



Artificial Intelligence and Machine Learning: A New Disruptive Force in Orthopaedics

Murali Poduval¹ · Avik Ghose² · Sanjeev Manchanda³ · Vaibhav Bagaria⁴ · Aniruddha Sinha²

Received: 10 August 2019 / Accepted: 18 September 2019 / Published online: 13 January 2020
© Indian Orthopaedics Association 2020

Abstract

Orthopaedics as a surgical discipline requires a combination of good clinical acumen, good surgical skill, a reasonable physical strength and most of all, good understanding of technology. The last few decades have seen rapid adoption of new technologies into orthopaedic practice, power tools, new implants, CAD–CAM design, 3-D printing, additive manufacturing just to name a few. The new disruption in orthopaedics in the current time and era is undoubtedly the advent of artificial intelligence and robotics. As these technologies take root and innovative applications continue to be incorporated into the main-stream orthopedics, as we know it today, it is imperative to look at and understand the basics of artificial intelligence and what work is being done in the field today. This article takes the form of a loosely structured narrative review and will introduce the reader to key concepts in the field of artificial intelligence as well as some of the directions in application of the same in orthopaedics. Some of the recent work has been summarised and we present our viewpoint at the conclusion as to why we must consider artificial intelligence as a disrupting positive influence on orthopaedic surgery.

Keywords Artificial intelligence · Orthopaedic surgery · Machine learning

Abbreviations

AI Artificial intelligence
DL Deep learning
ML Machine learning
CNN Convolutional neural network

ANN Artificial neural network
RNN Recurrent neural network

Introduction

“Ultimately humans and computers will work together—not against one another”—Satya Nadella CEO Microsoft corporation [1].

“We can only see a short distance ahead, but we can see plenty there that needs to be done.” Alan Turing [2].

“You could say the God of Genesis himself is a programmer: language not manipulation is his tool of creation. Words become worlds. Today, sitting on the couch with your laptop, you too can be a god. Imagine a universe and make it real. The laws of Physics are optional.” Pedro Domingo, The Master Algorithm [3].

Artificial intelligence, which was once a subject of science fiction, is now invading every part of our lives, and changing it. This is quite clear from the intelligent suggestions you receive on your phone from Amazon or Flipkart, the way your Netflix page opens up and the way your Kindle

✉ Murali Poduval
muralipoduval@gmail.com; murali.poduval@tcs.com

Avik Ghose
avik.ghose@tcs.com

Sanjeev Manchanda
sanjeev.manchanda@tcs.com

Vaibhav Bagaria
bagariavaibhav@gmail.com

Aniruddha Sinha
aniruddha.s@tcs.com

¹ Tata Consultancy Services, Unit 129/130, SDF V, SEEPZ, Andheri East, Mumbai 400093, India

² TCS Research and Innovation, Tata Consultancy Services, Kolkata 700160, India

³ TCS Research and Innovation, Tata Consultancy Services, Unit 129/130, SEEPZ, Andheri East, Mumbai 400096, India

⁴ Reliance HN Foundation Hospital Mumbai, Mumbai, India

reading syncs across devices. Algorithms and artificial intelligence drive the analytics behind your health applications on the phone, the food suggestions you get on Swiggy or Zomato and even the way your Gmail inbox organises itself into different categories. Natural language processing powered face recognition and fingerprint recognition has slowly and steadily enhanced our requirements of security on our personal devices. Artificial Intelligence powers our flight choices and the seamless connections across continents and cities. The growth of the mobile phone and mobile computing industry, availability of internet services at reasonable costs, wide acceptance of smart wearables such as watches and fitness bands and the huge market for mobile applications has spurred the need for smarter analytics to enhance the customer experiences as well as business directions and insights.

In comparison with the other industries, healthcare has been relatively slow in adopting artificial intelligence. Delivery of healthcare is dependent on a large number of factors, of which the most difficult to reproduce is the physician's experience and intuition and logical interpretation of the patient's condition by correlating the available clinical examination with radiology and other investigation reports. The diagnostic process is so complex we never hope to reproduce it in a machine. The incredible complexity of healthcare delivery is, strangely, what makes it a very fertile ground for application of artificial intelligence. But now, technology is changing how doctors interact with their implements, of how the instruments deliver information to the doctors and how the resultant interpretation is used in aiding the physician and the patient make an appropriate choice of treatment. Much like the aviation industry, where pilots have increased their efficiencies and accuracy and safety by flying with the help of instruments, it is time for doctors too to do the same [4].

In this realm of enhanced technology and digital innovation, orthopaedic surgery holds its own special place. Orthopaedic surgeons have been quick to adapt and refine new technologies and integrate them with their practice. The last half century has seen the exponential growth of the joint replacement industry, amazing refinements in trauma care, rapid strides in imaging technology, integration of navigation and three dimensional imaging into the operating room and scores of instrument and implant innovations which have made surgery safer, predictable and efficient. The current trends in orthopaedic surgery are about digitisation, artificial intelligence and smart robotics. There has been considerable interest in the literature and scientific forums about the utilisation of machine learning in various domains in orthopaedic surgery. This narrative review takes a brief look at the basics and defining principles of Artificial Intelligence (AI) and Machine learning (ML), starting from the roots, and explores some of the areas, where it is probably

making an impact. This is not a comprehensive review of the subject but a brief introduction of the subject and a look at some of the important work in the field.

The Background: History of Artificial Intelligence

In 1947, Alan Turing spoke at the London Mathematical society and in October 1950 published a detailed paper entitled "Computing Machinery and Intelligence" [2]. He wrote about what is known as the Turing test and the methods that could be used to consider a machine intelligent, a test which he called "The Imitation game". The paper also talks about the concept of a "Child Program": which could be educated by mutation or natural selection imposed by the examiner. Artificial intelligence by itself was not a new thought, experiments with machine learning date back to before Turing, but Turing laid the foundations of what we know today to be modern AI. Even in the current day, Turing's paper makes compelling reading.

In 1955, John McCarthy proposed a study at Dartmouth which was directed at studying the concept of artificial intelligence through a ten man 2 month workshop, which was subsequently held in the summer of 1956. The official origin of the name "Artificial Intelligence" is believed to date back to this proposal originally authored by John McCarthy of Dartmouth University, Marvin Minsky of Harvard, Nathaniel Rochester from IBM and Claude Shannon from Bell telephone laboratories [5, 6].

Many of the concepts used in artificial intelligence (AI) owe their roots to statistics and probability theory. Early computer algorithms developed were in the domains of heuristic research, computer vision and natural language processing as well as early primitive robotics. The initial interests in artificial intelligence did not produce tangible results and soon funds for research in the field dried out (these periods are referred to as the AI Winters) [6]. There has been an upsurge of AI and ML applications in all industries including healthcare over the last two decades. In general Artificial intelligence or AI is said to have four evolutionary stages [7], as depicted in Fig. 1. These are (a) *Reactive machines* that learn from data and react to changes in an intelligent world (b) *Limited memory machines* that learn from experience and can perform both prediction and forecasting (c) *Machines with theory of mind*—that can understand underlying behaviours and are capable of understanding and reacting to complex scenarios including human emotion. (4) *Self aware machines*: Machines that, such as humans, seem to hold a sense of purpose and will learn the purpose by observing the universe and body of knowledge around them. These can have opinion and cognitive biases just like human beings.

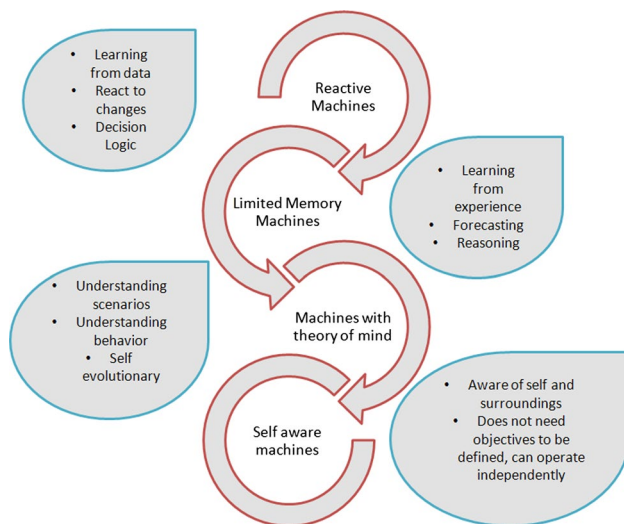


Fig. 1 Evolution of Artificial Intelligence

Healthcare providers, payers and life science CIOs (Chief Information Officer) listed machine learning and predictive analytics as the top game changing technologies in response to a Gartner survey [8]. Artificial Intelligence as a science is still evolving and is in the process of creating history in more ways than one.

