CP-54 COVID-19 MORALITY PREDICTION USING MACHINE LEARNING TECHNIQUES

Abstract

The COVID-19 pandemic sparked our research to explore and design a predictive model through Machine Learning algorithms to determine risk and mortality of COVID-19 admitted patients. Using a data set with over 90,000 patient admits and 20 clinical health features, this study aims to help prioritize patient care on patients that have a higher risk for COVID-19 being fatal based on health risk. The results reached up to 96 percent accuracy in predicting morality rate. Different machine learning techniques including Navies Bayes, Decision Trees, K-Nearest Neighbor, Support Vector Machine (SVM), Random Forrest and Multilayer Perceptron (MLP) were utilized in this study. Additionally, Feature Selection through ranking and ensemble methods of bagging, stacking and boosting were also used.

Introduction

Predicting which patients face a higher health risk and has a higher mortality rate and whether COVID-19 is a factor not only helps clinical decisions but helps in combating COVID-19 overall. COVID-19 pandemic has caused unimaginable devastation and suffering around the world. Machine learning methods and data mining used and applied can offer great promise for fast and accurate detection and prediction of COVID-19 compared to the traditional standard-of-care radiography and tomography images. In this study we use the "COVID-19 Mexico Patient Health Dataset" with 20 clinical features, our research applies models to determine the most accurate using WEKA to find morality rate predictions.

Research Question(s)

1. What is the role of Machine Learning in COVID-19 Morality Prediction?
2. What is the best/recommended machine learning model can help healthcare professionals in combating COVID-19?
3. Can feature selection techniques or applying ensemble learning techniques improve the accuracy of machine learning models in combating with COVID-19?
4. Explore how machine learning has battled COVID-19 and provide directions for future research on COVID-19.

Materials and Methods

- Feature Selection and Ranking
- Filtering
- Ensemble Learning

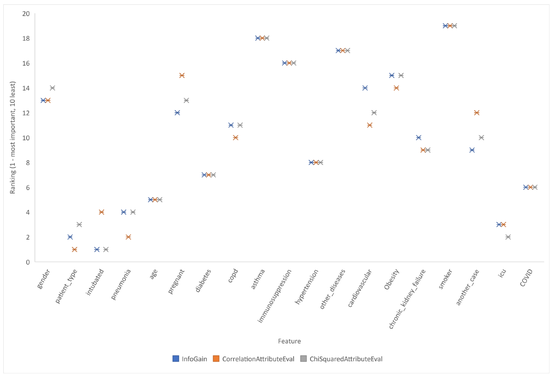

Fig.1 Ranking of Feature Methods

Results

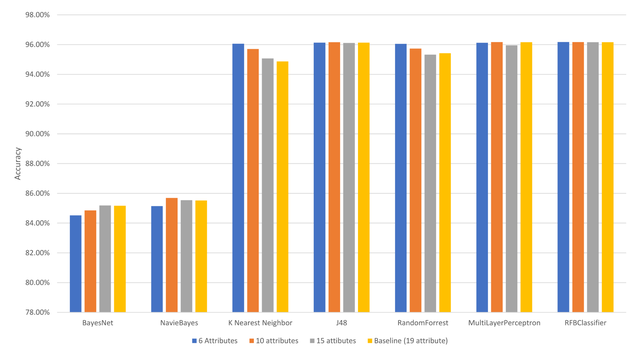

Fig 2. Results Using Different Machine Learning Techniques with Different Number of Features and 10 -fold cross validation

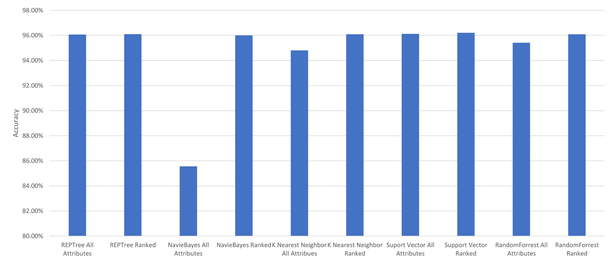

Fig 3. Results Using Different Machine Learning Techniques with Bagging and 10-fold cross validation

