J Med Syst (2017) 41: 69
DOI 10.1007/s10916-017-0715-6

SYSTEMS-LEVEL QUALITY IMPROVEMENT

# A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care

**Hamdan O. Alanazi**[1,2] · **Abdul Hanan Abdullah**[1] · **Kashif Naseer Qureshi**[3]

Received: 18 December 2016 / Accepted: 26 February 2017 / Published online: 11 March 2017


**Abstract** Recently, Artificial Intelligence (AI) has been used widely in medicine and health care sector. In machine learning, the classification or prediction is a major field of AI. Today, the study of existing predictive models based on machine learning methods is extremely active. Doctors need accurate predictions for the outcomes of their patients' diseases. In addition, for accurate predictions, timing is another significant factor that influences treatment decisions. In this paper, existing predictive models in medicine and health care have critically reviewed. Furthermore, the most famous machine learning methods have explained, and the confusion between a statistical approach and machine learning has clarified. A review of related literature reveals that the predictions of existing predictive models differ even when the same dataset is used. Therefore, existing predictive models are essential, and current methods must be improved.

**Keywords** Machine learning (ML) · Predictive model · Medicine and health care

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Kashif Naseer Qureshi
kashifnq@gmail.com

1 Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

2 Department of Medical Science Technology, Faculty of Applied Medical Science, Majmaah University, Al Majmaah, Kingdom of Saudi Arabia

3 Department of Computer Science, Bahria University, Islamabad, Pakistan

## Introduction

Recently, new information and communication technologies have changed the way of operations in all fields of life such as intelligent transportation systems, agriculture, education, and healthcare systems [1–3]. Artificial Intelligence (AI) has important applications for medicine and healthcare field; AI can be useful in clinical decision-making [4]. AI is defined as "the study and design of intelligent agents", where an intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success [5]. John McCarthy, who coined the term in 1955, defined it as "the science and engineering of making intelligent machines" [6]. AI can be defined as a subject that concerns computational methods for providing results or examples of behavior that are characteristics of human intelligence [7]. There are many advantages of using AI, such as flexibility, adaptability, pattern recognition, and fast computing [8]. The study of AI started approximately 25 years ago, and since that time, many brilliant computer scientists have performed AI research and have developed this field. More recently, AI has been used by medical doctors to attempt to resolve problems that are faced by the medical community. Based on knowledge that has already been gained, it appears auspicious that we can find solutions to some of the pressing problems that are faced today in the field of medicine. Machine learning is considered to be a very important discipline in the area of AI [9–12].

Machine learning methods can be categorized into supervised machine learning, unsupervised machine learning, semi-supervised machine learning and reinforcement machine learning. Currently, machine learning provides a substantial number of valuable tools for intelligent data analysis, data collection, data storage and uses large information systems. Machine learning technology is widely used for analyzing medical datasets. With machine learning, patient records and their accurate diagnoses are input into a computer program to execute a learning algorithm. The resulting classifier can subsequently be used to help physicians to diagnose new patients. In this manner, patient diagnosis can be accelerated while also becoming more accurate and reliable. Furthermore, a classifier can be used to educate student physicians on how to arrive at an accurate diagnosis. Predictive models based on machine learning provide the best support to clinicians' knowledge and experience. To reduce subjectivity, many expert systems have been created to codify and combine medical knowledge. Predictive methods can be integrated into these systems to reduce subjectivity while providing potentially useful new medical knowledge. As an example, Bratko and co-workers have created a system for understanding ECGs that uses models that are derived from expert systems [13]. When patients receive an inaccurate prognosis, it causes difficulty for the patients and their families [14]. According to doctors, time is a crucial factor in diagnoses, and arriving at an appropriate decision in a timely fashion can aid patients greatly. Thus, accurate prediction of patient outcomes poses is a challenge in healthcare [14, 15].

This paper aims to demonstrate predictive models in the field of healthcare. The study also discussed the recent challenges in the field and emphasizes to develop a novel machine learning model to provide accurate and dynamic predictions.

The rest of the paper is organized as follows: Section 2 provides research background. Section 3 presents examples of machine learning methods. Section 4 discusses existing predictive models in medicine and health care. In last section, paper concludes with future direction.

## Research Background

### Machine Learning

Machine learning refers to "computational methods for improving performance by mechanizing the acquisition of knowledge from experience" [16–18]. Machine learning and its related methods is a major branch of AI [9–12]. Machine learning is the essence of machine intelligence [19]. Machine learning aims to mimic the intelligent abilities of humans with machines. Representation and generalization are used in machine learning. Representations of data instances and functions evaluated on these instances are part of all machine learning methods. Generalization is the property according to which a machine learning method can provide predictions for previously unobserved data instances. Both supervised and unsupervised methods are used in machine learning [20].

### Supervised Machine Learning

Supervised learning is a machine learning approach that is deduced from labeled training data composed of a set of training examples. Each example in the supervised training dataset comprises a pair of input objectives, an input vector and a preferred output value (called a supervisory signal). In supervised machine learning, an algorithm analyzes the training data and produces an inferred function called a classifier. This deduced function should accurately predict the output value for any suitable input object. This approach means that the learning algorithm generalizes the training data to previously unobserved situations in a "reasonable" way. This task is called concept learning in human and animal psychology.

### Unsupervised Machine Learning

Unsupervised machine learning relates to a situation that attempts to find hidden structure in unmarked data. Because the examples given to the learner are unlabeled, there is no error or reward signal to evaluate a potential solution. The reward signal is a crucial factor that distinguishes between supervised and unsupervised machine learning. Unsupervised learning is related to density estimation in the field of statistics. Unsupervised learning algorithms are present in neural network (NN) models, including the self-organizing map (SOM) and adaptive resonance theory (ART).

