International Journal of Cloud Computing and Database Management

E-ISSN: 2707-5915
P-ISSN: 2707-5907
IJCCDM 2024; 5(2): 114-120
www.computersciencejournals.com/ijccdm
Received: 05-11-2024
Accepted: 30-12-2024

Dr. Md. Zillur Rahman
Faculty of MCA, SKMU, Dumka, Jharkhand, India

Corresponding Author:
Dr. Md. Zillur Rahman
Faculty of MCA, SKMU, Dumka, Jharkhand, India

A comprehensive review of machine learning: Techniques, applications and future research directions

Md. Zillur Rahman

DOI: https://doi.org/10.33545/27075907.2024.v5.i2b.80

Abstract
Machine learning (ML) is the empirical examination of algorithms and statistical models that enable computer systems to execute specified tasks autonomously, without explicit programming. Learning algorithms are used in several apps that we engage with every day. Each time an online search engine, such as Google, is used to explore the internet, its efficacy is mostly attributed to a learning algorithm that has been trained to rank web sites. These algorithms are used for diverse applications such as data mining, image processing, and predictive analytics, among others.
The primary benefit of using machine learning is that, once an algorithm has the necessary knowledge to process input, it may operate autonomously. This article provides a concise analysis and future outlook on the many uses of machine learning methods. The study examines literature on the uses of machine learning algorithms and methods across numerous domains, including medicine, agriculture, and science.

Keywords: Machine learning, knowledge, learning algorithms, google, image processing

1. Introduction
Machine learning focuses on the development and design of programs that autonomously enhance their performance via interactions inside the technological realm.
For instance, voice recognition, optical character recognition, autonomous vehicles, credit card fraud detection, and humanoid robots such as Sophia. Machine learning is a highly advanced and widely used technology nowadays. It is used in several domains such as medical research, computer science, cybersecurity, criminal investigation, data mining, artificial intelligence, and retail. Machine learning has enhanced our lives, making them cleaner and more comfortable. In the contemporary environment, we see several instances of machine learning around us.
Numerous mobile apps use machine learning to provide their services, including Uber, Google, Facebook, Twitter, and Snapchat, which we may cover later. Additionally, transportation employs machine learning in areas such as traffic signals and autonomous cars. Machine learning is undoubtedly enhancing our daily lives.

2. Basic Principle of Machine Learning
Machine learning algorithms aim to evaluate task performance via the use of examples and/or prior experience. Machine learning is often categorized into three primary types: supervised learning, unsupervised learning, and reinforcement learning.
Supervised machine learning fundamentally relies on the same principles as a conventional fitting procedure: it seeks to identify the unknown function that correlates known inputs with unknown outputs. The anticipated outcome for unidentified domain names is mostly dependent on the extrapolation of patterns identified in the labelled training data. Unsupervised learning involves identifying patterns in unlabelled data, such as in the grouping of samples. Ultimately, reinforcement learning addresses the challenge of identifying optimum or adequately suitable responses for a given scenario to maximize reward. It acquires knowledge via interactions.
Semi-supervised learning exists as an intermediary between supervised and unsupervised learning. In this instance, the algorithm is supplied with both unlabelled and categorized

data. Techniques in this area are most beneficial when data is incomplete and for learning representations.
Given that supervised learning is the predominant kind of machine learning in materials science, we will focus on it in the subsequent discussion.

3. Techniques in Machine Learning

Machine learning involves the development of algorithms that enable a machine to acquire knowledge autonomously. This is contingent upon the algorithm's input and intended result. Certain machine learning approaches will delineate how people can tackle a job. A number of mathematicians and programmers have developed answers about the methodologies and strategies for machine learning, as seen in the figure below.

