	65-20-15% split of training, validation, and holdout testing set	AUC = 83.44%, Sensitivity = 76.99%, Specificity = 77.94% AUC = 77.77%, Sensitivity = 67.01%, Specificity = 71.79%
Anifah, L. [103]	2013	X-ray	Gabor filter	GLCM	Self Organising Maps (SOM)	16,2% training/83,8% testing	Accuracy rate of 93.8% for KL-Grade 0, 70% for KL-Grade 1, 4% for KL-Grade 2, 10% for KL-Grade 3 and 88.9% for KL-Grade 4
Anifah, L. [105]	2018	X-ray	Gabor kernel	–	SOM	8,8% training/91,2% testing	40.52% acc for KL-Grade 2 & 36.21% for KL-Grade 0
Antony, J. [117]	2017	X-ray	FCN	FCN	CNN	70% training/30% validation, Multi-center validation	Multi-class classification accuracy 60.3%
Antony, J. [104]	2016	X-ray	Sobel horizontal image gradients, linear SVM	Pre-trained CNN (BVLC reference CaffeNet and VGG-*M*-128 networks)	Linear SVM	70% training (with 5F CV)/30% testing	Fitting a linear SVM produced 95.2% 5F CV and 94.2% test accuracy for knee joint detection; 57.6% accuracy in the muti-class KOA severity task (Grades 0–4)
Bayramoglou, N [108].	2019	X-ray	BoneFinder	Local Binary Patterns (LBP), Fractal Dimension (FD), Haralick features, Shannon entropy, and Histogram of Oriented Gradients (HOG)	Logistic regression	5F CV on OAI for training and validation in MOST data	AUC of 0.84
Chen, P. [110]	2019	X-ray	Customized one-stage YOLOv2 network	–	CNN models (VGG-19)	training, validation, and testing sets with a ratio of 7: 1:2.	69.7% acc
Gorriz, M. [118]	2019	X-ray	Trainable attention modules		CNN (VGG-16)	70% training/30% testing and 10% of the training data was kept for validation	64.3% acc
Gornale, Shivanand S. [112]	2017	X-ray	Images are cropped to 512 × 409 pixels and finally rescaled	Histogram of orientated gradients (HOG)	Multiclass SVM	Classification results validated by two experts that were in close agreement	Classification rate of 97.96% for Grade-0, 92.85% for Grade-1, 86.20% for Grade-2, 100% for Grade-3 & Grade-4
Liu, B. [119]	2020	X-ray	Region proposal network (RPN)	–	FLA (Faster R–CNN as original and our adjusted model as FLA)	5F CV	82.5% acc
Minciullo, L. [106]	2017	X-ray	PCA based combination of statistical shape and texture models	PCA- 3 stage Constrained Local Model	Indecisive Forest (IF) Optimised Indecisive Forest (OIF)	5F CV	87.61% acc 88.15% acc
	2017	X-ray	Shape Model	Statistical Shape Model (PCA)	Random Forest	5F CV	

(*continued on next page*)

Table 4 (*continued*)

Author	Year	Data	Localization of joints	Feature engineering	Learning Algorithm	Validation	Results
Minciullo, L. [107]							ROC AUC of 0.842 (binary) & 0.479 (5-class problem)
Navale, D. I. [113]	2016	X-ray	Dividing Image into Blocks	Texture analysis algorithm	SVM	71,4% training, 4,8% validation and 23,8% testing	For affected subjects' accuracy is 80%
Sharma, S. [114]	2016	X-ray	Cropping of images	Histogram method, GLCM and Canny Edge Detection Technique	SVM	75% training/25% testing.	95% acc
Tiulpin, A. [111]	2018	X-ray	FCN, as proposed in Antony 2017	–	CNN ResNet-34	67% training, 11% validation and 22% testing, multi-center validation	66.71% acc (multi-class Grades 0–4)
Tiulpin, A. [109]	2019	X-ray	Random forest regression voting approach implemented in a BoneFinder tool	–	An ensemble of deep residual networks with 50 layers, squeeze-excitation and ResNeXt blocks	5-fold subject-wise stratified CV	AUC of 0.98
von Tycowicz, C [120].	2019	X-ray	–	Shape Space, Graph Convolutional Filters	A multi-layer, feed-forward graph convolutional network	The data was split into training, validation, and test sets with a ratio of 2/3, 1/6, and 1/6, respectively	64.64% acc
Wahyuningrum, R. T [115].	2016	X-ray	Images were cropped around the knee properly	Contrast Limited Adaptive Histogram Equalization (CLAHE)-2DPCA/Structural 2-Dimensional Principal Component Analysis (S2DPCA)	SVM (Gaussian kernel)	3F CV	Up to 94.33% class accuracy for Grade 0
Wahyuningrum, R. T [116].	2019	X-ray	Manually cropping on the knee joint with dimensions of 400 × 100 pixels	CNN (VGG-16)	Long Short Term Memory (LSTM)	3F CV	75.28% acc

coding. Recurrent neural networks (RNN) is another important family of DL models that define unique topological connections between their neurons in order to encode temporal information in sequential data [64].

2. Methods

2.1. Literature search approach

This survey was based on research articles published between 2006 and 2020 using the search engines Scopus, PubMed and Google Scholar. During our search, we identified articles that used ML for the study of KOA by various techniques. Especially, for this search, the terms machine learning, deep learning and knee osteoarthritis were used. A prerequisite for the inclusion of an article in our research was the occurrence of one of the three terms mentioned as keywords, either in the title or in the abstract of each article.

2.2. Exclusion criteria

In the first instance, all articles retrieved and collected were examined for the title and the abstract by one of the authors. In order to reach our original goal, we excluded the following categories: non-English articles, postgraduate dissertations, doctoral dissertations, studies not involving people with knee osteoarthritis and studies using traditional technical statistics. All the selected articles have been presented either in journal papers or conferences. Finally, the rest of the authors reviewed again the titles and abstracts to ensure that they met the membership criteria.

2.3. Assessed outcomes

The studies, which are recorded in this article, were divided into four categories, namely (i) Predictions/Regression (13 studies), (ii) Classification (43 studies), (iii) Optimum post-treatment planning techniques (4 studies) and (iv) Segmentation (15 studies). The grouping was based on the technical characteristics of the ML methods and the application domain of each study.

Then, after separating the articles, the following information was extracted from each article: Author, Year of publication, Data (MRI, X-Ray, Kinetic and Kinematic data, Clinical data and Demographics), Feature Engineering approach, Learning Algorithm techniques, Validation and Results (evaluation of performance).