### Supervised Versus Unsupervised Learning

Theoretically, both methods of learning vary in terms of their underlying structure. In supervised learning, the model describes the outcome that one set of observations (inputs) has on other observations (output). Thus, the inputs and outputs that include mediating variables are at opposite ends of the causal chain. Nevertheless, in unsupervised learning, all of the observations are assumed to be caused by latent variables, and the observations are presumed to be at the end of the causal chain. This approach leaves the probability of the inputs undefined. However, if the inputs are modeled, then the missing inputs cause no difficulty because they can be deemed to be latent variables, as in unsupervised learning.

### Semi-Supervised Machine Learning

This machine learning technique uses both marked and unmarked data during the training period. Basically, an insignificant number of marked data is used with a large number of untagged data to produce a large improvement in the learning accuracy. The attainment of a tagged dataset for learning in difficult conditions normally entails a human expert. The work involved in the tagging procedure might make creating a fully labeled training set difficult and expensive, whereas obtaining an untagged dataset is relatively inexpensive. Hence, in specific cases, semi-supervised learning might be the best solution for solving a problem.

### Reinforcement Machine Learning

Reinforcement machine learning is an approach in machine learning that states what actions an agent should take in an environment to capitalize on the idea of a cumulative reward. Reinforcement machine learning has been extensively studied in other areas, such as game theory, control theory, operations research, information theory, simulation-based optimization, statistics, and genetic algorithms. The study of reinforcement learning methods is called approximate dynamic programming. This problem has been studied in the theory of optimal control, although most studies are concerned with the existence of optimal solutions and their characterization and not with learning.

In machine learning, the environment is typically formulated as a Markov decision process (MDP), and many reinforcement learning algorithms in this context are closely related to dynamic programming techniques. Reinforcement learning is different from standard supervised learning in that correct input/output pairs are never presented, nor are sub-optimal actions explicitly corrected. The primary focus is the on-line performance, which involves finding a balance between exploration (of uncharted territory) and exploitation (of current knowledge).

### Pattern Recognition

In the area of machine learning, pattern recognition is the capability for labeling a specific input. Classification is considered to be a part of pattern recognition, such as identification of whether a patient has lung cancer. Pattern recognition algorithms usually aim to perform the most likely matching of the input values. These algorithms can be categorized based on the type of learning, namely, supervised learning methods vs. unsupervised learning methods. A combination of these two methods is called a semi-supervised learning method, which utilizes both labeled and unlabeled data. This area of pattern recognition is widely applied in many areas, such as psychology, psychiatry, ethology, cognitive science, and computer science.

### Classification and Prediction

One of the most important goals for using machine learning methods is referred to as classification or prediction. The terms ‘classification’ and ‘prediction’ can be used interchangeably [21]. Classification or prediction can be used to extract models that describe important data classes or to predict future data trends [22]. Classification or prediction is the capability of making a generalization on a dataset, which means the capability of identifying new outcomes from past data. Classification and prediction are both sensitive to missing values [23].

### Clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that similar objects are in the same group (called a cluster). The aim of clustering is to divide a dataset into subsets (clusters) or to classify objects into different groups, and the data in each cluster has the same features. Clustering is used in many fields, such as pattern recognition, image analysis, information retrieval, and bioinformatics. Clustering is not an algorithm; instead, it is a technique for solving certain problems. The fuzzy C-means (FCM) algorithm and the subtractive clustering algorithm are the most famous clustering techniques. The terms clustering, automatic classification, and numerical taxonomy can be used interchangeably.

### Machine Learning versus Statistics

In an earlier period, parallel methods were developed for machine learning and statistics. Four statisticians published a book entitled “Classification and Regression Trees” in the mid-1980s [24]. These statistical techniques have been widely adapted by machine learning researchers to improve classification performance and to make the procedure computationally well-organized.

The difference between statistics and machine learning is that the former involves testing hypotheses, whereas the latter involves the task of building knowledge and storing it in some form in the computer. The knowledge can be stored in the form of mathematical models, algorithms, or anything that can assist in determining patterns or predicting outcomes.

## Examples of Famous Machine Learning Methods

Many machine learning methods have been introduced and developed by well-known scientists and researchers for classification and prediction. The most well-known machine learning methods are discussed in the following sections.

### Artificial Neural Networks (ANNs)

ANNs, which are a computational model inspired by the connectivity of neurons to animate nervous systems, are widely used as a method for classification and prediction. Over the past two decades, a great number of approaches have been proposed for classification and prediction using neural network methods [25]. During the past 30 years, research on

ANNs has had remarkable developments, and ANNs have been widely used (Ding et al., 2011). One of the reasons for their widespread use is their ability to perform mapping [26]. Figure 1 illustrates the basic diagram of ANNs. In this figure, each black circle is referred to as a neuron, which calculates a sum of the weights from the inputs and provides a summation function. In the case of using nonlinear functions in the network, any function can be used for the mapping from the training inputs to the training outputs; provided that a sufficient number of neurons exist in the network and that there is a sufficient number of training examples.

