

Healthcare Predictive Analytics Using Machine Learning and Deep Learning Techniques: A Survey

Mohammed Badawy (✉ mbadawy@pg.cu.edu.eg)

Cairo University

Nagy Ramadan

Cairo University

Hesham Ahmed Hefny

Cairo University

Research Article

Keywords: Healthcare Prediction, Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), Medical Diagnosis

Posted Date: July 28th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1885746/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Aim

This paper aims to present a comprehensive survey of existing machine learning and deep learning approaches utilized in healthcare prediction, as well as identify inherent obstacles to applying these approaches in the healthcare prediction domain.

Background

Healthcare prediction has been a significant factor in saving human lives in recent years. In the domain of healthcare, there is a rapid development of intelligent systems for analyzing complicated relationships among data and transforming them into real information for use in the prediction process. Consequently, artificial intelligence is rapidly transforming the healthcare industry. Thus comes the role of systems depending on machine learning as well as deep learning in the creation of steps that diagnose and predict diseases, whether from clinical data or based on images, that provide tremendous clinical support by simulating human perception and can even diagnose diseases that are difficult to detect by human intelligence.

Methods

The studies discussed in this paper have been presented in journals published by IEEE, Springer, and Elsevier. Machine learning, deep learning, healthcare, surgery, cardiology, radiology, hepatology, and nephrology are some of the terms used to search for these studies. The studies chosen for this survey are concerned with the use of machine learning as well as deep learning algorithms in healthcare prediction.

Results

A total of 40 working papers were selected and the methodology for each paper was clarified.

Conclusion

This paper presents a comprehensive survey as well as the current challenges in healthcare prediction. studies have shown that artificial intelligence plays a significant role in diseases diagnosing.

1. Background

Each day, human existence evolves, yet the health of each generation either improves or deteriorates. There are always uncertainties in life. Occasionally encounter a large number of individuals with fatal

health problems due to the late detection of diseases. Concerning the adult population, chronic liver disease would affect more than 50 million individuals worldwide. However, if the sickness is diagnosed early, it can be stopped. Disease prediction based on machine learning can be utilized to identify common diseases at an earlier stage. Currently, health is a secondary concern, which has led to numerous problems. Many patients cannot afford to see a doctor, and others are extremely busy and on a tight schedule, yet ignoring recurring symptoms for an extended length of time can have significant health repercussions [1].

A medical diagnosis is a form of problem-solving and is a crucial and significant issue in the actual world. Illness diagnosis is the process of translating observational evidence into disease names. The evidence comprises data received from evaluating a patient and substances generated from the patient; illnesses are conceptual medical entities that detect anomalies in the observed evidence [2].

Diseases are a global issue, thus medical specialists and researchers are exerting their utmost efforts to reduce disease-related mortality. In recent years, predictive analytic models are playing a pivotal role in the medical profession as a result of the increasing volume of healthcare data from a wide range of disparate and incompatible data sources. Nonetheless, processing, storing and analyzing the massive amount of historical data and the constant inflow of streaming data created by healthcare services has become an unprecedented challenge utilizing traditional database storage [3, 4, 5].

The concept of medical care is used to stress the organization and administration of curative care, which is a subset of healthcare [6]. The ecology of medical care was first introduced by White in 1961. White also proposed a framework for perceiving patterns of health concerning symptoms experienced in particular populations of interest, along with individual's choices in getting medical treatment. In this framework, it is able to calculate the proportion of the population who used medical services over a specific time period. The "ecology of medical care" theory has become widely accepted in academic circles over the past few decades [7].

Healthcare is the collective effort of society to ensure, provide, finance, and promote health. In the 20th century, there was a significant shift toward the ideal of wellness and the prevention of sickness and incapacity. The delivery of health care services entails organized public or private efforts to aid persons in regaining health and preventing disease and impairment [8]. Healthcare can be described as standardized rules that help evaluate actions or situations that affect decision-making [9].

Healthcare is a multidimensional system. The basic goal of healthcare is to diagnose and treat illnesses or disabilities. A healthcare system key component are health experts (physicians or nurses), health facilities (clinics, hospitals that provide medications and other diagnostic services), and a funding institution to support the first two [10].

With the introduction of systems based on computers, the digitalization of all medical records and the evaluation of clinical data in healthcare systems have come to be a widespread routine practice. The phrase "electronic health records" was chosen by the Institute of Medicine, a division of the National

Academies of Sciences, Engineering, and Medicine in 2003 to define the records that continued to enhance the healthcare sector for benefit of both the patients and physicians. Electronic Health Records (EHR) are "computerized medical records for patients that include all information in an individual's past, present, or future which occur in an electronic system used to capture, store, retrieve, and link data primarily to offer healthcare and health-related services," according to Murphy, Hanken, and Waters [10].

Daily, healthcare services produce an enormous amount of data, getting it increasingly complicated to analyze and handle it using "conventional ways." Using machine learning and deep learning, this data may be properly analyzed to generate actionable insights. In addition, genomics, medical data, social media data, environmental data, and other data sources can be used to supplement healthcare data. Figure 1 provides a visual picture of these data sources. The four key healthcare applications that can benefit from machine learning are prognosis, diagnosis, therapy, and clinical workflow, as outlined in the following section [11].

The long-term investment in developing novel technologies based on machine learning as well as deep learning techniques to improve the health of individuals via the prediction of future events reflects the increased interest in predictive analytics techniques to enhance healthcare. Clinical predictive models, as they have been formerly referred to, assisted in the diagnosis of persons with an increased probability of disease. These prediction algorithms are utilized to make clinical treatment decisions and counsel patients based on some patient characteristics [12].

Artificial Intelligence (AI) is a scientific field that successfully integrates computer science and large datasets to solve problems. It requires an understanding of computing to build tools and devices that offer desired behavior [13]. Figure 2 depicts machine learning and deep learning as subsets of AI.

Medical personnel are usually facing new problems, changing tasks, and frequent interruptions as a result of the system's dynamism and scalability. This variability often makes disease recognition a secondary concern for medical experts. Moreover, clinical interpretation of medical data is a challenging task from an epistemological point of view. This not only applies to professionals with extensive experience but also representatives, such as young physician assistants, with varied or little experience. The limited time available to medical personnel, the speedy progression of diseases, and the fluctuating patient dynamics all the time make diagnosis a particularly complex process. However, a precise method of diagnosis is critical to ensuring speedy treatment and thus ensuring patient safety [14].

1.1 Machine Learning

Machine learning (ML) is a subfield of AI that aims to develop predictive algorithms based on the idea that machines should have the capability to access data and learn on their own [15]. ML utilizes algorithms, methods, and processes to detect basic correlations within data and create descriptive and predictive tools that process those correlations. ML is usually associated with data mining, pattern

recognition, and deep learning. Although there are no clear boundaries between these areas and they often overlap, it is generally accepted that deep learning is a relatively new subfield of ML that uses extensive computational algorithms and large amounts of data to define complex relationships within data. As shown in Figure 3, ML algorithms can be divided into three categories: supervised learning, unsupervised learning, and reinforcement learning [16].

1.1.1 Supervised Learning

Supervised learning is an ML model for investigating the input-output correlation information of a system depending on a given set of training examples that are paired between the inputs and the outputs [17]. The model is trained with a labeled dataset. It matches how a student learns fundamental math from a teacher. This kind of learning requires labelled data with predicted correct answers based on algorithm output [18]. The most widely used supervised learning-based techniques include K-Nearest Neighbor, Naive Bayes, Support Vector Machines, Decision Trees, Random Forests, and Logistic Regression.

Linear Regression

Linear regression is a statistical method commonly used in predictive investigations. It succeeds in forecasting the dependent, output, variable (Y) based on the independent, input, variable (X). The connection between X and Y is represented as shown in eq. 1 assuming continuous, real, and numeric parameters.

$$Y = mX + c. \quad (1)$$

where m indicates the slope and c indicates the intercept. According to eq. 1, the association between the independent parameters (X) and the dependent parameters (Y) can be inferred. [19].

The advantage of linear regression is that it is straightforward to learn, and it is also easy to eliminate overfitting through regularization. One drawback of linear regression is that it is not convenient when it is applied to non-linear relationships. However, it is not recommended for most practical applications as it greatly simplifies real-world problems [20]. The implementation tools utilized in Linear Regression are Python, R, MATLAB, and Excel.

As shown in Figure 4, observations are highlighted in red, and random deviations' result (shown in green) from the basic relationship (shown in blue) between the independent variable (x) and the dependent variable (y) [21].

Logistic Regression

Logistic regression, also known as the logistic model, investigates the correlation between a large number of independent variables and a categorical dependent variable, and calculates the probability of an event by fitting the data to a logistic curve [22]. Discrete mean values must be binary, i.e., have only two outcomes: true or false, 0 or 1, yes or no, or either superscript or subscript. In logistic regression, categorical variables have to be predicted and classification problems to be solved. Logistic regression can be implemented utilizing various tools such as R, Python, Java, and MATLAB [19]. Logistic regression has many benefits, such as it shows the linear relationship between dependent and independent variables with the best results. It is also simple to understand. On the other hand, it can only predict numerical output, is not relevant to non-linear data, is sensitive to outliers [23].

Decision Tree

The Decision Tree (DT) is the most popular supervised learning methods used for classification. It combines the values of attributes based on their order either ascending or descending [24]. As a tree-based strategy, DT defines each path starting from the root by a data separating sequence until a Boolean conclusion is attained at the leaf node [25- 26]. DT is a hierarchical representation of knowledge interactions that contains nodes and links. When relations are employed to classify, nodes reflect purposes [27 - 28]. An example of DT is presented in Figure 5.

DTs have various drawbacks, such as increased complexity with increasing nomenclature, small modifications that may lead to a different architecture, and more processing time to train data [19]. The implementation tools used in DT are Python (Scikit-Learn), R Studio, Orange, KNIME, and Weka [23].

Random Forest

Random Forest (RF) It is a basic and most widely utilized algorithm that produces correct results most of the time. It may be utilized for classification and also regression. The program produces an ensemble of DTs and blends them [29].

In the RF classifier, the higher the number of trees in the forest, the more accurate the results. So, the RF has generated a collection of DTs called the forest and combined them to achieve more accurate prediction results. In RF, each DT is built only on a part of the given dataset and trained on approximations. The RF brings together several DTs to reach the optimal decision [19].

As indicated in Fig. 6. RF randomly selects a subset of features from the data and from each subset it generates n number of random trees [21]. RF will combine results from all DTs and provide them in the final output.

Two parameters are being used for tuning RF models: *mtry* - the count of randomly selected features to be considered in each division; and *ntree* - the model trees count. The *mtry* parameter has a trade-off: large values raise the correlation between trees but enhance the per-tree accuracy [30].

The RF works with a labeled dataset to do predictions and build a model. The final model is utilized to classify unlabeled data. The model integrates the concept of bagging with a random selection of traits to build variance-controlled DTs [31].

RF offers significant benefits. First, it can be utilized for determining the relevance of the variables in a regression and classification task [32,33]. This relevance is measured with a scale, based on the impurity drop at each node used for data segmentation [34]. Second, it automates missing values contained in the data and resolves the overfitting problem of DT. Finally, RF can efficiently handle huge data sets. On the other side, RF suffers from drawbacks such as it needs more computing and resources to generate the output results and it requires training effort due to the multiple DTs involved in it. The implementation tools used in RF are Python Scikit-Learn and R [19].

Support Vector Machine

The most popular supervised ML algorithm for classification issues and regression models is called Support Vector Machine (SVM). SVM is a linear model that offers solutions to issues that are both linear and nonlinear. as shown in Figure 7. Its foundation is the idea of margin calculation. The dataset is divided into several groups to build relations between them [19].

SVM is a statistical-based learning method that follows the principle of structural risk minimization and aims to locate decision bounds, also known as hyperplanes, that can optimally separate classes by finding a hyperplane in a usable N-dimensional space that explicitly classifies data points. [35, 36, 37]. SVM indicates the decision boundary between two classes by defining the value of each data point, in particular the support vector points placed on the boundary between the respective classes [38].

SVM has several advantages such as it works perfectly even with both semi-structured and unstructured data. Kernel trick is a strength point of SVM. Moreover, it can handle any complex problem with the right functionality and can also handle high-dimensional data. Furthermore, SVM generalization has less allocation risk. On the other hand, SVM has many downsides. Its model training time is increased on a large dataset. Choosing the right kernel function is also a difficult process. In addition, it is not working well with noisy data. Implementation tools used in SVM include SVMlight with C, LibSVM with Python, MATLAB or Ruby, SAS, Kernlab, Scikit-Learn, and Weka [23].

K - Nearest Neighbor

K-nearest neighbor (KNN) is an "instance-based learning" or non-generalized learning, which is often known as a "lazy learning" algorithm [39]. KNN is used for solving the classification problems. To anticipate the target label of the novel test data, KNN determines the distance of the nearest training data class labels with a new test data point in the existence of a K value, as shown in Figure 8. It then calculates the number of nearest data points using the K value and terminates the label of the new test data class. To determine the number of nearest-distance training data points, KNN usually sets the value of K among 0 and 1 [23].

KNN has many benefits such as it is sufficiently powerful if the size of training data is large. It is also simple and flexible with attributes and distance functions. Moreover, it can handle multi-class data sets. KNN has many drawbacks such as the difficulty of choosing the appropriate K value, it is very tedious to choose the distance function type for a particular dataset, and the computation cost being a little high due to the distance between all the training data points [31]. The implementation tools used in KNN are Python (Scikit-Learn), WEKA, R, KNIME, and Orange [23].

Naive Bayes

Naive Bayes (NB) focuses on the probabilistic model of Bayes' theorem and is simple to set up as the complex recursive parameter estimation is basically none, making it suitable for huge data sets [40]. NB determines the class membership degree based on a given class designation [41]. It scans the data once and thus classification is easy [42]. Simply, the NB classifier assumes that there is no relation between the presence of a particular feature in a class and the presence of any other characteristic. It is mainly targeted at the text classification industry [43].

NB has great benefits such as ease of implementation, can provide a good result even using fewer training data, can manage both continuous and discrete data, ideal to solve prediction of multiclass problems, and the irrelevant feature does not affect the prediction. NB, on the other hand, has the following drawbacks: it assumes that all features are independent which is not always viable in real-world problems, suffers the zero frequency problems, and the prediction of NB is not usually accurate. Implementation Tools: WEKA, Python, R Studio, and Mahout [19].

1.1.2 Unsupervised learning

Unlike supervised learning, there are no correct answers and no teachers in unsupervised learning [43]. It follows the concept that a machine can learn to understand complex processes and patterns on its own without external guidance. This approach is particularly useful in cases where experts have no knowledge of what to look for in the data and the data itself does not include the objectives. The machine predicts the outcome based on past experiences and learns to predict the real-valued outcome from the information previously provided, as shown in Figure 9.

Unsupervised learning is widely used in the processing of multimedia content, as clustering and partitioning of data in the lack of class labels is often a requirement [44]. Some of the most popular unsupervised learning-based approaches are k-means, Principal Component Analysis (PCA), and Apriori Algorithm.

k-means

The k means algorithm is the common portioning method [45] and one of the most popular unsupervised learning algorithms that deal with the well-known clustering problem. The procedure classifies a particular data set by a certain number of preselected (assuming k-sets) clusters [46]. The Pseudocode of the K-means algorithm is shown in Pseudocode 1.

Pseudocode 1: k-means Pseudocode

1. **Arrange** K points in the space represented by the clustered items. These points reflect the centroids of the first group.
2. **Set** each object of the group that has the nearest centroid.
3. After setting all the elements, the coordinates of the k centroids have to be recalculated.
4. **Repeat** Steps 2 and 3 until the centroids stop moving.