Theory: Definitions and Concepts in Artificial Intelligence

McCarthy defined artificial intelligence as “the science and engineering that tries to make machines intelligent, trying to get them to understand human language and to reach problems and goals as well as a human being” [9, 10] AI can be defined on two broad approaches, one is a human-centric approach which is an empirical approach based on human behaviour and hypotheses on the same. The other is a rational approach which requires a combination of mathematics and engineering [11]. Nilsson’s definition which has been used by the Sanford hundred year report states that “Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.” [6]. The term intelligence is defined by McCarthy as the computational part of the ability to achieve goals in the world [10]. Artificial intelligence then has multiple domains such as heuristics, automatic learning, computer vision, natural language processing and intelligent agents [9].

Here one must also define two more terms, Weak AI and Strong AI (Artificial General Intelligence). Weak AI is most of the AI which we see in practice, which is task based and narrower and defined in its scope, in other words

weak AI is programs that behave as if they are thinking. Strong AI or artificial general intelligence is AI which actually thinks, reasons and takes action. This is still far from reality.

Russel and Norvig [11] identified the concept of a rational ‘agent’ to be central to artificial intelligence. An agent is defined as anything that can perceive its environment through sensors and can act upon that environment through ‘actuators’. The agent thus interacts with the environment through the sensors and actuators. In artificial Intelligence the agent is an agent program. This agent is defined to be rational if it can maximise its performance measure based on the evidence (the percept sequence) and built-in knowledge (learnt knowledge). The agent program is, therefore, trained, it learns and then acts to provide the desired action.

The *algorithm* is the building block of AI. An algorithm is an instruction given to the computer in the form of a sequence or steps which would lead to the desired output. This is in the form of a precisely written code in a language the computer understands [3]. It is most important that, given an input, the algorithm must be consistent and produce expected results. As Domingo says “Scientists make theories, engineers make devices and computer scientists make algorithms, which are both theories and devices” [3]. Algorithms work together to produce complex actions and also learn from each other to create new algorithms. The task is incredibly complex and involves space complexity (on the machine), time complexity (must use time efficiently) and complexity of relating to human nature (wherein the algorithms may get too complex to comprehend and to correct) [3]. The preferred method in artificial intelligence is to clearly distinguish tasks. This is done by building learning agents which are capable of operating in unknown environments. This divides the functional aspects of the program into a learning component responsible for making improvements and the performance component which executes the action [11].

Machine learning refers to the science of creating methods for machines to learn and apply analytical techniques, using algorithms for analysis of data, and generate an output using other algorithms [9]. Different terms are used interchangeably for machine learning including pattern recognition, statistical modelling, data mining, knowledge discovery, predictive analytics, adaptive systems etc [3]. Machine learning thus becomes a set of techniques to enable AI [12]. Machine Learning has been applied in medical research to identify quantify, analyse and interpret the relationship between many known variables as well as to discover hitherto unknown variables that may be at play in the given scenario. The approach to machine learning differs from classical statistics essentially in terms of methodology [12].

Techniques of Machine Learning

Various methods of learning have been described and they are broadly described as [13]

- Inductive learning* learning from specific input–output pairs, the learning algorithm is told what the output should be given a standard input. The variables are identified, annotated and the result is provided in the training of the algorithm. The algorithm uses this knowledge gained to analyse input data to provide results in a real-world situation.
- Deductive or analytical* in which a general rule is applied to the data and then it progresses to identify and learn a hitherto unknown rule. Data are not labeled and specified nor are outputs provided in the training, the algorithm must sift, classify, analyse and interpret data to provide the necessary outputs.

More commonly, learning methods are described on the basis of feedback as supervised, unsupervised, semi-supervised or reinforced learning.

Supervised Machine Learning

A typical machine learning system for supervised machine learning process will take historical data with actual output as target. Historical data are pre-processed to make the data set suitable for learning and model building. Following pre-processing of data, data are divided into training and testing data sets. Different algorithms are suitable for different type

of problem solving. There are a large number of classification and deep learning neural network algorithms available to create suitable models and one uses the most suitable method for the problem at hand. Often multiple algorithms may be suitable for a problem. In such a scenario, parallel experimentation helps in identifying the most suitable algorithm. Once a suitable algorithm is identified, then the algorithm is trained using training data and its' performance is tested using test data. The training data are divided into two parts by the algorithm, the training data are used for training and learning, whereas the validation data set is used for internal validation. Parameter tuning and post-processing on model plays an important role in optimising the performance of the model. Different metrics are then generated to analyse the performance of these algorithms. Therefore, the training process generates a model for predictions after suitable validation and testing. This model is then deployed to a production environment for predictions. Predictions are made when pre-processed and as yet unused data are fed to the algorithm as inputs. The predictions are presented as an output to the user. In addition, user feedback is fed back into the training and learning process for improving the model based on latest outputs (Fig. 2).

Unsupervised Machine Learning

Unsupervised learning helps in categorising information that does not have labeled information, the training data set is, therefore, absent. The algorithm needs to categorise information based on its own logic to create clusters of raw input data. The interpretation and relationships so derived are used to process an output, as depicted in Fig. 3. Some applications

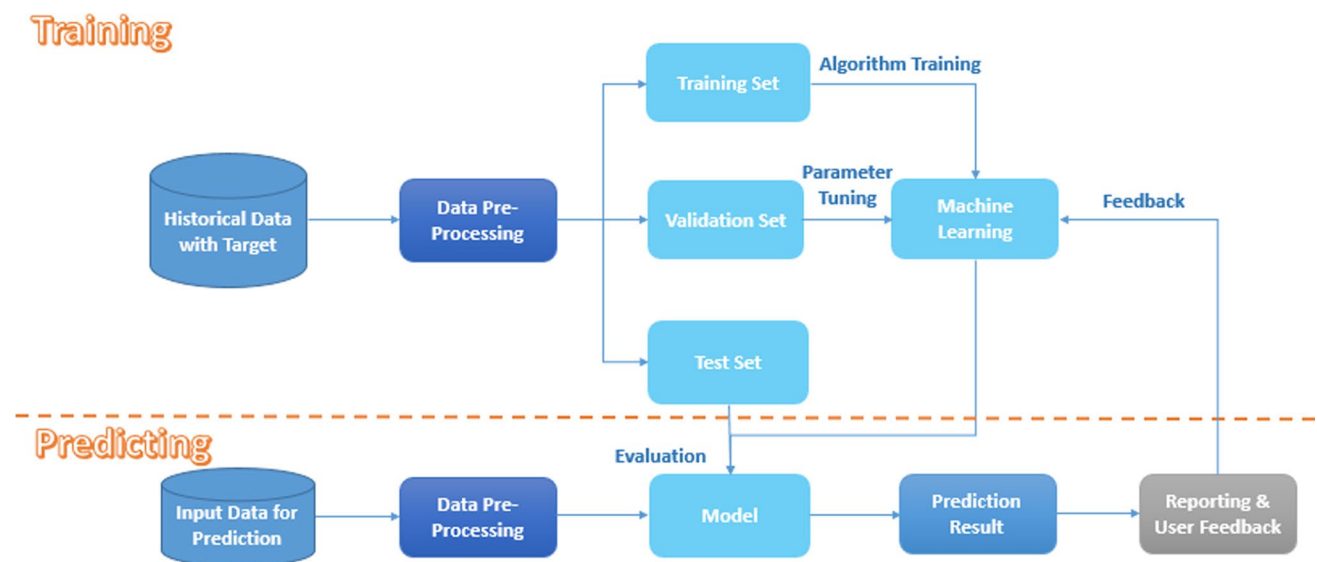


Fig. 2 Supervised learning: training prediction and feedback processing

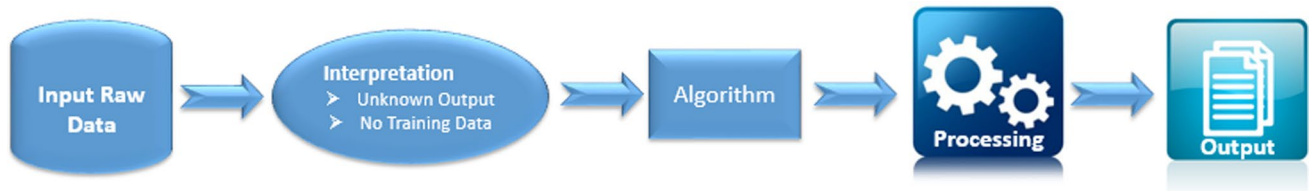


Fig. 3 Unsupervised learning process

of unsupervised learning are Clustering, Anomaly Detection etc.

