1-1-2022

A review of arthritis diagnosis techniques in artificial intelligence era: Current trends and research challenges

Maleeha Imtiaz
Edith Cowan University

Syed Afaq Ali Shah
Edith Cowan University, afaq.shah@ecu.edu.au

Zia ur Rehman

10.1016/j.neuri.2022.100079
Imtiaz, M., Shah, S. A. A., & ur Rahman, Z. (2022). A review of arthritis diagnosis techniques in artificial intelligence era: Current trends and research challenges. Neuroscience Informatics, 2, Article 100079. https://doi.org/10.1016/j.neuri.2022.100079

Neuroscience Informatics 2 (2022) 100079

Contents lists available at ScienceDirect

Neuroscience Informatics

www.elsevier.com/locate/neuri

Artificial Intelligence in Brain Informatics

A review of arthritis diagnosis techniques in artificial intelligence era: Current trends and research challenges

Maleeha Imtiaz [a], Syed Afaq Ali Shah [b,*], Zia ur Rehman [c]

[a] *Joondalup Health Campus, Australia*
[b] *Centre for AI and Machine Learning, Edith Cowan University, Australia*
[c] *Cairns Hospital, Cairns, QLD, Australia*

ARTICLE INFO

Article history:
Received 2 January 2022
Received in revised form 4 May 2022
Accepted 12 May 2022

Keywords:
Machine learning (ML)
Deep learning
Osteoarthritis
Rheumatoid arthritis

ABSTRACT

Deep learning, a branch of artificial intelligence, has achieved unprecedented performance in several domains including medicine to assist with efficient diagnosis of diseases, prediction of disease progression and pre-screening step for physicians. Due to its significant breakthroughs, deep learning is now being used for the diagnosis of arthritis, which is a chronic disease affecting young to aged population. This paper provides a survey of recent and the most representative deep learning techniques (published between 2018 to 2020) for the diagnosis of osteoarthritis and rheumatoid arthritis. The paper also reviews traditional machine learning methods (published 2015 onward) and their application for the diagnosis of these diseases. The paper identifies open problems and research gaps. We believe that deep learning can assist general practitioners and consultants to predict the course of the disease, make treatment propositions and appraise their potential benefits.

1. Introduction

Arthritis is a term which is used for various inflammatory conditions that affect different parts of the body such as joints, bones, and muscles. It can be of several types such as Osteoarthritis (OA), Rheumatoid Arthritis (RA), juvenile Arthritis, psoriatic arthritis, and gouty Arthritis, which can result in stiffness, pain, redness and swelling in the joints [47]. According to [5], it has been revealed that about 3.6 million (15%) of people are affected from arthritis which includes 17.9% females and 12.1% males. Moreover, 62% of patients affected from arthritis had Osteoarthritis, 12.7% had rheumatoid arthritis, and 32.1% had suffered from an unspecified form of arthritis. One in every seven Australians has Arthritis [6]. The prevalence of arthritis rises with age, primarily affecting the females (ABS, 2017). Moreover, higher mortality risk is also recorded in patients with rheumatoid arthritis (RA) as compared to the general population [22], [52].

Rheumatic diseases are chronic and fluctuating in nature, involving complicated and unclear etiology, which further intricates the treatment of this kind of arthritis [12], [57], [52]. Regardless, even from the invention of various biological and synthetic treatments for rheumatoid arthritis (RA), the decrease in disease progression is achieved only in a small subset of patients [23], [36]. Moreover, the clinical experiments for another rheumatic disease that is Osteoarthritis (OA) are not very fruitful due to different disease phenotypes involved in the disease. Therefore, the disease diagnosis at an early stage can slow down its progression, where diagnosis involves numerous imaging modalities such as X-rays, MRI and CT. However, diagnosis techniques, such as Kellgren-Lawrence (KL) grade suffer from subjectivity, as their accuracy heavily depends on the practitioner's experience [65]. Table 1 provides details of Kellgren and Lawrence grading for completion. In order to make the diagnosis process more systematic and reliable, computer-aided analysis and predictive modelling is required to overcome the human errors and for early disease detection in places where there are fewer experts available.

In addition, to advent an appropriate treatment for arthritis, a data-intensive investigation is essential where artificial intelligence (AI) can play a significant role in disease detection. Exclusively, machine learning (ML), a subfield of AI aims to design data-driven predictive models which possess the ability to learn from the experience regardless of the rules explicitly specified by individuals [23]. It uses methods, algorithms and processes to expose concealed associations within data and to produce prescriptive, descriptive and predictive tools in order to exploit these associations [27]. Additionally, the advancement of Machine Learning principles and Artificial Intelligence techniques has increased the productiv-

* Corresponding author.
E-mail address: afaq.shah@ecu.edu.au (S.A.A. Shah).

https://doi.org/10.1016/j.neuri.2022.100079

Table 1
Kellgren and Lawrence (KL) grading.

KL grade	Diagnosis
0	No features of osteophytes are present
1	Narrowing of joint space, doubtful OA
2	Certain narrowing of joint space, minor OA
3	Multiple osteophytes, sure joint space narrowing and some sclerotic areas, moderate OA
4	Large osteophytes, severe joint space narrowing, severe sclerosis and bone deformity, Severe OA

ity and effectiveness in medical imaging research [55]. Machine learning concepts, when applied to medical data, have great potential to improve disease diagnosis and early detection of diseases [11], [48], [57], [8]. In clinical settings, these techniques can help medical experts to analyse the disease in a better way to predict potential future issues and treat patients more effectively.

Machine learning algorithms are capable of learning useful data representations automatically [40], [50]. They can deal with a variety of data inputs such as genetic information, text e.g., electronic health records, patient cohorts and medical images. Furthermore, it can also learn from the knowledge available from clinical data and generate outcomes by recognising disease patterns, and features. Further, it can also help in optimising treatment strategies. Hence, it is quite evident that ML has helped in significantly filling the gap of automatic learning from clinical experience. Furthermore, Deep learning (DL), is a subfield of ML, which utilises multi-layered neural networks, intensive computational algorithms and big data [23], [46]. Over the last decade, both ML and DL have been used in the field of medicine for medical imaging, and it has been depicted that ML-based decision-making is superior to physicians' individual clinical trial decisions [23].

Inspired by the recent advent of artificial intelligence in medical field, this paper presents a survey of deep learning and traditional machine learning techniques for the diagnosis of osteoarthritis and rheumatoid arthritis. The paper also aims to identify the current challenges and open research problems in this area. In contrast to current review papers [23], [55], [24], [28] which mainly focus on a specific type of arthritis e.g., OA or RA and machine learning techniques only, this paper reviews deep learning as well as machine learning methods for the diagnosis of both OA and RA. In addition, this paper also provides detailed information about the publicly available datasets for RA and OA research (Section 4.2). This makes our survey paper different from the existing review articles.

The rest of the paper is organised as follows. Section 2 discusses most common arthritis types. Overview of the most popular machine and deep learning techniques is presented in Section 3. An overview of imaging techniques and arthritis datasets is provided in Section 4. Machine learning and deep learning approaches for the diagnosis of arthritis are presented in Section 5 and 6, respectively. Section 7 discusses some of the open research problems and research challenges. The paper is concluded in Section 8.

2. Arthritis and its types

Arthritis is a degenerative disorder associated with human joints that can result in disability. There are numerous types of arthritis such as rheumatoid arthritis, Osteoarthritis, Juvenile Arthritis, Psoriatic arthritis and gout arthritis. In the following, we will briefly discuss rheumatoid arthritis, osteoarthritis and Psoriatic arthritis.

2.1. Rheumatoid arthritis

Rheumatoid Arthritis (RA) is an autoimmune inflammatory disorder which involves multiple organs affecting one or more joints [27]. It is a disease with unclear etiology and a combination of genetic and environmental factors. The complex interactions of these factors affect disease development and progression [29]. In general, RA is categorised through morning stiffness and inflammation of joints that requires skills and experience for proper diagnosis of disease. In 1987, ACR (American College of Rheumatology) established a standard for diagnosis of rheumatoid arthritis based on morning stiffness, swelling of joints, but that was not appropriate for early disease analysis. Later in 2010, ACR/EULAR proposed a new criterion to make an early prediction of rheumatic patients [29], as early detection and treatment of rheumatoid arthritis can slow down the disease progression and also increase the chances of cure.

