



Designing Adaptive Mechanism for COVID-19 and Exacerbation in Cases of COPD Patients Using Machine Learning Approaches

Konan-Marcelin Kouamé*, Hamid Mcheick

Computer Science and Mathematics Department, University of Québec, Chicoutimi, Canada

Email address:

konan-marcelin.kouame1@uqac.ca (Konan-Marcelin M.), Hamid_Mcheick@uqac.ca (Hamid M.)

*Corresponding author

To cite this article:

Konan-Marcelin Kouamé, Hamid Mcheick. Designing Adaptive Mechanism for COVID-19 and Exacerbation in Cases of COPD Patients Using Machine Learning Approaches. *International Journal of Intelligent Information Systems*. Vol. 10, No. 5, 2021, pp. 81-97.

doi: 10.11648/j.ijis.20211005.11

Received: September 24, 2021; **Accepted:** October 21, 2021; **Published:** October 30, 2021

Abstract: The technology of machine learning has been widely applied in several domains and complex medical problems, specifically in chronic obstructive pulmonary disease (COPD). Researchers in the field of respiratory diseases confirm that people who suffer from COPD have high risks when exposed to COVID-19. The most common oncoming COPD exacerbations and COPD symptoms of COVID-19 are congruent. The distinction between COPD exacerbations and COVID-19 with COPD is nearly impossible without testing. This paper proposes a new powerful model for classifying COPD patients with exacerbations and those with COVID-19 using machine learning and deep learning algorithms. The major contribution of this research is the dynamic classification process based on the patient context that can help detect exacerbations or COVID-19 per period. Indeed, Five Machine Learning algorithms are trained, tested and a performant classification model is identified. This prediction model is then associated with a dynamic COPD patient context for monitoring the patient's health status. This model based on the dynamic adaptation mechanism combined with a classification contributes to identifying dynamically COPD exacerbations and COVID-19 symptoms for COPD patients. Indeed, periodically, data on a new patient is injected into the prediction model. At the output of the model, the patient is either classified in the exacerbation category, or classified in the COVID-19 category, or no category. By period. A dynamic dashboard of classified patients is available to help medical staff take appropriate decisions. This approach helps to follow the evolution of COPD patient comorbidities (exacerbation, COVID-19). Finally, classification would allow healthcare stakeholders to provide healthcare service according to the patient's status. The methodology of research consists of designing and implementing a dynamic model for classifying COPD patients. Since early intervention is associated with improved prognosis, with our solution, healthcare staff can identify COPD patients who are most at risk of developing exacerbation or COVID-19. Consequently, upon admission, this will ensure that these patients receive appropriate care as soon as possible.

Keywords: Software Adaptation Mechanism, Deep Learning, Exacerbation, COPD, COVID-19, Prediction

1. Introduction

Following the annual World Health Organization (WHO) statistics, 63% of deaths are attributed to chronic diseases [1]. Chronic obstructive pulmonary disease (COPD) is expected to become the third leading cause of death worldwide by 2030 [1]. The treatment of chronic diseases, such as cancer, COPD, asthma, arthritis, and diabetes, costs over 53 billion euros per year [1]. According to WHO, COPD is a combination of two

diseases – emphysema and chronic bronchitis – which cause breathing difficulties and can occur alone or together [2]. Unfortunately, COPD respiratory problems cannot be cured. Most patients experience a wide variety of symptoms and exacerbation attacks during their lives. COPD patients are at risk of developing onsets of complications or exacerbations. Therefore, the follow-up care of patients is essential to extend their lifespans.

The deadly novel coronavirus disease 2019 (COVID-19)

pandemic imposes close monitoring of people suffering from COPD since they are more vulnerable. However, COVID-19 is not always the cause of death in COPD patients, but it can introduce exacerbations. Some research projects related to the treatment and the prevention of exacerbations in COPD patients have been carried out [3]. Steill [2] indicated that the number of exacerbations in COPD patients at discharge is high, which leads to a high mortality rate in COPD patients. Researchers and specialists in the field of respiratory diseases have confirmed that people who suffer from COPD are at high risk when exposed to COVID-19 [4-10]. The signs of exacerbations and symptoms of COVID-19 are similar, and differences between COPD exacerbation and COVID-19 are not evident [5-9]. Therefore, classification mechanisms have been proposed to classify patients.

The novelty of this research is the dynamic approach based on mechanism adaptation in the process of classification that uses machine learning. In addition, a platform based on mechanism adaptation to interpret classification results dynamically is also proposed. Compared to other classification approaches, real-time adaptation is prominent in the proposed approach, offering different services depending on the classification results. This paper aims to propose these classification mechanisms to detect COVID-19 earlier to protect patients, reduce complications and help manage COPD patients anywhere and anytime.

The architecture to implement dynamic COPD is also a contribution of this research, which involves the combination of an adaptation model and dynamic COPD patient context with the classification process based on machine learning, deep learning model to follow the evolution of COPD patient comorbidities (Exacerbation, COVID-19). This paper is organized as follows. Section 2 presents the related works; Section 3 provides methods and tools; Section 4 introduces the COPD patient classification architecture model in predictive systems. Section 5 presents the experimental results and discussion. The conclusions and future work recommendations are given in Section 6.

2. Related Work

Deep learning (DL) is a subfield of machine learning (ML) based on multiple levels of learning through a hierarchy of features [11]. Researchers have increasingly adopted ML and DL in several areas, especially in the medical field for medical image analysis [11] and the classification of patient diseases.

In medical imaging, DL has enabled the researcher to produce quality results. Humphries et al. [12] proposed a solution based on DL to enable the automatic classification of emphysema patterns at CT. A DL algorithm using a convolutional NN and long short-term memory architecture was trained to classify the pattern of emphysema according to Fleischner criteria in a retrospective analysis of the genetic epidemiology of COPD (COPDGene) study [13]. In their research, Mohsen et al. [14] introduced an approach with DL using a convolutional NN. They used a classifier, which is a DL architecture, to divide a dataset of 66 brain MRIs into four

classes – normal, glioblastoma, sarcoma, and metastatic bronchogenic carcinoma tumors. These classification projects based on medical imaging did not use a dynamic classification approach and reasoning system.

Moreover, the diagnostic disease category includes many research projects that have used DL and ML approaches. Altan et al. [15] compared multiple machine-learning algorithms for the early diagnosis of COPD using multichannel lung sounds. Their study focused on analyzing multichannel lung sounds using statistical features of frequency modulations that were extracted using the Hilbert-Huang transform. Wang et al. [18] proposed a method for diagnosing COPD based on transfer learning called the balanced probability distribution (BPD) algorithm, which integrates instance- and feature-based transfers to improve the prediction accuracy of the model.

Esteban et al. [44] implemented an early warning system based on ML capabilities to predict an exacerbation but there is no approach based on a dynamic COPD context patient. In addition, Trappenburg et al. [45] also proposed a system based on ML algorithm to detect exacerbation using a dataset composed of a clinical COPD questionnaire and achieved a moderate detective capacity with $AUC = 75\%$. Similarly, Bertens et al. [46] identify patients at risk of developing severe COPD exacerbations.

Concerning the detection of diseases, many projects have been realized. Li et al. [20] developed a fully automatic framework to detect COVID-19 using chest CT and evaluate its performance. Liang et al. [21] showed that a DL-based survival model could predict the risk of COVID-19 patients developing critical illnesses based on clinical characteristics at admission. Alotaibi et al. [35] applied and evaluated the effectiveness of artificial NNs, support vector machines, and random forest regression using a variety of learning methods for early the prediction of severity using patient history and laboratory findings. Rhee et al. [36] also proposed a study that aimed to develop a prediction model of COPD acute exacerbation with big data using machine learning methods. In the literature, classification projects have addressed the problem of an *imbalanced* dataset. Indeed, Brownlee [40] presented two projects of imbalanced datasets: *mammography and Haberman Breast Cancer Classification*. In the article, Brownlee presented an approach based on classification with machine learning to detect cancer, explaining that detection is a popular example of an imbalanced classification problem because significantly more cases of non-cancer occur than actual cancer.

The goal of microcalcification classification research is to distinguish between microcalcifications and non-microcalcifications using the features for a given segmented object. Non-microcalcifications make up the negative case or majority class; microcalcifications are the positive case, or minority class. The author applied the best practices for imbalanced dataset projects, specifically for the mammography dataset. However, compared to our research criteria in the dynamic context of the patient, an adaptation model capable of dynamically monitoring the evolution of cancer did not improve. In addition, according to Brownlee's

[40] project “Haberman Breast Cancer Classification,” developing a probabilistic model is challenging in general, especially with a skew in the distribution of cases, referred to as an imbalanced dataset. The measure of this research was whether the patient survived for five years or longer. The Haberman dataset is standard for the study of imbalanced classification. Brownlee [40] explains in the dataset description that the breast cancer surgeries were conducted between 1958 and 1970 at the University of Chicago’s Billings Hospital. The dataset comprises 306 examples with three input variables: the age of the patient at the time of the operation, the two-digit year of the operation, the number of positive axillary nodes detected, which is a measure of cancer that has spread. Compared to our research within the dynamic context, an adaptation model was absent.