Semi-supervised Machine learning

Semi-Supervised Learning as name suggests, combines both labeled and unlabeled data. The algorithm uses partly labeled data to categorise unlabeled data. Semi-supervised learning has applications in MRI, CT-scan etc., where few labeled examples of images labeled by experts, help in clustering unlabeled examples. The deep learning neural networks work on small set of annotated examples to classify unlabeled data in more accurate way than unsupervised learning.

Artificial Neural Networks (ANN) and Deep Learning (DL)

These are layered and complex machine learning models that attempt to mimic the organisation of the human brain. The layered organisation of interconnected neurons produces an output which is the resultant of the collaboration of the neurons, each neuron producing an output which is weighted according to the experience it has collected throughout the period of its use. An ANN network typically has an input layer and multiple intermediate layers and finally an output layer [9]. There are two well known models of Deep learning, the convolutional neural network (CNN) and the Recurrent neural network (RNN). DL is premised on learning complex hierarchical representations from data that have multiple levels of abstraction. Input neurons activate the next layer when the input crosses a defined threshold value [12]. Deep learning models are extremely useful to filter and organise noisy and messy data such as sensor data and microphone inputs. DL methods help in refining and classifying the data which can then be used as an input to standardised Bayesian or regression methods [9, 12]. For example, Deep Learning methods as unsupervised learning has been successfully used in identifying phenotypical groups for targeted intervention in heart failure with normal ejection fraction [14]. Deep learning is applied to sift through large masses of EHR and EMR data to identify patterns, which may set the stage for precision and personalised

medicine. Rajakomar et al. [15] applied DL to raw EHR data of over 200,000 hospitalisations from two academic institutions. They demonstrated the effectiveness of deep learning models in predicting length of stay, diagnosis at discharge, mortality and re-admissions at different time points, outperforming all traditional predictive models.

Applications: The Machine Learning Pipeline (Algorithm Development and Maintenance)

A typical machine-learning workflows consist of steps, as provided in Fig. 4. Furthermore, steps in machine learning or deep learning involve training, validation, validation testing cycles prior to deployment, as illustrated in Fig. 5.

Application steps

Pre-processing

Interpolation and filtering is typically done on time-series data with high sampling such as sensor data to remove measurement noise, environmental noise and outliers in measurement. Sensor fusion techniques such as Kalman filters, complimentary filters etc. are used to combine measurements from two or more sensors to estimate the true value more

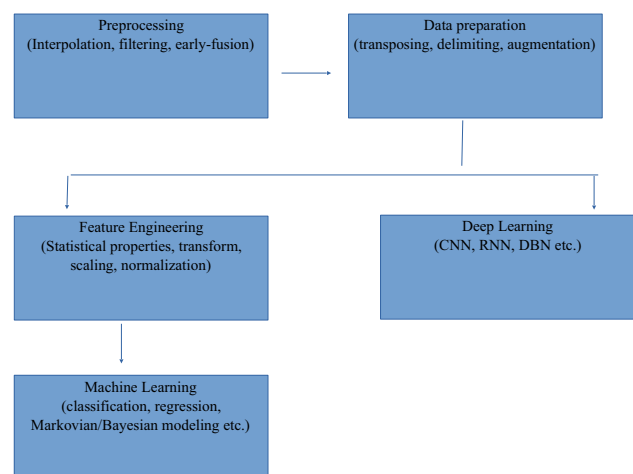


Fig. 4 A typical machine learning workflow

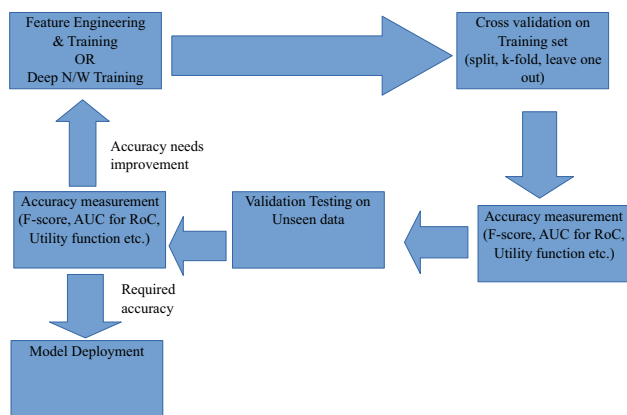


Fig. 5 The machine learning pipeline

closely. One good example is the MIT balance filter for fusing magnetometer and gyroscope data for inertial measurement systems [16].

Data Preparation

Before data can be provided to a machine-learning system, it needs to be neatly arranged in columns with each dimension separated by a delimiter. Such steps are typically considered under data preparation. Another problem that is often needed to handle in this step is data imbalance which is often handled by giving the minority class extra weight or using algorithms such as SMOTE [17].

Feature Engineering

Features are loosely defined as hidden properties in a data that have the three properties of independence, relevance and stability. Independence means that the property is not a linear or non-linear combination of other such properties or dimensions present in the data set. The property of relevance refers to the correlation of the property with class value or target variable value which is to be predicted using machine learning. The property of stability ensures that the feature is relatively free of environmental noise and sensor dependence which can be called reliability of a feature. Often these are referred to as the 3-Rs of maximum relevance, minimum redundancy and moderate reliability. There are many standard algorithms that provide a non-optimum check for features along these lines. Some notable ones include MRMR [18] and Feast [19]. It can be mathematically shown that finding an optimum solution is an NP-hard problem. Finally features are normalised using standard techniques to ensure no scaling problem exists in the data set due to different features having different dynamic ranges. Cross validation is the initial testing of ML accuracy performed on part of the training set.

Data Splitting

When there is substantial training data, it is often randomly “split” into percentage blocks for example 80–20%; the majority block is then used for training and the rest for validation purposes.

K-Fold

If data set is not so large, k-partitions of the data are made, for example, 5 partitions are made of which 4 are used for training the model whilst one is used for testing, The process is repeated till the whole set is exhausted. The mean and standard deviation of sensitivity and specificity over all the folds is estimated.

Leave One Out

This method is typically used when there are limited subjects with multiple trials such as in clinical trial studies. In such cases one can randomly keep one subject’s data out of training set and test on that subject and repeat the process till all subjects have been tested. This ensures minimum inter-subject variability. The mean and standard deviation of sensitivity and specificity over all the runs is then taken into consideration. Some of the commonly used metrics for validating an algorithm’s performance are:

F-Score

The sensitivity (true positive rate, recall, or probability of detection) of a machine-learning algorithm is defined as the number of positives that are actually defined as such. For example number of cats actually recognised as cats. Similarly, the specificity (true negative rate, precision) measures the proportion of actual negatives that are correctly identified as such. For examples number of dogs that were rejected as not being cats. The *F* score is the harmonic mean of sensitivity and specificity. It has a range of 0–1, where 1 means a perfect system. This is very effective in binary classification systems.

AUC of RoC

In an ROC (Receiver Operating Characteristic) curve the sensitivity is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter, such as tree depth for a decision tree. The area under

the ROC curve (AUC) is a measure of how well a parameter can distinguish between two classes.

Utility Function

This measure is used when an intersection of various features in a typically narrow band produces the ideal condition for a successful function. This is highly used in economics, where data are high dimensional and complex. This is also used in healthcare, where differential diagnosis is required for conditions have very general and overlapping symptoms such as GI tract infections, sepsis etc. Statistically speaking, If the data can be taped to real numbers, one can rank the data by ranking the real numbers and this mapping is called the utility function.

