Received: 5 December 2019 | Accepted: 1 December 2020

DOI: 10.1111/cpf.12686

REVIEW ARTICLE

Clinical Physiology and Functional Imaging

Machine learning for predictive data analytics in medicine: A review illustrated by cardiovascular and nuclear medicine examples

Antoine Jamin[1,2,3] | Pierre Abraham[4] | Anne Humeau-Heurtier[3]

[1]COTTOS Médical, Avrillé, France

[2]LERIA–Laboratoire d'Etude et de Recherche en Informatique d'Angers, Univ. Angers, Angers, France

[3]LARIS–Laboratoire Angevin de Recherche en Ingénierie des Systèmes, Univ. Angers, Angers, France

[4]Sports Medicine Department, UMR Mitovasc CNRS 6015 INSERM 1228, Angers University Hospital, Angers, France

Correspondence
Antoine Jamin, LERIA Département Informatique, Université d'Angers, 2 Boulevard de Lavoisier, 49045 Angers cedex 01, France.
Email: antoine.jamin@univ-angers.fr

Funding information
CIFRE, Grant/Award Number: 2017/1165

Abstract

The evidence-based medicine allows the physician to evaluate the risk–benefit ratio of a treatment through setting and data. Risk-based choices can be done by the doctor using different information. With the emergence of new technologies, a large amount of data is recorded offering interesting perspectives with machine learning for predictive data analytics. Machine learning is an ensemble of methods that process data to model a learning problem. Supervised machine learning algorithms consist in using annotated data to construct the model. This category allows to solve prediction data analytics problems. In this paper, we detail the use of supervised machine learning algorithms for predictive data analytics problems in medicine. In the medical field, data can be split into two categories: medical images and other data. For brevity, our review deals with any kind of medical data excluding images. In this article, we offer a discussion around four supervised machine learning approaches: information-based, similarity-based, probability-based and error-based approaches. Each method is illustrated with detailed cardiovascular and nuclear medicine examples. Our review shows that model ensemble (ME) and support vector machine (SVM) methods are the most popular. SVM, ME and artificial neural networks often lead to better results than those given by other algorithms. In the coming years, more studies, more data, more tools and more methods will, for sure, be proposed.

KEYWORDS

error-based approach, information-based approach, prediction, probability-based approach, similarity-based approach, supervised algorithm

1 | INTRODUCTION

With the evidence-based medicine, doctors have settings and data from which they can evaluate the risk–benefit ratio of a treatment. They can also analyse different information to make risk-based choices. With the emergence of new technologies in medicine, a large amount of data is recorded offering interesting perspectives with machine learning for predictive data analytics. However, the collection, storage and use of data raise ethical and security issues. Internet of things (IoT) is a field that groups all the systems that can be identified, perform measurements and do tasks while being open on the Internet. It is an emerging field (Stojkoska & Trivodaliev, 2017; Whitmore et al., 2015) that leads to various IoT applications in healthcare (Islam et al., 2015; Yin et al., 2016). Pavel et al. (2013) proposed to divide the future of IoT health care into four parts: patient-based applications, physician-based applications, treatment and prevention

and reduction of medical errors. IoT devices record and store a large amount of data (Dimitrov, 2016). The hospital information system (HIS) is another source of medical data that allows to store information about patients and hospital (Khalifa & Alswailem, 2015). The emergence of new technologies has increased the number and the type of data inside HIS. Many research studies use HIS data. With the increase of data, data processing is a current challenge (Chen et al., 2015; Jin et al., 2015). This is why many new data processing algorithms are still proposed (Fabris et al., 2017; Satija et al., 2018).

To predict a behaviour, that can help in decision-making, it is possible to use machine learning algorithms to create predictive models (Cheng et al., 2008; Kelleher et al., 2015). Machine learning is a process that extracts models from data. It is a multidisciplinary area that is used in different fields: bank, logistics, medicine, trade, … (Barboza et al., 2017; Portugal et al., 2017). In medicine, predictive data analytics is a crucial challenge to improve the diagnosis and the monitoring of patients. Machine learning for predictive data analytics in medicine is now used in many fields: oncology (Adegoke et al., 2017; Ammad-Ud-Din et al., 2017; Armero et al., 2016; Borisov et al., 2017; Coley et al., 2017; Hoogendoorn et al., 2016; Kim & Cho, 2015; Nagarajan & Upreti, 2017; Schwartzi et al., 2015), neurology (Ertuğrul et al., 2016; Jeon et al., 2017; Khan et al., 2014; Kim et al., 2015; Kramer et al., 2017; Tripoliti et al., 2013; Xia et al., 2015; Yuvaraj et al., 2014), geriatric (Deschamps et al., 2016; Fabris et al., 2016; Ivascu et al., 2017; Kabeshova et al., 2016); Wan et al., 2015), epidemiology (Khanna & Sharma, 2018; Modu et al., 2017; Wang et al., 2016), pharmacology (Bakal et al., 2018; Bendtsen et al., 2017; Huang et al., 2017; Luo et al., 2015; Oztaner et al., 2015), … (Alghamdi et al., 2016; Delibašić et al., 2018; Hu et al., 2017; Jarmulski et al., 2018; Jing et al., 2016; Montoye et al., 2017; Oztekin et al., 2018; Saleh et al., 2017; Sanz et al., 2017).

In this review, we will detail the use of machine learning for medical predictive data analytics problems (Deo, 2015) and this will be illustrated with examples from the cardiovascular and the nuclear medicine fields. Medical data can be split into two categories: medical images (Billings et al., 2017; Collij et al., 2016; Mirzaei et al., 2016; Wernick et al., 2010) and other data as medical time series, data acquired through a questionnaire, … (Bzdok & Meyer-Lindenberg, 2017; Leung et al., 2016; Nagarajan & Upreti, 2017; Samant & Agarwal, 2018). For concision purposes, our review deals with any kind of medical data excluding images.

To create this review, we performed three bibliographic researches on the Scopus platform. All the results were analysed, and we selected 57 papers from different fields of medicine. The papers, from the cardiovascular field, that were selected were those in which methods are detailed and compared. Recent papers from the nuclear medicine field are also detailed and compared.

2 | REVIEW ON SUPERVISED MACHINE LEARNING METHODS

This part is organized to present four machine learning approaches: information-based (Section 2.1), similarity-based (Section 2.2), probability-based (Section 2.3) and error-based (Section 2.4) machine learning approaches.

2.1 | Information-based approaches

In this part, we present methods that use information theory to build prediction models. Information theory was introduced in Ref. (Shannon, 1948). The quantity of information of a source is defined as (Shannon, 1948):

$$Q(t) = -\log(p(t)), \tag{1}$$

where $Q(t)$ is the information quantity of a set and $p(t)$ the probability that an event t will appear.

In this part, we develop three methods: the iterative dichotomizer 3 (Section 2.1.1), the regression tree (Section 2.1.2) and the model ensembles (Section 2.1.3).

2.1.1 | Iterative dichotomizer 3 algorithm

The iterative dichotomizer 3 (ID3) is an algorithm used to construct the shallowest decision tree. It uses decision trees, entropy and information gain, that are detailed below.

Decision Trees

Prediction with a decision tree (DT) consists in testing—in a certain order—some descriptive features to predict a state or a value.

In Figure 1, one has to start with testing the first descriptive feature $F1$. If the value of the query instance is $Q1$, the process goes to the left branch, and if it is $Q2$, the process goes to the right branch. This operation has to be repeated until the leaf node LN_i, that is the prediction.

Shannon's entropy model and information gain

Entropy is the quantification of signal irregularity. Entropy is defined as (Shannon, 1948):

$$H(t) = -\sum_{i=1}^{I} (p(t=i) \times \log_s(p(t=i))), \tag{2}$$

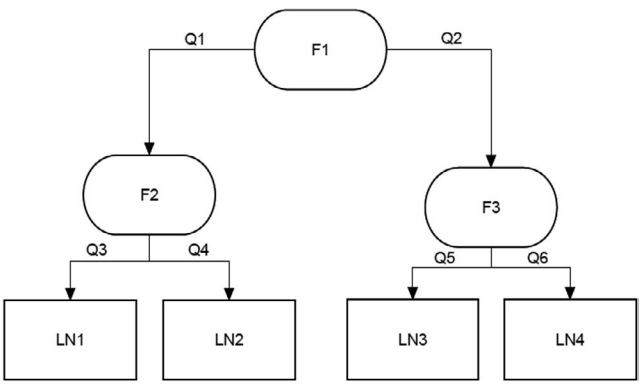

FIGURE 1 Simple example of a decision tree

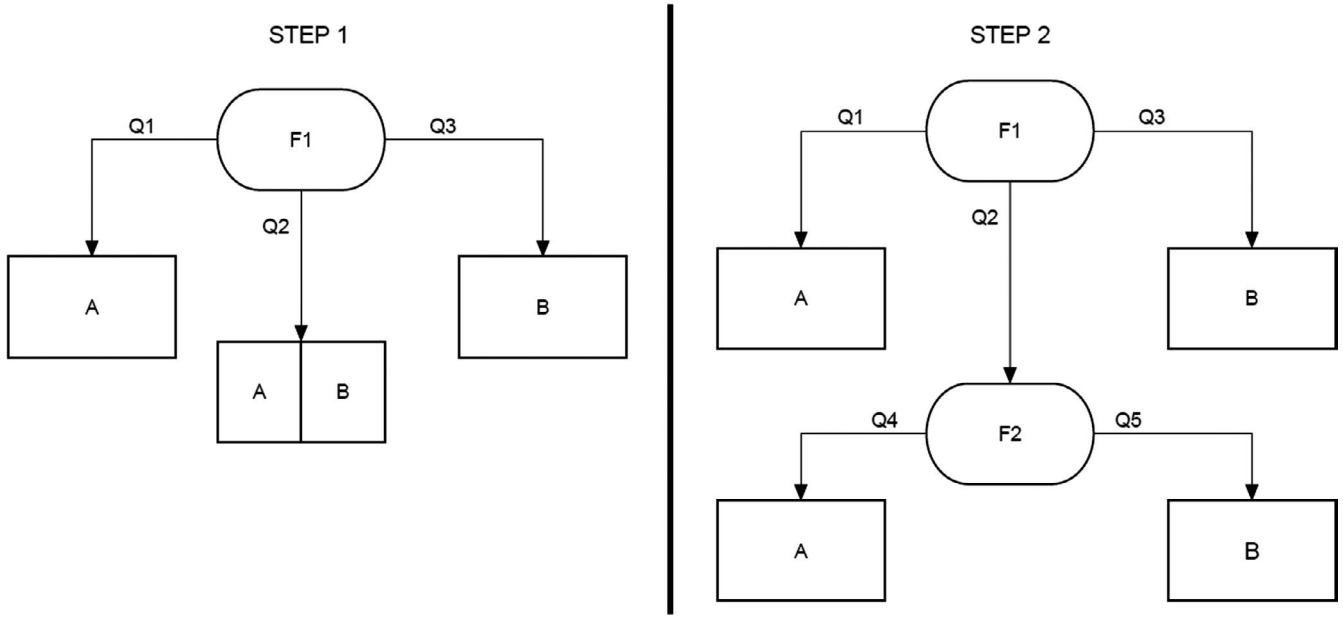

FIGURE 2 Simple example of ID3 algorithm

where $H(t)$ is the entropy of an element t, $p(t = \mathrm{i})$ is the probability that the outcome of randomly selected t is equal to i, l is the number of different values that are in the set, and s is the logarithmic base used.

