IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH

Received November 11, 2019, accepted November 26, 2019, date of publication December 6, 2019,
date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958047

Bayesian Non-Parametric Classification With Tree-Based Feature Transformation for NIPPV Efficacy Prediction in COPD Patients

YANG WENG[1], YIN FANG[1], HAIYING YAN[2], YANG YANG[2], AND WENXING HONG[3]

[1]College of Mathematics, Sichuan University, Chengdu 610064, China
[2]Department of Respiration and Critical Care Medicine, Sichuan Provincial People's Hospital, Chengdu 610064, China
[3]School of Aerospace Engineering, Xiamen University, Xiamen 361005, China

Corresponding author: Wenxing Hong (hwx@xmu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0830300, in part by the Science and Technology Program of Fujian, China, under Grant 2018H0035, and in part by the Science and Technology Program of Xiamen, China, under Grant 3502Z20183011.

ABSTRACT Non-invasive positive pressure ventilation (NIPPV) is a life-saving approach which was developed to reduce the complications of endotracheal intubation and invasive ventilation in patients with chronic obstructive pulmonary disease (COPD). However, it has a certain probability of invalid. Failure of NIPPV will lead to an increase in mortality, which highlights the importance of rational diagnosis about the need for NIPPV therapy. In order to avoid delaying endotracheal intubation, we proposed a hybrid model which combine tree-based feature transformation with Bayesian non-parametric classification, to predict whether the patient should adopt NIPPV based on the their own physical condition. We delved into the feature importance and justified the rationality of using tree-based feature transformation. The proposed gaussian process classification (GPC) with gradient boosting decision tree (GBDT) feature transformation model has shown state-of-the-art results on both the NIPPV dataset and two simulated datasets with larger sample size. For critically ill COPD patients, the proposed method provides diagnostic assistance for physicians' decision making and avoids delaying endotracheal intubation or mechanical ventilation.

INDEX TERMS COPD, NIPPV, therapeutic efficacy prediction, GBDT, Gaussian process classification.

I. INTRODUCTION

Chronic obstructive pulmonary disease (COPD) [1] is a chronic inflammatory lung disease that leads to obstructed airflow from the lungs. This disease is caused by prolonged exposure to particulate matter or irritating gases, usually from cigarette smoke [2]. It is a leading cause of chronic morbidity and mortality worldwide. Many people are afflicted with COPD for years and die prematurely of it or its complications [3]. People with COPD have symptoms of breathing difficulty, sputum production, cough and wheezing. Also, their risk of developing lung cancer, heart disease and a variety of other infectious complications is increasing.

COPD is controllable and preventable. With proper management, most people with COPD can achieve good symptom control and quality of life, as well as reduced risk of other associated conditions. The treatment of stable COPD always includes oxygen therapy and smoking cessation. Although these measures can manage COPD, acute exacerbation still occurs due to respiratory failure. Endotracheal intubation and mechanical ventilation can be a life-saving procedure for patients with acute exacerbation of this disease [4]. However, the use of artificial airways may lead to infectious complications and injury to the trachea. Non-invasive positive pressure ventilation (NIPPV) [5] is an alternative approach that was developed to avoid these complications in patients with acute respiratory failure. And the hypercapnic ventilatory failure that occurs in patients with this disorder seems to respond well to non-invasive ventilation [6]. Previous studies suggested that NIPPV can reduce the need for endotracheal intubation and the length of the hospital stay [7]. A reduction in mortality with this approach in patients with COPD were also reported, in which patients who cannot tolerate this treatment were excluded from the comparison [8].

With the gradual maturity of NIPPV, its therapeutic efficacy have generated increasing concern. In general,

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Peng.

conscious patients receive NIPPV, while unconscious ones are directly treated with endotracheal intubation. Effective efficacy of NIPPV refers to good symptom control. Although NIPPV has the advantage of reducing the complications of endotracheal intubation and invasive ventilation, it is not always effective. If its therapeutic efficacy is judged to be ineffective, it means that endotracheal intubation must be performed. Failure of NIPPV will lead to an increase in mortality, which highlights the importance of rational diagnosis about the need for NIPPV therapy. Physicians need to comprehensively consider the clinical characteristics of patients, as there is always a coupling relationship between one another. Although they have already known which features are more important, they cannot judge with only a single characteristic. In order to avoid delaying endotracheal intubation and bringing about serious consequences attribute to ineffective NIPPV treatment [9], we hope to analyze whether the patient should adopt NIPPV based on the their own physical condition. It is important to note that an NIPPV efficacy predictive model is highly desirable to provide diagnostic decision making assistance for physicians treating patients with COPD. So far, however, few research has been carried out on NIPPV therapeutic efficacy prediction in patients with COPD.

As the therapeutic effect is annotated with binary labels, effective and ineffective, supervised machine learning algorithms are suitable for solving this problem of efficacy prediction. Machine learning and data-driven approaches are becoming significant in many areas. Usage of effective statistical models that capture the complex data dependencies is one of the crucial factors driving successful applications. Inductive learning [10] laid the theoretical foundation for machine learning. There are many successful application scenarios based on inductive learning, such as speech recognition [11], computer vision [12] and machine translation [13]. Predictive model is constructed by training inputs and outputs in existing sample. In our study, we define patient's clinical characteristics as inputs and therapeutic efficacy as outputs. When new patients encounter, we input their relevant characteristics in predictive model to obtain the prediction of NIPPV efficacy. Due to the complex pathogenesis of COPD, simple classifiers may not make good predictions. Given the high cost of data collection, we collected the NIPPV dataset with 144 cases from Sichuan Provincial People's Hospital in the past two years. One of the important things is to gather out the valuable features, which capture key information that dominate other types of features. According to this demand, we utilize tree-based method to achieve feature transformation. Because of the small size and the nonlinearity of our NIPPV dataset, we choose non-parametric methods, with which we do not have to worry about whether it is possible for the model to fit the data, to make our predictions. In summary, we propose a hybrid model for NIPPV therapeutic efficacy prediction that concatenate tree-based feature transformation method and Bayesian non-parametric classifier. Specific contents about models are detailed as follows.

Tree-based learning algorithms are considered to be one of the best and widely used supervised learning methods. Quinlan introduced decision tree [14] as a decision support tool, whose set of splitting rules used to segment the predictor space can be summarized in a tree. They empower predictive models with high accuracy, stability and ease of interpretation. Unlike linear models, they map non-linear relationships quite well. Other tree-based methods are also being popularly used in all kinds of data science problems. Among the machine learning methods used in practice, gradient boosting decision tree (GBDT) [15] is one technique that shines in many applications. Tree boosting has been shown to give state-of-the-art results on many standard classification benchmarks [16]. The impact of eXtreme Gradient Boosting (XGBoost) [17], a scalable system for tree boosting, has been widely recognized in a number of machine learning and data mining challenges. LambdaMART [18], a variant of tree boosting for ranking, achieves state-of-the-art result for ranking problems.