The Pros and Cons of Using AI in Medicine

Artificial Intelligence and machine learning have been used in many domains in medicine, the more publicised of these being oncology and cardiology. Jiang et al. surveyed the current status of AI in healthcare under four headings namely motivations for applying AI in healthcare, the data types which must be analysed by AI, the mechanisms needed for AI to produce clinically meaningful results, and the disease types being currently tackled by AI-based methods [20]. Obermeyer [21] identified areas of disruption that can occur in healthcare due to implementation of AI which are that ML will dramatically improve prognosis; however, the algorithms will need many years more of data acquisition before it can be sensitive and specific enough. Obermeyer predicts that machine learning will replace a lot of the work that pathologists and radiologists do today and reduce diagnostic error and bring about better accuracy. Reddy et al. identified four critical areas of maximum influence for implementation of AI in healthcare [22]. These are healthcare administration, clinical decision support, patient monitoring and healthcare interventions. We are today faced with very large volumes of data coming from the healthcare system and to make effective use of these data, we will need methods based on machine learning to help us understand and utilise the hidden and known correlations and connections in these data. In addition, AI and ML can reduce the load on over-worked clinicians by doing much of the documentation work required in a medical practice and also many of the routine and repetitive jobs.

There are a number of problems that are emerging with the gradual introduction of artificial intelligence-based methods in clinical medicine. Some of these are

(a) *Regulatory and legal* The FDA has defined steps to regulate the use of software as a medical device and

is in the process of setting up standards for the development, validation and monitoring of these solutions. The International Medical Device Regulators Forum has defined SaMD (Software as a Medical Device) as any software used for one or more medical purposes that perform these purposes without being part of a hardware medical device [23]. The FDA recognised the current medical device regulation was not designed for technologies such as ML and AI. It published a “Proposed regulatory framework for modifications to AI/ML-based software as a medical device” and sought public opinion the document. The comprehensive program proposes a pre-certification program, and a change control plan which is predetermined in the pre-market submissions itself. Transparency in the changes to the software and periodic updates are also part of the FDA’s proposed regulatory pathway [23–25]. Similar changes are occurring across regulatory bodies in other parts of the world too.

- (b) *Ethical and medico-legal contexts* We are seeing the use of ML in traditionally rule-based approaches such as safe drug prescription and scoring methods as well as clinical decision support such as survival estimates and prognosis and risk estimation [26]. There is a likely to be conflicting opinions on the medico-legal validity of decisions made with the support of AI-based systems. The justifications for use and of not using these systems, will need specific directions to be set into the methodology of development of algorithms, management of algorithms and re-training of the algorithm and end user, and clear instructions for usage.
- (c) *Distributional shift and black box decision making* The lack of adequate data as well as inappropriate sampling can substantially influence the performance and generalisability of algorithms. Overfitting data, spurious correlations, under-representation of populations, and the inevitable opacity of the decision making and output process (black box decision making) have all raised concerns about the universal applicability and generalisability of artificial intelligence and machine-learning-based decision support systems. Even with systems such as the IBM Watson oncology it has been pointed out that the system can perform better on commoner cases, wherein it is the uncommon case which the doctor demands help with the decision making process [27].

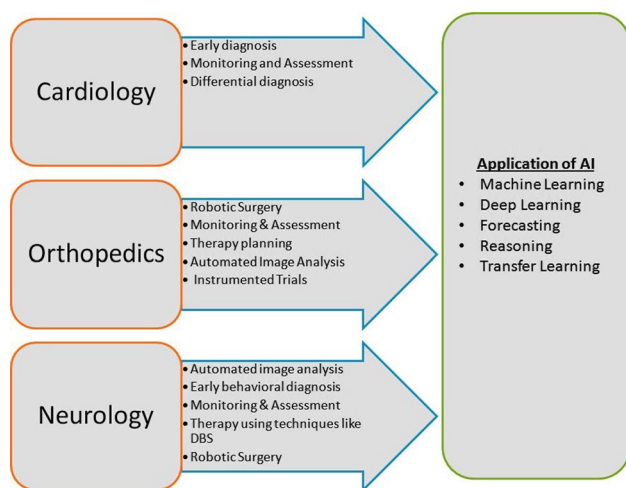


Fig. 6 Some applications of Machine Learning in healthcare

Machine-Learning Applications in the Field of Healthcare

The application of ML and AI has been extensively reported in the field of cardiology, neurology and oncology, as given in Fig. 6. In cardiology, application of machine-learning techniques has been found useful in prediction of coronary artery disease, in interpreting electrocardiograms, in interpreting echocardiograms and also in identifying phenotypes in a disease population [12, 14, 28–30].

Similarly, in neurology various ML-based algorithms [31] are being used to monitor the progression of diseases in neurodegenerative diseases such as Parkinson’s and Alzheimer’s. The disease in itself is known to fluctuate in its course and its response to drugs which is traditionally monitored using a diary maintained by the patient or his relatives. There has been substantial work involving accelerometer-based wearable sensor to monitor daily activity [32–34] of PD subjects to facilitate and fine tune medical therapy and rehabilitation, as well as prevent relapses, falls and complications.

In oncology, ML has found greater application [35–39]. With the greater understanding of cancers and the evolution of phenotypes and predictive biomarkers in targeted cancer therapy, ML has evolved as a powerful tool to sift through the varying types of data, link to real-world evidence, identify correlations and suggest clinical trials and therapies based on collective inputs and analysis. This has been demonstrated by the IBM Watson for Oncology system in many areas including breast cancer and gastric cancer. Watson can retrieve the most-applicable treatment plan based on tumour characteristics, overall health and preferences and link it to available evidence to support the choice [38]. Personalised and precision oncology hinges upon ML as the facilitating factor. Recently, AI-based applications

have been used to identify skin cancer and for identifying nodules in chest radiographs. These examples show how prevalent ML has been becoming in the healthcare industry. We consider it beyond our scope to discuss further in other medical domains. Instead we shall focus on some of the use cases for ML in our subject Orthopaedic surgery. Whilst we look at some of the work published in recent times, we aim to keep it regionwise anatomically for ease of understanding rather than divide it by methods used for machine learning.

Artificial Intelligence in Orthopaedic Surgery

“Will intelligent machines revolutionise orthopaedic imaging?” Asked Berg in an editorial in the *Acta Orthopaedica Scandinavia* in 2017 [40]. In the same issue Olzak et al. presented their research on applying ML to orthopaedic trauma radiographs with surprisingly good results comparable to radiologists [41]. Since then, the orthopaedic evidence base has seen the appearance of a number of studies utilising machine learning and artificial intelligence on databases ranging from imaging data to patient registries. Cabitza et al. reviewed the published literature on the subject of applications of machine learning in orthopaedic surgery [42]. They were able to identify 70 papers using either machine learning or deep learning as a methodology applied to clinical orthopaedics including fracture detection, spinal pathology assessments, skeletal bone age detection, shoulder strength assessment, gait classification, osteoarthritis prediction and detection, optimal injection point localisation, ACL/PCL detection and bone and cartilage image segmentation.

Kim et al. trained and validated ML models on a ACS-NSQIP (American College of Surgeons–National Surgical Quality Improvement Program) database to attempt to precisely predict mortality, venous thromboembolism, cardiac complications and wound complications following posterior lumbar fusion [43]. Both machine-learning models (ANN and LR) outperformed the ASA standard for producing each complication. The authors demonstrated the ability of using ML on a small data set to predict complications with low occurrences using appropriate and carefully applied techniques of machine learning. In another study Pereira et al. [44] used three methods, a classic scoring system, a nomogram-based method and a boosting algorithm (method of machine learning) to predict survival in metastatic spine disease. Survival was predicted better by the nomogram as compared to the classic scoring algorithm at 30 days, 90 days, and 365 days. Boosting algorithms were more accurate on sample data. However, on test data sets, it was slightly worse as compared to the nomogram. The researchers were also able to identify white cell count, haemoglobin and previous systemic therapy as three new factors associated with survival.

