

# **Predicting the Correct Procedure of COVID-19 Patients in Hospitals Using Machine Learning**

التنبؤ بالإجراءات المناسبة لمرضى الكورونا في المستشفيات باستخدام تعلم الآلة

By

**Hiba Jaleel Talib**

**Supervisor:**

**Dr. Mohammad Al-Azzeh**

*Full Professor*

**Co-Supervisor:**

**Dr. Mohammad Hijjawi**

## **Abstract**

Since the beginning of the COVID-19 (Coronavirus Disease 2019) global pandemic, there is an urgent need to avoid the spread of COVID-19 in all the world. Despite the development of several vaccines that do not seem to be yet widely distributed to cover most countries, especially those with experiencing second waves of the disease and those struggling with the emergence of new more transmissible and/or more virulent variants of the virus. Health officials warn that designated treatment centers are also overwhelmed with COVID-19 patients, which increases pressure on hospitals and all kinds of health care system resources. Therefore, there is a desperate need to reduce such pressure and avoid extra costs on patients and hospitals. The solution that we provide in this thesis aims at identifying the correct procedure for COVID-19 patients when they visit hospitals. Specifically, we propose using machine-learning algorithms like Naïve Bayes and Support Vector Machine to find the best procedure for each patient based on his health status. This solution is supposed to reduce crowdedness in hospitals, the cost of treatments, relieving pressure on healthcare systems, especially on human resources. The result for each algorithm was compared and evaluated using classification accuracy measures to determine which algorithm has a better accuracy measure, we compare the result for each algorithm once we run the experiment with all features and record the accuracy measure such as (Classification Accuracy, Sensitivity, Specificity), We noticed that the results of the two algorithms were fluctuating and converging making it difficult to determine which algorithm was the most accurate. So, we used the feature that we ranked by Gain Ratio algorithm to rerun the experiment for each algorithm and recorded the results for accuracy measures to determine the best accuracy upon such reduction in the number of features and we concluded that when the rank was 6 to all features selected the result was stabilization., and when the rank was from (2-6) features the result was the worst. Also, the results and the accuracy measures results were improved upon using the SVM rather than NB. Then we identified the algorithm of the best performance. We built hypotheses using a nonparametric statistical test (Mann-Whitney test) and the Boxplot to visualize the performance results. We concluded that the SVM and NB algorithms have the same performance. Except for a slight difference in AUC performance measure between the two models.

**Keywords:** Machine Learning (ML), Support Vector Machine, Naïve Bayes, COVID-19, Mann-Whitney