



## Review article

## Role of machine learning in medical research: A survey

Arunim Garg, Vijay Mago\*

Department of Computer Science, Lakehead University, Canada

## ARTICLE INFO

## Article history:

Received 1 July 2020

Received in revised form 4 January 2021

Accepted 19 January 2021

Available online 2 February 2021

## MSC:

00-01

99-00

## Keywords:

Medical research

Machine learning

Deep learning

Medical data

## ABSTRACT

Machine learning is one of the essential and effective tools in analyzing highly complex medical data. With vast amounts of medical data being generated, there is an urgent need to effectively use this data to benefit the medical and health care sectors all across the world. This survey paper presents a systematic literature review for the investigation of various machine learning techniques used for numerous medical applications which are published in highly reputable venues in recent years. Considering only the recent work, we are able to survey the current machine learning and deep learning models that are being used for medical data. This literature review identifies a clear shift of artificial intelligence techniques used in the medical domain, with deep learning methods taking precedence over machine learning methods.

© 2021 Elsevier Inc. All rights reserved.

## Contents

1. Introduction.....	1
1.1. Motivation .....	2
1.2. Survey structure.....	2
2. Methods.....	3
2.1. Search strategy.....	3
2.2. Selection strategy.....	3
3. Medical data.....	4
4. Machine learning techniques.....	5
4.1. Support vector machines .....	6
4.2. Clustering.....	6
4.2.1. K-means clustering .....	7
4.3. Decision trees (DT) .....	7
4.4. Random forests .....	7
4.5. K-Nearest Neighbors.....	8
4.6. Naïve Bayes .....	9
4.7. Deep learning .....	9
4.7.1. Neural Networks (NNs) .....	9
4.7.2. Convolutional Neural Networks (CNNs) .....	10
4.8. Other recent medical applications of ML and DL.....	11
5. Results.....	11
6. Conclusion .....	15
Declaration of competing interest.....	16
Acknowledgments .....	16
Appendix A. ....	16
Appendix B. ....	16
References .....	16

\* Corresponding author.

E-mail address: [vmago@lakeheadu.ca](mailto:vmago@lakeheadu.ca) (V. Mago).

## 1. Introduction

With the ever-increasing generation of medical data, the need for analyzing the data continues to grow [1]. For example,

medical data that originates from disease diagnosis can be analyzed to save significant amount of expenditure. McKinsey reported that 80% of American healthcare spending is on chronic disease treatment and 50% of Americans suffer from chronic diseases [2]. The prevalence of chronic diseases, in America and other countries with good living standards, is growing [3]. In China, 86% of disease-related deaths are connected to chronic diseases [3]. Early detection of diseases in the process of diagnosis and prognosis is critical for healthcare. With a huge influx of medical data being developed and maintained for research [4], medical data is being stored in electronic health records (EHRs) [3]. Machine learning (ML) models can automatically extract relevant features from huge datasets of patient records stored in EHRs [3]. Using the relevant features from the EHRs, the ML model can then further help in the detection of diseases by monitoring the data and predicting potential diseases.

Machine learning and deep learning (DL) help a computer model “adapt to new circumstances and to detect and extrapolate patterns” [5] by learning from raw data. A model is said to be learning if the performance measure of that model increases for a task with experience [6], and if the model can be more useful in performing the task [7]. For a specific task, a ML model would be fed examples containing various features as inputs and the labels as the desired output. For instance, pixels from digitized images of a human brain read by pathologists are converted into features, and the traces of cancer cells present in the brain are set to be as the labels [8]. The learning model is then trained using different types of learning techniques, in this case object detection techniques, to predict the cancer cells reliably from a new, never-seen input of a human brain. In the field of medicine, ML models learn the domain knowledge from the data anywhere from doctor-level observations (doctor notes) to medical imaging (computerized tomography scans or CT scans) [9]. Algorithms parse through all the data to extract only the descriptors (features) which would help accurately predict the outcome. A visualization for the process of machine learning starting with data preparation to model evaluation can be seen in Fig. 1.

Machine learning adapts several learning approaches which are supervised and unsupervised learning, which are differentiated based on of presence of feedback. Supervised learning is when the model is offered a set of examples with respective labels and the model learns to identify the correct label [10]. Classification and regression are the most common types of supervised learning. Classification models take data as the input to the classifier and assign the data into a specific class, usually between two classes [11]. Classification models like support vector machines have consistently shown accurate results for various disease diagnosis tasks [12]. Whereas, unsupervised learning is when the labels for the inputs are not provided. The model learns the similarity among the inputs and tries to predict the label as an output based on the similarities [10]. Clustering is an example of unsupervised learning, where the algorithm groups similar inputs into a cluster. As an example, fuzzy c-means clustering has been successfully used for breast cancer diagnosis [13]. An extensive number of ML techniques, following either of these approaches, are being used for medical and healthcare applications. Deep learning, an improved form of machine learning, implements higher number of layers than a typical machine learning model to extract deeper features from the data which enables the generation of more refined predictions, as explained in Section 4.7.

Recent examples of artificial intelligence techniques being used in the field of medical domain identifies a clear shift from machine learning methods to deep learning methods. In this literature review, we show the methodology for searching and selecting the papers that are reviewed to identify the role of machine learning in medical research.

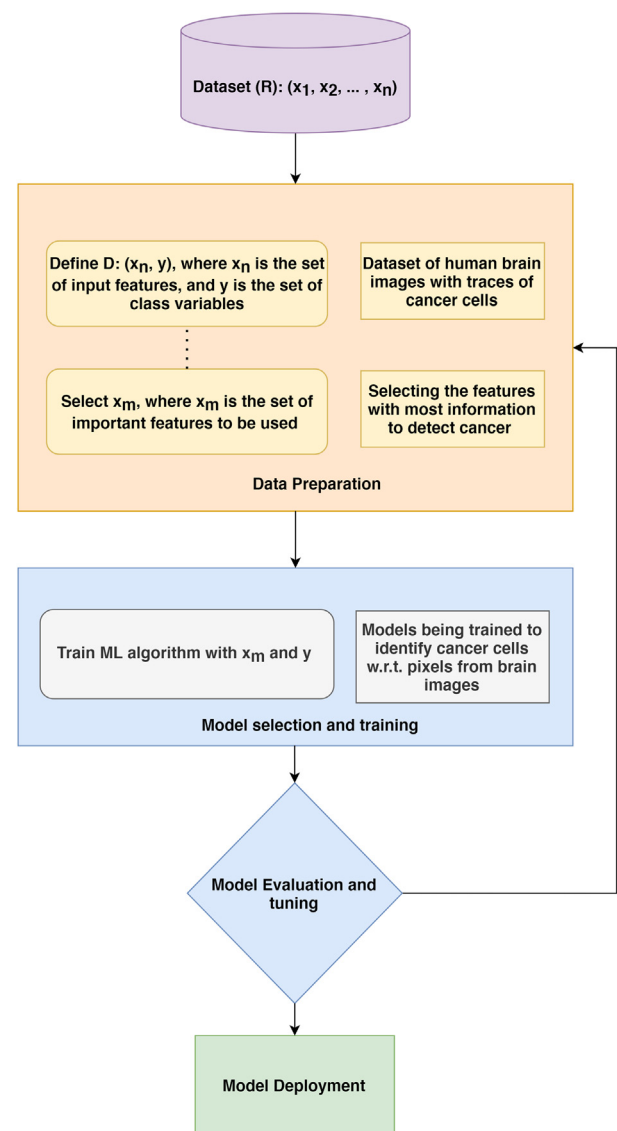


Fig. 1. Development process of a machine learning model in medical domain.

### 1.1. Motivation

Machine learning has seen a tremendous amount of research in the past few decades. With discoveries in the area of artificial intelligence (AI) and ever-increasing technological development, novel ML and DL techniques can analyze medical data more efficiently. Benefitted from this overall growth, researchers are building better ML and DL models to deal with even more complex and abundant medical data. Our motivation for this literature review is to analyze the role of ML in medical research in recent years. After observing the ML techniques used in medical research, we can follow the type of shift that is taking place in medical research. Furthermore, this survey can benefit researchers to review the specific kind of ML technique used for specific medical applications in the past five years.

### 1.2. Survey structure

The visual structure of the survey is displayed in Fig. 2. In Section 2, various methodologies of how the sources for this survey paper have been searched and selected have been listed

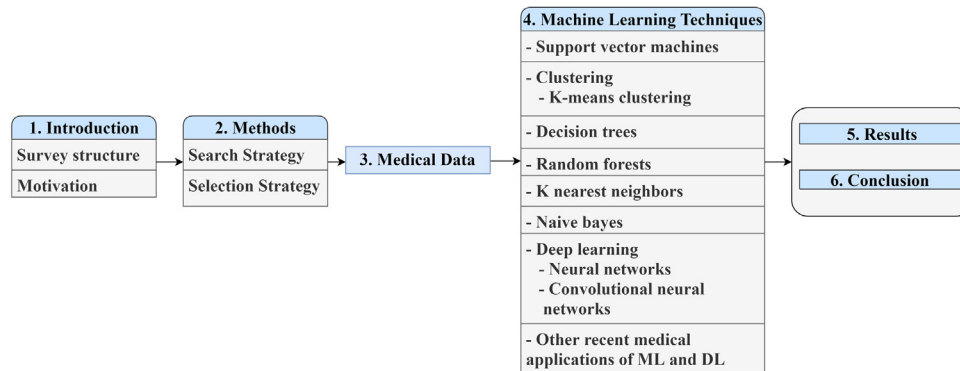


Fig. 2. Structure of this article.

and visualized. In Section 3, the paper highlights the importance, challenges and types of medical data that is currently prevalent and being analyzed by the various ML techniques. This survey paper will then focus on the ML techniques that are being used currently in the field of medicine for various medical applications in Section 4. In Section 5 of this paper, the findings from this survey paper have been listed and visualized. A brief analysis of ML techniques surveyed within the boundaries of this article is also provided in this section. Finally, in Section 6 we conclude our survey with a brief summary.

## 2. Methods

We conducted a systematic literature review (SLR) which involved an extensive search for the relevant research articles and survey papers in the field of medicine. A SLR is an unbiased and a well-established methodology to identify, analyze, and interpret all available data related to a field that is unrepeatable (to a degree) [14]. For a successful SLR to be conducted, there needs to be an understanding of the methodologies prior to conducting the review [14,15]. Methodologies used in the search and selection strategy for the extraction of the papers used for this survey paper follows preset standard criteria, explained in the following subsections.

### 2.1. Search strategy

Machine learning, in the field of computer science, is an umbrella term which incorporates different types of learning like supervised, unsupervised, semi-supervised, and deep learning. Searching the term ‘machine learning’ in the popular database corpora like Google Scholar offered us approximately 4.5 million articles.

To narrow our search and to accommodate all the available ML techniques, for this survey paper, we employed a comprehensive search strategy. We used the following key terms in Google Scholar:

- “medical data” AND “machine learning”
- “medical data” AND “neural networks” - here, we listed machine learning techniques in addition to ‘neural networks’ which are commonly used for medical data analysis like convolution neural network, support vector machine, decision trees and many more.
- the articles published after 2015 were taken into consideration (with some exceptions mentioned in the next section)

After narrowing our search by using more keywords, we were able to condense the number of articles to thousands after which we used a specific selection strategy to choose the appropriate sources for our survey.