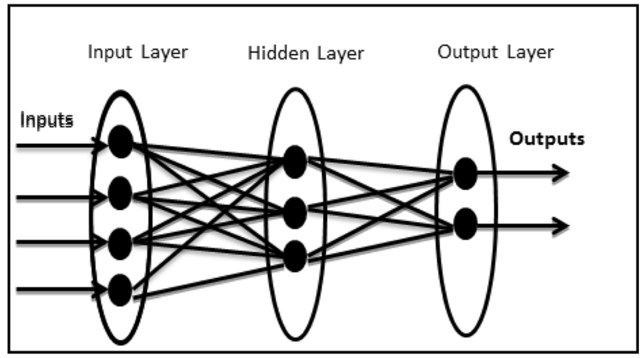


**Fig. 1** Basic Scheme of ANNs

### Support Vector Machines (SVMs)

SVMs were introduced by Cortes and Vapnik [27]. SVMs are a type of supervised machine learning method that explores datasets to identify outcomes; this approach is widely used for binary classification and prediction. A simple SVM can analyze a group of inputs from a dataset to forecast two outputs. SVMs are one type of linear binary classifier. For example, suppose that a dataset has a large number of instances and that every instance belongs to one of two classes. The SVM method creates a model that allocates the instances to one of the two classes. SVMs represent the instances in the form of points in a space that can map the instances into two classes that are separated by a gap. The new instances are also mapped into the space to be predicted and classified into one of the two classes according to their positions with respect to the gap. Moreover, SVMs can use the 'kernel trick', which can map the instances into a high-dimensional space to provide nonlinear prediction or classification [28].

### Naive Bayes Method

Bhargavi and Jyothi [29] explain that the naive Bayes method (NB) is a simple probabilistic method that assumes that the existence or absence of a specific feature of a class is independent of and unrelated to the other features, given the class variables. Despite the naive design and apparently oversimplified assumptions, NB classifiers are successfully applied to many complex real-world problems. As an example, a vehicle can be considered a car if it has an oval body, possesses an engine, has seats, and is approximately 2 m in length. Even if these traits depend on each other or on the occurrence of other features, an NB classifier considers all of these traits to independently add to the probability that an instance belongs to a category, such as a vehicle belonging to the category 'car' or a fruit belonging to the category 'apple'.

NB classifiers can be efficiently taught to classify using supervised learning. In several realistic applications, parameter approximation in NB models has been performed using the maximum likelihood method [30]. Thus, one can work with the NB model without using any Bayesian algorithm. Despite their naive design and apparently oversimplified assumptions, NB classifiers have worked quite well in many complex real-world situations. Analysis of the Bayesian classification problem indicated that there are some theoretical reasons for the apparently unreasonable efficacy of NB classifiers [29]. One notable feature of NB classifiers is that they require only a small amount of training data to approximate the parameters that are essential for categorization. Furthermore, the entire covariance matrix does not need to be ascertained; instead, only the variances of the variables for each class must be determined.

### Decision Trees

Decision trees (DTs) are another basic type of classification algorithm. DTs can solve a classification problem by continually dividing the input space to build a tree on which the nodes are as pure as possible and contain points of a single class. A new test point is categorized by moving from the top to the bottom of the tree, at each point taking a single (side) branch of the tree. This approach begins from the first node, which is the root, to the last node, which is a terminal. DTs are considered simple methods; however, they perform very well in prediction and classification applications. An increased number of inputs in the datasets increase the computational complexity. Thus, large datasets require a very large amount of memory. The C4.5 method [31] is widely used. According to [32], the software package C5.0, which combines a variant of AdaBoost with C4.5, is one of the best packages available.

### Logistic Regression (LR)

LR is a type of regression that predicts the probability of an occurrence by fitting data to a logistic function. The LR method can address a discrete or categorical value. There are two types of LR methods, namely, the binomial LR method and the multinomial LR method. The outcomes of the binary LR method (also called the binomial LR method) must have only two forms, such as "true" or "false". However, the outcomes of the multinomial LR method can be of more than two forms, such as “hot”, “normal”, or “cold”. The outcomes in the binomial LR method are usually changed to be simply “0”, which can be called a “non-case”, and “1”, which can be called a “case”. The mean of the case distribution can be indicated by M. The variance, which is equivalent to the product of all of the cases, can be represented as P. The proportion of outcomes that are cases is PC, and PN represents the non-cases, which is equivalent to (1-PC). Thus, SD, the standard deviation, can be calculated as the square root of the product of all of the cases (P). The LR method can forecast the chances that a data instance is a case or a non-case. The chances can be calculated using the deviation of the probability of cases and non-cases, as well. The chances ratio (CR) can be measured as the size of an effect in LR to find the membership, which can be classified into the cases or non-cases categories. The CR can be calculated by dividing the chances of the cases and the non-cases. Zero is the minimum value of the chances ratio, and it does not have a maximum value [33].

### K-Nearest Neighbor (K-NN) Algorithm

In machine learning, the k-NN algorithm is a technique for categorizing objects that are based on the closest training examples in the feature space. k-NN is a type of instance-based learning, or lazy learning, in which the function is estimated only locally and all of the calculations are delayed up to the prediction or classification. The k-NN algorithm is considered one of the simplest machine learning methods. An entity can be categorized by the knowledge contributed by its neighbors. An object is assigned to one of the classes of its nearest neighbors using a k value. Weighting the influences of the nearest neighbors is very valuable: this approach allows a nearest neighbor to influence more than a remote neighbor. The weight of the neighbors can be obtained by dividing by the distance. During the training of the k-NN algorithm, the neighbors are organized from a set of entities. Thus, the k-NN method is sensitive to the dataset. The rules of the k-NN algorithm yield the decision as a result [34].

## A Survey of Existing Predictive Models in Medicine and Health Care

A survey of the existing predictive models has been performed. This survey is categorized based on the developed models, as described below.

### Neural Networks

Liu, et al. [35] developed a predictive model for predicting brain death that is called EANN AAN. The experiments in this study focused on patients who sustain a severe level of brain death, which was determined based on the GCS, in the Taiwan University hospital. Two forecasting models were developed. The first forecasting model has 11 inputs, and the second model has 14 inputs. The investigators used an ROC for testing and evaluating the predictive model. The root mean square error (RMSE) of the first predictive model was 0.38, whereas the RMSE of the second predictive model, with 14 inputs, was 0.171.