Finally, other authors addressed classification issues on COVID-19 patients and explained that the sudden deterioration of patients with COVID-19 into critical illness is a major concern [41]. The solution was to identify these patients early. A DL-based survival model could achieve this by predicting the risk of COVID-19 patients developing critical illness based on clinical characteristics at admission. They developed this model using a cohort of 1590 patients from 575 medical centers, with an internal validation performance of concordance index of 0.89. Compared to our criteria, this project had limits, including the lack of patient context and model adaptation.

Table 1 makes a comparison of five main existing classification research projects with our research objectives or criteria.

Table 1. Limits of existing classification research projects.

Classification projects	Research criteria		
	Dynamic patient context	Adaptation model	Classification process with ML and DL algorithms
COPDGene [13]	NO	NO	YES
Mammography [40]	NO	NO	YES
Haberman Breast Cancer Classification solution [40]	NO	NO	YES
Early triage of critically ill COVID-19 patients [41]	NO	NO	YES
Detect COVID-19 using chest CT [21]	NO	NO	YES

3. Methods and Tools

3.1. Research Approach

The approach of the research project is structured in three

steps: i) Determination of the best performing classification model; ii) Association of the COPD patient context with the prediction model and iii) Adaptation mechanism for monitoring the patient's health status. Figure 1 describes these steps.

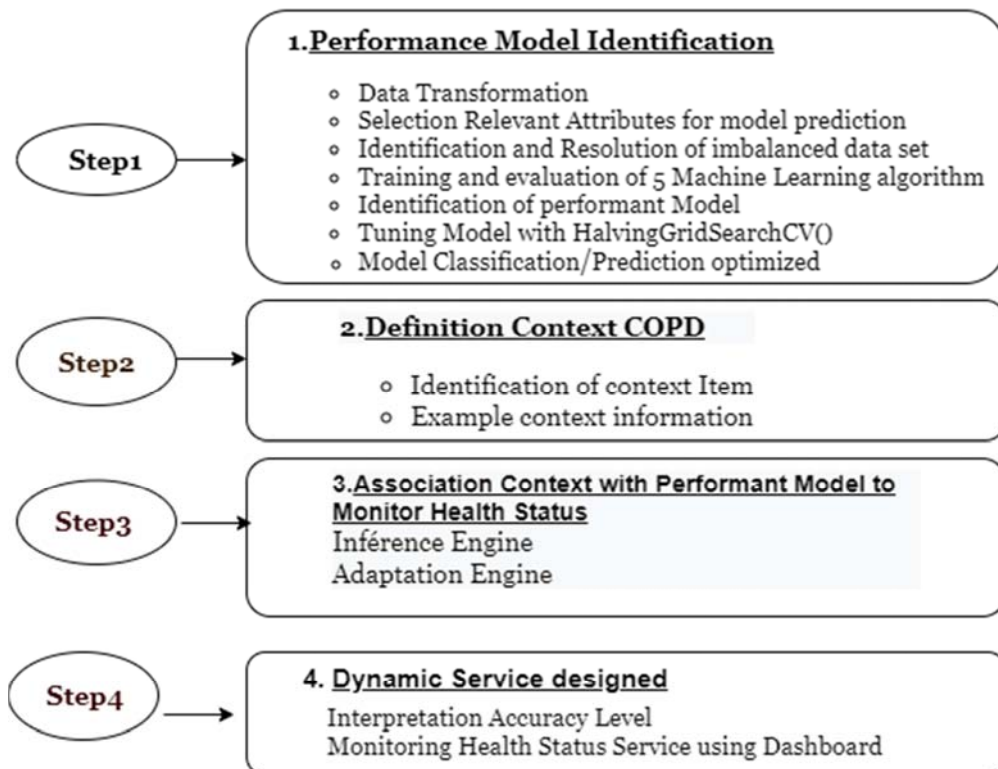


Figure 1. Steps of Research.

3.2. Tools and Classification Process

During this research, the Python and Google Colab framework tools (TensorFlow, Kera, panda, ...) were used to implement the steps of this project classification. The Google Colab package was our platform. Imbalanced-Learn and DL Libraries are also used. Google Collab is used for four reasons:

- i. Google Collab with Python programming language is widely used for machine learning research projects.

- ii. Google Collab is smoothly integrated with Python.
- iii. The powerful and faster computation with Python facilitates the estimation of machine learning models.
- iv. Google Collab contains all the necessary libraries for programming and evaluating the models.

Table 2 presents a description of tools used per classification step.

Table 2. Tools per classification step.

Steps	Activities	Tools/Libraries
01	Package installation	pip install scikit-learn==0.24.1; install tf-nightly; install delayed;
02	Initialization: importation library step	Numpy; tensorflow; pandas; matplotlib.pyplot; seaborn; pylab; maploplib; google.colab; xgboost;
03	Data processing: data loading step	read_excel;
04	Feature selection: recursive feature elimination, mutual information gain	sklearn.svm;
05	Setting step: variable settings, functions for creating the dataset for training the model, split data in training and testing; create dataset	feature_column; create_dataset (data frame, batch_size=32); data frame; dataframe.copy; create_dataset.
06	Parameter tuning algorithms and cross-validation	sklearn.model_selection; sklearn.experimental; metrics
07	Evaluation model	Eight Metrics
08	Prediction model for new data	

3.3. Classification Process

The first step in this research is to identify the best-performing classification model. To achieve this, identify the models to be evaluated in the first activity. Many classification algorithms exist, but how to choose the best algorithm for the classification process is an issue. To choose

the algorithms, three criteria are proposed: 1. widely used, 2. easy to implement, 3. compatible with unbalanced data. In this research, five algorithms to perform the classification of COPD patients are implemented. The algorithms were evaluated with these three criteria and the result was the identification of five algorithms described in table 3 below.

Table 3. Algorithms used for the classification model.

No	Algorithms	Description
1	Gradient boosting machine (GBM)	Machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees [24]
3	Support vector machine (SVM)	Set of supervised learning techniques designed to solve discrimination and regression problems., generalizations of linear classifiers [25]
4	Random forest (RF)	A technique that performs learning on multiple decision trees trained on slightly different data subsets [26]
5	K-nearest neighbors (KNN)	Nonparametric method used for classification and regression [27]
6	Linear discriminant analysis (LDA)	A method that involves explaining and predicting an individual's membership in a predefined class (group) based on characteristics measured using predictive variables [28]

The classification process used during this research involved five steps: i) exploration of the dataset, ii) assessment of the imbalance of the dataset, iii) execution of the taxonomy of metrics and evaluation model, iv) prediction of new data. The purpose of this process is to develop classifiers to detect the exacerbation and cases of COVID-19 from a COPD patient database. Exploring the data is crucial to obtain a summary of its distribution, which may identify imbalanced dataset intuition problems.

3.3.1. Data Processing Step

Generally, the different steps to implement classic classifications are data transformation, feature selection, model implementation with a machine learning algorithm, metrics definition, choosing the best model, and prediction

with new data. A similar approach is used but taking into account the imbalanced dataset problem through the choice of the metrics. The purpose of this model was to develop classifiers to detect, on one hand, the exacerbation of COPD and on the other, the exacerbation of COPD and COVID 19. Two databases are used during process classification: a COPD database of 2500 COPD patients with or without exacerbation and a COVID-19 database containing symptoms of patients suffering from COVID-19 provided by COVID-19 dataset clearinghouse a repository for public data sets relating to the COVID-19 pandemic. Since one of our goals was to develop a classifier algorithm that could detect the exacerbation and cases of COVID-19. An combine approach based on the two databases using the propensity score matching method is implemented. The matching was realized based on the

non-symptoms common variables in the two databases: sex, age, and weight. After combining, a single database of 5500 patients suffering from either COVID-19 or COPD exacerbation, or none of both is obtained. his dataset to develop supervised learning algorithms capable of detecting exacerbation and COVID-19 is built. In the first step of data processing, our datasets were cleaned to account for missing values. ifferent algorithms are evaluated and the MissForest algorithm is choose [29].

3.3.2. Relevant Feature Selection Methods

Feature Selection (FS) is one of the core concepts

applicable in machine learning that hugely impacts the performance of a prediction model. The data features used to train a machine learning model have a significant influence on the performance of the model. Irrelevant or partially relevant features can negatively impact model performance. However, by following feature selection techniques, it is possible to automatically select those features, which contribute most to the prediction label. There are two main categories of feature selection methods: Wrapper methods and filter methods. Table 4 below presents some of the Relevant Feature Selection Methods.

Table 4. Feature Selection Methods.

Feature Selection Method	Description
1. Recursive Feature Selection (RFE)	This method selects features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features and the importance of each feature is obtained either through any specific attribute or callable. Then, the least important features are pruned from the current set of features [43]
2. Forward Feature Selection (FFS)	It selects an attribute to include in the prediction and estimates the new performance after the addition of this attribute using cross-validation. If the performance increases, the attribute is added to the selected feature set, else it is removed.
3. Backward Feature Elimination (BFE)	It takes the full set of attributes as input and, in each round; removes an attribute from the set. For each removed attribute, the performance is estimated
4. Weight by Information Gain	This method calculates the weight of attributes concerning the target attribute by using the information gain. The higher the weight of an attribute, the more relevant it is considered
5. Weight by Information Gain Ratio	This method will assign higher weights to the attributes that have more information about the label class and lower weights to the attributes that have less information.
6. Weight by Uncertainty	the weight of each attribute is estimated based on symmetrical uncertainty between this attribute and the label attribute.
7. Weight by Chi-Square Statistic	This method is usually used to evaluate the likelihood of correlation or association between variables using their frequency distribution.