To make prediction with entropy, the information gain (IG) has to be computed with a three-step algorithm:

Compute the entropy (see Equation 2) of the original data set. This gives a measure of the information that allows to organize the data set into sets.

Compute the remaining entropy, defined as:

$$\mathrm{rem}\,(d, D) = \sum_{l \in \mathrm{levels}\,(d)} \frac{|D_{d=l}|}{|D|} \times H\left(t, D_{d=l}\right), \tag{3}$$

D is the data set, d is the descriptive feature, l is the target feature level, t is the target feature, and $H(t, D_{d=l})$ is the entropy of partition $D_{d=l}$.

Compute IG defined as:

$$\mathrm{IG}\,(d, D) = H\,(t, D) - \mathrm{rem}\,(d, D)\,. \tag{4}$$

IG has some drawbacks; in particular, it does not respect the correlation between the descriptive feature and the target feature (Quinlan, 1986). To overcome the limitations of IG, information gain ratio (GR) has later been introduced in (Quinlan, 1993). It is computed as:

$$\mathrm{GR}\,(d, D) = \frac{\mathrm{IG}\,(d, D)}{H\,(d, D)} \tag{5}$$

where IG(d,D) is the information gain of feature d for the data set D (see Equation 4), and $H(d,D)$ is the entropy of the data set D with feature d (see Equation 2).

There are many measures of impurity, similar to IG or GR, that can be used in the ID3 algorithm. One of them is the Gini index, computed as (Breiman et al., 1984):

$$\mathrm{Gini}\ (t, D) = 1 - l\mathrm{levelst}) \sum p\,(t = l)^2, \tag{6}$$

where D is a data set with a target feature t, levels (t) is the set of levels in the domain of the target feature, and $p(t = \mathrm{l})$ is the probability of instance D having the target level l.

Algorithm

ID3 is an algorithm to construct the shallowest decision tree. It is computed on a training data set. This algorithm allows to split the data set with the descriptive features. To find the best feature, one has to choose the one that has the highest IG. If a partition has different target feature levels, another descriptive feature is used to redivide the rest of the data set. A leaf node is constructed where a partition has the same target level. This algorithm is stopped when each branch ends by a leaf node.

In Figure 2, we illustrate the ID3 algorithm with a two-step example:

- STEP 1: Choose the descriptive feature $F1$ (with the highest IG) and one partition is established per feature level ($Q1$, $Q2$, and $Q3$). Two partitions have only one target level value (A and B);

therefore, a leaf node has to be constructed. However, a partition groups two target levels, so it must be redivided in step 2.

- STEP 2: Choose another descriptive feature *F2* (with the highest IG) and one partition is created per feature level (*Q*4 and *Q*5). Each partition has only one target level (*A* and *B*), and a leaf node is constructed for each branch.

The main advantage of the ID3 algorithm is that it offers understandable prediction rules. However, it is not adapted for a small sample.

2.1.2 | Regression tree

To predict a continuous target, it is necessary to use a regression tree. The entropy measure in the ID3 algorithm is replaced by a variance calculation computed as:

$$\mathrm{var}(t,D) = \frac{\sum_{i=1}^{n}(t_i - \bar{t})^2}{n-1}, \tag{7}$$

where D is the data set that has reached the node, n is the number of instances in D, $\bar{t}$ is the mean of the target feature for the data set D, and t_i iterates across the target value of each instance in D.

The feature which has the lowest weighted variance for the target feature has to be chosen to construct the tree. The weighted variance is computed as:

$$d[\mathrm{best}] = arg_{d\in d[]}\, min\, \mathrm{levels}(d) \sum \frac{|D_{d=l}|}{D} \times \mathrm{var}(t, D_{d=l}), \tag{8}$$

where var (t, $D_{d=l}$) is the variance of the target feature in the partition of data set D containing the instances where $d = l$, $|D_{d=l}|$ is the size of this partition, and $|D|$ is the size of the data set.

The prediction value is obtained with the mean target feature value of the leaf node indicated by the descriptive feature of a query instance.

The main advantage of regression tree is that it does not require data preprocessing. However, it needs large data set to train the model and a small change can considerably increase the complexity of the model.

2.1.3 | Model ensembles

Definition

To make a prediction based on information, it is possible to use a model ensemble method (Zhou, 2012). It allows to create a set of models (like decision trees or other machine learning methods) where all outputs of each model are grouped. The prediction of a model, of an ensemble, does not depend on any other prediction model.

The ensemble respects these two rules: each model uses the same data set in input (but it can be modified later), and all the model predictions (outputs) are used to make a final prediction (with a voting method).

There are two common kinds of approaches to create ensembles: boosting and bagging, detailed afterwards (Quinlan, 1996).

One of the most used model ensembles is the random forest (RF) that is an ensemble of decision trees (Breiman, 2001). This is why, in sections 3 and 4, we separate random forest and model ensemble.

Boosting

Boosting creates new models that are added to the ensemble, and each new model is biased to pay more attention to instances than the previous misclassified instances (Schapire, 1999).

Firstly, a weight w_i is affected to each instance of the data set. After that, the following steps are applied for a predefined number of times:

1. Apply a classifier on the data set.
2. Calculate the total error ε that quantifies the misclassified instances of the training data set. The ε value is calculated by summing the weight of misclassified instances.
3. Increase the weight of misclassified instances: $w[i] \leftarrow w[i] \times \frac{1}{2\times\varepsilon}$. Decrease the weight of instances correctly classified by the model: $w[i] \leftarrow w[i] \times \frac{1}{2\times(1-\varepsilon)}$
4. Calculate the confidence factor: $\alpha = \frac{1}{2} \times \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$.

The number of iterations is defined to reduce the misclassification error. It often ranges from 100 to 1,000 (Mease & Wyner, 2008).

For categorical target feature, the model ensemble returns the majority target level using a weighted (defined with the confidence factor of each model) vote. For a continuous target feature, the ensemble returns the weighted mean.

Bagging/Bootstrap aggregating

Bagging (or bootstrap aggregating) method allows to establish random samples to create a model per sample. To sample the data set, the sampling with replacement is used (it is possible to have the same instance in the sample). The ensemble makes predictions by returning the majority vote or the median (better for continuous target features).

The main advantages of a model ensemble method (like random forest) are: (a) it works for categorical and continuous variables, (b) no preprocessing data is necessary, and (c) it can handle missing and noising data. However, it has a high complexity and it needs a longer training period.

2.2 | Similarity-based approaches

In this part, we present methods that compare features using a distance measurement to group data sets. These methods use a feature

space, detailed in Section 2.2.1, that represents features on a coordinate system. Algorithms are applied to group features in the space by their similarity. In Section 2.2.2, we develop the nearest neighbour algorithm and some related methods: *k* nearest neighbour and *k* − *d* tree.

2.2.1 | Feature space and similarity

A feature space represents descriptive features on a coordinate system. One dimension corresponds to a descriptive feature of the data set.

Figure 3 shows two examples of feature spaces. For more than 3 features, it is possible to have a 2D or 3D representation using a reduction method based on correlation between features, as does the principal component analysis.

A distance measure has to be used to calculate the similarity between data in the feature space. Lots of distance metrics exist:

- Euclidean $(a, b) = \sqrt{\sum_{i=1}^{m} (a[i] - b[i])^2}$, with *a* and *b* two data sets and *m* the number of descriptive features.
- Manhattan $(a, b) = \sum_{i=1}^{m} |a[i] - b[i]|$, with *a* and *b* two data sets and *m* the number of descriptive features.
- Minkowski $(a, b) = \left(\sum_{i=1}^{m} \left| a[i] - b[i] \right|^{p} \right)^{\frac{1}{p}}$, with *a* and *b* two data sets, *m* the number of descriptive features, and $p = 1,2,\ldots,\infty$ (larger values of *p* place more emphasis on large differences between feature values).

These feature spaces and metrics are used in machine learning algorithms based on similarity listed in the following section.

2.2.2 | Nearest neighbour algorithm and variants

The nearest neighbour algorithm (KNN) is based on results of a training data set. The algorithm groups data of a feature space that have the same target level and that are directly close. The distance measure used to find the nearest neighbour is the Euclidean distance (see section 2.2.1).

To generalize this algorithm, it is just necessary to follow the four-step algorithm (illustrated by Figure 4):

1. Represent the feature space of the training data set.
2. Decompose the space into regions where each region contains all the points in the space whose distance is less than the distance to any other point. This step is the Voronoi tessellation.
3. Generate the decision boundary by aggregating the neighbouring Voronoi regions that have the same target level.
4. Make the prediction of a new data according to the target level of the boundary in which the point is in the feature space.

The nearest neighbour algorithm has a good efficiency for clean data and for reasonably sized data sets. To dilute the dependency of the algorithm on individual instance, the *k* nearest neighbours algorithm has been introduced. The nearest neighbour algorithm, previously defined, is a *k* nearest neighbours algorithm with $k = 1$. In step 3 of the previous algorithm, the decision boundary is generated by aggregating Voronoi regions that have the same target level as their *k* neighbours. It is necessary to set the value of *k* to adapt the algorithm to the training data set. A high value of *k* increases the risk of loosing the true pattern.