2.2. Osteoarthritis

Osteoarthritis (OA) is one of the most common musculoskeletal conditions which can result in significant disability among patients. Knee OA is ranked as the 11th highest cause of disability worldwide [65]. OA is a disorder which can cause articular cartilage breakdown. Cartilage is a smooth, steady layer that ensures an effortless movement of knee joints. In OA, cartilage is erupted, lose elasticity and becomes feeble [17]. In general, it affects the joints in the knee, hip, spine, and feet. According to [15], firstly, OA can occur either due to genetic reasons or ageing factors. Secondly, OA can be seen in early stages of life due to some injury, diabetes, obesity, athletics or in rheumatoid arthritis patients. The primary symptoms of OA include joint pain and trouble in joint movement, joint stiffness during the morning or after a long rest [16]. Usually, due to unclear etiology OA is not diagnosed at a later stage for an effective treatment, and sometimes expensive and invasive joint replacement surgery is required [65]. However, early detection of disease can slow down its progression. The current OA evaluation of OA is based on the combination of clinical examination, symptoms and radiographic assessment techniques such as X-ray, MRI, and CT as required [62]. Besides various proposed methods of OA diagnosis, Kellgren-Lawrence (KL) grading system is a gold clinical standard for classifying individual joints into five grades based on OA severity [17].

2.3. Psoriatic arthritis

Psoriatic arthritis is a type of arthritis that affects people with psoriasis — a disease that features red skin patches surmounted with silvery scales. Most people develop psoriasis first and are diagnosed with psoriatic arthritis later, but sometimes joint problems may start before skin patches appear. The main signs and symptoms of psoriatic arthritis are joint pain, stiffness, and swelling. This disease can have effect on any part of body, including the tips of fingers and the spine, and can range from relatively mild to serious [8].

In this paper, we will focus on Rheumatoid Arthritis and Osteoarthritis, which are the most common types of arthritis and chronic diseases [69].

3. Machine learning overview

Machine learning algorithms can be categorised into two types i.e., supervised and unsupervised learning. In **supervised learning**, the machine learning model looks for the relationship between input variables i.e., a set of features and output variables that output classes or labels. It then computes a function capable of predicting output value for a set of unlabelled input values [27]. Hence, supervised learning uses labelled data to train models. On the other hand, **unsupervised learning** recognises the underlying patterns and structure in data, and does not require class labels [27]. Thus, learning with unlabelled data is known as unsupervised learning. In the following, we discuss the most popular supervised and unsupervised learning techniques including deep learning methods.

3.1. Supervised techniques

3.1.1. K-nearest neighbours

K-nearest neighbours (KNN) is a supervised learning algorithm that is used for both classification and regression. It is used to predict a new sample through K-closest samples from the training set [30]. The output varies accordingly depending on whether it is used for classification or regression. In k-NN classification, the samples are classified by the majority vote of its neighbours and assigned a class most common among k samples with the most identical features [23]. In k-NN regression, the outcome is a property value of the object i.e., the average of k nearest neighbours.

3.1.2. Support vector machine

SVM is a traditional supervised machine learning model that is used for classification tasks. Generally, it is used in binary problems, however, with the help of various kernels such as polynomial, it is able to handle multi-class problems as well. SVM takes training samples as input and separates them into different categories for classification using hyperplanes. Therefore, SVM can be used as a model capable of assigning categories to newer samples [30]. SVM is trained to discover the best probable separation of distinct categories by using different function weights such as polynomial [23].

3.1.3. Decision trees

The Decision trees are tree-like models, which consist of decisions and their likely consequences [23]. In contrast to other machine learning methods, the decision trees incorporate the classification functions and the collection of features within one model [70].

3.1.4. Random forest

Random forest [20] is an improvement over decision trees [23]. It is an ensemble classifier for machine learning consisting of several local classifiers and regression tree classifiers. It classifies the input samples based on majority votes of all the trees and results in lower variance and lower bias. It has been used to rate the individual predictors and to pick the best performing models [70].

Random Forest method helps in the reduction of predictors and related variables. Importantly, sub-trees' grouping property allows the Random Forest to address correlation and interaction between variables adequately. The Random Forest approach selects two-thirds of the data to create each of the trees and then uses the remaining one-third to determine misclassification. While the Random Forest algorithm has established several potential trees that perform well in selecting important predictors to identify patients with RA or non-RA, however, it is a “black-box” process, as it is difficult to induce the clear classification rules and interpret the model's predictions [70].

3.1.5. Artificial neural networks

Artificial neural networks (ANNs) are computational models which are inspired by human biological nervous systems and contain parts like neurons and have a layered structure [38]. ANN is mostly used for supervised learning tasks such as classification of labelled data. Artificial neural networks are also capable of performing regression tasks. A typical ANN contains an input layer, hidden layers and output layer. These layers consist of neurons and are connected by different weights [2]. Artificial neural networks have shown to achieve better results compared to traditional methods such as Logistic Regression and Support Vector machines.

3.2. Unsupervised techniques

In the following, we discuss the most representative unsupervised learning methods including the most popular autoencoder and generative adversarial networks.

3.2.1. K-means clustering

K-means clustering is an unsupervised learning technique [37]. In this method, data is clustered into k clusters, reducing the separation space that is irregular with each cluster [54]. A centroid is selected, and classification is achieved depending on the distance between the centroid and adjacent data [69].

3.2.2. Reinforcement learning

Reinforcement Learning is an unsupervised technique, which is based on rewarding desired behaviours and/or punishing undesired ones [58]. Reinforcement learning is about an agent, which is able to perceive and interpret its environment, take actions and learn through trial and error.

In reinforcement learning, a method of rewarding desired behaviours and punishing negative behaviours is devised. This method assigns positive values to the desired actions to encourage the agent and negative values to undesired behaviours. This programs the agent to seek long-term and maximum overall reward to achieve an optimal solution [58].

3.3. Deep learning techniques

Deep learning further expands the capabilities of ANNs by utilising deep neural networks in order to map input data to the desired outcome [27]. In deep learning, the data representations are learned automatically by deep neural networks, which consist of numerous consecutive layers and several basic non-linear operations known as neurons [23]. These architectures can process different kind of input data, for instance, medical images, text, or combination of both. With the deep neural network, the performance of image processing methods in particular has been significantly improved [51]. With the availability of large scale dataset such as ImageNet, it has become possible to train data hungry deep neural networks to recognise and classify a wide range of images and objects [55]. The major difference between an artificial neural network (ANN) and Deep neural network structure is that a deep neural network contains more than one hidden layer [45].

3.3.1. Auto-encoder

An Auto-Encoder (AE) consists of two parts i.e., an encoder and a decoder [51]. Both the encoder and the decoder have hidden layers each, with a shared third layer (the central hidden layer). The encoder part of the auto-encoder finds a compact low dimensional meaningful representation of the input data. The encoder parameters are learnt by combining the encoder with the decoder and jointly training the encoder-decoder structure to reconstruct the input data by minimization of a cost function. The decoder can therefore be defined as a combination of layers joined together by a non-linear activation function which reconstruct the input from the encoder output.

3.3.2. Generative adversarial networks

Generative Adversarial Networks (GANs) [14] were first introduced by Goodfellow et al., in 2014. The main idea behind a GAN is to have two competing neural network models. The first model is called generator, which takes noise as input and generates samples. The other neural network, called discriminator, receives samples from both the generator (i.e., fake data) and the training data (i.e., real data), and discriminates between the two sources [25]. These two networks undergo a continuous learning process, where the generator learns to produce more realistic samples, and the discriminator learns to get better at distinguishing generated data from real data. These two networks are trained simultaneously with the aim that this training will drive the generated samples to be indistinguishable from real data. One of the advantages of GANs is that they can back-propagate the gradient information from the discriminator back to the generator network. The generator, therefore, knows how to adapt its parameters in order to produce output data that can fool the discriminator.

3.3.3. Convolutional neural networks

Convolutional Neural networks are different from typical ANNs as they are generally used for pattern recognition [45]. CNN consists of convolutional layers and pooling layers to learn image specific features and generate feature maps [25]. CNNs have shown much better results for computer vision applications than traditional ANNs [45].

The convolutional neural network is a popular architecture for most image recognition, classification, and detection tasks. It has been used for improved treatment of rheumatic and musculoskeletal diseases (RMDs) in patients. Researchers have used CNNs to detect bone erosions. Similar networks have been used in Doppler Ultrasound images to measure the synovitis disease. One problem with DCNN is that they need enormous quantities of training data and precise training parameters tuning [18].

3.3.4. Fully convolutional network

A fully convolutional network (FCN) has been developed for semantic segmentation of images. FCN is an improved model compared to previous architectures. It uses multiresolution layer combinations and extends classification capability of modern convolutional networks for segmentation. FCN has been reported to achieve high accuracy and is efficient. FCN performs pixel-wise prediction and requires supervised pre-training. FCN can be built by using different deep neural networks such as AlexNet, VGG net and GoogleNet and adapt those into fully convolutional networks to perform the segmentation tasks [35].