Indeed, to select the best explanatory variables to predict the target variable for the model, recursive feature selection (RFE) algorithm [31] among other filter methods is implemented because our dataset contained categorical, quantitative variables and widely implemented with '*class sklearn.feature_selection.RFE (estimator, *, n_features_to_select=None, step=1, verbose=0, importance_getter='auto')*' [29] in python environment.

3.3.3. Datasets

i. Dataset exacerbations

To detect the exacerbation, a training phase of different machine learning algorithms on our exacerbation data is executed. The exacerbation database is a file composed of 2500 COPD patients. These attributes are structured in three categories. Attributes related to patient's medical profile, medical history, and vital signs and symptoms parameters. The sample size of COPD patients was 2500. Since there is a small number of variables, it is decided to keep all the variables even though the feature selection algorithm ranks them according to importance.

To detect the exacerbation and COVID-19, a training of the same machine learning algorithms on our combined dataset, which was obtained through propensity score matching between the COVID-19 and COPD with exacerbation datasets is essential. Indeed, patients who were in the common support of the matching in the analysis are kept. The feature selection led us to keep all variables in the dataset.

ii. Imbalanced Problem and Taxonomy of Classifier

Evaluation Metrics

Imbalanced classification is a classification predictive modeling problem where the distribution of examples across the classes is not equal [40]. According to Brownlee [40], imbalanced classification is also defined by a dataset with a skewed class distribution. The distribution may range in severity – 1:2, 1:10, 1:100, or even 1:1000 – as explained by [40].

After exploring the dataset, it may be helpful to know how imbalanced the dataset actually is. The class distribution in the combined COVID-Exacerbation dataset was not balanced, as shown in table 5 in appendix section (54% with no COVID or Exacerbation, 30% with exacerbation and 16% with COVID-19). Therefore, the choice of evaluation metrics wisely to avoid bias is important.

Taxonomy metrics help us to understand the different group of metrics, importance, and criteria to make the best choice for evaluating the candidate Model. Indeed, the choice of the wrong metric to evaluate our models, can generate a poor model, or in the worst case, be misled about the expected performance of your model [40].

For imbalance, metrics like classification accuracy lose their meanings, and alternate methods for evaluating predictions on imbalanced examples are required to solve this foundational challenge. Thus, for evaluating predictions of our models on COPD with the Exacerbation dataset, the list of the metrics described in Table 5 are used. The choice of these eight metrics is realized because they are not biased by the '*Imbalanced dataset problem*' and widely used.

Table 5. Metrics for binary classification.

Abbreviation	Metric	Definition*
SNS	Sensitivity	$\frac{TP}{TP+TN}$
SPC	Specificity	$\frac{TN}{TN+FP}$
PRC	Precision	$\frac{TP}{TP+FP}$
NPV	Negative predictive value or recall	$\frac{TN}{TN+FN}$
ACC	Accuracy	$\frac{TP+TN+FP+FN}{TP+TN+FP+FN}$
F1	F1 measure	$2 * \frac{PRC * SNS}{PRC + SNS}$
GM	Geometric mean	$\sqrt{SNS * SPC}$
MCC	Matthews correlation coefficient	$\frac{TP+TN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$

* TP is true positive; TN is true negative; FP is false positive; FN is false negative; PP predicting positive; PN predicting negative.

Table 5 presents metrics for binary classification.

The evaluation of the prediction of our models on the combined COVID-19 and COPD dataset needed metrics that take into account the imbalanced background problem. Following the Truica and Leordeanu [39] approach, the weights of each class using the weighted versions of accuracy (WA), precision (WP), and recall (WR) are implemented. The weighted version of these measures calculates metrics for each class, computes their average, weighted by the number of true instances for class (Table 6).

Table 6. Metrics for multi-class classification.

Symbol	Metrics	Defined As
WA	Weighted Accuracy	$\frac{1}{n} \sum_{i=1}^3 n_i * \frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}$
WP	Weighted Precision	$\frac{1}{n} \sum_{i=1}^3 n_i * \frac{TP_i}{TP_i + FP_i}$
WR	Weighted Recall	$\frac{1}{n} \sum_{i=1}^3 n_i * \frac{TP_i}{TP_i + FN_i}$

Candidate models using repeated stratified *k-fold cross-validation* is evaluated. Stratification activity of the splits using the *StratifiedKFold* class that supports stratified k-fold cross-validation is realized. The objective of this approach is to provide the best general estimate of model performance that is not too optimistically biased, at least

compared to a single train-test split. Repetition value $k = 50$ to best estimate confidence intervals for the mean value of the metrics is proposed.

3.3.4. Modelling, Classification, and Prediction for New Data

i. Modelling, Classification

A combined dataset to develop supervised learning algorithms capable of detecting exacerbation and COVID-19 is implemented. These models used the following classification algorithms: GBM [34], SVM [25], KNN [27] and LDA [28]. ML models have been chosen because DL models are more efficient in the presence of a large amount of data. They are only 2550 and 5500 patients in our datasets, too small in this case to apply Deep Learning models. The tuning parameters were obtained by the Halving Grid Search algorithm (Table 7).

ii. Prediction for New Data and tuning parameter

Evaluation of the performance of the models allowed the identification of the best model. The characteristics of the optimization parameters of these prediction models for the new COPD patient data are presented in the table below. Table 7 presents examples of characteristics of the optimization parameters for three ML algorithms:

Table 7. Tuning parameters from (Halving) Grid Search Algorithm.

Model	Exacerbation Predictions	Exacerbation COVID 19 Predictions
GBM	learning_rate=0.01, n=1000	learning_rate=0.005, n = 1000
SVM	C=0.1, gamma = 0.001, kernel='poly'	C=1, gamma = 0.1, kernel='poly'
Random Forest	criterion='entropy', max_depth=8, max_features='log2', n_estimators=200	criterion='entropy', max_depth=5, max_features='sqrt', n_estimators=200
KNN*	algorithm='auto', leaf_size=30, metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=65, p=2, weights='distance'	algorithm='brute', leaf_size=30, metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=47, p=2, weights='distance'

* search for tuning parameters for GBM, SVM, Random Forest has been realized with Halving Grid Search Algorithm and KNN with Grid Search Algorithm.

4. Architecture Model in Predictive Systems Proposal

4.1. COPD Patients Classification Architecture

The Architecture Model in the predictive system proposal is

based on the telemedicine approach where COPD patient data was collected from their residences and hospitalization rooms. Indeed, this architecture was composed of three components: data acquisition, classification, and results and presents the different components of the classification process, along with the input data and output data of this classification. Healthcare staff performed an investigation and entered data into Google

forms [4]. In the previous approach, vital sign data were collected with a sensor installed into the patient body, and profile patient data were collected from the patient medical records, and symptom data were collected from a questionnaire distributed with Google forms.

The data were periodically collected from the three data sources: patient profiles, questionnaires, and sensors. These data were saved in a real-time database that was to be used for the prediction. A part of this database would be used to train the prediction models, and another part would be used for validation tests. For each period of data of the patient context, the model was trained and then validated by the test model. The main goal was to demonstrate the dynamicity of our

classification based on patient context information.

Figure 1 presents the classification architecture model. With this architecture, the evolution of COPD patient health state could be followed. For example, at the period $T_0 = 01/05/2021$, COPD patient 1 had exacerbation but no symptoms of COVID-19 according to the classification process. After two weeks, at period $T_1 = 15/05/2021$, a classification process was also performed, and the result was that patient had COVID-19 but no exacerbation. Therefore, the dynamic context of COPD patients combined with the classification process, as proposed by this architecture, allowed us to follow COPD patients' health state. Figure 2 corresponds to an architecture having three main components.

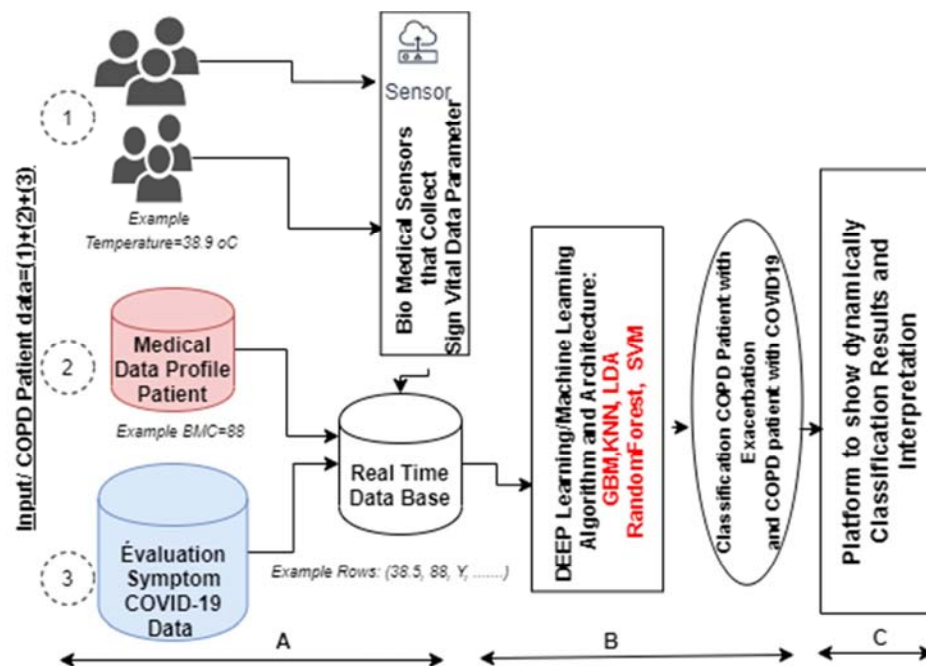


Figure 2. Classification architecture model.