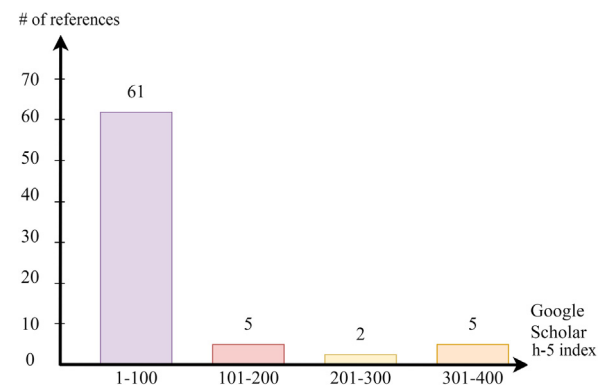


Fig. 3. Distribution of Google Scholar h-5 index over the frequency count of the references.

### 2.2. Selection strategy

Our selection strategy was based on five different criteria. There is an abundance of papers being published in the vast amounts of publication venues in all different research fields. For this survey paper, we set a standard for selecting the papers only published in the most reputable publication venues. The criterion for defining the reputable venues was to, first, keep track of the Google Scholar metric – h5 Index – which they define on their website as follows, “h5-index is the h-index for articles published in the last 5 complete years. It is the largest number h such that h articles published in 2014–2018 have at least h citations each” [16]. We aimed to have a Google Scholar h5-index ranking of 40 or higher, with an exception of 18 references.

Second, we also kept track of the SCIMago journal index (SJR) [17] for each publication venue. They define their rank to show the “journal’s impact, influence or prestige. It expresses the average number of weighted citations received in the selected year by the documents published in the journal in the three previous years”. SJR index also gives a quartile (Q) rank from Q1 to Q4, Q1 being the top quartile, for the subject categories the journal publish their articles in. We aimed to have a SJR index of 40 or higher, with an exception of 2 references. Our priority in this selection strategy was also to have references of Q1 quartile.

As seen in Appendix B, we have collected all the appropriate information regarding our methodologies. The table includes the authors and the year the article was published, the publication venue, its Google scholar h-5 index and SJR index (and SJR quartile), and the number of citations the article has received as of April 4, 2020.

The trend of the distribution of Google scholar h-5 index and SJR index for all our sources can be seen in Fig. 3 and Fig. 4,

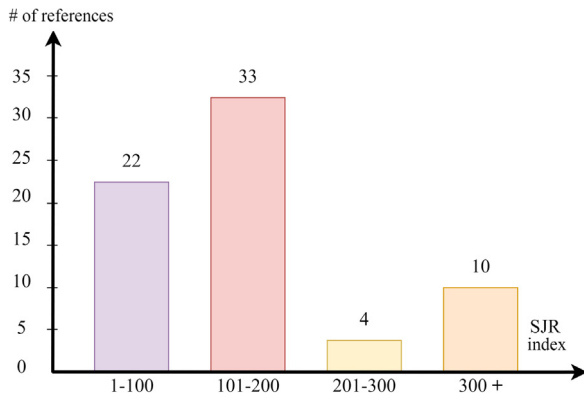


Fig. 4. Distribution of SJR index over the frequency count of the references.

Distribution of references by quartile of SJR ranking

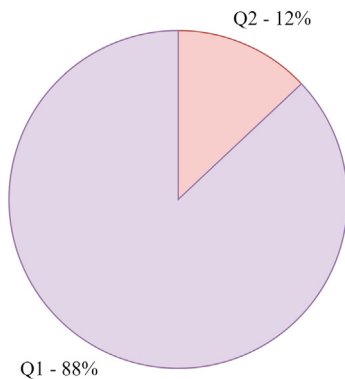


Fig. 5. Distribution of the quartile of SJR index.

respectively. Most of the sources that were selected for this survey paper, as intended, are from the past 5 years. In addition, 87.9% of our sources in this survey paper are from venues with SJR Q1 quartile, as shown in Fig. 5. Out of all the sources, only 9 of the sources were published in Q2 venues [14,18–25], since they either passed our other selection criterion or were needed for our survey.

To ensure a topical survey, we considered the year of publication to be significant. The papers that we chose to include in our survey are from 2015 onward. However, out of all the sources referenced in our article, we have published work from roughly 2015 and onward, with exceptions to the articles which were either published by highly reputed authors, indicative by their citations as of April 4, 2020; for example, Geoffrey Hinton’s ImageNet article [26] has 59633 citations or written during the inception of their particular topic, like for the K-nearest neighbors technique we chose a 1967 paper with 11,770 citations [27], which is when the first paper was published for this topic. Other exceptions made for this selection strategy were due to the consideration of the relevant topics to our survey that the sources tackled. The sources that were cited before 2015 for the reasons mentioned above include [6,7,15,18,19,26–53].

We, also, aimed to choose highly-cited articles to ensure that we survey and study widely-accepted and influential work, with the exception of the few papers which were recently published (late 2019, or early 2020). In Figs. 6 and 7, we have visualized the distribution of the years our sources were published in and number of citations each reference received as of aforementioned date.

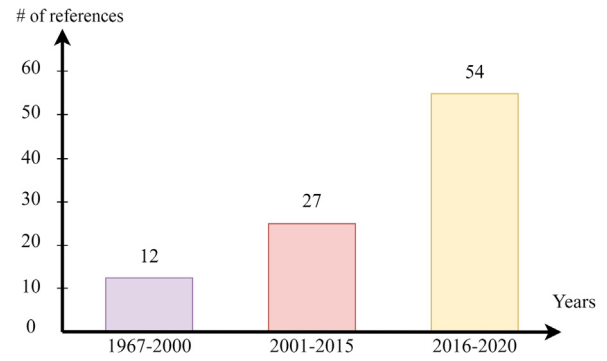


Fig. 6. Distribution of references by the year of publication.

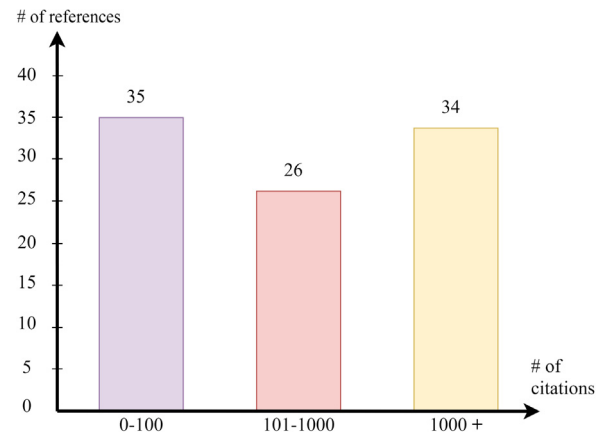


Fig. 7. Distribution of references by the number of citations.

### 3. Medical data

Medical data is available from a variety of sources including disease prognosis, public health records, biometric data, medical imaging, and is rapidly growing in this age of big data [14]. Moreover, analyzing this data and using it for appropriate applications can bring a significant positive change in the medical industry. As summarized by Chen, M., et al. (2014) [4], effective utilization of the medical data potentially “increases the USA (United States of America) medical industry to USD 300 billion, and thus reducing the healthcare expenditure by 8%”. Medical data is complex, by its nature or how it has been collected. It constitutes class imbalance and has complex and high dimensionality since there exists a lot of parameters and variables. Poor collection of medical data results in missing data (patient records) [14].

Medical data exists in various forms like clinical data which includes patient records, and bio-medical data which includes data related to genomics, bio-medicine and drug research and development. Bio-medical data is embedded with complex biological phenomena which can be revealed by constructing equally complex and intricate analytical models [4,54]. Clinical medical data, also, exists in plethora. As reported by Chen, M., et al. [4], clinical data stored as electronic health records is in thousands of terabytes with companies in various agencies which not only includes the medical history of the patients but also image data from CT scans and x-rays. Clinical or medical data, generally exists in two forms, structured or unstructured data [25]. Structured clinical data can be like demographics of the patient, living habits of the patient (drinking, smoking, or eating habits), and disease history. Whereas, unstructured clinical data would consist of doctor–patient conversation or doctor’s investigation records [3].

**Table 1**  
Machine learning techniques for medical applications.

Paper	Topic	Machine learning techniques									
		NN	CNN	SVM	K-means	DT	Naive Bayes	Clustering	Novel ML techniques	RF	KNN
Z. Wang, et al. (2019)	Medical Diagnosis	x									
Shaikhina, T., et al. (2017)	Medical Data Classification	x									
Frid-Adar, M., et al. (2018)	Brain lesion segmentation		x								
Shin, H. C., et al. (2016)	Lymph node detection		x								
Shen, L., et al. (2016)	Various datasets			x							
Khan, S. et al (2017)	UCI datasets				x						
Wiwie, C., et al. (2015)	Various datasets							x			
T. Le (2019)	UCI datasets	x			x						
Cheplygina, V. (2019)	Medical vs non-medical input data		x								
Greenspan, H., et al (2016)	Medical image classification		x								
Nilashi, M., et al (2017)	UCI datasets					x					
Wang, Z., et al (2012)	Electronic Health Records								x		
Dallora, A. L., et al. (2017)	Prognosis of dementia	x		x		x	x				
Taylor, R. A., et al. (2018)	Uninary Tract infection			x							
Leha, A., et al. (2019)	Pulmonary hypertension			x		x					
Meiring, C., et al. (2018)	ICU patients mortality rate	x		x		x				x	
Hasnain, Z., et al. (2019)	Bladder cancer patients			x						x	x
Gao X., et al. (2015)	Cataract classification		x								
Setio, A.A.A., et al. (2016)	Nodules in chest CT scans		x								
Nie, D., et al. (2016)	Brain tumor: high-grade gliomas		x								
Payer, C., et al. (2016)	Image localization		x								
Zheng Y., et al. (2015)	Carotid artery bifurcation detection		x								
Ghesu F. C., et al. (2016)	Localization of aortic valve	x									
Kukar, M. et al.	Femoral neck fracture recovery						x				
Kamnitsas, K., et al. (2017)	Brain lesion segmentation		x								
Chen, M., et al. (2017)	Cerebral infraction disease		x			x	x				x
van Grinsven, et al. (2016)	Hemorrhages detection in color fundus		x								
Ramalingam V. V., et al. (2018)	Heart Disease			x		x				x	x
Chang, Y., et al. (2019)	Medical image denoising		x								
Yelipe, U., et al. (2018)	9 benchmark datasets							x			
Zhu, M., et al. (2018)	Weight Voting Random Forests							x		x	
Wang, J., et al. (2015)	Expenditures of diagnosed diabetics							x		x	
Goto, T., et al. (2018)	Disposition of asthma in ED							x			
<b>Total:</b>	<b>33</b>	<b>6</b>	<b>13</b>	<b>7</b>	<b>2</b>	<b>6</b>	<b>3</b>	<b>5</b>	<b>1</b>	<b>5</b>	<b>3</b>

Analyzing medical data faces a lot of challenges including its low quality (most of the data is unstructured, and is hard to implement for its intended use), the presence of unlabeled data, and inconsistent or falsified collection of the data, which are embedded in the nature of the data itself [14,28,55,56]. Inherently, medical data has other disadvantages with respect to its analysis like privacy, and being of heterogeneous nature [52]. ML techniques are used to overcome these challenges, in addition to its analysis. ML models specifically curated for medical data mining are in place which help reduce the affect of the above-mentioned constraints [53,56]. It can help with extracting useful information from the unstructured data and simplifying it into structured form, labeling the unlabeled data, and also work with data of high variability of data types. In addition to overcoming these challenges, ML models have many applications in medical data analysis, specifically, disease and treatment, public health and research applications, and population management [57].

