Predictive Analytics Dashboard for Monitoring Patients in Advanced Stages of COPD

Riad Alharbey
Claremont Graduate University
Riad.alharbey@cgu.edu

Abstract
Chronic Obstructive Pulmonary Disease (COPD) becomes a highly prevalent disease that impacts both patients and healthcare systems. Patients in the advanced stages of COPD can be vulnerable to acute exacerbations that can be fatal. The exacerbations can affect hospitals’ resources caused by frequent admissions and readmissions. Caregivers want a reliable mechanism that can help them manage COPD patients remotely and predict the risks of exacerbation in advance before COPD episode is about to occur. It is vital to develop effective tools to provide much-needed assistance for the elderly COPD patients. Tools involve understanding the patient’s behavior and providing indicators that assist the decision-making process. In this paper, I discuss the design and implementation of a feed forward multi-layer neural network for prediction of COPD acute exacerbations. I used the backpropagation algorithm to train the ANN network. The results show a positive correlation between the computed outputs and the desired outputs with respect to COPD Clinical Questionnaire.

1. Introduction

Chronic obstructive Pulmonary Disease (COPD) is a highly prevalent disease worldwide [1]. By 2020, the disease is expected to be the third leading cause of death, and costs about $49 billion annually. COPD symptoms vary depending on the amount of damages in the lung. Function Pulmonary Tests are meant to measure how the patient’s lung works [2].

According to Chronic Obstructive Lung Disease (GOLD) [1], COPD stages have been classified primarily, based on the values of Forced Expiratory Volume tests, into four stages [2]. Stage 1 of COPD has mild symptoms, like some shortness of breath. Most likely, there is no cough or mucus. In this stage, FEVs values are usually greater than 80%. In stage 2, the shortness level of breath can be moderate when the patients are exerted with activities or exercises. FEVs values in stage 2 fall between 50 and 80 percent. The values of FEVs in stage 3 are between 30 and 50 percent and show signs of severe symptoms. In stage 4, the quality of life for COPD patients is reduced where the values of FEVs are less than 30 percent. When elderly patients reach stage 3 or stage 4, they are at high risk for exacerbations. These exacerbations have extremely negative impacts on the patients’ quality of life. In the same vein, patients with high level of exacerbations can impact the hospital sources caused by frequent admissions. Predicting what causes the exacerbations and making an effort to avoid them can decrease the amount of episodes and hospital readmissions. Monitoring COPD patient’s symptoms and collecting pertinent information can assist caregivers and healthcare professionals to take action at the right time. Patients in the advanced stages of COPD are susceptible to the acute exacerbation that can be fatal. Caregivers want a reliable mechanism that can help them manage COPD patients remotely and predict the risks of exacerbation before COPD episode is about to occur.

By designing a predictive analytics tool, exacerbations can be expected and hence the predictive method can allow early therapeutic interventions. The multi-layer artificial neural network (ANN) with backpropagation learning algorithm is often used for prediction process as it is easy to train the proposed model[3]. Early intervention provides a means for both patients and caregivers to stay better informed on the patient’s health status and to act appropriately. The input of the model is the symptoms and physiological information where the outputs of the model would feed the dashboard automatically. The dashboard shows the level of COPD exacerbations. The goal of this research is to answer the following research...
questions: What are the symptoms and vitals to be monitored for COPD patients in their homes to help prediction? How can COPD exacerbations in (stage3 and stage4) be predicted? What are the indicators needed for the intervention, and how are they prioritized?

In this paper, I discuss the design and evaluation of a feed forward multi-layer neural network for a prediction of COPD exacerbations. I used the backpropagation algorithm to train the ANN network. Further, I discussed the role of intervention that can be provided by designing a dashboard for the caregivers.

I generated the COPD data randomly in light of COPD Clinical Questionnaire scores (CCQ).

2. Related work

It is essential to develop effective tools and systems for caregivers so that they can provide much-needed assistance to the chronic obstructive pulmonary disease (COPD) patients [3]. Tools involve understanding the patient statuses and providing indicators that assist the decision-making process. Related work discusses the issue of prevalence of COPD exacerbations by delivering different analytical models. These models have been performed based on the hospitals' inputs and measurement values [4]–[6]. The contribution of these studies aims to detect the COPD risk by using a number of fundamental indicators of COPD such as dyspnea, chronic cough and sputum production [5]. Past studies address several prediction models as a means to improve patient awareness and ultimately improve the utility of health care system management by providing useful educational means.