To improve efficiency in research of the nearest neighbour, it is possible to use a binary space partition tree: the *k* − *d* tree. This

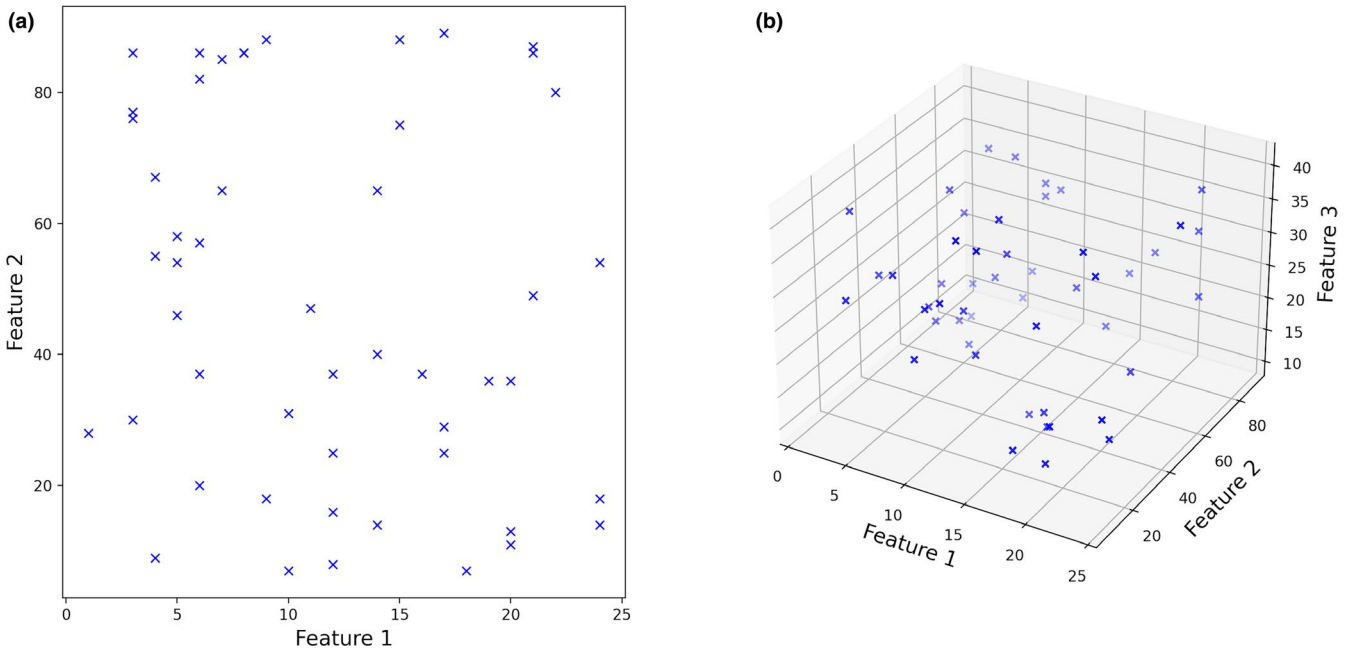

FIGURE 3 Examples of feature space. Fifty data from a data set are used to illustrate these two examples, with two descriptive features for (a) and three for (b)

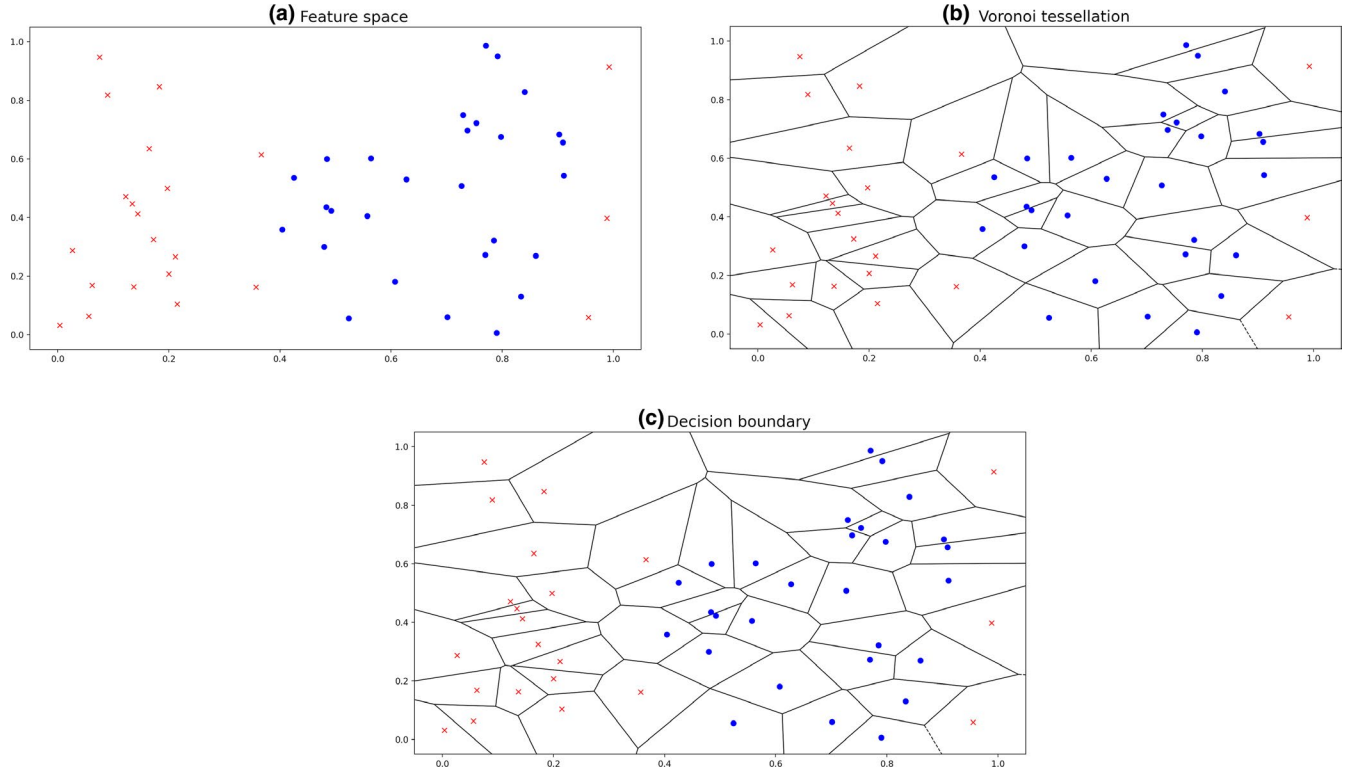

FIGURE 4 Example of the nearest neighbour algorithm. (a) is the feature space representation of fifty data from a training data set with 2 target level represented by red cross and blue circle. (b) is the Voronoi tessellation obtained with the training data set. In (c), the decision boundary which groups the same target level (circle) is coloured in yellow

method consists in creating a binary tree where each node contains one point of dimension *k*. Each node separates the feature space into two. For example, if one node divides the space along the *x*-axis, all the points with *x* coordinate (a descriptive feature of the feature space) higher than the node coordinate will be in the right branch of the node and the others will be in the left branch.

For continuous target, the *k* nearest neighbour algorithm evolves by averaging the *k* points closest to the point whose value is to be predicted.

With these similarity-based machine learning algorithms, all features should have the same variance. For this purpose, normalization techniques can be used.

Fuzzy KNN

In the KNN algorithm, each of the sample vectors is considered equally important. This property makes completion difficult when the sample sets overlap. This is why fuzzy KNN approach was introduced (Keller et al., 1985). This method consists in applying the fuzzy set theory, introduced by (Zadeh, 1965), into KNN algorithm.

The fuzzy set theory divides the sample vectors $\boldsymbol{X} = \{x_1, \ldots, x_n\}$ into *c* fuzzy partitions. For $i = 1,\ldots,c$ and $j = 1,\ldots,n$, the fuzzy membership matrix is shown by $\boldsymbol{U}$. U_{ij} is the fuzzy membership degree of x_j in class *i* (Nikoo et al., 2018).

The fuzzy KNN algorithm is defined by Algorithm 1. The result of this algorithm differs from the classical KNN procedure because it assigns the vector to a particular class.

Algorithm 1 Fuzzy KNN algorithm. This algorithm is a transcription of flowchart and algorithm presented by (Keller et al., 1985; Nikoo et al., 2018).

Let $W = \{x_1, x_2, \ldots, x_n\}$ be a set of n labelled samples.
Let $u_i(x)$ be the assigned membership of the vector x and u_{ij} be the membership in the ith class of the jth vector of the labelled sample.
INPUT : sample x (unknown sample).
Set K ($1 \leq K \leq n$), where K is the number of nearest neighbours, and n the number of labelled samples.
Initialize $i = 1$
while K Nearest neighbours to x is not determined **do**
if $i \leq K$ **then**
Include $u_i(x)$ in the set of KNN
else
Search and find x_i closer to x than any previous nearest neighbour
Delete the farthest of the KNN
Include x_i in the set of KNN
end if
$i = i+1$
end while

Initialize $i = 1$
while Fuzzy membership degree of all classes is not determined do
Compute $u_i(x)$ using the following equation:

$$u_i(x) = \frac{\sum_{j=1}^{K} \frac{u_{ij}}{\|x - x_j\|^{\frac{2}{m-1}}}}{\sum_{j=1}^{K} \frac{1}{\|x - x_j\|^{\frac{2}{m-1}}}} \quad (9)$$

where K is the predefined number of nearest neighbour and m is a constant parameter that determines the weight of each nearest neighbour in the calculation of the fuzzy membership value.
$i = i + 1$
end while

The main advantages of KNN and fuzzy KNN algorithms are that they are easy to understand, they do not need training step because they do not construct model, and they can be used for classification and regression purposes. However, they need homogeneous data, they are slow algorithms, and they are outlier-sensitive.

2.3 | Probability-based approaches

In this part, we review methods that use the probability theory and particularly Bayes' theorem. These approaches are focused on the probability of a feature level.

Bayes' theorem defines the conditional probability of an event as:

$$p(X|Y) = \frac{p(Y|X)p(X)}{p(Y)}, \quad (10)$$

where X and Y are events, $p(X|Y)$ and $p(Y|X)$ are, respectively, the likelihood of event X occurring given that Y is true and event Y, while X is true, and $p(X)$ and $p(Y)$ are the likelihoods of observing X and Y.

In this part, we develop four methods: Bayesian prediction (Section 2.3.1), naive Bayes model (Section 2.3.2), Bayesian networks (Section 2.3.3) and artificial neural network (Section 2.3.4).

2.3.1 | Bayesian prediction

The Bayesian prediction uses the likelihood that a target feature takes a specific level. The probabilities are calculated from a training data set. The prediction is calculated by taking the level value of the maximal conditional probability. The standard equation is:

$$M(q) = \arg_{l \in \text{levels}(t)} max\, p(t = l|q), \quad (11)$$

where $M(q)$ is the prediction of the model M, q the query composed by all the descriptive features, t is the target feature, and levels (t) the levels that the target can take.

With Bayes' theorem and as $p(q)$ is independent of l, the previous equation can be simplified (Mitchell, 1997):

$$M(q) = arg_{l \in \text{levels}(t)} max\, p(q|t = l) \times p(t = l). \quad (12)$$

2.3.2 | Naive Bayes model

It is important to consider the dependence between descriptors of q: generally, they are not purely independent. Conditional independence must be introduced in the calculation of the Bayesian prediction to obtain the naive Bayes Model (NB) defined as:

$$M(q) = arg_{l \in \text{levels}(t)} max\, (\, (\prod_{i=1}^{m} p(q[i]t = l) \times p(t = l)\,), \quad (13)$$

where t is the target feature, levels (t) the levels that the target can take, and q the query composed by all the descriptive features.