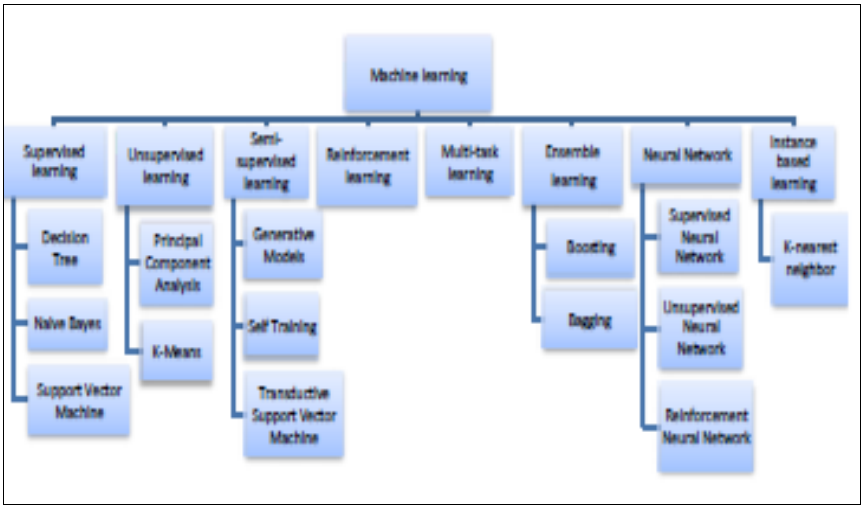

Fig 1: Machine Learning Techniques

The major machine learning approaches are discussed here, along with their corresponding applications.

Supervised Learning: In supervised learning, a proper classification is pre-assigned to train a data sample from the data source. It may also be seen as a formalization of a specific concept of learning from examples, when there exists an input and a desired result. The learner, perhaps a computer program, is supplied with a training set and a test set of data. The trainer must acquire knowledge from the labelled training set, which will be used to accurately identify the unlabelled instances in the test set. The supervised learning method requires external support. The workflow of the supervised learning approach is shown below.

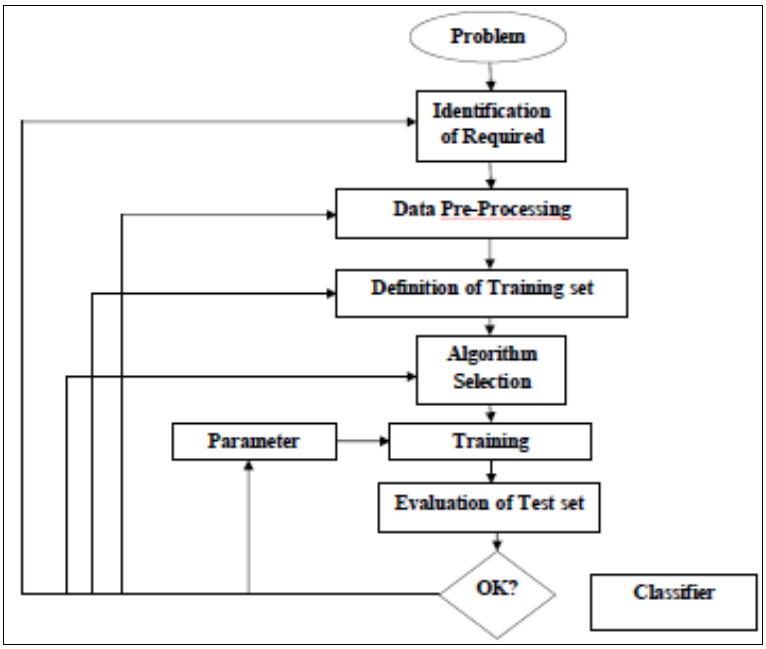

Fig 2: Workflow of supervised machine learning technique

Three main categories of supervised learning algorithms are delineated below:

Decision Tree: Decision trees categorize qualities by organizing them according to their distinct values. Each of these decision trees has nodes and branches and is mostly used for categorization. A node signifies properties inside a group, while the branch denotes a value that the node may assume. A decision tree example is shown below.