3. Results

3.1. Predictions/regression

Despite the fact that OA field has been relatively slow adopting advanced analytical models compared to other fields, nowadays many studies focus on developing ML prediction models for KOA based on medical imaging (Magnetic Resonance Imaging (MRI), X-ray), clinical information, self-reported and biomechanical data.

3.1.1. Data sources

Imaging technologies (either MRI or X-ray) were incorporated into the majority of advanced analytical models to predict knee articular cartilage morphology with accuracies varying from 76.1% up to 92% ([65–69]). Recently, the combination of multimodal data (medical images with clinical or biomechanical data) has formed the basis for more powerful and efficient models. To enhance the quality of the available raw data or overcome the curse of dimensionality, a number of sophisticated algorithms were reported in the literature including: (i) LASSO [70], Topological Data Analysis [71], Recursive feature elimination (RFE) [72], PCA [73] for dimensionality reduction or (ii) CNN [74] to extract new more informative deep features for images. The major finding of these studies was that the accuracy of image-based prediction

Table 5
Studies with ML-driven post-treatment planning techniques of KOA.

Author	Year	Data	Feature engineering	Learning Algorithm	Validation	Results
Chen, H·P. [123]	2016	Biomechanical data	Tilt angle calculation and initial posture classification algorithm	Multi-layer SVM	10-fold cross validation	90.6% on layer-1 SVM & 92.7% on layer-2 SVM
Huang, P·C. [124]	2017	Biomechanical data	Sequential forward feature selection (SFS)	Multi-class SVM	10-fold cross validation	Accuracy for rehabilitation exercises recognition is 100% and for motion identification is 97.7%.
Levinger, P. [121]	2009	Biomechanical data	SVM	SVM	LOOCV	Accuracy of 100% for the training set and 88.89% for the test set
Wittevrongel, B [122].	2015	Biomechanical data	k-equal frequency binning	Decision tree & Rule sets	LOOCV	Best accuracy 92.9% & 76.5% respectively

of KOA progression can be improved if it is complemented with data sources such as clinical data, self-reported and biomechanical data.

3.1.2. Learning techniques

Due to their efficiency and predictive performance, ensemble algorithms (RF or Gradient Boosting) were selected in five out of the twelve (12) studies in this category. However, a significant number of studies employed simpler models (e.g. linear regression models [65,75] or logistic regression [71]) to implement the regression or prediction task. Non-linear SVMs were also investigated in four (4) papers ([67,69,72,76]) and this choice could be attributed to the fact that they are relatively efficient in low and medium size feature spaces and that they generalize well. More complex learning (and subsequently more difficult to handle) approaches were finally tested in some studies ([67,69,70]) using NN-based architectures such as Artificial neural networks (ANNs) and CNNs.

3.1.3. Validation

In the majority of those studies, validation has been performed with n-fold cross validation. Hold-out (typically 70%/30% for training/testing) and Leave-one-out cross-validation (LOOCV) have also been observed as a validation approach in some of the studies. It is worth noting that Tiulpin et al. [74] used an independent test set (acquired in another center) for validation. An overview with all the studies including prediction models of KOA are shown in Table 2:

3.2. Classification

This section presents the outcomes of our survey on the application of classification models on the field of KOA research. It is worthwhile to note the plurality of different datasets along with the heterogeneity of data types used by each study. The identified data sources are: biomechanical data (kinematic-kinetic data and EMG signals), osteoarthritic outcome score, demographic characteristics, some gene polymorphisms, radiographs, X-ray and MRI. For this reason, we are grouping the studies into two categories, the first for biomechanical data-scores and the second for images.

3.2.1. Biomechanical data and discrete variables

3.2.1.1. Data sources. Biomechanical data were the most widely used source of information in the reported studies including kinematic-kinetic data and electromyography signals. Furthermore, clinical data consisting of self-reported, osteoarthritic outcome scores, demographic characteristics and some gene polymorphisms were used as additional sources complementing the biomechanical features.

3.2.1.2. Feature engineering. Feature extraction and dimensionality reduction have been applied to improve the predictive capabilities of the learning models as well as to increase their computational efficiency. A variety of algorithms and techniques were reported in the literature including: (i) Simulated annealing (SA) [78], Genetic algorithms (GAs) [78], Discrete wavelet transform (DWT) ([79,80]), Wavelet Packet [81], SVM-based Fuzzy criteria [81] and Mahalanobis Distance algorithm [82] for feature selection and/or extraction (ii) Probabilistic PCA (PPCA) [83] and PCA ([84–87]) for dimensionality reduction and (iii) feature subsets exploration or use of time-domain statistical features ([88,89]) to lead in more powerful learning models. PCA has been observed to be the most popular feature engineering technique due to its simplicity and easiness to handle.

3.2.1.3. Learning techniques. A variety of machine learning models were used for implementing the detection and/or classification tasks. KNNs and SVMs were the most frequently selected algorithms being tested in (7) out of nineteen studies in this subcategory. Furthermore, RF [90], DT [82], Dempster Shafer Theory [78,85], Bayes classifier [83] and Discriminant analysis [84]) were also investigated. Finally, the use of deep learning techniques (e.g. ANNs [85,91], PNNs [92], MLPs [82,89, 93] or CNNs [86,89]) was limited due to the nature of the available training datasets (heterogeneous features and small sample sizes).

An overview of the aforementioned studies is shown in Table 3:

3.2.2. Medical images

Medical images form a crucial source of information in the KOA research. The types of medical imaging that have been analysed in this survey were either MRI or X-ray. According to our knowledge only six studies have been presented in the literature, until now, that reported the development of MRI data analysis methodologies for the diagnosis of KOA. Only one of the aforementioned studies adopted a deep learning approach applying directly learning algorithms (CNN and specifically MRNet) on the available images without the inclusion of any feature selection technique [97]. The rest of the reported studies employed a number of feature engineering techniques prior to the application of the learning models. Discrete wavelet transform, Gray level Co-occurrence Matrix (GLCM) and PCA are among the algorithms that were used to either extract new features or reduce the feature space dimensionality. As regards the learning part, NNs ([98,99]), SVM [100,101] and LDA [102] were the most commonly employed models for early detection and diagnosis of KOA.

Localization of joints was a crucial task in the reported X-ray applications. Numerous approaches of varying complexity were applied such as filtering (Gabor, Sobel) ([103–105]), statistical shape/texture analysis ([106,107]), fully automated software tools (Bonefinder ([108,109])) or more sophisticated deep learning networks including YOLO and FCN ([110,111]). In some cases, manual cropping was also performed ([112–116]). PCA and GLCM were again selected in many of the reported papers to generate small and informative feature subsets, whereas several recent studies adopted CNN-based methodologies as an alternative for the feature extraction task. Deep learning networks (e.g. VGG-19, VGG-16, DenseNet, ResNet-34 and LSTM) were also involved in several studies acting as the main learning algorithm. State-of-the-Art ML models such as SVMs were finally selected in a few Xray-based studies to drive the decision making process. In most of the cases, validation was performed via k-fold CV and hold-out whereas some studies adopted more robust validation strategies (cross-center validation). The main characteristics of the reported image-based classification studies are shown in Table 4.