In the same field, Rughani, et al. [36] aspired to create ANNs to be used as predictive models in neurosurgery and compared their predictive ability with those of regression models and clinicians. Another group of researchers, Iselin, et al. [37] used the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) to predict psychosocial function in children with traumatic brain injuries (TBIs).

Güler, et al. [38] developed a predictive model for assessing the severity of TBI using ANNs. In this research, the instances in the dataset were 32 patients with different demographic characteristics who sustained TBIs. The accuracy of this predictive model was 91%.

### Support Vector Machine

Chen, et al. [39] developed a predictive model that uses the SVM method to predict and estimate the intracranial pressure (ICP), which was derived from essential measurements that were obtained from different sources. More specifically, the essential variables included midline shift measurements and texture features that were obtained from computed tomography (CT) slices together with demographic information about the patients. Another input that was considered was the Injury Severity Score. Once all of the traits or features from the slices had been put into the model, the most essential features were chosen using the feature selection scheme.

Researchers later developed and evaluated several new descriptors of evolutionary information for sequence-based prediction of DNA- and RNA-binding residues using SVMs [40]. Their findings showed that the SVM classifiers achieved 77.3% sensitivity and 79.3% specificity for the prediction of DNA-binding residues and 71.6% sensitivity and 78.7% specificity for the RNA-binding site predictions. This finding demonstrated that the newly created SVM classifier was better and more accurate than the existing models.

Another research group [41] applied advanced quantitative MRI techniques (T1, T2 mapping, and diffusion tensor MRI) to 24 mild TBI patients and 20 matched controls. They applied an SVM to categorize the quantitative MRI data and found that univariate categorization was unproductive because of characteristics that are shared by the patient and the controls. However, multi-parametric categorization attained a sensitivity of 88% and a specificity of 75%. Additionally, Zhao, et al. [42] developed an SVM for T-cell epitope prediction with an MHC type I restricted T-cell clone. However, they demonstrated that ANNs and DT classifiers had better positive predictive values than the SVM.

### Decision Trees

Studies performed by Low, et al. [43] used DT and LR analysis on variables that included the ICP, mean arterial pressure (MAP), cerebral perfusion pressure (CPP), pressure reactivity index (PRx), multimodal monitoring parameters for assessing brain tissue oxygenation (PbtO2), and micro dialysis parameters to predict outcomes based on a dichotomized Glasgow Outcome Scale (GOS). Their predictive models achieved prediction accuracies of up to 80%.

Choi, et al. [44] used another technique for prediction called the tree technique. These authors used more than 500 serious head injury patients from the Medical College of Virginia hospitals for their studies. The findings of their studies demonstrated that the tree technique had an accuracy rate of 77.7%, which was higher than the accuracies obtained through standard prediction methods.

McQuatt, et al. [45] used DT techniques to forecast results for head injury patients. This work was based on patient data from the Edinburgh Royal Infirmary, which contained both background (demographic) data and temporal (physiological) data. Later, Flemming, et al. [46] conducted a DT analysis that combined clinical and CT scan features to predict the GOS of one (death), two (vegetative state), or three (dependence) at discharge. They analyzed 81 patients who had spontaneous lobar hemorrhage that presented within 48 h of early neurologic signs.

### Other Models

Kalpakis, et al. [47] used permutation entropy to calculate the complexity of continuous vital signs that were recorded from patients with TBI. Using permutation entropy that was calculated from early vital signs (the initial 10–20% of the patient's hospital stay time), classifiers were built to predict in-hospital mortality, and their mobility was measured by the three-month Extended GOS (GOSE). It was found that 60 patients with severe TBI produced a skewed dataset when they were evaluated for accuracy, sensitivity, and specificity. Using the early vital sign data, the overall prediction accuracy obtained using leave-one-out cross-validation was 91.67% for mortality and 76.67% for the three-month GOSE in the testing datasets.

Kuo, et al. [48] developed a predictive model that uses an LR method. They developed the predictive model to predict the outcomes of TBIs. There were 13 input variables, and 84 patients with TBIs were used in the dataset.

Pignolo and Lagani [49] performed a comparative study on dissimilar machine learning classifiers, i.e., C4.5, an SVM, NB, and k-NN, for prognosis results on VS after TBI. The researchers found that only the SVM classifier was an appropriate model for clinical use.

Sut and Simsek [50] built a predictive model to forecast the mortality of head injuries. For their predictive model, the receiver operating characteristic curve demonstrated that the AUC of the BTCR method was 0.954 and that the accuracy was 93.0%. The area under the curve of the classification and regression tree was 0.801, and the accuracy was 91.1%.

Pignolo and Lagani [49] attempted to develop a predictive model for the prognosis of the results of VS after TBI using a DT, an SVM, NB, and k-NN. The AUC for the first machine learning method, which was the DT, was 0.84, whereas the AUC of the SVM was 0.81. The AUC for the third machine learning method, which was NB, was 0.91, and the AUC of k-NN was 0.88.

Ji, et al. [51] developed predictive models for predicting the outcomes of TBIs using ML methods. These authors compared the accuracy of LR, AdaBoost, C4.5, CART, SVM, and RBF ANN. The most accurate method in their research was RBF ANN, which had an accuracy of 79.04%.

Chen, et al. [39] developed a prognostic model for delayed diagnosis. These authors subsequently contrasted the outcome, which was based on the accuracy of their model, with the approaches of the SVM, an NN, a rough set (Rosetta), and LR. The AUC of their developed model, which they called Affinity, was 0.894, and the AUC of the NN was 0.881.