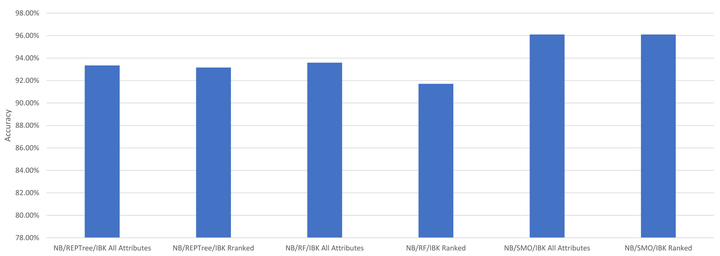

Fig 4. Results Using Different Machine Learning Techniques with Stacking and 10 -fold cross validation

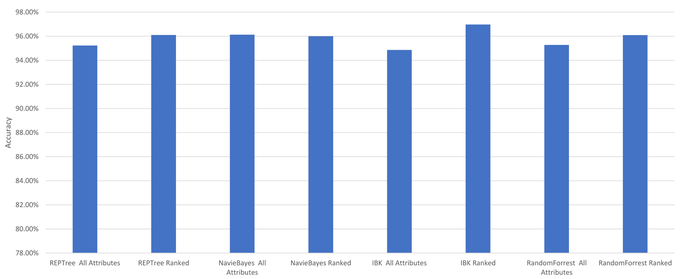

Fig 5. Results Using Different Machine Learning Techniques with Boosting and 10 -fold cross validation

Conclusions

In feature ranking, it was determined that health risks that contribute to a higher mortality rate included: if the patient was intubated, patient type, if admitted to ICU, diagnosis of pneumonia, the age of the patient and if the COVID-19 result was positive. Following these were diabetes, hypertension and chronic kidney failure.

In relation to number of features selected, only slight difference were observed in each experiment topping at 96% accuracy in multiple methods (MLP, J48, RBF Classifier)

With ensemble techniques, stacking method created a more robust model using Naive Bayes as the base level 0 classifier, combining it with Support Vector Machines and K-Nearest Neighbor with 10-fold cross validation producing a 96% accuracy. Bagging also reached 96% accuracy.

Acknowledgments

Special thanks to co-author Dr. Pouriyeh in direction on this project

We will be submitting the results to ICTS4eHealth in the 26th IEEE Symposium on Computers and Communications (https://icts4ehealth.icar.cnr.it/)

Contact Information

Teams:

Lindsay Schirato: lschirat@kennesaw.students.edu
Dwayne Flanders: dflande4@kennesaw.students.edu
Kennedy Makina: kmakina@kennesaw.students.edu

Advisor:

Dr. Seyeadmin Pouriyeh: spouriye@kennesaw.edu

References

[1] Matta, D. M., & Saraf, M. K. (2020). Prediction of COVID-19 using Machine Learning Techniques.

[2] Agbelusi, O., & Olayemi, O. C. (2020). Prediction of Mortality Rate of COVID-19 Patients using Machine Learning Techniques in Nigeria. *International Journal of Computer Science and Software Engineering*, *9*(5), 30-34.

[3] Langer, T., Favarato, M., Giudici, R., Bassi, G., Garberi, R., Villa, F., ... & Fumagalli, R. (2020). Development of machine learning models to predict RT-PCR results for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in patients with influenza-like symptoms using only basic clinical data. *Scandinavian journal of trauma, resuscitation and emergency medicine*, *28*(1), 1-14.

Project Website: https://sites.google.com/view/covid-19-spring2021/home

Data set: https://www.kaggle.com/marianarfranklin/mexico-covid19-clinical-data

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING

Author(s): Lindsay Schirato, Dwayne Flanders & Kennedy Makina
Advisor(s): Dr. Seyedamin Pouriyeh