While significant work has been carried out, there is still a chance for improvement [6] specially, for early prediction COPD exacerbation at-home. Recent research has investigated particular indicators that may lead to such exacerbation [4], [7], [8]. They are similar in some aspect to the proposed solution in this paper, but they lack critical component of integration by providing a holistic approach from which patients and caregivers would benefit or react. This holistic approach provides a dynamic solution that supports early detection of COPD exacerbation in advanced stages. The approach should combine the mechanism of prediction with a Dashboard in a dynamic way to enable a fast decision. The dashboards become an intelligent enabler in a Medical Decision Support System that is designed to assist the clinical decision [8]. Heart Failure analysis Dashboard has been designed and combined with a useful device for automatic set of patient's clinical inputs [9].

Machine learning classifiers are being applied to provide rigorous means for prediction chronic risk factors. Several classifiers have been applied to help the diagnostic of COPD using Normal Classifier, K nearest neighbor (KNN), decision trees, artificial neural networks (PNN), and support vector machines (SVM) [10], [11]. Backpropagation algorithm most popular classifier for performing an efficient supervised learning. The algorithm has not been used yet to tackle the issue of prediction the COPD exacerbations. Backpropagation uses intuitive and simple equations that make the algorithm easier to be adjusted and traced [12].

Dashboard proposed in this paper involves interactive and dynamic display that allows dynamic interaction. By combining the prediction algorithm with display, caregivers can clearly grasp the underlying information. The prediction system is considered grounding activities of a Dashboard [9] for the remote monitoring COPD patients.

3. Design

This paper follows design science research (DSR) method where the design includes both a product (artifact) and a process [13]. Three artifacts are designed and evaluated. The design processes for building these artifacts are established to attain and meet the objectives and the desired ends.

3.1. Data generation

The first artifact is a method that aims to generate the data randomly based on reliable and the valid Clinical COPD Questionnaire (CCQ). Process for building and evaluating the method of generating random COPD data is established to meet the objective by having reliable data, which eliminate the source of bias assignments.

3.1.1. Building process of generating random data

- Defining the COPD attributes that are needed for the study based on the CCQ questionnaire.
- Defining the data type, data range and mean for each attribute in order to identify the scope of the data.
- Generating random values with respect to the CCQ specification.
- Computing the CCQ scores in order to identify the exacerbation level of COPD. This will be used
as a desired output in the supervised learning algorithm.

3.2. Design the prediction algorithm

To predict the COPD exacerbations before they occur, I have designed a prediction algorithm as an artifact. This algorithm aims to predict the patient's risk of getting an exacerbation. The multi-layer Artificial Neural Network (ANN) algorithm is applied for prediction. The learning process of multi-layer neural network is conducted by using a supervised backpropagation learning algorithm. With the supervised learning, I used a simple data set to train the neural network about its problem domain with respect to the desired outputs [10]. Backpropagation algorithm is used to determine how the neural connection weights are adjusted based on the difference between the actual and desired outputs. As can be seen in the Figure 1, the data generated in the first stage for the COPD-diagnosed patients are used to train and test the multi-layer feed-forward neural network classifier in order to enable early prediction of exacerbations. The data consist of ten attributes. These attributes are classified under three main domains (Symptoms State, Functional State and Mental State).