Güler, et al. [38] developed a system for detecting the severity of TBIs using fuzzy logic. A total of 26 TBI patients with different demographic characteristics participated in this study. Electroencephalography, trauma, and GCS scores were used as inputs for the system, and the results indicated a noteworthy connection between the findings of the neurologists and the system output for normal, mild, and severe electroencephalography tracing data. Implementing this system will help neurologists to arrive at a quick decision for the degree of trauma using electroencephalography. The system presented a similarity of 88.98% between the neurologists' final comments and the system's findings.

Pang, et al. [52] developed five predictive models for predicting the severity of head injuries using datasets that contained 500 and 13 patients. The accuracy of the first predictive model, which was a decision tree, was 73.1%. The second predictive model, which was a LR method, had an accuracy of 70.51%. The ACC of the Bayesian Network was 65.67%. The accuracy of the ANN method was 63.38%. The accuracy of the last predictive model of this research, the discriminant analysis method, was 69.39%. Mac Donald, et al. [53] employed regression to verify whether diffusion tensor imaging can identify traumatic axonal injuries.

Studies undertaken by Steyerberg, et al. [54] aimed to establish the validity of six models for forecasting mortality after severe or moderate TBIs. Their findings highlighted the need to conduct external validation of the prognostic models. They found that LR models, which are based on large samples, can be used to classify TBI patients.

Cremer, et al. [55] aimed to develop a predictive model that is based on LR for predicting the outcomes of the severity of TBIs. The dataset used for this model contained 304 patients. The ACC of this predictive model for predicting three outcomes was 66%.

The primary objective of Amantini, et al. [56] was to develop a predictive model for the prognosis of somatosensory evoked potentials after a TBI. These authors used a dataset that was composed of 60 instances.

Hukkelhoven, et al. [57] developed a predictive model for predicting the outcomes of the severity of TBIs. Their dataset contained 2269 patients. The area under the curve for this predictive model was greater than 0.80.

Newgard, et al. [58] developed a predictive model for forecasting the outcomes of patients who require special attention. The sensitivity of this predictive model was 0.94, and the specificity was 0.063.

Rovlias and Kotsou [59] created a predictive model that was based on the classification and regression tree method for forecasting the severity of TBIs. The dataset of this research was composed of 345 patients with TBIs. The accuracy of this predictive model was 86.84%.

Andrews, et al. [60] developed predictive models for forecasting the outcomes of TBIs. These authors used the LR and discriminants analysis methods. They used a dataset that was composed of 69 patients with TBIs for one year.

[32] proposed predictive models, including an LR model and an ANN Model, to predict outcomes for patients with TBIs. A sensitivity of 88% and a specificity of 80% was found in the ROC curve of the ANN model. In contrast, the LR model achieved a sensitivity of 73% and a specificity of 68% in its ROC curve.

Lavrač [61] discussed some predictive models that can predict outcomes from medical sets. Kampfl, et al. [62] stressed that the prediction of a vegetative state after a TBI should be improved. The dataset used by these authors contained 80 patients with TBIs. They found that the existing variables could predict recovery. They suggested that MRI can be used as a variable for the prognosis of the recovery.

Lang, et al. [63] attempted to predict the outcome (dead versus alive) after severe head injury. A total of 1066 consecutive patients with GCS scores of 8 or less during the first 24 h after injury were arbitrarily divided into two groups. Data from the first group ($n = 799$) and data from the second group ($n = 267$) were used to build the models. It was found that the six-month mortality rate was 63.5%, and it was further established that the age, GCS scores, and hypotension are very important in predicting the outcomes.

Combes, et al. [64] undertook a study to determine whether the predictors could be determined early after admission to allow unfavorable outcomes to be predicted within 48 h of a severe head injury. In this study, 198 successive comatose patients who were hospitalized from 1989 to 1992 were used. LR analysis demonstrated that a combination of age, the best motor response score from the GCS, and hypoxia provided a good prediction model of unfavorable outcomes (the sensitivity was 0.93).

Choi, et al. [65] developed a predictive model that was based on using LR. They used a dataset that was composed of 786 patients with TBIs. The accuracy of this model was 94%.

Selladurai, et al. [66] developed a predictive model to predict the severity of TBIs. In their dataset, they had 109 patients.

Choi, et al. [67] analyzed a dataset that was composed of 532 patients in an attempt to find the strongest variables for predicting the outcomes of TBIs.

Choi, et al. [68] developed a predictive model that uses the LR method for predicting the outcomes of the patients with TBIs. The accuracy of this developed model was 80%.

**Table 1** Different accuracies of existing predictive models with the same dataset

| No. | Author(s) and Year | MLP | RBF | C4.5 | SVM | NB | K-NN |
|---|---|---|---|---|---|---|---|
| 1 | Pignolo and Lagani [49] | Not Used | Not Used | Sen. =0.94 and Spec. =0.58 | Sen. =0.97 and Spec. =0.65 | Sen. =0.83 and Spec. =0.77 | Sen. =0.97 and Spec. =0.44 |
| 2 | Ji, et al. [51] | Not Used | AUC = 79.04% | AUC = 75.2% | AUC = 79% | Not Used | Not Used |
| 3 | Li, et al. [77] | Sen. =0.88 and Spec. =0.80 | Sen. =0.80 and Spec. =0.80 | Not Used | Not Used | Not Used | Not Used |
| 4 | Pang, et al. [52] | ACC = 67.2354% | Not Used | Not Used | Not Used | Not Used | Not Used |

| No. | AdaBoost | CART | Log. Reg. | Dec. Tree | Discr. Analysis | Bayesian Network |
|---|---|---|---|---|---|---|
| 1 | Not Used | Not Used | Not Used | Not Used | Not Used | Not Used |
| 2 | AUC 73% | AUC =77.6% | AUC = 72.9% | Not Used | Not Used | Not Used |
| 3 | Not Used | Not Used | Sen. =0.73 and Spec. =0.68 | Not Used | Not Used | Not Used |
| 4 | Not Used | Not Used | ACC = 67.4856% | ACC = 67.1082% | ACC = 63.944 3% | ACC = 64.9531% |