Various ML techniques that are surveyed in this paper, employ numerous medical datasets to test the models, as listed in Table 1 in the Results section (Section 5). The datasets include

image datasets like CT scans, and 3-dimensional brain images or tabular datasets from UCI [58] (University of California Irvine) or benchmark datasets from KEEL [59] (Knowledge Extraction based on Evolutionary Learning) like Australian, Cleveland, Ecoli, German, Hepatitis, Wine, Iris, Pima and Wisconsin datasets. There are various other datasets specific to the medical field included in this survey. In the next section, the focus is to elaborate different ML and DL techniques that use medical data for its analysis.

#### 4. Machine learning techniques

For machine learning in the medical domain, finding appropriate features (descriptors) for data to predict the correct classification (label) is one of the essential tasks. Conventionally, for example, medical image interpretation was performed by highly trained physicians and other medical or healthcare professionals. They would develop meaningful descriptors for certain output labels for a medical dataset, on the basis of their expert knowledge in the domain [60]. But, with enormous amounts of data being generated everyday, various ML techniques can have prediction



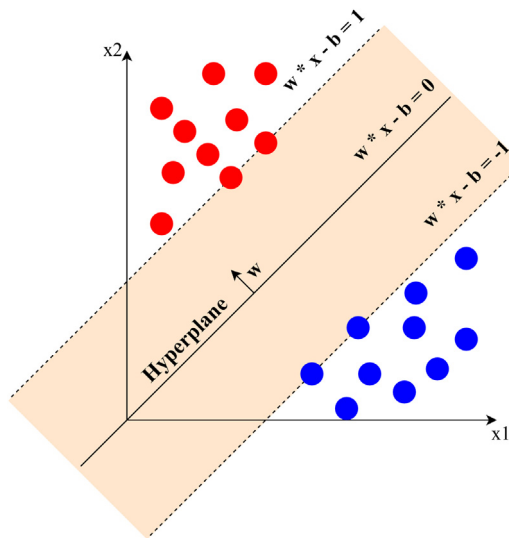


Fig. 8. Support Vector Machine producing a linear solution [19].

almost as accurate as human experts of the field [60]. For cancer research, ML models have proven to accurately mine relevant features from clinical imaging arrays to help detect early indicators in cancer treatment [61]. ML can help physicians in clinical applications as well. Prediction from ML techniques can augment the work of clinicians in the process of prognosis, diagnosis, image examination and treatment [9,62].

There are several different ML algorithms which deal with tasks of prediction, classification, clustering, and regression, among many others. The following subsections provide various ML techniques used for medical applications in the domain of medicine and health care. Table 2 in the Appendix A summarizes the mathematical aspects of various machine learning techniques.

#### 4.1. Support vector machines

Support vector machines (SVMs) are supervised ML models used for binary classification problems in numerous fields, especially in the field of medicine. A typical SVM model maps the data points into a higher dimensional space where it employs a hyperplane to separate the two classes with maximum distance in between [18,33]. Maximizing the distance between the two classes, with a linearly separable hyperplane, is a quadratic programming problem and with no prior knowledge of the mapping, a SVM uses dot product functions in the feature space in the map called kernels to find the most optimal hyperplane separator, which can be seen in Eq. (1), where  $w^T$  is the weight vector for the input vector  $x$ , and  $b$  is the bias. Also, as seen in Fig. 8, the linear plane ( $w * x - b = 0$ ) separates the data points into two classes, where the position of the data points is dependent on their respective weights and bias.

$$w^T x + b = 0 \quad (1)$$

Originally, support vector machines were introduced for linear binary classifications. However, they can produce non-linear solutions as well. Kernel functions can add more features in the higher dimension to separate the data into classes [19]. SVMs implement a data-driven algorithm approach and they are most efficient with datasets that have smaller samples compared to the variables, which is why they are popular in the field of medicine for applications in prognosis and diagnosis of diseases [10]. In this

SLR, SVMs are present in 7 out of the 33 selected studies, as listed in Table 1.

For dementia prognosis research [10], SVMs are employed for binary classifications tasks to classify if mild cognitively impaired patients developed the Alzheimer's disease [10]. SVMs are also used for other medical classification problems like predicting urinary tract infections [63], diagnosis of pulmonary hypertension [64] and predicting the optimal intensive care outcome [65]. By tuning a SVM with appropriate kernel functions and optimization algorithms, and using *fruit fly optimization algorithm*, the classification accuracy and processing time was better for medical datasets of breast cancer, diabetes, Parkinson and thyroid diseases [12]. Huang, S., et al. (2018) reviews that SVMs are largely being used in cancer genomic research leading the discovery of new cancer biomarkers, new drug targets, and a better understanding of cancer driver genes [19]. Due to SVMs' versatility in binary classification tasks and the liberty of choosing appropriate parameters specific to medical data, SVMs prove to be beneficial for even future cancer genomic applications.

#### 4.2. Clustering

Clustering is way of finding similarities and patterns within a set of data and organize them into clusters [37]. Without the presence of labels, clustering algorithm groups unlabeled data into meaningful clusters in such a way that similar data is categorized into separate clusters [44]. A typical clustering algorithm involves the steps of pattern representation, definition of the proximity measure depending on the domain, and clustering [37]. The input data that needs to be clustered can be numerical, categorical or mix of both types [66]. Clustering has achieved great success in finding structure in medical datasets [66]. In terms of proximity measure, several types of metrics are defined depending on the domain the data is from. For k-means clustering, the measure of Euclidean distance is incorporated for calculating the distance between the data points. The groups are then decided for the data points that are closer, having more similarity [66]. Clustering is based on different techniques, some of which are hierarchical, and partitioning [45], which are used to group the data into homogeneous clusters using empirical measures of similarity [46].

In this SLR, clustering methods are present in 5 out of the 33 selected studies, as listed in Table 1. Clustering methods like k-means are widely used to classify unstructured and unlabeled medical and biological data of gene expressions to protein domains [67]. C-means clustering, a small variant to k-means clustering, has shown strong results for classifying UCI medical datasets [58] of breast cancer and liver disease [13]. Here, fuzzy c-means clustering method, in which a single data observation can belong to more than one cluster, is used as a pre-classifier to group the data before the clusters are used to train a more advanced and tweaked clustering model. C-means clustering method follows the same procedure as k-means, whereas the number of clusters are fixed.

A class-based clustering method proposed by Yelipe, U., et al. (2018) [20] called class-based clustering classifier (CBCC) is based on computing the Euclidean distance and fuzzy measures to classify the medical records into clusters for 9 benchmark datasets like Australian, Cleveland, Ecoli, German, Hepatitis, Wine, Iris, Pima and Wisconsin datasets [59]. Using a similar strategy for data imputation in the aforementioned datasets using Euclidean and fuzzy measures, the authors also came up with a class-based clustering imputation technique which combined with CBCC brought better results than other popular classifiers like SVM and K-nearest neighbors.

#### 4.2.1. K-means clustering

K-means clustering is a robust unsupervised clustering method. K-means algorithm uses the criteria of squared error, like Euclidean distance measure, for calculating the distance between the data points for the process of grouping. K-means follows the typical process of clustering with first initializing the data into random clusters. The randomly assigned clusters are given a cluster center point. Each data point moves closer to a cluster center using a distance measure in each iteration of the algorithm. It, then, keeps on reassigning the clusters based on the distance measure (similarity) between the data points until it reaches a predefined threshold for a convergence criterion by minimizing the dissimilarity. Simultaneously, cluster centers are also reassigned based on the current cluster memberships [37].

The algorithm is built to minimize a cost function  $J(C)$ , seen in Eq. (2), where  $i$  is the number of data points and  $j$  is the cluster. Euclidean distance is used to measure the similarity between the data points and  $\hat{\mu}_j$  denotes mean vector distance associated with the cluster  $j$ , which can also be viewed as the center of the cluster  $j$ .

$$J(C) = \sum_{j=1}^K \sum_{C(i)=j} \|x_i - \hat{\mu}_j\|^2 \quad (2)$$

Clustering like k-means, can achieve better results if there is more information of the domain available in addition to the raw data itself [47]. This methodology is essential for highly complex domain like medicine since the high-dimensional nature of the data. Another modification to k-means clustering which has proven to be highly effective in achieving better results is global k-means algorithm [48]. This approach takes the dataset in an online stream and the cluster center is added sequentially through a deterministic global search procedure. K-means algorithm is implemented for the whole dataset in that manner  $N$  times (dataset of size  $N$ ).

In this SLR, K-means clustering methods are present in 2 out of the 33 selected studies, as listed in Table 1. In [13], authors present a modification to K-means clustering method, C-means clustering. The paper shows strong results for breast cancer and liver disease diagnosis using C-means clustering. Khanmohammadi, S. et al. (2017) in the 2nd paper, [66] proposed an improved k-means clustering algorithm which computed better results for 10 popular and common public medical datasets from UCI 2016 [58], including breast cancer dataset, Parkinson dataset, lung cancer dataset among many others. This approach produced compellingly better results compared to other clustering algorithms.

#### 4.3. Decision trees (DT)

Another machine learning model which takes a supervised learning approach, most popular for binary classification, is decision trees, which interpolates learned knowledge from a dataset into a tree which is governed by if-then rules [10]. Each node in the tree represents the learning variable which recursively checks how accurately can each node classify the labeled data, by calculating the information gain and entropy of each node. This learning process leads to selecting the best node as the parent node and the children nodes carry the possible values of the selected input data [10]. The whole process is recursively repeated until there can be no more splits. With no prior knowledge, DTs can find the most appropriate features for a particular dataset.

$$Entropy(S) = - \sum_{i=1}^c p_i \log_2 p_i \quad (3)$$

In Eq. (3),  $p_i$  refers to the probability of an element present in class  $i$ . This formula is used to compute the entropy of the data, which measures the certainty or disorder. Entropy, along with information gain, helps deciding where the data should be split to have a balanced distribution of classes, i.e., a perfect classification. The higher the entropy, higher the level of disorder, leading to the fact that there is high variance in the dataset. The role of a decision tree would be to reduce the level of disorder or uncertainty. Furthermore, Eq. (3) can then be used to compute the information gain of a particular feature in the dataset, as seen in Eq. (4).

$$Gain(S, A) = Entropy(S) - Entropy(S/A) \quad (4)$$

The entropy measure in dataset  $S$  given feature  $A$  is subtracted from the entropy of dataset  $S$  before any changes are made, to measure the reduction in entropy of  $S$  given an additional information about  $A$ . The higher the information gain, the more information a feature contains for the classification of the dataset into the available classes.

Entropy can measure the variance in a dataset which can further help calculating the information gain. Information gain is equal to the amount of entropy removed. More the information gain, the better is the split for the classification task. Neither of entropy or information gain alone can determine a perfect split in the dataset.

In this SLR, DTs are present in 6 out of the 33 selected studies, as listed in Table 1. DTs are being used in the medical field especially for prognostic studies [10], disease prediction [3,68], prediction of hypertension [64], intensive care outcome prediction [65], and heart disease prediction [69].