Jamaludin et al. applied deep learning techniques to reading T2 weighted sagittal lumbar MRI images, automating the identification of disc spaces, grading the degenerative changes such as spondylolisthesis and central canal stenosis and comparing them to what experienced radiologists would do [45]. The CNN-based model performed almost as well as experienced radiologists on the test data. The advantage of the deep learning model was that it did not need labeling and feature description, and with the addition of coronal and axial views the model could gain in accuracy and reliability. A distinct advantage is the avoidance of arbitrary scores. Though applied to T2 sagittal images here, it could easily be expanded to include the entire set of MRI scans [45].

Oncology has seen extensive application of deep learning and machine-learning techniques and orthopaedic oncology has been no exception. Recognising that purely image-based prediction for pathological fractures is inadequate, Oh et al. [46] used machine learning on CT imaging and clinical features to derive predictions for pathological femoral fractures in metastatic lung cancer and compared the model with one that used CT features alone. The machine-learning model that included clinical features showed superior predictive accuracy as compared to the model using CT features alone, thus reinforcing the ability of machine learning to use multivariate data and generate the best possible predictive path.

Survival estimates in patients with long bone metastases were studied using the application of a boosting algorithm on data from patients operated for long bone fractures and compared against a classic scoring system and a nomogram at 30 days, 90 days and 1 year time stamps [47]. The machine-learning algorithm proved superior in all training data sets, but in test data sets its performance was slightly less than the nomogram and the authors recommended the nomogram as simpler to use. Five year survival in chondrosarcoma was estimated applying the SORG (Skeletal Oncology Research Group) algorithm [48, 49]. Thio et al. [48] used data from the SEER data set (Survival Epidemiology and End Result) and applied machine-learning methods on demographics, tumour characteristics, treatment and outcome data. An application usable on a mobile phone, tablet or laptop with the outcome of interest being 5 year survival, was then deployed using the best performing Bayesian model. This was probably the first of its kind freely available online predictive tool. The algorithm was externally validated by Bongers et al. [49] who used institutional data from two tertiary-level institutions to validate the performance of the algorithm. They found the algorithm to systematically overestimate the survival in the institutional data set. However, the algorithm overestimated survival to a lesser extent on a smaller supplementary data set that had less than 5 year survival data available. Tools such as PathFx are available online to personalise bone cancer treatment. The ability of the PathFx tool to predict survival at several

points in patients undergoing surgery or palliative treatment for metastatic bone disease, using a multivariate tool modelled on Bayesian and Random Forrest techniques, has been tested on diverse patient populations with success [50–52]. The model predicted 1, 3, 6 and 12 month survival with 90% accuracy in a Japanese cohort (Asian) [51], it performed well in an Italian population when compared against the training data set (United States) and the first external validation (Scandinavian) [50, 51]. Nandra et al. used Bayesian belief networks to predict 1 year survival in bone sarcomas [53] and found them to be a useful decision support tool. These studies are reassuring and add strength to the premise that machine learning has potential in areas which enable both patient and doctor with wide spread implications in selecting appropriate treatment as well as avoiding inappropriate interventions.

In sports medicine, newer areas have emerged with the availability of wearables which can track the athletes' movements and physiology in real time. Along with the availability of large registry data, the potential to use machine-learning analytics to improve performance as well as pro-actively prevent injuries has been gaining ground [54]. The use of accelerometers, heart rate monitoring devices, RFID (Radio Frequency Identity) trackers, GPS (Global Positioning System) and camera-based motion-tracking systems devices in innovative ways helps determine baseline fitness, energy consumption, performance and quantification of motion. Applied to the available data on injuries and performance, analytics can drive the development of optimal training programs for Elite athletes as well as minimise the risk of injury and loss of play time [50].

Application of machine learning to automate reading of orthopaedic trauma radiographs may significantly reduce the load on emergency room physicians. The seminal paper by Olzak et al. [41] studied the use of artificial intelligence in analysing orthopaedic trauma radiographs and if it could be better than humans. Using a large database of hand wrist and ankle radiographs with associated radiology reports, and four identified outcomes of laterality, exam view, fracture and body part; five known deep learning networks were applied to the data-taking fracture to be the primary outcome and the others secondary. The performance of the model was compared against that of two senior orthopaedic surgeons on the same test data. All networks performed well reaching 99% accuracy when identifying body part, 90% on the laterality and 95% on the exam view but on detecting fractures the accuracy was greater with certain deeper networks reaching a maximum of 83%. In another study, a machine-learning algorithm was applied to the T2 weighted maps of the central medial femoral condyle using data from the Osteoarthritis Initiative [55]. The aim was to classify these cartilage maps and predict progression to clinically symptomatic osteoarthritis as evinced by a change in the

WOMAC (Western Ontario and MacMaster) score over 3 years. The authors found that the algorithm was able to classify the t2 weighted cartilage maps obtained before the onset of clinical osteoarthritis to predict the onset of osteoarthritis with 75% accuracy. Schmaranzer et al. developed a deep learning convolutional network to automate the 3-D segmentation of hip cartilage models in biochemical MRI of the hip done in symptomatic patients with structural hip deformities [56]. They found the fully automated method almost as good as the manual method and the indices generated in perfect concordance with two human observers.

Bevenino et al. developed a deep learning model to predict the likelihood of amputation in combat-related open calcaneal fractures and compared it to a standard logical regression model. They found the deep learning method 30% more accurate and better suited to clinical use than the logistical regression model [57]. In an interesting application of machine-learning methods, Menendez et al. applied machine-learning-based natural language processing to explore sentiment in negative patient comments following Total Shoulder Arthroplasty. They identified patient-related factors associated with negative comments and attempted to correlate them to peri-operative outcomes and traditional measures of patient satisfaction [58]. The data mined from the single institution single surgeon database were classified into four categories using machine-learning-based natural language processing into four groups, positive(62%), negative (32%), mixed (5%) and neutral. They found a common theme of room conditions followed by time management and pain management amongst others in the negative comments. This application presents interesting possibilities in the analysis of post surgical PROM (Patient-Related Outcome Measure) surveys in determining quality and satisfaction after orthopaedic surgery.

In total joint arthroplasty a number of recent papers have explored the application of machine-learning methods. Fontana et al. applied three different supervised machine-learning models to hospital registry data to predict which patients would achieve a less than minimal clinically important difference (MCID) in four PROMs 2 years after total joint arthroplasty [59]. They also sought to identify how the predictive ability changed with the addition of more information and which variables affected the predictive ability of the models [59]. They incrementally considered predictors before the decision to undergo surgery, before surgery, before discharge and after discharge and evaluated model performance on a test data set comprising 25% of the data set excluded from the modeling. They reported fair to good performance on pre-surgical data and found that machine learning has good predictive power in predicting MCID including before surgical decision and before surgery data, and that this predictive power did not change significantly if surgical

and post-surgical data were included as well. The value of this model in planning post surgical monitoring and rehabilitation is relevant and more such studies validated on diverse populations would help develop finer models. Harris et al. explored the premise whether machine learning could provide simple easy to use tools to predict 30 day mortality and morbidity after total joint arthroplasty [60, 61]. The internal validation was most accurate for cardiac complications and mortality [60]. On further validation studies [61], they were able to develop fairly accurate models predicting mortality and cardiac complications but not the rarer complications such as re-operation and deep infection. They attributed this to the elective nature of surgery, where patients are pre-optimised already, dichotomy of several patient data, intra-operative and postoperative events that cause complications and are not part of the model, as well as variables that are not easily incorporated into the model [60]. Recent papers have applied machine learning on pre-operative hospital data to predict inpatient stays and patient specific payments for inpatient care with the objective of creating a risk adjusted payment model for total hip and knee arthroplasty [62, 63]. These models showed excellent predictability of length of stay with the application of naive Bayesian algorithms using basic pre-operative co-morbidity data, but as complexity of the case increased, accuracy for predicting payment decreased proportionately in THA, whereas in TKA, the proportionate predicted costs increased by 3, 10 and 15% for moderate, severe and extreme risk populations. The Cleveland Clinic Group spoke about the establishment of a machine-learning arthroplasty lab recognising that machine-learning algorithms are the best way for surgeons who want to make the best use of data for optimising patient and healthcare outcomes [64]. The authors used machine learning at their institution to establish patient specific risk adjusted payment models. Taking it one step further the authors have used a knee sleeve to monitor step count, range of motion, exercise plan compliance, activity level and opioid use. This motivational aid is used to capture data for future analysis [64]. The use of machine learning in conjunction with finite element modelling techniques in an attempt to optimise the short stem femoral implant, to minimise stress shielding and optimise function was described by Stojadinovic et al. [65] opening up new avenues to look at intelligent design of implants.