K means have several benefits such as being more computationally efficient than hierarchical grouping in case of large variables. It provides more compact clusters than hierarchical ones when small k is used. Also, the ease of implementation and comprehension of assembly results is another benefit. However, K-Means have disadvantages such as the difficulty of predicting the value of K. Also, as different starting sections lead to various final combinations, the performance is affected. It is accurate for raw points and local optimization, and there is no single solution for a given K value - so the average of the K value must be run multiple times (20-100 times) and then pick the results with the minimum J [20].

Principal Component Analysis

In modern data analysis, Principal component analysis (PCA) is an essential tool as it provides a guide for extracting the most important information from a dataset, compressing the data size by keeping only those important features without losing much information, and simplifying the description of a data set [47, 48].

PCA is frequently used to reduce data dimensions before applying classification models. Moreover, unsupervised methods, such as dimensionality reduction or clustering algorithms, are commonly used for

data visualizations, detection of common trends or behaviors, and decreasing the data quantity to name a few only [49].

PCA converts the 2D data into 1D data. This is done by changing the set of variables into new variables known as principal components (PC) which are orthogonal [24]. In PCA data dimensions are reduced to make calculations faster and easier. To illustrate how PCA works, let's consider an example of 2D data. When this data is plotted on a graph, it will take two axes. Applying PCA the data turns into 1D. This process is illustrated in Figure 10 [50].

Apriori

Apriori algorithm is considered an important algorithm, that was first introduced by R. Agrawal and R. Srikant, and published in [51,52].

The principle of the Apriori algorithm is to represent the filter generation strategy. It creates a filter element set $(k + 1)$ based on the repeated k element groups. Apriori uses an iterative strategy called planar search, where k item sets are employed to explore $(k+1)$ item sets. First, the set of repeating 1-items is produced by scanning the dataset to collect the number for each item, then collecting items that meet the minimum support. The resulting group is called L_1 . Then L_1 is used to find L_2 , the recursively set of two elements is used to find L_3 , and so on until no repeated k element groups are found. Finding every L_k needs a full dataset scan. To improve production efficiency at the level-wise of repeated element groups, a key property called the Apriori property is used to reduce the search space. Apriori property states that all non-empty subsets of a recursive element group must be iterative. A two-step technique is used to identify groups of common elements: join and prune activities [53].

Although it is simple, the Apriori algorithm suffers from several drawbacks. The main limitation is the costly wasted time to contain a large number of candidate sets with a lot of redundant item sets. It also suffers from low minimum support or large item sets and multiple rounds of data are needed for data mining which usually results in irrelevant items, in addition to difficulties in discovering individual elements of events [54, 55].

1.1.3. Reinforcement learning

Reinforcement learning (RL) is different supervised learning and unsupervised learning. It is a goal-oriented learning approach. RL is closely related to an agent (controller) that takes the responsibility for the learning process to achieve a goal. The agent, in particular, chooses actions, and as a result, the environment changes its state and returns rewards. Positive or negative numerical values are used as rewards. An agent's goal is to maximize the rewards accumulated over time. A job is a complete environment specification that identifies how to generate rewards [56]. Some of the most popular reinforcement learning-based algorithms are the Q-Learning algorithm and the Monte-Carlo Tree Search (MCTS).

Q-Learning

Q-Learning is a type of model-free RL. It can be considered an asynchronous dynamic programming approach. It enables agents to learn how to operate optimally in Markovian domains by exploring the effects of actions, without the need to generate domain maps [57]. It represented an incremental method of dynamic programming that imposed low computing requirements. It works through the successive improvement of the assessment of individual activity quality in particular states [58].

In information theory, Q-learning is strongly employed, and other related investigations are underway. Recently, Q-learning combined with information theory has been employed in different disciplines such as Natural Language Processing (NLP), pattern recognition, anomaly detection, and image classification [61, 62, 63,64]. Moreover, a framework has been created to provide a satisfying response based on the user's utterance using RL in a voice interaction system [65]. Furthermore, a high-resolution deep learning-based prediction system for local rainfall has been constructed [66].

The advantage of developmental Q-learning is that it is possible to identify the reward value effectively on a given multi-agent environment method as agents in ant Q-learning are interacting with each other. The problem with Q-learning is that its output can stuck in the local minimum as agents just take the shortest path [67].

Monte Carlo Tree Search

Monte Carlo Tree Search (MCTS) is an effective technique for solving sequential selection problems. Its strategy is based on a smart tree search that balances exploration and exploitation. MCTS presents random samples in the form of simulations and keeps activity statistics for better-educated choices in each future iteration. MCTS is a decision-making algorithm that is employed in searching trees-like huge complex regions. In such trees, each node refers to a state, which is also referred to as problem configuration, while edges represent transitions from one state to another [68].

The MCTS is related directly to cases that can be represented by a Markov decision process (MDP), which is a type of discrete-time random control process. Some modifications of the MCTS make it possible to apply it to Partially Observable Markov Decision Processes (POMDP) [69]. Recently, MCTS coupled with deep RL became the base of AlphaGo developed by Google DeepMind and documented in [70]. The basic MCTS method is conceptually simple, as shown in Figure 11.

Tree 1 is constructed progressively and unevenly. The tree policy is utilized to get the critical node of the current tree for each iteration of the method. The tree strategy seeks to strike a balance between exploration and exploitation concerns. Then, from the specified node, simulation 2 is run, and the search tree is then updated according to the obtained results. This comprises adding a child node that matches

the specified node's activity and updating its ancestor's statistics. During this simulation, movements are performed based on some default policy, which in its simplest case is to make uniform random movements. The benefit of MCTS is that there is no need to evaluate the values of the intermediate state, which significantly minimizes the amount of required knowledge in the field [72].

1.2 Deep Learning

Over the past decades, ML has had a significant impact on our daily lives with examples including efficient computer vision, web search, and recognition of optical characters. In addition, by applying ML approaches, AI at the human level has also been improved [73, 74, 75]. However, when it comes to the mechanisms of human information processing (such as sound and vision), the performance of traditional ML algorithms is far from satisfactory. The idea of Deep Learning (DL) was formed in the late 20th inspired by the deep hierarchical structures of human voice recognition and production systems. DL breaks have been introduced in 2006 when Hinton built a deep structured learning architecture called Deep Belief Network (DBN) [76].

The performance of classifiers using DL has been extensively improved with an increased amount of data compared to classical learning methods. Figure 12 shows the performance of classic ML algorithms and DL methods [77]. The performance of typical ML algorithms becomes stable when they reach the training data threshold, but DL upturns their performance as the amount of data increases [78].

DL (deep ML, or deep structured learning) is a subset of ML which involves a collection of algorithms attempting to represent high-level abstractions for data through a model that has complicated structures or otherwise, composed of numerous non-linear transformations. The most important characteristic of DL is the depth of the network. Another essential aspect of DL is the ability to replace handcrafted features generated by efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction [79].

DL has significantly advanced the latest technologies in a variety of applications, including machine translation, speech, and visual object recognition, NLP, and text automation, through the use of multi-layer Artificial Neural Networks (ANNs) [16].

Different DL designs in the past two decades give the enormous potential for employment in various sectors such as automatic voice recognition, computer vision, NLP, and bioinformatics. This section discusses the most common architectures of DL such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and Recurrent Convolution Neural networks (RCNNs) [80].

Convolutional Neural Network

CNNs are special types of neural networks inspired by the human visual cortex and used in computer vision. It is an automatic feed-forward neural network in which information transfers exclusively in the

forward direction [81]. CNN is frequently applied in face recognition, human organ localization, text analysis, and biological image recognition [82].

Since CNN was first created in 1989, it has done well in disease diagnosis over the past three decades [83]. Figure 13 depicts the general architecture of a CNN composed of feature extractors and a classifier. Each layer of the network accepts the output of the previous layer as input and passes it on to the next layer in feature extraction layers. A typical CNN architecture consists of three types of layers: convolution, pooling, and classification. There are two types of layers at the network's low and middle levels: convolutional layers and pooling layers. Even-numbered layers are used for convolutions, while odd-numbered layers are used for pooling operations. The convolution and pooling layers' output nodes are categorized in a two-dimensional plane called feature mapping. Each layer level is typically generated by combining one or more previous layers [84].

CNN has a lot of benefits, including a human optical processing system, greatly improved 2D and 3D image processing structure, and is effective in learning and extracting abstract information from 2D information. The max-pooling layer in CNN is efficient in absorbing shape anisotropy. Furthermore, they are constructed from sparse connections with paired weights and contain far fewer parameters than a fully connected network of equal size. CNNs are trained using a gradient-based learning algorithm and are less susceptible to the diminishing gradient problem because the gradient-based approach trains the entire network to directly reduce the error criterion, allowing CNNs to provide highly optimized weights [84].

Long Short Term Memory

LSTM is a special type of Recurrent Neural Networks (RNNs) with internal memory and multiplicative gates. Since the original LSTM introduction in 1997 by Sepp Hochreiter and Jürgen Schmidhuber, a variety of LSTM cell configurations have been described [92].

LSTM has contributed to the development of well-known software such as Alexa, Siri, Cortana, Google Translate, and Google voice assistant [93]. LSTM is an implementation of RNN with a special connection between nodes. The special components within the LSTM unit include the input, output, and forget gates. Figure 14 depicts a single LSTM cell.

LSTM is an RNN module that handles gradient loss problems. In general, RNN uses LSTM to eliminate propagation errors. This allows the RNN to learn over multiple time steps. LSTM is characterized by cells that hold information outside the recurring network. This cell enables the RNN to learn over many time steps. The basic principle of LSTMs is the state of the cell, which contains information outside the recurrent network. A cell is similar to a memory in a computer, which decides when data should be stored, written, read, or erased via the LSTM gateway [94]. Many network architectures use LSTM such as

bidirectional LSTM, hierarchical and attention-based LSTM, convolutional LSTM, autoencoder LSTM, network LSTM, cross-modal, and relational LSTM [95].

Bidirectional LSTM networks move the state vector forward and backward in both directions. This implies that dependencies must be taken into account in both temporal directions. As a result of inverse state propagation, the expected future correlations can be included in the network's current output [96]. investigates and analyses this because bidirectional LSTM networks encapsulate spatially and temporally scattered information and can tolerate incomplete inputs via a flexible cell-state vector propagation communication mechanism. Based on the detected gaps in data, this filtering mechanism reidentifies the connections between cells for each data sequence. Figure 15 depicts the architecture. A bidirectional network is used in this study to process properties from multiple dimensions into a parallel and integrated architecture [95].

Hierarchical LSTM networks solve multidimensional problems by breaking them down into sub-problems and organizing them in a hierarchical structure. This has the advantage of focusing on a single or multiple sub-problems. This is accomplished by adjusting the weights within the network in order to generate a certain level of interest [95]. A weighting-based attention mechanism that analyses and filters input sequences is also used in hierarchical LSTM networks for long-term dependency prediction [97].

Convolutional LSTM reduces and filters input data collected over a longer period of time using convolution operations applied in LSTM networks or the LSTM cell architecture directly. Furthermore, due to their distinct characteristics, convolutional LSTM networks are useful for modelling many quantities such as spatially and temporally distributed relationships. However, many quantities can be expected collectively in terms of reduced feature representation. Decoding or decoherence layers are required to predict different output quantities not as features but based on their parent units [95].

The LSTM autoencoder solves the problem of predicting high-dimensional parameters by shrinking and expanding the network [98]. The autoencoder architecture is separately trained with the aim of accurate reconstruction of the input data as reported in [99]. Only the encoder is used during testing and commissioning to extract the low-dimensional properties that are transmitted to the LSTM. The LSTM was extended to multimodal prediction using this strategy. To compress the input data and cell states, the encoder and decoder are directly integrated into the LSTM cell architecture. This combined reduction improves the flow of information in the cell and results in an improved cell state update mechanism for both short-term and long-term dependency [95].

Grid Long Short-Term Memory is a network of LSTM cells organized into a multidimensional grid that can be applied to sequences, vectors, or higher dimensional data like images [100]. Grid LSTM has connections to eg the spatial or temporal dimensions of input sequences. Thus, connections of different dimensions within cells extend the normal flow of information. As a result, Grid LSTM is appropriate for the parallel prediction of several output quantities that may be independent, linear, or non-linear. The network's dimensions and structure are influenced by the nature of the input data and the goal of the prediction [101].

A novel method for the collaborative prediction of numerous quantities is the cross-modal and associative LSTM. It uses a number of standard LSTMs to separately model different quantities. To calculate the dependencies of the quantities, these LSTM streams communicate with one another via recursive connections. The chosen layers' outputs are added as new inputs to the layers before and after them in other streams. Consequently, a multimodal forecast can be made. The benefit of this approach is that the correlation vectors that are produced have the same dimensions as the input vectors. As a result, neither the parameter space nor the computation time increase [102].

Recurrent Convolution Neural Network

CNN is a key method for handling various computers vision challenges. In recent years, a new generation of CNNs has been developed, the Recurrent Convolution Neural Network (RCNN), which is inspired by large-scale recurrent connections in the visual systems of animals. The Recurrent Convolutional Layer (RCL) is the main feature of RCNN, which integrates repetitive connections among neurons in the normal convolutional layer. With the increase in the number of repetitive computations, the Receptive Domains (RFs) of neurons in the RCL expand infinitely, which is contrary to biological facts [103].

The RCNN prototype was proposed by Ming Liang & Xiaolin Hu [104, 105], the structure is illustrated in Figure 16, in which both forward and redundant connections have local connectivity and weights shared between distinct sites. This design is quite similar to the Recurrent Multi-Layer Perceptron (RMLP) concept which is often used for dynamic control [106, 107] (Fig. 17, middle). Similar to the distinction between MLP and CNN, the primary distinction is that in RMLP, common local connections are used in place of full connections. For this reason, the proposed model is known as RCNN [108].

The main unit of RCNN is the RCL. RCLs develop through discrete time steps. RCNN offers three basic advantages. First, it allows each unit to accommodate background information in an arbitrarily wide area in the current layer. Second, recursive connections improve the depth of the network while keeping the number of mutable parameters constant through weight sharing. This is consistent with the trend of modern CNN architecture to grow deeper with a relatively limited number of parameters. The third aspect of RCNN is the time exposed in RCNN which is a CNN with many paths between the input layer and the output layer, which makes learning simple. On one hand, having longer paths makes it possible for the model to learn very complex features. On the other hand, having shorter paths may improve the inverse gradient during training [103].

The primary goals of this work are to present a comprehensive overview of the key machine learning as well as deep learning techniques employed in healthcare prediction, as well as to identify the obstacles that machine learning and deep learning face in healthcare prediction.

The rest of this paper is structured as follows:

- Section 2 presents a survey methodology
- Section 3 gives a literature survey of the machine learning and deep learning techniques used in healthcare prediction.
- Section 4 summarizes the advantage and limitations of the techniques discussed in section 3.
- Finally, Section 6 outlines the conclusions.

2. Survey Methodology

The studies discussed in this paper have been presented and published in high-quality journals and international conferences published by IEEE, Springer, and Elsevier. Machine learning, deep learning, healthcare, surgery, cardiology, radiology, hepatology, and nephrology are some of the terms used to search for these studies. The studies chosen for this survey are concerned with the use of machine learning as well as deep learning algorithms in healthcare prediction. For this survey, empirical and review articles on the topics were considered.

2.1 Survey Structure

This section discusses existing research efforts that healthcare prediction using various techniques in ML and DL. This survey gives a detailed discussion about the methods and algorithms which are used for predictions, performance metrics, and tools of their model.