#### 4.4. Random forests

Random Forests (RFs) are collection of decision trees which together vote for the classification task, the majority 'votes' determine the respective class for the unknown data point [39,70]. The random forest finds the average of predictions of the trees in it. For a random forest with  $m$  trees with individual weights  $W_j$  the prediction is calculated using Eq. (5). The deeper a tree grows, it causes more overfitting in the training process, i.e., for a small change in the input, there is a higher or bigger variance [24]. RFs takes in vectorized input and each decision tree tries to classify a different part of the input vector. RFs combine the functionality of the decision trees by having the input vector go through each decision tree in the forest, and having each decision tree classify the input vector from the specific part of the vector they get as an input [24]. Then, the RF chooses, either the highest 'voted' class as the classification outcome for the input vector or it takes the average of all the 'votes'. Since, RFs incorporate numerous decision trees, the variance (overfitting) problem from one specific decision tree is, hence, avoided. The voting rule can be changed if there is a huge class imbalance [49].

$$\hat{y} = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n W_j(x_i, x') \quad (5)$$

As seen in Fig. 9, the RFs consists of three different decision trees which takes different parts of the same subset of the dataset as the input vector, to classify the data into class A or class B. RFs can achieve high classification accuracy, compared to standard classification techniques, whose base technique is voting for the class [71]. RFs are also good at handling class imbalance, given that the imbalance is not extraordinarily high [39]. RFs are used in different medical domains including gene selection for gene expression studies [50]. In this SLR, RFs are present in 5 out of the 33 selected studies, as listed in Table 1.

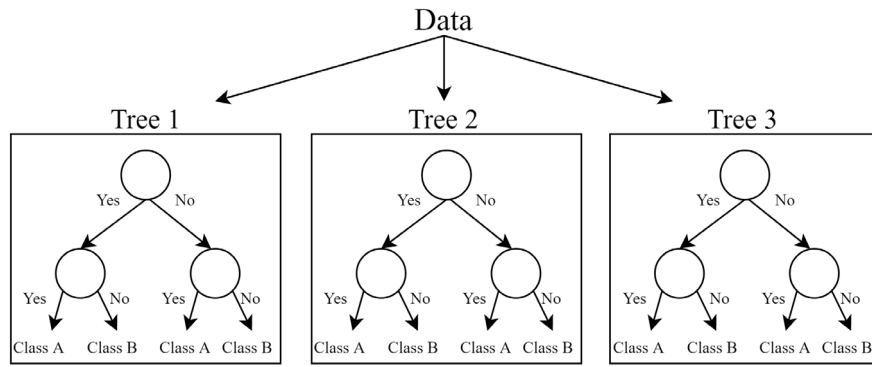


Fig. 9. Illustration of a random forests with three different decision trees [24].

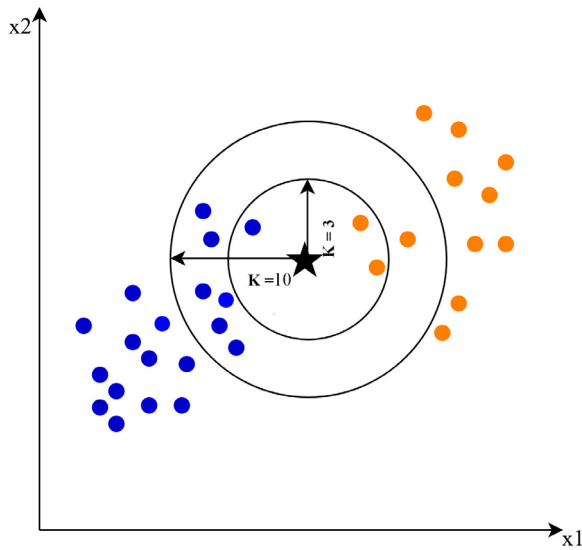


Fig. 10. Illustration of KNN algorithm [24]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To solve the problem of high class imbalance when using RFs, Zhu, M., et al. (2018) [71] proposed a weight voting RF model, where every vote of each classifier (decision tree) is multiplied by a weight to reinforce the authenticity of the respective classifier [71]. This helps with the issue of class imbalance, since with this method, the lowest majority classes will get a weighted proportional identification compared to the higher majority classes. This model was tested on the UCI 2016 medical datasets [58], like Mammographic dataset, Diagnostic Wisconsin Breast Cancer dataset, and Colic dataset, which possess great class imbalance, for which the model was able to get better accurate results in identifying both majority and minority classes.

Wang, J., et al. (2015) [72] used RF model with numerous decision trees with no correlation with each other to avoid overfitting in the prediction of medical expenditures of diagnosed diabetics in the USA. According to Goto, T., et al. (2018) [73], RFs performed really well for prediction the disposition of asthma or chronic obstructive pulmonary disease patients in the emergency department (ED) at the hospitals. This classification task involved two clinical outcomes of either critical care or hospitalization among the ED patients with the aforementioned conditions. Although, for the prediction of critical care another ML technique – boosting – performed the best, but for the prediction of the hospitalization prediction, RF achieved the highest accuracy.

#### 4.5. K-Nearest Neighbors

K-Nearest Neighbors (KNN) is another classification technique that works on the idea of assigning the label of a classified data point to an unclassified data point nearest to it [27]. Starting with the unclassified class data point as the input vector in the feature space, it is assigned to the class in which majority of its K nearest data points belong to [10], seen in Fig. 10.

In Fig. 10, the 'star' data point is assigned to 'orange' class when  $K = 3$  since the majority of data points when  $k = 3$  are of class 'orange'. Whereas, when  $K = 10$ , the unknown data point is classified as 'blue' since the majority of its 10 nearest data points belong to 'blue' class.

To measure the distance between the data points, like any other clustering algorithm, KNN mostly employs Euclidean distance ( $D_E$ ) (Eq. (6)), or Manhattan distance ( $D_M$ ) (Eq. (7)) measures. These measures are used when the data points are the nature of continuous variables, but for categorical variables Hamming distance ( $D_H$ ) measure (Eq. (8)), where if  $x = y$ ,  $D_H$  is 0, otherwise  $D_H$  will be greater than 0) is used.

$$D_E = \left( \sum_{i=1}^k (x_i - y_i)^2 \right)^{1/2} \quad (6)$$

$$D_M = \sum_{i=1}^k |x_i - y_i| \quad (7)$$

$$D_H = \sum_{i=0}^k |x_i - y_i|^2 \quad (8)$$

KNN being one of the more straight-forward ML classifications techniques to employ although, since it considers all features to be of equal weight, it is not an appropriate algorithm to be used for data with numerous attributes [10]. The best strategy would be to use feature selection methods with KNN if data has large number of superfluous attributes.

Fuzzy KNN algorithm, introduced by Keller, J.M. and Gray, M.R. [38], tackles the problem of each labeled sample being given the same priority in selecting the class membership for the unknown or unclassified data point. In this tweaked version of the base KNN algorithm, the unknown data points are assigned a sample vector instead of a class. In this way, the data point does not get automatically assigned to a class just on the basis of the membership measure of the nearest neighbors, instead it gets assigned to a sample vector which is further examined and the data point is classified after a satisfactory class is chosen [38]. For example, if a vector is assigned a membership of 0.9 in one class and 0.1 membership for another, the vector will be satisfyingly assigned to the first class. But, if a vector gets 0.85 membership in one class, and 0.65 and 0.04 memberships in other classes, there



is no clear classification in this case. Even though, the third class can be disregarded relatively justifiably, but there would need to be further examination to get a true classification for the vector.

In this SLR, KNNs are present in 3 out of the 33 selected studies, as listed in Table 1. Zhang, Z. (2016) [74] provides mathematical evidence that preparing the medical data to be used for a KNN algorithm is of high importance. Setting hyperparameters specific to the medical data with the right value for K is essential, which can be verified by checking the diagnostic performance of the KNN model. The authors claim that the 30% of the times KNN was used medical datasets surveyed in [24], KNN achieved best accuracy results. Chen, M., et al. (2017) [3] used structured patient data from hospitals which included demographics of the patients, living habits, examination results like blood tests, and diseases, with KNN, Naïve Bayes, and decision tree ML models to predict the risk of cerebral infraction disease.

#### 4.6. Naïve Bayes

Another classification machine learning technique, Naïve Bayes is a data classification technique where there is a prediction of a class based on a instance from a set of features describing that instance, given the information about the other instances. Naïve Bayes bases its foundation on the Bayes theorem. Bayes theorem uses a simple yet effective rule that calculates the probability of an event based on the information gained related to that event, as mathematically shown in Eq. (9).

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (9)$$

In this Equation,  $P(x|c)$  is the conditional probability of event  $c$  based on the information known about  $x$ . Calculating the probability of each feature based on the information gained from other features, helps build independence among the features, i.e., the features are not correlated to one another until to the point of interdependently contributing to probability [69]. Naïve Bayes is an important technique to increase the classification accuracy by eliminating the unrelated features [69].

Naïve Bayes, thus, can be used effectively for feature selection [75]. It has shown substantial results for text classification in the medical field, as a powerful pre-processing task, which helps improve scalability, efficiency, and accuracy of the text classifier [75].

In this SLR, Naïve Bayes methods are present in 3 out of the 33 selected studies, as listed in Table 1. For medical field, selecting domain-specific features is highly important and thus effective when using medical datasets, Naïve Bayes proves to be extremely simple yet effective, since its sensitivity to the feature selection. Kukar, M. et al. [12] compared Naïve Bayes with KNN and other classifiers in prognosis of the femoral neck fracture recovery to find that Naïve Bayes showed the highest accuracy.

#### 4.7. Deep learning

A highly-incorporated machine learning method in the medical field is DL. DL is a tool that is being substantially used for image acquisition and image interpretation over ML in the fields of general imaging and computer vision [76]. Complementary to the structure of ML models, DL models implement higher number of layers to create abstraction, extract deeper features, understand complex concepts by integrating simpler concepts, and generate more refined predictions from the data [31,77].

Deep learning has contributed significantly to medical research in recent years [76], mainly for image classification, object detection and image segmentation. Traditional learning approaches highly depend on hand-crafted features to be implemented for the learning process, for example lung cancer tumor detection requires hand crafted structural features to be

extracted [78]. But, to satisfy the diversity in the patient-to-patient data with numerous features, the novel ML or in this case DL techniques trump the traditional learning techniques due to former's ability to sift through various features and perform analysis on big data [76].

Due to popularization of DL, researchers have understood how to apply DL with multimodality data. Single modalities like audio, video, image or text are efficiently used with ML and DL. DL now also uses more than one modality together (e.g., audio-visual or AV data). Jiquan N., et al. (2019) created models that work on AV Letters datasets [79] for audio-visual speech classification [30]. Multimodal DL is also being used for various applications in the field of health informatics and healthcare. Most of medical data is produced using more than one channels like magnetic resonance images which is easily handled by DL models like convolutional neural networks (CNN) [80].