Other areas that have been explored with machine learning include prediction of non-unions [66], and gait pattern prediction and analysis [67–69].

As we see from the few examples above, the possibilities are limitless and we are only seeing the tip of the iceberg. Whilst we write and read this, many more approaches are being tried out in the field of orthopaedics. We have been using logistical regression for our models for many

years, especially those that predict risk, survival, mortality and morbidity. We feel that these are areas which will show promise with ML and AI applications in the near term.

Discussion

As we enter an exciting age of AI and robotics, it has been said it is a brave new world [70]. As surgeons, we inherently believe in value derived from patient outcomes, surgical innovations, implant designs and best practices in the field [71]. The precision that is promised by AI in our ability to deliver optimised care is indeed something to look forward to. Whether it be survival, prediction of costs, assisting in image diagnostics, clinical decision support or even implant design and improvement, the avenues we can see are tremendous and varied. We are now looking at artificial intelligence-based technologies sitting at the top of the Gartner Hype Curve [72]. However, it is quite unlikely to fall into the trough of disillusionment and then the plateau of enlightenment followed by productivity [73]. We can anticipate increasing integration of these technologies into the workplace driven by the need for value of care and patient entered outcome evaluation. The most valuable area that is developing is image analysis, where AI is showing promising results in reading X-rays and other imaging data. Clinical decision support on the strength of analysis of varied types of data such as imaging, EMR and EHR data and treatment documents with the aid of deep learning and natural language processing, is already showing promising results. The papers from Cleveland clinic [62–64] have shown how machine learning can help predict stays and develop risk adjusted payment models. These are huge strides forward in our quest for optimised care at reasonable costs and with reduced complications. Enabled with masses of wearable data, we can envision a future that is wearable enabled and data driven to provide precision and personalised treatment for our patients. As an editorial comment in the Journal of Arthroplasty recently pointed out, whilst our patients may demand the same degree of ease and convenience and personalisation from their medical treatment, as they are experiencing in their personal lives, they also realise the situation is different, where their health is at stake [74]. There is also a fear that machines will overrun the doctors. This fear though rampant, is at the moment at least, ill founded. The kind of artificial intelligence that is needed for this does not exist and is still decades away at the earliest. As Obermeyer [21] has said, medical practice has always required doctors to handle huge volumes of data, and the ability of doctors to handle this increasingly complex data, sets good doctors apart from the mediocre ones. ML provides doctors the unique opportunity to understand their patient better and to use the best option [21]. Clinicians must train themselves

to use these methods effectively and improve their practice. There is no doubt that AI is all set to replace much of the diagnostic work and in some years, may even become the standard of care. Hence an ethical, moral and legal framework needs to be in place for the development, implementation and maintenance and upgrade of these algorithms. AI can also be misdirected by bias and inherent inability to translate features and relations from a narrower database to a larger population [75]. What does the future really hold? Does it envisage machines replacing doctors? Not in the near future, it does not look like it will, although we see great amounts of automation in the way we work. What we will see, however, is that we will be submerged in huge mountains of data in this increasingly connected world and workplace. We will have to face the reams of EMR and EHR data, wearable-based monitoring data, app-based patient outcome data, imaging data, surgical videos and procedural data, literature and multiple complex volumes of imaging data, which we already are finding it difficult to handle and interpret. In such a developing workplace, we will see the gradual intrusion and permeation of AI and ML, helping us sift through data, find correlations, interpret and conclude from the data. We will find algorithms simplify the paperwork we need to do to administer our work and practice and payments. We will see algorithms filtering out those patients who need the most attention and directing our interest the right way. Many more ways we can list how algorithms will enrich the healthcare industry. The downsides we have listed already in brief and they themselves would serve a separate paper, but it seems reasonable to conclude, as many more have before, and many who continue to do so, to reassure the skeptics, and embrace this brave new universe. Doctors need to play an active and interactive role with engineers in developing, tailoring, implementing and managing algorithms in this domain. We need doctors to take responsibility and train algorithms and interpret the validity and usage of algorithms before they are released into the practice domain. We are already seeing this synergy. We need a note of caution too, whilst we may consider medicine to be a rule based, evidence based, rational activity based on well-defined conditions, in reality it involves a lot more than that in actual practice [76]. There is reasoning, there are values, empathy, relationships, advice and reassurances. There is experiential learning, intuitive responses based on real time and real-world understanding of the environment in which the patient is living and working in, which will be difficult to incorporate at the current time in ML technologies. It is for us to reason and understand together the best directions and applications that ML can bring us to improve what we do best, care for the patient.

Mr Nadella, the CEO of Microsoft has laid down some principles and goals for AI which we as an industry and society need to debate on. These apply as much to our domain of

patient care as to other domains in the real world. In brief, these are (a) AI must be designed to assist humanity (b) AI must be transparent -one must know how it works, men must know about the machines, ethics and design must go hand in hand c) Maximise effectiveness without destroying the dignity of people—the tech industry should not dictate the values or virtues of the future that should preserve cultural commitments and diversity. (d) AI must be designed for intelligent privacy. (e) Design AI for algorithmic responsibility so that humans can undo unintended damage. (f) AI must avoid bias, one needs to ensure proper representative research to avoid bias. Mr Nadella also talked about the characteristics that humans need to develop to be able to stay relevant in the age of AI [1, 9]. These include (a) Empathy—this is difficult to replicate in machines and will be valuable in the human-AI world. (b) Education; One will need knowledge and skills to implement new technologies on a large scale. For us in the field of medicine, it will mean a change in the basic medical curriculum which will need to incorporate the knowledge of using algorithms intuitively in their practice. (c) Creativity—the enhanced capabilities provided by machines will continue to augment and improve our capabilities.; and last but not the least (d) Judgement and responsibility—to accept that a decision made by a machine still means a human has to be ultimately responsible [1, 9].

Conclusion

The world of algorithms brings with it a lot of expectations and also apprehensions and fears. AI and ML have demonstrated their efficacy in well selected and conducted examples, and the utility of these algorithms to augment diagnostics and clinical care is slowly getting well established. In orthopaedics, prognostication of outcomes, prediction of costs and optimisation of care, image analysis, surgical implant design, survival analysis are all areas being looked into. We can expect the technology to spread rapidly and more insights to emerge especially from the large and long running implant registries in Europe and North America. We can also expect insights and changes in personalised orthopaedic care on the basis of patterns derived by deep learning algorithms from EMR and EHR data. In short, there are exciting times ahead and the way we practice is set to change, and we need to get prepared well by training ourselves and our colleagues, participating in technology development and using it well to augment our clinical practice and patient care. In this, AI is truly a positive and welcome disrupting force in orthopaedic surgery.

Author contributions Concepts: MP, AG, SM, VB, and AS. Design: MP, AG, SM, VB, and AS. Definition of intellectual content: MP, AG,

SM, VB, and AS. Literature search: MP, AG, SM, VB, and AS. Clinical studies: MP, AG, SM, VB, and AS. Experimental studies: not available. Data acquisition: none. Data analysis: none. Statistical analysis: not available. Manuscript preparation: MP, AG, SM, VB, and AS. Manuscript editing: MP, AG, SM, VB, and AS. Manuscript review: MP, AG, SM, VB, and AS. Guarantor: MP, AG, SM, VB, and AS.

Compliance with Ethical Standards

Conflict of interest The authors declare no conflict of interests in relation to the content published here which is entirely educative and informative.