This architecture has three main components: Context and data sources (Input) is component A; classification process is component B; the platform to show the results of the classification (Output) is component C. Figure 3 presents the HealthCare Interface to follow the Distinction System COPD patient with exacerbation and COPD patient with COVID-19 symptoms.

HealthCare Interface to monitor Dynamically Distinction System COPD patient with exacerbation and COPD patient with COVID-19 Symptoms	
1. Sample Information	
Start:	End:
Sample Size:	
2. Prediction Information	
New COPD Patient Num:	
COVID :	
EXACERBATION:	Accuracy:
Interpretation Service:	
3. Relevant Algorithm Information	

Copy Right KKM Reserved 2021 PhD Research

Figure 3. HealthCare Interface to monitor Distinction System COPD patients with exacerbation and COPD patients with COVID-19 symptoms.

The objective of this interface is to dynamically interpret the classification results for physicians. Thus, the medical staff does not necessarily need to be familiar with the models.

4.1.1. Context and Data Sources (Input)

All data collected from three data sources: Sensor data source, profile, and investigation data source, were merged into a unique real-time database. Sensors are installed on COPD patient Body. The real-time database was populated dynamically from these three sources (Sensor data source, profile, and investigation data source) per period collected.

After the results of the classification process were obtained, the result was saved in a data object, which provides information in a database. The interface to monitor Distinction System COPD patients with exacerbation and COPD patients with COVID-19 symptoms displayed three pieces of information: 1) sample information, 2) prediction information and 3) relevant algorithm information. Figure 4 presents the database for distinction COPD patients with exacerbation and those with COVID-19 symptoms.

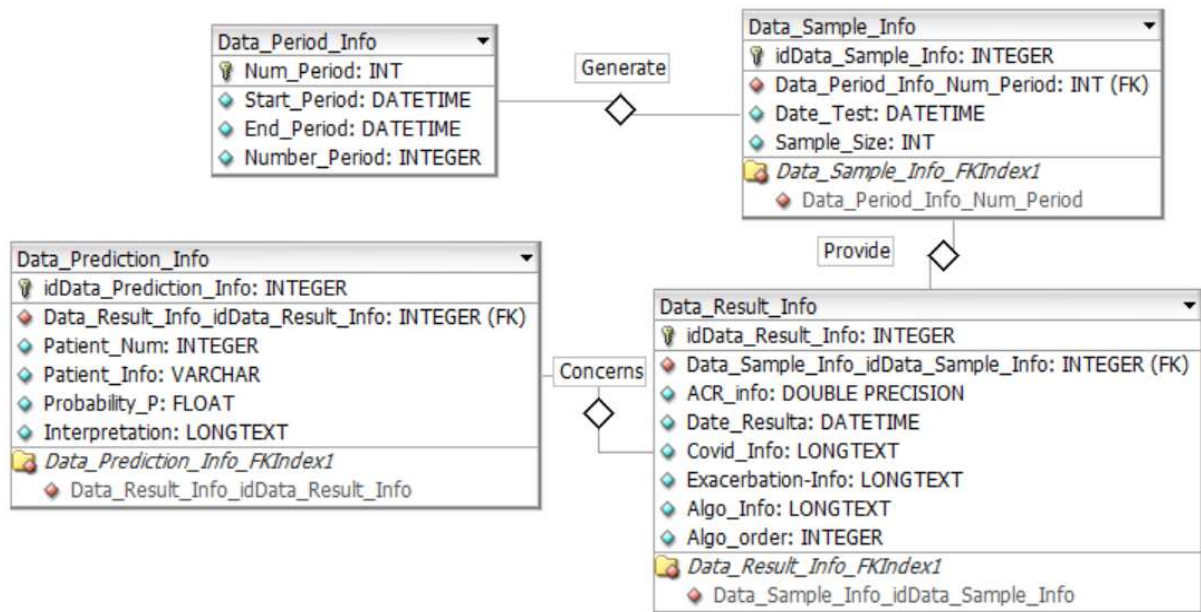


Figure 4. Database for interpretation system.

4.1.2. Classification Process

The data transformation processes are: i) data cleaning to delete duplicate rows or redundant columns, ii) outlier detection and removal, iii) missing value identification and imputation, iv) feature selection with statistics and models, and v) estimation and classification reporting.

4.1.3. Results of Classification (Output)

After the classification process, the result allowed us to determine if a COPD patient had an exacerbation or COVID-19. Therefore, for a new patient, the classification model should propose a patient classification with accuracy. Then, an application would interpret this classification dynamically according to the rate proposed by healthcare staff. The objective of this interpretation was to help the physician better understand DL and machine learning technology. With this interpretation, the physician could focus only on care. Concretely, our solution could help health staff in the following ways:

1. Possibility to follow COPD patient health state evolution with dynamic context.
2. Advanced knowledge is not needed to understand classification results thanks to a platform interpreting the

results.

3. Alert service proposed on mechanism adaptation if exacerbation or COVID-19 percentage was high, middle, or minor.
4. The dashboard integrated into the interpretation interface allowed for remote patient monitoring.

Based on this information, the system identification facilitated medical assistance. Indeed, healthcare staff could easily provide medical assistance.

4.2. Adaptation Mechanism Delivering and Service Accuracy Interpretation (Range)

Many classification projects exist, but our approach classification is dynamic. Unlike the concept of the “black box” in machine learning, our model was explainable. The interpreted the classification results component is based on an adaptation system. Therefore, a physician did not need to understand the classification model. In different periods, data from different sources were merged in real-time databases, like Datawarehouse. Then, a program based on Python extracted information for training in the database and test database. Metric SN (sensitivity) is chosen for platform interpretation because the literature review found that SN is

more effective for imbalanced dataset problems. Finally, in each period, the adaptation mechanism delivered service interpretation depending on current context information (range).

The main components of the adaptation mechanism of Figure 4 are the following: context information, including patients' information, sensors, classification results and accuracy (SN); reasoning (adaptation and inference engine); services (service accuracy interpretation offered by the application of context-aware tools); and rule-based information (rule to interpret accuracy for healthcare staff).

Accuracy is considered as context information of our

dynamic classification system. Indeed, at each period $T0$, the system extracted data collected from the three data sources, performed classification with an accuracy of $SN1 = \text{Range1}$. For another period $2T0$, the system extracted data collected from the three sources and realized a classification $SN2 = \text{Range2}$, and so on. The reasoning component is the engine adaptation program that proposes services interpretation depending on current context data (SN). The reasoning technique is a rule-based approach. Figure 5 presents this adaptation mechanism to interpret the classification rate. The adaptation mechanism could help medical staff easily understand the classification results.

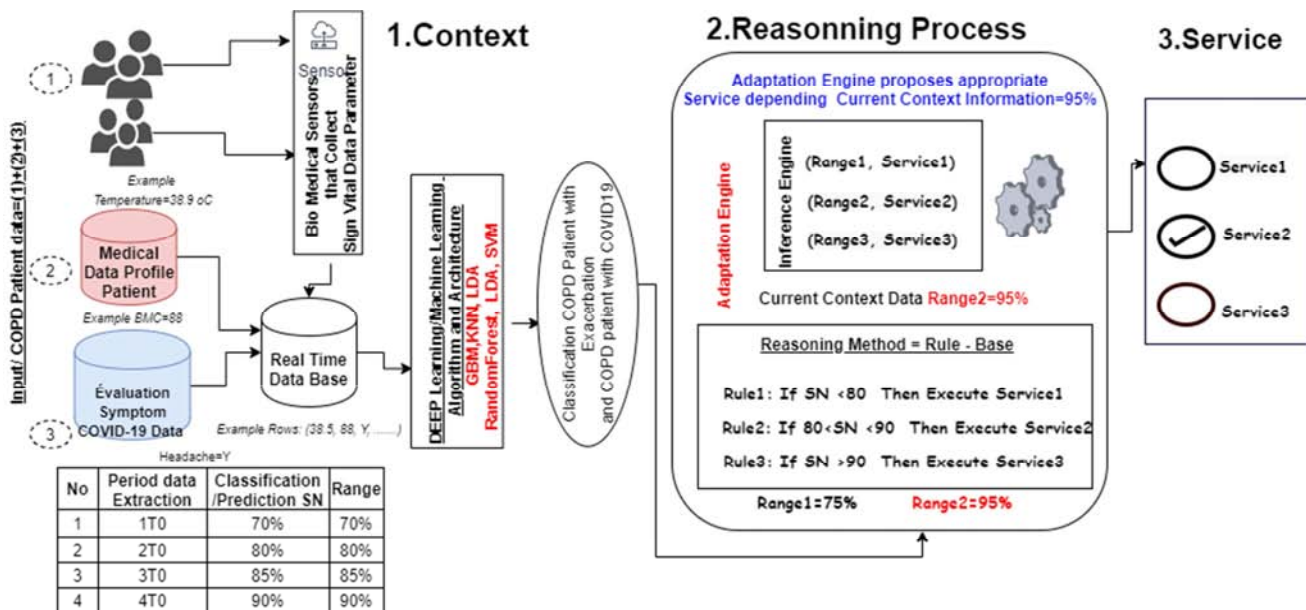


Figure 5. Adaptation mechanism to interpret the classification rate.