DL models have more layers than a traditional ML model, CNN being one of the best representations of a deep layered learning model [31]. A typical CNN architecture, like VGG16, has 16 layers, each layer referred to as a feature map. The feature map of the input layer is a 3D matrix of pixel intensities for different color channels (e.g. RGB – red, green, blue). The feature map of any internal layer is an induced multi-channel image, whose 'pixel' can be viewed as a specific feature. Every neuron is connected with a small portion of adjacent neurons from the previous layer (receptive field). Different types of transformations [26,34,35] can be conducted on feature maps, such as filtering and pooling. Filtering (convolution) operation convolutes a filter matrix (learned weights) with the values of a receptive field of neurons and takes a activation function (such as sigmoid, rectified linear unit or ReLU) to obtain final responses. ReLU function, for instance, would return the positive part of the argument that is passed through it, as seen in Eq. (10). Pooling operation, such as max pooling, average pooling, L2-pooling and local contrast normalization [36], summaries the responses of a receptive field into one value to produce more robust feature descriptions.

$$ReLU(x) = \max(0, x) \quad (10)$$

##### 4.7.1. Neural Networks (NNs)

NNs form the basis of mainly all deep learning models. NNs comprises of neurons which comprises of an activation function, weight and bias parameters. The input values get linearly computed with the weight and the bias in the neurons. Element-wise non-linearity is then performed on the result by a activation function, like sigmoid function, as shown in Eq. (11), where  $w$  and  $b$  are the weight and the bias respectively. These transformations occur in the neurons in the first few layers of the network, followed by fully-connected layers network, like a multi-layered feed-forward propagation network [43,81]. NNs can solve any function in existence, simple or complicated [82]. Ensembles of similar neural networks can be implemented to minimize the prediction error on previously unseen data points [42].

$$a = \sigma(w^T x + b) \quad (11)$$

Backpropagation is a ML technique of fine-tuning the weights in the neurons of artificial neural networks using gradient descent. Using the usual feedforward technique, constituting of error function and learning rate, the model learns the patterns in the raw data in the training phase. Consecutively, after feedforwarding step, the model calculates a gradient of the error function with respect to the weights of the multi-layer stack of modules going into the network. The model, then, internally tries to go against the gradient descent to reduce the error between the desired output and the calculated output to ultimately predict the representation in the classification or regression problem [31].

This technique uses a sigmoidal activation function due to its design compatibility [41].

With the complexity and versatility that NNs possess, they are extensively used in the medical field. In this SLR, NNs are present in 6 out of the 33 selected studies, as listed in Table 1. An ingeniously engineered NN is proposed by [83], which predicts the compressive strength (CS) of femoral trabecular bone in patients suffering from severe osteoarthritis. Using a *tanh* activation function for the hidden layers and linear activation function for the output values, this NN also uses a backpropagation algorithm with the mean square error cost function to gauge the difference between the CS values and the actual CS values.

Wang, Z., et al. (2019) [84] developed an effective and efficient multi-instance neural network for medical diagnosis classification tasks, specifically trained on low-quality and incomplete datasets of outpatient records. The two datasets included Traditional Chinese Medicine dataset for the diagnosis of meridian obstruction and Western Medicine dataset for the diagnosis of schizophrenia relapse. Using a multi-instance NN, they captured inter-instance features and then extracted the key features to be trained and used for the final classification in their model. Their novel model showed better results than traditional ML models for this task of medical diagnosis especially using incomplete and low-quality data. Furthermore, T. Le (2019) [13], proposed a combination of NN with k-means clustering algorithm to improve diagnostic accuracy which was tested on breast cancer and liver disease classification datasets from UCI 2016 [58].

#### 4.7.2. Convolutional Neural Networks (CNNs)

Traditionally, for a successful object detection, there are three sub-tasks, namely, informative region selection, feature extraction and classification [85]. With deep learning and CNNs, these sub-tasks become easier since the development in the architecture and computational resources is growing.

CNNs are a strong DL tool that have been widely incorporated in the field of computer vision and medical image analysis [31]. CNNs have a deep network of convolution, pooling and fully-connected layers that operate on images to extract features in a completely supervised environment, thus helping humans by automatically devising the features [86]. CNNs can learn complex features from available data, mostly image data, which allows them to detect complex disease specific features [86]. For instance, CNNs are successfully used in prediction of chronic diseases outbreak in the most disease-frequent cities of China using hospital image data [3].

As discussed earlier in the DL section, CNN architecture is designed to have many layers for the training process which occurs in a sequential manner involving numerous iterations (or epochs) to learn selective features and optimize the network parameters with each epoch [87]. For a common image classification task of predicting a diseased region in a human body, the network is given specific pixels of a image to be trained upon and find discriminative features of the diseased areas. A typical CNN structure would constitute a number of convolution and pooling layer sets with fully-connected layers attached at the end for the classification.

As shown in Fig. 11, a convolution layer consists of several convolution kernels, which compute different feature maps for the input images. More specifically, each neuron of a feature map is connected to the neurons in the previous layer. The new feature map is obtained by first convolving the input with a trained kernel from the previous feature map and then applying an element-wise nonlinear activation function, typically *sigmoid*, *tanh* or *ReLU* activation, on the convolved results [88]. Several feature maps are obtained by using several convolution kernels. CNNs can non-linearly combine a plethora of descriptors, often more than the data observations [9]. Sequentially, activation

functions then help introduce the non-linearity in the network, which is appropriate for the CNNs to detect non-linear features in the data [32]. Pooling layers, then, reduce the dimensionality of the features by reducing the number of free parameters and introduce spatial invariance in the network [86]. Finally, the fully connected layers connect all the neurons from the previous layers to the current layer to extract high-level reasoning [88]. An appropriate score is measured which is then matched in the output layer to its respective class.

In this SLR, CNNs are present in 13 out of the 33 selected studies, as listed in Table 1, making DL the highest frequently used technique in our survey combined with NNs. Since, CNNs are so versatile in their structure and usage, researchers are developing domain-specific CNNs to help in medical application like thoraco-abdominal lymph node (LN) detection [89] and human liver lesion classification [90]. Diseased or abnormal human tissues are detected by classification methods by usually using hand-crafted features fed to the CNN model [78]. The classifier compares the mapping of the feature vectors to the probability of the presence of a lesion. Greenspan, H. (2016) [78] surveys various CNNs used for image classification and segmentation like detection of colonic polyps on CT colonography, cerebral microbleeds, and interstitial lung diseases in CT scans and histopathological images.

Cheplygina, V. (2019) [91] conducted a study on various types of CNNs like CIFAR10, AlexNet, Caltech101, etc., to understand the importance of input data, whether domain specific input data would affect the performances of CNNs compared with non-domain specific data. Gao X., et al. (2015) [92] proposed a deep CNN network for grading nuclear cataracts in slit-lamp images, where CNN filters were pre-trained, which as they claim helps in extracting the contextual information making the factor of image size used in the CNNs independent.

Setio, A.A.A., et al. (2016) [93] used a multi-stream CNN network to classify areas of interest in a chest CT scan into a cancerous or non-cancerous nodules. Various augmented images of the chest CT scan were fed into the CNN in different streams simultaneously, only to combine all the features together for the fully connected layers to finish the classification. Nie, D., et al. (2016) [21] used a 3-D CNN to assess survival in patients suffering from high-grade gliomas. These human brain images depicting the high-grade gliomas, used for the training of this 3-D CNN, were also 3-dimensional. The diverse nature of the functionality of CNNs carries forward to not only 2-dimensional images, which constitute the majority of the medical images present in the datasets in the domain of image classification, and segmentation, but also to 3-dimensional images.

Furthermore, Kamnitsas, K., et al. (2017) [94] also created a 11-layered deep CNN network to handle the complex task of brain lesion segmentation. For 3-dimensional images, they ingeniously incorporated the processing of adjacent images passes to one pass, which would cut down on the computational burden of 3-dimensional convolution. M. J. J. P. van Grinsven, et al. (2016) [86] created an appropriately tweaked CNN to lower the training complexity for medical image dataset by dynamically selecting misclassified negative samples during training for detection of hemorrhages in color fundus images. Chang, Y., et al. (2019) [95] created a two-stage CNN model for medical noise removal via image decomposition. Generally, to denoise medical data, most of the denoising models either completely disregard the image or work on deleting the residual noise in the images. The CNN model proposed by Chang, Y., et al. (2019), treats both image and noise equally in terms of CNN training and by doing that, the image sub-network for training gets enough information from the noise sub-network to be considered for the denoising task. Using a CNN model for medical image denoising task proved to be much better than the current state-of-the-art methods used for image denoising.

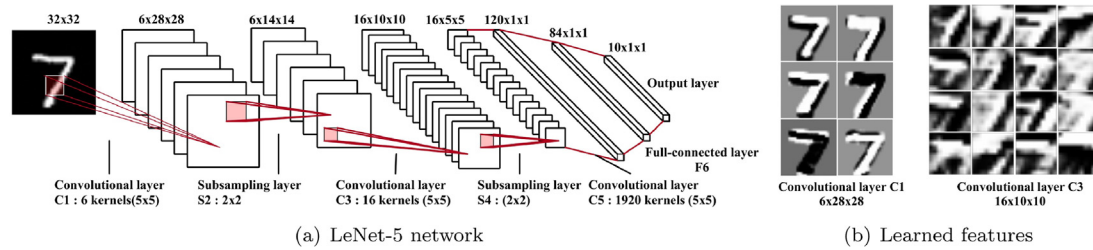


Fig. 11. Structure of LeNet-5 Network [32].

Image localization is also a big task in the medical imaging domain. Localizing areas or points of interest from infected organs and tissues is a preliminary and essential task for cancer detection [81]. Payer, C., et al. (2016) [22] created a CNN network to directly identify a location by training the CNN with the location as the ground truth input data represented by a Gaussian function. For localization tasks in 3-dimensional images, the complexity of the problem increases than for 2-dimensional images, exponentially [81]. Zheng Y., et al. (2015) [23] worked on this problem by incorporating 3 different 1-dimensional convolutions in a CNN instead of 1 3-dimensional convolution for carotid artery bifurcation detection in CT data. Whereas, Ghesu F. C., et al. (2016) [96] worked on reducing the complexity of a 3-dimensional localization task by using a sparse adaptive deep neural network powered by marginal space learning of the aortic valve in 3D transeophageal echocardiogram.

#### 4.8. Other recent medical applications of ML and DL

A novel machine learning algorithm, *Semi-supervised Set Covering Machine*, was proposed by Wang, Z., et al. [29] to detect the presence of coronary angiogram results and ovarian cancer diagnoses from unstructured text in the General Practice Research Database. This technique used for only coronary angiogram results and ovarian cancer diagnoses can be further used for different disease-related datasets of electronic health records from the medical field, since the researchers did not use linguistic rules, rather developed the algorithm based on testing the model on pre-classified datasets for the purpose of binary classification.

Dallora, A. L., et al. (2017) [10] surveyed various traditional and novel ML techniques used for prognosis of dementia. They found support vector machines, decision trees, Bayesian networks and artificial neural networks were the more prevalent ML techniques being used in this particular field of medical research. Taylor, R. A., et al. (2018) [63] proposed several different ML algorithms like support vector machines, logistic regression, neural networks, and RFs to predict urinary tract infection in a large diverse dataset from emergency department patients. These models, as they claim, were used for their ability to model non-linear associations, and resiliency to overfitting.