This method is effective for categorical prediction. However, it does not work for the prediction of continuous target, because one has to calculate how often the event occurred but a continuous feature can have an infinite number of values in its domain.

Sometimes, a model is unable to return a prediction. To solve this problem, it is possible to smooth the probability in the naive Bayes method with a smoothing probability method like Laplace smoothing.

For continuous sources, the probability calculation of descriptors, in the naive Bayes model, with an estimation has to be replaced by a probability density function (PDF):

1. plot the histogram of the continuous descriptors for each target level.
2. choose the PDF corresponding to histograms. Many kinds of PDF exist: exponential, normal, mixture of Gaussian, student-t, ...
3. set parameter of the selected distance to the descriptive feature values of the data set.

It is also possible to convert a continuous feature in a categorical feature with binning, a method to separate a data set of continuous data to a number of bins defined by the binning frequency.

The main advantages of naive Bayes algorithm are that it is easy to implement, it requires less training data, and it can be used with missing values and new integrated data. However, it is not adapted for data set with a small number of features.

2.3.3 | Bayesian networks

Bayesian network (BN) is a graph-based method that models the structural relationships between features. A network is composed of 3 elements:

- node that represents a feature,
- edge that represents the connectivity between nodes. This connection represents the influence and the conditional independence between nodes, and
- conditional probability tables (CPT) that represents the probability distribution of the feature.

Bayesian network is built with a training data set. The main advantages of the Bayesian networks method are that it is easily understandable, it can use different kinds of data, and it can be used with missing values and new integrated data. However, it is sometimes limited with continuous data, and it needs large data sets.

Figure 5 shows a Bayesian network example built to predict a target level.

2.3.4 | Artificial neural network

Artificial neural network (ANN) is a computing system that models a complex real-word problem with an ensemble of functions. It is composed by three types of layers: input, hidden and output. Each hidden layer represents a function that describes a part of the system. All the layers are connected to form an ANN. They are usually optimized by probabilistic learning methods, specifically Bayesian. For each system, there is at least one different ANN. Figure 6 shows a Bayesian network example.

A neuron (hidden layer) is defined by two functions:

- Activation function: $a_j(t+1)=f(a_j(t), p_j(t), \theta_j)$, where a_j is the neuron state, $p_j(t)$ is the input and θ_j the threshold.
- Output function: $o_j(t) = f_{out}(a_j(t))$.

The neuron input, $p_j(t)$, is also called propagation function that computes the input value from the outputs of predecessor neurons.

The main advantages of the ANN method are that it has the capacity for parallel processing, and it can be used with incomplete data sets. However, it is not easy to understand.

2.4 | Error-based approaches

In this part, we present methods that consist in building a parameterized model that minimizes the total error across the prediction. We focus our study on three methods: linear regression (Section 2.4.1), logistic regression (Section 2.4.2) and support vector machine (Section 2.4.3).

2.4.1 | Linear regression

To predict continuous target (temperature, blood-pressure, …), it is possible to use a simple linear regression (LR). This method consists in modelling a training data set by a line ($y = ax + b$) that minimizes the error function (distance between points and line).

It exists lots of error functions, the most frequently used is the sum of squared errors (L_2) that calculates the distance between points of the training data set (D) and the values obtained by the linear model (M). It is defined by:

$$L_2(M,D) = \frac{1}{2}\sum_{i=1}^{n}(t_i - M(d_i))^2, \quad (14)$$

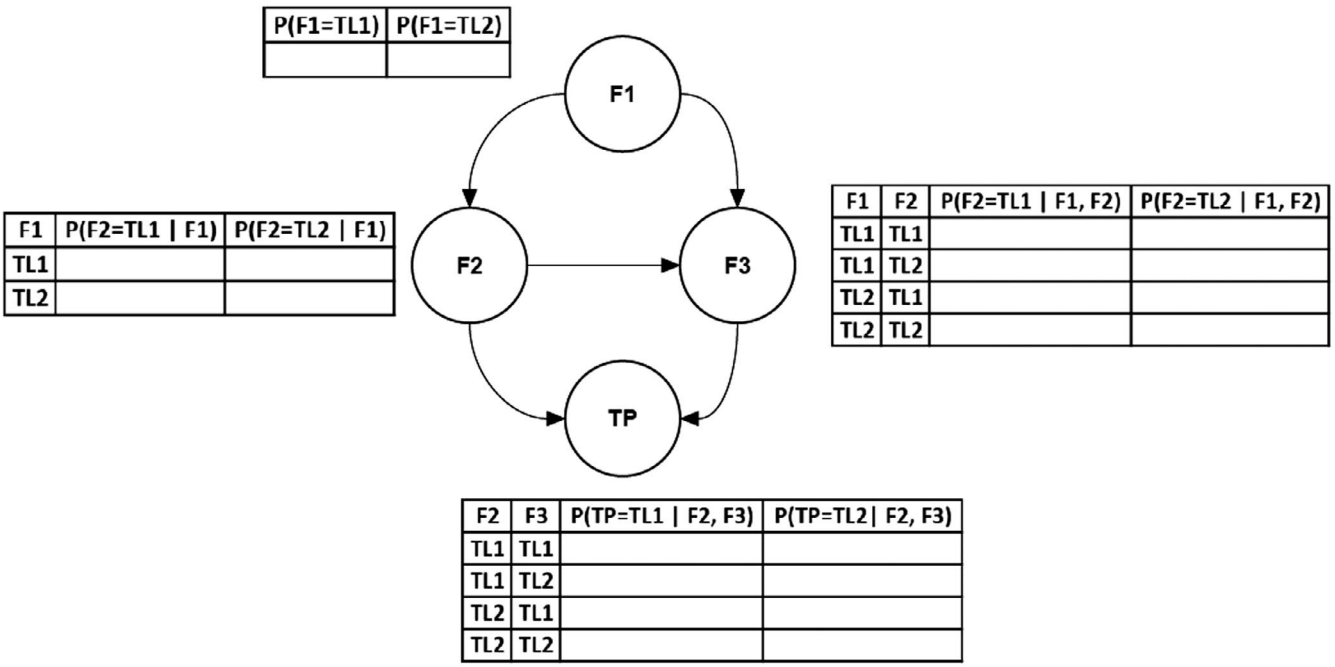

FIGURE 5 Example of a Bayesian network. F1, F2 and F3 are features. TL1 and TL2 are target levels. TP is the target level to predict

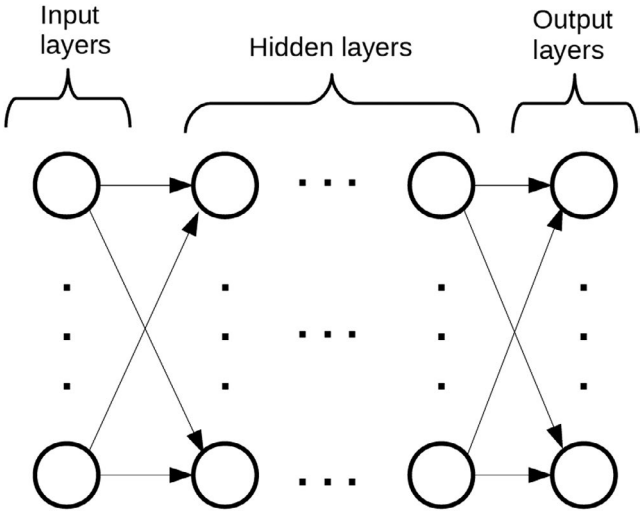

FIGURE 6 Example of an artificial neural network

where t_i is the target feature, $\boldsymbol{d}$ the descriptive features, and n the size of the training data set.

Often machine learning problems have more than one descriptive feature. The multivariable linear regression model ($\boldsymbol{M}$) is defined as:

$$M(\boldsymbol{d}) = \boldsymbol{w}[0] + \sum_{j=1}^{m} \boldsymbol{w}[j] \times d[j], \tag{15}$$

where $\boldsymbol{d}$ is a vector of m descriptive features, and $\boldsymbol{w}$ is a vector of $m - 1$ model parameters.

The previous equation can be simplified by integrating $d[0] = 1$:

$$M(\boldsymbol{d}) = \boldsymbol{w} \cdot \boldsymbol{d}. \tag{16}$$

To find the best value of $\boldsymbol{w}$, it is possible to use the gradient descent method. It consists, firstly, in choosing a random value of $\boldsymbol{w}$ and calculating L_2. Then, the value of $\boldsymbol{w}$ is adjusted till getting the global minimum on the error surface (L_2 according to $\boldsymbol{w}$).

It is also possible to predict categorical target with a linear regression. The modelled line, obtained by the regression, is used to separate the points with different target values of a feature space.

Figure 7 shows two examples of linear regression. The first one is for a continuous target prediction, and the second one is for a categorical target prediction.

The main advantage of linear regression is that it is the best algorithm when the relation between independent and dependent variables is linear. However, often these relations are not really clear and only linear.

2.4.2 | Logistic regression

It is also possible to approximate a training data set by a non-linear function. The most frequently used function is the logistic regression defined by:

$$M(d) = \frac{1}{1 + e^{-\boldsymbol{w} \cdot \boldsymbol{d}}} \tag{17}$$

To find the best parameter $\boldsymbol{w}$, the gradient descent method, as defined previously, can be used.

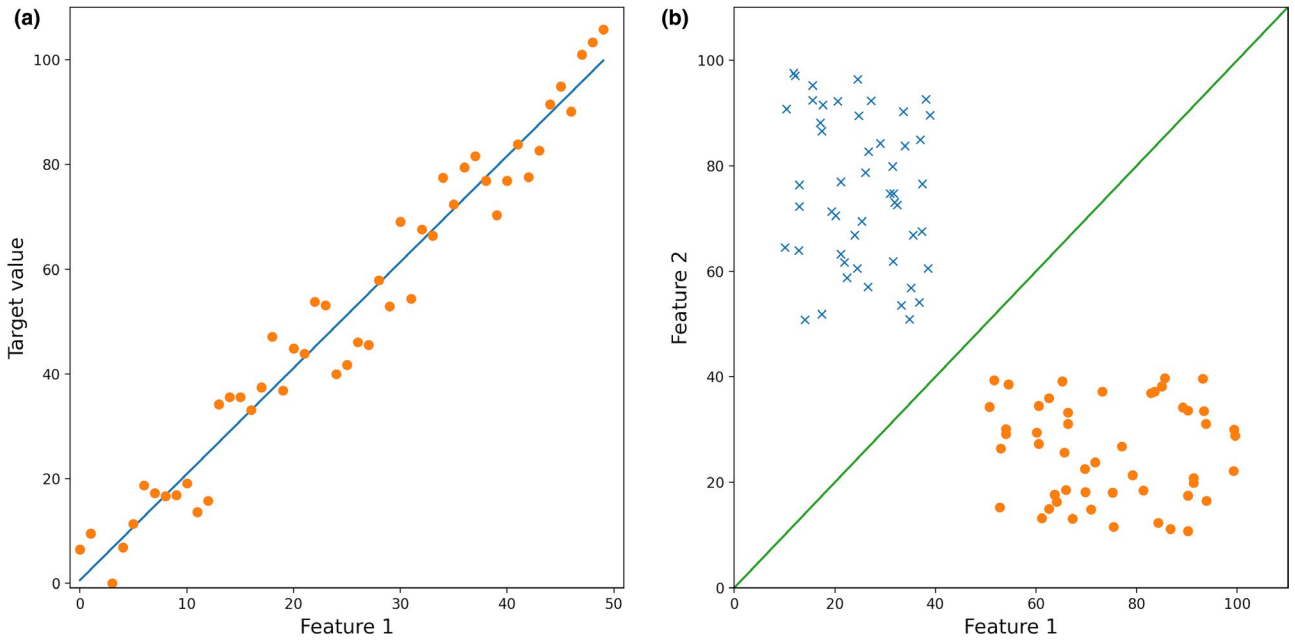

FIGURE 7 Two examples of linear regression. (a) for a continuous target prediction and (b) for a categorical target prediction

To predict categorical target with more than two levels, it is also possible to use logistic regression. It is necessary to create a model with the logistic regression for each target level.