2.1.1 ML-based Healthcare Prediction

In [109], the authors utilized a framework to create and assess ML classification models such as Logistic Regression, KNN, SVM, and RF for the prediction of diabetes patients. ML method was implemented on the Pima Indian Diabetes Database (PIDD) which has 768 rows and 9 columns. The forecast accuracy delivers 83 percent accuracy. Results of the implementation approach indicate how the Logistic Regression outperformed other algorithms of ML. The results indicated that only a structured dataset was selected but unstructured data are not considered, also model should be implemented in other healthcare domains like heart disease, and COVID-19, finally other factors should be considered for diabetes prediction, like family history of diabetes, smoking habits, and physical inactivity.

In [110], The authors developed a diagnosis system focusing on 4 prediction algorithm models (RF, SVM, NB, DT) to predict diabetes using two various databases (Frankfurt Hospital in Germany and PIDD provided by the UCI ML repository). the SVM algorithm performed with an accuracy of 83.1 percent. There

are some aspects of this study that need to be improved, such as using a DL approach to predict diabetes may lead to achieving better results, furthermore, the model should be tested in other healthcare domains such as heart disease and COVID-19 prediction.

In [111], the authors proposed three ML methods (Logistic Regression - DT - Boosted RF) to assess the COVID-19 OpenData Resources from Mexico and Brazil. To predict rescue and death, the proposed model incorporates just the COVID-19 patient's geographical, social, and economic conditions, as well as clinical risk factors, medical reports, and demographic data. On the dataset utilized, the model for Mexico has a 93 percent accuracy, and an F1 score is 0.79. On the other hand, on the used dataset, the Brazil model has a 69 percent accuracy and an F1 score is 0.75. The three ML algorithms have been examined and the acquired results showed that Logistic Regression is the best way of processing data. The authors should be concerned about the usage of authentication and privacy management of the created data.

In [112], The authors introduced a new model for predicting type 2 diabetes utilizing a network approach as well as ML techniques (Logistic Regression, SVM, NB, KNN, Decision Tree, RF, XGBoost, and ANN). To predict the risk of type 2 diabetes, the healthcare data of 1,028 type 2 diabetes patients and 1,028 non-type 2 diabetes patients were extracted from de-identified data. The experimental findings reveal the models' effectiveness with an Area Under Curve (AUC) varied from 0.79 to 0.91. The RF model achieved higher accuracy than others. This study relies only on the dataset providing hospital admission and discharges summaries from one insurance company. External hospital visits and information from other insurance companies are missing for people with many insurance providers.

In [113], The author proposed a healthcare management system that patients could use to schedule appointments with doctors and verify their prescriptions. It gives support for ML to detect ailments and determine medicines. ML models including DT, RF, logistic regression, and NB classifiers are applied to the datasets of diabetes, heart disease, chronic kidney disease, and liver. The results showed that among all the other models, logistic regression had the highest accuracy of 98.5 percent in the heart dataset. while the least accuracy is of the DT classifier which came out to be 92 percent. In the liver dataset the logistic regression with maximum accuracy of 75.17 percent among all others. In the chronic renal disease dataset, the logistic regression, RF, and Gaussian NB, all performed well with an accuracy of 1. In the diabetes dataset random forest with maximum accuracy of 83.67 percent. The authors should include a hospital directory as then various hospitals and clinics can be accessed through a single portal. Additionally, image datasets should be included to allow image processing of reports and the deployment of DL to detect diseases.

In [114], the authors developed an ML model to predict the occurrence of Type 2 Diabetes in the following year ($Y + 1$) using factors in the present year (Y). Between 2013 and 2018, the dataset was obtained as an electronic health record from a private medical institute. authors applied logistic regression, RF, SVM, XGBoost, and ensemble ML algorithms to predict the outcome of non-diabetic, prediabetes, and diabetes. Feature selection was applied to choose the three classes efficiently. FPG, HbA1c, triglycerides, BMI, gamma-GTP, gender, age, uric acid, smoking, drinking, physical activity, and family history were among

the features selected. According to the experimental results, the maximum accuracy was 73 percent from RF, while the lowest was 71 percent from the logistic regression model. The authors presented a model that used only one dataset. As a result, additional data sources should be applied to verify the models developed in this study.

In [115], the authors categorized the diabetes dataset using SVM and NB algorithms coupled with feature selection for enhancing the accuracies of the model. PIDD is taken from the UCI Repository for analysis. For training and testing purposes the authors employed the K-fold cross-validation model, the SVM classifier was performing better than the NB method it offers around 91 percent correct predictions, however, the authors acknowledge that they need to extend to the latest dataset that will contain additional attributes and rows.

In [116], the authors introduced an unsupervised ML algorithm K-means clustering for the UCI heart disease dataset to detect heart disease in the early stage. PCA is used for dimensionality reduction. The outcome of the method demonstrates early cardiac disease prediction with 94.06 percent accuracy. The authors should apply the proposed technique using more than one algorithm and use more than one dataset.

In [117], the authors constructed a predictive model for the classification of diabetes data using the logistic regression classification technique. the dataset includes 459 patients for training data and 128 cases for testing data. The prediction accuracy using logistic regression was obtained at 92 percent. The main limitation of this research is that the authors have not compared the model with other diabetes prediction algorithms and so it cannot be confirmed.

In [118], the authors developed a prediction model that analyses the user's symptoms and predicts the disease using ML algorithms (DT classifier, RF classifier, and NB classifier) to solve health-related problems by allowing professionals to predict diseases at an early stage. A dataset is a sample of 4920 patient records with 41 illnesses diagnosed. A total of 41 disorders were included as a dependent variable. All of the algorithms achieved the same accuracy score of 95.12%. The authors noticed that overfitting occurred when all 132 symptoms from the original dataset were assessed instead of 95 symptoms. i.e., the tree appears to remember the dataset provided and thus fails to classify new data. As a result, just 95 symptoms were assessed during the data-cleansing process, with the best ones being chosen.

In [119], the authors built a decision-making system that assists practitioners to anticipate cardiac problems in exact classification through a simpler method and will deliver automated predictions about the condition of the patient's heart. implemented 4 algorithms (KNN, RF, DT, and NB), all these algorithms were used in the Cleveland Heart Disease dataset. The accuracy varies for different classification methods. The maximum accuracy is given when they utilized the KNN algorithm with the Correlation factor which is almost 94 percent. The authors should extend the presented technique to leverage more than one dataset and forecast different diseases.

In [120], the authors applied three classification methods (NB, SVM, DT, and KNN) to the Cleveland dataset consisting of 303 cases and 76 attributes. Of these 76 traits, only 14 attributes are chosen for testing. authors performed data preprocessing to remove noisy data. The KNN obtained the greatest accuracy with 90.79 percent. To improve the accuracy of early heart disease prediction, the authors need to use more sophisticated models.

In [121], the authors proposed a model to predict heart disease utilizing a cardiovascular dataset used in this model and classified by using supervised ML algorithms (DT, NB, Logistic Regression, RF, SVM, and KNN). The results reveal that the DT classification model predicted cardiovascular disorders better than other algorithms with an accuracy of 73 percent. the authors highlighted that the ensemble ML techniques employing the CVD dataset can generate a better illness prediction model.

In [122], the authors attempted to increase the accuracy of heart disease prediction by applying a Logistic Regression using a healthcare dataset to determine whether patients have heart illness problems or not. The dataset was acquired from an ongoing cardiovascular study on people of the town of Framingham, Massachusetts. The model reached an accuracy prediction of 87 percent. the authors acknowledge the model could be improved with more data and the use of more ML models.

In [123], the author introduced an accurate classification to examine the breast cancer data with a total of 569 rows and 32 columns, because breast cancer affects one in every 28 women in India. Similarly employing a heart disease dataset and Lung cancer dataset, this research offered A novel way to function selection. This method of selection is based on genetic algorithms mixed with the SVM classification. The classifier results are Lung cancer 81.8182, Diabetes 78.9272. noticed that size, kind, and source of data used are not indicated.

In [124], the authors, predicted the risk factors that cause heart disease using the K-means clustering algorithm and analyzed with a visualization tool using a Cleveland heart disease dataset with 76 features of 303 patients, holds 209 records with 8 attributes such as age, chest pain type, blood pressure, blood glucose level, ECG in rest, heart rate as well as four types of chest pain. The authors forecast cardiac diseases by taking into consideration the primary characteristics of four types of chest discomfort solely and K-means clustering is a common unsupervised ML technique.

In [125], the authors aimed to report on the benefits of various DM methods and proven heart disease survival prediction models. From the observations, the authors proposed that Logistic Regression and NB achieved the highest accuracy when performed on a high dimensional dataset on the Cleveland hospital dataset and DT and RF produce better results on small dimensional datasets. RF delivers more accuracy than the DT classifier as the algorithm is an optimized learning algorithm. The author mentioned that this work can be extended to other ML algorithms, the model could be developed in a distributed environment such as Map-Reduce, Apache Mahout, HBase, etc.

In [126] the authors proposed a single algorithm named hybridization, that combines used techniques into one single algorithm, The presented Method has three phases, preprocessing phase, classification

phase, and diagnosis phase. They employed the Cleveland database and algorithms NB, SVM, KNN, NN, J4.8, RF, and GA. NB and SVM always perform better than others, whereas others depend on the specified features. results attained an accuracy of 89.2 percent. Authors need to enhance accuracy, better accuracy is the key goal. Notice that the dataset is little, hence the system was not able to train adequately, so the accuracy of the method was bad.

In [127], the authors presented a study concentrated on the utilization of clinical data for liver disease prediction and investigate several ways of representing such data through this analysis by utilizing six algorithms Logistics Regression, KNN, DT, SVM, NB, and RF. The original dataset was taken from the northeast of Andhra Pradesh, India. includes 583 liver patient's data whereas 75.64 percent are male and 24.36 percent are female. The analysis result indicated that the Logistics Regression classifier delivers the most increased order exactness of 75 percent depending on the f1 measure to forecast the liver illness and NB gives the least precision of 53 percent. Authors merely studied a few prominent supervised ML algorithms; more algorithms can be picked to create an increasingly exact model of liver disease prediction and performance can be steadily improved.

In [128], the authors aimed to predict coronary heart disease (CHD) based on historical medical data using ML technology. The goal of this study is to use three supervised learning approaches, NB, SVM, and DT, to find correlations in CHD data that could aid improve prediction rates. The dataset contains a retrospective sample of males from KEEL, a high-risk heart disease location in the Western Cape of South Africa. the model utilized NB, SVM, and DT. NB achieved the most accuracy among the three models. SVM and DT J48 outperformed NB with a specificity rate of 82 percent but showed to have an inadequate sensitivity rate of less than 50 percent.

In [129], the authors applied data mining and network analysis techniques in hospital admission and discharge data to analyze the disease or comorbidity footprints of chronic patients. A chronic disease risk prediction framework was created and evaluated in the Australian healthcare system to predict type 2 diabetes risk. Using a private healthcare funds dataset from Australia that spans six years and three different predictive algorithms (regression, parameter optimization, and DT). The accuracy of the prediction ranges from 82 to 87 percent. The hospital admission and discharge summary is the dataset's source. As a result, it does not provide information about general physician visits or future diagnoses.

2.1.2 DL-based Healthcare Prediction

In [130], the authors proposed a system for predicting the patients with the more common inveterate diseases with the help of the DL algorithms such as CNN for auto feature extraction and illness prediction so, they used KNN for distance calculation to locate the exact matching in the dataset and the outcome of the final prediction of the sickness. A combination of disease symptoms was made for the structure of the dataset, the living habits of a person, and also the specifies attaches to doctor consultations which are acceptable in this general disease prediction. In this study, the Indian chronic kidney disease dataset was utilized that comprises 400 occurrences, 24 characteristics, and 2 classes were restored from the UCI ML store. At last, a comparative study of the proposed system with other algorithms such as NB, DT, and

logistic regression has been demonstrated in this study. The findings showed that the proposed system gives an accuracy of 95 percent which is higher than the other two methods. So, the proposed technique should be applied using more than one dataset.

In [131], the authors developed a DL approach that uses chest radiography images to differentiate between patients with mild, pneumonia, and COVID-19 infections, providing a valid mechanism for COVID-19 diagnosis. To increase the intensity of the chest X-ray image and eliminate noise, image-enhancing techniques were used in the proposed system. Two distinct DL approaches based on a pertained neural network model (ResNet-50) for COVID-19 identification utilizing Chest X-ray (CXR) pictures are proposed in this work to minimize overfitting and increase the overall capabilities of the suggested DL systems. The authors emphasized that tests using a vast and hard dataset encompassing several COVID-19 cases are necessary to establish the efficacy of the suggested system.

In [132], the authors presented a Cuckoo search-based deep LSTM classifier for disease prediction. The deep convLSTM classifier is used in the cuckoo search optimization, which is a nature-inspired method for accurately predicting disease by transferring information and therefore reducing time consumption. The PIMA dataset is used to predict the onset of diabetes. The National Institute of Diabetes and Digestive and Kidney Diseases provided the data. The dataset is made up of independent variables including insulin level, age, and BMI index, as well as one dependent variable. The new technique was compared to traditional methods, and the results showed that the proposed method achieved 97.591 percent accuracy, 95.874 percent sensitivity, and 97.094 percent specificity, respectively. authors noticed more datasets are needed, as well as new approaches to improve the classifier's effectiveness.

In [133], the authors presented a wavelet-based convolutional neural network to handle data limitations in this time of COVID-19 fast emergence. By investigating the influence of discrete wavelet transform decomposition up to 4-levels, the model demonstrated the capability of multi-resolution analysis for detecting COVID-19 Chest X-rays. The wavelet sub-bands are the CNN's inputs at each decomposition level. COVID-Chest X-ray-12 is a collection of 1,944 chest X-ray pictures divided into 12 groups that were compiled from two open-source datasets (National Institute Health containing several X-rays of pneumonia-related diseases where the COVID-19 dataset is collected from Radiology Society North America). COVID-Neuro wavelet, a suggested model, was trained alongside other well-known ImageNet pre-trained models on COVID-CXR-12. the authors acknowledge they hope to investigate the effects of other wavelet functions besides the Haar wavelet.

In [134], the authors developed a CNN framework for COVID-19 identification utilizing computed tomography images is suggested. The proposed framework employer a public CT dataset of 2482 CT images from patients of both classifications. the system attained an accuracy of 96.16 percent and recall of 95.41 percent after training using only 20 percent of the dataset. The authors stated that the use of the framework should be extended to multimodal medical pictures in the future.

In [135], the authors performed multi-disease prediction for intelligent clinical decision support by deploying a long short-term memory network and enhancing it with two processes to conduct multi-label

classification based on patients' clinical visit records. a massive data set of electronic health records collected from a prominent hospital in southeast china. The suggested LSTM approach outperforms several standard and DL models in predicting future disease diagnoses, according to model evaluation results. The F1 score rises from 78.9% and 86.4 percent, respectively, with the state-of-the-art conventional and DL models, to 88.0 percent with the suggested technique. The authors stated that the model prediction performance may be enhanced further by including new input variables and that to reduce computational complexity, the method only uses one data source.

In [136], the authors introduced an approach to creating a supervised ANN structure based on the subnets (the group of neurons) instead of layers, in the cases of low datasets, this effectively predicted the disease. The model was evaluated using textual data and compared to Multilayer Perceptron's (MLPs) as well as LSTM recurrent neural network models using three small-scale publicly accessible benchmark datasets. On the Iris dataset, the experimental findings for classification reached 97 percent accuracy, compared to 92 percent for RNN (LSTM) with three layers, and the model had a lower error rate, 81, than RNN (LSTM) and MLP on the diabetic dataset, while RNN (LSTM) has a high error rate of 84. For larger datasets, however, this method is useless. This model is useless because not implement our model on large textual and image datasets.