Leha, A., et al. (2019) [64] applied five ML algorithms including RF of classification trees, RF of regression trees, lasso penalized logistic regression, boosted classification trees, and support vector machines for prediction of pulmonary hypertension in a dataset constituting patients with invasively determined pulmonary artery pressure (PAP) with corresponding echocardiographic estimations of PAP obtained within 24 h. Meiring, C., et al. (2018) [65] used different ML techniques to predict the mortality rate of patients in intensive care unit over time. Among the ML techniques used for this research problem, the most effective techniques for the highest prediction percent were deep neural networks, support vector machines, and RFs, which over-time outperformed logistic regression models. Hasnain, Z., et al. (2019) [97] used support vector machines, bagged support vector machines, K-nearest neighbors, RFs, and a mixture

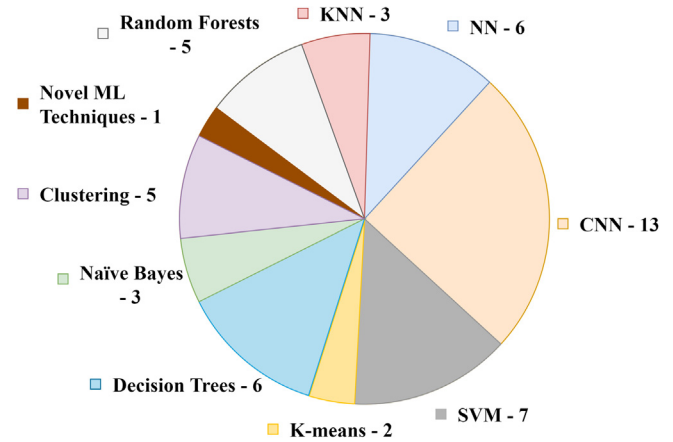


Fig. 12. Distribution of machine learning techniques in surveyed papers.

of combinations among these ML techniques to predict post-cystectomy recurrence and survival in bladder cancer patients. Ramalingam V. V., et al. (2018) [69] surveyed various traditional ML techniques including support vector machines, random forests, K nearest neighbors, and decision trees to predict heart- and cardiovascular-related diseases.

## 5. Results

This review's objective is to analyze the role of machine learning (and deep learning) in medical research in recent years. We surveyed 33 sources where a ML or DL technique is used for a medical application, as seen in detail in Fig. 12. The criteria that are followed for choosing these papers are mentioned in the Methods Section 2. Furthermore, the papers that resulted in the survey criteria are tabulated in Table 1, along with the general theme of the medical application and the ML technique used in the paper.

Table 1 tracks the author and the publication year of the article for which the ML technique was used. There were two major trends that were derived from this analysis. First, most frequent technique used in the field of medicine is CNN. The tasks that are followed for choosing these papers are mentioned in the Methods Section 2. Furthermore, the papers that resulted in the survey criteria are tabulated in Table 1, along with the general theme of the medical application and the ML technique used in the paper.

Furthermore, traditional ML techniques like SVM, K-means, and clustering techniques, are used to analyze the tabular numeric data from popular datasets like UCI datasets, KEEL benchmark datasets. As expected, these techniques with domain-

**Table 2**

Formulae.

<b>Machine learning Techniques</b>		
<b>Support Vector Machines</b>		
1	Linear Kernel	$w^T x + b = 0$
2	Polynomial	$(x^T x_i + 1)^p$
3	Radial basis-function	$\exp\left(-\frac{1}{2\sigma^2} \ x - x_i\ ^2\right)$
4	Multiquadric kernel	$(\ x - \bar{x}\ ^2 + \sigma^2)^{1/2}$
5	Laplace kernel	$\exp(-\ x - \bar{x}\ ^2 / \sigma^2)^{1/2}$
6	Laplace RBF kernel	$\exp(-\sigma \ x - \bar{x}\ )$
7	Power exponential kernel	$\exp(-\ x - \bar{x}\ ^2 / 2\sigma^2)^d$
8	ANOVA RB kernel	$\sum_{k=1}^n \exp(-\sigma(x^k - \bar{x}^k)^2)^d$
9	Cauchy kernel	$1/(1 + \ x - \bar{x}\ ^2/\sigma)$
10	Two-layer perceptron	$\tanh(\beta_0 x^T x_i + \beta_1)$
11	Sigmoid Kernel	$\exp(-\sigma \ x - \bar{x}\ )$
<b>Clustering</b>		
12	K-means clustering	$J(C) = \sum_{j=1}^K \sum_{C(i)=j} \ x_i - \hat{\mu}_j\ ^2$
13	Fuzzy C-means Clustering	$\sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2$
14	Mountain Clustering	$m_{new}(v) = m(v) - m(c_i) \exp\left[-\frac{\ v - c_i\ ^2}{2\beta^2}\right]$
15	Subtractive Clustering	$D_i = D_i - D_{c_i} \exp\left[-\frac{\ x_i - x_{c_i}\ ^2}{(\frac{r_b}{2})^2}\right]$
16	Probabilistic Clustering	$\Pr\left(\frac{X}{C}\right) = \prod_{i=1:N} \sum_{j=1:k} r_j \Pr\left(\frac{x_i}{c_j}\right)$
17	Robust Clustering Algorithm for Categorical Data	$E = \sum_{j=1:k}  C_j  \times \sum_{x,y \in C_j} \text{link}(x, y) /  C_j ^{1+2f(\theta)}$
<b>Decision Trees</b>		
18	Entropy	$\text{Entropy}(S) = - \sum_{i=1}^c p_i \log_2 p_i$
19	Gain	$\text{Gain}(S, A) = \text{Entropy}(S) - \text{Entropy}(S/A)$
<b>Random Forests</b>		
20	Random Forests	$\hat{y} = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n W_j(x_i, x')$
<b>KNN</b>		
21	Euclidean Distance	$D_E = \left(\sum_{i=1}^k (x_i - y_i)^2\right)^{1/2}$
22	Manhattan Distance	$D_M = \sum_{i=1}^k  (x_i - y_i)^2 $
23	Hamming distance	$D_H = \sum_{i=0}^k  (x_i - y_i)^2 $
<b>Naïve Bayes</b>		
24	Naïve Bayes	$P(c x) = \frac{P(x c)P(c)}{P(x)}$
<b>Deep learning</b>		
25	CNN	$X_k^l = \sigma(W_k^{l-1} * X^{l-1} + b_k^{l-1})$
26	RNN	$h_t = \sigma(W_{x_t} + R h_{t-1} + b)$
27	Neural Network - Multi-layer perceptron	$f(x; \Theta) = \sigma(W^T \sigma(W^T \dots \sigma(W^T x + b)) + b)$
28	ReLU Function	$f(x) = x^+ = \max(0, x)$
29	Sigmoid Function	$a = \sigma(w^T x + b)$
30	Stochastic Gradient Descent	$\arg \min_{\Theta} - \sum_{n=1}^N \log[P(y_n   x_n; \Theta)]$

specific modifications use the aforementioned datasets as a standard to validate their accuracy and efficiency. But, in fact, more

DL techniques were seen to be better in prediction tasks for these datasets. Finally, for the analysis of more complex and

**Table 3**

Table of References.

References #	Authors	Year	Publication Venue	Google Scholar Index	SJR	Quartile	# of Citations
[1]	Scruggs , S. B., Watson, K., et al.	2015	Circulation Research	122	306	Q1	51
[2]	P. Groves, B. Kayyali, et al.	2016					593
[3]	M. Chen, Y. Hao, et al.	2017	IEEE Access	89	56	Q1	104
[4]	M. Chen, S. Mao, et al.	2014	Mobile Network Applications	30	79	Q1	2471
[5]	Stuart Russell and Peter Norvig.	2016	Book				33864
[6]	Mitchell, T. M.	1997					1612
[7]	Michie, D., Spiegelhalter, D. J., et al.	1994					3873
[8]	Rajkomar, A., Dean, J., et al.	2019	The New England journal of medicine	352	933	Q1	143
[9]	Obermeyer, Z., & Emanuel, E. J.	2016	The New England journal of medicine	352	933	Q1	762
[10]	Dallora, A. L., Eivazzadeh, S., et al.	2017	PLOS One	94	195	Q1	16
[11]	Alpaydin, E.	2020					7407
[12]	Shen, L., Chen, H., et al.	2016	Knowledge-Based Systems	79	94	Q1	113
[13]	T. Le	2019	IEEE Access	89	56	Q1	2
[83]	Shaikhina, T., & Khovanova, N. A.	2017	Artificial Intelligence in Medicine	34	74	Q1	53
[66]	Khanmo hammadi, S., Adibeig, N., et al.	2017	Expert Systems with Applications	105	162	Q1	80
[67]	Wiwie, C., Baumbach, J., et al.	2015	Nature Methods	152	257	Q1	134
[63]	Taylor, R. A., Moore, C. L., et al.	2018	PLOS One	94	195	Q1	24
[64]	Leha, A., Hellenkamp, K., et al.	2019	PLOS One	94	195	Q1	57
[91]	Cheplygina, V.	2019	Elsevier: Current Opinion in biomedical eng	14			4
[78]	Greenspan, H., Van Ginneken, B., et al.	2016	IEEE transactions on medical imaging	74	195	Q1	811
[28]	Halevy A, Norvig P, et al.	2009	IEEE Intelligent Systems		111	Q1	1096
[14]	Lee, C. H., & Yoon, H. J.	2017	Kidney Research and Clinical Practice	17	13	Q2	185
[15]	Kitchenham B, Charters S.	2007					4070
[16]			Google Scholar				
[17]	SCImago, (n.d.). SJR						
[18]	Suykens, J., Vandewalle, J.	1999	Neural Processing letters	26	43	Q2	4583
[19]	Huang, S., Cai, N., et al.	2018	Cancer genomics and proteomics		32	Q2	75
[20]	Yelipe, U. R., Porika, S., et al.	2018	Computers & Electrical Engine				
[21]	Nie D., Zhang H., et al.	2016	Notes in computer science		374	Q2	122
[22]	C. Payer, D. Stern, et al.	2016	Notes in computer science		374	Q2	116
[23]	Zheng Y., Liu D., et al.	2015	Notes in computer science		374	Q2	105
[24]	Uddin, S., Khan, A., et al.	2019	BMC Medical Informatics and Decision Making	38	59	Q2	
[25]	Schaekerman n, M., et al.	2019	ACM	80	132	Q1	2
[26]	A. Krizhevsky, I. Sutskever, and G. E. Hinton	2012	Communica -tions of the ACM	70	189	Q1	57520
[27]	Cover T, Hart P.	1967	IEEE transactions on information theory	87	249	Q1	11770
[29]	Wang, Z., Shah, A. D., et al.	2012	PLOS One	94	195	Q1	79
[30]	Ngiam, J., Khosla, A., et al.	2011					2082

(continued on next page)



Table 3 (continued).