Table 6
Segmentation techniques applied on the KOA research.

Author	Year	Data	Feature engineering	Learning Algorithm	Validation	Results
Ababneh, S·Y. [136]	2010	MRI	Disjoint (non-overlapping) block-wise scanning, two-pass block discovery	A graph-cut based segmentation algorithm	30 images from the OAI database	96% acc
Ambellan, F. [126]	2019	MRI	–	Combination of Statistical Shape models (SSMs) and 2D/3D CNN	Datasets: (i) SKI10, (ii) OAI Imorphics and (iii) OAI ZIB	(i) 74.0 ± 7.7 Total score (ii) For femoral cartilage the DSC is 89.4%; for baseline and 89.1% (iii) The DSC is 98.6% for femoral bone, 98.5% for tibial bone, 89.9% for femoral cartilage, and 85.6% for tibial cartilage
Gan, H·S. [125]	2017	MRI	k-means clustering algorithm, Fuzzy c-mean	Flexible seeds labelling method	Manual validation by two experts on 10 images	Dice's reproducibility of 0.80 for observer 1 and 0.82 for observer 2
Gornale, Shivanand S. [132]	2019	X-Ray	ROI extraction using Sobel, Prewitt edge detection, Computation of basic statistical features	Otsu's Segmentation, Texture based Segmentation and KNN	532 digital Knee X-ray images	The accuracy rate of 91.16% for Sobel method, 96.80% for Otsu's method, 94.92% for texture method and 97.55% for Prewitt method is obtained
Kashyap, S. [134]	2016	MRI	Extraction of 3D Haar-like features from volume of interest (VOI)	LOGISMOS, just-enough interaction (JEI) as post-processing and Random Forest Classifier	The data from OAI were divided into two training sets with 15 and 13 which were used to train the NAF and the second RF classifier. 53 data-sets were used for testing	Border positioning errors (mm) Femur signed 0.03 ± 0.19 Femur unsigned 0.55 ± 0.11 Tibia signed 0.10 ± 0.17 Tibia unsigned 0.61 ± 0.14, For RF classifier: Femur signed −0.06 ± 0.18 Femur unsigned 0.56 ± 0.11 Tibia signed 0.16 ± 0.24 Tibia unsigned 0.65 ± 0.17
Kashyap, S. [135]	2018	MRI	Neighborhood Approximation Forests k-means clustering	Hierarchical Random Forest Classifier and LOGISMOS	108 MRIs from baseline, and 12 month follow-up scans of 54 patients	Cartilage surface positioning errors (in mm) of 4D Femur signed 0.01 ± 0.18 Femur unsigned 0.53 ± 0.11 at Baseline
Marstal, K. [133]	2011	MRI	Histogram equalization, extraction of similarity features from neighboring patches and PCA	K-means	MRI scans from 50 subjects (25 for training)	Average sensitivity, specificity and dice similarity coefficient of 0.853 ± 0.093, 0.999 ± 0.001, 0.800 ± 0.106 and 0.831 ± 0.095, 0.999 ± 0.001, 0.777 ± 0.054 on tibial and femoral cartilages respectively
Panfilov, E. [129]	2019	MRI	–	Deep learning U-net with two modern regularization techniques, namely, supervised mixup and UDA	5-fold cross-validation. Dataset A: 88 MRI images, Dataset B: 108 MRI images and Dataset C: 44 MRI images	Mean of volumetric DSCs is 0.907 (U-net + mixup, Dataset A) for femoral cartilage and DSCs is 0.821 (U-net + UDA2, Dataset C).
Park, S·H. [137]	2009	MRI	Combined Intensity and Shape Priors	Iterative Local Branch-and-mincut	LOOV on 8 3D MRI images	Average similarity index over 0.80 for normal participants and 0.75, 0.67, and 0.64 for participants with established knee OA
Swanson, M. S. [138]	2010	MRI	Manual selection of seed points, histogram and fitted Gaussian curves of the region	Threshold operation followed by conditional dilation and post-processing	Validation on 10 normal knees images and 14 knees with OA	Mean similarity Index 0.64–0.80
Tack, A. [127]	2018	MRI	2D U-net followed by statistical shape models of menisci	CNN (3D U-Net)	Validation on 5 different datasets of MRI images from OAI with 2F CV	DSCs was 83.8% for medial menisci (MM) and 88.9% for lateral menisci (LM) at baseline, and 83.1% and 88.3% at 12-month follow-up.
Tack, A. [128]	2019	MRI	–	3D CNN (3D U-Nets)	MRI data of 1378 subjects from the OAI (2F CV)	Accuracy of 88.02 ± 4.62 for medial tibial cartilage (MTC) and 91.27 ± 2.33 for lateral tibial cartilage (LTC) at baseline and 87.43 ± 4.02 and 90.78 ± 2.42 at 12-months follow-up
Tamez-Pena, J. G [139].	2012	MRI	Manual creation of atlases by experts using CiPAS	Multi-atlas segmentation using CiPAS platform	LOO on 48 MRI images	DSC 0.88 and 0.84 for the femoral and tibial cartilage
Tiulpin, A. [131]	2017	X-ray	Anatomically-based joint area proposal and Histogram of Oriented Gradients	SVM	The images from MOST were used to create training (991), validation (110) and test sets (473), Jyvaskyla (93), OKOA (77)	Mean intersection over the union equals to: 0.84 (MOST), 0.79 (Jyvaskyla) and 0.78 (OKOA).
Tiulpin, A. [130]	2019	X-ray	ROI localization using low-costs annotations	Hourglass-like encoder-decoder models for landmark localization	5-fold patient-wise cross-validation split stratified by a KL grade (748 knee joints in total)	Presicion 92.11 ± 0.34 at 2.5 mm