5. Experimental Results and Discussion

The experimental phase addressed three aspects: i) identification of COPD patients with exacerbation and COPD patients without exacerbation from the exacerbation database, ii) identification of COPD patients with exacerbation and COPD patients with COVID-19 from the COVID-19 exacerbation database and iii) design of an adaptation model that could consider the dynamic COPD patient context to dynamically distinguish COPD patients with exacerbation and COPD patients with COVID-19 based on the previously defined prediction model.

5.1. Performance of Exacerbations Prediction Model

Concerning the first approach of this research, to detect individuals suffering from exacerbation, previous models based on a dataset composed of COPD patients suffering from an exacerbation, and normal COPD patients without exacerbation are trained. Table 8 presents the performance of the different models on the test data: KNN; LDA; GBM; RF and SVM classifiers. Figure 6 presents a model performance comparison. The Information in Table 8 was from the results of training model activity on dataset exacerbation.

Table 8. Performance of exacerbation prediction models.

	SNS	SPC	PRC	NPV	ACC	F1	GM	MCC
GBM	0.207 ± 0.068	0.767 ± 0.066	0.813 ± 0.124	0.5 ± 0.139	0.8 ± 0.064	0.33 ± 0.09	0.414 ± 0.044	0.086 ± 0
SVM	0.243 ± 0.072	0.731 ± 0.07	0.648 ± 0.118	0.552 ± 0.133	0.753 ± 0.058	0.354 ± 0.085	0.434 ± 0.037	0.073 ± 0
RF	0.271 ± 0.066	0.706 ± 0.064	0.748 ± 0.101	0.665 ± 0.128	0.814 ± 0.064	0.398 ± 0.078	0.448 ± 0.033	0.078 ± 0
KNN	0.106 ± 0.003	0.962 ± 0.119	0.606 ± 0.32	0.668 ± 0.092	0.662 ± 0.264	0.176 ± 0.005	0.322 ± 0.181	0.111 ± 0.025
LDA	0.232 ± 0.091	0.729 ± 0.086	0.67 ± 0.107	0.541 ± 0.142	0.76 ± 0.063	0.344 ± 0.1	0.43 ± 0.046	0.076 ± 0

Note – Values are reported as mean ± standard deviation from 50-FOLD cross-validation.

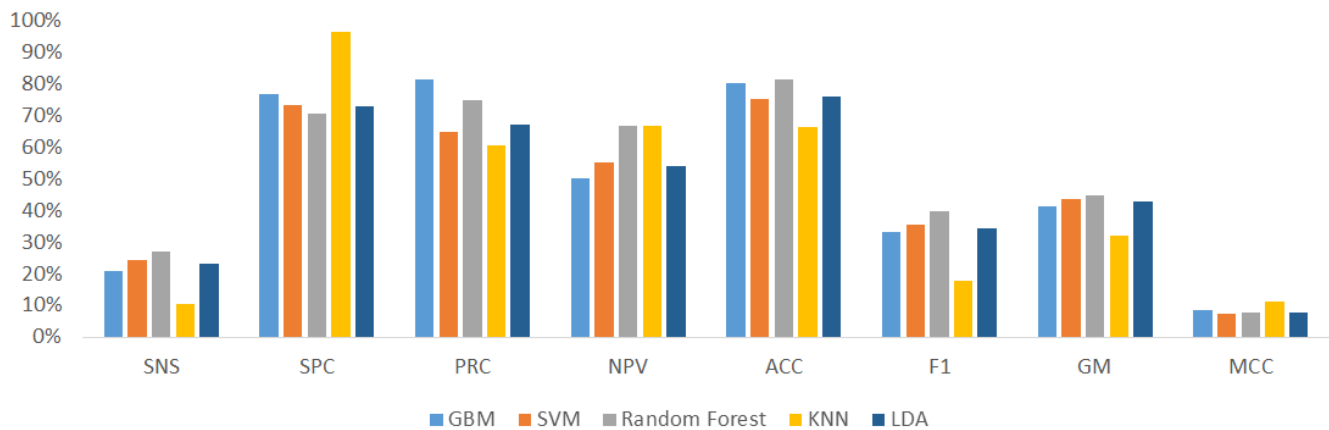


Figure 6. Model performance comparison COPD exacerbation.

All the models were evaluated by eight metrics. Since our dataset was well balanced, we chose the best model based on accuracy. The standard deviations are nearly equal through models, so we could focus on mean values. Thus, the final model was the RF model, which had the highest accuracy rate ($81.4 \pm 6.4\%$).

5.2. Performance of Exacerbations Prediction Model Versus COVID-19

Table 9. Performance of prediction models on the combined COVID-19 Exacerbation dataset.

	Overall Accuracy	WP	WR	WA
GBM	0.776 ± 0.04	0.879 ± 0.016	0.517 ± 0.05	0.837 ± 0.033
SVM	0.798 ± 0.036	0.851 ± 0.031	0.462 ± 0.048	0.863 ± 0.027
RF	0.802 ± 0.036	0.865 ± 0.022	0.473 ± 0.044	0.869 ± 0.024
KNN	0.794 ± 0.131	0.788 ± 0.002	0.79 ± 0.001	0.785 ± 0
LDA	0.776 ± 0.046	0.869 ± 0.022	0.512 ± 0.047	0.837 ± 0.034

Note – Values are reported as mean \pm standard deviation from 50-FOLD cross-validation.

Regarding the second approach, to detect COVID-19 from individuals suffering from exacerbation, models are trained on a dataset composed of COPD patients suffering from exacerbation or not and COVID-19 patients with the performance of KNN; LDA; GBM; RF, and SVM classifiers. Table 9 presents the Performance of prediction models on the combined COVID-19 Exacerbation dataset.

The evaluation of the performance was based on the weighted accuracy.

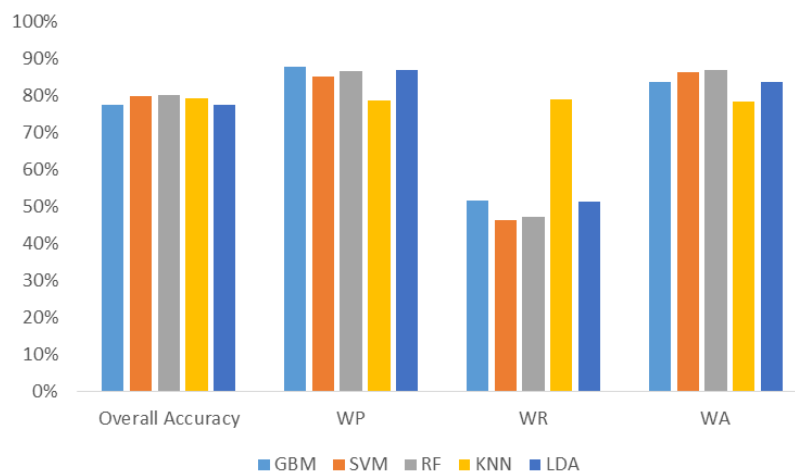


Figure 7. Model performance comparison exacerbation vs COVID-19.

The performance indicators of the models lead us to choose RF as a predictive algorithm. Figure 7 presents the model performance comparison. This algorithm gave us the highest weighted accuracy rate of 86.9%.

Final Algorithm and Prediction Exacerbation or COVID-19 for new patient.

After identifying the best algorithm step for classification, it is now possible to make a prediction for a new COPD patient. Figure 8 describes the prediction approach. Inputs are information of new COPD patients and outputs are results of classifications. Results show whether the new patient belongs exacerbation category or the COVID-19 category. Figure 8 Presents an example of Prediction COVID-19 for new COPD patient.

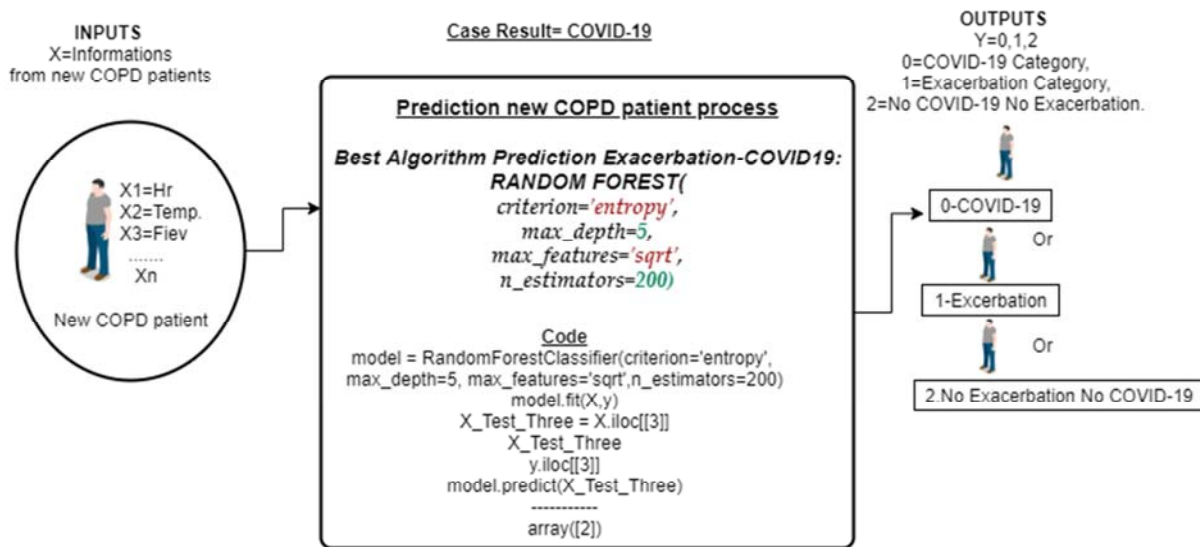


Figure 8. Prediction COVID-19 for new COPD patient process.