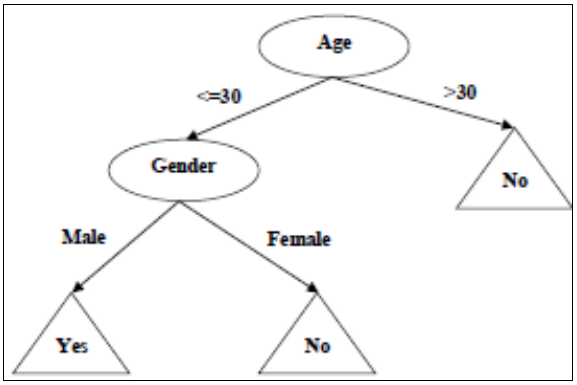

Fig 3: Decision Tree

Naive Bayes: This method is extensively used and is a focal point in the text classification sector. It is used for clustering and classification objectives. Conditional probability is fundamental to the Naive Bayes method, since it constructs trees based on the likelihood of occurrence. These trees may also be considered Bayesian Networks. An example is shown below.

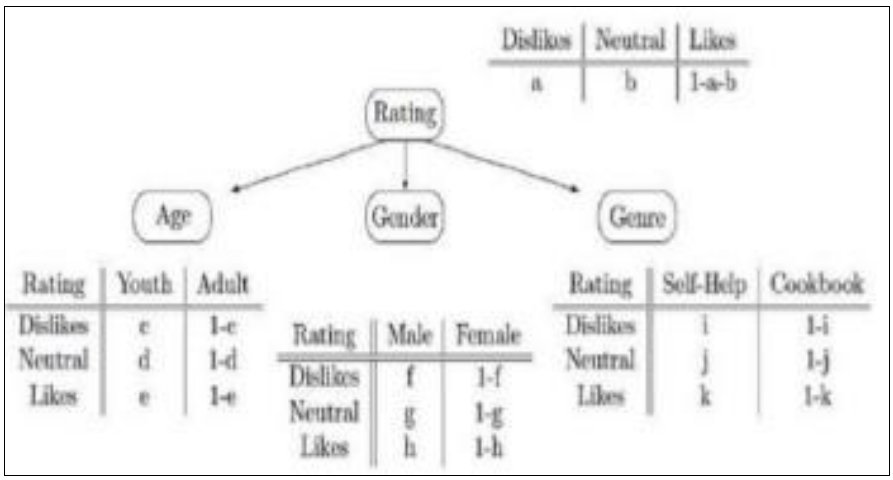

Fig 4: Bayesian Network

Support Vector Machine: This method is a prevalent and extensively used machine learning approach, mostly employed for classification. Support Vector Machine use margin computations to delineate boundaries between classes. The gap between margins and classes is consistently substantial to minimize classification mistake. The graphic below illustrates a functioning support vector machine.

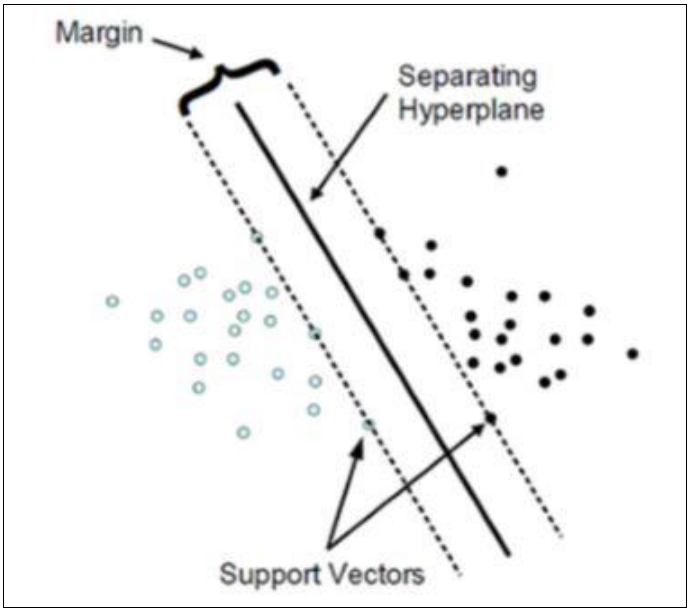

Fig 5: A working Support Vector Machine

An application of the supervised learning methodology involves using supervised data from the Stanford Natural Language Inference datasets to train and demonstrate that universal sentence representations routinely surpass unsupervised techniques.