## Summary

Supervised machine learning methods have been used for classification or prediction [69]. In recent research studies, SVM, ANN, NB, and DT machine learning methods have been widely used [70]. A study performed by Vink and de Haan [71] demonstrated that AdaBoost is the best machine learning method. However, a study by Ji, et al. [51] stated that the ANN method is more accurate than LR, AdaBoost, C4.5, CART, and SVM. A study undertaken by Wang, et al. [72] found that the SVM was more accurate than NB. However, further studies made by Zhao, et al. [42] compared the SVM and found that ANN and DT classifiers had better positive predictive values than the SVM. More conflicts arose when other researchers stated that the SVM is more accurate than ANNs and DTs [73]. Other researchers have claimed that SVMs do not guarantee optimality of the resulting classifier or a globally optimal classification performance [74–76]. In fact, different machine learning methods provide different levels of accuracy for the same dataset. Existing models have conflicting results, as shown in Table 1; consequently, it is important that a new model for accurate and dynamic predictions be developed.

## Conclusions

Studies of the predictions of patient outcomes are significant because they can help doctors make accurate clinical decisions. This paper highlights that timing is another significant factor that influences clinical decisions. This review demonstrates that existing machine learning methods provide different accuracies when using the same dataset. In future work, a new machine learning model to provide an accurate and dynamic prediction should be developed.

**Acknowledgment** This research is supported by Ministry of Education Higher Malaysia (MOHE) and conducted in collaboration with Research Management Center (RMC) at Universiti Teknologi Malaysia (UTM) under VOT NUMBER:R.J130000.7828.4F708.

## References

1. Qureshi, K. N., Abdullah, A. H., and Anwar, R. W., Wireless sensor based hybrid architecture for vehicular ad hoc networks, *TELKOMNIKA (Telecommunication Computing Electronics and Control),* vol. 12, pp. 942–949, 2014.
2. Qureshi, K.N., and Abdullah, A.H., Localization-based system challenges in vehicular ad hoc networks: survey. *Smart Computing Review*. 4:515–528, 2014.
3. Qureshi, K.N., and Abdullah, A.H., Adaptation of wireless sensor network in industries and their architecture, standards and applications. *World Applied Sciences Journal*. 30:1218–1223, 2014.
4. Signorini, D.F., Andrews, P.J., Jones, P.A., Wardlaw, J.M., and Miller, J.D., Predicting survival using simple clinical variables: a case study in traumatic brain injury. *J. Neurol. Neurosurg. Psychiatry*. 66:20–25, 1999.
5. Komninos, N., Intelligent cities: towards interactive and global innovation environments. *International Journal of Innovation and Regional Development*. 1:337–355, 2009.
6. Nilsson, N. J. (2014). Principles of artificial intelligence. Burlington, MA, Morgan Kaufmann.
7. Charniak, E., Riesbeck, C. K., McDermott, D. V., and Meehan, J. R., Artificial intelligence programming: Psychology Press, 2014.
8. Kumar, G., Kumar, K., and Sachdeva, M., The use of artificial intelligence based techniques for intrusion detection: a review. *Artif. Intell. Rev.* 34:369–387, 2010.
9. Bratko, I., Michalski, R. S., and Kubat, M., Machine learning and data mining: methods and applications, 1999.
10. Michie, D., Spiegelhalter, D. J., and Taylor, C. C., Machine learning, neural and statistical classification, 1994.
11. Mitchell, T., The role of unlabeled data in supervised learning, in *Proceedings of the sixth international colloquium on cognitive science*, 1999, pp. 2–11.
12. Shavlik, J. W. and Dietterich T. G., Readings in machine learning: Morgan Kaufmann, 1990.
13. Bratko, I., Mozetič, I., and Lavrač, N., *KARDIO: a study in deep and qualitative knowledge for expert systems*: MIT Press, 1990.
14. Beca, J., Cox, P., Taylor, M., Bohn, D., Butt, W., Logan, W., et al., Somatosensory evoked potentials for prediction of outcome in acute severe brain injury. *J. Pediatr.* 126:44–49, 1995.
15. Alanazi, H.O., Abdullah, A.H., and Al Juma, M., A critical review for an accurate and dynamic prediction for the outcomes of traumatic brain injury based on Glasgow outcome scale. *J. Med. Sci.* 13:244, 2013.
16. Agrawal, D., A comprehensive study of data mining and application, *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),* vol. 2, pp. pp: 249–252, 2013.
17. Jackson, J., Data mining; a conceptual overview. *Commun. Assoc. Inf. Syst.* 8:19, 2002.
18. Maojo, V. and Sanandrés, J., A survey of data mining techniques, In *International Symposium on Medical Data Analysis*, 2000, pp. 17–22.
19. McDonald, C., Machine learning: a survey of current techniques. *Artif. Intell. Rev.* 3:243–280, 1989.
20. Villalba, S.D., and Cunningham, P., An evaluation of dimension reduction techniques for one-class classification. *Artif. Intell. Rev.* 27:273–294, 2007.
21. Shekhar, S., Schrater, P.R., Vatsavai, R.R., Wu, W., and Chawla, S., Spatial contextual classification and prediction models for mining geospatial data. *IEEE Trans. Multimedia*. 4:174–188, 2002.
22. Kotsiantis, S.B., Zaharakis, I.D., and Pintelas, P.E., Machine learning: a review of classification and combining techniques. *Artif. Intell. Rev.* 26:159–190, 2006.
23. Juhola, M., and Laurikkala, J., Missing values: how many can they be to preserve classification reliability? *Artif. Intell. Rev.* 40:231–245, 2013.
24. Loh, W.Y., Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 1:14–23, 2011.
25. Izeboudjen, N., Larbes, C., and Farah, A., A new classification approach for neural networks hardware: from standards chips to embedded systems on chip. *Artif. Intell. Rev.* 41:491–534, 2014.
26. Haykin, S.S., *Kalman filtering and neural networks*. Online Library, Wiley, 2001.
27. Cortes, C., and Vapnik, V., Support-vector networks. *Mach. Learn.* 20:273–297, 1995.