3.3.5. U-net

U-net architecture is used mainly for medical image segmentation. It has few parameters compared to other segmentation architectures. The input given to a network is a whole image and matching segmentation masks are created through a series of trainable weights. It has been revealed by [43] that bounding boxes generated by U-net outperform template matching methods for image localisation tasks.

3.3.6. RetinaNet

RetinaNet is another popular architecture, which has been used for joint localisation. It is quite efficient in placing the bounding boxes accurately in order to detect joints [49]. RetinaNet was developed to fix the problem of foreground and background imbalance as the previous approaches did not address it properly. RetinaNet addressed the issue of extreme foreground and background imbalance with dense detectors. The RetinaNet uses a novel Focal Loss to fix the imbalance issue. This resulted in high accuracy and in less computation time [33].

Although deep neural networks provide outputs with high accuracy, it is important to generalise these models to avoid one of the major issues such as overfitting. Therefore, fine-tuning the model for generalizability is very important. In addition to this, deep neural networks can also be seen as a black box, which can make the medical practitioners question their reliability and methodology.

4. Imaging techniques and arthritis data

In the first part of this section, we briefly discuss commonly used imaging techniques for the diagnosis of RA and OA. We will then discuss arthritis data, which has been used in the literature and the publicly available datasets.

4.1. Imaging techniques

There is a wide range of imaging modalities involved in the diagnosis of rheumatoid arthritis (RA) and Osteoarthritis (OA) such as plain radiographs (X-ray images), ultrasonography, magnetic resonance imaging (MRI), computer tomography (CT). In the following, we briefly discuss these imaging techniques.

4.1.1. X-ray

X-ray imaging is the most widely accessible and used tool for the diagnosis of the knee OA, as it is a non-invasive method. It is comparatively inexpensive, rapid and easy to assess imaging technique in order to monitor disease progression [67]. In addition to this, X-ray imaging can reflect the variations in the structure of bones at early stage [34].

4.1.2. Computed tomography (CT)

CT is a 3D volume image of an organ. It is the most commonly used method to image the human body. In addition, CT also involves the injection of contrast agents and contact with significant ionising radiations that are the negatives of CT [26].

4.1.3. Magnetic resonance imaging (MRI)

The non-invasive magnetic resonance imaging is also appropriate and widely used for rheumatoid arthritis diagnosis [13]. It offers a precise diagnosis of the core pathophysiologic phenomena that occur in the myocardium of individuals suffering from RA e.g., myocarditis, vasculitis, and macro-/microcoronary artery disease [26]. The disadvantage of this imagining modality is that it is expensive and inefficient for patients with retained metallic medical prostheses.

4.1.4. Positron emission tomography (PET)

PET is a nuclear imaging modality that utilises radioactive material. It is generally taken up at places of active inflammation and offers improved visualisation of the target lesion [26]. This technique perceives acceptance of positron-emitting radiotracers and allows exact measurement of volumes along with quantification of blood flow. Its disadvantages include high cost and ionising radiation [13].

4.1.5. Carotid ultrasound (CUS)

CUS is currently the most effective non-invasive technique that provides the most robust and confirmed assessments of development of RA. It is also cost-effective imaging modality that gives disease information at an early stage of RA patients.

4.2. Arthritis data and datasets

4.2.1. Rheumatoid arthritis

Since arthritis is a disease which affects different areas of the body such as knee, hands and hips, the data collection method and popular dataset sources significantly vary between them. Firstly, various studies have tried to understand RA, but they have consistently relied on one or few hospital data to do the task. Kim et al., [27] used medical records to identify patients with Rheumatoid Arthritis. Only 2% of 9,482 patients responded and consented on the review of their medical data. In studies carried out by Yoo et al., [69] and Lezcano-Valverde et al., [31] clinical data was used. In former approach, the study involved data from 60 patients from Euji University Hospital, and for the latter the researchers utilised data from two datasets including 1461 records from Hospital Clinics San Carlos and 280 from Universitario de La Princesa Early Arthritis Register Longitudinal study dataset for Rheumatoid Arthritis patients.

Murakami et al., [39] utilised hand radiographs selected by expert radiologists, however, the study only has 159 cases in total out of which 30 of them were used for verification of the Deep Convolutional Neural Network. Orange et al., [44] reported the identification of three distinct synovial subtypes based on the synovial gene signatures of patients with RA. These labels were used to design a histologic scoring algorithm in which the histologic scores correlated with clinical parameters such as ESR and C-reactive protein (CRP) level. The authors selected 14 histologic features from 129 synovial samples (123 RA and six osteoarthritis [OA] patients) and the 500 most variably expressed genes in 45 synovial samples (from 39 RA and six OA patients). Gene-expression-driven subgrouping was explored by k-means clustering, in which n objects are partitioned into k clusters, with each object belonging to the cluster with the nearest mean. Clustering was most robust at 3 and this subgrouping was validated by principal component analysis, but not in an independent dataset. Three subgroups comprising high-inflammatory, low-inflammatory, and mixed subtypes were designated based on their gene patterns and enriched ontology. The aim of the study was to determine the synchrony between synovial histologic features and genomic subtype, thereby yielding a convenient histology-based approach to characterization of synovial tissue. To this end, a leave-one-out cross-validation SVM classifier was implemented. The aim of an SVM is to find a decision hyperplane that separates data points of different classes with a maximal margin (i.e., the maximal distance to the nearest training data points). The model's performance in separating both the high and the low inflammatory subtypes from the other subtypes was relatively good, however, their evaluation dataset was small that resulted into overfitting of SVM.

Some other studies have performed analysis on a smaller number of patients. For instance, Singh et al., [54] used numerical data from 60 anonymous patients. Further, Ureten et al., [67] used datasets from two hospitals, where one of the datasets is used for training, and the other was used for testing. This extra data helped to understand the validity and applicability of the research. The data was attained from UH and SNH Cohorts, respectively. To add on, various studies started to use medical image data for analysis, such as [67] used radiographs from the outpatient clinic of medical faculty and 50 used ultrasounds, which contains 1342 Doppler US Images.

Yoo et al., [69] investigated the clinical data to predict Rheumatoid Arthritis (RA). The study was conducted on 60 RA patients where data was provided by Euji University Hospital. During this study, it was found that the typical diagnosis criteria of RA diagnosis recommended by ACR (American College of Rheumatology), established in 1987, was insufficient for early detection of the disease. Therefore, to help rheumatologists the clinical data including Rheumatoid factor (RF), Anti CCP, SJC and ESR was utilized in order to make early predictions. The data of patients was analyzed using k-means clustering to examine the threshold value of four factors where RA factor (RF)$>$7, Anti CCP$>$18, SJ$>$4 and ESR$>$25. These factors were used to predict RA. The research concluded that for K=4 the achieved accuracy was 84%. In addition to this, selecting two factors and finding their association is higher than selecting only one parameter. The K-means algorithm exhibited that RA can be predicted from two out of these four factors and if one of the two factors i.e., RF and AC are positive rheumatic disease can occur.

4.2.2. *Osteoarthritis*

For Osteoarthritis, since the problem is different from Rheumatoid Arthritis, different type of data collection, processing and analysis is performed. There has been a lot of progress due to the presence of publicly available image databases such as OAI (The Osteoarthritis Initiative) and MOST (Multicentre Osteoarthritis Study). These datasets make the training and testing of deep neural networks feasible.

Thomson et al., [61] used OAI dataset of 4796 participants, out of which 500 were used for performance testing. In 2016, Antony et al., also used X-ray Images from OAI dataset to evaluate their technique on 4,476 images. There have been some studies that have used smaller datasets such as, Gornale et al., [15], [16], which only utilised 200 and 207 knee X-ray Images for training and testing, respectively.

Gornale et al., [17] used 616 X-ray images from various hospitals. Tiulpin et al., [65] used multiple sources such as MOST, Central Finland Centre Hospital and OKOA to acquire 1574, 93 and 77 knee radiographs, respectively. Few other studies such as Tiulpin et al., [64], Tiulpin et al., [62] and Antony et al., [3] have combined the MOST and OAI datasets, and split them into training and test sets. This helped to determine whether their proposed framework is able to handle big data and possible to apply on general populations.

Furthermore, due to the advantages of 3D images, few studies started to use MRIs such as Tolpadi et al., [66]. They used both 2D radiographs and 3D MRI Images from 4,796 patients to attain high performance with the help of 3D CNNs.