References #	Authors	Year	Publication Venue	Google Scholar Index	SJR	Quartile	# of Citations
[31]	Y. LeCun, Y. Bengio, G. Hinton	2015	Nature	368	1096	Q1	23 141
[32]	Y. LeCun, L. Bottou, Y., et al.	1998	Proceedings of the IEEE	80	250	Q1	24 780
[33]	Chun-Fu Lin and Sheng-De Wang	2002	IEEE Transactions on Neural Networks	96	180	Q1	691
[34]	M. Oquab, L. Bottou, I. Laptev, J. Sivic et al.	2014	Proceedings of the NIPS				55
[35]	M. Oquab, L. Bottou, et al.	2014	IEEE Conference on Computer Vision & Pat Rec	240	264	Q1	2188
[36]	K. Kavukcuoglu, R. Fergus, Y. LeCun et al.	2009	IEEE Conference on Computer Vision & Pat Rec	240	264	Q1	364
[37]	A.K. Jain, M.N. Murty, et al.	1999	ACM Computing Surveys	80	132	Q1	16 185
[38]	Keller JM, Gray MR, et al.	1985	IEEE Transactions on Systems, Man, and Cybernetics	100	156	Q1	2169
[39]	Breiman L.	2001	Machine Learning	40	135	Q1	56 732
[40]	Brereton, P., Kitchenham, B. A., et al.	2007	Journal of Systems and Software	52	94	Q1	1524
[41]	Hecht-Nielsen, R.	1992					3619
[42]	Hansen, L. K., & Salamon, P.	1990	IEEE trans. on patt. ana. and machine intelli.	127	326	Q1	4346
[43]	Beale, H. D., Demuth, H. B., & Hagan, M. T.	1996					9788
[44]	Xu, R., & Wunsch, D.	2008					1262
[45]	Rokach, L., & Maimon, O.	2019	ACM	80	132	Q1	2
[46]	Johnson, S. C.	1967	Psychometrika	26	68	Q1	5266
[47]	Wagstaff, K., Cardie, C., et al.	2001	ICML	135	54	Q1	2828
[48]	Likas, A., Vlassis, N., et al.	2003	Pattern Recognition	79	180	Q1	1688
[49]	Liaw, A., & Wiener, M.	2002					11 829
[50]	Díaz-Uriarte, R., & De Andres, S. A.	2006	BMC bioinformatics	61	183	Q1	2406
[51]	Murphy, K. P.	2012					6725
[52]	Cios, K. J., & Moore, G. W.	2002	Artificial intelligence in medicine	34	74	Q1	748
[53]	Lavrac, N.	1999	Artificial intelligence in medicine	34	74	Q1	63
[54]	Grossman, R. L., Health, A. P.	2016	N Engl J Med	352	933	Q1	430
[55]	Vayena, E., Blasimme, A., & Cohen, I. G.	2018	PLOS Medicine	94	195	Q1	58
[56]	Dinov, I. D.	2016	Gigasience	46	32	Q1	85
[57]			Nature Reviews Cardiology	74	105	Q1	222
[58]			UCI datasets				
[59]			KEEL benchmark datasets				
[60]	Shen, D., Wu, G., et al.	2017	Annual Review of Biomedical Engineering		120	Q1	974
[61]	M. Zhou, J. Scott, et al.	2018	American Journal of Neuroradiology	61	162	Q1	33
[62]	Chen, J. H., & Asch, S. M.	2017	The New England journal of medicine	352	933	Q1	265
[65]	Meiring, C., Dixit, A., et al.	2018	PLOS One	94	195	Q1	8
[68]	Nilashi, M., Ibrahim, et al.	2017	Computers and Chemical Engineering - Elsevier		124	Q1	51
[69]	Ramalingam, V V, Dandapath, Ayantan, et al.	2018	International Journal of Eng & Tech	13	44	Q1	6
[70]	Erickson, B. J., Korfiatis, P., et al.	2017	Radiographics	51	151	Q1	278

(continued on next page)

**Table 3** (continued).

References #	Authors	Year	Publication Venue	Google Scholar Index	SJR	Quartile	# of Citations
[71]	M. Zhu et al.	2018	IEEE Access	89	56	Q1	4
[72]	Jing Wang, Leiyu Shi,	2015	International Journal for Quality in Health Care	32	81	Q1	
[73]	Goto, T., Camargo, C. A., et al.	2018	American Journal of Emergency Medicine	38	78	Q1	15
[74]	Zhang Z.	2016	Annals of translational medicine	46	31	Q1	72
[75]	K. J. D'souza and Z. Ansari	2018	IEEE Conf on Cloud Computin Emerg Mark	11			0
[76]	Razzak, M. I., Naz, S., et al.	2018	Springer International Publishing - Book				165
[77]	Goodfellow, I., Bengio, Y., & Courville, A.	2016	MIT press				15 723
[79]			AV Letters Dataset				
[80]	Ravi, D., Wong, C., et al.	2016	IEEE journal of biomedical and health informatics	66	104	Q1	548
[81]	Litjens, G., Kooi, T., et al.	2017	Medical Image Analysis - Elsevier	61	113	Q1	2957
[82]	Nielsen, M.A.	2015	Determination press				1945
[84]	Z. Wang, J. Poon, et al.	2019	International Joint Conference on Neu Net	36	59	Q1	1
[85]	Zhong-Qiu Zhao, Peng Zheng, et al.	2018	Archive				233
[86]	Van Grinsven, M., Van Ginneken, B., et al.	2016	IEEE Transactions on Medical Imaging	74	195	Q1	200
[87]	Schmidhuber	2015	Neural Networks - Elsevier	57	128	Q1	8214
[88]	Gu, J., Wang, Z., et al.	2018	Pattern Recognition	79	180	Q1	791
[89]	Shin, H. C., Roth, H. R., et al.	2016	IEEE Transactions on Medical Imaging	74	195	Q1	1735
[90]	Frid-Adar, M., Diamant, I.,	2018	Archive				173
[92]	Gao X., Lin S., et al.	2015	IEEE Transactions on Biomedical Engineering	73	172	Q1	74
[93]	A.A.A. Setio, F. Ciompi, et al.	2016	IEEE Transactions on Medical Imaging	74	195	Q1	529
[94]	Kamnitsas, K., Ledig, C.,	2017	Medical Image Analysis - Elsevier	61	113	Q1	1134
[95]	Y. Chang, L. Yan, et al.	2019	IEEE Transactions on Instru and Measur	51	100	Q1	0
[96]	F.C. Ghesu, E. Krubasik, et al.	2016	IEEE Transactions on Medical Imaging	74	195	Q1	88
[97]	Hasnain, Z., Mason, J., et al.	2019	PLOS One	94	195	Q1	4

domain-specific data, like medical image data (CT scans) of brain lesion segmentation, carotid artery bifurcation detection, and brain tumor detection and medical diagnosis and prognosis data, we observed DL to be preferred more.

## 6. Conclusion

With the emergence of big data, machine learning techniques are used to learn, analyze, and extrapolate details in the medical research. This survey provides a comprehensive overview of the ML techniques including support vector machines, K-means clustering, decision trees, random forests, Naïve Bayes, K nearest neighbors, neural networks, and convolution neural networks, that are being used for various types of medical data and applications in the recent years. A systemic literature review was conducted to search and select research articles from highly reputed and relevant journals in the recent years. The papers selected to survey and document in this survey paper are from journals with a high SJR and google scholar h-5 index. Other

criteria like year of publication and number of citations were also taken into consideration in the selection process.

Further this survey paper discusses about the current influx of medical data and the challenges that medical data faces with respect to its analysis. Medical data, due to its complex nature, needs intricate analysis techniques. ML techniques, but more recently DL models, have proven to understand medical image, and multi-variate numerical data. More recently, neural networks (and convolution neural networks) have been used for medical image segmentation and classification like brain lesion segmentation, lymph node detection, classification of nodules in human chest from CT scans, brain tumor classification, among many other applications. For other medical applications like medical diagnosis, and dementia prognosis, DL models are also being used more. In fact, for tabular datasets from UCI and KEEL, we have seen applications being shifted from traditional ML techniques to DL. Using empirical evidence, we have concluded that, in the recent years, DL has been preferred more (19 out of 33 papers) by the researchers to work with medical data. We believe that this survey paper can be of high importance for the researchers

to observe relevant details and trends about the current research in the computer science and the medical research.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The authors would like to extend their gratitude to the research team in the DaTALab at Lakehead University for their support; in particular, Abhijit Rao, Dhivya Chandrasekaran and Punardeep Sikka for their feedback and revisions on this publication. We would also like to thank Lakehead University, CASES, and the financial support from NOAMA RP-544 C-18-19 and NSERC RGPIN/05377-2017.

### Appendix A

See Table 2.

### Appendix B

See Table 3.

### References

- [1] S.B. Scruggs, K. Watson, A.I. Su, H. Hermjakob, J.R. Yates III, M.L. Lindsey, P. Ping, Harnessing the heart of big data, *Circ. Res.* 116 (7) (2015) 1115–1119.
- [2] P. Groves, B. Kayyali, D. Knott, S.V. Kuiken, The 'big Data' revolution in Healthcare: Accelerating Value and Innovation, Center for US Health System Reform Business Technology Office, 2016.
- [3] M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, Disease prediction by machine learning over big data from healthcare communities, *Ieee Access* 5 (2017) 8869–8879.
- [4] M. Chen, S. Mao, Y. Liu, Big data: A survey, *Mob. Netw. Appl.* 19 (2) (2014) 171–209.
- [5] S. Russell, P. Norvig, *Artificial intelligence: a modern approach*, 2002.
- [6] T.M. Mitchell, *Machine Learning*, McGraw-Hill, New York, 1997.
- [7] D. Michie, D.J. Spiegelhalter, C. Taylor, et al., *Machine learning*, *Neural Stat. Classif.* 13 (1994) (1994) 1–298.
- [8] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, *New Engl. J. Med.* 380 (14) (2019) 1347–1358.
- [9] Z. Obermeyer, E.J. Emanuel, Predicting the future—big data, machine learning, and clinical medicine, *New England J. Med.* 375 (13) (2016) 1216.
- [10] A.L. Dallora, S. Eivazzadeh, E. Mendes, J. Berglund, P. Anderberg, Machine learning and microsimulation techniques on the prognosis of dementia: A systematic literature review, *PLoS One* 12 (6) (2017) e0179804.
- [11] E. Alpaydin, *Introduction to Machine Learning*, MIT Press, 2020.
- [12] L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang, H. Li, B. Yang, D. Liu, Evolving support vector machines using fruit fly optimization for medical data classification, *Knowl.-Based Syst.* 96 (2016) 61–75.
- [13] T.-L. Le, Fuzzy C-means clustering interval type-2 cerebellar model articulation neural network for medical data classification, *IEEE Access* 7 (2019) 20967–20973.
- [14] C. Lee, H. Yoon, Medical big data: promise and challenges, *Kidney Res. Clin. Pract.* 36 (1) (2017) 3–11.
- [15] B. Kitchenham, S. Charters, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, Citeseer, 2007.
- [16] Google scholar, [https://scholar.google.com/citations?view\\_op=top\\_venues&hl=en](https://scholar.google.com/citations?view_op=top_venues&hl=en).
- [17] SCImago journal and country rank [portal], <https://www.scimagojr.com>.
- [18] J.A. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.* 9 (3) (1999) 293–300.
- [19] S. Huang, N. Cai, P.P. Pacheco, S. Narrandes, Y. Wang, W. Xu, Applications of support vector machine (SVM) learning in cancer genomics, *Cancer Genomics-Proteomics* 15 (1) (2018) 41–51.
- [20] U. Yelipe, S. Porika, M. Golla, An efficient approach for imputation and classification of medical data values using class-based clustering of medical records, *Comput. Electr. Eng.* 66 (2018) 487–504.
- [21] D. Nie, H. Zhang, E. Adeli, L. Liu, D. Shen, 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 212–220.
- [22] C. Payer, D. Štern, H. Bischof, M. Urschler, Regressing heatmaps for multiple landmark localization using CNNs, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2016, pp. 230–238.
- [23] Y. Zheng, D. Liu, B. Georgescu, H. Nguyen, D. Comaniciu, 3D deep learning for efficient and robust landmark detection in volumetric data, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 565–572.
- [24] S. Uddin, A. Khan, M.E. Hossain, M.A. Moni, Comparing different supervised machine learning algorithms for disease prediction, *BMC Med. Inf. Decis. Mak.* 19 (1) (2019) 1–16.
- [25] M. Schaeckermann, G. Beaton, M. Habib, A. Lim, K. Larson, E. Law, Understanding expert disagreement in medical data analysis through structured adjudication, *Proc. ACM Hum.-Comput. Interact.* 3 (CSCW) (2019) 1–23.
- [26] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM* 60 (6) (2017) 84–90.
- [27] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory* 13 (1) (1967) 21–27.
- [28] A. Halevy, P. Norvig, F. Pereira, The unreasonable effectiveness of data, *IEEE Intell. Syst.* 24 (2) (2009) 8–12.
- [29] Z. Wang, A.D. Shah, A.R. Tate, S. Denaxas, J. Shawe-Taylor, H. Hemingway, Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning, *PLoS One* 7 (1) (2012) e30412.
- [30] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, in: *ICML*, 2011.
- [31] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *nature* 521 (7553) (2015) 436–444.
- [32] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, *Proc. IEEE* 86 (11) (1998) 2278–2324.
- [33] C.-F. Lin, S.-D. Wang, Fuzzy support vector machines, *IEEE Trans. Neural Netw.* 13 (2) (2002) 464–471.
- [34] M. Oquab, L. Bottou, I. Laptev, J. Sivic, et al., Weakly supervised object recognition with convolutional neural networks, in: *Proc. of NIPS*, 2014, pp. 1545–15963.
- [35] M. Oquab, L. Bottou, I. Laptev, J. Sivic, Learning and transferring mid-level image representations using convolutional neural networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1717–1724.
- [36] K. Kavukcuoglu, M. Ranzato, R. Fergus, Y. LeCun, Learning invariant features through topographic filter maps, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2009, pp. 1605–1612.
- [37] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, *ACM Comput. Surv.* 31 (3) (1999) 264–323.
- [38] J.M. Keller, M.R. Gray, J.A. Givens, A fuzzy k-nearest neighbor algorithm, *IEEE Trans. Syst. Man Cybern.* (4) (1985) 580–585.
- [39] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [40] P. Brereton, B.A. Kitchenham, D. Budgen, M. Turner, M. Khalil, Lessons from applying the systematic literature review process within the software engineering domain, *J. Syst. Softw.* 80 (4) (2007) 571–583.
- [41] R. Hecht-Nielsen, Theory of the backpropagation neural network, in: *Neural Networks for Perception*, Elsevier, 1992, pp. 65–93.
- [42] L.K. Hansen, P. Salamon, Neural network ensembles, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (10) (1990) 993–1001.
- [43] H.B. Demuth, M.H. Beale, O. De Jess, M.T. Hagan, *Neural Network Design*, Martin Hagan, 2014.
- [44] R. Xu, D. Wunsch, *Clustering*, Vol. 10, John Wiley & Sons, 2008.
- [45] O. Maimon, L. Rokach, *Data Mining and Knowledge Discovery Handbook*, Springer, 2005.
- [46] S.C. Johnson, Hierarchical clustering schemes, *Psychometrika* 32 (3) (1967) 241–254.
- [47] K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl, et al., Constrained k-means clustering with background knowledge, in: *ICML*, Vol. 1, 2001, pp. 577–584.
- [48] A. Likas, N. Vlassis, J.J. Verbeek, The global k-means clustering algorithm, *Pattern Recogn.* 36 (2) (2003) 451–461.
- [49] A. Liaw, M. Wiener, et al., Classification and regression by randomforest, *R news* 2 (3) (2002) 18–22.
- [50] R. Díaz-Uriarte, S.A. De Andres, Gene selection and classification of microarray data using random forest, *BMC Bioinf.* 7 (1) (2006) 3.
- [51] K.P. Murphy, *Machine Learning: a Probabilistic Perspective*, MIT Press, 2012.
- [52] K.J. Cios, G.W. Moore, Uniqueness of medical data mining, *Artif. Intell. Med.* 26 (1–2) (2002) 1–24.
- [53] N. Lavrač, Machine learning for data mining in medicine, in: *Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making*, Springer, 1999, pp. 47–62.