Features and data determine the ceiling of machine learning, while models and algorithms just try to approximate this upper limit. Practically, the features that can be directly used in machine learning models are often not enough. So the validity of models depends on the valuable features mined from raw data. Once we get the valuable features which captured key information and the right model, other factors play small roles. Besides acting as a stand-alone predictor, tree-based methods are also incorporated into generating tree features which are included as inputs of classifier [19]. Ad Click-through-rate (CTR) prediction in Facebook [20] used GBDT for non-linear feature transformation and fed them to Logistic regression (LR) [21] for the final prediction, proving that boosted decision trees are very powerful input feature transformations. It is also the defacto choice of ensemble method and is applied in challenges such as the Netflix prize [22]. Boosting neural networks with gradient boosting decision trees turns out to be the best among eight ensemble CTR estimation models [23].

Bayesian methods represent an important component of statistics, which achieved significant developments in machine learning [24]. In order to make sophisticated inference of hidden factors and predictions, prior knowledge is placed on uncertain evidence with incomplete information and randomness in Bayesian methods. Because of its flexibility, adaptivity and scalability, Bayesian plays a crucial role in protecting high-capacity models against overfitting, and allowing models adaptively updating their capacity. However, the application of Bayesian methods often stuck in computation and needs to be solved by approximate inference methods.

Parametric Bayesian comprises classic methodology for prior and posterior distributions in models with a finite, fixed number of parameters regardless of how the data changes. This restriction may limit the model capacity, especially for applications where it may be difficult or even counterproductive to predetermine the number of parameters. Take Gaussian

mixture model [25] as an example, it may not fit the variable dataset comes under a slightly changed distribution well if we choose the fixed number of clusters. In addition, the estimation and testing of parameters is based on the precondition of large samples and the assumption of normal distribution, which leads to the difference between the asymptotic distribution of statistics and the true distribution of raw data. Considering these limitations, it would be ideal to find a clustering model with automatic mechanism to figure out the unknown number of clusters. Feature representation learning [26] also has similar requirements on automatically figuring out the dimension of latent features and topological structure among features at different abstraction levels.

Bayesian non-parametric methods provide a solution to such needs on automatic model selection and adaptation using non-parametric models, as opposed to parametric models. Non-parametric models are completely driven by data, and distribution-free. By defining stochastic processes on rich measure spaces, Bayesian non-parametric approach fit a single model that can adapt its complexity to the data. Further, it allows the complexity to grow as more data comes, which means the dimension of the parameter space in a non-parametric approach should change with sample size. However, non-parametric models also have some drawbacks. It cannot extrapolate data that was not observed in the past and may lead to curse of dimensionality. Fortunately, the dataset we collected from hospital has a wide coverage of age and other clinical characteristics. Also, the number of features is small, which avoid such problems.

Gaussian Process (GP) [27], as a Bayesian non-parametric methodology, have become a promising scheme for the problem of classification/regression in recent years. GP methods are genuine probabilistic models that naturally give predictive probabilities for classification problems and provide information on how certain we are about the answer. As no weights are required to be estimated in non-parametric method, the non-linear functions are defined by GP priors with associated covariance functions for GP-based methods. The hyper-parameters of the covariance functions, which models classification as a GP, can be learned automatically from data instead of manually setting. Challis *et al.* [28] applied Bayesian Gaussian Process Logistic Regression for disease classification. Dhall and Goecke [29] proposed an expression intensity estimation method based on Gaussian process regression (GPR). Yuan *et al.* [30] combined local binary pattern (LBP) like features, kernel principal component analysis (KPCA), and GPR to propose a novel data processing pipeline for smoke detection. However, exact inference evaluation are intractable for GP based models. So several successful methods have been proposed for approximately integrating over the latent function values, such as the Laplace approximation [31], expectation propagation (EP) [32], Markov Chain Monte Carlo (MCMC) [33], and variational approximations [34].

In this paper, we concatenate tree-based feature transformation method and Bayesian non-parametric classifier to propose a hybrid approach for NIPPV therapeutic efficacy prediction. In particular, we transform features into a higher dimensional, sparse space by GBDT and include them as inputs of GPC. We calculate feature importance to justify the rationality of using tree-based feature transformation. Further, our hybrid model is flexible and extensible, so we can replace the first step of it with other similar tree-based methods, such as random forest and XGBoost, for further improvement. We show on a NIPPV dataset our hybrid model outperforms all baseline approaches working under the same conditions. Considering that the sample size of the NIPPV dataset is insufficient, we construct two simulated datasets with larger sample size and conduct additional experiments. Results have demonstrated that Bayesian non-parametric model based on tree-based feature transformation are superior to baseline in terms of capturing the leading features which include critical information about the patient or their symptom and enhancing the quality of classification. Several fundamental parameters impact the final prediction performance of our hybrid model. We then explore the fundamental parameters tuning in order to improve the quality of our training.

It is also important to note that we propose an NIPPV efficacy predictive model based on machine learning algorithms to provide diagnostic assistance for physicians treating patients with COPD. With our hybrid model, physicians can determine whether patients should receive NIPPV treatment to avoid delaying endotracheal intubation or mechanical ventilation according to their clinical characteristics. Especially for critically ill patients with urgent treatment time, physicians need a tool to aid decision making, which highlights the importance of our predictive model.

The rest of the paper is structured as follows. In Section II we describe exploratory analysis. We present our method in Section III. The experiments and results are detailed in Section IV. In Section V we conclude.

II. DATA SET AND EXPLORATORY ANALYSIS

We derived our dataset from the NIPPV efficacy data of people with COPD from the Department of Respiratory Medicine of Sichuan Provincial People's Hospital in 2016-2018. We collect 144 cases with 8 features: gender, age, illness duration and physiological indicators which include respiratory rate, potential of hydrogen (pH), partial pressure of oxygen (pO_2), partial pressure of carbon dioxide (pCO_2) and plasma albumin content. Except for GBDT, our work includes other methods which cannot tolerate missing values. So we omitted 3 cases, one of which is an outlier and the other two lost their features or outcomes. Preprocessing of the dataset consists of missing variables deletion and data standardization. After this phase, 141 cases remain for the NIPPV dataset construction.