In [137], the authors presented a novel AI and Internet of Things (IoT) convergence-based disease detection model for a smart healthcare system. Data collection, reprocessing, categorization, and parameter optimization are all stages of the proposed model. IoT devices, such as wearables and sensors, collect data, which AI algorithms then use to diagnose diseases. The forest technique is then used to remove any outliers found in the patient data. Healthcare data was used to assess the performance of the CSO-LSTM model. During the study, the CSO-LSTM model had a maximum accuracy of 96.16 percent on heart disease diagnoses and 97.26 percent on diabetes diagnoses. This method offered a greater prediction accuracy for heart disease and diabetes diagnosis, but there was no feature selection mechanism, hence it requires extensive computations.

In [138], the authors focused on the coronavirus epidemic, which constitutes a daily threat to global health. The majority of their research was aimed at detecting disease in people whose Xrays had been selected as potential COVID-19 candidates. Chest x-rays of people with COVID-19, viral pneumonia, and healthy people are included in the dataset. The study compared the performance of two DL algorithms, namely CNN and RNN. DL techniques were used to evaluate a total of 657 chest X-ray images for the diagnosis of COVID-19. VGG19 is the most successful model, with a 95% accuracy rate. The VGG19 model successfully categorizes COVID-19 patients, healthy individuals, and viral pneumonia cases. The dataset's most failing approach is InceptionV3. The success percentage can be improved, according to the authors, by improving data collection. In addition to chest radiography, lung tomography can be used. The success ratio and performance can be enhanced by creating numerous DL models.

In [139], the authors developed a method based on the RNN algorithm for predicting blood glucose levels for diabetics a maximum of one hour in the future, which required the patient's glucose level history. The

Ohio T1DM dataset for blood glucose level prediction, which included blood glucose level values for six people with type 1 diabetes, was used to train and assess the approach. The distribution features were further honed with the use of studies that revealed the procedure's certainty estimate nature. The authors point out that they can only evaluate prediction goals with enough glucose level history, thus they can't anticipate the beginning levels after a gap, which doesn't improve the prediction's quality.

In [140], the authors used an 18-layer residual CNN pre-trained on ImageNet with a different anomaly detection mechanism for the classification of COVID-19 to construct a new deep anomaly detection model for speedy, reliable screening. On the X-ray dataset, which contains 100 images from 70 COVID-19 persons and 1431 images from 1008 non-COVID-19 pneumonia subjects, the model obtains a sensitivity of 90.00 percent specificity of 87.84 percent or sensitivity of 96.00 percent specificity of 70.65 percent. The authors noted that the model still has certain flaws, such as missing 4% of COVID-19 cases and having a 30% false-positive rate. In addition, more clinical data is required to confirm and improve the model's usefulness.

In [141], the authors developed COVIDX-Net, a novel DL framework that allows radiologists to diagnose COVID-19 in X-ray images automatically. Seven algorithms (MobileNetV2, ResNetV2, VGG19, DenseNet201, InceptionV3, Inception, and Xception) were evaluated using a small dataset of 50 photos (MobileNetV2, ResNetV2, VGG19, DenseNet201, InceptionV3, Inception, and Xception). Each deep neural network model can classify the patient's status as a negative or positive COVID-19 case based on the normalized intensities of the X-ray image. The f1-scores for the VGG19 and Dense Convolutional Network (DenseNet) models were 0.89 and 0.91, respectively. With f1-scores of 0.67, the InceptionV3 model has the weakest classification performance.

In [142], The authors created a DL approach for delivering 30-minute predictions about future glucose levels based on a Dilated RNN (DRNN). The performance of the DRNN models was evaluated using data from two electronic health records datasets: OhioT1DM from clinical trials and the in-silicon dataset from the UVA-Padova simulator. It outperformed established glucose prediction approaches such as Neural Networks (NNs), Support Vector Regression (SVR), and autoregressive models (ARX) (ARX). The results demonstrated that it significantly improved glucose prediction performance, although there are still some limits, such as the authors' creation of a data-driven model that heavily relies on past EHR. The quality of the data has a significant impact on the accuracy of the prediction. The number of clinical datasets is limited, however, often restricted. Because certain data fields are manually entered, they are occasionally incorrect.

In [143], the authors utilized a deep neural network to discover 15,099 stroke patients, researchers were able to predict stroke death based on medical history and human behaviors utilizing large-scale electronic health information. The Korea Centers for Disease Control and Prevention collected data from 2013 to 2016 and found that there are around 150 hospitals in the country, all having more than 100 beds. Gender, age, type of insurance, mode of admission, necessary brain surgery, area, length of hospital stay, hospital location, number of hospital beds, stroke kind, and CCI were among the 11 variables in the DL

model. To automatically create features from the data and identify risk factors for stroke, researchers used a DNN/scaled Principal Component Analysis (PCA). 15,099 people with a history of stroke were enrolled in the study. The data were divided into a training set (66%) and a testing set (34%), with 30 percent of the samples used for validation in the training set. DNN is used to examine the variables of interest, while scaled PCA is utilized to improve the DNN's continuous inputs. This study sensitivity, specificity, and AUC values were respectively 64.32 percent, 85.56 percent, and 83.48 percent.

In [144] the authors proposed a glucose forecasting approach called (GluNet) that used a personalized DNN for forecasting the probabilistic distribution of short-term measurements having Type 1 diabetes based on their historical data that involved insulin doses, meal information, glucose measurements, and various factors. It utilized the newest DL techniques consisting of four components: post-processing, dilated Convolution Neural Network (CNN), label recovery/ transform, and data pre-processing. authors run the models on the subjects from the OhioT1DM datasets. The outcomes revealed significant enhancements over the previous procedures via a comprehensive comparison concerning the and Root Mean Square Error (RMSE) having a time lag of 60 mins Prediction Horizons (PH) and RMSE having a small time lag for the case of prediction horizons in the virtual adult participants. If the PH is properly matched to the lag between input and output, the user may learn the control of the system more frequently and it achieves good performance. Additionally, GluNet was validated on two clinical data sets. It attained an RMSE with a time lag of 60 mins PH and RMSE with a time lag of 30 mins PH. The authors point out that the model does not consider physiological knowledge, and that they need to test GluNet with larger prediction horizons and use it to predict overnight hypoglycemia.

In [145], the authors proposed the Short-Term Blood Glucose Prediction Model (VMD-IPSO-LSTM), which is a short-term strategy for predicting blood glucose (VMD-IPSO-LSTM). Initially, the Intrinsic Modal Functions (IMF) in various frequency bands were obtained using the Variational Modal Decomposition (VMD) technique, which deconstructed the blood glucose content. The short and long-term memory networks then constructed a prediction mechanism for each blood glucose component Intrinsic Modal Functions (IMF). Because the time window length, learning rate, and neuron count are difficult to set, the upgraded PSO approach optimized these parameters. The improved LSTM network anticipated each IMF, and the projected subsequence was superimposed in the final step to arrive at the ultimate prediction result. The data of 56 participants were chosen as experimental data among 451 diabetic Mellitus patients. The experiments revealed that it improved prediction accuracy at "30 minutes, 45 minutes, and 60 minutes." The RMSE and MAPE were lower than the "VMD-PSO-LSTM, VMD-LSTM, and LSTM," indicating that the suggested model is effective. The longer time it took to anticipate blood glucose levels and the higher accuracy of the predictions gave patients and doctors more time to improve the effectiveness of diabetes therapy and manage blood glucose levels. The authors noted that they still faced challenges, such as an increase in calculation volume and operation time. The time it takes to estimate glucose levels in the short term will be reduced.

In [146], The authors presented a paradigm for primary COVID-19 detection using a radiology review of chest radiography or chest X-ray, to reduce diagnosis time and human error. The researchers used a

dataset of chest X-rays from verified COVID-19 patients (408 photographs), confirmed pneumonia patients (4273 images), and healthy people (1590 images) to perform a three-class image classification (1590 images). There are 6271 people in total in the dataset. To fulfill this image categorization problem, the authors plan to use CNN and transfer learning. For all of the folds of data, the model's accuracy ranged from 93.90 percent to 98.37 percent. Even the lowest level of accuracy, 93.90 percent, is still quite good. The authors will face a restriction, particularly when it comes to adopting such a model on a large scale for practical usage.

In [147], the authors proposed DL models for predicting the number of COVID-19 positive cases in Indian states. The Ministry of Health and Family Welfare dataset contains time-series data for 32 individual confirmed COVID-19 cases in each of the states (28) and union territories (4) since March 14, 2020. This dataset was used to conduct an exploratory analysis of the increase in the number of positive cases in India. As prediction models, RNN-based LSTMs are used. Deep LSTM, convolutional LSTM, and bi-directional LSTM models were tested on 32 states/union territories, and the model with the best accuracy was chosen based on absolute error. Bi-directional LSTM produced the best performance in terms of prediction errors, while convolutional LSTM produced the worst performance. For all states, daily and weekly forecasts were calculated, and bi-LSTM produced accurate results (error less than 3%) for short-term prediction (1–3 days).

In [148], the authors suggested a new type 1 diabetes prediction technique based on CNNs and DL to improve the robustness and accuracy of type 1 diabetes prediction. It was all about figuring out how to extract the behavioral pattern. Numerous observations of identical behaviors were used to fill in the gaps in the data. The suggested model was trained and verified using data from 759 people with type 1 diabetes who visited Sheffield Teaching Hospitals between 2013 and 2015. A subject's type 1 diabetes test, demographic data (age, gender, years with diabetes), and the final 84 days (12 weeks) of Self-Monitored Blood Glucose (SMBG) measurements preceding the test formed each item in the training set. In the presence of insufficient data and certain physiological specificities, prediction accuracy deteriorates, according to the authors.

In [149], the authors constructed a machine learning technique using the PIDD by NIDDK. PID's participants are all female and at least 21 years old. PID comprises 768 incidences, with 268 samples diagnosed as diabetic and 500 samples not diagnosed as diabetic. The eight most important characteristics that led to diabetes prediction. The accuracy of functional classifiers such as ANN, NB, DT, and DL is between 90 and 98 percent. On the PIMA dataset, DL had the best results for diabetes onset among the four, with an accuracy rate of 98.07 percent. The technique uses a variety of classifiers to accurately predict the disease, but it failed to diagnose it at an early stage.

3. Future Directions

The adoption of ML, as well as DL models, produced a massive impact in all areas and mainly in the healthcare domain. However, with the availability of numerous reports and scans, human decision-

making remains the only option for diagnosis. This could lead to inaccurate diagnoses due to preference in human decisions, as a result, many human lives are saved. Therefore, researchers identified numerous methods for automating diagnosis in healthcare in different specialties. Research has shown that healthcare professionals are gradually facing the technology for a variety of purposes, such as collecting patient records or diagnostics. Looking at AI in terms of medical decision-making and data management. The presented study has drawbacks that can be solved by using more investigations in the future, for example, there is a possibility that other approaches are present which are not included in this survey. Next, the use of search keywords like "AI", "ML", "DL", and "Healthcare" could be common and eliminate the interesting researches. Furthermore, this survey investigated 40 scientific papers, since the topic of research is new, the evaluation of more research papers may yield more interesting results.

4. Conclusion

The use of machine learning, as well as deep learning algorithms for healthcare prediction, has the potential ability to change the way traditional healthcare services are delivered. In the case of machine learning and deep learning applications, healthcare data is deemed the most significant component that contributes to medical-care systems. This paper aims to convey a rich discussion with medical care staff about how AI can be helpful to them to increase the quality of work. A total of 40 working papers covering the period from 2019 to 2022 were selected and the methodology for each paper was clarified. studies have shown that artificial intelligence plays a significant role in diagnosing diseases accurately and helps to anticipate healthcare and analyze health data by linking hundreds of clinical records and rebuilding a patient's history using this data. Therefore, there is a need for more studies to improve the links of AI with data quality management considerations in healthcare.

Abbreviations

AI: Artificial Intelligence; ML: Machine learning; DT: Decision Tree; EHR: Electronic Health Records; RF: Random Forest; SVM: Support Vector Machine; KNN: K - Nearest Neighbor; NB: Naive Bayes; RL: Reinforcement learning; NLP: Natural Language Processing; MCTS: Monte Carlo Tree Search; POMDP: Partially Observable Markov Decision Processes; DL: Deep Learning; DBN: Deep Belief Network; ANNs: Artificial Neural Networks; CNNs: Convolutional Neural Networks; LSTM: Long Short-Term Memory; RCNNs: Recurrent Convolution Neural networks; RNNs: Recurrent Neural Networks; RCL: Recurrent Convolutional Layer; RFs: Receptive Domains; RMLP: Recurrent Multi-Layer Perceptron; PIDD: Pima Indian Diabetes Database; CHD: Coronary Heart Disease; CXR: chest X-ray; MLPs: Multilayer Perceptron's; LSTM: Long Short-Term Memory; IOT: Internet of Things; DRNN: Dilated RNN; NNs: Neural Networks; SVR: Support Vector Regression; PCA: Principal Component Analysis; PH: Prediction Horizons; RMSE: Root Mean Square Error; IMF: Intrinsic Modal Functions; VMD: Variational Modal Decomposition; IMF: Intrinsic Modal Functions; SMBG: Self-Monitored Blood Glucose.

Declarations

Supplementary Information

Not applicable.

Acknowledgments

Not applicable.

Authors' contributions

All authors have participated equally in this work.

Funding

Not applicable.

Availability of data and materials

The corresponding author can provide the material used and data analyzed on request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests. All authors approved the final manuscript.

References

1. Hema Latha, M., Ramakrishna, A., Sudarsha Chakravarthi Reddy, B., Venkateswarlu, C., & Yamini Saraswathi, S. (2022). Disease Prediction by Stacking Algorithms Over Big Data from Healthcare Communities. In *Intelligent Manufacturing and Energy Sustainability* (pp. 355-363). Springer, Singapore.
2. Elmahdy, H. N. (2014). Medical Diagnosis Enhancements through Artificial Intelligence.
3. Van Calster, B., Wynants, L., Timmerman, D., Steyerberg, E. W., & Collins, G. S. (2019). Predictive analytics in health care: how can we know it works?. *Journal of the American Medical Informatics Association*, 26(12), 1651-1654.
4. Prasan Kumar Sahoo, Suvendu Kumar Mohapatra, and Shih-Lin Wu. "SLA based healthcare big data analysis and computing in cloud network". In: *Journal of Parallel and Distributed Computing* 119 (2018), pp. 121–135.
5. Vivekanandan Thanigaivasan, Swathi J Narayanan, and N Ch Sriman Narayana Iyengar. "Analysis of Parallel SVM Based Classification Technique on Healthcare using Big Data Management in Cloud Storage". In: *Recent Patents on Computer Science* 11.3 (2018), pp. 169–178.
6. Yichuan Wang et al. "An integrated big data analytics-enabled transformation model: Application to health care". In: *Information & Management* 55.1 (2018), pp. 64–79.
7. Omran S. Habib. "Primary health care and health care administration". In: *Recent Patents on University of Basrah* 11.3 (2016), pp. 1–57.
8. Xiong, X., Cao, X., & Luo, L. (2021). The ecology of medical care in Shanghai. *BMC Health Services Research*, 21(1), 1-9.
9. Burazeri G, Zaletel-Kragelj L In book: *Health: Systems – Lifestyles – Policies (Volume I)* Edition: 2nd Chapter: The role and organization of health care systems
10. Marzorati, C., & Pravettoni, G. (2017). Value as the key concept in the health care system: how it has influenced medical practice and clinical decision-making processes. *Journal of Multidisciplinary Healthcare*, 10, 101.
11. Qayyum, A., Qadir, J., Bilal, M., & Al-Fuqaha, A. (2020). Secure and robust machine learning for healthcare: A survey. *IEEE Reviews in Biomedical Engineering*, 14, 156-180.
12. El Seddawy, A. B., Moawad, R., & Hana, M. A. (2018). Applying Data Mining Techniques in CRM.
13. Malik, M., Khatana, R., & Kaushik, A. (2021, October). Machine Learning With Health Care: A perspective. In *Journal of Physics: Conference Series* (Vol. 2040, No. 1, p. 012022). IOP Publishing.
14. Mirbabaie, M., Stieglitz, S., & Frick, N. R. (2021). Artificial intelligence in disease diagnostics: a critical review and classification on the current state of research guiding future direction. *Health and Technology*, 11(4), 693-731.