References

- Nadella, S. (2016). The partnership of the future, SLATE: June 28 2016. <https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-society-challenges.html>. Accessed: 19 Jun 2019.
- Turing, A. M. (2019). Computing Machinery and Intelligence, *Mind*, New Series, Vol. 59, No. 236 (Oct., 1950), pp. 433–460 Published by: Oxford University Press on behalf of the Mind Association available at: <http://www.jstor.org/stable/2251299>. Accessed: 19 Jun 2019.
- Domingo, P. (2017). The machine learning revolution. In *The master algorithm: How the quest for the ultimate learning machine will remake our world* (pp. 1–22). UK: Penguin Random House.
- Woodson, J. (2019). Decades Ago, Pilots Learned to “Fly by Instruments.” Doctors Need to Do the Same [Internet]. *Harvard Business Review*. 2019 [cited 23 June 2019]. <https://hbr.org/2018/03/decades-ago-pilots-learned-to-fly-by-instrument-s-doctors-need-to-do-the-same>.
- McCarthy, J., Marvin, L., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. August 31, 1955. <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>. Accessed June 23 2019.
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., & Hager, G., et al. (2019). Artificial Intelligence and Life in 2030.”One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA, September 2016. <http://ai100.stanford.edu/2016-report>. Accessed June 23 2019.
- Hintz A. (2019). Understanding the four types of AI, from reactive robots to self-aware beings [Internet]. *The Conversation*. 2019 [cited 28 July 2019]. Available from: <https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616>. Accessed 28 July 2018.
- Cool vendors in healthcare artificial intelligence. <https://www.gartner.com/document/3913322> Accessed June 23 2017
- González, G. C., Núñez-Valdez, E., García-Díaz, V., Pelayo, G., Bustelo, C., & Cueva-Lovelle, J. (2019). A Review of Artificial Intelligence in the Internet of Things. *International Journal of Interactive Multimedia and Artificial Intelligence*, 5(4), 9.
- McCarthy, J. (2019). What is AI? <http://jmc.stanford.edu/articles/whatisai/whatisai.pdf>. Accessed 23 June 2019.
- Russel, S. J., & Norvig, P. (2015). Introduction. In *Artificial intelligence: A modern approach* (3rd ed., pp. 1–3). New Delhi: Pearson India Education Services Pvt Ltd.
- Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., et al. (2018). Artificial intelligence in cardiology. *Journal of the American College of Cardiology*, 71(23), 2668–2679.

13. Russel, S. J., & Norvig, P. (2015). Learning from examples. *Artificial intelligence: A modern approach* (3rd ed., pp. 706–781). New Delhi: Pearson India Education Services Pvt Ltd.
14. Deo, R. C. (2015). Machine learning in medicine. *Circulation*, *132*(20), 1920–1930.
15. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., et al. (2018). *NPJ Digit Med.*, *1*, 18.
16. Colton, S., & Mentor, F. R. C. (2007). “The balance filter.” Presentation, Massachusetts Institute of Technology (2007). http://dl.amobbs.com/bbs_upload782111/files_44/ourdev_665531S2JZG6.pdf. Accessed 14 Dec 2019.
17. Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research.*, *16*, 321–357.
18. Mundra, P., & Rajapakse, J. (2010). SVM-RFE with MRMR filter for gene selection. *IEEE Transactions on NanoBioscience.*, *9*(1), 31–37.
19. Pai-shun, T., Chun-Chen, T., Pin-Yu, C., Ya-Yun, L., & Shin-Ming, C. (2019). FEAST: An automated feature selection framework for compilation tasks. [Internet]. Arxiv.org. Arxiv 2016 [cited 28 July 2019]. Available from: <https://arxiv.org/pdf/1610.09543.pdf>.
20. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*, *2*(4), 230–243.
21. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *New England Journal of Medicine*, *375*(13), 1216–1219.
22. Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, *112*(1), 22–28.
23. “Software as a medical device (SaMD)”. (2019). <https://www.fda.gov/medical-devices/digital-health/software-medical-device-samd>. Accessed 28 June 2019.
24. “Artificial Intelligence and machine learning in SaMD”. (2019). <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device>. Accessed 28 June 2019.
25. Developing Software Pre-certification program: A Working Model”. (2019). <https://www.fda.gov/media/113802/download>. Accessed 28 June 2019.
26. Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Qual Saf*, *28*(3), 231–237.
27. Keikes, L., Medlock, S., van de Berg, D. J., Zhang, S., Guicherit, O. R., Punt, C. J. A., et al. (2018). The first steps in the evaluation of a “black-box” decision support tool: a protocol and feasibility study for the evaluation of Watson for Oncology. *Journal Of Clinical and Translational Research*, *3*(Suppl 3), 411–423.
28. Sharma, S., & Seth, U. (2017). Artificial intelligence in cardiology. *Journal of the Practice of Cardiovascular Sciences*, *3*(3), 158.
29. Bonderman, D. (2017). Artificial intelligence in cardiology. *Wiener Klinische Wochenschrift*, *129*(23–24), 866–868.
30. Tajik, A. J. (2016). machine learning for echocardiographic imaging: embarking on another incredible journey. *Journal of the American College of Cardiology*, *68*(21), 2296–2298.
31. Dijkstra, B., Zijlstra, W., Scherder, E., & Kamsma, Y. (2008). Detection of walking periods and number of steps in older adults and patients with Parkinson’s disease: accuracy of a pedometer and an accelerometry-based method. *Age and Ageing*, *37*(4), 436–441.
32. Herman, T., Weiss, A., Brozgol, M., Giladi, N., & Hausdorff, J. M. (2014). Gait and balance in Parkinson’s disease subtypes: objective measures and classification considerations. *Journal of Neurology*, *261*(12), 2401–2410.
33. Weiss, A., Herman, T., Giladi, N., & Hausdorff, J. M. (2015). New evidence for gait abnormalities among Parkinson’s disease patients who suffer from freezing of gait: insights using a body-fixed sensor worn for 3 days. *Journal of Neural Transmission (Vienna)*, *122*(3), 403–410.
34. Rakhim, P., & Lan, K. C. (2016). Gait monitoring for early neurological disorder detection using sensors in a smartphone: validation and a case study of Parkinsonism. *Telemedicine and e-Health*, *22*(1), 75–81.
35. Azuaje, F. (2019). Artificial intelligence for precision oncology: beyond patient stratification. *NPJ Precision Oncology*, *3*, 6.
36. Curioni-Fontecedro, A. (2017). A new era of oncology through artificial intelligence. *ESMO Open*, *2*(2), e000198.
37. Kim, Y. Y., Oh, S. J., Chun, Y. S., Lee, W. K., & Park, H. K. (2018). Gene expression assay and Watson for Oncology for optimization of treatment in ER-positive, HER2-negative breast cancer. *PLoS One*, *13*(7), e0200100.
38. Malin, J. L. (2013). Envisioning Watson as a rapid-learning system for oncology. *Journal of Oncology Practice*, *9*(3), 155–157.
39. Choi, Y. I., Chung, J. W., Kim, K. O., Kwon, K. A., Kim, Y. J., Park, D. K., et al. (2019). Concordance rate between clinicians and watson for oncology among patients with advanced gastric cancer: Early, real-world experience in Korea. *Canadian Journal of Gastroenterology and Hepatology*, *2019*, 8072928.
40. Berg, H. E. (2017). Will intelligent machine learning revolutionize orthopedic imaging? *Acta Orthopaedica*, *88*(6), 577.
41. Olczak, J., Fahlberg, N., Maki, A., Razavian, A. S., Jilert, A., Stark, A., et al. (2017). Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthopaedica*, *88*(6), 581–586.
42. Cabitza, F., Locoro, A., & Banfi, G. (2018). Machine learning in orthopedics: A literature review. *Frontiers in Bioengineering and Biotechnology*, *6*, 75.
43. Kim, J. S., Merrill, R. K., Arvind, V., Kaji, D., Pasik, S. D., Nwachukwu, C. C., et al. (2018). Examining the ability of artificial neural networks machine learning models to accurately predict complications following posterior lumbar spine fusion. *Spine (Phila Pa 1976)*, *43*(12), 853–860.
44. Paulino Pereira, N. R., Janssen, S. J., van Dijk, E., Harris, M. B., Hornicek, F. J., Ferrone, M. L., et al. (2016). Development of a prognostic survival algorithm for patients with metastatic spine disease. *Journal of Bone and Joint Surgery American*, *98*(21), 1767–1776.
45. Jamaludin, A., Lootus, M., Kadir, T., Zisserman, A., Urban, J., Battié, M. C., et al. (2017). ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: Automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *European Spine Journal*, *26*(5), 1374–1383.
46. Oh, E., Seo, S. W., Yoon, Y. C., Kim, D. W., Kwon, S., & Yoon, S. (2017). Prediction of pathologic femoral fractures in patients with lung cancer using machine learning algorithms: Comparison of computed tomography-based radiological features with clinical features versus without clinical features. *Journal of Orthopaedic Surgery (Hong Kong)*, *25*(2), 2309499017716243.
47. Janssen, S. J., van der Heijden, A. S., van Dijke, M., Ready, J. E., Raskin, K. A., Ferrone, M. L., et al. (2015). 2015 Marshall urist young investigator award: Prognostication in patients with long bone metastases: Does a boosting algorithm improve survival estimates? *Clinical Orthopaedics and Related Research*, *473*(10), 3112–3121.
48. Thio, Q. C. B. S., Karhade, A. V., Ogink, P. T., Raskin, K. A., De Amorim, Bernstein K., Lozano Calderon, S. A., et al. (2018). Can machine-learning techniques be used for 5-year survival prediction of patients with chondrosarcoma? *Clinical Orthopaedics and Related Research*, *476*(10), 2040–2048.