Extensive research about NIPPV investigate the patient's physical condition by analyzing the patient's pulmonary function indicators such as forced vital capacity (FVC) and forced expiratory volume in one second (FEV_1) [35]. However, different from them, the condition of patients with

FIGURE 1. **Matrix of plots with the NIPPV dataset.**

COPD in China is generally more serious than that of foreign countries. Pulmonary function test requires patients to inhale deeply, then quickly blow with force and continue for six seconds without interruption. This procedure need strong compliance, which is not suitable for critically ill patients. The results of blood gases before and after NIPPV treatment showed that pO_2 and pH were significantly increased while pCO_2 was the opposite [36], [37]. pO_2 is a measurement of oxygen pressure in arterial blood. pCO_2 is one of several tests used to measure arterial blood gasses in people with lung disease and other illnesses, which evaluates how well $CO2$ moves from the lungs into the blood. If the partial pressure of oxygen and carbon dioxide is normal, the molecules will move from the alveoli into the blood and back as they should. Changes in that pressure can result in getting too little oxygen in the blood or accumulating too much carbon dioxide in the blood. Having too much carbon dioxide is called hypercapnia, a condition common in people with late-stage COPD [38], which will lead to low pH values. Previous studies suggested that these indicators bear on the therapeutic effect of NIPPV, since it has been recognized as a significant advance in the management of acute hypercapnic respiratory failure [39]. Apart from above, previous experiments also include body mass index (BMI) as an index of clinical characteristics [40], [41]. But most hospitals in China do not have such hospital beds with the function of weight measurement, and critical patients are too heavy to move, making it difficult to obtain BMI. However, BMI mainly reflects the nutritional status of patients, which is also closely related to plasma albumin content. Therefore, we choose plasma albumin as a surrogate, given that it is more accurate, more objective, and more valuable.

In order to observe the relationship between variables and explore the characteristics of each feature, we describe it explicitly in the form of pictures. In Figure 1, we make a matrix of plots with the NIPPV dataset. Effective and ineffective treatments are represented in blue and pink respectively. In the upper part of the matrix, we compute the coefficient of correlation between continuous variables on the overall data and on different efficacy data respectively. Box plots are displayed on continuous and categorical data. Similarly, we show this two forms of data with scatter plot and histogram respectively in the lower part of the matrix. As for diagonal, we apply density plot for continuous data and bar chart for discrete. From box plots, we can see that the NIPPV efficacy tends to be effective as the respiratory rate decrease, while its density plot and scatter plot has the similar characteristic. From the medical point of view, NIPPV generally works in patients with acute respiratory failure. Although not as obvious as respiratory rate, pO_2 has the same trend. It is also illustrated in Figure 1 that the higher the plasma albumin content, the more effective of NIPPV. The contribution of other features cannot be intuitively obtain from this figure, which need further analysis later.

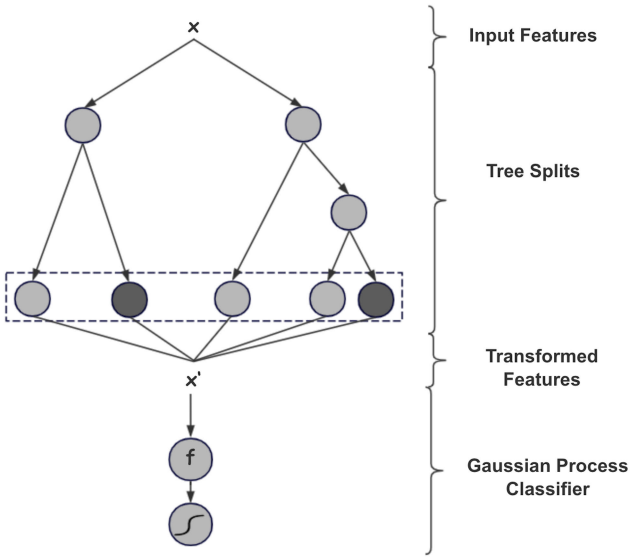

FIGURE 2. Hybrid model structure. Fit an ensemble of trees on the training set to transform input features into a higher dimensional, sparse space. Then train a Gaussian process classifier model on these features.

III. BAYESIAN NON-PARAMETRIC BASED PREDICTION MODEL

In this section we present a hybrid model structure: the concatenation of trees and a Bayesian non-parametric classifier, illustrated in Figure 2. In Section III-A, we fit an ensemble of trees on the training set. Each individual tree is treated as a categorical feature which takes the index of the leaf an instance ends up falling in as value. Therefore, input features are transformed into a higher dimensional, sparse space. Then train a Bayesian non-parametric classifier model on these features. Symbol f represents the latent function with a Gaussian process prior and then be "squashed" through the logistic function, which plays a crucial role in making predictions and detailed in Section III-B.

A. FEATURE TRANSFORMATION

The most important thing is to have the right features: those capturing critical information about the patient or symptom that dominate other types of features. Once we have the appropriate model with the right features, other factors make a small difference [20].

We use tree-based methods to generate tree features which are included as inputs of classifier. The leaf indices of each tree in the ensemble are then encoded in a one-hot fashion. Each sample passes through the decisions of each tree of the ensemble and ends up in one leaf per tree, and the sample is encoded by setting feature values for these leaves to 1 and the other feature values to 0.

For instance, consider the tree-baed model illustrated in Figure 2 with two subtrees, where the first subtree has two leaves and the second three leaves. It can be seen from the figure that a sample ends up in leaf 2 in the first subtree and leaf 3 in second subtree, represented as deepen opaque dots. The overall input to the Gaussian process classifier will be the binary-value vector [0, 1, 0, 0, 1], where the first two entries correspond to the leaves of the first subtree and last three represent those of the second subtree.

In each learning iteration, a new tree is created to model the residual of previous trees. We can regard tree-based transformation as a supervised feature encoding which converts a real-valued vector into a compact binary-valued vector. The traversing from root node to a leaf node reveals a rule on certain features. During the training phase, both root nodes and leaf nodes are local optimal features. With this characteristic of tree-based model, we input the transformed features into classifier so as to improve the fitting ability of classification.

B. BAYESIAN NON-PARAMETRIC CLASSIFICATION

Given a training set $\mathcal{D}$ of n observations, $\mathcal{D} = \{(\mathbf{x}_i, y_i)\,|i = 1, \ldots, n\}$. Here, $\mathbf{x}$ denotes the input generated from tree-based feature transformation and we use the labels $y = +1$ and $y = -1$ to distinguish two classes. We can write $\mathcal{D} = (X, \mathbf{y})$, as the column vector inputs for all n cases are aggregated in the design matrix X, and the targets are collected in the vector $\mathbf{y}$.

As can be seen from Figure 2, a GP prior is placed over the latent function $f(x)$, whose value will then be compressed into 0 to 1 through logistic function to obtain a prior on $\kappa(\mathbf{x}) \triangleq p(y = +1|\mathbf{x}) = \sigma(f(\mathbf{x}))$. Note that the purpose of $f(x)$ is simply to allow a easier formulation of the model, and the computational goal pursued will be to integrate out f. Usually, for notational simplicity we suppose the mean function of the GP prior to be zero.