Wang et al., [68] used TSE Images and DESS Images from 718 case-control patients (274: Male, 444: Female). Yoo et al., [69] used KNHANES V-1, and bilateral radiographs were accessed for participants who are over the age of 50. The study involved 2665 participants and performed an external validation of 4731 participants from the OAI dataset. Liu et al., [34] used a dataset collected at a hospital in Shanghai that consisted of 2,770 X-ray images. Von et al., [49] used 15,364 hip joint scans to model the severity of hip osteoarthritis.

5. Machine learning for the diagnosis of arthritis

One of the major and challenging steps in automatic diagnosis of arthritis is the detection of joints in images such as X-rays, MRIs and CT scans. In the following, we review some of the recent ML based joint detection techniques followed by ML techniques for the diagnosis of arthritis.

5.1. *ML techniques for joint detection*

Shamir et al., [53] proposed a template matching method for detecting the knee joint centre in an image patch of 20$\times$20 pixels. In their approach, the x-ray image was downscaled to 10% of the actual image size and then exposed to histogram equalisation for normalising the intensity. Next, each input image was scanned through a sliding window of 20$\times$20 to calculate the Euclidean distance. The window that recorded smallest Euclidean distance was declared as the knee joint centre. Although the method was easy to implement, but detection accuracy was low and was slow for big data [3].

The research conducted by [1] used Fully convolutional n[3]. The size of the input image was selected as 200×300 in order to preserve the aspect ratio on the basis of mean aspect ratio (1.6) for all the extracted regions of interest. They used Knee X-ray images and graded the severity level of the impairments according to the Kellgren and Lawrence criteria (an ordinal scale of five points). Elastic Net (EN) and Random Forest (RF) approaches were used to build predictive models using patients' clinical assessment data. X-ray images were used to train a convolutional neural network. Linear mixed-effect (LMM) models have been used to construct the relation between the two knees. For the CNN, EN and RF models, the root mean square error was 0.77, 0.97 and 0.94, respectively. Overall, the LMM reveals close predictive accuracy to the EN regression. However, this multi-stage pipeline approach, which extracts and crops knee joints from the x-ray image requires a lot of memory [34].

[9] used YOLOv2 for knee detection by keeping the initial knee joint size similar to the actual knee joint size where the initial knee size was obtained through clustering on all the available training knee joints. In YOLOv2 object detection is considered as a regression problem that enhances height, width, centre coordinates along with confidence score for every bounding box located in all the grid centres. Their experimental results demonstrated that YOLOv2 performed better than FCN [4] and HOG-SVM [65].

Norman et al., [43] applied U-net model to localise knee joints. The x-ray radiographs were pre-processed by splitting it into left and right knees by dividing the image in the middle and then left side of the image was flipped. Next, a 2-D cross-correlation template method was used to generate bounding boxes around the knee (approximately 500 images). These 500 bounding boxes were then processed to ensure that the template extracted the knee joint area correctly. In addition to this, 450 scans of the localised knee joints were used to train the U-net network. The result of this model was then verified manually on a new dataset of 500 knee images. Their experimental results showed that bounding boxes generated by U-net localised the knee correctly. U-net recorded accuracy of 98.3% for 1000 randomly sampled cases which indicated an improvement over the baseline template matching technique. They next applied deep learning on radiographs to predict OA severity. Interestingly, the authors grouped together KL grade 0 and KL grade 1 as they believed that clinical response is the same for both.

Tiulpin et al., [62] proposed a regression-based deep learning approach to address the issues related to anatomical landmark localisation within knee x-ray radiographs for different stages of Osteoarthritis. According to the authors, the landmark localisation is divided into two sub-tasks i.e., the localisation of region of interest (ROI) and self-localisation of landmarks. The former is deployed for a complete analysis of knee images, and the later is utilised for bone shape and texture analysis. Moreover, manual annotation without prior information on knee anatomy is a minor challenge, which even becomes more complex with increasing OA severity. In their study, initially, they trained a model to localise ROIs while using low-cost annotations within a bilateral radiograph. Later, the model was trained over localised ROIs in order to predict 16 anatomical landmarks on femur and tibia. They also used the VGG image annotation tool to perform annotations and implemented BoneFinder tool for annotations. This technique used the hourglass convolutional network to localise landmarks and soft-argmax layer to evaluate every landmark point directly. The performance of the method was evaluated on two independent datasets, and it demonstrated a better performance compared to state-of-art baseline methods.

Liu et al., [34] deployed region proposal network (RPN) to localise joints in the input X-ray images. The traditional object detection methods usually rely on inexpensive features and economic inference schemes which are computationally expensive. Fully convolutional network eliminates this bottleneck and generates high-quality region proposals which are further utilised by fast R-CNN for object detection. The network consists of 101 layers where each convolutional layer is followed by batch normalisation and ReLU activation layer in order to avoid overfitting issue. In addition to this, the convolutional and max-pooling layers abstract features and generate shared convolutional feature maps used by both RPN and Fast R-CNN. These feature maps serve as input to RPN and to produce an output of region proposals. Once the RPN inferencing scores and coordinates of all the regions are computed, non-maximal suppression (NMS) is applied to remove the redundancy during detection. After applying NMS, the anchor box with highest RPN score was kept. Their proposed approach was able to design an end to end deep learning model for the diagnosis of knee osteoarthritis.

Tolpadi et al., [66] used MRI Images on DenseNet-121 to predict the knee replacement probability in the next five years of a patient's life. The study used OAI dataset, which contained both Imaging and Non-Imaging variables. They adopted the same approach as in [43] to pre-process the radiographic MRI images by cropping them to 500×500 region across the centre of the knee joint. Next, 2D cross-correlation template matching was used to identify bounding box. Subsequently, they trained a U-net architecture that identified posteroanterior radiographs from the OAI study.

Hoang et al., [21] used random forest Regression voting constrained local model technique implemented in BoneFinder tool similar to [62]. Next, based on localised anatomical landmarks and cropped ROIs of 140 mm×140 mm around the knee joint image (at most two ROIs per image), they executed standardisation of each ROI by aligning tibial plateau horizontally. Images were fragmented, flipped, and eventually intensities were normalised. In [4] and [64] training set statistics were used to normalise input data. However, in their study, the authors used a mean of 0.5 and a standard deviation of 0.5.

In a recent study [49], RetinaNet was trained to detect the left and right hip joint within a radiograph. Furthermore, the depicted image (input size 640×640) of one hip was cropped, contrast-stretched and rescaled to 224×224 pixels. A pretrained DenseNet 161 served as a shared convolutional feature extractor with multitask loss function for the implementation of multitask neural network. Each fully connected layer was trained for every radiographic feature of OA to attain final evaluation. The results showed that RetinaNet placed bounding boxes accurately for all the joint images and the model had accuracy of 80.8%.

In a technique proposed by Thomas et al., [60], the image was split to generate left and right knee joint and all the images of the dataset were resized to 299×299 in order to provide a consistent input to neural network model. Initially, two model architectures were considered as the original images varied in resolution, pixel sizes and scale. During image augmentation, the original images were cropped, zoomed in, up scaled, noise was added, flipped horizontally and adjusted contrast so that generated images could follow the distribution of images in original dataset. In addition to this, the training set images were replaced by numerous altered versions (for instance mirroring an image and alteration of contrast for converting a right knee of high contrast to a left knee of low contrast), a model was prepared to do predictions for new images with different parameters than those traced in original non-augmented training set.

5.2. ML techniques for arthritis diagnosis

Lezcano-Valverde et al., [31] demonstrated the use of machine learning method in the development and validation of a predictive model for rheumatoid arthritis mortality based on demographic and clinical variables. The Random Survival Forests (RSF) a non-parametric approach is used to overcome the challenges of traditional survival techniques such as restricted assumptions, proportional hazards, parametric and non-linear effects which results in overfitting. RSF method creates multiple decision trees through bootstrap sampling and generates cumulative hazards function (CHF) from every single tree that are then averaged out to produce an ensemble CHF. Two different datasets were used in this study. One is from Hospital Clinico San Carlos RA cohort (HCSC-RAC) and includes 1,461 patients. It is a daily clinical practice cohort collected as a clinical diagnosis of RA by rheumatologists. It is used for the training of the model. The other dataset is from Hospital Universitario de La Princesa Early Arthritis Register Longitudinal (PEARL), which consists of 280 RA patients. It is used for the model validation. The demographic and clinical variables were collected in first two years of RA diagnosis. It is observed that the mortality rate was 22.1% and 14.6% for HCSC-RA and PERAL in a follow-up time of 4.3 and 5 years respectively. The variables such as age at diagnosis, median ESR (erythrocyte sedimentation rate) and count of hospital admissions showed higher predictive capacity. The prediction errors for training cohort was 0.187 and for validation cohort was 0.233. Overall, potential mortality risk factors have been identified. They were successful in developing a prediction model for RA mortality which has allowed to identify subgroups with higher mortality risks. For further studies, external validation and specific interventions can be applied to reduce the mortality risks for the subgroups with high mortality risk.