49. Bongers, M. E. R., Thio, Q. C. B. S., Karhade, A. V., Storm, M. L., Raskin, K. A., Lozano Calderon, S. A., et al. (2019). Does the SORG Algorithm predict 5-year survival in patients with chondrosarcoma? An external validation. *Clinical Orthopaedics and Related Research*, 477, 2296–2303.
50. Piccioli, A., Spinelli, M. S., Forsberg, J. A., Wedin, R., Healey, J. H., Ippolito, V., et al. (2015). How do we estimate survival? External validation of a tool for survival estimation in patients with metastatic bone disease-decision analysis and comparison of three international patient populations. *BMC Cancer*, 22(15), 424.
51. Ogura, K., Gokita, T., Shinoda, Y., Kawano, H., Takagi, T., Ae, K., et al. (2017). Can a multivariate model for survival estimation in skeletal metastases (PATHFx) be externally validated using Japanese patients? *Clinical Orthopaedics and Related Research*, 475(9), 2263–2270.
52. Forsberg, J. A., Wedin, R., Boland, P. J., & Healey, J. H. (2017). Can we estimate short- and intermediate-term survival in patients undergoing surgery for metastatic bone disease? *Clinical Orthopaedics and Related Research*, 475(4), 1252–1261.
53. Nandra, R., Parry, M., Forsberg, J., & Grimer, R. (2017). Can a Bayesian belief network be used to estimate 1-year survival in patients with bone sarcomas? *Clinical Orthopaedics and Related Research*, 475(6), 1681–1689.
54. Sikka, R. S., Baer, M., Raja, A., Stuart, M., & Tompkins, M. (2019). Analytics in sports medicine: Implications and responsibilities that accompany the era of big data. *Journal of Bone and Joint Surgery American*, 101(3), 276–283.
55. Ashinsky, B. G., Bouhrara, M., Coletta, C. E., Lehallier, B., Urish, K. L., Lin, P. C., et al. (2017). Predicting early symptomatic osteoarthritis in the human knee using machine learning classification of magnetic resonance images from the osteoarthritis initiative. *Journal of Orthopaedic Research*, 35(10), 2243–2250.
56. Schmaranzer, F., Helfenstein, R., Zeng, G., Lerch, T. D., Novais, E. N., Wylie, J. D., et al. (2019). Automatic MRI-based three-dimensional models of hip cartilage provide improved morphologic and biochemical analysis. *Clinical Orthopaedics and Related Research*, 477(5), 1036–1052.
57. Bevevino, A. J., Dickens, J. F., Potter, B. K., Dworak, T., Gordon, W., & Forsberg, J. A. (2014). A model to predict limb salvage in severe combat-related open calcaneus fractures. *Clinical Orthopaedics and Related Research*, 472(10), 3002–3009.
58. Menendez, M. E., Shaker, J., Lawler, S. M., Ring, D., & Jawa, A. (2019). Negative patient-experience comments after total shoulder arthroplasty. *Journal of Bone and Joint Surgery American*, 101(4), 330–337.
59. Fontana, M. A., Lyman, S., Sarker, G. K., Padgett, D. E., & MacLean, C. H. (2019). Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clinical Orthopaedics and Related Research*, 477(6), 1267–1279.
60. Harris, A. H. S., Kuo, A. C., Weng, Y., Trickey, A. W., Bowe, T., & Giori, N. J. (2019). Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? *Clinical Orthopaedics and Related Research*, 477(2), 452–460.
61. Harris, A. H., Kuo, A. C., Bowe, T., Gupta, S., Nordin, D., & Giori, N. J. (2018). Prediction models for 30-day mortality and complications after total knee and hip arthroplasties for veteran health administration patients with osteoarthritis. *Journal of Arthroplasty*, 33(5), 1539–1545.
62. Ramkumar, P. N., Navarro, S. M., Haeberle, H. S., Karnuta, J. M., Mont, M. A., Iannotti, J. P., et al. (2019). Development and validation of a machine learning algorithm after primary total hip arthroplasty: Applications to length of stay and payment models. *Journal of Arthroplasty*, 34(4), 632–637.
63. Navarro, S. M., Wang, E. Y., Haeberle, H. S., Mont, M. A., Krebs, V. E., Patterson, B. M., et al. (2018). Machine learning and primary total knee arthroplasty: Patient forecasting for a patient-specific payment model. *Journal of Arthroplasty*, 33(12), 3617–3623.
64. Ramkumar, P. N., Haeberle, H. S., Bloomfield, M. R., Schaffer, J. L., Kamath, A. F., Patterson, B. M., et al. (2019). Artificial intelligence and arthroplasty at a single institution: Real-world applications of machine learning to big data, value-based care, mobile health, and remote patient monitoring. *The Journal of Arthroplasty*. <https://doi.org/10.1016/j.arth.2019.06.018>.
65. Cilla, M., Borgiani, E., Martínez, J., Duda, G. N., & Checa, S. (2017). Machine learning techniques for the optimization of joint replacements: Application to a short-stem hip implant. *PLoS One*, 12(9), e0183755.
66. Stojadinovic, A., Kyle Potter, B., Eberhardt, J., Shawen, S. B., Andersen, R. C., Forsberg, J. A., et al. (2011). Development of a prognostic naive Bayesian classifier for successful treatment of nonunions. *Journal of Bone and Joint Surgery. American Volume*, 93(2), 187–194.
67. Begg, R., & Kamruzzaman, J. (2005). A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics*, 38(3), 401–408.
68. Joyseeree, R., Abou Sabha, R., & Mueller, H. (2015). Applying machine learning to gait analysis data for disease identification. *Studies in Health Technology and Informatics*, 210, 850–854.
69. Sayed, M. (2018). Biometric gait recognition based on machine learning algorithms. *Journal of Computer Science*, 14(7), 1064–1073.
70. Parsley, B. S. (2018). Robotics in orthopedics: A brave new world. *Journal of Arthroplasty*, 33(8), 2355–2357.
71. Levy, J. C. (2019). Don't lose sight of the outcome: Commentary on an article by Mariano E. Menendez, MD, et al.: "Negative Patient-Experience Comments After Total Shoulder Arthroplasty". *J Bone Joint Surg Am*, 101(4), e15.
72. Trends emerge in the Gartner Hype cycle for emerging technologies. (2018). <https://www.gartner.com/smarterwithgartner/5-trends-emerge-in-gartner-hype-cycle-for-emerging-technologies-2018/>. Accessed 21 July 2018.
73. Bini, S. A. (2018). Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *Journal of Arthroplasty*, 33(8), 2358–2361.
74. Froimson, M. I. (2018). Digital Health and advanced technology in arthroplasty. *Journal of Arthroplasty*, 33(8), 2344.
75. Loh, E. (2018). Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health. *BMJ Leader*, 2(2), 59–63.
76. Jones, L. D., Golan, D., Hanna, S. A., & Ramachandran, M. (2018). Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern? *Bone Joint Res*, 7(3), 223–225.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.