The main advantages of logistic regression algorithm are that it is very easy to implement and very efficient to train. However, it is vulnerable to overfitting.

2.4.3 | Support vector machine

The support vector machine (SVM) is an approach that combines a regression with a decision boundary or also a separating hyperplan. The aim of this method is to find the maximum decision boundary that leads to the best separation of the data set. A support vector is a set of features used to define the decision boundary. There is one support vector for each level of the target feature.

For a linear separation, the separating hyperplan is defined as:

$$w_0 + \boldsymbol{w} \cdot \boldsymbol{d} = 0. \tag{18}$$

A − 1 level is affected for negative target ($w_0 + \boldsymbol{w}\,\boldsymbol{d} < 0$), and a + 1 level is affected for positive target ($w_0 + \boldsymbol{w}\cdot\boldsymbol{d} > 0$). The space between −1 and + 1 is the margin.

The SVM model for a linear separation is calculated with:

$$M_{\alpha,w_0}(q) = \sum_{i=1}^{s} (t_i \times \alpha_i \times (d_i \cdot \boldsymbol{q}) + w_0), \tag{19}$$

where $\boldsymbol{q}$ is the descriptive features, d_i are s support vectors (composed of descriptive and target features), w_0 is the first weight of the decision boundary, and $\boldsymbol{\alpha}$ is the Lagrange multipliers.

The optimization of the decision boundary is solved by a constrained quadratic optimization problem.

For data sets that are not linearly separable, it is also possible to use basis functions with logistic regression models with SVM. The model, in this case, is calculated as:

$$M_{\alpha,\phi,w_0}(\boldsymbol{q}) = \sum_{i=1}^{s} (t_i \times \alpha[i] \times (\phi(d_i) \cdot \phi(q)) + w_0), \tag{20}$$

where ϕ is a set of basis functions applied to the descriptive features.

To generalize the SVM model, a kernel function has to be used:

- Linear: kernel $(\boldsymbol{d}, \boldsymbol{q}) = \boldsymbol{d}\cdot\boldsymbol{q} + c$.
- Polynomial: kernel $(\boldsymbol{d}, \boldsymbol{q}) = (\boldsymbol{d}\cdot\boldsymbol{q} + 1)^p$.
- Gaussian radial basis: kernel $(\boldsymbol{d}, \boldsymbol{q}) = \exp(-\gamma||\boldsymbol{d} - \boldsymbol{q}||^2)$.

The model is defined as:

$$M_{\alpha,\text{kernel},w_0}(\boldsymbol{q}) = \sum_{i=1}^{s} (t_i \times \alpha[i] \times \text{kernal}\ (d_i, \boldsymbol{q}) + w_0). \tag{21}$$

The main advantages of SVM algorithm are that it is easy to use when no information about the data is given, and it limits the overfitting risks. However, it can be very difficult to choose the best kernel function.

3 | REVIEW ON MACHINE LEARNING APPLICATIONS FOR PREDICTIVE DATA ANALYTICS IN CARDIOVASCULAR MEDICINE

Machine learning in cardiovascular medicine is widespread used to understand particular phenotypes and to adapt some treatments (Mathur et al., 2020). The main challenge, outlined by Marthur *et al*., for machine learning in cardiovascular medicine is the reproducibility of the models obtained, which is strongly influenced by the experimental conditions, the interpretation of the results and the impact of data bias (Mathur et al., 2020). In this part, we detail the thirteen methods found in the cardiovascular field with the bibliographic search.

Medical data give information about: the patient, a treatment, a context, … They are often extracted from the HIS.

Gupta et al. (2017) applied NB, DT, RF, SVM and ME on a data set of 303 data to predict heart diseases. The authors showed that ME is more efficient than the other four methods (NB: 86%, DT: 78%, RF: 81%, SVM: 83% and ME: 88%).

Sankar and Sridhar (2017) proposed to use naive Bayes on 40 genes data to predict diabetes. This approach is very innovative but results are highly dependent on the type of diabetes.

Lo et al. (2016) used NB, ME, RF, DT, SVM, ANN, KNN and a personalized voting algorithm on a data set of 920 patients to predict the coronary artery disease. For each method, the accuracy is more than 75%, the sensitivity is more than 77%, and the specificity is more than 65%. The personalized voting algorithm gives the best results (accuracy: 79.0%, sensitivity: 81.7% and specificity: 75.6%).

Bashir et al. (2015) created a model ensemble using BagMOOV algorithm on a data set of 138 patients to predict and analyse heart disease. This method gives better results than individual classifiers (NB, LR and SVM) but it is worse in computation time than individual classifiers (ME: 84.78%, NB: 71.74%, LR: 78.26% and SVM: 81.16%).

Letham et al. (2015) compared an approach based on a probability model, called Bayesian rule lists (BRL), with individual classifiers (DT, LR, SVM and RF) on a data set of 1733 records to predict stroke. BRL gives better results than the other four classifiers (BRL: 77.15%, DT: 62.3%, LR: 74.7%, SVM: 74.6% and RF: 76.85%).

Kadi and Idri (2015) used a decision tree algorithm (C4.5) on a data set of 178 records to diagnose and predict cardiovascular dysautonomias. The accuracy of the algorithm was 98.54% for training set and 97.76% for testing set but it was tested on a small data set.

Xu et al. (2017) used RF, NB, ID3 and ME algorithms on a data set of 403 testers to predict diabetes risks. The ID3 algorithm is easy to implement but the data imbalance perturbs the model, and the ME

algorithm is more efficient (RF: 73.0%, NB: 73.9%, ID3: 67.1%, ME: 74.4%).

Verde et al. (2016) used Bayesian model on a data set of 339 patients, extracted from 6 studies, to predict effects of diabetes treatment. This study showed that the assessment of treatment effect is possible. However, predictions validity is highly dependent on the observational data quality.

Krikunov et al. (2016) used SVM on a data set of 3,980 episodes to predict a treatment status of acute coronary syndrome episodes. The accuracy for patient's outcome prediction was 65%; for patient's reanimation time, it was 89%, and for patient length of stay, it was 56%. Therefore, this model is efficient but the accuracy can be improved (Sobradillo et al., 2011).

Pradeep and Naveen (2016) used a decision tree algorithm (J48) on a current blood glucose level data set to predict the diabetes status. This method is easy to implement and has a good precision (70%).

Tylman et al. (2016) created a Bayesian network from ECG data to predict cardiovascular disease in admission wards and accident and emergency departments. This model is based on medical knowledge of physicians but it needs to be tested in various medical environments.

Chetty et al. (2015) applied a fuzzy KNN algorithm on two online data sets to predict, on patients, diabetes and liver disorder. The fuzzy KNN algorithm allows better accuracy (99.1%) than classical KNN (96.6%) on these data sets.

Maslennikov et al. (2015) implemented a Bayesian network on a data set of 1,193 patients to predict the risk factor of coronary artery disease. This approach is efficient (80%) but it should be improved to have better accuracy (Gruber & Ben-Gal, 2012).

4 | REVIEW ON MACHINE LEARNING APPLICATIONS FOR PREDICTIVE DATA ANALYTICS IN NUCLEAR MEDICINE

Machine learning is emergent in nuclear medicine and offers great potentials and challenges (Uribe et al., 2019). It offers the opportunity to save time for clinicians by performing repetitive tasks, generating preliminary reports and detecting what is not visible to the human eye (Aktolun, 2019). The main challenges of machine learning in nuclear medicine, outlined by Aktolun, are the insufficient amount of data to train generated model and the labelling of the data set by experts that takes a lot of time (Aktolun, 2019). The most frequently encountered data are radiomics that include high-throughput computation of quantitative medical images features (Ibrahim et al., 2019). To present examples on image-based data, we include herein applications where data correspond to features extracted from image processing procedures. For concision purposes, we focus only on features extracted from nuclear medicine images. In this part, we detail the eleven methods found in the nuclear medicine field.

Ingrisch et al. (2018) used a combination of decision trees, called random survival forest, on a data set of various medical data (demographic, tumour localization, baseline liver function, …) recorded on 366 hepatic tumour patients to predict their response to radio-embolization. The authors obtained a better predictive value for baseline cholinesterase and bilirubin than previous tested model (Cox proportional hazards model).

Papp et al. (2018) created a model based on geometric probability on a data set of features (voxel properties), extracted from positron emission tomography (PET) images and recorded on 70 patients with glioma to predict their survival. The authors showed that ex vivo features are important to increase the sensitivity and the specificity of the 16 tested models. They obtained a sensitivity greater than 86% and a specificity greater than 92%.

Jeong et al. (2019) used a model based on principal component analysis, SVM, RF and gradient boosting on a data set of textural features, extracted from F-fluorodeoxyglucose (F-FDG) PET images, recorded on 70 patients with high-grade osteosarcoma to predict the response to neoadjuvant chemotherapy (NAC). The authors showed that the proposed model can illustrate changing between before and after NAC with a sensitivity greater than 81% and a specificity greater than 71%.

Li et al. (2019) developed a model based on RF on a data set of features (intensity, texture, spatial, …), extracted from F-FDG PET images and recorded on 41 patients with suspected relapsed acute leukaemia to predict bone marrow involvement (BMI). The authors showed that their model is more accurate (83%) than visual analysis (59%) to predict BMI.