Inference process contains two steps. Firstly, computing the distribution of the latent variable $f_* \triangleq f(\mathbf{x}_*)$ corresponding to a test case x_*

$$p(f_*|X, \mathbf{y}, \mathbf{x}_*) = \int p(f_*|X, \mathbf{x}_*, \mathbf{f})\, p(\mathbf{f}|X, \mathbf{y}) d\mathbf{f}, \tag{1}$$

where $p(\mathbf{f}|\mathbf{X}, \mathbf{y}) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|X)/p(\mathbf{y}|\mathbf{X})$ is the posterior of the latent variables and $\mathbf{f} = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))^\top$ represents Gaussian process latent function values. Secondly, using this distribution over the latent f_* to produce a probabilistic prediction

$$\overline{\kappa}_* \triangleq p(y_* = +1|X, \mathbf{y}, \mathbf{x}_*) = \int \sigma(f_*)\, p(f_*|X, \mathbf{y}, \mathbf{x}_*)\, df_*. \tag{2}$$

In classification the non-Gaussian likelihood $p(\mathbf{y}|\mathbf{f})$ in eq.(1) makes the integral analytically intractable. Similarly, eq.(2) can hardly be intractable analytically. Therefore, we use the Laplace approximation as represented by Williams and Rasmussen [42] to approximate the non-Gaussian joint posterior with a Gaussian one.

Doing a second Taylor expansion of $\log p(\mathbf{f}|X, \mathbf{y})$ around the maximum of the posterior, we obtain a Gaussian approximation

$$q(\mathbf{f}|X, \mathbf{y}) = \mathcal{N}(\mathbf{f}|\hat{\mathbf{f}}, A^{-1}) \propto \exp(-\frac{1}{2}(\mathbf{f} - \hat{\mathbf{f}})^\top A(\mathbf{f} - \hat{\mathbf{f}})), \tag{3}$$

where $\hat{\mathbf{f}} = \text{argmax}_{\mathbf{f}}\, p(\mathbf{f}|X, \mathbf{y})$ and $A = -\nabla\nabla \log p(\mathbf{f}|X, \mathbf{y})|_{\mathbf{f}=\hat{\mathbf{f}}}$ is the Hessian of the negative log posterior at that point.

By Bayes' rule the posterior over the latent variables is given by $p(\mathbf{f}|\mathbf{X}, \mathbf{y}) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|X)/p(\mathbf{y}|\mathbf{X})$. We need only consider the un-normalized posterior as $p(\mathbf{y}|\mathbf{X})$ is independent of $\mathbf{f}$. Taking a logarithm of $p(\mathbf{f}|\mathbf{X}, \mathbf{y})$ we obtain

$$\begin{aligned}\Psi(\mathbf{f}) &\triangleq \log p(\mathbf{y}|\mathbf{f}) + \log p(\mathbf{f}|X) \\ &= \log p(\mathbf{y}|\mathbf{f}) - \frac{1}{2}\mathbf{f}^\top K^{-1}\mathbf{f} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi, \quad (4)\end{aligned}$$

where $K \triangleq K(X, X)$ represents the $n \times n$ covariance matrix and $\log p(\mathbf{f}|X) = -\frac{1}{2}\mathbf{f}^\top K^{-1}\mathbf{f} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi$ holds under the assumption of the GP prior $\mathbf{f}|X \sim \mathcal{N}(\mathbf{0}, K)$.

Differentiating eq.(4) gives

$$\nabla\Psi(\mathbf{f}) = \nabla \log p(\mathbf{y}|\mathbf{f}) - K^{-1}\mathbf{f}, \quad (5)$$

$$\nabla\nabla\Psi(\mathbf{f}) = \nabla\nabla \log p(\mathbf{y}|\mathbf{f}) - K^{-1} = -W - K^{-1}, \quad (6)$$

where $W \triangleq -\nabla\nabla \log p(\mathbf{y}|\mathbf{f})$ is a diagonal since the likelihood factorizes over cases.

We use Newton's method to find the maximum of Ψ, and at the maximum of $\Psi(\mathbf{f})$ we have

$$\nabla\Psi = \mathbf{0} \Longrightarrow \hat{\mathbf{f}} = K(\nabla \log p(\mathbf{y}|\hat{\mathbf{f}})). \quad (7)$$

After finding the maximum posterior $\hat{\mathbf{f}}$, we specify the Laplace approximation to the posterior as a Gaussian with mean $\hat{\mathbf{f}}$ and covariance matrix given by the negative inverse Hessian of Ψ from eq.(6) as

$$q(\mathbf{f}|X, \mathbf{y}) = \mathcal{N}\left(\hat{\mathbf{f}}, \left(K^{-1} + W\right)^{-1}\right). \quad (8)$$

The posterior mean for f_* can be expressed by combining the GP predictive mean with eq.(7) into

$$\mathbb{E}_q\left[f_*|X, \mathbf{y}, \mathbf{x}_*\right] = \mathbf{k}\left(\mathbf{x}_*\right)^\top K^{-1}\hat{\mathbf{f}} = \mathbf{k}\left(\mathbf{x}_*\right)^\top \nabla \log p(\mathbf{y}|\hat{\mathbf{f}}), \quad (9)$$

note that we write $\mathbf{k}\left(\mathbf{x}_*\right) = \mathbf{k}_*$ to denote the vector of covariances between the test point and the n training points $K(X, x_*)$.

We also compute the variance of $f_*|X, \mathbf{y}$ under the Gaussian approximation

$$\mathbb{V}_q\left[f_*|X, \mathbf{y}, \mathbf{x}_*\right] = k\left(\mathbf{x}_*, \mathbf{x}_*\right) - \mathbf{k}_*^\top\left(K + W^{-1}\right)^{-1}\mathbf{k}_*. \quad (10)$$

Given the mean and variance of f_*, we make predictions by computing

$$\overline{\kappa}_* \simeq \mathbb{E}_q\left[\pi_*|X, \mathbf{y}, \mathbf{x}_*\right] = \int \sigma\left(f_*\right) q\left(f_*|X, \mathbf{y}, \mathbf{x}_*\right) df_*. \quad (11)$$

where $q\left(f_*|X, \mathbf{y}, \mathbf{x}_*\right)$ is Gaussian with mean and variance given by equations (9) and (10) respectively.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. FEATURE IMPORTANCE

We delve into the importance of our features in the NIPPV dataset in order to justify the rationality of using tree-based feature transformation. A benefit of using tree-based methods like GBDT is that they can automatically estimate feature importance from a trained predictive model. Importance provides a score that indicates how useful each feature is in the construction of the boosted decision trees within the model. Theoretically, the more a feature is used to make key decisions with trees, the higher its score. Importance of a single decision tree is calculated by the amount of performance improvement for each split point, which weighted by the number of observations the node is responsible for. The performance metric can be the purity used to select the split point or another more specific error function. Then, take an average of all decision trees within the model to obtain the score of feature importance.