15. Singh, G., Al'Aref, S. J., Van Assen, M., Kim, T. S., van Rosendael, A., Kolli, K. K., ... & Min, J. K. (2018). Machine learning in cardiac CT: basic concepts and contemporary data. *Journal of Cardiovascular Computed Tomography*, 12(3), 192-201.
16. Kim, K. J., & Tagkopoulos, I. (2019). Application of machine learning in rheumatic disease research. *The Korean journal of internal medicine*, 34(4), 708.
17. Liu, B. (2011). Supervised learning. In *Web data mining* (pp. 63-132). Springer, Berlin, Heidelberg.
18. Haykin, S., & Lippmann, R. (1994). Neural networks, a comprehensive foundation. *International journal of neural systems*, 5(4), 363-364.
19. Gupta, M. (2022). A Comparative Study on Supervised Machine Learning Algorithm. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, Volume 10.
20. Ray, S. (2019, February). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE.
21. Srivastava, A., Saini, S., & Gupta, D. (2019, June). Comparison of various machine learning techniques and its uses in different fields. In *2019 3rd International conference on electronics, communication and aerospace technology (ICECA)* (pp. 81-86). IEEE.
22. Park, H. A. (2013). An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing*, 43(2), 154-164.
23. Obulesu, O., Mahendra, M., & ThrilokReddy, M. (2018, July). Machine learning techniques and tools: A survey. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 605-611). IEEE.
24. Dhall, D., Kaur, R., & Juneja, M. (2020). Machine learning: a review of the algorithms and its applications. *Proceedings of ICRIC 2019*, 47-63.
25. F. Yang, "An Extended Idea about Decision Trees," in *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*, Dec. 2019, pp. 349–354, doi: 10.1109/CSCI49370.2019.00068.
26. A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2670–2679, Apr. 2015, doi: 10.1016/j.eswa.2014.11.009.
27. A. Shamim, H. Hussain, and Maqbool Uddin Shaikh, "A framework for generation of rules from decision tree and decision table," in *2010 International Conference on Information and Emerging Technologies*,
28. A. S. Eesa, A. M. Abdulazeez, and Z. Orman, "A DIDS Based on The Combination of Cuttlefish Algorithm and Decision Tree," *Science Journal of University of Zakho*, vol. 5, no. 4, pp. 313–318, 2017.
29. Bakyarani, S., Srimathi, H., & Bagavandas, M., (2019). "A Survey Of Machine Learning Algorithms In Health Care" *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 8, ISSUE 11*.

30. Resende, P. A. A., & Drummond, A. C. (2018). A survey of random forest based methods for intrusion detection systems. *ACM Computing Surveys (CSUR)*, 51(3), 1-36.
31. Breiman, L. Random forests. *Mach. Learn.* 2001, 45, 5–32.
32. Barandiaran, I. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 1998, 20, 832–844.
33. Hofmann, H.; Klinkenberg, R. *RapidMiner: Data Mining Use Cases and Business Analytics Applications*; CRC Press Taylor & Francis Group: Boca Raton, FL, USA, 2013.
34. Chow, C.K.; Liu, C. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* 1968, 14, 462–467.
35. C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, 1998.
36. J. Han and M. Kamber, *Data Mining Concepts and Techniques*, Elsevier, 2011.
37. Vapnik, Corinna Cortes and Vladimir, *Support Vector Network*, *Machine Learning*, vol. 20, pp. 273-297, 1995.
38. Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in using IoT with machine learning in health prediction system. *Forecasting*, 3(1), 181-206.
39. Sarker, I. H. (2021). *Machine learning: Algorithms, real-world applications and research directions*. *SN Computer Science*, 2(3), 1-21.
40. Ting, K. and Zheng, Z. "Improving the Performance of Boosting for Naïve Bayesian Classification. *Methodologies for Knowledge Discovery and Data Mining*", 1999, pp.296-298.
41. Kaur, R., & Juneja, M. (2016). A survey of different imaging modalities for renal cancer. *Indian J Sci Technol*, 9(44), 1-6.
42. Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018, March). Machine learning in healthcare: A review. In *2018 Second international conference on electronics, communication and aerospace technology (ICECA)* (pp. 910-914). IEEE.
43. Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR)*. [Internet], 9, 381-386.
44. Greene, D., Cunningham, P., & Mayer, R. (2008). Unsupervised learning and clustering. In *Machine learning techniques for multimedia* (pp. 51-90). Springer, Berlin, Heidelberg.
45. Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc..
46. Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, 1(6), 90-95.
47. Shlens, J. (2014). A tutorial on principal component analysis. arXiv preprint arXiv:1404.1100.
48. Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., ... & Laishram, M. (2017). Multivariate statistical data analysis-principal component analysis (PCA). *International Journal of Livestock Research*, 7(5), 60-78.

49. Kamani, M. M., Haddadpour, F., Forsati, R., & Mahdavi, M. (2022). Efficient fair principal component analysis. *Machine Learning*, 1-32.
50. Dey, A.: Machine learning algorithms: a review. *Int. J. Comput. Sci. Inf. Technol.* 7(3), 1174–1179 (2016).
51. Agrawal.R, Imielinski.T, Swami.A. Mining Association rules between Sets of Items in large Databases[CI In Proceedings of the ACM-SIGMOD Conference on Management of Data,1993:207-216.
52. Agrawal.R, SrikantR. Fast Algorithms for Mining Association Rules in Large Databases[Clin Proceedings of the 20th International Conference on Very Large Databases, I 994:487-499.
53. Singh, J., Ram, H., & Sodhi, D. J. (2013). Improving efficiency of apriori algorithm using transaction reduction. *International Journal of Scientific and Research Publications*, 3(1), 1-4.
54. Al-Maolegi, M., & Arkok, B. (2014). An improved Apriori algorithm for association rules. *arXiv preprint arXiv:1403.3948*.
55. Abaya, S. A. (2012). Association rule mining based on Apriori algorithm in minimizing candidate generation. *International Journal of Scientific & Engineering Research*, 3(7), 1-4.
56. Coronato, A., Naeem, M., De Pietro, G., & Paragliola, G. (2020). Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine*, 109, 101964.
57. Watlnds, C.J.C.H. (1989). Learning from delayed rewards. PhD Thesis, University of Cambridge, England.
58. Watkins, C.J.C.H. (1989). Learning from delayed rewards. PhD Thesis, University of Cambridge, England.
59. Watkin, C. J. C. H., & Dayan, P. (1992). Technical note Q-learning. *Machine Learning*, 8(3), 279-292.
60. Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. *IEEE Access*, 7, 133653-133667.
61. A. Achille and S. Soatto, "Information dropout: Learning optimal representations through noisy computation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
62. G. Williams, N. Wagener, B. Goldfain, P. Drews, J.M. Rehg, B. Boots, and E.A. Theodorou, "Information theoretic mpc for model-based reinforcement learning," In *Robotics and Automation (ICRA)*, IEEE International Conference on, pp. 1714–1721, 2017.
63. J.T. Wilkes and C.R. Gallistel, "Information theory, memory, prediction, and timing in associative learning," *Computational Models of Brain and Behavior*, pp. 481–492, 2017.
64. Jang, B., Kim, M., Harerimana, G., & Kim, J. W. (2019). Q-learning algorithms: A comprehensive classification and applications. *IEEE Access*, 7, 133653-133667.
65. Y. An, Y. Wang, and H. Meng, "Multi-task deep learning for user intention understanding in speech interaction systems," 2017.
66. X. Shi, Z. Gao, L. Lausen, H. Wang, D.Y. Yeung, W.K. Wong, and W.C. Woo, "Deep learning for precipitation nowcasting: A benchmark and a new model," *Advances in Neural Information*

- Processing Systems,” pp. 5622–5632, 2017.
67. Chia-Feng Juang and Chun-Ming Lu. 2009. Ant colony optimization incorporated with fuzzy Q-learning for reinforcement fuzzy control. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 39, 3 (2009), 597–608.
 68. Świechowski, M., Godlewski, K., Sawicki, B., & Mańdziuk, J. (2021). Monte Carlo tree search: A review of recent modifications and applications. *arXiv preprint arXiv:2103.04931*.
 69. Lizotte DJ, Laber EB (2016) Multi-Objective Markov Decision Processes for Data-Driven Decision Support. *Journal of Machine Learning Research* 17:211:1–211:28
 70. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van Den Driessche G, Schrittwieser J, Antonoglou I, Panneershelvam V, Lanctot M, et al. (2016) Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* 529(7587):484–489
 71. H. Baier and P. D. Drake, “The power of forgetting: Improving the lastgood- reply policy in Monte Carlo Go,” *IEEE Trans. Comput. Intell. AI Games*, vol. 2, no. 4, pp. 303–309, Dec. 2010.
 72. Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P, ... & Colton, S. (2012). A survey of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in games*, 4(1), 1-43.
 73. Z. H. Ling, S. Y. Kang, H. Zen, A. Senior, M. Schuster, X. J. Qian, H. M. Meng, and L. Deng, “Deep learning for acoustic modelling in parametric speech generation: A systematic review of existing techniques and future trends,” *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 35–52, 2015.
 74. J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
 75. D. Yu, and L. Deng, “Deep learning and its applications to signal and information processing [exploratory dsp],” *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 145–154, 2011.
 76. G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.
 77. Goyal, P., Pandey, S., Jain, K.: Introduction to natural language processing and deep learning. In: *Deep Learning for Natural Language Processing*, pp. 1–74. Springer, Berlin (2018).
 78. Mathew, A., Amudha, P., & Sivakumari, S. (2020, February). Deep learning techniques: an overview. In *International Conference on Advanced machine learning technologies and applications* (pp. 599–608). Springer, Singapore.
 79. Yoshua Bengio, Ian J. Goodfellow, and Aaron Courville. *Deep learning*. Book in preparation for MIT Press, 2015.
 80. Gomes, L. (2014). Machine-learning maestro michael jordan on the delusions of big data and other huge engineering efforts. *IEEE spectrum*, 20.
 81. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708.

82. Yap, M.H.; Pons, G.; Martí, J.; Ganau, S.; Sentís, M.; Zwiggelaar, R.; Davison, A.K.; Marti, R. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J. Biomed. Health Inform.* 2017, 22, 1218–1226.
83. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. 22. Hayashi, Y. The right direction needed to develop white-box deep learning in radiology, pathology, and ophthalmology: A short review. *Front. Robot. AI* 2019, 6, 24.
84. Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, 8(3), 292.
85. Apaydin, H., Feizi, H., Sattari, M. T., Colak, M. S., Shamsirband, S., & Chau, K. W. (2020). Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting. *Water*, 12(5), 1500.
86. Ganatra, N., & Patel, A. (2018). A comprehensive study of deep learning architectures, applications and tools. *International Journal of Computer Sciences and Engineering*, 6(12), 701-705.
87. Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). Ieee.
88. Y. S. P. & F. P. Bengio, " Learning long-term dependencies with gradient descent is difficult," *IEEE Trans*, vol. *Neural Networks* 5, p. 157–166 , 1994.
89. S. & L. B. & Y. S. Min, "Deep Learning in Bioinformatics. Briefings in Bioinformatics.," 18. 10.1093/bib/bbw068, 2016.
90. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*; 2010. p. 249–56.
91. Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8(1), 1-74.
92. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
93. Smagulova, K., & James, A. P. (2019). A survey on LSTM memristive neural network architectures and applications. *The European Physical Journal Special Topics*, 228(10), 2313-2324.
94. Setyanto, A., Laksito, A., Alarfaj, F., Alreshoodi, M., Oyong, I., Hayaty, M., ... & Kurniasari, L. (2022). Arabic Language Opinion Mining Based on Long Short-Term Memory (LSTM). *Applied Sciences*, 12(9), 4140.
95. Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. *Procedia CIRP*, 99, 650-655.
96. Cui, Z., Ke, R., Pu, Z., and Wang, Y., Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. *arXiv preprint arXiv:1801.02143*, 2018.

97. Villegas, R., Yang, J., Zou, Y., Sohn, S., Lin, X., and Lee, H., Learning to generate long-term future via hierarchical prediction. in Proceedings of the 34th International Conference on Machine Learning- Volume 70, 2017, pp.3560–3569.
98. Gensler, A., Henze, J., Sick, B., and Raabe, N., Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. in 2016 IEEE international conference on systems, man, and cybernetics (SMC), 2016, pp. 2858–2865.
99. Lindemann, B., Fesenmayr, F., Jazdi, N., and Weyrich, M., Anomaly detection in discrete manufacturing using self-learning approaches. *Procedia CIRP*, vol. 79, 2019, pp. 313–318.
100. Kalchbrenner, N., Danihelka, I., and Graves, A., Grid long short-term memory. arXiv preprint arXiv:1507.01526, 2015.
101. Cheng, B., Xu, X., Zeng, Y., Ren, J., and Jung, S., Pedestrian trajectory prediction via the Social-Grid LSTM model. *The Journal of Engineering*, vol. 2018, no. 16, 2018, pp. 1468–1474.
102. Veličković, P. et al., Cross-modal recurrent models for weight objective prediction from multimodal time-series data. in Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare, 2018, pp. 178–186.
103. Wang, J., & Hu, X. (2021). Convolutional neural networks with gated recurrent connections. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
104. M. Liang and X. Hu, “Recurrent convolutional neural network for object recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3367–3375.
105. M. Liang, X. Hu, and Z. B., “Convolutional neural networks with intra-layer recurrent connections for scene labeling,” in Proc. Neural Inf. Process. Syst., 2015, pp. 937–945.
106. B. Fernandez, A. G. Parlos, and W. Tsai. Nonlinear dynamic system identification using artificial neural networks (anns). In International Joint Conference on Neural Networks (IJCNN), pages 133–141, 1990.
107. G. V. Puskorius and L. A. Feldkamp. Neurocontrol of nonlinear dynamical systems with kalman filter trained recurrent networks. *IEEE Transactions on Neural Networks*, 5(2):279–297, 1994.
108. D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1. chapter Learning Internal Representations by Error Propagation, pages 318–362. MIT Press, 1986.
109. Krishnamoorthi, R., Joshi, S., Almarzouki, H. Z., Shukla, P. K., Rizwan, A., Kalpana, C., & Tiwari, B. (2022). A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. *Journal of Healthcare Engineering*, 2022.
110. Edeh, M. O., Khalaf, O. I., Tavera, C. A., Tayeb, S., Ghouali, S., Abdulsahib, G. M., ... & Louni, A. (2022). A Classification Algorithm-Based Hybrid Diabetes Prediction Model. *Frontiers in Public Health*, 10.
111. Iwendi, C., Huescas, C. G. Y., Chakraborty, C., & Mohan, S. (2022). COVID-19 health analysis and prediction using machine learning algorithms for Mexico and Brazil patients. *Journal of Experimental & Theoretical Artificial Intelligence*, 1-21.