Gornale et al., [15] proposed a computer aided analysis of knee OA using active contour segmentation method and K-NN is used to classify various computed features. Dataset consists of 207 knee X-ray images of individuals with different ages, gender, blood group and occupations. In this investigation, image acquisition and pre-processing is performed followed by image segmentation, which is carried out using Active Contour Segmentation method (Chan-Vese and Edge methods are used). Next, image enhancement (Contrast Adjustment technique) is performed to get better quality image. Then various features like Shape features, Statistical features, First-four moments, Haralick features, Texture analysis features and Zernike moments are computed, and classification is performed using the K-nearest neighbour classifier. The reported classification accuracy rate is 88.88% and their technique classifies whether given image is normal or affected.

Gornale et al., [16] implemented a semi-automated method for the diagnosis of Knee OA. Dataset used in their study consists of 200 Knee X-ray images collected from various hospitals based on age, gender, blood group and occupation. In their proposed technique, first image acquisition and pre-processing are performed and then image is segmented using Active Contour segmentation method (Chan-Vese Edge methods). After that image enhancement technique (Contrast adjustment) is used to improve the image quality. Later, Various features such as Haralick, Statistical, First four moments, Texture and Shape are computed. These features are further classified using Random Forest classifier. According to authors, as only radiological assessment of knee X-ray has been considered for this investigation, therefore, a misclassification rate is observed while considering individuals. The classification accuracy rate of their technique is 87.92%, when features are merged together. In future, both clinical symptoms and radiographic assessments need to be considered to develop a detection method for a better classification rate.

In another work, Gornale et al., [17] developed a machine vision method for diagnosis of Knee OA using region based and active shape model. The feature computation involves histogram of oriented gradient (HOG) method. The computed gradients are classified using multiclass SVM classifier to examine OA based on KL (Kellgren Lawrence) grading system. The dataset consists of 616 digital knee X-ray images collected from various Hospitals and Diagnostic Centers based on numerous attributes such as age, gender, blood group, occupation and weight of patients. The images are of dimension 1345×2451 and DICOM Standard (Digital Imaging and Communications in Medicine). Two distinguished medical experts assigned KL grades to each Knee X-ray image. Initially, X-ray images were prepossessed and segmented using implicit active contour algorithm. HOG features were extracted from these X-ray images for further processing and then these features were classified using SVM. The classification rate of 97.96% for Grade-0, 92.85% for Grade-1, 86.20% for Grade-2, 100% for Grade-3 and Grade-4 is obtained, respectively. The classification results validated by the two experts are in close agreement which is 94.96% and 94.64% respectively. Moreover, the proposed method yielded better results with accuracy of 95% compared to methods that used active contour segmentation to acquire ROI and Random Forest Classifier with recognition rate of 87.92% and K-NN Classifier with classification rate of 88.88%.

Thomson et al., [61] implemented a Random Forest Regression Voting Constrained Local Model (RFCLM) for the detection of bone positions and for locating the primary landmarks around the tibia and femur efficiently. They combined features from bone shape and image texture in tibia from 500 X-ray images and then applied two random forests using a weighted approach. Their reported results showed that the combination of shape and texture-based models provide a significant improvement in overall classification performance as the accuracy value increased from 0.789 to 0.849 when using both shape and texture instead of the shape only.

Tiulpin et al., [65] used HOG feature descriptor for preprocessing and applied linear SVM which was pre-trained on radiographic images from three sources (MOST dataset - 1,574 Images, Central Finland centre hospital – 93 and OKOA - 77). This study proposed a system which can be applied for large scale analysis. There is, however, one limitation of this study. The images were annotated by one person and therefore could have a potential bias. In addition to this, more research on localisation is required to be conducted to improve this method.

Subramoniam et al., [56] used SVM with fused kernel functions for classifying 130 (30 normal and 100 abnormal) radiographic images as the haralick features extracted from ROI was one-dimensional data and not sufficient for the appropriate classification. Therefore, kernel functions (such as linear, Polynomial and radial basis) were used to map the one-dimensional data to higher dimensional in order to make classification effective. KL grading system was used in order to grade Articular Space (AS) between bones as the reduction in space indicates OA. Further, only the cases under grade 0, grade 1 and grade 2 were considered and rest all were discarded as joint space narrowing was easily predictable without using any supporting tool. During classification, SVM classifier used hyperplane classification to classify features extracted from ROI of knee joint into normal AS and abnormal AS joints. The results depicted that better classification results were obtained with RBF kernel functions. Further, the classification accuracy was improved by aggregating extracted features and cascading kernel functions.

6. Deep learning techniques for arthritis diagnosis

Besides traditional machine learning approaches, advanced deep learning techniques have also been used for the diagnosis of arthritis. In the following, we discuss the relevant deep learning techniques. (See also Table 2.)

Antony et al., [4] applied Convolutional neural networks on Radiographs (4,476 Participants) to obtain Knee Severity using Kellgren and Lawrence Grade. The classification results with features extracted using pre-trained CNNs (VGG16, VGG-M-128 and

Table 2
Summary of machine and deep learning techniques for the diagnosis of arthritis (in chronological order).

Reference	Year	Technique	Dataset size No. of patients or images	Modality*
Thomson et al., [61]	2015	RFRVCL/SVM	500 images	X-ray
Subramoniam et al., [56]	2015	SVM	130 images	Radiographs
Antony et al., [4]	2016	CNN	4476 patients	MRI
Gornale et al., [15]	2016	Active Contour Segmentation	200 images	X-ray
Gornale et al., [16]	2016	KNN	207 images	X-ray
Yoo et al., [69]	2016	k-means	60 patients	EHR
Gornale et al., [17]	2017	HoG/SVM	616 images	X-ray
Antony et al., [3]	2017	FCN/CNN	7502 patients	MRI
Lezcano-Valverde et al., [31]	2017	Random Survival Forest	1741 patients	EHR
Murakami et al., [39]	2018	CNN	30 patients	Radiographs
Tang et al., [59]	2018	CNN	–	Ultrasound
Orange et al., [44]	2018	SVM	129 patients	Tissue Samples
Norgeot et al., [42]	2019	LSTM	820 patients	EHR
Hemalatha et al., [19]	2019	CNN	–	Ultrasound
Abedin et al., [1]	2019	ElasticNet/Random Forest/CNN	4796 patients	MRI
Chen et al., [9]	2019	Yolo v2	–	X-ray
Tiulpin et al., [63]	2019	ResNet-34	–	Radiographs
Norman et al., [43]	2019	U-Net	500 images	X-ray
Tiuplin et al., [62]	2019	Deep Learning Regression	–	X-ray
Liu et al., [34]	2020	Fast RCNN	–	X-ray
Tolpadi et al., [63]	2020	DenseNet-121		MRI
Hoang et al., [21]	2020	Forest Regression Voting	–	X-ray
Bonaretti et al., [7]	2020	- Extended Phase Graph (EPG)	10 patients	MRI
Dang et al., [10]	2020	CNN	200 patients	X-ray
Von et al., [49]	2020	RetinaNet	–	X-ray
Chen et al., [9]	2019	CNN	–	X-ray
Nguyen et al., [41]	2020	Siamese Network	500 images	X-ray

* As reported in the original paper.

BVLC Caffenet) were shown to be better than OA classification using Wndchrm, which is an open source utility for biological image analysis. Moreover, features from conv4 layer, pool5 layer of VGG-M-128 net and conv5 layer, pool5 layer of BVLC Caffenet had higher classification accuracy than fully connected layers (fc6 and fc7) of VGGnets and Caffenet. Later, multi-class classification through linear SVM was performed, and again CNN features outperformed Wndchrm tool, and classification accuracy of convolutional as well as pooling layers was better than a fully connected layer. In addition to this, BVLC CaffeNet and VGG-M-128 networks were fine-tuned by repacing the top fully connected layer, which further improved multi-class classification where fine-tuned BVLC performed slightly better than VGG-M-128. Moreover, authors argued that treating KL grades as discrete variables can cause classification issues and incorrect predictions. Thus, they proposed more appropriate measure to assess the performance of Knee OA severity by using a continuous evaluation metric such as mean squared error. The pre-trained BVLC CaffeNet model was fine-tuned using both classification loss (cross-entropy on softmax outputs) and regression loss (mean squared error) to compare performance of knee OA severity assessment. The results revealed that it reduced both mean squared error and improved multi-class classification accuracy of the model. The MSE was significantly lower in the case of CNN Regression loss (that is 0.504) than CNN classification network and Wndchrm (that is 0.836 and 2.459 respectively). To evaluate that regression loss provides better classification accuracy, the network trained with classification loss and trained with regression loss were also compared. It was reported that the multiclass classification accuracy of networks fine-tuned for regression loss was 59.6% and it is 43% in case of classification loss. Hence, the network trained with regression performed better than the network trained with classification loss. This is because of the fact that regression provides more information to network about the relationships between KL grades and allows them to better generalise for the unseen data.