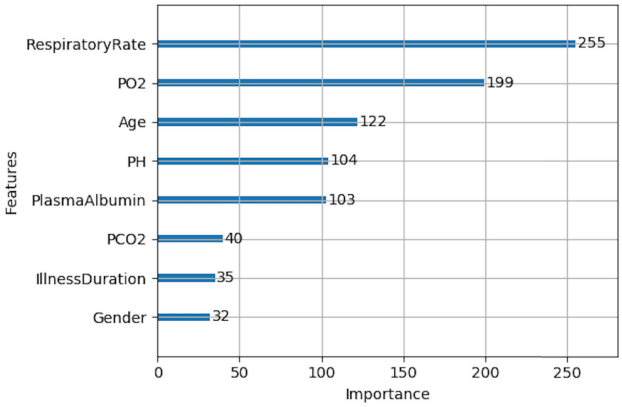

FIGURE 3. Feature importance given by GBDT. Calculate the importance of features in the NIPPV dataset, which measures the contribution of each feature is in the construction of the boosted decision trees.

As can be seen from Figure 3, all features in the NIPPV dataset are ranked by importance. Respiratory rate is the most important, followed by pO_2. Age, pH, and plasma protein also have considerable importance, arranged behind them in turn. We analyze from a medical perspective as follows. Respiratory rate is intuitive and real-time. Although pH and pO_2 are also important considerations, due to the need of blood tests, it takes at least half an hour to get the test results. The impact of age also need to be taken into account. As the senses of the elderly are relatively sluggish, he will be more tolerant of NIPPV's discomfort while the tolerance of young people is worse since the sensitivity of his face and throat. Therefore, the cooperation of young people is worse. In theory, the greater the patient's age, the better the compliance, then the efficacy will be guaranteed. However, it is also depend on their condition.

Physicians need to analyze the coupling relationships between the clinical characteristics. This process is complicated but extremely important for physicians to make accurate judgments. In this case, our machine learning model assists in medical diagnosis by identifying potential relationships between variables and predicting the effectiveness of NIPPV based on the physiological indicators of patients. This result of feature importance given by GBDT is generally consistent with the exploratory data analysis in Section II, which indicates that tree-based method really can capture the valuable features that dominate others.

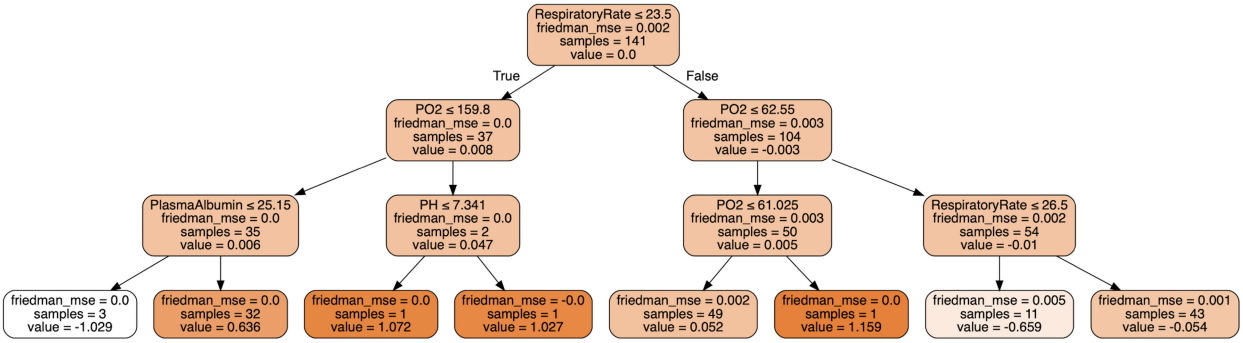

FIGURE 4. One of the tree trained by GBDT.

TABLE 1. Predictive performance comparison of several methods on three datasets.

Methods	NIPPV dataset				Simulated dataset1				Simulated dataset2			
	Acc.	p	r	F_1	*Acc.*	p	r	F_1	*Acc.*	p	r	F_1
LR[1]	0.7386	0.7621	0.7571	0.7563	0.8301	0.8285	0.8318	0.8274	0.8269	0.8364	0.8186	0.8255
SVC[2]	0.5534	0.5534	1.0000	0.7123	0.8301	0.8417	0.8157	0.8247	0.8199	0.8274	0.8147	0.8193
GBDT[3]	0.5979	0.5850	0.9714	0.7272	0.8288	0.8472	0.8127	0.8290	0.8289	0.8465	0.8167	0.8290
GPC[4]	0.7719	0.8034	0.7929	0.7934	0.8361	0.8386	0.8358	0.8337	0.8248	0.8363	0.8106	0.8219
GBDT-LR	0.8250	0.8667	0.9333	0.8867	0.8862	0.9050	0.8800	0.8853	0.8723	0.8849	0.8944	0.8871
GBDT-SVC	0.7833	0.8167	0.9500	0.8467	0.8714	0.9181	0.8533	0.8694	0.8799	0.9247	0.8427	0.8755
GBDT-GPC	**0.8667**	**0.9000**	**0.9500**	**0.9067**	**0.9218**	**0.9348**	**0.9233**	**0.9244**	**0.9167**	**0.9333**	**0.9333**	**0.9199**
RF[5]-GPC	0.8583	0.8667	0.9500	0.8967	0.9087	0.9071	0.9500	0.9253	0.9183	0.9133	0.9099	0.9072
XGB[6]-GPC	0.8500	0.8833	0.9000	0.8600	**0.9385**	**0.9633**	**0.9099**	**0.9290**	0.8985	0.9183	0.8750	0.8904

[1]Logistic regression; [2]Support vector machine; [3]Gradient boosting decision tree;[4]Gaussian process classification; [5]Random forest; [6]eXtreme gradient boosting.

According to the idea of boosting, each step of GBDT uses a decision tree to fit the residual of current learning and then obtain a new tree. Figure 4 shows one of the decision tree trained by GBDT. The decision process of the tree is actually a simulation of physicians' diagnosis, whose splitting rules take into account the importance of the features and Friedman mean squared error [15] of each step. Each node represents a feature, each branch represents a decision and each leaf represents an outcome. As illustrated in this figure, each sample goes through the decision tree and ends up in one leaf. It turns out that the features displayed on the nodes are important for determining whether a patient should receive NIPPV treatment. As an ensembling technique, GBDT combined several decision trees as shown in Figure 4 to yield a powerful model, in an iterative fashion.