112. Lu, H., Uddin, S., Hajati, F., Moni, M. A., & Khushi, M. (2022). A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Applied Intelligence*, 52(3), 2411-2422.
113. Chugh, M., Johari, R., & Goel, A. (2022). MATHS: Machine Learning Techniques in Healthcare System. In *International Conference on Innovative Computing and Communications* (pp. 693-702). Springer, Singapore.
114. Deberneh, H. M., & Kim, I. (2021). Prediction of Type 2 Diabetes Based on Machine Learning Algorithm. *International Journal of Environmental Research and Public Health*, 18(6), 3317.
115. Gupta, S., Verma, H. K., & Bhardwaj, D. (2021). Classification of diabetes using Naive Bayes and support vector machine as a technique. In *Operations Management and Systems Engineering* (pp. 365-376). Springer, Singapore.
116. Islam, M. T., Raza, S. R., & Kibria, M. G. (2020, December). Early prediction of heart disease using PCA and hybrid genetic algorithm with k-means. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)* (pp. 1-6). IEEE.
117. Qawqzeh, Y. K., Bajahzar, A. S., Jemmali, M., Otoom, M. M., & Thaljaoui, A. (2020). Classification of diabetes using photoplethysmogram (PPG) waveform analysis: logistic regression modeling. *BioMed Research International*, 2020.
118. Grampurohit, S., & Sagarnal, C. (2020, June). Disease prediction using machine learning algorithms. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-7). IEEE.
119. Moturi, S., & Srikanth Vemuru, D. S. (2020). Classification model for prediction of heart disease using correlation coefficient technique. *International Journal*, 9(2).
120. D. Shah, S. Patel and S. Bharati, "Heart Disease Prediction using Machine Learning Techniques," Springer Nature Singapore Pte Ltd, (2020).
121. Princy, R. J. P., Parthasarathy, S., Jose, P. S. H., Lakshminarayanan, A. R., & Jeganathan, S. (2020, May). Prediction of Cardiac Disease using Supervised Machine Learning Algorithms. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 570-575). IEEE.
122. Saw, M., Saxena, T., Kaithwas, S., Yadav, R., & Lal, N. (2020, January). Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE.
123. Soni, V. D. (2020). Chronic Disease Detection Model Using Machine Learning Techniques. *International Journal of Scientific & Technology Research*, 9(9), 262-266.
124. Indrakumari, R., Poongodi, T., & Jena, S. R. (2020). Heart disease prediction using exploratory data analysis. *Procedia Computer Science*, 173, 130-139.
125. Wu, C. S. M., Badshah, M., & Bhagwat, V. (2019, July). Heart disease prediction using data mining techniques. In *Proceedings of the 2019 2nd international conference on data science and information technology* (pp. 7-11).
126. Tarawneh, M., & Embarak, O. (2019, February). Hybrid approach for heart disease prediction using data mining techniques. In *International Conference on Emerging Internetworking, Data & Web*

- Technologies (pp. 447-454). Springer, Cham.
127. Rahman, A. S., Shamrat, F. J. M., Tasnim, Z., Roy, J., & Hossain, S. A. (2019). A comparative study on liver disease prediction using supervised machine learning algorithms. *International Journal of Scientific & Technology Research*, 8(11), 419-422.
 128. Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A., & Singh, G. (2019, July). Prediction of coronary heart disease using machine learning: An experimental analysis. In *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies* (pp. 51-56).
 129. Khan, A., Uddin, S., & Srinivasan, U. (2019). Chronic disease prediction using administrative data and graph theory: The case of type 2 diabetes. *Expert Systems with Applications*, 136, 230-241.
 130. Alanazi, R. (2022). Identification and Prediction of Chronic Diseases Using Machine Learning Approach. *Journal of Healthcare Engineering*, 2022.
 131. Gouda, W., Almurafteh, M., Humayun, M., & Jhanjhi, N. Z. (2022, February). Detection of COVID-19 Based on Chest X-rays Using Deep Learning. In *Healthcare* (Vol. 10, No. 2, p. 343). MDPI.
 132. Kumar, A., Satyanarayana Reddy, S. S., Mahommad, G. B., Khan, B., & Sharma, R. (2022). Smart Healthcare: Disease Prediction Using the Cuckoo-Enabled Deep Classifier in IoT Framework. *Scientific Programming*, 2022.
 133. Li, J. P., Nneji, G. U., James, E. C., Chikwendu, I. A., Ejiyi, C. J., Oluwasanmi, A., & Mgbejime, G. T. (2021, August). The capability of multi resolution analysis: A case study of COVID-19 diagnosis. In *2021 4th International Conference on Pattern Recognition and Artificial Intelligence (PRAI)* (pp. 236-242). IEEE.
 134. Al Rahhal, M. M., Bazi, Y., Jomaa, R. M., Zuair, M., & Al Ajlan, N. (2021). Deep learning approach for COVID-19 detection in computed tomography images. *Cmc-Computers Materials & Continua*, 2093-2110.
 135. Men, L., Ilk, N., Tang, X., & Liu, Y. (2021). Multi-disease prediction using LSTM recurrent neural networks. *Expert Systems with Applications*, 177, 114905.
 136. Ahmad, U., Song, H., Bilal, A., Mahmood, S., Alazab, M., Jolfaei, A., ... & Saeed, U. (2021). A novel deep learning model to secure internet of things in healthcare. In *Machine intelligence and big data analytics for cybersecurity applications* (pp. 341-353). Springer, Cham.
 137. Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence and Internet of Things enabled disease diagnosis model for smart healthcare systems. *IEEE Access*, 9, 45137-45146.
 138. Sevi, M., & Aydin, İ. (2020, October). COVID-19 detection using deep learning methods. In *2020 International conference on data analytics for business and industry: way towards a sustainable economy (ICDABI)* (pp. 1-6). IEEE.
 139. Martinsson, J., Schliep, A., Eliasson, B., & Mogren, O. (2020). Blood glucose prediction with variance estimation using recurrent neural networks. *Journal of Healthcare Informatics Research*, 4(1), 1-18.
 140. Zhang, J., Xie, Y., Li, Y., Shen, C., & Xia, Y. (2020). Covid-19 screening on chest x-ray images using deep learning based anomaly detection. *arXiv preprint arXiv:2003.12338*, 27.

- [141] Hemdan, E. E. D., Shouman, M. A., & Karar, M. E. (2020). Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. arXiv preprint arXiv:2003.11055.
- [142] Zhu, T., Li, K., Chen, J., Herrero, P., & Georgiou, P. (2020). Dilated recurrent neural networks for glucose forecasting in type 1 diabetes. *Journal of Healthcare Informatics Research*, 4(3), 308-324.
- [143] Cheon, S., Kim, J., & Lim, J. (2019). The use of deep learning to predict stroke patient mortality. *International journal of environmental research and public health*, 16(11), 1876.
- [144] Li, K., Liu, C., Zhu, T., Herrero, P., & Georgiou, P. (2020). GluNet: A deep learning framework for accurate glucose forecasting. *IEEE journal of biomedical and health informatics*, 24(2), 414-423.
- [145] Wang, W., Tong, M., & Yu, M. (2020). Blood glucose prediction with VMD and LSTM optimized by improved particle swarm optimization. *IEEE Access*, 8, 217908-217916.
- [146] Rashid, N., Hossain, M. A. F., Ali, M., Sukanya, M. I., Mahmud, T., & Fattah, S. A. (2020, November). Transfer Learning Based Method for COVID-19 Detection From Chest X-ray Images. In *2020 IEEE REGION 10 CONFERENCE (TENCON)* (pp. 585-590). IEEE.
- [147] Arora, P., Kumar, H., & Panigrahi, B. K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, Solitons & Fractals*, 139, 110017.
- [148] Zaitcev, A., Eissa, M. R., Hui, Z., Good, T., Elliott, J., & Benaissa, M. (2020). A Deep Neural Network Application for Improved Prediction of HbA_{1c} in Type 1 Diabetes. *IEEE journal of biomedical and health informatics*, 24(10), 2932-2941.
- [149] Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders*, 19(1), 391-403.

Figures



Figure 1

Illustration of heterogeneous sources contributing to healthcare data [11].

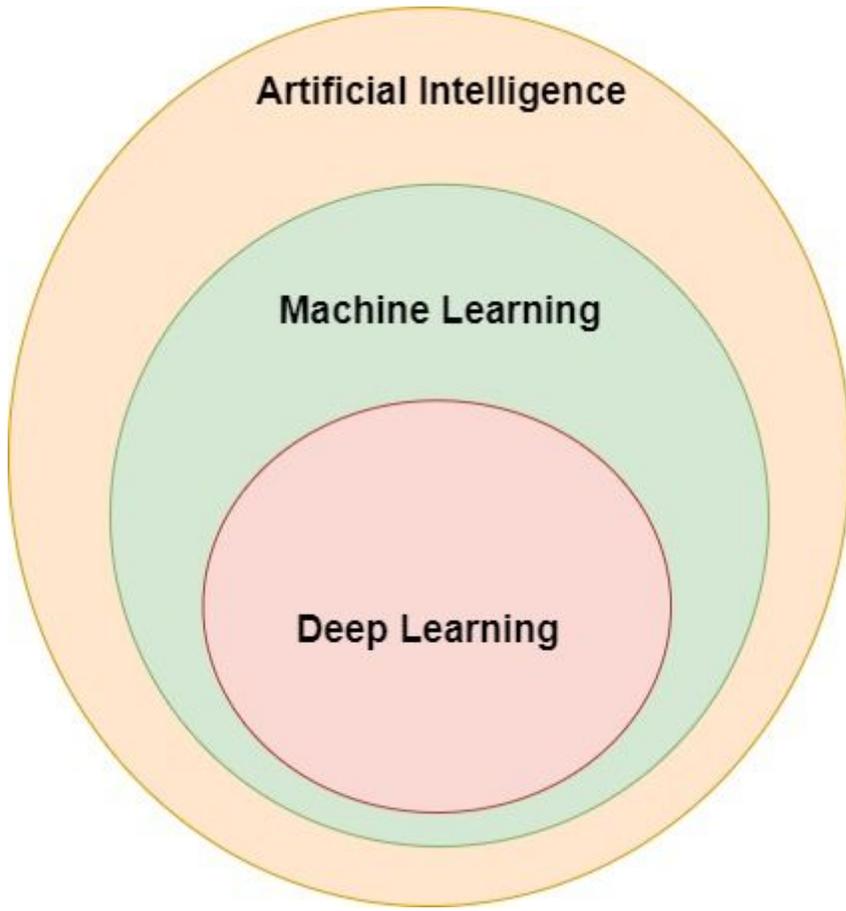


Figure 2

AI, ML, and DL.

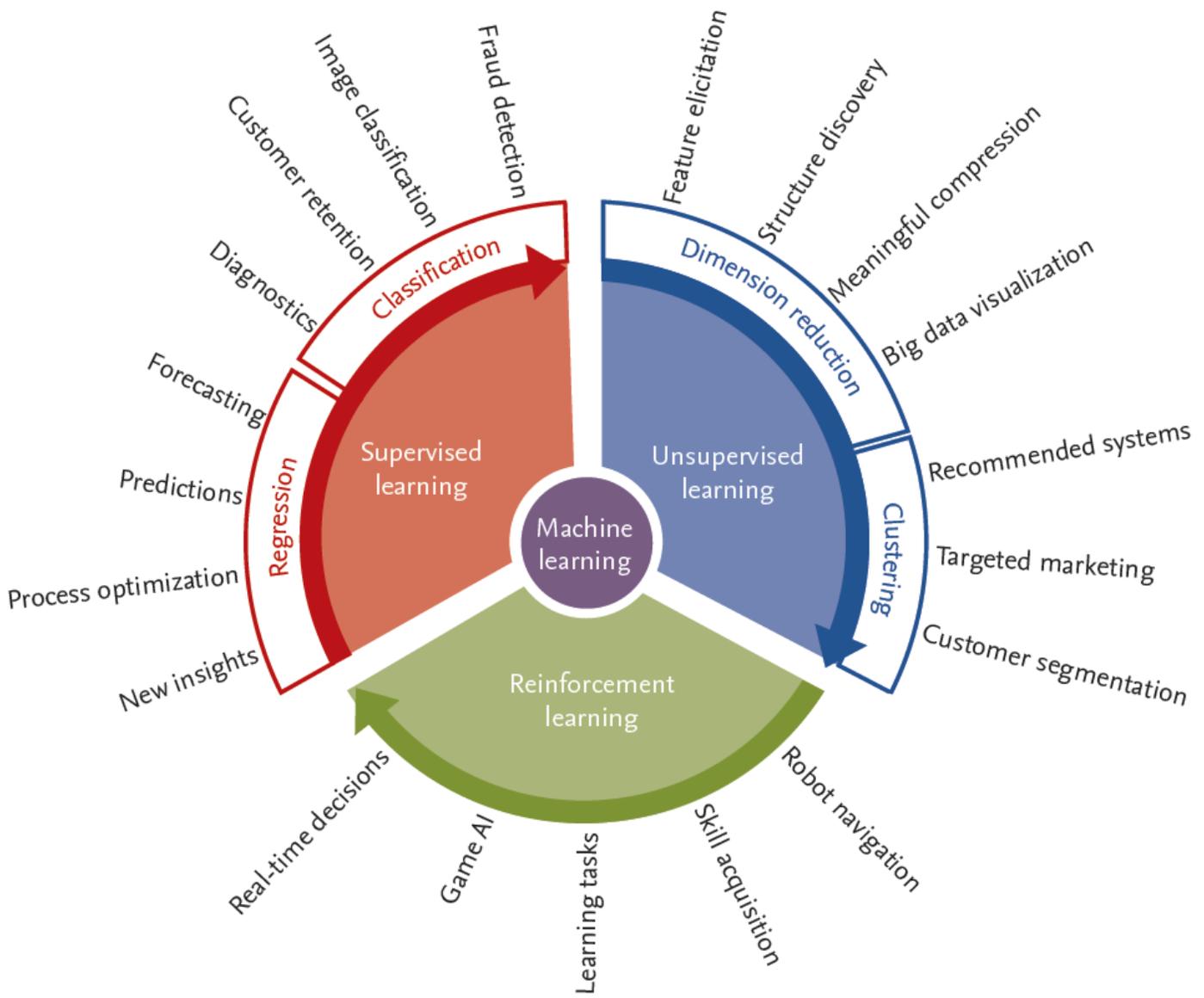


Figure 3

Different types of Machine Learning algorithms [16].