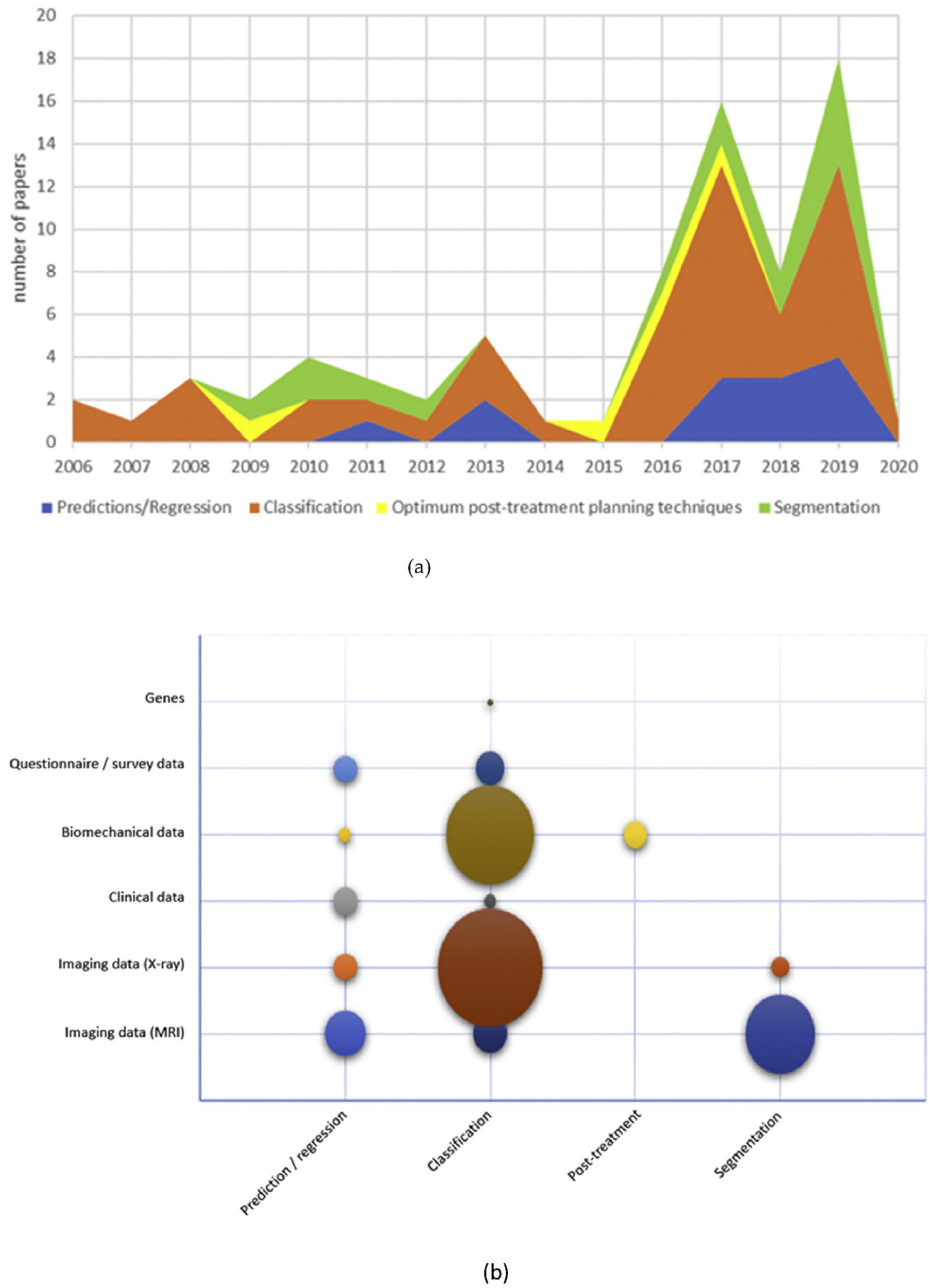

Fig. 2. a) A temporal evolution chart depicting the number of papers per category published each year since the year 2006 and included in the survey, b) Bubble chart showing a distribution of the papers considered in this survey arranged according to the data sources utilized in each survey category.

3.3. Optimum post-treatment planning techniques

As concluded in this survey, there is a lack of studies on the development of ML based decision support systems (DSS) for the post-treatment stage of KOA. According to our knowledge, the first attempt in that direction was made in 2009 in Ref. [121] where the authors presented an approach for detecting recovery from knee replacement surgery using gait spatio-temporal parameters. Their main aim was to investigate if the classifier could detect changes at 2 and 12 months following knee replacement surgery. The proposed method achieved to: (i) detect improvements in gait function and (ii) recognize gait parameters that are altered due to KOA. In Ref. [122], the authors tackled the task of selecting the appropriate gait re-training strategy as a ML problem and presented interpretable learning models. Using the trained models, a specialist was able to know which technique would work best for a specific patient. Online segmentation for KOA rehabilitation monitoring was also investigtaed in Ref. [123]. The novelty of this system was the real-time feedback to patients and physiotherapists. Finally, a SVM-based human motion identification for rehabilitation exercise assessment of KOA was proposed in Ref. [124] using biomechanical data with reliable results (up to 100% in recognizing the types of rehabilitation exercises and over 97.7% in motion identification). In the majority of the reported studies, the SVM technique was applied (in three out of four reports) on biomechanical data leading to even perfect identification rates (up to 100%). The validation was performed with 10-fold cross validation or with the leave one out (LOO) cross-validation approach. The studies with the ML-empowered post-treatment planning techniques of KOA are shown in Table 5.

3.4. Segmentation

Image segmentation is the process of changing the representation of an image into meaningful segments. MR image segmentation for KOA is typically performed by clinicians following a manual, laborious, time-consuming process that is prone to subjective diagnosis error. Therefore, many studies have focused on interactive, semi or fully automated cartilage segmentation to assist the medical research in KOA. At this point, it should be mentioned that even in the case of ML and especially in supervised learning approaches, a researcher/doctor still needs to label the images, hence the developed trained model is prone to the subjectivity.

3.4.1. Landmark localization and shape modelling

To increase the performance of medical image segmentation techniques, landmark localization and shape modelling have been utilized as preliminary tools before the application of ML or DL. As recorded, landmark localization took place by using either hourglass-like encoder-decoder models or with manual cropping and selection of seed points. Furthermore, a number of shape modelling tools were employed to extract informative shape-relevant characteristics from the available images including: Statistical Shape Models (SSMs), Combined Intensity, Shape Priors, Histogram of Oriented Gradients (HoG) and edge detectors.

3.4.2. Segmentation

Segmentation was accomplished employing either interactive or (semi- and/or fully) automated approaches. Flexible seeds labelling applied on MRI data [125] was the dominant approach on the integrative segmentation category. To enable automation on the segmentation tasks, advanced DL-based techniques were adopted (e.g. CNN ([126–128]), unsupervised domain adaptation DL [129] and DNN [130]) or even state-of-the-art ML techniques such as SVM ([131]), KNN ([132,133]) and RF ([134,135]). Finally, more traditional segmentation approached were also proposed including: two-pass block discovery mechanism [136], Iterative Local Branch-and-mincut [137], Gaussian fit model [138] and multi-atlas segmentation (MAS) [139].