B. EXPERIMENT RESULTS

To demonstrate performance of our method, we compared the proposed method with existing methods on the NIPPV dataset. We implement LR, SVC, GBDT and GPC without feature transformation to set up the baseline. As for the associated evaluation metrics, accuracy [43] and precision [44] are the metrics most often applied, together usually, with recall [45] and F_1 [46]. In order to evaluate the quality of the classifiers, we applied 10-fold cross-validation (CV). In this process, we obtain the average of each metrics mentioned above. All data will be involved in training and prediction, which can effectively avoid overfitting and eliminate randomness.

Table 1 shows the results of our experiments on three datasets. For convenience, the results of the baseline methods are displayed in the top four rows. Bayesian non-parametric model GPC obtained the best performance among the four methods. The rows marked "GBDT-" apply GBDT in generating tree features before classification. GBDT-LR and GBDT-SVC achieved improvements over pure classification. Considering that GBDT with parametric models have already obtain significant improvements, and are therefore higher baselines, the performance improvements of GBDT-GPC are very encouraging.

Since the expensive cost of collect such NIPPV treatment for COPD patients data, the sample size of real data is small. In order to validate our proposed method on datasets with larger sample size, we generate two random 2-class classification simulated datasets, which contain 500 cases with 8 features and 1000 cases with 20 features, from the

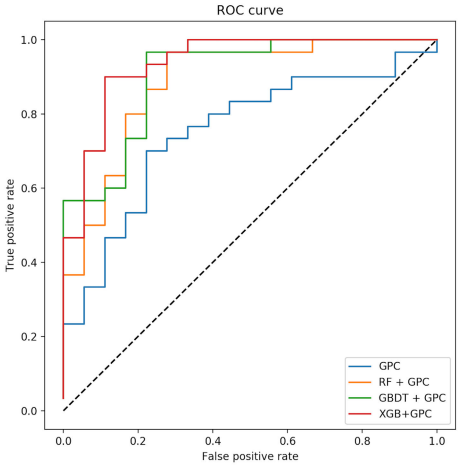

FIGURE 5. Performance of hybrid model with other tree-based feature transformations. Replace the first step with random forest and XGBoost. Draw their ROC curves for performance comparison.

machine learning library scikit-learn. In the middle and right half of Table 1, evaluation metrics of baseline models are higher than the results on the NIPPV dataset. Similar to the left half, the performance of models is significantly improved after the joint of tree-based feature transformation, in which GBDT-GPC works best. The results demonstrate that our method has powerful generalization performance and the ability of dealing with data of different size, which is even one order of magnitude larger than the original sample size.

In addition to studying the findings as discussed above, it is interesting to see how different methods of tree-based feature transformation affect the performance. In Figure 5 we show an overview of the different experimental results. Receiver Operating Characteristics (ROC) curve [47] is a graphical plot that illustrates the diagnostic ability of a binary classifier system. Higher the Area Under Curve (AUC), better the model is at distinguishing between classes. As expected, GPC with tree-based feature transformations are easier to perform well, as reflected in the figure by the large AUC. The AUC of GBDT-GPC, RF-GPC and XGB-GPC reached 0.9037, 0.8852 and 0.9370 respectively, while GPC was only 0.7519. This indicates that our hybrid model is flexible and extensible, as we can replace the first step with RF or XGBoost. The performance of these two methods are shown in the last two rows of Table 1. It can be seen that both of them achieved substantial performance improvements, with XGB-GPC even obtain higher accuracy than GBDT-GPC.

C. TRAINING OPTIMIZATIONS

The next task is to optimize the performance and accuracy in training phase. In this part, we share several accuracy-critical factors have proven effective for our method.

Theoretically, the more trees in the model the longer the time required to make a prediction. We vary the number of boosted trees from 1 to 130 and study the impact of the

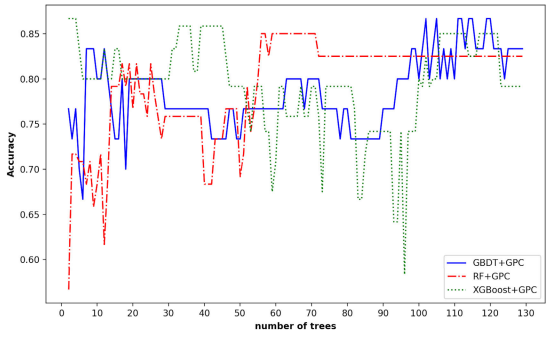

FIGURE 6. Predictive performance comparison with different number of boosting trees.

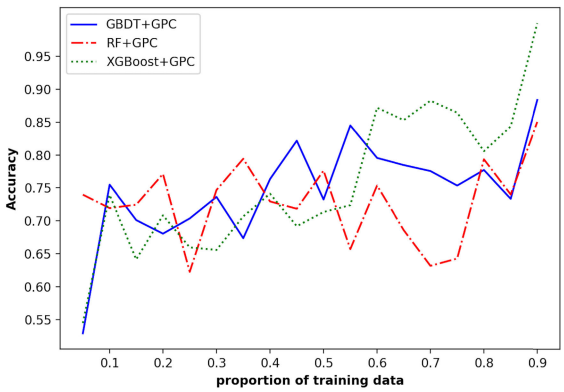

FIGURE 7. Predictive performance comparison with different size of training set.

number of trees on estimation accuracy. The experimental results are shown in Figure 6. The accuracy of the three methods fluctuated greatly before the number of trees reached 100. With the number above 100, the performance is relatively better and tends to be stable. However, the raise in the number does not cause a significant increase in performance and add training cost. Therefore, it is reasonable to control the number of trees in the range of 100 to 120.

In the following, we do some study to explore the effect of different proportions of training data. It is shown in Figure 7 that when the proportion is less than 10%, the performance of the two models, GBDT-GPC and XGBoost-GPC, is poor. As the training size improves from 10% to 85%, the accuracy is unstable since it fluctuates drastically. The performance of the tree models is greatly enhanced when the proportion is higher than 85%, which may be due to the fact that our medical data is difficult to obtain and the size of dataset is small.

From the above results illustrated in Figure 3, top 5 features are responsible for more than half of total feature importance, while the last 3 contribute less than 13%. Based on this finding, we further experiment with keeping 1 to 8 features, and evaluate how the accuracy is effected. As shown in Figure 8, the performance of XGBoost achieved huge improvement since the number of features greater than 3. It may be caused by the gradually joint of the most importance features.

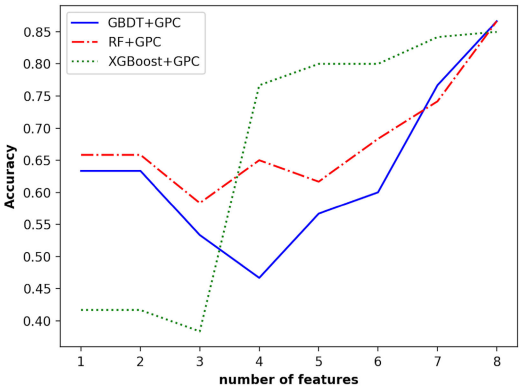

FIGURE 8. **Predictive performance comparison with different number of features.**

The performance of three model peaks at 8, indicating that all features are included for better results, although some of them are less important.