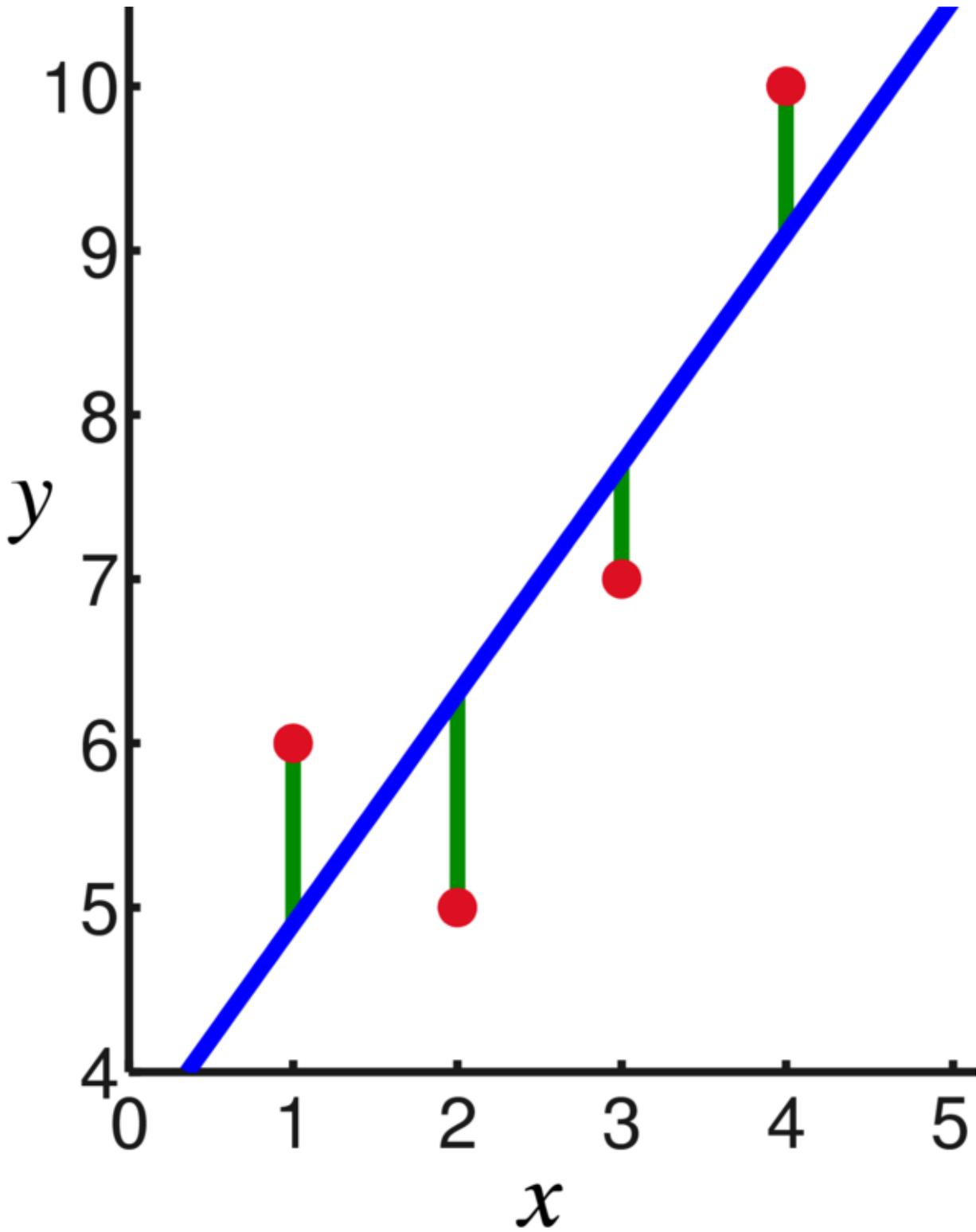


Figure 4

Linear Regression model [21].

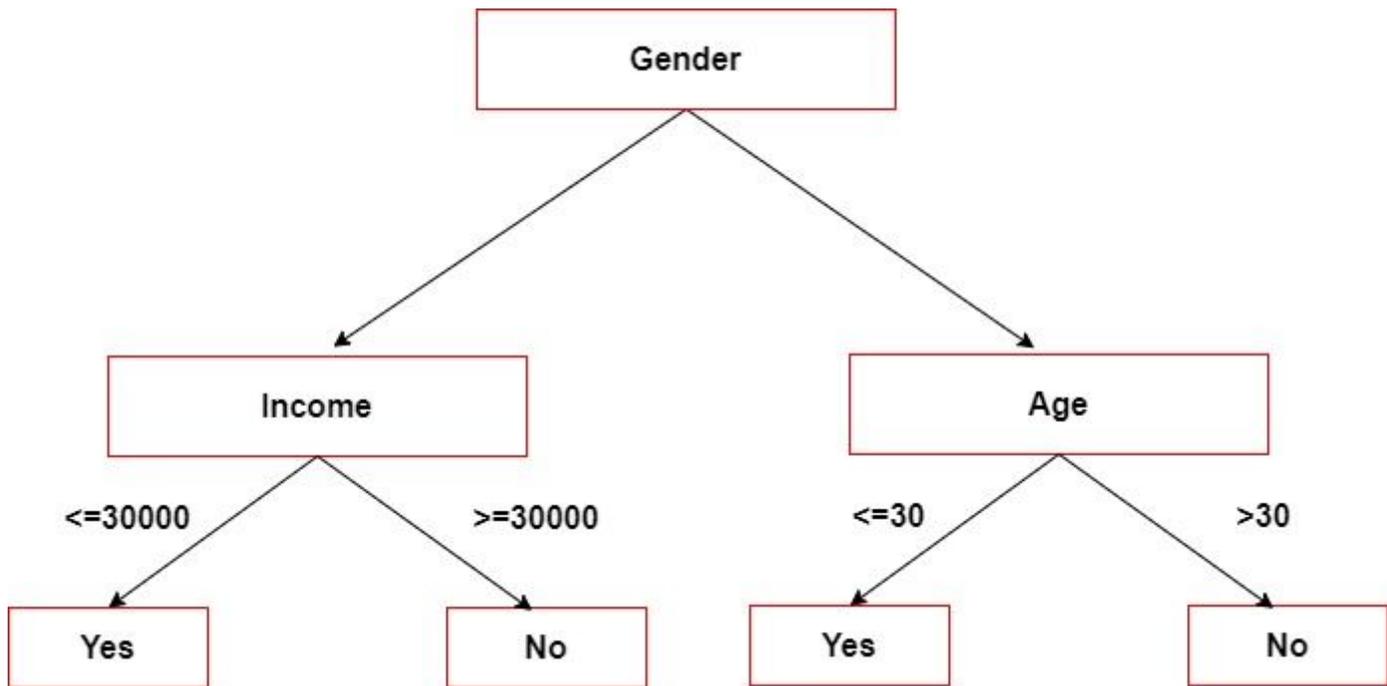


Figure 5

Example of a DT.

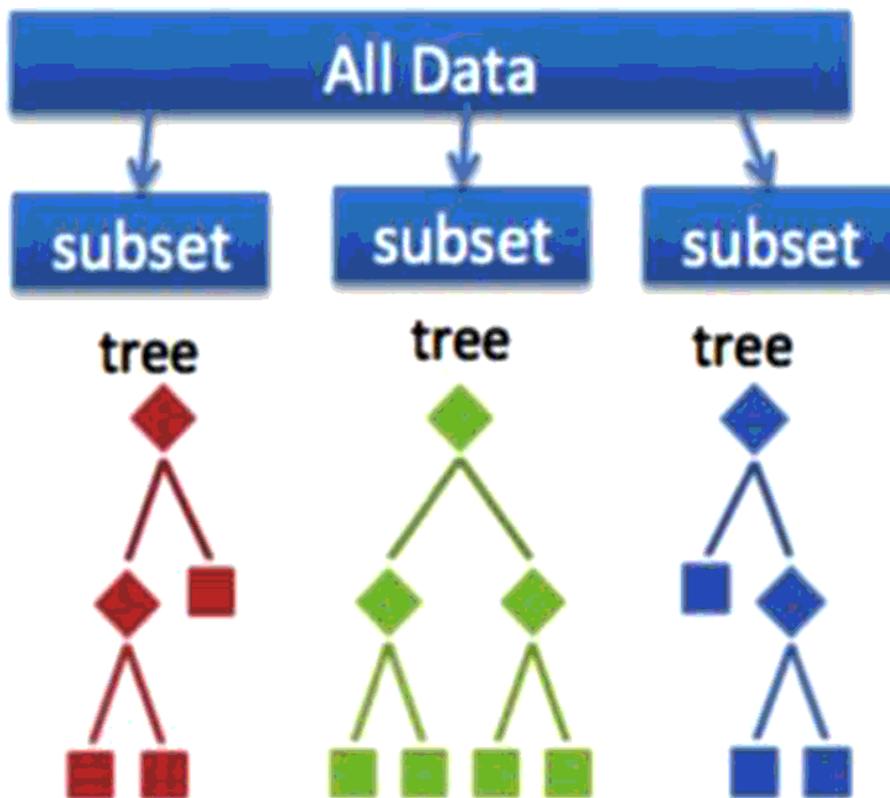


Figure 6

Introduction to SVM

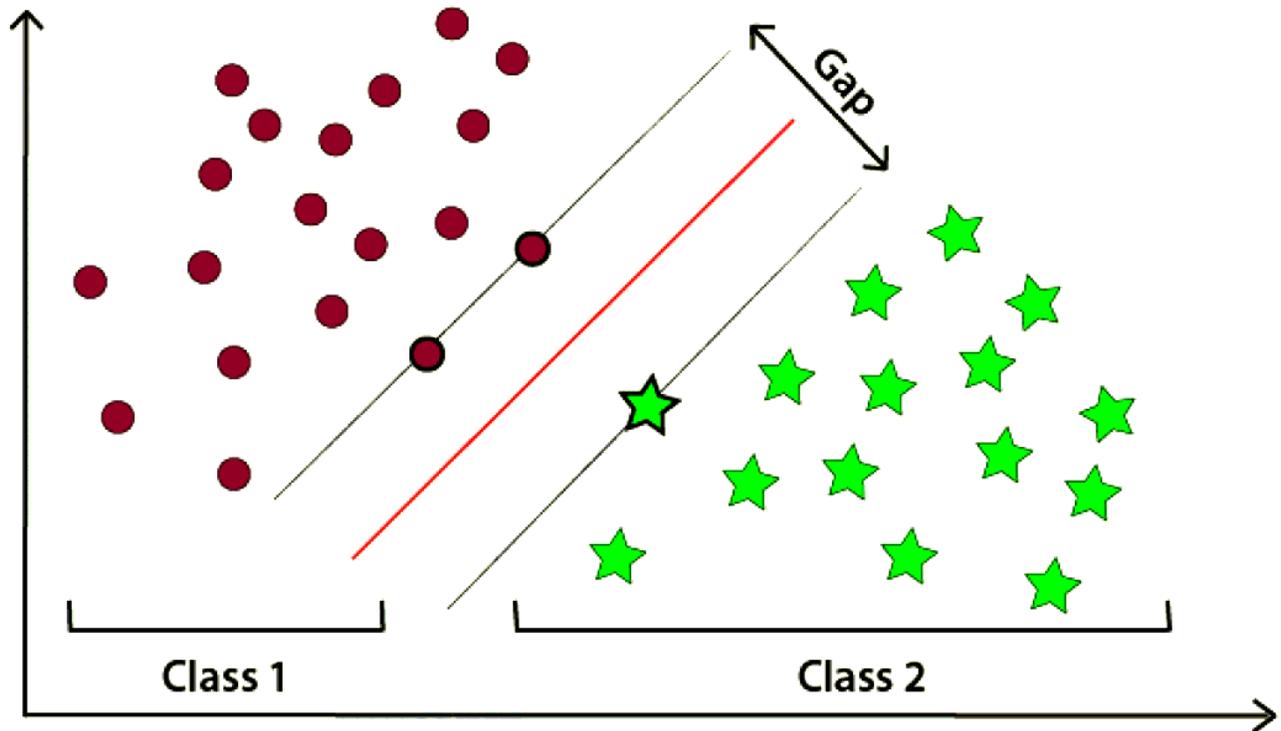


Figure 7

Support Vector Machine [19].

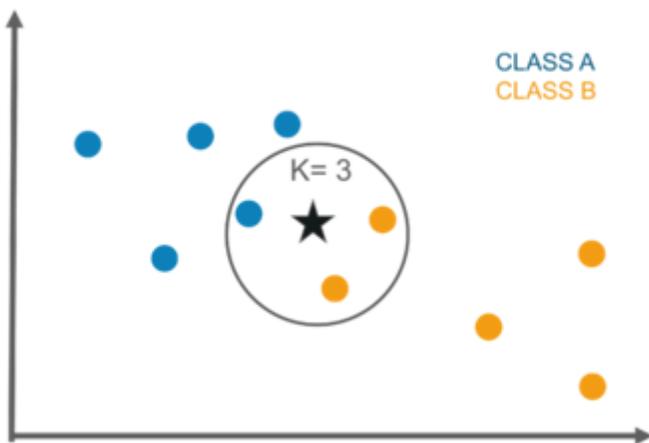


Figure 8

K-Nearest Neighbor [19].

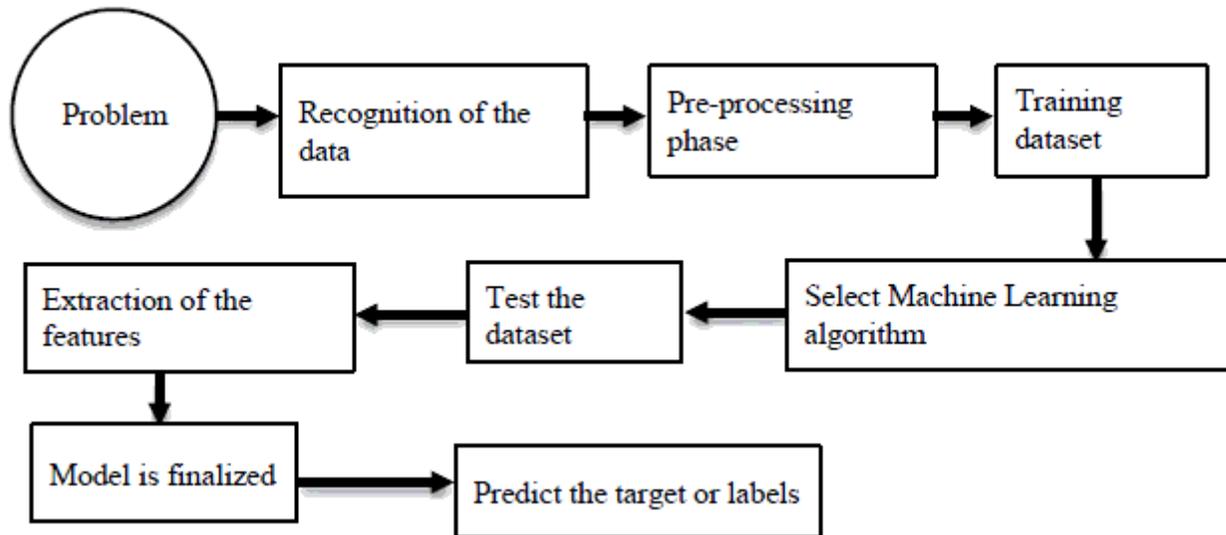


Figure 9

Workflow of unsupervised learning [24].

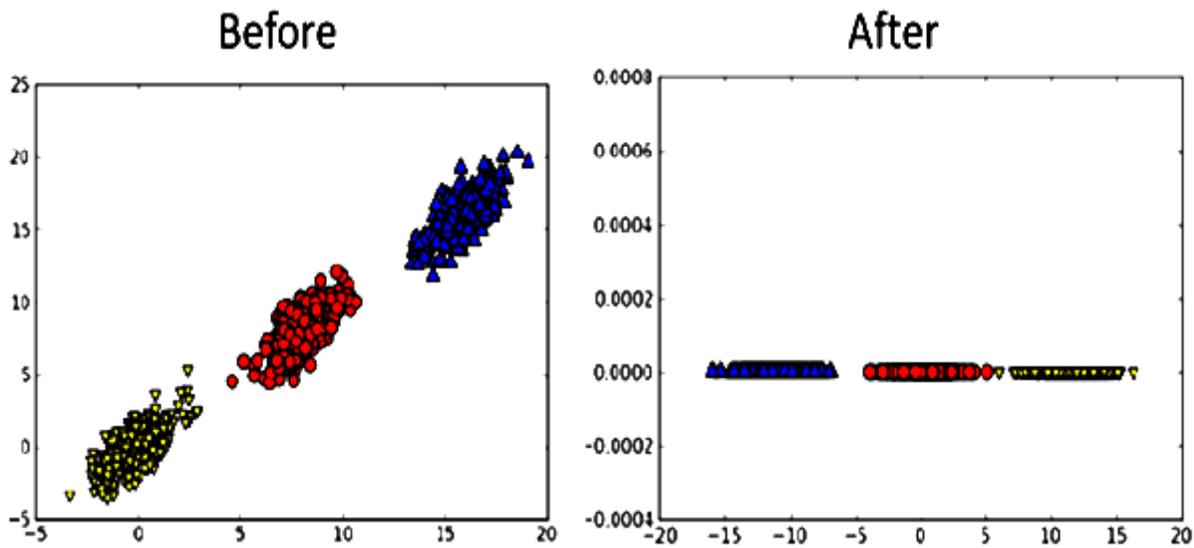


Figure 10

Visualization of data before and after applying PCA [50].