3.4.3. Validation

OAI and MOST were the most-used databases to validate the performance of the aforementioned segmentation approaches. Validation was performed using k-Fold CV, LOOV or even manual assessment from experts.

An overview of all the identified KOA segmentation studies of our survey is given in Table 6:

4. Discussion and conclusions

Our literature survey outlined the current usage of machine learning methods in KOA diagnosis and prediction challenge. Fig. 2 shows an increasing trend of ML-related studies and papers in the field of KOA indicating the need for (i) enhancing our understanding about the onset and progression of the disease and (ii) new data-driven tools that could enable early diagnosis and prediction of KOA. ML could play a key role towards these directions extracting valuable knowledge from various types of clinical data (biomechanical parameters, images, kinematics) and finding new solutions that utilize data from the greatest possible variety of sources.

Data has to be seen as an asset being one of the most important and instructive assets of the healthcare industry. In KOA research, several data sources have been considered as inputs forming powerful multi-dimensional training and testing data sets. Medical Imaging is one of the dominant data sources of the sector with MRI and X-ray images being typically employed in the majority of the papers of our survey (25 and 25 papers out of 75 used MRI and X-ray, respectively). Biomechanical parameters were also investigated in 21 studies demonstrating a big potential to be useful input data in KOA diagnosis, prognosis and the post-treatment planning. Finally, other complementary data sources have been also considered in KOA research in several papers including pain, outcome scores, demographics, generic attributes and genes (Fig. 2).

Feature engineering algorithms were applied on the available clinical data to either reduce the input feature dimensionality or extract new informative parameters from the raw data. PCA was employed in a number of papers to compress 3D kinematic time-series, ground reaction forces and MRI/X-ray images into more compact representations. Time domain and time-frequency domain features (e.g. DWT or Wavelet packet) were also extracted from GRF or EMG signals. GLCM was proved to be a quite popular technique for extracting textural features in studies where MRI or X-ray images are considered as inputs. A number of feature selection techniques has been also employed to select the most informative features from the pool of the available or extracted parameters. Partial least squares, simulated annealing, random selection and sequential forward FS were among the techniques that were used to reduce the feature dimensionality of the initial space so as to increase the computational efficiency as well as generalisation capability of the subsequent classification or regressing models. Pre-trained CNN models were finally employed to extract valuable information for clinical images.

As far as the type of the ML models that were reported in our survey, SVMs were proved to be the most frequently used model in all the survey categories. Four (4) SVM-based studies were identified in the knee OA prediction survey, whereas another ten (10) papers made use of SVM for classification purposes including biomechanical discrete parameters or images (mostly MRI and X-ray). Moreover, SVM was also employed in three (3) out of the four (4) papers in the post-treatment survey. The choice of SVM could be attributed to the fact that they generalize well in practice and that are computationally effective in high dimensional spaces. Neural networks were the second most frequent technique with three (3) studies reported for knee OA prediction and eighteen (18) applications of NN-based models in the OA classification survey. Convolutional neural networks were finally considered in studies where clinical images were used as inputs. CNN-based approaches were either employed for feature extraction and/or for quantifying the severity of knee OA.

Nowadays biomedical research and clinical practices on KOA are struggling to cope with the growing complexity of interactions with the gained knowledge being fragmented and associated either with molecular/cellular processes or with tissue and organ phenotype changes related to clinical symptoms. Therefore, KOA is a big data problem in terms of the big data complexity and not the data size as it has been commonly considered in the literature. To tackle this huge complexity challenge, a multidisciplinary research approach should be proposed in the future across many disciplines: biomedical modelling via mechanistic analyses at various scales to capture locally the available knowledge into predictive simulations; medical imaging and sensing technologies to produce quantitative data about the patient's anatomy and physiology; data processing to extract from such data information that in some cases is not immediately available; big data analytics and computational intelligence tools that will generate personalised ‘hyper-models’ under the operational conditions imposed by clinical usage. Machine learning can explore massive design spaces to identify correlations and multiscale modelling can predict system dynamics to identify causality. This has the potential to lead to the development of individually tailored treatments to maximize the efficacy of treatment. Research work at the intersection of machine learning and KOA offers great promise for improving clinical decision-making, and accelerating relevant intervention programs. To enable appropriate adoption of advanced learning algorithms and stay tuned with the new developments in ML/DL that are embracing research to other medical fields, open data, tools, and discussions must be forceful encouraged within the KOA research community.

Declaration of Competing Interest

None.

Acknowledgments

This work has received funding from the European Community's H2020 Programme, under grant agreement Nr. 777159 (OACTIVE).