V. CONCLUSION

NIPPV is an alternative approach of endotracheal intubation and mechanical ventilation to avoid infectious complications and injury to the trachea in patients with acute respiratory failure. But it is not always effective. In order to avoid delaying endotracheal intubation due to ineffective NIPPV treatment, we introduce a hybrid model to predict whether the patient should adopt NIPPV based on the their own physical condition.

Efficacy prediction of NIPPV can be regarded as a specific kind of binary classification. We improve the performance of classifier by concatenating tree-based feature transformation method and Bayesian non-parametric model. Specifically, we generate tree features by GBDT and then model classification as a GP with no assumptions about data structures. We delved into the feature importance and justified that tree-based method really can capture the valuable features that dominate others. The experiments carried out both on the NIPPV dataset and two simulated datasets with lager sample size, which validated the powerful generalization performance and robustness of our method. GBDT-GPC have demonstrated state-of-the-art performance on NIPPV therapeutic efficacy prediction task. Further, we can replace the first step of our hybrid model with tree-based method like random forest and XGBoost for further improvement, since it is flexible and extensible. We analyzed the effect of fundamental parameters tuning on the final prediction performance. With our hybrid model, physicians can determine whether patients should receive NIPPV treatment according to their clinical characteristics. Especially for critically ill patients, it provides diagnostic assistance and avoids delaying endotracheal intubation or mechanical ventilation.

REFERENCES

[1] A. Tee, "Chronic obstructive pulmonary disease (copd): Not a cigarette only pulmonary disease," *Ann. Acad. Med.*, vol. 46, no. 11, pp. 415–416, 2017.
[2] M. T. Dransfield, K. M. Kunisaki, M. J. Strand, A. Anzueto, S. P. Bhatt, R. P. Bowler, G. J. Criner, J. L. Curtis, N. A. Hanania, and H. Nath, "Acute exacerbations and lung function loss in smokers with and without chronic obstructive pulmonary disease," *Amer. J. Respiratory Crit. Care Med.*, vol. 195, no. 3, pp. 324–330, 2017.
[3] J. Obi, A. Mehari, and R. Gillum, "Mortality related to chronic obstructive pulmonary disease and co-morbidities in the United States, a multiple causes of death analysis," *COPD, J. Chronic Obstructive Pulmonary Disease*, vol. 15, no. 2, pp. 200–205, 2018.
[4] A. Holm and P. Dreyer, "Intensive care unit patients' experience of being conscious during endotracheal intubation and mechanical ventilation," *Nursing Crit. Care*, vol. 22, no. 2, pp. 81–88, 2017.
[5] E. Brambilla, R. F. Doherty, and P. R. Kwok, "Patient interface and non-invasive positive pressure ventilating method," U.S. Patent 9 782 553, Oct. 10, 2017.
[6] B. M. Raju, S. Jotkar, M. Prathyusha, S. Goswami, M. Dube, and A. Singh, "Effectiveness of non-invasive positive pressure ventilation for acute exacerbation of chronic obstructive pulmonary disease," *Int. J.*, vol. 5, no. 2, p. 102, 2018.
[7] A. Gomez, M. Weerasekara, P. Wickramaarachchi, K. Prathapasinghe, and A. W. Wickramanayaka, "Comparison of continuous positive airway pressure and non-invasive positive pressure ventilation as modes of non-invasive respiratory support for neonates in a level III neonatal intensive care unit," *Sri Lanka J. Child Health*, vol. 47, no. 3, pp. 242–248, 2018.
[8] X. Jiang, H. Xiao, R. Segal, W. C. Mobley, and H. Park, "Trends in readmission rates, hospital charges, and mortality for patients with chronic obstructive pulmonary disease (copd) in Florida from 2009 to 2014," *Clin. Therapeutics*, vol. 40, no. 4, pp. 613–626, 2018.
[9] R. Scala and L. Pisani, "Noninvasive ventilation in acute respiratory failure: Which recipe for success?" *Eur. Respiratory Rev.*, vol. 27, no. 149, 2018, Art. no. 180029.
[10] R. S. Michalski, "A theory and methodology of inductive learning," *Machine Learning*. Berlin, Germany: Springer, 1983, pp. 83–134.
[11] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.
[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[13] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, and R. Zens, "Moses: Open source toolkit for statistical machine translation," in *Proc. 45th Annu. Meeting Assoc. Comput. Linguistics Companion*, 2007, pp. 177–180.
[14] J. R. Quinlan, "Induction of decision trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.
[15] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, 2001.
[16] P. Li, "Robust logitboost and adaptive base class (ABC) logitboost," 2012, *arXiv:1203.3491*. [Online]. Available: https://arxiv.org/abs/1203.3491
[17] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
[18] C. J. Burges, "From ranknet to lambdarank to lambdamart: An overview," *Learning*, vol. 11, p. 81, Jun. 2010.
[19] X. Wang, X. He, F. Feng, L. Nie, and T.-S. Chua, "Tem: Tree-enhanced embedding model for explainable recommendation," in *Proc. World Wide Web Conf.*, 2018, pp. 1543–1552.
[20] X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, and S. Bowers, "Practical lessons from predicting clicks on ADS at Facebook," in *Proc. 8th Int. Workshop Data Mining Online Advertising*, 2014, pp. 1–9.
[21] D. R. Cox, "The regression analysis of binary sequences," *J. Roy. Stat. Soc., B (Methodol.)*, vol. 20, no. 2, pp. 215–232, 1958.
[22] J. Bennett and S. Lanning, "The netflix prize," in *Proc. KDD Cup Workshop*, New York, NY, USA, 2007, p. 35.
[23] X. Ling, W. Deng, C. Gu, H. Zhou, C. Li, and F. Sun, "Model ensemble for click prediction in bing search ADS," in *Proc. 26th Int. Conf. World Wide Web Companion*, 2017, pp. 689–698.
[24] J. Zhu, J. Chen, W. Hu, and B. Zhang, "Big learning with Bayesian methods," *Nat. Sci. Rev.*, vol. 4, no. 4, pp. 627–651, 2017.
[25] S. Newcomb, "A generalized theory of the combination of observations so as to obtain the best result," *Amer. J. Math.*, vol. 8, no. 4, pp. 343–366, 1886.