Figure 9 presents an example of Prediction Exacerbation for new COPD patient 10.

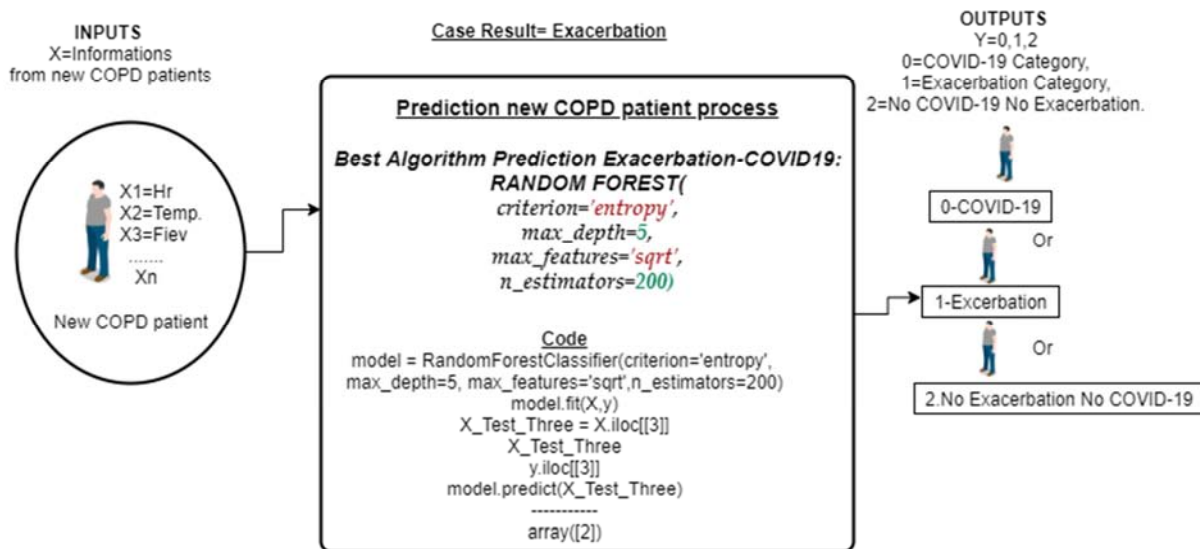


Figure 9. Prediction exacerbation for new COPD patient process.

After identifying the best and final algorithm for prediction, a dynamic context associated with the prediction final model to monitor patient health status by period is proposed.

5.3. Dynamic Approach

This dynamic health status monitoring framework is composed of three parts: COPD Patient Background, Exacerbation&COVID-19 Prediction Model, Patient Health Status Prediction Dashboard by Period. Indeed, for each period, the values of the COPD patients' parameters are collected from

the sensors installed on the COPD patients. The data from the sensors are stored in a database in real-time. Then for each time period, the data on each patient are extracted and injected into a prediction model. The prediction results of all patient data extracted during this period are displayed on the board of the emergency physician, pulmonologist, or medical staff in general. Based on these dashboards that change dynamically per period, the physician can take the appropriate actions to protect the patient from developing exacerbations or COVID-19. Figure 10 describes the different steps to monitor dynamically COPD patient Health Status.

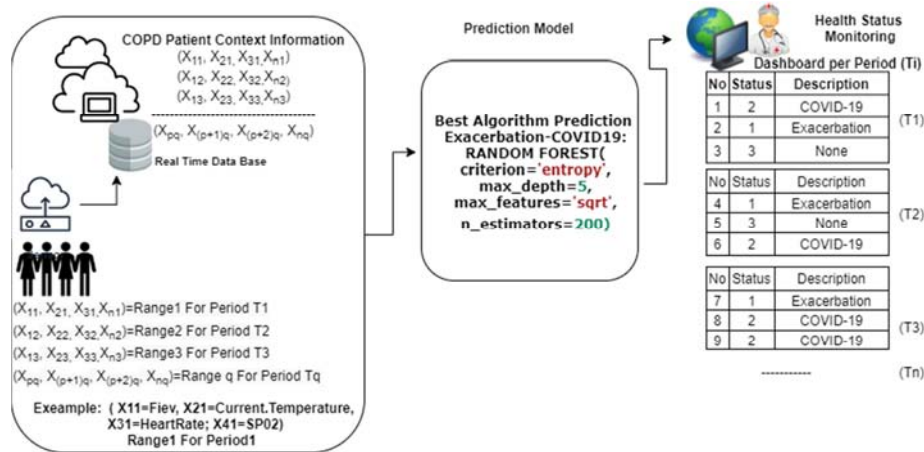


Figure 10. Framework to monitor dynamically COPD patient Health Status.

After the presentation of the framework for monitoring the health status of COPD patients, it is important to note that all the proposed solutions can be adapted to other chronic diseases.

5.4. Applicability to Other Diseases

This classification system can also be used for other diseases with the following steps:

- Identify patient context items, such as vital parameters Sign.
- Propose rules to interpret classification results, and link these rules to the information in the context (range, accuracy of the results).
- Define the relevant comorbidities and symptoms to facilitate patient classification.
- Identify relevant attributes that better describe comorbidities.
- Collect data based on relevant attributes in a database.
- Identify performant model prediction.
- Combine model prediction with context patient.

- Provide Dashboard to display new patient classification.
- Modify the program of the adaptation and inference engine depending on current context information.
- Modify the adaptation mechanism.

The architecture of the classification process does not need to be modified. The context information just needs to be changed according to the parameters identified for characterizing the disease. Overall, these are minor modifications because several parts of the Python prediction model are based on standard Machine Learning (ML), so DL Algorithms and classifiers can be reused.

5.5. Comparison of the Prototype to Other Tools

Our system is based on the com of classification of COPD patients with COVID-19 symptoms and those with exacerbation. This system is dynamic and uses adaptation mechanisms fitted to the COPD patient context. Table 10 presents a comparison of our system to others. Our prototype meets the needs of the ubiquitous system: a dynamic, adaptable, and context-aware approach.

Table 10. Comparison of our system to others.

Advantages/Features	Our Prototype	Prediction of Severity COVID-19 [35]	COPD Acute Exacerbation [36]	Classification of Exacerbation COPDGene [13]
Dynamic patient context	x			
Imbalanced dataset addressed	x			
Prediction approach	x	x	x	x
Combined machine learning algorithm and the dynamic patient context in the classification process	x	x		
Interface to interpret dynamically classification result	x			

6. Conclusions and Future Work

In this paper, after training and testing five Machine learning (ML) algorithms, a dynamic classification approach that combines the best prediction model of COPD patient classification (exacerbation, COVID-19) with dynamic COPD patient context is proposed. Indeed, periodically, data on a new patient is injected into the prediction model. At the output of the model, the patient is either classified in the exacerbation category,

or classified in the COVID-19 category, or no category. by period. Then a dynamic dashboard of classified patients, is implemented to help the medical staff to make appropriate decisions. In comparison with other classification approaches, our solution could easily dynamically distinguish between COPD patients with exacerbations and those with COVID-19 symptoms. The main contributions are the prediction of COVID-19 based on the dynamic ML models, the dynamic interpretation of accuracy, monitoring dynamically health status based on dashboard per period. Since early intervention is associated with improved

prognosis, with our solution, healthcare staff can identify COPD patients who are most at risk of developing exacerbations or COVID-19. Consequently, upon admission, these patients can receive appropriate care as soon as possible. The innovation of this research is also the

combined adaptation model, using dynamic COPD patient context with the performant prediction model based on machine learning to follow the evolution of COPD patient comorbidities (exacerbation, COVID-19).

Our future work will investigate the relationship between dynamic context and accuracy improvement in a classification process.

Author Contributions

KM designed the presented model. KM, and HA analyzed

the classification model. KM wrote the manuscript in consultation with HA. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest

The authors declare that there is no conflict of interest associated with this publication.

Funding

This work was sponsored by Natural Sciences and Engineering Research Council of Canada (NSERC) and the computer science department of the University of Québec at Chicoutimi (Québec), Canada.

Appendix

Table 11. Feature Selection and Information Gain in Exacerbation prediction.

Variables	RFE
Gender	True
Age	False
Short of Breath	True
Short of Breath_1	True
Cough	True
Wheezing	True
Weight	False
Current FEV1	True
Current Temperature	True
Current Heart Rate	True
Sputum	True
Exacerbation	True

Table 12. Feature Selection and Information Gain in Exacerbation - COVID-19 prediction.