Rastegar et al. (2020) used a model based on RF and KNN on a data set of features (grey level, intensity, histogram, …), extracted from dual-energy x-ray absorptiometry images and recorded on 147 patients to predict osteoporosis and osteopenia. The authors evaluated the performance of their models with the area under the receiver operator characteristic (ROC) curve (AUC). They showed that the best efficient tested model has an AUC equal to 0.78 when comparing healthy subjects with subjects with osteoporosis.

Ahn et al. (2019) compared five machine learning methods (RF, ANN, NB, LR and SVM) on a data set of features (textural and shape features), extracted from PET computed tomography (CT) images and recorded on 93 patients with non-small-cell lung cancer to predict the prognostic value of PET. Authors show that RF obtains the better results (AUC = 0.96, accuracy = 90% and precision = 91%) than the other four tested methods.

Shayesteh et al. (2019) compared four machine learning methods (SVM, ANN, BN and KNN) and model based on different combinations of the four machine learning methods on a data set of features (shape, intensity, texture, …), extracted from magnetic resonance imaging (MRI) and recorded on 98 patients with rectal cancer to predict their chemotherapy response. The authors showed that the model that combines the four methods has the highest performance (AUC greater than 0.9 and accuracy greater than 82%).

Qian et al. (2019) compared six machine learning methods (Adaboost, KNN, DT, NB, RF and SVM) on a data set of features (shape, texture, wavelet, …), extracted from MRI and recorded on 142 patients with brain metastases to classify, preoperatively,

glioblastoma. The authors showed that SVM gives better results (AUC greater than 0.9) than the other tested methods.

Blum et al. (2018) used SVM on a data set of diuresis renogram recorded on 55 patients to predict uteropelvic junction obstruction. This method has a good accuracy (93%), a good sensitivity (91%) and a good specificity (96%).

Ellmann et al. (2019) used an average-model neural network on a data set of features (dynamic contrast, enhancement sequences, ...), extracted from PET CT and recorded on 28 rats with breast cancer to predict the early metastatic disease. This method has a good accuracy (86%), a good sensitivity (83%) and a good specificity (90%). However, its adaptation to human model must be tested.

Wottschel et al. (2019) used SVM on a data set of features (volume, probability, ...), extracted from brain MRI and various medical data (demographic and clinical measures) and recorded on 400 patients (in several locations) with clinical isolated syndrome suggestive of multiple sclerosis to predict a second clinical episode. The method accuracy is influenced by the part of data set considered (multi or single centre). It was always between 65% and 71%.

5 | CONCLUSION

We reviewed machine learning for predictive data analyses through four approaches: information-based, similarity-based, probability-based and error-based approaches. Fuzzy set theory, detailed in Section 2.2.2, can be applied to other machine learning methods. This theory could be a solution where the sample sets overlap.

In our bibliographic search, the two most commonly used methods are ME and SVM. SVM, ME and ANN frequently lead to better results than the other methods. The cardiovascular domain (with diabetes) is the one that leads to the largest number of references in our bibliographic search. The most commonly used data in machine learning for predictive data analytics applications are detailed in sections 3 and 4; they correspond to data frequently extracted from HIS for cardiovascular medicine and radiomics for nuclear medicine.

The difficulty in machine learning problems is the data. First, the acquisition has to be standardized for the study. Second, the processing has to be adapted to the data and the problem. A lot of machine learning solutions start with a preprocessing step to improve the algorithm accuracy.

Machine learning applications are growing in the medical field. In the coming years, more studies, more data, more tools and more methods will, for sure, be proposed.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

FUNDING INFORMATION

A CIFRE grant No 2017/1165 was awarded by ANRT (France) to the company COTTOS Médical to support the work of graduate student A. Jamin.

ORCID

Antoine Jamin https://orcid.org/0000-0003-0473-3669
Pierre Abraham https://orcid.org/0000-0001-7053-2801
Anne Humeau-Heurtier https://orcid.org/0000-0002-6289-0040

REFERENCES

Adegoke, V. F., Chen, D., Banissi, E., & Barikzai, S. (2017). Prediction of breast cancer survivability using ensemble algorithms. In *2017 International Conference on Smart Systems and Technologies (SST)* (pp. 223–231). https://doi.org/10.1109/SST.2017.8188699

Ahn, H. K., Lee, H., Kim, S. G., & Hyun, S. H. (2019). Pre-treatment 18F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer. *Clinical Radiology*, *74*(6), 467–473. https://doi.org/10.1016/j.crad.2019.02.008

Aktolun, C. (2019). Artificial intelligence and radiomics in nuclear medicine: Potentials and challenges. *European Journal of Nuclear Medicine and Molecular Imaging*, *46*(13), 2731–2736. https://doi.org/10.1007/s00259-019-04593-0

Alghamdi, W., Stamate, D., Vang, K., Stahl, D., Colizzi, M., Tripoli, G., Quattrone, D., Ajnakina, O., Murray, R. M., & Forti, M. D. (2016). A prediction modelling and pattern detection approach for the first-episode psychosis associated to Cannabis use. In *2016 15th IEEE International Conference on Machine Learning and Appli-cations (ICMLA)* (pp. 825–830). https://doi.org/10.1109/ICMLA.2016.0148

Ammad-Ud-Din, M., Khan, S. A., Wennerberg, K., & Aittokallio, T. (2017). Systematic identification of feature combinations for predicting drug response with Bayesian multi-view multi-task linear regression. *Bioinformatics (Oxford, England)*, *33*, i359–i368. https://doi.org/10.1093/bioinformatics/btx266

Armero, C., Forné, C., Rué, M., Forte, A., Perpiñán, H., Gómez, G., & Baré, M. (2016). Bayesian joint ordinal and survival modeling for breast cancer risk assessment. *Statistics in Medicine*, *35*, 5267–5282. https://doi.org/10.1002/sim.7065

Bakal, G., Talari, P., Kakani, E. V., & Kavuluru, R. (2018). Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative relations. *Journal of Biomedical Informatics*, *82*, 189–199. https://doi.org/10.1016/j.jbi.2018.05.003

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Bashir, S., Qamar, U., & Hassan Khan, F. (2015). BagMOOV: A novel ensemble for heart disease prediction bootstrap aggregation with multi-objective optimized voting. *Australasian Physical & Engineering Sciences in Medicine*, *38*(2), 305–323. https://doi.org/10.1007/s13246-015-0337-6

Bendtsen, C., Degasperi, A., Ahlberg, E., & Carlsson, L. (2017). Improving machine learning in early drug discovery. *Annals of Mathematics and Artificial Intelligence*, *81*, 155–166. https://doi.org/10.1007/s10472-017-9541-2

Billings, J. M., Eder, M., Flood, W. C., Singh Dhami, D., Natarajan, S., & Whitlow, C. T. (2017). Machine learning applications to resting-state functional MR imaging analysis. *Functional Connectivity*, *27*, 609–620. https://doi.org/10.1016/j.nic.2017.06.010

Blum, E. S., Porras, A. R., Biggs, E., Tabrizi, P. R., Sussman, R. D., Sprague, B. M., Shalaby-Rana, E., Majd, M., Pohl, H. G., & Linguraru, M. G. (2018). Early detection of ureteropelvic junction obstruction using signal analysis and machine learning: A dynamic solution to a dynamic problem. *Journal of Urology*, *199*(3), 847–852. https://doi.org/10.1016/j.juro.2017.09.147

Borisov, N., Tkachev, V., Muchnik, I., & Buzdin, A. (2017). Individual drug treatment prediction in oncology based on machine learning using cell culture gene expression data. In *Proceedings of the 2017*

International Conference on Computational Biology and Bioinformatics (pp. 1–6). https://doi.org/10.1145/3155077.3155078

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classiffication and Regression Trees*, 1st Edition. Taylor and Francis. https://doi.org/10.1201/9781315139470

Bzdok, D., & Meyer-Lindenberg, A. (2017). Machine learning for precision psychiatry: Opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*, 223–230. https://doi.org/10.1016/j.bpsc.2017.11.007

Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: Literature review and challenges, data mining for the internet of things: Literature review and challenges. *International Journal of Distributed Sensor Networks*, *11*, 431047. https://doi.org/10.1155/2015/431047

Cheng, J., Tegge, A. N., & Baldi, P. (2008). Machine learning methods for protein structure prediction. *IEEE Reviews in Biomedical Engineering*, *1*, 41–49. https://doi.org/10.1109/RBME.2008.2008239

Chetty, N., Vaisla, K. S., & Patil, N. (2015). An improved method for disease prediction using fuzzy approach. In *2015 Second International Conference on Advances in Computing and Communication Engineering* (pp. 568–572). https://doi.org/10.1109/ICACCE.2015.67

Coley, R. Y., Fisher, A. J., Mamawala, M., Carter, H. B., Pienta, K. J., & Zeger, S. L. (2017). A Bayesian hierarchical model for prediction of latent health states from multiple data sources with application to active surveillance of prostate cancer. *Biometrics*, *73*, 625–634. https://doi.org/10.1111/biom.12577

Collij, L. E., Heeman, F., Kuijer, J. P. A., Ossenkoppele, R., Benedictus, M. R., Möller, C., Verfaillie, S. C. J., Sanz-Arigita, E. J., van Berckel, B. N. M., van der Flier, W. M., Scheltens, P., Barkhof, F., & Wink, A. M. (2016). Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease. *Radiology*, *281*(3), 865–875. https://doi.org/10.1148/radiol.2016152703

Delibašić, B., Radovanović, S., Jovanović, M., Bohanec, M., & Suknović, M. (2018). Integrating knowledge from DEX hierarchies into a logistic regression stacking model for predicting ski injuries. *Journal of Decision Systems*, *27*(sup1), 201–208 https://doi.org/10.1080/12460125.2018.1460164

Deo, R. C. (2015). Machine learning in medicine. *Circulation*, *132*(20), 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593

Deschamps, T., Le Goff, C. G., Berrut, G., Cornu, C., & Mignardot, J. B. (2016). A decision model to predict the risk of the first fall onset. *Experimental Gerontology*, *81*, 51–55. https://doi.org/10.1016/j.exger.2016.04.016

Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare Informatics Research*, *22*(3), 156–163. https://doi.org/10.4258/hir.2016.22.3.156

Ellmann, S., Seyler, L., Evers, J., Heinen, H., Bozec, A., Prante, O., Kuwert, T., Uder, M., & Bäuerle, T. (2019). Prediction of early metastatic disease in experimental breast cancer bone metastasis by combining PET/CT and MRI parameters to a Model-Averaged Neural Network. *Bone*, *120*, 254–261. https://doi.org/10.1016/j.bone.2018.11.008

Ertuğrul, Ö. F., Kaya, Y., Tekin, R., & Almalı, M. N. (2016). Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait. *Expert Systems with Applications*, *56*, 156–163. https://doi.org/10.1016/j.eswa.2016.03.018