[26] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[27] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer School on Machine Learning*. Berlin, Germany: Springer, 2003, pp. 63–71.
[28] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, "Gaussian process classification of Alzheimer's disease and mild cognitive impairment from resting-state fMRI," *NeuroImage*, vol. 112, pp. 232–243, May 2015.
[29] A. Dhall and R. Goecke, "Group expression intensity estimation in videos via Gaussian processes," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, 2012, pp. 3525–3528.
[30] F. Yuan, X. Xia, J. Shi, H. Li, and G. Li, "Non-linear dimensionality reduction and Gaussian process based classification method for smoke detection," *IEEE Access*, vol. 5, pp. 6833–6841, 2017.
[31] C. K. I. Williams and D. Barber, "Bayesian classification with Gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1342–1351, Dec. 1998.
[32] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. 17th Conf. Uncertainty Artif. Intell.* San Mateo, CA, USA: Morgan Kaufmann, 2001, pp. 362–369.
[33] R. M. Neal, "Monte Carlo implementation of Gaussian process models for Bayesian regression and classification," 1997, *arXiv:physics/9701026*. [Online]. Available: https://arxiv.org/abs/physics/9701026
[34] M. N. Gibbs and D. J. C. Mackay, "Variational Gaussian process classifiers," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1458–1464, Nov. 2000.
[35] T. Köhnlein, W. Windisch, D. Köhler, A. Drabik, J. Geiseler, S. Hartl, O. Karg, G. Laier-Groeneveld, S. Nava, and B. Schönhofer, "Non-invasive positive pressure ventilation for the treatment of severe stable chronic obstructive pulmonary disease: A prospective, multicentre, randomised, controlled clinical trial," *Lancet Respiratory Med.*, vol. 2, no. 9, pp. 698–705, 2014.
[36] L. Cong, L. Zhou, H. Liu, and J. Wang, "Outcomes of high-flow nasal cannula versus non-invasive positive pressure ventilation for patients with acute exacerbations of chronic obstructive pulmonary disease," *Int. J. Clin. Exp. Med.*, vol. 12, no. 8, pp. 10863–10867, 2019.
[37] A. S. Salah, M. S. Sobh, A. S. Shaker, and T. H. Hamdy, "Role of non invasive positive pressure ventilation as a weaning method in mechanically ventilated COPD patients," *Zagazig Univ. Med. J.*, vol. 23, no. 5, pp. 311–321, 2018.
[38] L. Pisani, L. Fasano, N. Corcione, V. Comellini, M. A. Musti, M. Brandao, D. Bottone, E. Calderini, P. Navalesi, and S. Nava, "Change in pulmonary mechanics and the effect on breathing pattern of high flow oxygen therapy in stable hypercapnic copd," *Thorax*, vol. 72, no. 4, pp. 373–375, 2017.
[39] C. R. Osadnik, V. S. Tee, K. V. Carson-Chahhoud, J. Picot, J. A. Wedzicha, and B. J. Smith, "Non-invasive ventilation for the management of acute hypercapnic respiratory failure due to exacerbation of chronic obstructive pulmonary disease," *Cochrane Database Syst. Rev.*, vol. 7, no. 7, 2017, Art. no. CD004104.
[40] R. Chen, L. Guan, W. Wu, Z. Yang, X. Li, Q. Luo, Z. Liang, F. Wang, B. Guo, and Y. Huo, "The chinese version of the severe respiratory insufficiency questionnaire for patients with chronic hypercapnic chronic obstructive pulmonary disease receiving non-invasive positive pressure ventilation," *BMJ Open*, vol. 7, no. 8, 2017, Art. no. e017712.
[41] L. Zhou, L. Guan, W. Wu, X. Li, X. Chen, B. Guo, Y. Huo, J. Xu, Y. Yang, and R. Chen, "High-pressure versus low-pressure home non-invasive positive pressure ventilation with built-in software in patients with stable hypercapnic COPD: A pilot study," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 16728.
[42] C. K. Williams and C. E. Rasmussen, *Gaussian Processes in Machine Learning*, vol. 2. Cambridge, MA, USA: MIT Press, 2006.
[43] P. Annesi, D. Croce, and R. Basili, "Semantic compositionality in tree kernels," in *Proc. 23rd ACM Int. Conf. Knowl. Manage.*, 2014, pp. 1029–1038.
[44] S. Fernando and M. Stevenson, "A semantic similarity approach to paraphrase detection," in *Proc. 11th Annu. Res. Colloq. UK Special Interest Group Comput. Linguistics*, 2008, pp. 45–52.
[45] B. Hu, Z. Lu, H. Li, and Q. Chen, "Convolutional neural network architectures for matching natural language sentences," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2042–2050.
[46] R. Mihalcea, C. Corley, and C. Strapparava, "Corpus-based and knowledge-based measures of text semantic similarity," in *Proc. AAAI*, vol. 6, 2006, pp. 775–780.
[47] L. B. Lusted, "Signal detectability and medical decision-making," *Science*, vol. 171, no. 3977, pp. 1217–1219, 1971.

YANG WENG received the B.S. and Ph.D. degrees from the Department of Mathematics, Sichuan University, Chengdu, China, in 2001 and 2006, respectively. Since 2006, he has been with the College of Mathematics, Sichuan University, where he is currently an Associate Professor. He was a Postdoctoral Fellow with Nanyang Technological University, Singapore, from August 2008 to July 2010. His current research interests include statistic machine learning and nonparametric Bayesian inference.

YIN FANG received the B.E. degree from Sichuan University, Chengdu, China, in 2017, where she is currently a Graduate Student with the Department of Mathematics. Her current research interests include Bayesian non-parametrics and natural language processing.

HAIYING YAN received the B.S. degree from the School of Clinical Medical Sciences, Southwest Medical University, in 2002. Since then, she has been with Sichuan Provincial People's Hospital, where she is currently the Deputy Chief Physician of the Department of Respiration and Critical Care Medicine. Her current research interests include COPD and chronic disease management.

YANG YANG is currently the Director of the Department of Respiratory and Critical Care Medicine, Sichuan Provincial Academy of Medical Sciences, Sichuan Provincial People's Hospital. He was the Principal Investigator of several provincial-level scientific research projects and a number of international multicenter clinical drug research projects. He has published more than ten articles in the core journals of this research area.

WENXING HONG received the Ph.D. degree in system engineering from Xiamen University, China, in 2010. Since 2010, he has been with the Department of Automation, Xiamen University, China, where he is currently an Associate Professor. He is also the Dean of the Research Center for Systems and Control, Xiamen University. He is the author or coauthor of more than 25 articles. He has led and participated in more than 15 research projects and funds, including National Natural Science Foundation of China. His current research interests are in the area of data mining, big data, artificial intelligence, recommendation systems, and FinTech. He is a member of the CCF. He had served as the General Secretary for the International Conference on Computer Science and Education (ICCSE) and Fujian Systems Engineering Society, from 2006 and 2010.