- Creating multi-layer ANN by identifying the number of neurons in each layer. The number of neurons in the first layer correspond to the input attributes in the data set, which in this case, ten neurons. For the hidden layer, k-fold cross validation can be applied to define the optimal number of hidden layer neurons. The output layer consists of only one neuron.
- Determining the size of training set and testing set. In this study, we have five data sets generated randomly. The first data set is designated for training, the others are chosen for testing.
- Normalizing the desired output data (CCQ score). Since the backpropagation neural network uses the sigmoid function, which yields a value between 0 and 1, the desired output does not fall in this range. Thus, I found data normalization crucial to address the problem and fits the data between the sigmoid function output's ranges.
- Applying the backpropagation training algorithm in the first data set as it works using following procedures:
  1. Initialize weights with random values.
  2. Read the input values and the desired outputs.
  3. Working forward through the layers, the actual output is computed with respect to the input values and their weights assigned to them.
  4. The error (delta) is computed simply as the difference between actual and desired output. This procedure is conducted in the output layer. The errors are used to correct the weights for the connection to the previous layer (hidden layer).
  5. Change the weights by working backward from the output layer through the hidden layers.
  6. Testing the ANN using the remaining four data sets. Denormalizing the output data that will be displayed and demonstrated in the dashboard.

3.3. Design the dashboard

The third artifact is the dashboard (instantiation) that displays the risk predictions. Section 3.3.1 describes in detail how the dashboard is designed.

3.3.1. Building process for the dashboard

Process for building the dashboard as an instantiation is established to meet the objective of clear visualization of patients’ and provide decisions that
allow an early intervention. To keep abreast of dashboard standard, building the dashboard is established according to Pauwels et al [14] principles.

- Determining what kind of dashboard is suitable in term of caregivers’ specification of display. The display can be either static or dynamic. This process is intended not only for aesthetic display, but also to make it easy for caregivers to grasp the underlying information.

- Identifying the measures needed for the decision. Applying the Anthonisen’s criteria [15] of COPD can be applied to determine whether the exacerbation is caused by Symptoms, Functional State or Mental State. This process can be used if the dashboard involves interactive and dynamic display.

Several design principles and knowledge can be emerged from the process of designing the artifacts:

- Although data are simulated, the process of randomly generating the data can be adopted to be conducting in an academic research involving Machine Learning. Myriad number of research studies use simulations for the evaluation process. Yet, simulation is rarely used for mocking the data. That is, process of mocking the data using reasonable simulation methods could accelerate the activity of conducting research though the accuracy is questionable.

- The process of building a prediction algorithm from the Clinical COPD Questionnaires (CCQ) could be adopted for other chronic diseases. Most of the diseases can be diagnosed using valid and reliable surveys. Therefore, I am hoping that the prediction process that has been initiated in this research study would be used as a reference for health informatics research studies.

- Integrating the Dashboard with the prediction algorithm leads to utility as it enables an automatic interactive classifier that supports the early detection of COPD exacerbations and benefit both caregivers and patients.

4. Result

The results represent the testing process conducted in the ANN model and operated in the allocated testing data sets (from the second data set until the fourth data set). The trend lines in Figure 2 and Figure 3 show a positive correlation between the computed output line and the desired output line. CCQ scores represent both the desired outputs and computed output. The increase in value of CCQ score indicates high level of exacerbations that can trigger.

![Figure 2: Testing result –Data set 2](image)

![Figure 3: Testing result –Data set 4](image)

Five data sets have been generated randomly. In each data set, the generator method produced 52 records that correspond to the numbers of the weeks when a patient fills out the clinical COPD questionnaire. One data set is used for training and the others are used for testing. The backpropagation learning rule was applied in the designated training data set based on the gradient descent formula. The weights are initialized with random values, and then they were transformed in a direction that reduced the error. The learning process was iteratively repeated until the training reached the desired goal. The algorithm set the desired goal to 0.02 as the value reflected the minimum error value between computed output and the desired output.

5. Discussion

In this section, we discuss three important concepts that identify the legitimacy of this research. The intervention rules for the caregivers will be addressed. The process of evaluating the performance
of the backpropagation neural network predictor is discussed as well. Third, we will discuss the possibility of emerging the dynamic interactive predictors. The design process and the evaluation process are grounded and supported based on the prediction theory of classification. The theory of prediction is crucial in this study as it addresses the basic prediction system and its impact on classification success of implication and evaluation for learning algorithm design, and learning execution.