Feature	RFE
Fever	True
Dry cough	True
Difficulty in breathing	True
Gender	True
Age	True

The column RFE specifies the selection of variables for the model, TRUE if the variable is selected, FALSE otherwise. The column Information Gain presents the global quantity of information carried by a variable. This information reflects the importance of the variable in the model. Table 13 below describes the column Information Gain for each variable.

Table 13. Description of quantitative features.

	Exacerbation		Total	Sig
	Yes	No		
Age	65.30591	64.95993	65.08	
Weight	162.124	163.7863	163.21	
Current FEV1 Predicted	54.29817	47.30443	49.7	***
Current Heart Rate	85.21023	87.34433	86.61	***
Current Pulse Ox	92.0904	90.43615	91	***
Current Temperature	98.9297	98.9905	98.95	**

Table 13 presents the quantitative features that belonged to patients in our first dataset. The two columns of Exacerbation Yes and No represent the mean values of the features in our dataset according to whether the individuals suffered from an exacerbation. The Total column shows the mean values of the features in the overall sample. Finally, the Sig column provides information on the significance of the test of differences in means between the quantitative features and the exacerbation variable. The confidence interval is 95% for the different information described in the table.

Table 14. Description of the qualitative features.

	Exacerbation		Total	Total	Sig
	N	Y			
	%	%	%	Effective	
Gender					
Female	50.9	49.1	49.7	1248	
Male	49.1	50.9	50.3	1262	
Short of breath					***
Less Than Usual	38	15.3	23.1	580	
More Than Usual	27.5	66.2	52.9	1328	
More than usual	0.1	0	0	1	
Same As Usual	34.4	18.5	23.9	601	
Cough					***
Less Than Usual	47	9.5	22.4	562	
More Than Usual	19.5	69.5	52.3	1313	
Same As Usual	33.5	21	25.3	635	
Wheezing					***
Less Than Usual	42.5	13.7	23.6	592	
More Than Usual	2.5	38.6	26.2	658	
Same As Usual	54.9	47.7	50.2	1260	
Sputum					***
Both Increased Sputum Production And Change in Sputum Color	5.8	25	18.4	462	
Change in Sputum Color	10.7	23.7	19.2	482	
Increased Sputum Production	8.2	22.5	17.6	442	
No Change In Sputum	75.3	28.8	44.8	1124	
Short of Breath-2					***
Less Than Usual	38	15.3	23.1	580	
More Than Usual	27.5	66.2	52.9	1328	
More than usual	0.1	0	0	1	
Same As Usual	34.4	18.5	23.9	601	
Total	50	50	100%	2550	

Significance levels of the association test – *10%, **5%, ***1%.

Table 15 presents the description of the qualitative features.

Table 15. Description of the combined sample COVID-19 – Exacerbation.

	No COVID No Exacerbation		Exacerbation		COVID-19		Total		Sig
	Effective	%	Effective	%	Effective	%	Effective	%	
Temperature									***
No	167	41%	1567	39%	799	20%	4036	100%	
Yes	132	90%	80	6%	64	4%	1464	100%	
Difficulty in Breathing									***
No	1452	58%	423	17%	617	25%	2492	100%	
Yes	1538	51%	1224	41%	246	8%	3008	100%	
Gender									***
No	878	41%	809	38%	439	21%	2126	100%	
Yes	2112	63%	838	25%	424	13%	3374	100%	
Age									***
25–60	2990	77%	593	15%	304	8%	3887	100%	
60 years and over	0	0%	1054	65%	559	35%	1613	100%	
Total	2990	54%	1647	30%	863	16%	5500	100%	

Table 16. Description of selected attributes for exacerbation prediction.

Variables	Values	Codification
Gender	Male, Female	1,2
Age	Numeric	Numeric
Short of Breath	Less Than, Same As Usual, More Than Usual	1,2,3
Short of Breath_1	Less Than, Same As Usual, More Than Usual	1,2,3
Cough	Less Than, Same As Usual, More Than Usual	1,2,3
Wheezing	Less Than, Same As Usual, More Than Usual	1,2,3
Current FEV1	Numeric	Numeric
Current Temperature	Numeric	Numeric
Current Heart Rate	Numeric	Numeric
Sputum	Both Increased sputum production and Change in Sputum, Change in Sputum Color, Increased Sputum Production, No Change in Sputum	1,2,3,4
Exacerbation	Yes, No	1,0

Table 17. Description of selected attributes for exacerbation prediction.

Variables	Values	Codification
Gender	Male, Female	1,0
Age	[25,60], [60,+]	4,5
Dry Cough	Yes, No	1,0
Difficulty in Breathing	Yes, No	1,0
Cough	Yes, No	1,0
Disease	COVID-19, Exacerbation, No COVID-19 and No Exacerbation	0,1,2

Table 18. Tuning Parameters.

Model	Exacerbation Predictions	Exacerbation COVID 19 Predictions
GBM	learning_rate=0.01, n=1000	learning_rate=0.005, n = 1000
SVM	C=0.1, gamma = 0.001, kernel='poly'	C=1, gamma = 0.1, kernel='poly'
Random Forest	criterion='entropy', max_depth=8, max_features='log2', n_estimators=200	criterion='entropy', max_depth=5, max_features='sqrt', n_estimators=200
KNN	algorithm='auto', leaf_size=30, metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=65, p=2, weights='distance'	algorithm='brute', leaf_size=30, metric='minkowski', metric_params=None, n_jobs=None, n_neighbors=47, p=2, weights='distance'
LDA	Default	Default

Table 19. Performance of exacerbation vs COVID-19 prediction models.

	Precision A	Precision B	Precision C	Recall A	Recall B	Recall C	Accuracy A	Accuracy B	Accuracy C
GBM	1±0	0,613±0,091	0,24±0,134	0,645±0,055	0,215±0,048	0,04±0,025	0,835±0,038	0,832±0,031	0,885±0,033
SVM	0,979±0,041	0,595±0,095	0,516±0,131	0,608±0,05	0,211±0,041	0,087±0,028	0,869±0,029	0,83±0,037	0,898±0,026
Random Forest	1±0	0,616±0,087	0,421±0,139	0,613±0,052	0,215±0,048	0,072±0,025	0,879±0,028	0,833±0,031	0,892±0,035
KNN	0,852±0,002	0,698±0,005	0,729±0,025	0,926±0	0,699±0,011	0,47±0,02	0,873±0,002	0,823±0,001	0,415±0,002
LDA	1±0	0,578±0,096	0,288±0,151	0,645±0,051	0,201±0,054	0,048±0,032	0,835±0,036	0,832±0,04	0,885±0,032

Code Sample – Feature Selection

```

from sklearn.ensemble import GradientBoostingClassifier
estimator_GBM = GradientBoostingClassifier()
selector = RFE(estimator_GBM, n_features_to_select=10, step=1, verbose=0,)
selector = selector.fit(X, y)
selector.support_, selector.ranking_, X.info()

```

Code Sample – K-Fold Cross Validation

```

from sklearn.metrics import classification_report, confusion_matrix
from sklearn.model_selection import cross_val_score
from sklearn import metrics
accuracy_50 = cross_val_score(model, X, y, cv=KFold(n_splits=50,shuffle=True),scoring='accuracy')
precision_50 = cross_val_score(model, X, y, cv=KFold(n_splits=50,shuffle=True),scoring='precision_macro')
recall_50 = cross_val_score(model, X, y, cv=KFold(n_splits=50,shuffle=True),scoring='recall_macro')
f1_50 = cross_val_score(model, X, y, cv=KFold(n_splits=50,shuffle=True),scoring='f1_macro')
confusion_matrix_50 = cross_validate(model, X, y, cv=KFold(n_splits=50,shuffle=True),scoring=confusion_matrix_scorer)

```

Code Sample – Tuning Parameters – Halving Grid Search

```

from sklearn.model_selection import GridSearchCV
from sklearn.experimental import enable_halving_search_cv
from sklearn.model_selection import HalvingGridSearchCV
from sklearn.ensemble import RandomForestClassifier
base_estimator = RandomForestClassifier()
param_grid = {
    'n_estimators': [200, 500],
    'max_features': ['auto', 'sqrt', 'log2'],
    'max_depth': [4,5,6,7,8],
    'criterion': ['gini', 'entropy']
}
HGSCV = HalvingGridSearchCV(base_estimator, param_grid, cv=5, factor=4, min_resources=20).fit(X, y)
from sklearn.ensemble import GradientBoostingClassifier
base_estimator = GradientBoostingClassifier()
param_grid = {
    'learning_rate': [0.15, 0.1, 0.05, 0.01, 0.005, 0.001],

```

```
'n_estimators': [100,250,500,750,1000,1250,1500,1750]
}
HGSCV = HalvingGridSearchCV(base_estimator, param_grid, cv=5, factor=4, min_resources=20).fit(X, y)
from sklearn.svm import SVC
base_estimator = SVC()
param_grid = {'C': [0.1, 1, 10, 100], 'gamma': [1, 0.1, 0.01, 0.001], 'kernel': ['rbf', 'poly', 'sigmoid']}
HGSCV = HalvingGridSearchCV(base_estimator, param_grid, cv=5, factor=4, min_resources=20).fit(X, y)
```