References

[1] M.J. Lespasio, et al., Knee osteoarthritis: a primer, Perm. J. (2017) 21.
[2] A. Courties, J. Sellam, F. Berenbaum, Metabolic syndrome-associated osteoarthritis, Curr. Opin. Rheumatol. 29 (2) (2017) 214–222.
[3] F. Cabitza, A. Locoro, G. Banfi, Machine learning in orthopedics: a literature review, Frontiers Bioengin. Biotechn. (2018) 6.
[4] A.C. Staugaard, Robotics and AI: an Introduction to Applied Machine Intelligence, Prentice-Hall Englewood Cliffs, 1987.
[5] Q.-H. Ye, et al., Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning, Nat. Med. 9 (4) (2003) 416.
[6] C. Helma, et al., Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds, J. Chem. Inf. Comput. Sci. 44 (4) (2004) 1402–1411.
[7] P. Larranaga, et al., Machine learning in bioinformatics, Briefings Bioinf. 7 (1) (2006) 86–112.
[8] C. Voyant, et al., Machine learning methods for solar radiation forecasting, Rev. Renew. Ene. 105 (2017) 569–582.
[9] J. Behmann, et al., A review of advanced machine learning methods for the detection of biotic stress in precision crop protection, Precis. Agric. 16 (3) (2015) 239–260.
[10] S. Mullainathan, J. Spiess, Machine learning: an applied econometric approach, J. Econ. Perspect. 31 (2) (2017) 87–106.
[11] A. Jamshidi, J.-P. Pelletier, J. Martel-Pelletier, Machine-learning-based patient-specific prediction models for knee osteoarthritis, Nat. Rev. Rheumatol. 15 (1) (2019) 49–60.
[12] S. Kluzek, T.A. Mattei, Machine-learning for osteoarthritis research, Osteoarthritis Cartilage 27 (7) (2019) 977–978.
[13] A. Zheng, A. Casari, Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists, O'Reilly Media, Inc, 2018.
[14] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436.
[15] P. Dayan, M. Sahani, G. Deback, Unsupervised Learning. The MIT Encyclopedia of the Cognitive Sciences, 1999, pp. 857–859.
[16] M. Noroozi, P. Favaro, Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles, Cham: Springer International Publishing, 2016.
[17] D.R. Cox, The regression analysis of binary sequences, J. Roy. Stat. Soc. B 20 (2) (1958) 215–242.
[18] M.A. Efroymson, Multiple regression analysis, Math. methods Digit. Comput 1 (1960) 191–203.
[19] B.D.I. Craven, S.M. N, Ordinary least-squares regression, SAGE Dict. Quant. Manag. Res (2011) 224–228.
[20] J.H. Friedman, Multivariate adaptive regression splines, Ann. Stat. 19 (1) (1991) 1–67.
[21] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Am. Stat. Assoc. 74 (1979) 829–836.
[22] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2 (11) (1901) 559–572.
[23] H. Wold, Partial least squares, in: Encyclopedia of Statistical Sciences, vol. 6, 1985, pp. 581–591.
[24] R.A. Fisher, The use of multiple measures in taxonomic problems, Ann. Eugen 7 (1936) 179–188.
[25] R.C. Tryon, Communality of a variable: formulation by cluster analysis, Psychometrika 22 (3) (1957) 241–260.
[26] P. Lloyd, S, Least Squares Quantization in PCM's, vol. 28, 1982, pp. 129–136.
[27] S.C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (3) (1967) 241–254.
[28] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Stat. Soc. B 39 (1) (1977) 1–38.
[29] S.J.N.P. Russell, Artificial Intelligence: A Modern Approach, vol. 9, 1995.
[30] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers Inc. 552, 1988.
[31] R.O. Duda, Pattern Classification and Scene Analysis, in: Richard O. Duda, Peter E. Hart (Eds.), Wiley, , New York, 1973. P.E. Hart.
[32] R.E. Neapolitan, Models for reasoning under uncertainty, Appl. Artif. Intell. 1 (4) (1987) 337–366.
[33] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[34] R.E. Schapire, A brief introduction to boosting, in: Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2, Morgan Kaufmann Publishers Inc, Stockholm, Sweden, 1999, pp. 1401–1406.
[35] Y.S. Freund, R. E., experiments with a new boosting algorithm, in: International Conference on Machine Learning, 1996, pp. 148–156.
[36] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.
[37] W.A. Belson, Matching and prediction on the principle of biological classification, J. Royal Stat. Soci. Series C (Applied Statistics) 8 (2) (1959) 65–75.
[38] L.F. Breiman, J. H, R.A. Olshen, C.J. Stone, in: Classification and Regression Trees, vol. 19, 1984.
[39] G.V. Kass, An exploratory technique for investigating large quantities of categorical data, Journal of the Royal Statistical Society. Series C (Applied Statistics) 29 (2) (1980) 119–127.
[40] J.H. Friedman, Stochastic gradient boosting, Comput. Stat. Data Anal. 38 (4) (2002) 367–378.
[41] D.S. Broomhead, D. Lowe, Multivariable functional interpolation and adaptive networks, Complex Syst. 2 (1988) 321–355.
[42] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain, Psychol. Rev. 65 (6) (1958) 386–408.
[43] S. Linnainmaa, Taylor expansion of the accumulated rounding error, BIT Numerical Mathematics 16 (2) (1976) 146–160.
[44] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: the RPROP algorithm, in: IEEE International Conference on Neural Networks, 1993.
[45] R. Hecht-Nielsen, Counterpropagation networks, Appl. Optic. 26 (23) (1987) 4979–4983.
[46] J.R. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man, and Cybernetics 23 (3) (1993) 665–685.
[47] W. Melssen, R. Wehrens, L. Buydens, Supervised Kohonen networks for classification problems, Chemometr. Intell. Lab. Syst. 83 (2) (2006) 99–113.
[48] J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. Unit. States Am. 79 (8) (1982) 2554–2558.
[49] S.K. Pal, S. Mitra, Multilayer perceptron, fuzzy sets, and classification, IEEE Trans. Neural Network. 3 (5) (1992) 683–697.
[50] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications. Neurocomputing 70 (1) (2006) 489–501.
[51] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[52] I. Goodfellow, Y. Bengio, A. Courville, Regularization for deep learning, Deep Learning (2016) 216–261.
[53] R. Salakhutdinov, G. Hinton, Deep Boltzmann machines, in: Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, D. David Van and W. Max, Editors. 2009, PMLR: Proceedings of Machine Learning Research, 2009, pp. 448–455.
[54] P. Vincent, et al., Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (2010) 3371–3408.
[55] E. Fix, J.L. Hodges, Discriminatory analysis. Nonparametric discrimination: consistency properties, Int. Stat. Rev./Rev. Int. Stat. 57 (3) (1989) 238–247.
[56] C.G. Atkeson, A.W. Moore, S. Schaal, Locally weighted learning. Artificial intelligence review 11 (1) (1997) 11–73.
[57] T. Kohonen, Learning vector quantization, Neural Network. 1 (1988) 303.
[58] T. Kohonen, The self-organizing map. Proceedings of the IEEE 78 (9) (1990) 1464–1480.
[59] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[60] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27.
[61] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Process. Lett. 9 (3) (1999) 293–300.
[62] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Network. 61 (2015) 85–117.
[63] L. Zhang, et al., A review on deep learning applications in prognostics and health management, IEEE Access 7 (2019) 162415–162438.
[64] R. Jozefowicz, W. Zaremba, I. Sutskever, An empirical exploration of recurrent network architectures, in: International Conference on Machine Learning, 2015.
[65] C. Donoghue, et al., Manifold learning for automatically predicting articular cartilage morphology in the knee with data from the osteoarthritis initiative (OAI), in: Progress in Biomedical Optics and Imaging - Proceedings of SPIE, 2011.
[66] J. Marques, et al., Diagnosis of osteoarthritis and prognosis of tibial cartilage loss by quantification of tibia trabecular bone from MRI, Magn. Reson. Med. 70 (2) (2013) 568–575.
[67] Y. Du, J. Shan, M. Zhang, Knee osteoarthritis prediction on MR images using cartilage damage index and machine learning methods, in: Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, vol. 2017, BIBM, 2017.
[68] B.G. Ashinsky, et al., Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative, J. Orthop. Res. 35 (10) (2017) 2243–2250.
[69] Y. Du, et al., A novel method to predict knee osteoarthritis progression on MRI using machine learning methods, IEEE Trans. NanoBioscience 17 (3) (2018) 228–236.
[70] J. Abedin, et al., Predicting knee osteoarthritis severity: comparative modeling based on patient's data and plain X-ray images, Sci. Rep. 9 (1) (2019) 5761.
[71] V. Pedoia, et al., MRI and biomechanics multidimensional data analysis reveals R2 -R1rho as an early predictor of cartilage lesion progression in knee osteoarthritis, J. Magn. Reson. Imag. 47 (1) (2018) 78–90.
[72] P. Widera, et al., Multi-classifier prediction of knee osteoarthritis progression from incomplete imbalanced longitudinal data, Preprint at (2019). https://Arxiv.org/Abs/1909.13408.
[73] A. Nelson, et al., A machine learning approach to knee osteoarthritis phenotyping: data from the FNIH Biomarkers Consortium, Osteoarthritis Cartilage 27 (7) (2019) 994–1001.