5.1. Prediction rules

This research aims to predict the level of COPD exacerbations and episodes before they trigger. Backpropagation classifier was applied to provide rigorous means for prediction COPD risk factors. The supervised backpropagation neural network predictor has been applied by training the existing data to predict the CCQ scores. The caregivers should benefit from the predictor and act upon the results. By applying Anthonisen criteria [15], caregiver can benefit from the predictor suggestions. Anthonisen criteria of the COPD exacerbations are classified based on the symptom state category. We included these criteria along with the functional state as the rules of creating and training the backpropagation neural network predictor.

5.2. Evaluation

This research follows design science research method guideline [16]. Design science is a part of information system research cycle that creates and evaluates IT artifacts intended to solve identified organizational problems. Thus, the evaluation stage is the central stage in this research. The performance for the classifier can be evaluated by using several methods. Since the backpropagation is considered a supervised learning algorithm, we can apply accuracy analysis and recall analysis. In this research, our goal is to predict the exacerbation level based on CCQ questionnaire. We considered the CCQ scores as a desired output for the neural network algorithm. That is, the CCQ score attribute is assigned to be a continuous variable. As a result, neither confusion matrix nor recall analysis can be effective for such evolution. We decided to evaluate the predictor using mean square error (MSE). The MSE is an efficient criterion that can be used to evaluate the performance of a predictor or an estimator. It is basically computed based on the variance of forecast error. As can be seen in the Table 1, the values of MSEs are significant. The MSEs have been computed on three the assigned testing data sets.

<table>
<thead>
<tr>
<th>FILENAME</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TESTDATASET2</td>
<td>0.1312</td>
</tr>
<tr>
<td>TESTDATASET4</td>
<td>0.1380</td>
</tr>
<tr>
<td>TESTDATASET5</td>
<td>0.1041</td>
</tr>
</tbody>
</table>

Table 1: MSE values to evaluate the performance of the predictor (BK-ANN)

5.3. Dynamic interactive predictors

The predictor algorithm was designed in a flexible way, as it allows predicting different variables. The algorithm helps the caregiver to select an appropriate attribute as the desired output. In another words, different ANN models can provide different suggestions for predicting COPD exacerbation. The rules of the prediction may change based on the context of geographic locations where the COPD symptoms vary. For example, a caregiver in the USA may be interested in Anthonisen criteria to predict the exacerbation whereas a caregiver in another country may be interested in different criteria or formula. The flexibility of our algorithm demonstrates a good practical implication.

6. Conclusion

I designed and built a mechanism to monitor COPD patients remotely and predict the risks of COPD exacerbation in advance. I introduced a solution of improvement that can enhance the predictor with constructing the dynamic interactive predictors. The ANN backpropagation predictor algorithm was designed in a flexible way, as it endorsed predicting different variables. Designing the predictor was guided by Anthonisen’s criteria. Applying the Anthonisen’s criteria of COPD was applied to determine whether the exacerbation is caused by Symptoms State, Functional State or
Mental State. The design process and the evaluation process are grounded and supported based on the prediction theory of classification. The theory of prediction is crucial in this study as it addresses the basic prediction system and its impact on classification success of implication and evaluation for learning algorithm design, and learning execution. The method of generating COPD symptoms data is directed by the theory of random processes, which describes a group of random variables. The method aims to generate the data randomly in light of reliable and the valid Clinical COPD questionnaire (CCQ). Five data sets have been generated randomly. In each data set, the generator method produced 52 records that corresponded to the numbers of the weeks when a patient fills out the clinical COPD questionnaire. One data set is used for training and the others are used for testing. I used the backpropagation algorithm to train the ANN network. The algorithm introduced the improvement of prediction COPD exacerbations. Backpropagation uses intuitive and simple equations that make the algorithm easier to adjust and trace. The results show a positive correlation between the computed output and the desired output.

During the process of conducting this study, I faced the following challenges:

- The lack of access to the actual COPD data decreased the chance of getting contextual insights of symptoms and their causes. This is deemed to be the main hurdle of conducting this research. Mobile application can be designed as a tool to facilitate the process of data gathering. Hence, it provides an opportunity for researcher to build upon this work to integrate and add an Alert mechanism system with the predictive model.

- As it is the case with most of neural network algorithms, the backpropagation works only with numerical data. The categorical type must be converted to numerical data. Although data normalization can solve the issue, the output values are rarely computed to produce the exact same desired values.

7. References