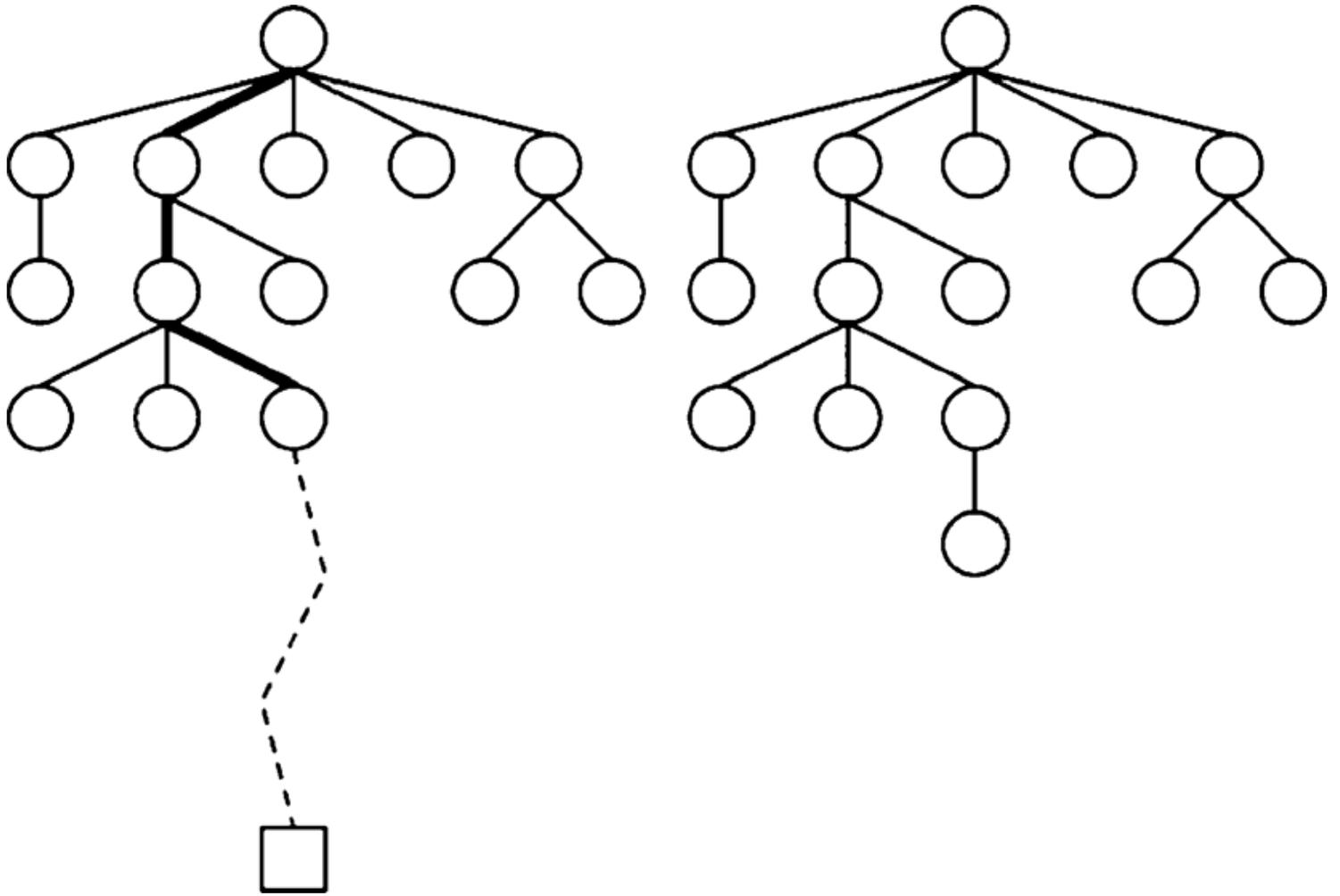


Figure 11

The basic MCTS process [71].

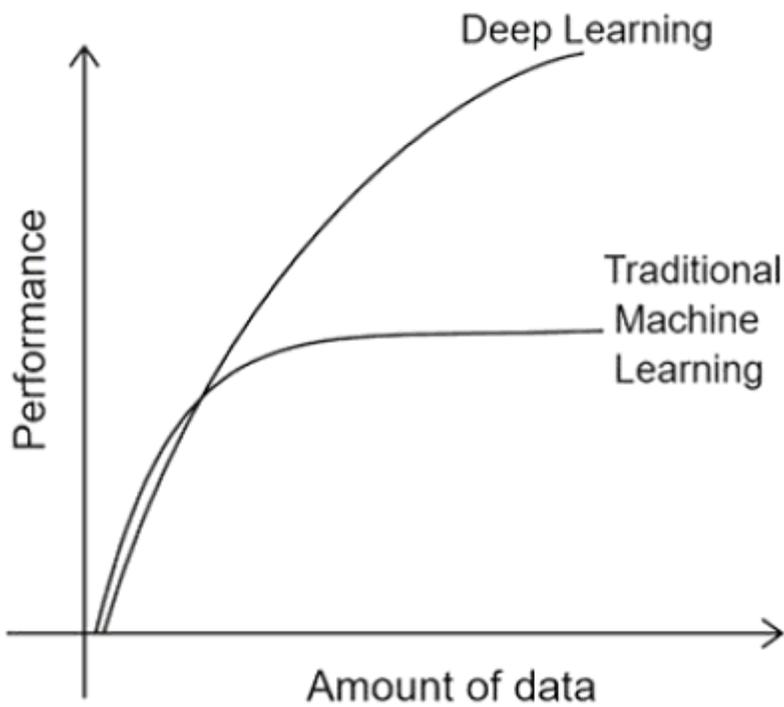


Figure 12

The performance of deep learning concerning the amount of data [78].

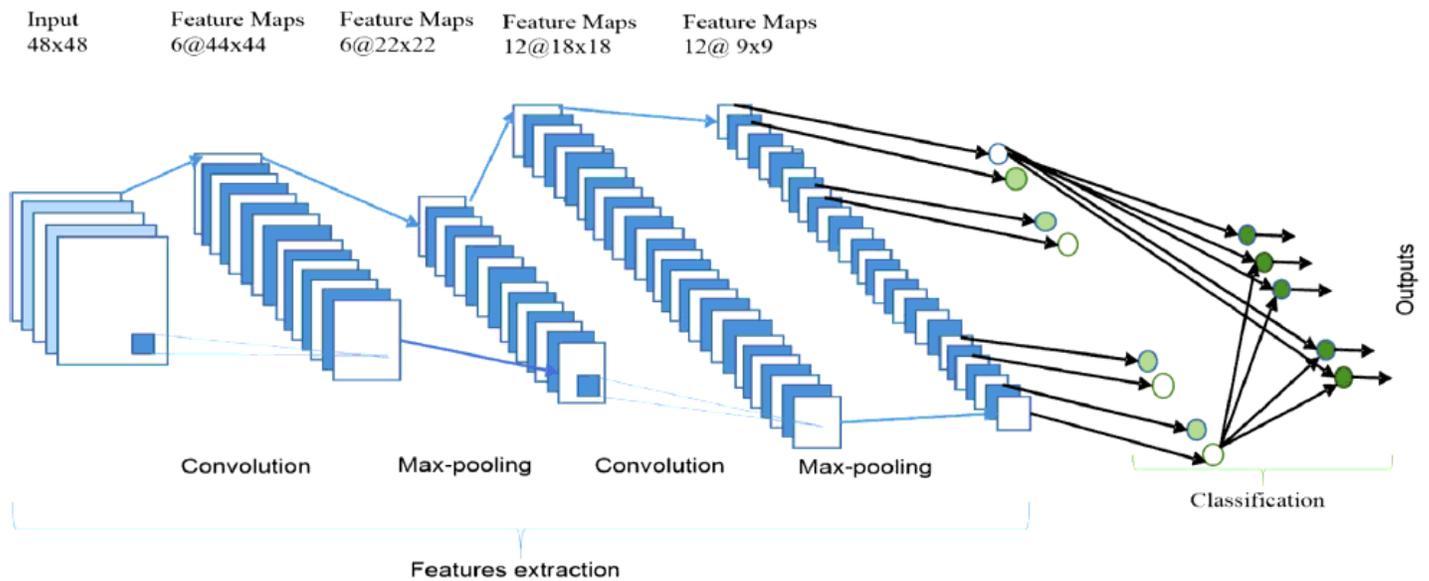


Figure 13

The architecture of CNN [84].

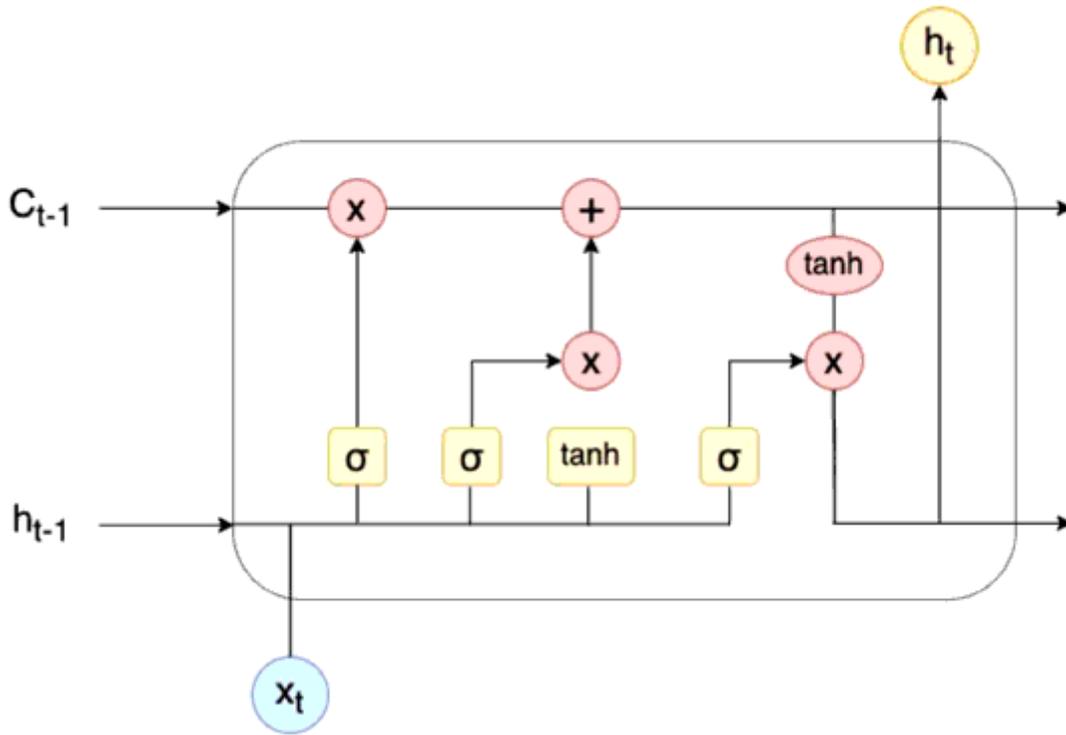


Figure 14

LSTM unit. [94].

where

x_t = Input vector at the time t .

h_{t-1} = Previous Hidden state.

c_{t-1} = Previous Memory state.

h_t = Current Hidden state.

c_t = Current Memory state.

$[x]$ = Multiplication operation.

$[+]$ = Addition operation.

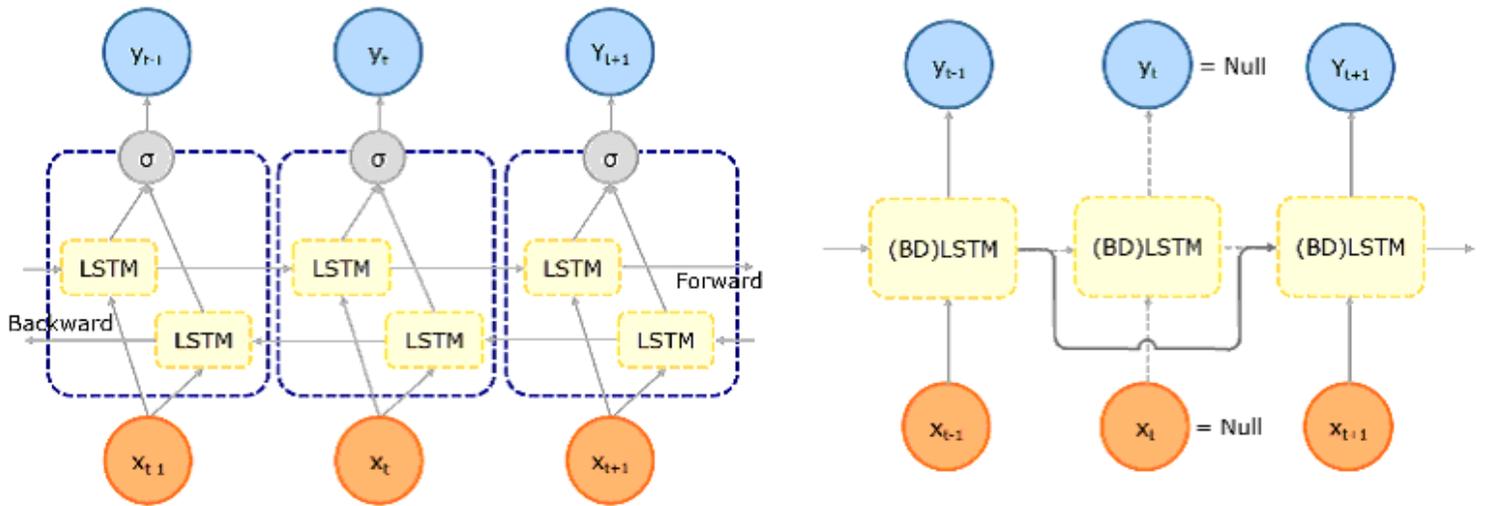


Figure 15

(left) Bidirectional LSTM and (right) filter mechanism for processing incomplete data [96].

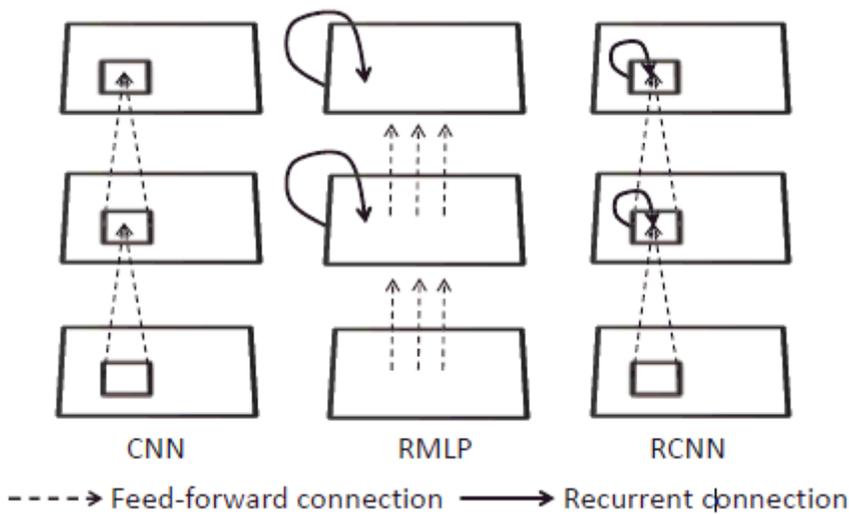


Figure 16

Illustration of the architectures of CNN, RMLP, and RCNN [97].