[74] A. Tiulpin, et al., Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data, Sci. Rep. 9 (1) (2019) 1–11.
[75] E. Halilaj, et al., Modeling and predicting osteoarthritis progression: data from the osteoarthritis initiative, Osteoarthritis Cartilage 26 (12) (2018) 1643–1650.
[76] T.K. Yoo, et al., Interpretation of movement during stair ascent for predicting severity and prognosis of knee osteoarthritis in elderly women using support vector machine, in: Conf Proc IEEE Eng Med Biol Soc, vol. 2013, 2013, pp. 192–196.
[77] N. Lazzarini, et al., A machine learning approach for the identification of new biomarkers for knee osteoarthritis development in overweight and obese women, Osteoarthritis Cartilage 25 (12) (2017) 2014–2021.
[78] M.J. Beynon, L. Jones, C.A. Holt, Classification of osteoarthritic and normal knee function using three-dimensional motion analysis and the Dempster-Shafer theory of evidence, IEEE Trans. Syst. Man Cybern. Syst. Hum. 36 (1) (2006) 173–186.
[79] N. Mezghani, et al., Hierarchical analysis and classification of asymptomatic and knee osteoarthritis gait patterns using a wavelet representation of kinetic data and the nearest neighbor classifier, J. Mech. Med. Biol. 8 (1) (2008) 45–54.
[80] N. Mezghani, et al., Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier, IEEE Trans. Biomed. Eng. 55 (3) (2008) 1230–1232.
[81] S.P. Moustakidis, J.B. Theocharis, G. Giakas, A fuzzy decision tree-based SVM classifier for assessing osteoarthritis severity using ground reaction force measurements, Med. Eng. Phys. 32 (10) (2010) 1145–1160.
[82] N. Şen Köktaş, et al., A multi-classifier for grading knee osteoarthritis using gait analysis, Pattern Recogn. Lett. 31 (9) (2010) 898–904.
[83] M. Kotti, et al., The complexity of human walking: a knee osteoarthritis study, PloS One 9 (9) (2014), e107325.
[84] K.J. Deluzio, J.L. Astephen, Biomechanical features of gait waveform data associated with knee osteoarthritis: an application of principal component analysis, Gait Posture 25 (2007) 86–93.
[85] L. Jones, C.A. Holt, M.J. Beynon, Reduction, classification and ranking of motion analysis data: an application to osteoarthritic and normal knee function data, Comput. Methods Biomech. Biomed. Eng. 11 (1) (2008) 31–40.
[86] J. Lim, J. Kim, S. Cheon, A deep neural network-based method for early detection of osteoarthritis using statistical data. International journal of environmental research and public health 16 (7) (2019) 1281.
[87] A. Phinyomark, et al., Gender differences in gait kinematics for patients with knee osteoarthritis, BMC Muscoskel. Disord. 17 (2016) 157.
[88] S. Moustakidis, et al., Application of machine intelligence for osteoarthritis classification: a classical implementation and a quantum perspective, Quantum Machine Intelligence 1 (3) (2019) 73–86.
[89] J. de Dieu Uwisengeyimana, T. Ibrikci, Diagnosing knee osteoarthritis using artificial neural networks and deep learning, Biomedical Statistics and Informatics 2 (3) (2017) 95.
[90] M. Kotti, et al., Detecting knee osteoarthritis and its discriminating parameters using random forests, Med. Eng. Phys. 43 (2017) 19–29.
[91] T.K. Yoo, et al., Simple scoring system and artificial neural network for knee osteoarthritis risk prediction: a cross-sectional study, PloS One 11 (2) (2016), e0148724.
[92] Ö. Aksehirli, et al., Knee osteoarthritis diagnosis using support vector machine and probabilistic neural network, 2013.
[93] N. Şen Köktaş, N. Yalabik, G. Yavuzer, Ensemble classifiers for medical diagnosis of knee osteoarthritis using gait data, in: Proceedings - 5th International Conference on Machine Learning and Applications, ICMLA, vol. 2006, 2006.
[94] M.J. Long, et al., Predicting knee osteoarthritis risk in injured populations, Clin. BioMech. 47 (2017) 87–95.
[95] J. McBride, et al., Neural network analysis of gait biomechanical data for classification of knee osteoarthritis, in: Proceedings of the 2011 Biomedical Sciences and Engineering Conference: Image Informatics and Analytics in Biomedicine, BSEC, vol. 2011, 2011.
[96] N. Mezghani, et al., Mechanical biomarkers of medial compartment knee osteoarthritis diagnosis and severity grading: discovery phase, J. Biomech. 52 (2017) 106–112.
[97] N. Bien, et al., Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet, PLoS Med. 15 (11) (2018), e1002699.
[98] C.Z. En, T.T. Swee, Computer-aided knee osteoarthritis classification system using artificial neural netwrok (ANN), J. Med. Imag. Hea. Inf 3 (4) (2013) 561–565.
[99] V. Pedoia, et al., Diagnosing osteoarthritis from T2 maps using deep learning: an analysis of the entire Osteoarthritis Initiative baseline cohort, Osteoarthritis Cartilage 27 (7) (2019) 1002–1010.
[100] S. Kubkaddi, K. Ravikumar, Early detection of knee osteoarthritis using SVM classifier, IJSEAT 5 (3) (2017) 259–262.
[101] A. Kumarv, A.K. Jayanthy, Classification of MRI images in 2D coronal view and measurement of articular cartilage thickness for early detection of knee osteoarthritis, in: 2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016-Proceedings, 2017.
[102] J. Marques, L.K.H. Clemmensen, E. Dam, Diagnosis and Prognosis of Ostheoarthritis by Texture Analysis Using Sparse Linear Models, 2012.
[103] L. Anifah, et al., Osteoarthritis classification using self organizing map based on gabor kernel and contrast-limited adaptive histogram equalization, Open Biomed. Eng. J. 7 (2013) 18–28.
[104] J. Antony, et al., Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016.
[105] L. Anifah, et al., Osteoarthritis severity determination using self organizing map based gabor kernel, in: IOP Conference Series: Materials Science and Engineering, 2018.
[106] L. Minciullo, et al., Indecisive trees for classification and prediction of knee osteoarthritis, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, pp. 283–290.
[107] L. Minciullo, T. Cootes, Fully automated shape analysis for detection of Osteoarthritis from lateral knee radiographs, in: Proceedings - International Conference on Pattern Recognition, 2017.
[108] N. Bayramoglu, et al., Adaptive segmentation of knee radiographs for selecting the optimal ROI in texture analysis, Preprint at (2019). https://Arxiv.org/Abs/1908.07736.
[109] A. Tiulpin, S. Saarakkala, Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks, Preprint at (2019). https://Arxiv.org/Abs/1907.08020.
[110] P. Chen, et al., Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss, Comput. Med. Imag. Graph. 75 (2019) 84–92.
[111] A. Tiulpin, et al., Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach, Sci. Rep. 8 (1) (2018) 1727.
[112] S.S. Gornale, et al., Determination of osteoarthritis using histogram of oriented Gradients and multiclass SVM, Int. J. Image Graph. Signal Process. (12) (2017) 9.
[113] D.I. Navale, R.S. Hegadi, N. Mendgudli, Block based texture analysis approach for knee osteoarthritis identification using SVM, in: 2015 IEEE International WIE Conference on Electrical and Computer Engineering, WIECON-ECE, vol. 2015, 2016.
[114] S. Sharma, S.S. Virk, V. Jain, Detection of osteoarthritis using SVM classifications, in: Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom, vol. 2016, 2016.
[115] R.T. Wahyuningrum, et al., A novel hybrid of S2DPCA and SVM for knee osteoarthritis classification, in: 2016 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications, CIVEMSA 2016-Proceedings, 2016.
[116] R.T. Wahyuningrum, et al., A new approach to classify knee osteoarthritis severity from radiographic images based on CNN-LSTM method, in: 2019 IEEE 10th International Conference on Awareness Science and Technology (iCAST), IEEE, 2019.
[117] J. Antony, et al., Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2017, pp. 376–390.
[118] M. Górriz, et al., Assessing Knee OA Severity with CNN attention-based end-to-end architectures, Preprint at (2019). https://Arxiv.org/Abs/1908.08856.
[119] B. Liu, J. Luo, H. Huang, Toward automatic quantification of knee osteoarthritis severity using improved Faster R-CNN, Int. J. Comp. Ass. Rad. Sur. (2020) 1–10.
[120] C. von Tycowicz, Towards shape-based knee osteoarthritis classification using graph convolutional networks, Preprint at (2019). https://Arxiv.org/Abs/1910.06119.
[121] P. Levinger, et al., The application of support vector machines for detecting recovery from knee replacement surgery using spatio-temporal gait parameters, Gait Posture 29 (1) (2009) 91–96.
[122] B. Wittevrongel, et al., Predicting Gait Retraining Strategies for Knee Osteoarthritis, 2015.
[123] H.P. Chen, et al., Online segmentation with multi-layer SVM for knee osteoarthritis rehabilitation monitoring, in: BSN 2016-13th Annual Body Sensor Networks Conference, 2016.
[124] P.C. Huang, et al., Human motion identification for rehabilitation exercise assessment of knee osteoarthritis, in: Proceedings of the 2017 IEEE International Conference on Applied System Innovation: Applied System Innovation for Modern Technology, ICASI, vol. 2017, 2017.
[125] H.S. Gan, et al., Flexible non cartilage seeds for osteoarthritic magnetic resoance image of knee: data from the osteoarthritis initiative, in: IECBES 2016-IEEE-EMBS Conference on Biomedical Engineering and Sciences, 2017.
[126] F. Ambellan, et al., Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: data from the Osteoarthritis Initiative, Med. Image Anal. 52 (2019) 109–118.
[127] A. Tack, A. Mukhopadhyay, S. Zachow, Knee menisci segmentation using convolutional neural networks: data from the osteoarthritis initiative, Osteoarthritis Cartilage 26 (5) (2018) 680–688.
[128] A. Tack, S. Zachow, Accurate automated volumetry of cartilage of the knee using convolutional neural networks: data from the osteoarthritis initiative, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE, 2019.
[129] E. Panfilov, et al., Improving robustness of deep learning based knee MRI segmentation: mixup and adversarial domain adaptation, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019.
[130] A. Tiulpin, I. Melekhov, S. Saarakkala, KNEEL: knee anatomical landmark localization using hourglass networks, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019.
[131] A. Tiulpin, et al., A novel method for automatic localization of joint area on knee plain radiographs, in: Scandinavian Conference on Image Analysis, Springer, 2017.