Unsupervised Learning: This method is more challenging than supervised learning. This occurs when we instruct the computer to learn a task without providing explicit instructions on how to execute it. This learning strategy does not provide categorization but rather makes judgments that optimize rewards. Certain self-organized neural networks use unsupervised learning methods to identify hidden patterns in unlabelled data inputs. The benefit of this absence of guidance is that it enables the algorithm to retrospectively identify patterns that were not previously acknowledged. The unsupervised learning approach extracts a limited number of characteristics from a dataset and then employs these learnt features to classify fresh data upon its introduction. The unsupervised learning approach is mostly used for clustering and dimensionality reduction. An example is shown below.
Two primary algorithms for the supervised learning approach are presented below:

K-Means Clustering: It is commonly referred to as grouping, is an unsupervised learning technique that autonomously forms groups. K-means derives its name from the formation of k-distance clusters, whereby objects with like features are grouped together. The center of the cluster is derived from the mean of the values inside the cluster.

Principal Component Analysis: In this method, the dimensionality of the data is minimized to expedite and simplify computations. For instance, when principal component analysis is used on two-dimensional data, it will be reduced to one-dimensional data.

Unsupervised learning methods may be used on phrases to enhance embeddings. The unsupervised learning approach may discover phases and their transitions in systems; principle component analysis extracts the original data, while clustering identifies the phases.

Semi-Supervised Learning: This method integrates both supervised and unsupervised learning approaches. The unlabelled data is readily available, however identifying the labelled data is laborious and challenging. This approach is prevalent in the area of data mining. The following discusses some methods for the semi-supervised learning methodology.

1. **Generative Models:** This is one of the first semi-supervised learning techniques, whereby models such as Gaussian mixture models are used as a mixed distribution to infer a structure. The mixed component may be seen within the unlabelled data.
2. **Self-Training:** In this method, the classifier is taught to independently learn by supplying it with a subset of labelled data, where the labelled instances and the predicted labels are combined in the training set, followed by iterative repetition of the procedure.
3. **Transductive Support Vector Machine:** This algorithm is an extension of the Support Vector Machine algorithm that incorporates both labelled and unlabelled data. This method is used to categorize unlabelled data in a manner that maximizes the margin between labelled and unlabelled data.

Reinforcement Learning: This technique's beneficial outcomes are contingent upon the acts undertaken. The learner lacks awareness of the appropriate action to do unless presented with a specific scenario. The learner's behaviours influence future circumstances. The following is a paradigm for reinforcement learning.