28. Weinberger, K. Q., Sha, F., and Saul, L. K., Learning a kernel matrix for nonlinear dimensionality reduction, In *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 106.
29. Bhargavi, P., and Jyothi, S., Applying naive bayes data mining technique for classification of agricultural land soils. *International journal of computer science and network security*. 9:117–122, 2009.
30. Jordan, A., On discriminative vs. generative classifiers: a comparison of logistic regression and naive bayes. *Adv. Neural Inf. Proces. Syst.* 14:841, 2002.
31. Li, J. and Wong, L., Using rules to analyse bio-medical data: a comparison between C4. 5 and PCL, In *International Conference on Web-Age Information Management*, 2003, pp. 254–265.
32. Freund, Y. and Mason, L., The alternating decision tree learning algorithm, In *icml*, 1999, pp. 124–133.
33. Hosmer, D. W. Jr and Lemeshow, S., *Applied logistic regression*: Wiley, 2004.
34. Pathak, A., Sehgal, M., and Christopher, D., A Study on Selective Data Mining Algorithms, *International Journal of Computer Science Issues (IJCSI),* vol. 8, 2011.
35. Liu, Q., Cui, X., Abbod, M.F., Huang, S.-J., Han, Y.-Y., and Shieh, J.-S., Brain death prediction based on ensembled artificial neural networks in neurosurgical intensive care unit. *J. Taiwan Inst. Chem. Eng.* 42:97–107, 2011.
36. Rughani, A.I., Dumont, T.M., Lu, Z., Bongard, J., Horgan, M.A., Penar, P.L., et al., Use of an artificial neural network to predict head injury outcome: clinical article. *J. Neurosurg.* 113:585–590, 2010.
37. Iselin, G., Le Brocque, R., Kenardy, J., Anderson, V., and McKinlay, L., Which method of posttraumatic stress disorder classification best predicts psychosocial function in children with traumatic brain injury? *J. Anxiety Disord.* 24:774–779, 2010.
38. Güler, I., Tunca, A., and Gülbandilar, E., Detection of traumatic brain injuries using fuzzy logic algorithm. *Expert Systems with Applications*. 34:1312–1317, 2008.
39. Chen, Y.-W., Larbani, M., Hsieh, C.-Y., and Chen, C.-W., Introduction of affinity set and its application in data-mining example of delayed diagnosis. *Expert Systems with Applications*. 36: 10883–10889, 2009.
40. Wang, L., and Brown, S.J., BindN: a web-based tool for efficient prediction of DNA and RNA binding sites in amino acid sequences. *Nucleic Acids Res.* 34:W243–W248, 2006.
41. Aribisala, B. S., Cowie, C. J., He, J., Wood, J., Mendelow, A. D., Mitchell, P. *et al.*, Multi-parametric classification of traumatic brain injury patients using automatic analysis of quantitative MRI scans, in *International Workshop on Medical Imaging and Virtual Reality*, 2010, pp. 51–59.
42. Zhao, Y., Pinilla, C., Valmori, D., Martin, R., and Simon, R., Application of support vector machines for T-cell epitopes prediction. *Bioinformatics*. 19:1978–1984, 2003.
43. Low, D., Kuralmani, V., Ng, S.K., Lee, K.K., Ng, I., and Ang, B.T., Prediction of outcome utilizing both physiological and biochemical parameters in severe head injury. *J. Neurotrauma*. 26:1177–1182, 2009.
44. Choi, S.C., Muizelaar, J.P., Barnes, T.Y., Marmarou, A., Brooks, D.M., and Young, H.F., Prediction tree for severely head-injured patients. *J. Neurosurg.* 75:251–255, 1991.
45. A. McQuatt, P. Andrews, D. Sleeman, V. Corruble, and P. Jones, The analysis of head injury data using decision tree techniques, in *Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making*, 1999, pp. 336–345.
46. Flemming, K.D., Wijdicks, E.F., and Li, H., Can we predict poor outcome at presentation in patients with lobar hemorrhage? *Cerebrovasc. Dis.* 11:183–189, 2001.
47. Kalpakis, K., Yang, S., Hu, P. F., Mackenzie, C. F., Stansbury, L. G., Stein, D. M., *et al.*, Outcome prediction for patients with severe traumatic brain injury using permutation entropy analysis of electronic vital signs data, in *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, 2012, pp. 415–426.
48. Kuo, J.-R., Lo, C.-J., Lu, C.-L., Chio, C.-C., Wang, C.-C., and Lin, K.-C., Prognostic predictors of outcome in an operative series in traumatic brain injury patients. *J. Formos. Med. Assoc.* 110:258–264, 2011.
49. Pignolo, L., and Lagani, V., Prediction of outcome in the vegetative state by machine learning algorithms: a model for clinicians? *J. Softw. Eng. Appl.* 4:388, 2011.
50. Sut, N., and Simsek, O., Comparison of regression tree data mining methods for prediction of mortality in head injury. *Expert systems with applications*. 38:15534–15539, 2011.
51. Ji, S.-Y., Smith, R., Huynh, T., and Najarian, K., A comparative analysis of multi-level computer-assisted decision making systems for traumatic injuries. *BMC Med. Inform. Decis. Mak.* 9:1, 2009.
52. Pang, B.C., Kuralmani, V., Joshi, R., Hongli, Y., Lee, K.K., Ang, B.T., et al., Hybrid outcome prediction model for severe traumatic brain injury. *J. Neurotrauma*. 24:136–146, 2007.
53. Mac Donald, C., Dikranian, K., Song, S., Bayly, P., Holtzman, D., and Brody, D., Detection of traumatic axonal injury with diffusion tensor imaging in a mouse model of traumatic brain injury. *Exp. Neurol.* 205:116–131, 2007.
54. Steyerberg, E.W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G.S., et al., Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med*. 5:e165, 2008.
55. Cremer, O.L., Moons, K.G., van Dijk, G.W., van Balen, P., and Kalkman, C.J., Prognosis following severe head injury: development and validation of a model for prediction of death, disability, and functional recovery. *J. Trauma Acute Care Surg.* 61:1484–1491, 2006.
56. Amantini, A., Grippo, A., Fossi, S., Cesaretti, C., Piccioli, A., Peris, A., et al., Prediction of 'awakening'and outcome in prolonged acute coma from severe traumatic brain injury: evidence for validity of short latency SEPs. *Clin. Neurophysiol.* 116:229–235, 2005.
57. Hukkelhoven, C.W., Steyerberg, E.W., Habbema, J.D.F., Farace, E., Marmarou, A., Murray, G.D., et al., Predicting outcome after traumatic brain injury: development and validation of a prognostic score based on admission characteristics. *J. Neurotrauma*. 22: 1025–1039, 2005.
58. Newgard, C., Hedges, J., Stone, J., Lenfesty, B., Diggs, B., Arthur, M., et al., Derivation of a clinical decision rule to guide the interhospital transfer of patients with blunt traumatic brain injury. *Emerg. Med. J.* 22:855–860, 2005.
59. Rovlias, A., and Kotsou, S., Classification and regression tree for prediction of outcome after severe head injury using simple clinical and laboratory variables. *J. Neurosurg.* 21: 886–893, 2004.
60. Andrews, P.J., Sleeman, D.H., Statham, P.F., McQuatt, A., Corruble, V., Jones, P.A., et al., Predicting recovery in patients suffering from traumatic brain injury by using admission variables and physiological data: a comparison between decision tree analysis and logistic regression. *J. Neurosurg.* 97:326–336, 2002.
61. Lavrač, N., Selected techniques for data mining in medicine. *Artif. Intell. Med.* 16:3–23, 1999.
62. Kampfl, A., Schmutzhard, E., Franz, G., Pfausler, B., Haring, H.-P., Ulmer, H., et al., Prediction of recovery from post-traumatic vegetative state with cerebral magnetic-resonance imaging. *The Lancet*. 351:1763–1767, 1998.
63. Lang, E.W., Pitts, L.H., Damron, S.L., and Rutledge, R., Outcome after severe head injury: an analysis of prediction based upon comparison of neural network versus logistic regression analysis. *Neurol. Res.* 19:274–280, 1997.