[132] S.S. Gornale, et al., Study of segmentation techniques for assessment of osteoarthritis in knee X-ray images, Int. J. Im. Graph. Sig. Process.(IJIGSP) 11 (2) (2019) 48–57.
[133] K. Marstal, et al., Semi-automatic segmentation of knee osteoarthritic cartilage in magnetic resonance images, in: Proceedings Elmar - International Symposium Electronics in Marine, 2011.
[134] S. Kashyap, et al., Automated segmentation of knee MRI using hierarchical classifiers and just enough interaction based learning: data from osteoarthritis initiative, Med Image Comput Comput Assist Interv 9901 (2016) 344–351.
[135] S. Kashyap, et al., Learning-based cost functions for 3-D and 4-D multi-surface multi-object segmentation of knee MRI: data from the osteoarthritis initiative, IEEE Trans Med Imaging vol. 37 (5) (2018) 1103–1113.
[136] S.Y. Ababneh, M.N. Gurcan, An automated content-based segmentation framework: application to MR images of knee for osteoarthritis research, in: 2010 IEEE International Conference on Electro/Information Technology, EIT2010, 2010.
[137] S.H. Park, et al., Fully automatic 3-D segmentation of knee bone compartments by iterative local branch-and-mincut on MR images from Osteoarthritis Initiative (OAI), in: Proceedings - International Conference on Image Processing, ICIP, 2009.
[138] M.S. Swanson, et al., Semi-automated segmentation to assess the lateral meniscus in normal and osteoarthritic knees, Osteoarthritis Cartilage 18 (3) (2010) 344–353.
[139] J.G. Tamez-Pena, et al., Unsupervised segmentation and quantification of anatomical knee features: data from the Osteoarthritis Initiative, IEEE Trans. Biomed. Eng. 59 (4) (2012) 1177–1186.