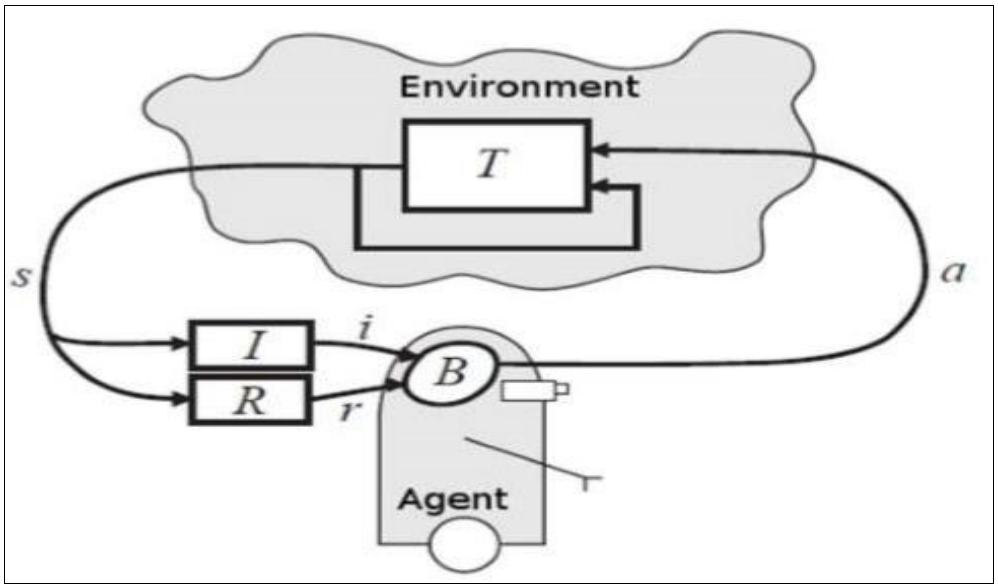

Fig 6: Reinforcement Learning Model

Multitask Learning: The algorithm retains the methodology and steps used to resolve a specific issue and the rationale behind its result. The method and technique are used to provide solutions to additional duties or issues. It may also be referred to as a transfer mechanism. Collaborative learning among students facilitates contemporaneous acquisition of knowledge, significantly expediting the learning process compared to alone study. A benchmark of four clinical predictions is provided using data from a publicly accessible medical information

database, where multitask learning, deep supervision, and data-specific adjustments are used to enhance the performance of neural models.

Ensemble Learning: This approach involves the amalgamation of individual learners to create a singular learner. For instance, a decision tree, naive Bayes, and neural network may be integrated to create an ensemble learning model. The collaboration of learners yields superior results compared to those of an individual learner. Ensemble learning comprises two algorithms.

1. **Boosting:** This method aggregates weak learners to form a single robust learner. It reduces bias and variance.
2. **Bagging:** This method, often referred to as bootstrap aggregating, is used when a machine learning system necessitates enhanced accuracy and stability.

Machine learning models are developed and presented via ensemble learning techniques, drawing on a substantial archive of malware samples and apps sourced from an antivirus provider. This aims to mitigate malware attacks on the Android platform.

Neural Network Learning: This technique is derived from the notion of neurons, which consist of four components: dendrites, nucleus, soma, and axon. The interconnection of neurons is referred to as a neural network. The artificial neural network exhibits consistent behaviour. A schematic of an artificial neural network is shown below.

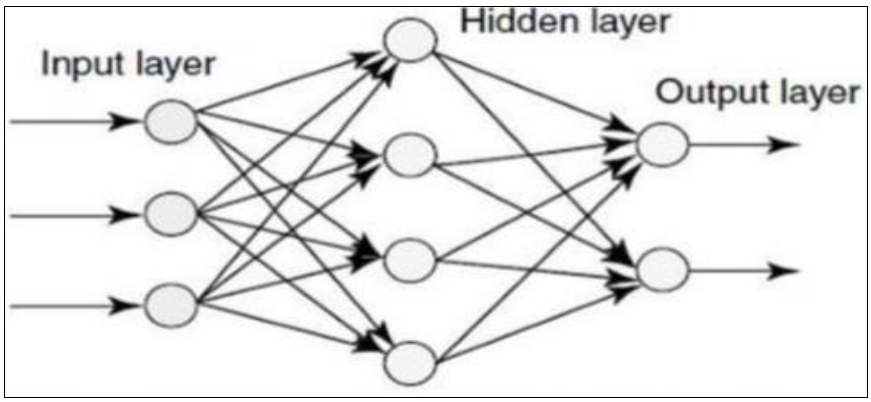

Fig 7: Artificial Neural Network

The input layer receives data, the hidden layer processes it, and the output layer delivers the computed result. Artificial Neural Networks may be categorized into three types: supervised, unsupervised, and reinforcement neural networks. Neural networks are potent and effective for challenging learning tasks such as image, audio, and natural language processing.

Instance-Based Learning: This technique involves the learner being acquainted just with a certain sort of pattern, which it attempts to apply to freshly introduced data. The strategy is referred to as lazy since it postpones action until the test data is received, subsequently using it with the training data. The complexity increases with the growth of data. An example of the instance-based learning method is the k-nearest neighbour algorithm. The instance-based learning method is used in agriculture to segment crop photographs, allowing for the automated differentiation of green textured crops from other varieties.