In another study, Antony et al., [3] applied the combination of FCN and CNN where FCN was used for Knee Detection, and CNN was used for both classification $(0, 1, 2, 3, 4)$ and regression (0 to 4) tasks to predict the KL grade severity. The classification results are compared with Wndchrm, and this model trained from scratch outperformed Wndchrm with accuracy of 60.3% and MSE 0.898. Even the results are better than their previously reported methods that used BVLC CaffeNet for classifying Knee OA X-rays through transfer learning. These improvements are because of the lightweight architecture of the network with less (5.4 million) parameters compared to 62 million parameters of BVLC. When the network is jointly trained for classification and regression of knee images, the learning curves show a decrease in training and validation losses, and also an increase in training and validation accuracies over the training. In addition, an improvement in the multi-class classification accuracy was observed for network jointly trained for classification and regression as compared to the previous method. The confusion matrix indicated that the classification of Knee OA images conditioned on KL grade 1 was problematic because of small variations.

Chen et al., [9] applied two deep convolutional neural networks (CNN) to automatically assess the severity of knee OA and used the Kellgren-Lawrence (KL) grading system. First, a customised one-stage YOLOv2 network was proposed to detect knee joints, in X-ray images with slight variations. Secondly, the most popular CNN models, including ResNet, VGG, and DenseNet (different variants) as well as InceptionV3, have been fine tuned to classify the detected knee joint images. They also developed new adjustable ordinal loss function which improved the classification accuracy and reduced the MAE of all classification models. The size of X-ray images of the knee is 2048×2560, which was very large for YOLOv2. Therefore, it was resized to the size of 256×320 for all the x-ray images. The results were obtained by using the cross-entropy loss and the proposed ordinal loss. In all comparative classifiers, the ordinal loss obtains greater accuracy and lower MAE in manually cropped knee joints. Most classifiers (except ResNet-152, VGG-16bn, InceptionV3) get higher accuracy at automatically detected knee joints. All classifiers, however, get lower MAE. These findings indicate the superior impact of a proposed ordinal loss function for KL grading of the knee. Some classifiers, such as ResNet-101 (MAE: 0.408, 0.391, accuracy: 65.5% and 66.7%) and VGG-16 (MAE: 0.356, 0.358, accuracy: 68.5% and 69.1%), get even greater accuracy and lower MAE for automatically identified knee joints. The fine-tuned VGG-19 model achieves the best classification performance for knee KL grading compared to the ResNet or DenseNet variants, validating that the performance of CNN models was highly dependent on the training task. The fine-tuned VGG-19 model with the proposed ordinal failure on the knee grading function obtains the best classification accuracy of 69.7 per cent and average absolute error (MAE) of 0.344.

Norgeot et al., [42] proposed long short term memory (LSTM) based technique for the diagnosis and predicting the progression of RA disease. In their study, electronic health record (EHR) of patients was used to train and test the LSTM. The dataset used in their study consists of 820 patients.

Li et al., [32] used Convolutional Siamese Neural networks to score knee severity using KL grading scale. The study takes fine-tuned ResNet-34 as a baseline and shows novelties such as generating a new approach that uses Siamese networks to reduce the learnable parameters and making the model less sensitive to noise. The study was able to use radiographs from both MOST and OAI datasets. However, only the 5, 10- and 15-degree beam angle images were used from the MOST dataset. For processing the images, rather than learning images in similarity metric between the pairs, the study used symmetry in the image itself and the network was able to learn the identical weights for image sides. The Siamese network was able to reduce the number of parameters that were needed to be learned. This helped the model to contain itself to features that a human expert would focus on. The model achieved a multiclass accuracy of 66.71% and AUC of 0.93 on OAI dataset. However, the qualitative assessment on test set revealed that the fine-tuned model can learn features that are not useful or are irrelevant. The study also used GradCAM to determine whether the model has been learning the correct features. One of the biggest strengths of this approach is that the model and the results reproduced as both the dataset and the implementation (code) are publicly available.

Tiulpin et al., [63] developed a multimodal pipeline to generate output prediction on Osteoarthritis progression. The researchers proposed models for the pipeline which can take the input from raw images to generate results. The best results were generated by the combination of Convolutional Neural network and other features such as Age, Sex, BMI, Injury, Surgery, WOMAC score and KL-grade. Next, it was fused with Gradient Boosted Machines. The best model received the AUC of 0.81 (0.79-0.82) and an AP of 0.70 (0.68-0.72). The researchers also used GradCAM to visualize that the model was learning correctly.

Bonaretti et al., [7] presented the mean characteristics of the OAI Control and Incidence Cohort of the selected patients. All subjects in OA groups were monitored for the symptomatic progression of KL grades of 0 baselines. In WOMAC, the control group had a decrease of 0.1 over 36 months, suggesting a lack of development. The OA group's symptomatic progression had a change in WOMAC of 25, indicating a consistent development of the OA symptoms over a follow-up span of 36 months. The OA group's symptomatic development was significantly higher in the theme age and BMI than in the control group (age: 56, BMI: 25, Mean change in WOMAC: -0.1, Mean baseline KL: 0) and (age: 59, BMI: 29, Mean change in WOMAC: -25, Mean baseline KL: 0).

Nguyen et al., [41] proposed a method which consists of two parts: a) A novel variant of Siamese network and b) a novel Deep SSL technique. In their proposed approach, the Siamese model focuses on the medial and lateral sides of the examined knee. They selected three architectures including, GAP, SAM-VH, SAM-HV according to their average BAs. The best one among these (i.e., SAM-HV) has been selected as the basic model of SSL approaches. Their reported results show that their SAM architecture performed better than the Baseline SL model for all dataset settings. In fact, their SAM-HV architecture was 9% better than the BAs Baseline. In the case of 500 samples per KL, SAM-HV model exceeded the baseline model by 6%. The accuracy of their early detection OA system (which implies KL $= 2$) with 500 and 1000 marks per grade KL, was 58% and 74%, respectively.

Yoo et al., [69] built a scoring system and improved ANN model using predictive factors such as sex, age, body mass index, educational status, hypertension, moderate physical activity, and knee pain. Both the scoring system and ANN predicted radiographic knee OA (AUC 0.73 vs. 0.81, $p < 0.001$) and symptomatic knee OA (AUC 0.88 vs. 0.94, $p < 0.001$) with strong discriminative capability in internal validation. All the scoring system and ANN showed lower discriminative potential in predicting radiographic knee OA (AUC 0.62 versus 0.67, $p < 0.001$) and symptomatic knee OA (AUC 0.7) in external validation. For the OAI population, the scoring system predicted radiographic and symptomatic knee OA with the AUCs of 0.62 and 0.70 respectively, and the ANN with the AUCs of 0.66 and 0.76, respectively. Table 2 provides summary (in chronological order) of machine and deep learning techniques for the diagnosis of arthritis.

7. Open problems and research challenges

In view of the above, it can be noted that most of the techniques use the publicly available datasets including MOST and OAI for the diagnosis and prediction of arthritis. We also found that several recent papers have made their code publicly available, thus making it easier for other researchers to reproduce the model and experimental results. The availability of code and evaluation on publicly available datasets can contribute towards the future research and innovation in the field of arthritis diagnosis.

We have also noticed that the recent studies have started to use MRI images for research as they can provide better results than simple 2D images. While acquisition of MRIs is expensive, however, these scans can provide valuable information to machine learning techniques. From the deep learning point of view, several interesting and novel architectures have been developed to address the challenging problem of the arthritis diagnosis.

Despite of this progress, there are still some research challenges which need to be addressed. These are briefly discussed below.

7.1. Deep learning models

Nowadays, deep learning is dominating the field of vision and it has achieved remarkable improvements in this context. However, it has not received much attention in addressing the challenges particular to early diagnosis of arthritis and prediction. By reviewing the current techniques, the number of deep learning methods proposed in the literature is too few.

7.2. Data scarcity - large scale dataset

Data scarcity is a serious issue in medical researcher due to patients' privacy and several other factors such as ethic approvals. Currently, there is no large scale dataset publicly available for the training of data hungry deep learning models. The current techniques have been evaluated on small datasets, which contain only few hundred images. Deep learning algorithms perform better on a large scale dataset and availability of large scale OA and RA datasets can help in the early diagnosis of these diseases and improve the overall diagnostic system performance.