Fabris, F., Freitas, A. A., & Tullet, J. M. A. (2016). An extensive empirical comparison of probabilistic hierarchical classifiers in datasets of ageing-related genes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *13*, 1045–1058. https://doi.org/10.1109/TCBB.2015.2505288

Fabris, F., Magalhães, J. P. D., & Freitas, A. A. (2017). A review of supervised machine learning applied to ageing research. *Biogerontology*, *18*(2), 171–188. https://doi.org/10.1007/s10522-017-9683-y

Gruber, A., & Ben-Gal, I. (2012). Efficient Bayesian network learning for system optimization in reliability engineering. *Quality Technology & Quantitative Management*, *9*, 97–114. https://doi.org/10.1080/16843703.2012.11673280

Gupta, N., Ahuja, N., Malhotra, S., Bala, A., & Kaur, G. (2017). Intelligent heart disease prediction in cloud environment through ensembling. *Expert Systems*, *34*, e12207. https://doi.org/10.1111/exsy.12207

Kop R., Hoogendoorn M., Teije A., Büchner F. L., Slottje P., Moons L. M. G., Numans M. E. (2016). Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. *Computers in Biology and Medicine*, *76*, 30–38. https://doi.org/10.1016/j.compbiomed.2016.06.019

Hu, Y.-H., Tai, C.-T., Chen, S.-C.-C., Lee, H.-W., & Sung, S.-F. (2017). Predicting return visits to the emergency department for pediatric patients: Applying supervised learning techniques to the Taiwan National Health Insurance Research Database. *Computer Methods and Programs in Biomedicine*, *144*, 105–112. https://doi.org/10.1016/j.cmpb.2017.03.022

Huang, X., Sun, Y., Trow, P., Chatterjee, S., Chakravartty, A., Tian, L., & Devanarayan, V. (2017). Patient subgroup identification for clinical drug development. *Statistics in Medicine*, *36*, 1414–1428. https://doi.org/10.1002/sim.7236

Ibrahim, A., Vallières, M., Woodruff, H., Primakov, S., Beheshti, M., Keek, S., Refaee, T., Sanduleanu, S., Walsh, S., Morin, O., Lambin, P., Hustinx, R., & Mottaghy, F. M. (2019). Radiomics analysis for clinical decision support in nuclear medicine. *Seminars in Nuclear Medicine*, *49*(5), 438–449. https://doi.org/10.1053/j.semnuclmed.2019.06.005

Ingrisch, M., Schöppe, F., Paprottka, K., Fabritius, M., Strobl, F. F., De Toni, E. N., Ilhan, H., Todica, A., Michl, M., & Paprottka, P. M. (2018). Prediction of 90 Y radioembolization outcome from pretherapeutic factors with random survival forests. *Journal of Nuclear Medicine*, *59*, 769–773. https://doi.org/10.2967/jnumed.117.200758

Islam, S. M. R., Kwak, D., Kabir, M. H., Hossain, M., & Kwak, K. S. (2015). The internet of things for health care: A comprehensive survey. *IEEE Access*, *3*, 678–708.. https://doi.org/10.1109/ACCESS.2015.2437951

Ivascu, T., Cincar, K., Dinis, A., & Negru, V. (2017). Activities of daily living and falls recognition and classification from the wearable sensors data. In *2017 E-Health and Bioengineering Conference (EHB)* (pp. 627–630). https://doi.org/10.1109/EHB.2017.7995502

Jarmulski, W., Wieczorkowska, A., Trzaska, M., Ciszek, M., & Paczek, L. (2018). Machine learning models for predicting patients survival after liver transplantation. *Computer Science*, *19*, 223. https://doi.org/10.7494/csci.2018.19.2.2746

Jeon, H., Lee, W., Park, H., Lee, H. J., Kim, S. K., Kim, H. B., Jeon, B., & Suk Park, K. (2017). High-accuracy automatic classification of Parkinsonian tremor severity using machine learning method. *Physiological Measurement*, *38*, 1980. https://doi.org/10.1088/1361-6579/aa8e1f

Jeong, S. Y., Kim, W., Byun, B. H., Kong, C.-B., Song, W. S., Lim, I., Lim, S. M., & Woo, S.-K. (2019). Prediction of chemotherapy response of osteosarcoma using baseline 18f-FDG textural features machine learning approaches with PCA. *Contrast Media & Molecular Imaging*, *2019*, 3515080. https://doi.org/10.1155/2019/3515080

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of Big Data research. *Big Data Research*, *2*, 59–64. https://doi.org/10.1016/j.bdr.2015.01.006

Jing, B., Meng, H. B., Yang, S. C., Zhao, D. S., & Shang, X. Y. (2016). The diagnostic model of neonatal respiratory distress syndrome based on intelligent algorithm. In *2016 8th International Conference on Information Technology in Medicine and Education (ITME)* (pp. 319–323). https://doi.org/10.1109/ITME.2016.0078

Kabeshova, A., Launay, C. P., Gromov, V. A., Fantino, B., Levinoff, E. J., Allali, G., & Beauchet, O. (2016). Falling in the elderly: Do statistical models matter for performance criteria of fall prediction? Results from two large population-based studies. *European Journal of Internal Medicine*, *27*, 48–56. https://doi.org/10.1016/j.ejim.2015.11.019

Kadi, I., & Idri, A. (2015). A decision tree-based approach for cardiovascular dysautonomias diagnosis: A case study. In *2015 IEEE Symposium Series on Computational Intelligence* (pp. 816–823). https://doi.org/10.1109/SSCI.2015.121

Kelleher, J. D., Mac Namee, B., & D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: Algorithms, worked examples, and case studies*, 1st Edition, . MIT Press.

Keller, J. M., Gray, M. R., & Givens, J. A. (1985). A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-15*(4), 580–585. https://doi.org/10.1109/TSMC.1985.6313426

Khalifa, M., & Alswailem, O. (2015). Hospital information systems (HIS) acceptance and satisfaction: A case study of a tertiary care hospital. *Procedia Computer Science*, *63*, 198–204. https://doi.org/10.1016/j.procs.2015.08.334

Khan, T., Westin, J., & Dougherty, M. (2014). Classification of speech intelligibility in Parkinson's disease. *Biocybernetics and Biomedical Engineering*, *34*, 35–45. https://doi.org/10.1016/j.bbe.2013.10.003

Khanna, D., & Sharma, A. (2018). Kernel-Based Naive Bayes classifier for medical predictions. In *Intelligent Engineering Informatics* (pp. 91–101). Springer. https://doi.org/10.1007/978-981-10-7566-7_10

Kim, H., Ji Lee, H., Lee, W., Kwon, S., Kyong Kim, S., Seon Jeon, H., Park, H., Won Shin, C., Jin Yi, W., Jeon, B. S., & Park, K. S. (2015). Unconstrained detection of freezing of Gait in Parkinson's disease patients using smartphone. *Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2015. pp. 3751–3754. https://doi.org/10.1109/EMBC.2015.7319209

Kim, K.-J., & Cho, S.-B. (2015). Meta-classifiers for high-dimensional, small sample classiffication for gene expression analysis. *Pattern Analysis and Applications*, *18*, 553–569. https://doi.org/10.1007/s10044-014-0369-7

Kramer, D., Veeranki, S., Hayn, D., Quehenberger, F., Leodolter, W., Jagsch, C., & Schreier, G. (2017). Development and validation of a multivariable prediction model for the occurrence of delirium in hospitalized gerontopsychiatry and internal medicine patients. *Studies in Health Technology and Informatics*, *236*, 32–39. https://doi.org/10.3233/978-1-61499-759-7-32

Krikunov, A. V., Bolgova, E. V., Krotov, E., Abuhay, T. M., Yakovlev, A. N., & Kovalchuk, S. V. (2016). Complex data-driven predictive modeling in personalized clinical decision support for acute coronary syndrome episodes. *Procedia Computer Science*, *80*, 518–529. https://doi.org/10.1016/j.procs.2016.05.332

Letham, B., Rudin, C., McCormick, T. H., & Madigan, D. (2015). Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model. *The Annals of Applied Statistics*, *9*, 1350–1371. https://doi.org/10.1214/15-AOAS848

Leung, M. K. K., Delong, A., Alipanahi, B., & Frey, B. J. (2016). Machine learning in genomic medicine: A review of computational problems and data sets. *Proceedings of the IEEE*, *104*, 176–197. https://doi.org/10.1109/JPROC.2015.2494198

Li, H., Xu, C., Xin, B., Zheng, C., Zhao, Y., Hao, K., Wang, Q., Wahl, R. L., Wang, X., & Zhou, Y. (2019). 18f-FDG PET/CT radiomic analysis with machine learning for identifying bone marrow involvement in the patients with suspected relapsed acute leukemia. *Theranostics*, *9*(16), 4730–4739. https://doi.org/10.7150/thno.33841

Lo, Y.-T., Fujita, H., & Pai, T.-W. (2016). Prediction of coronary artery disease based on ensemble learning approaches and co-expressed observations. *Journal of Mechanics in Medicine and Biology*, *16*, 1640010. https://doi.org/10.1142/S0219519416400108

Luo, G., Lu, F., Jiang, L., Cai, Y., & Zhang, Y. (2015). Virtual screening of cytochrome P450 2a6 inhibitors from traditional Chinese medicine using support vector machine and molecular docking. In *2015 8th International Conference on Biomedical Engineering and Informatics (BMEI)* (pp. 505–509). https://doi.org/10.1109/BMEI.2015.7401557

Maslennikov, E. D., Sulimov, A. V., Savkin, I. A., Evdokimova, M. A., Zateyshchikov, D. A., Nosikov, V. V., & Sulimov, V. B. (2015). An intuitive risk factors search algorithm: Usage of the Bayesian network technique in personalized medicine. *Journal of Applied Statistics*, *42*, 71–87. https://doi.org/10.1080/02664763.2014.934664

Mathur, P., Srivastava, S., Xu, X., & Mehta, J. (2020). Artificial intelligence, machine learning, and cardiovascular disease. *Clinical Medicine Insights. Cardiology*, *14*, 1179546820927404. https://doi.org/10.1177/1179546820927404

Mease, D., & Wyner, A. (2008). Evidence contrary to the statistical view of boosting. *Journal of Machine Learning Research*, *9*, 131–156.

Mirzaei, G., Adeli, A., & Adeli, H. (2016). Imaging and machine learning techniques for diagnosis of Alzheimer's disease. *Reviews in the Neurosciences*, *27*(8), 857–870. https://doi.org/10.1515/revneuro-2016-0029

Mitchell, T. M. (1997). *Machine Learning*, 1st ed. McGraw-Hill Inc.