4. Applications of Machine Learning

Machine Learning has been ingrained in our everyday lives. In some kind or uniquely, we are using AI to manage our lives. Machine Learning is a subset of Artificial Intelligence. We often fail to see the profound integration of Machine Learning into our lives; thus, let us examine some actual applications of this technology.

Product Suggestions on E-Commerce Platforms

When we shop on an online website, we often notice that the next time we log in, it displays recommendations for comparable products and combinations for purchase. The webpage must dispatch emails about the corresponding merchandise. When using a mobile application for shopping, alerts from the app likely include discounts, coupon codes, similar goods, and often provided recommendations, as well as suggestions based on searched items and other related information.

All the aforementioned tips aim to enhance and simplify the buying experience, encouraging customers to revisit their website. All of this is accomplished using machine learning. It examines the consumer profile, wish list, orders, and cart contents to provide item predictions.

Facial Recognition in Photo Tagging

Certain internet packages in conjunction with Facebook identify the user associated with the friends likely shown in the image. The guy then tags his buddy with the proposal. On our mobile phones, photographs often provide tagging choices including the names of those present in the image from our contact list. This functionality is enabled by the use of the Machine Learning Facial Recognition algorithm. This collection of rules operates inside internet protocols and many image tagging applications.

Speech Recognition

Speech Recognition Devices such as Google Alexa and Amazon Echo may provide us with information mostly based on our spoken inquiries. When we request them to set an alarm, search for a phrase, play music, or provide flight schedules, they recognize our commands, do online searches, and then provide spoken responses. This capability is activated with the use of ML Speech Recognition Algorithms. These algorithms collect data, process the information, and improve it depending on the user's previous interactions with the gadgets.

Route and Traffic Recommendations

Google Maps Applications such as Google Maps facilitate the identification of optimal routes to reach our goal. These tips pertain to computations derived from external data about speed, vehicle locations, etc. It will retain all the data on a relevant server. Machine learning techniques assist in congestion and directional analysis.

Price Recommendations for Online Cab Booking

Duration Cab booking applications, like Uber and OLA, use Machine Learning for fare suggestions at certain times of the day. The charge increases and rate decreases are mostly derived from data collected from prior reservations and input into machine learning algorithms. These applications subsequently provide taxi booking pricing according with the rider's desire.

5.Research Directions

Our investigation into machine learning algorithms for intelligent data analysis and applications presents several research challenges in the field. In this part, we outline and examine the obstacles encountered, as well as the prospective research prospects and future initiatives.

The efficacy and efficiency of a machine learning solution are contingent upon the data's type and qualities, as well as the performance of the learning algorithms. Gathering data in pertinent fields, such as cybersecurity, IoT, healthcare, and agriculture, as stated in the section "Applications of Machine Learning," is not a simple task, despite the contemporary internet facilitating the generation of vast quantities of data at a high frequency.

Therefore, gathering relevant data for targeted machine learning applications, such as smart city initiatives, and their administration is essential for further analysis.

Consequently, a comprehensive examination of data gathering methodologies is essential when engaging with real-world data.

Furthermore, the historical data may have several unclear values, absent values, outliers, and irrelevant data. The machine learning methods addressed in Section "Machine Learning Tasks and Algorithms" significantly influence data quality and availability for training, hence affecting the resulting model. Consequently, effectively cleaning and pre-processing the varied data obtained from several sources is a formidable challenge. Consequently, it is essential to alter or expand current pre-processing methods, or to propose new data preparation approaches, to use learning algorithms successfully within the relevant application area.

Numerous machine learning methods are available to analyse data and extract insights, as outlined in Section "Machine Learning Tasks and Algorithms." Consequently, choosing an appropriate learning algorithm that aligns with the intended application is difficult. The output of various learning algorithms may change based on the qualities of the data. Choosing an inappropriate learning algorithm may yield unforeseen results, perhaps resulting in wasted effort and diminished model efficacy and accuracy.