7.3. Data imbalance

Data imbalance is another challenge in machine learning based arthritis diagnosis. The available datasets are imbalanced and the research papers do not mention how they deal with data imbalance. This is also a significant issue for deep learning. Therefore, having a balanced dataset and information on how to handle the imbalanced medical dataset would be very useful.

7.4. RA and OA research gaps

The current literature review suggests that a lot of work has been done on osteoarthritis particularly knee OA. However, there are few studies on other types of OA such as hip osteoarthritis. In addition, machine learning techniques for the diagnosis of rheumatoid arthritis are also very few. There is a room for research in these areas. A combination of imaging technology and electronic health record for the diagnosis of RA and other types of OA will be a good research direction.

8. Conclusion

Early diagnosis of arthritis, including OA and RA, and the ability to track disease progression is challenging. Progress is needed to help medical practitioners and researchers make efficient and reliable decisions within a short time. Accurate predictive modelling for arthritis progression might be difficult to achieve without advanced machine learning (ML) and deep learning (DL) techniques.

In view of the above, our literature survey discussed the current machine learning and deep learning techniques for the diagnosis of OA and RA. The article also highlights open problems and current research challenges. It can be noted that there is a rising trend of ML/DL-related studies and papers in the field of arthritis. This indicates the need for enhancing our understanding about the onset and progression of the disease, and new data-driven tools that could enable early diagnosis and prediction of arthritis. ML and DL could play a key role towards these directions extracting valuable knowledge from various types of clinical data and finding new solutions that utilize data from the greatest possible variety of sources.

Efficient and reliable screening of patients with early arthritis and patients who will progress rapidly using prediction models is important, not only from a medical and patient standpoint but also for the pharmaceutical industry, scientific community and society in general. Such screening could be used as a tool to guide clinical decision making, representing a major advancement towards attaining precision medicine.

In developing predictive modelling approaches, huge dataset is needed particularly for RA research. In addition, to increase the prediction accuracy and interpretability of arthritis prediction models, advanced deep learning techniques are required to be developed. We believe that there are a lot of research opportunities particularly in the artificial intelligence/deep learning domain for the diagnosis of OA and RA that may assist physicians to predict the course of the diagnosis and suggest appropriate treatment for the patients.

Human and animal rights

The authors declare that the work described has not involved experimentation on humans or animals.

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

Declaration of competing interest

The authors declare that they have no known competing financial or personal relationships that could be viewed as influencing the work reported in this paper.

Acknowledgements

This research is supported by Edith Cowan University, Australia.

References

[1] J. Abedin, J. Antony, K. McGuinness, K. Moran, N.E. O'Connor, D. Rebholz-Schuhmann, J. Newell, Predicting knee osteoarthritis severity: comparative modeling based on patient's data and plain x-ray images, Sci. Rep. 9 (2019) 1–11.

[2] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: a survey, Heliyon 4 (2018) e00938.

[3] J. Antony, K. McGuinness, K. Moran, N.E. O'Connor, Automatic detection of knee joints and quantification of knee osteoarthritis severity using convolutional neural networks, in: International Conference on Machine Learning and Data Mining in Pattern Recognition, Springer, 2017, pp. 376–390.

[4] J. Antony, K. McGuinness, N.E. O'Connor, K. Moran, Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 1195–1200.

[5] Australian Bureau of Statistics, National health survey: first results, https://www.abs.gov.au/ausstats/, 2020. (Accessed 4 May 2020).

[6] Australian Institute of Health and Welfare, Arthritis.

[7] S. Bonaretti, G.E. Gold, G.S. Beaupre, pykneer: an image analysis workflow for open and reproducible research on femoral knee cartilage, PLoS ONE 15 (2020) e0226501.

[8] A. Chattopadhyay, M. Maitra, Mri-based brain tumor image detection using cnn based deep learning method, Neurosci. Inform. (2022) 100060.

[9] P. Chen, L. Gao, X. Shi, K. Allen, L. Yang, Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss, Comput. Med. Imaging Graph. 75 (2019) 84–92.

[10] S.D.H. Dang, L. Allison, Using deep learning to assign rheumatoid arthritis scores, in: 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), IEEE, 2020, pp. 399–402.

[11] A. Deodhar, M. Rozycki, C. Garges, O. Shukla, T. Arndt, T. Grabowsky, Y. Park, Use of machine learning techniques in the development and refinement of a predictive model for early diagnosis of ankylosing spondylitis, Clin. Rheumatol. (2019) 1–8.

[12] R.I. Doewes, L. Gangadhar, S. Subburaj, An overview on stress neurobiology: fundamental concepts and its consequences, Neurosci. Inform. 1 (2021) 100011.

[13] G.J. Fent, J.P. Greenwood, S. Plein, M.H. Buch, The role of noninvasive cardiovascular imaging in the assessment of cardiovascular risk in rheumatoid arthritis: where we are and where we need to be, Ann. Rheum. Dis. 76 (2017) 1169–1175.

[14] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, arXiv preprint, arXiv:1406.2661, 2014.

[15] S.S. Gornale, P.U. Patravali, R.R. Manza, Computer assisted analysis and systemization of knee osteoarthritis using digital x-ray images, in: Proceedings of 2nd International Conference on Cognitive Knowledge Engineering (ICKE), 2016, pp. 207–212.

[16] S.S. Gornale, P.U. Patravali, R.R. Manza, Detection of osteoarthritis using knee x-ray image analyses: a machine vision based approach, Int. J. Comput. Appl. 145 (2016).

[17] S.S. Gornale, P.U. Patravali, K.S. Marathe, P.S. Hiremath, Determination of osteoarthritis using histogram of oriented gradients and multiclass svm, Int. J. Image Graph. Signal Process. 9 (2017).

[18] J. Gutiérrez-Martínez, C. Pineda, H. Sandoval, A. Bernal-González, Computer-aided diagnosis in rheumatic diseases using ultrasound: an overview, Clin. Rheumatol. (2020) 1–13.

[19] R. Hemalatha, V. Vijaybaskar, T. Thamizhvani, Automatic localization of anatomical regions in medical ultrasound images of rheumatoid arthritis using deep learning, Proc. Inst. Mech. Eng., H J. Eng. Med. 233 (2019) 657–667.

[20] T.K. Ho, Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, IEEE, 1995, pp. 278–282.