References

- [1] Chronic respiratory diseases. Available online: www.who.int/respiratory (accessed on 10-03-2021).
- [2] Stiell, I. G.; Perry, J. J.; Clement, C. M.; Brison, R. J.; Rowe, B. H.; Aaron, S. D.; McRae, A. D.; Borgundvaag, B.; Calder, L. A.; Forster, A. J.; et al. Clinical validation of a risk scale for serious outcomes among patients with chronic obstructive pulmonary disease managed in the emergency department. *Can. Med. Assoc. J.* 2018, 190, E1406–E1413, doi: 10.1503/cmaj.180232.
- [3] Ajami, H., & Mcheick, H. (2018). Ontology-based model to support ubiquitous healthcare systems for COPD patients. *Electronics*, 7 (12), 371.
- [4] Kouamé, K. M., & Mcheick, H. (2021). An Ontological Approach for Early Detection of Suspected COVID-19 among COPD Patients. *Applied System Innovation*, 4 (1), 21.
- [5] Tal-Singer, R., & Crapo, J. D. (2020). COPD at the time of COVID-19: a COPD Foundation perspective. *Chronic Obstructive Pulmonary Diseases: Journal of the COPD Foundation*, 7 (2), 73.
- [6] Olloquequi, J. (2020). COVID-19 Susceptibility in chronic obstructive pulmonary disease. *European journal of clinical investigation*, 50 (10), e13382.
- [7] Alqahtani, J. S., Oyelade, T., Aldhahir, A. M., Alghamdi, S. M., Almhadi, M., Alqahtani, A. S.,... & Hurst, J. R. (2020). Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: a rapid systematic review and meta-analysis. *PloS one*, 15 (5), e0233147.
- [8] Leung, J. M., Niihura, M., Yang, C. W. T., & Sin, D. D. (2020). COVID-19 and COPD. *European Respiratory Journal*, 56 (2).
- [9] Deslée, G., Zysman, M., Burgel, P. R., Perez, T., Boyer, L., Gonzalez, J., & Roche, N. (2020). Chronic obstructive pulmonary disease and the COVID-19 pandemic: Reciprocal challenges. *Respiratory medicine and research*, 78, 100764.
- [10] Attaway, A. (2020). Management of patients with COPD during the COVID-19 pandemic. *Cleveland clin. j. med.*
- [11] Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K.,... & Dean, J. (2019). A guide to DL in healthcare. *Nature medicine*, 25 (1), 24-29.
- [12] Humphries, S. M., Notary, A. M., Centeno, J. P., Strand, M. J., Crapo, J. D., Silverman, E. K., & Genetic Epidemiology of COPD (COPDGene) Investigators. (2020). DL enables automatic classification of emphysema pattern at CT. *Radiology*, 294 (2), 434-444.
- [13] J. Ying et al., "Classification of Exacerbation Frequency in the COPDGene Cohort Using DL With Deep Belief Networks," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 6, pp. 1805-1813, June 2020, doi: 10.1109/JBHI.2016.2642944.
- [14] Mohsen, H., El-Dahshan, E. S. A., El-Horbaty, E. S. M., & Salem, A. B. M. (2018). Classification using DL neural networks for brain tumors. *Future Computing and Informatics Journal*, 3 (1), 68-71.
- [15] Altan, G., Kutlu, Y., & Allahverdi, N. (2019). DL on computerized analysis of chronic obstructive pulmonary disease. *IEEE journal of biomedical and health informatics*, 24 (5), 1344-1350.
- [16] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., & Lungren, M. P. (2018). DL for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practising radiologists. *PLoS medicine*, 15 (11), e1002686.
- [17] Negahdar, M., & Beymer, D. (2019, March). Lung tissue characterization for emphysema differential diagnosis using deep convolutional neural networks. In *Medical Imaging 2019: Computer-Aided Diagnosis* (Vol. 10950, p. 109503R). International Society for Optics and Photonics.
- [18] Wang, Q., Wang, H., Wang, L., & Yu, F. (2020). Diagnosis of Chronic Obstructive Pulmonary Disease Based on Transfer Learning. *IEEE Access*, 8, 47370-47.
- [19] Peng, J., Chen, C., Zhou, M., Xie, X., Zhou, Y., & Luo, C. H. (2020). A machine-learning approach to forecast aggravation risk in patients with acute exacerbation of chronic obstructive pulmonary disease with clinical indicators. *Scientific reports*, 10 (1), 1-9.
- [20] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., & Xia, J. (2020). Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*.
- [21] Liang, W., Yao, J., Chen, A., Lv, Q., Zanin, M., Liu, J., & He, J. (2020). Early triage of critically ill COVID-19 patients using DL. *Nature communications*, 11 (1), 1-7.
- [22] Nature communications, 11 (1), 1-7. Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B.,... & Xia, J. (2020). Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*.
- [23] Fadlullah, Z. M., Tang, F., Mao, B., Kato, N., Akashi, O., Inoue, T., & Mizutani, K. (2017). State-of-the-art DL: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems. *IEEE Communications Surveys & Tutorials*, 19 (4), 2432-2455.
- [24] Khemphila, A., & Boonjing, V. (2011, August). Heart disease classification using neural network and feature selection. In *2011 21st International Conference on Systems Engineering* (pp. 406-409). IEEE.

- [25] Bhatia, S., Prakash, P., & Pillai, G. N. (2008, October). SVM based decision support system for heart disease classification with integer-coded genetic algorithm to select critical features. In *Proceedings of the world congress on engineering and computer science* (pp. 34-38).
- [26] Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., Rueckert, D., & Alzheimer's Disease Neuroimaging Initiative. (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*, 65, 167-175.
- [27] Sinha, P., & Sinha, P. (2015). Comparative study of chronic kidney disease prediction using KNN and SVM. *International Journal of Engineering Research and Technology*, 4 (12), 608-12. To provide personalized care to chronic patients at home. *Journal of biomedical informatics*, 46 (3), 516-529.
- [28] Banu, G. R. (2016). Predicting thyroid disease using linear discriminant analysis (LDA) data mining technique. *Commun. Appl. Electron. (CAE)*, 4, 4-6.
- [29] Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28 (1), 112-118.
- [30] Azhagusundari, B., & Thanamani, A. S. (2013). Feature selection based on information gain. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2 (2), 18-21.
- [31] Richhariya, B., Tanveer, M., Rashid, A. H., & Alzheimer's Disease Neuroimaging Initiative. (2020). Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE). *Biomedical Signal Processing and Control*, 59, 101903.
- [32] Brownlee, J. (2019). Information Gain and Mutual Information for Machine Learning. *Preuzeto*, 18, 2020.
- [33] Python tutorial. Available online: <https://data-flair.training/blogs/python-tutorials-home/> (accessed on 21/03/2021).
- [34] Brownlee, J. (2016). *Machine Learning mastery with python*. Machine Learning Mastery Pty Ltd, 527, 100-120.
- [35] Alotaibi, A., Shiblee, M., & Alshahrani, A. (2021). Prediction of Severity of COVID-19-Infected Patients Using Machine Learning Techniques. *Computers*, 10 (3), 31.
- [36] Rhee, C. K., Kim, J. W., Yoo, K. H., & Jung, K. S. (2020). Prediction model of COPD acute exacerbation with big data by Machine Learning methods.
- [37] Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22 (1), 31-72.
- [38] He and E. A. Garcia, *Learning from Imbalanced Data*, IEEE Transactions on Knowledge and Data Engineering, 21 (2009), No. 9, 1263-1284.
- [39] Truica, C. O., & Leordeanu, C. A. (2017). Classification of an imbalanced data set using decision tree algorithms. *Univ. Politech. Bucharest Sci. Bull. Ser. C Electr. Eng. Comput. Sci*, 79, 69-84.
- [40] Brownlee, J. (2020). *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*. Machine Learning Mastery.
- [41] Amalia Luque et al., The impact of class imbalance in classification performance metrics based on the binary confusion matrix, *Pattern Recognition*, Elsevier, 91 (2019) 216–231.
- [42] Alberto Fernández et al., Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches, *Knowledge-Based Systems*, Elsevier, 42 (2013) 97–110.
- [43] Scikit-learn: Machine Learning in Python, Pedregosa et al., *JMLR* 12, pp. 2825-2830, 2011.
- [44] Esteban, Cristobal and Moraza, Javier and Esteban, Cristobal and Sancho, Fernando and Aburto, Myriam and Aramburu, Amaia and Goiria, Begona and Garcia-Loizaga, Amaia and Capelastegui, Alberto, "Machine learning for COPD exacerbation prediction," *European Respiratory Journal*, vol. 46, 2015.
- [45] Jaap CA Trappenburg, Irene Touwen, Gerdien H de Weert-van Oene, Jean Bourbeau, Evelyn M Monninkhof, Theo JM Verheij, Jan- Willem J Lammers, Augustinus JP Schrijvers, "Detecting exacerbations using the Clinical COPD Questionnaire," *Health Qual Life Outcomes*, vol. 8, no. 102, 2010.
- [46] Bertens, L. C. M.; Reitsma, J. B.; Moons, K. G. M.; van, M. Y.; Lammers, J. J.; Broekhuizen, B.; Hoes, A. W.; Ru, tten F. H., "Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease," *Int J Chron Obstruct Pulmon Dis.*, vol. 8, p. 493499, 2013.