The methodologies outlined in the section "Machine Learning Tasks and Algorithms" may be directly used to address many real-world challenges across several domains, including cybersecurity, smart cities, and healthcare, as highlighted in the section "Applications of Machine Learning."

The hybrid learning model, such as the combination of methods, modification or augmentation of current learning approaches, or the creation of new learning methods, may provide a promising avenue for future research in this field.

The final success of a machine learning system and its applications mostly relies on the data and the learning algorithms. If the data is inadequate for learning, such as being non-representative, low-quality, including irrelevant characteristics, or insufficient in number for training, the machine learning models may become ineffective or yield diminished accuracy. Consequently, efficiently processing data and managing various learning algorithms are crucial for a machine learning-based solution and ultimately developing intelligent applications.

6. Conclusion

This study presents a thorough examination of machine learning for intelligent data analysis and its applications. In alignment with our objective, we have succinctly examined the use of several machine learning methodologies in addressing numerous real-world challenges.

We assert that our study on machine learning-based solutions presents a promising avenue and may serve as a reference for prospective research and applications for both academic and industrial experts, as well as for decision-makers, from a technological perspective. This review provided an overview of machine learning, including its algorithms and methodologies. Machine learning is a rapidly emerging technology used lately to address numerous challenges. It encompasses numerous algorithms, including supervised, semi-supervised, unsupervised, and reinforcement algorithms, among others.

7. References

1. Chatterjee D, Ghosh S, Brady PR, Kapadia SJ, Miller AL, Nissanke S, Pannarale F; c2019.
2. Chen D, Mak BKW. Multitask learning of deep neural networks for low-resource speech recognition. IEEE/ACM Trans Audio Speech Lang Process. 2015;23(7):1172-1183.
3. Fu GS, Levin-Schwartz Y, Lin QH, Zhang D. Machine learning for medical imaging. J Healthc Eng; c2019.
4. Gondia A, Siam A, El-Dakhakhni W, Nassar AH. Machine learning algorithms for construction projects delay risk prediction. J Constr Eng Manag. 2019;146(1):04019085.
5. Harutyunyan H, Khachatrian H, Kale DC, Steeg GV, Galstyan A. Multitask learning and benchmarking with clinical time series data. arXiv preprint arXiv:1703.07771. 2017.
6. Harwath D, Torralba A, Glass J. Unsupervised learning of spoken language with visual context. In: Advances in Neural Information Processing Systems. 2016. p. 1858-1866.
7. Law MT, Traboulsee AL, Li DK, Carruthers RL, Freedman MS, Kolind SH, Tam R. Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. Mult Scler J Exp Transl Clin. 2019;5(4):2055217319885983.
8. Learned-Miller EG. Introduction to supervised learning. I: Department of Computer Science, University of Massachusetts; c2014.

9. Lemm S, Blankertz B, Dickhaus T, Müller KR. Introduction to machine learning for brain imaging. Neuroimage. 2011;56(2):387-399.
10. Khasawneh KN, Ozsoy M, Donovick C, Abu-Ghazaleh N, Ponomarev D. Ensemble learning for low-level hardware-supported malware detection. In: International Symposium on Recent Advances in Intrusion Detection. Springer, Cham; 2015. p. 3-25.
11. King AJ, Cooper GF, Clermont G, Hochheiser H, Hauskrecht M, Sittig DF, Visweswaran S. Using machine learning to selectively highlight patient information. J Biomed Inform. 2019;103327.
12. Lin E, Tsai SJ. Machine learning in neural networks. In: Frontiers in Psychiatry. Springer, Singapore; 2019. p. 127-137.
13. Lison P. An introduction to machine learning. Lang Technol Group (LTG). 2015;1(35).
14. Liu T, Yang Y, Huang GB, Yeo YK, Lin Z. Driver distraction detection using semi-supervised machine learning. IEEE Trans Intell Transp Syst. 2015;17(4):1108-1111.