64. Combes, P., Fauvage, B., Colonna, M., Passagia, J., Chirossel, J., and Jacquot, C., Severe head injuries: an outcome prediction and survival analysis. *Intensive Care Med.* 22:1391–1395, 1996.
65. Choi, S.C., Barnes, T.Y., Bullock, R., Germanson, T.A., Marmarou, A., and Young, H.F., Temporal profile of outcomes in severe head injury. *J. Neurosurg.* 81:169–173, 1994.
66. Selladurai, B.M., Jayakumar, R., Tan, Y.Y., and Low, H.C., Outcome prediction in early management of severe head injury: an experience in Malaysia. *Br. J. Neurosurg.* 6:549–557, 1992.
67. Choi, S.C., Narayan, R.K., Anderson, R.L., and Ward, J.D., Enhanced specificity of prognosis in severe head injury. *J. Neurosurg.* 69:381–385, 1988.
68. Choi, S.C., Ward, J.D., and Becker, D.P., Chart for outcome prediction in severe head injury. *J. Neurosurg.* 59:294–297, 1983.
69. García-Borroto, M., Martínez-Trinidad, J.F., and Carrasco-Ochoa, J.A., A survey of emerging patterns for supervised classification. *Artif. Intell. Rev.* 42:705–721, 2014.
70. Polat, H., Du, W., Renckes, S., and Oysal, Y., Private predictions on hidden Markov models. *Artif. Intell. Rev.* 34:53–72, 2010.
71. Vink, J.P., and de Haan, G., Comparison of machine learning techniques for target detection. *Artif. Intell. Rev.* 43:125–139, 2015.
72. Wang, L., Huang, C., Yang, M.Q., and Yang, J.Y., BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. *BMC Syst. Biol.* 4:1, 2010.
73. Foody, G.M., and Mathur, A., The use of small training sets containing mixed pixels for accurate hard image classification: training on mixed spectral responses for classification by a SVM. *Remote Sens. Environ.* 103:179–189, 2006.
74. Kim, H.-C., Pang, S., Je, H.-M., Kim, D., and Bang, S.Y., Constructing support vector machine ensemble. *Pattern Recognit.* 36:2757–2767, 2003.
75. M. Li and I. K. Sethi, SVM-Based Classifier Design with Controlled Confidence, in *ICPR (1)*, 2004, pp. 164–167.
76. Li, M., and Sethi, I.K., Confidence-based classifier design. *Pattern Recognit.* 39:1230–1240, 2006.
77. Li, Y.-C., Liu, L., Chiu, W.-T., and Jian, W.-S., Neural network modeling for surgical decisions on traumatic brain injury patients. *Int. J. Med. Inform.* 57:1–9, 2000.