- [54] R.L. Grossman, A.P. Heath, V. Ferretti, H.E. Varmus, D.R. Lowy, W.A. Kibbe, L.M. Staudt, Toward a shared vision for cancer genomic data, *New Engl. J. Med.* 375 (12) (2016) 1109–1112.
- [55] E. Vayena, A. Blasimme, I.G. Cohen, Machine learning in medicine: Addressing ethical challenges, *PLoS Med.* 15 (11) (2018) e1002689.
- [56] I.D. Dinov, Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data, *Gigascience* 5 (1) (2016) s13742–016.
- [57] J.S. Rumsfeld, K.E. Joynt, T.M. Maddox, Big data analytics to improve cardiovascular care: promise and challenges, *Nat. Rev. Cardiol.* 13 (6) (2016) 350.
- [58] UCI datasets, <https://archive.ics.uci.edu/ml/datasets.php>.
- [59] KEEL datasets, <https://sci2s.ugr.es/keel/datasets.php#sub1>.
- [60] D. Shen, G. Wu, H.-I. Suk, Deep learning in medical image analysis, *Ann. Rev. Biomed. Eng.* 19 (2017) 221–248.
- [61] M. Zhou, J. Scott, B. Chaudhury, L. Hall, D. Goldgof, K.W. Yeom, M. Iv, Y. Ou, J. Kalpathy-Cramer, S. Napel, et al., Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches, *Am. J. Neuroradiol.* 39 (2) (2018) 208–216.
- [62] J.H. Chen, S.M. Asch, Machine learning and prediction in medicine—beyond the peak of inflated expectations, *New England J. Med.* 376 (26) (2017) 2507.
- [63] R.A. Taylor, C.L. Moore, K.-H. Cheung, C. Brandt, Predicting urinary tract infections in the emergency department with machine learning, *PLoS One* 13 (3) (2018) e0194085.
- [64] A. Leha, K. Hellenkamp, B. Unsöld, S. Mushemi-Blake, A.M. Shah, G. Hasenfuß, T. Seidler, A machine learning approach for the prediction of pulmonary hypertension, *PLoS One* 14 (10) (2019) e0224453.
- [65] C. Meiring, A. Dixit, S. Harris, N.S. MacCallum, D.A. Brealey, P.J. Watkinson, A. Jones, S. Ashworth, R. Beale, S.J. Brett, et al., Optimal intensive care outcome prediction over time using machine learning, *PLoS One* 13 (11) (2018) e0206862.
- [66] S. Khanmohammadi, N. Adibeig, S. Shanehbandy, An improved overlapping k-means clustering method for medical applications, *Expert Syst. Appl.* 67 (2017) 12–18.
- [67] C. Wiwie, J. Baumbach, R. Röttger, Comparing the performance of biomedical clustering methods, *Nat. Methods* 12 (11) (2015) 1033.
- [68] M. Nilashi, O. bin Ibrahim, H. Ahmadi, L. Shahmoradi, An analytical method for diseases prediction using machine learning techniques, *Comput. Chem. Eng.* 106 (2017) 212–223.
- [69] V. Ramalingam, A. Dandapath, M.K. Raja, Heart disease prediction using machine learning techniques: a survey, *Int. J. Eng. Technol.* 7 (2.8) (2018) 684–687.
- [70] B.J. Erickson, P. Korfiatis, Z. Akkus, T.L. Kline, Machine learning for medical imaging, *Radiographics* 37 (2) (2017) 505–515.
- [71] M. Zhu, J. Xia, X. Jin, M. Yan, G. Cai, J. Yan, G. Ning, Class weights random forest algorithm for processing class imbalanced medical data, *IEEE Access* 6 (2018) 4641–4652.
- [72] J. Wang, L. Shi, Prediction of medical expenditures of diagnosed diabetics and the assessment of its related factors using a random forest model, *MEPS 2000–2015*, *In. J. Qual. Health Care* 32 (2) (2020) 99–112.
- [73] T. Goto, C.A. Camargo Jr, M.K. Faridi, B.J. Yun, K. Hasegawa, Machine learning approaches for predicting disposition of asthma and COPD exacerbations in the ED, *Am. J. Emerg. Med.* 36 (9) (2018) 1650–1654.
- [74] Z. Zhang, Introduction to machine learning: k-nearest neighbors, *Ann. Translational Med.* 4 (11) (2016).
- [75] K.J. D'souza, Z. Ansari, Big data science in building medical data classifier using Naïve Bayes model, in: 2018 IEEE International Conference on Cloud Computing in Emerging Markets, CCEM, IEEE, 2018, pp. 76–80.
- [76] M.I. Razzak, S. Naz, A. Zaib, Deep learning for medical image processing: Overview, challenges and the future, in: *Classification in BioApps*, Springer, 2018, pp. 323–350.
- [77] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, *Deep Learning*, Vol. 1, (2) MIT press Cambridge, 2016.
- [78] H. Greenspan, B. Van Ginneken, R.M. Summers, Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1153–1159.
- [79] AV letters database, <http://www.ee.surrey.ac.uk/Projects/LiLiR/datasets/avletters1/index.html>.
- [80] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.-Z. Yang, Deep learning for health informatics, *IEEE J. Biomed. Health Inf.* 21 (1) (2016) 4–21.
- [81] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (2017) 60–88.
- [82] M.A. Nielsen, *Neural Networks and Deep Learning*, Vol. 2018, Determination press San Francisco, CA, 2015.
- [83] T. Shaikhina, N.A. Khovanova, Handling limited datasets with neural networks in medical applications: A small-data approach, *Artif. Intell. Med.* 75 (2017) 51–63.
- [84] Z. Wang, J. Poon, S. Sun, S. Poon, Attention-based multi-instance neural network for medical diagnosis from incomplete and low quality data, in: 2019 International Joint Conference on Neural Networks, IJCNN, IEEE, 2019, pp. 1–8.
- [85] Z.-Q. Zhao, P. Zheng, S.-t. Xu, X. Wu, Object detection with deep learning: A review, *IEEE Trans. Neural Netw. Learn. Syst.* 30 (11) (2019) 3212–3232.
- [86] M.J. Van Grinsven, B. van Ginneken, C.B. Hoyng, T. Theelen, C.I. Sánchez, Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1273–1284.
- [87] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Netw.* 61 (2015) 85–117.
- [88] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, et al., Recent advances in convolutional neural networks, *Pattern Recognit.* 77 (2018) 354–377.
- [89] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1285–1298.
- [90] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification, *Neurocomputing* 321 (2018) 321–331.
- [91] V. Cheplygina, Cats or CAT scans: transfer learning from natural or medical image source data sets? *Curr. Opinion Biomed. Eng.* 9 (2019) 21–27.
- [92] X. Gao, S. Lin, T.Y. Wong, Automatic feature learning to grade nuclear cataracts based on deep learning, *IEEE Trans. Biomed. Eng.* 62 (11) (2015) 2693–2701.
- [93] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.J. Van Riel, M.M.W. Wille, M. Naqibullah, C.I. Sánchez, B. van Ginneken, Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1160–1169.
- [94] K. Kamnitsas, C. Ledig, V.F. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, *Med. Image Anal.* 36 (2017) 61–78.
- [95] Y. Chang, L. Yan, M. Chen, H. Fang, S. Zhong, Two-stage convolutional neural network for medical noise removal via image decomposition, *IEEE Trans. Instrum. Meas.* 69 (6) (2019) 2707–2721.
- [96] F.C. Ghesu, E. Krubasik, B. Georgescu, V. Singh, Y. Zheng, J. Hornegger, D. Comaniciu, Marginal space deep learning: efficient architecture for volumetric image parsing, *IEEE Trans. Med. Imaging* 35 (5) (2016) 1217–1228.
- [97] Z. Hasnain, J. Mason, K. Gill, G. Miranda, I.S. Gill, P. Kuhn, P.K. Newton, Machine learning models for predicting post-cystectomy recurrence and survival in bladder cancer patients, *PLoS One* 14 (2) (2019) e0210976.