Modu, B., Polovina, N., Lan, Y., Konur, S., Asyhari, A., & Peng, Y. (2017). Towards a predictive analytics-based intelligent malaria outbreak warning system. *Applied Sciences*, *7*(8), 836. https://doi.org/10.3390/app7080836

Montoye, A. H. K., Begum, M., Henning, Z., , & Pfeiffer, K. A. (2017). Comparison of linear and non-linear models for predicting energy expenditure from raw accelerometer data. *Physiological Measurement*, *38*(2), 343–357. https://doi.org/10.1088/1361-6579/38/2/343

Nagarajan, R., & Upreti, M. (2017). An ensemble predictive modeling framework for breast cancer classification. Systems approaches for identifying disease genes and drug. *Methods*, *131*, 128–134. https://doi.org/10.1016/j.ymeth.2017.07.011

Nikoo, M. R., Kerachian, R., & Alizadeh, M. R. (2018). A fuzzy KNN-based model for significant wave height prediction in large lakes. *Oceanologia*, *60*,153–168. https://doi.org/10.1016/j.oceano.2017.09.003

Oztaner, S. M., Taskaya Temizel, T., Erdem, S. R., & Ozer, M. (2015). A Bayesian estimation framework for pharmacogenomics driven warfarin dosing: A comparative study. *IEEE Journal of Biomedical and Health Informatics*, *19*(5), 1724–1733. https://doi.org/10.1109/JBHI.2014.2336974

Oztekin, A., Al-Ebbini, L., Sevkli, Z., & Delen, D. (2018). A decision analytic approach to predicting quality of life for lung transplant recipients: A hybrid genetic algorithms-based methodology. *European Journal of Operational Research*, *266*(2), 639–651. https://doi.org/10.1016/j.ejor.2017.09.034

Papp, L., Pötsch, N., Grahovac, M., Schmidbauer, V., Woehrer, A., Preusser, M., Mitterhauser, M., Kiesel, B., Wadsak, W., Beyer, T., Hacker, M., & Traub-Weidinger, T. (2018). Glioma survival prediction with combined analysis of in vivo 11c-MET PET features, ex vivo features, and patient features by supervised machine learning. *Journal of Nuclear Medicine*, *59*(6), 892–899. https://doi.org/10.2967/jnumed.117.202267

Pavel, M., Jimison, H. B., Wactlar, H. D., Hayes, T. L., Barkis, W., Skapik, J., & Kaye, J. (2013). The role of technology and engineering models in transforming healthcare. *IEEE Reviews in Biomedical Engineering*, *6*, 156–177. https://doi.org/10.1109/RBME.2012.2222636

Portugal, I., Alencar, P., & Cowan, D. (2017). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, *97*, 205–227. https://doi.org/10.1016/j.eswa.2017.12.020

Pradeep, K. R., & Naveen, N. C. (2016). Predictive analysis of diabetes using J48 algorithm of classification techniques. In *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)* (pp. 347–352). https://doi.org/10.1109/IC3I.2016.7917987

Qian, Z., Li, Y., Wang, Y., Li, L., Li, R., Wang, K., Li, S., Tang, K., Zhang, C., Fan, X., Chen, B., & Li, W. (2019). Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classiffers. *Cancer Letters*, *451*, 128–135. https://doi.org/10.1016/j.canlet.2019.02.054

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/BF00116251

Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc.

Quinlan, J. R. (1996) Boosting, and C4.S. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence* Vol. 1, (pp. 725–730).

Rastegar, S., Vaziri, M., Qasempour, Y., Akhash, M. R., Abdalvand, N., Shiri, I., Abdollahi, H., & Zaidi, H. (2020). Radiomics for classification of bone mineral loss: A machine learning study. *Diagnostic and Interventional Imaging*, *101*(9), 599–610. https://doi.org/10.1016/j.diii.2020.01.008

Saleh, L., Mcheick, H., Ajami, H., Mili, H., & Dargham, J. (2017). Comparison of Machine Learning Algorithms to Increase Prediction Accuracy of COPD Domain. In Enhanced Quality of Life and Smart Living, 247–254.

Samant, P., & Agarwal, R. (2018). Machine learning techniques for medical diagnosis of diabetes using iris images. *Computer Methods and Programs in Biomedicine*, *157*, 121–128. https://doi.org/10.1016/j.cmpb.2018.01.004

Sankar, A., & Sridhar, R. (2017). Gene interactions and inuences identi_cation for diabetes from bio-medical literature. In *2017 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0576–0582). https://doi.org/10.1109/ICCSP.2017.8286424

Sanz, J., Paternain, D., Galar, M., Fernandez, J., Reyero, D., & Belzunegui, T. (2017). A new survival status prediction system for severe trauma patients based on a multiple classifier system. *Computer Methods and Programs in Biomedicine*, *142*, 1–8. https://doi.org/10.1016/j.cmpb.2017.02.011

Satija, U., Ramkumar, B., & Manikandan, M. S. (2018). A Review of Signal Processing Techniques for Electrocardiogram Signal Quality Assessment. *IEEE Reviews in Biomedical Engineering*, *11*, 36–52.. https://doi.org/10.1109/RBME.2018.2810957

Schapire, R. E. (1999). A brief introduction to boosting. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, Vol. 2, (pp. 1401–1406).

Schwartzi, M., Parkl, M., Phanl, J. H., & Wang, M. D. (2015). Integration of multimodal RNA-seq data for prediction of kidney cancer survival. *Proceedings IEEE International Conference on Bioinformatics and Biomedicine*, *2015*, 1591–1595. https://doi.org/10.1109/bibm.2015.7359913

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shayesteh, S. P., Alikhassi, A., Fard Esfahani, A., Miraie, M., Geramifar, P., Bitarafan-rajabi, A., & Haddad, P. (2019). Neo-adjuvant chemoradiotherapy response prediction using MRI based ensemble learning method in rectal cancer patients. *Physica Medica*, *62*, 111–119. https://doi.org/10.1016/j.ejmp.2019.03.013

Sobradillo, P., Pozo, F., & Agustí, Á. (2011). P4 medicine: The future around the corner. *Archivos De Bronconeumología ((English Edition))*, *47*(1), 35–40. https://doi.org/10.1016/S1579-2129(11)70006-4

Stojkoska, B. L. R., & Trivodaliev, K. V. (2017). A review of internet of things for smart home: Challenges and solutions. *Journal of Cleaner Production*, *140*, 1454–1464. https://doi.org/10.1016/j.jclepro.2016.10.006

Tripoliti, E. E., Tzallas, A. T., Tsipouras, M. G., Rigas, G., Bougia, P., Leontiou, M., Konitsiotis, S., Chondrogiorgi, M., Tsouli, S., & Fotiadis, D. I. (2013). Automatic detection of freezing of gait events in patients with Parkinson's disease. *Computer Methods and Programs in Biomedicine*, , *110*(1), 12–26. https://doi.org/10.1016/j.cmpb.2012.10.016

Tylman, W., Waszyrowski, T., Napieralski, A., Kamiński, M., Trafidło, T., Kulesza, Z., Kotas, R., Marciniak, P., Tomala, R., & Wenerski, M. (2016). Real-time prediction of acute cardiovascular events using hardware-implemented Bayesian networks. *Computers in Biology and Medicine*, *69*, 245–253. https://doi.org/10.1016/j.compbiomed.2015.08.015

Uribe, C. F., Mathotaarachchi, S., Gaudet, V., Smith, K. C., Rosa-Neto, P., Bénard, F., Black, S. E., & Zukotynski, K. (2019). Machine learning in nuclear medicine: Part 1—Introduction. *Journal of Nuclear Medicine*, *60*(4), 451–458. https://doi.org/10.2967/jnumed.118.223495

Verde, P. E., Ohmann, C., Morbach, S., & Icks, A. (2016). Bayesian evidence synthesis for exploring generalizability of treatment effects: A case study of combining randomized and non-randomized results in diabetes. *Statistics in Medicine*, *35*, 1654–1675. https://doi.org/10.1002/sim.6809

Wan, C., Freitas, A. A., & de Magalhães, J. P. (2015). Predicting the pro-longevity or anti-longevity effect of model organism genes with new hierarchical feature selection methods. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *12*, 262–275. https://doi.org/10.1109/tcbb.2014.2355218

Wang, X., Yang, B., Huang, J., Chen, H., Gu, X., Bai, Y., & Du, Z. (2016). IASM: A system for the intelligent active surveillance of malaria. *Computational and Mathematical Methods in Medicine*, *2016*, 11. https://doi.org/10.1155/2016/2080937

Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C. (2010). Machine learning in medical imaging. *IEEE Signal Processing Magazine*, *27*(4), 25–38.. https://doi.org/10.1109/MSP.2010.936730

Whitmore, A., Agarwal, A., & Da Xu, L. (2015). The Internet of Things | A survey of topics and trends. *Information Systems Frontiers*, *17*(2), 261–274. https://doi.org/10.1007/s10796-014-9489-2

Wottschel, V., Chard, D. T., Enzinger, C., Filippi, M., Frederiksen, J. L., Gasperini, C., Giorgio, A., Rocca, M. A., Rovira, A., De Stefano, N., Tintoré, M., Alexander, D. C., Barkhof, F., & Ciccarelli, O. (2019). SVM recursive feature elimination analyses of structural brain MRI predicts near-term relapses in patients with clinically isolated syndromes suggestive of multiple sclerosis. *NeuroImage: Clinical*, *24*, 102011. https://doi.org/10.1016/j.nicl.2019.102011

Xia, Y., Gao, Q., & Ye, Q. (2015). Classification of gait rhythm signals between patients with neuro-degenerative diseases and normal subjects: Experiments with statistical features and different classification models. *Biomedical Signal Processing and Control*, *18*, 254–262. https://doi.org/10.1016/j.bspc.2015.02.002

Xu, W., Zhang, J., Zhang, Q., & Wei, X. (2017). Risk prediction of type II diabetes based on random forest model. In *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (pp. 382–386). https://doi.org/10.1109/AEEICB.2017.7972337

Yin, Y., Zeng, Y., Chen, X., & Fan, Y. (2016). The internet of things in healthcare: An overview. *Journal of Industrial Information Integration*, *1*, 3–13. https://doi.org/10.1016/j.jii.2016.03.004

Yuvaraj, R., Murugappan, M., Mohamed Ibrahim, N., Sundaraj, K., Iqbal Omar, M., Mohamad, K., & Palaniappan, R. (2014). Detection of emotions in Parkinson's disease using higher order spectral features from brain's electrical activity. *Biomedical Signal Processing and Control*, *14*, 108–116. https://doi.org/10.1016/j.bspc.2014.07.005

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*, 1st ed. Chapman & Hall/CRC.

How to cite this article: Jamin A, Abraham P, Humeau-Heurtier A. Machine learning for predictive data analytics in medicine: A review illustrated by cardiovascular and nuclear medicine examples. *Clin Physiol Funct Imaging*. 2021;41:113–127. https://doi.org/10.1111/cpf.12686