[21] H. Hoang Nguyen, S. Saarakkala, M. Blaschko, A. Tiulpin, Semixup: in- and out-of-manifold regularization for deep semi-supervised knee osteoarthritis severity grading from plain radiographs, arXiv:2003.01944, 2020.
[22] J. Van den Hoek, H. Boshuizen, L. Roorda, G. Tijhuis, M. Nurmohamed, G. van den Bos, J. Dekker, Mortality in patients with rheumatoid arthritis: a 15-year prospective cohort study, Rheumatol. Int. 37 (2017) 487–493.
[23] M. Hügle, P. Omoumi, J.M. van Laar, J. Boedecker, T. Hügle, Applied machine learning and artificial intelligence in rheumatology, Rheumatol. Adv. Pract. 4 (2020) rkaa005.
[24] A. Jamshidi, J.P. Pelletier, J. Martel-Pelletier, Machine-learning-based patient-specific prediction models for knee osteoarthritis, Nat. Rev. Rheumatol. 15 (2019) 49–60.
[25] S. Khan, H. Rahmani, S.A.A. Shah, M. Bennamoun, A guide to convolutional neural networks for computer vision, in: Synthesis Lectures on Computer Vision, vol. 8, 2018, pp. 1–207.
[26] N.N. Khanna, A.D. Jamthikar, D. Gupta, M. Piga, L. Saba, C. Carcassi, A.A. Giannopoulos, A. Nicolaides, J.R. Laird, H.S. Suri, et al., Rheumatoid arthritis: atherosclerosis imaging and cardiovascular risk assessment using machine and deep learning–based tissue characterization, Curr. Atheroscler. Rep. 21 (2019) 7.
[27] K.J. Kim, I. Tagkopoulos, Application of machine learning in rheumatic disease research, Korean J. Int. Med. 34 (2019) 708.
[28] C. Kokkotis, S. Moustakidis, E. Papageorgiou, G. Giakas, D. Tsaopoulos, Machine learning in knee osteoarthritis: a review, Osteoarthr. Cartil. Open 2 (2020) 100069.
[29] M. Kourilovitch, C. Galarza-Maldonado, E. Ortiz-Prado, Diagnosis and classification of rheumatoid arthritis, J. Autoimmun. 48 (2014) 26–30.
[30] M. Kuhn, K. Johnson, Applied Predictive Modeling, vol. 26, Springer, 2013.
[31] J.M. Lezcano-Valverde, F. Salazar, L. León, E. Toledano, J.A. Jover, B. Fernandez-Gutierrez, E. Soudah, I. González-Álvaro, L. Abasolo, L. Rodriguez-Rodriguez, Development and validation of a multivariate predictive model for rheumatoid arthritis mortality using a machine learning approach, Sci. Rep. 7 (2017) 1–10.
[32] M.D. Li, K. Chang, B. Bearce, C.Y. Chang, A.J. Huang, J.P. Campbell, J.M. Brown, P. Singh, K.V. Hoebel, D. Erdoğmuş, et al., Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging, npj Digit. Med. 3 (2020) 1–9.
[33] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[34] B. Liu, J. Luo, H. Huang, Toward automatic quantification of knee osteoarthritis severity using improved faster r-cnn, Int. J. Comput. Assisted Radiol. Surg. 15 (2020) 457–466.
[35] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[36] R.M. Loureiro, J. Collin, D.V. Sumi, L.C. Araújo, R.W. Murakoshi, R.L.E. Gomes, M.M. Daniel, Postoperative ct findings of orthognathic surgery and its complications: a guide for radiologists, J. Neuroradiol. (2021).
[37] J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Oakland, CA, USA, 1967, pp. 281–297.
[38] S.B. Maind, P. Wankar, et al., Research paper on basic of artificial neural network, Int. J. Recent Innov. Trends Comput. Commun. 2 (2014) 96–100.
[39] S. Murakami, K. Hatano, J. Tan, H. Kim, T. Aoki, Automatic identification of bone erosions in rheumatoid arthritis from hand radiographs based on deep convolutional neural network, Multimed. Tools Appl. 77 (2018) 10921–10937.
[40] U. Nadeem, S.A.A. Shah, M. Bennamoun, R. Togneri, F. Sohel, Real time surveillance for low resolution and limited-data scenarios: an image set classification approach, Inf. Sci. (2021).
[41] H.H. Nguyen, S. Saarakkala, M.B. Blaschko, A. Tiulpin, Annotation-efficient deep semi-supervised learning for automatic knee osteoarthritis severity diagnosis from plain radiographs, in: Online Proceedings Medical Imaging Meets NeurIPS Workshop 2020, NeurIPS, 2020.
[42] B. Norgeot, B.S. Glicksberg, L. Trupin, D. Lituiev, M. Gianfrancesco, B. Oskotsky, G. Schmajuk, J. Yazdany, A.J. Butte, Assessment of a deep learning model based on electronic health record data to forecast clinical outcomes in patients with rheumatoid arthritis, JAMA Netw. Open 2 (2019) e190606.
[43] B. Norman, V. Pedoia, A. Noworolski, T.M. Link, S. Majumdar, Applying densely connected convolutional neural networks for staging osteoarthritis severity from plain radiographs, J. Digit. Imaging 32 (2019) 471–477.
[44] D.E. Orange, P. Agius, E.F. DiCarlo, N. Robine, H. Geiger, J. Szymonifka, M. McNamara, R. Cummings, K.M. Andersen, S. Mirza, et al., Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data, Arthritis Rheumatol. 70 (2018) 690–701.
[45] K. O'Shea, R. Nash, An introduction to convolutional neural networks, arXiv preprint, arXiv:1511.08458, 2015.
[46] S.D. Pande, P.P. Jadhav, R. Joshi, A.D. Sawant, V. Muddebihalkar, S. Rathod, M.N. Gurav, S. Das, Digitization of handwritten Devanagari text using cnn transfer learning – a better customer service support, Neurosci. Inform. 2 (2022) 100016.
[47] M. Roques, F. Tanchoux, L. Calvière, L. Cuinat, V. Lubrano, E. Uro-Coste, C. Cognard, V. Larrue, F. Bonneville, Mri with dwi helps in depicting rheumatoid meningitis, J. Neuroradiol. 41 (2014) 275–277.
[48] A. Sarkar, A. Singh, R. Chakraborty, A deep learning-based comparative study to track mental depression from eeg data, Neurosci. Inform. (2022) 100039.
[49] C.E. von Schacky, J.H. Sohn, F. Liu, E. Ozhinsky, P.M. Jungmann, L. Nardo, M. Posadzy, S.C. Foreman, M.C. Nevitt, T.M. Link, et al., Development and validation of a multitask deep learning model for severity grading of hip osteoarthritis features on radiographs, Radiology 295 (2020) 136–145.
[50] S.A. Shah, U. Nadeem, M. Bennamoun, F. Sohel, R. Togneri, Efficient image set classification using linear regression based image reconstruction, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 99–108.
[51] S.A.A. Shah, M. Bennamoun, F. Boussaid, Iterative deep learning for image set based face and object recognition, Neurocomputing 174 (2016) 866–874.
[52] S.A.A. Shah, K. Yahya, G. Mubashar, A. Bais, Quantification and visualization of mri cartilage of the knee: a simplified approach, in: 2010 6th International Conference on Emerging Technologies (ICET), IEEE, 2010, pp. 175–180.
[53] L. Shamir, S.M. Ling, W. Scott, M. Hochberg, L. Ferrucci, I.G. Goldberg, Early detection of radiographic knee osteoarthritis using computer-aided analysis, Osteoarthr. Cartil. 17 (2009) 1307–1312.
[54] U.V. Singh, E. Gupta, T. Choudhury, Detection of rheumatoid arthritis using machine learning, in: 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), IEEE, 2019, pp. 25–29.
[55] B. Stoel, Use of artificial intelligence in imaging in rheumatology–current status and future perspectives, RMD Open 6 (2020) e001063.
[56] M. Subramoniam, A non-invasive method for analysis of arthritis inflammations by using image segmentation algorithm, in: 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], IEEE, 2015, pp. 1–4.
[57] A. Sujith, G.S. Sajja, V. Mahalakshmi, S. Nuhmani, B. Prasanalakshmi, Systematic review of smart health monitoring using deep learning and artificial intelligence, Neurosci. Inform. 2 (2022) 100028.
[58] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.
[59] J. Tang, Z. Jin, X. Zhou, H. Chu, J. Yuan, M. Wu, Q. Cheng, X. Wang, Grading of rheumatoid arthritis on ultrasound images with deep convolutional neural network, in: 2018 IEEE International Ultrasonics Symposium (IUS), IEEE, 2018, pp. 1–4.
[60] K.A. Thomas, Ł. Kidziński, E. Halilaj, S.L. Fleming, G.R. Venkataraman, E.H. Oei, G.E. Gold, S.L. Delp, Automated classification of radiographic knee osteoarthritis severity using deep neural networks, Radiol. Artif. Intell. 2 (2020) e190065.
[61] J. Thomson, T. O'Neill, D. Felson, T. Cootes, Automated shape and texture analysis for detection of osteoarthritis from radiographs of the knee, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 127–134.
[62] 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.
[63] A. Tiulpin, S. Saarakkala, Automatic grading of individual knee osteoarthritis features in plain radiographs using deep convolutional neural networks, Diagnostics 10 (2020) 932.
[64] A. Tiulpin, J. Thevenot, E. Rahtu, P. Lehenkari, S. Saarakkala, Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach, Sci. Rep. 8 (2018) 1–10.
[65] A. Tiulpin, J. Thevenot, E. Rahtu, S. Saarakkala, A novel method for automatic localization of joint area on knee plain radiographs, in: Scandinavian Conference on Image Analysis, Springer, 2017, pp. 290–301.
[66] A.A. Tolpadi, J.J. Lee, V. Pedoia, S. Majumdar, Deep learning predicts total knee replacement from magnetic resonance images, Sci. Rep. 10 (2020) 1–12.
[67] K. Üreten, H. Erbay, H.H. Maraş, Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network, Clin. Rheumatol. 39 (2020) 969–974.
[68] C. Wang, Z. Peng, Design and implementation of an object detection system using faster r-cnn, in: 2019 International Conference on Robots & Intelligent System (ICRIS), IEEE, 2019, pp. 204–206.
[69] J. Yoo, M.K. Lim, C. Ihm, E.S. Choi, M.S. Kang, A study on prediction of rheumatoid arthritis using machine learning, Int. J. Appl. Eng. Res. 12 (2017) 9858–9862.
[70] S.M. Zhou, F. Fernandez-Gutierrez, J. Kennedy, R. Cooksey, M. Atkinson, S. Denaxas, S. Siebert, W.G. Dixon, T.W. O'Neill, E. Choy, et al., Defining disease phenotypes in primary care electronic health records by a machine learning approach: a case study in identifying rheumatoid arthritis, PLoS ONE 11 (2016) e0154515.