

# Evaluation Of Infectious Diseases Outbreak Among Prisoners Using Machine Learning

**Kofoworola O. Fapohunda**

Department of Biomedical Engineering  
Louisiana Tech University  
United State  
[Kof003@latech.edu](mailto:Kof003@latech.edu)

**Mary Nations**

Department of Biomedical Engineering  
Louisiana Tech University  
United State  
[Med041@latech.edu](mailto:Med041@latech.edu)

**James Vieux**

Department of Biomedical Engineering  
Louisiana Tech University  
United State  
[jpv009@latech.edu](mailto:jpv009@latech.edu)

**Daniel Davis**

Department of Biomedical Engineering  
Louisiana Tech University  
United State  
[dwd015@latech.edu](mailto:dwd015@latech.edu)

**Rhoda O. Adewunmi**

Department of Public health  
North Dakota State University  
United State  
[rhoda.adewunmi@ndsu.edu](mailto:rhoda.adewunmi@ndsu.edu)

**Temitayo E. Balogun**

Department of Computer Science  
Ekiti State Polytechnic, Isan-Ekiti  
Nigeria  
[tebalogun@ekspoly.edu.ng](mailto:tebalogun@ekspoly.edu.ng)

**Samuel G. Faluyi**

Department of Computer Science  
Ekiti State Polytechnic, Isan-Ekiti  
Nigeria  
[sgfaluyi@ekspoly.edu.ng](mailto:sgfaluyi@ekspoly.edu.ng)

**Abstract**— Prisoners' health is highly vital as everyone else health is important and thus requires proper management. Most especially management of infectious diseases among people in prisons necessitates forecast outbreaks. This study concentrates on using decision tree (machine learning approach) algorithm for the prediction of infectious diseases adopting the COVID dataset collected from prisons. The J48 decision tree algorithm was used for the formulation of the model and was evaluated using three different validation methods: 10-fold cross-validation, holdout, and 3-way split. The performance metric used was an accuracy measure along with a confusion matrix to show how the instances were classified and misclassified. The result shows that the decision tree with 10-fold cross-validation performed better than the decision tree with holdout and 3-way split. The study concluded that forecasting outbreaks of infectious diseases is appropriate and necessary, most especially in prisons and machine learning techniques adoption make it more effective and efficient.

**Keywords**—Prisoners' health; Infectious diseases; Decision tree; validation method; performance metrics

## I. INTRODUCTION

One of the major responsibilities of prisons as organizations is to see to the appropriate health management of inmates in the yard. It is essential to note that inmates' health is highly vital, this is the fact

that every human has the right to life. This is in connection with the UN Basic Principles for the Treatment of Prisoners, which states that convicts "shall have access to the health services available in the country without discrimination on the grounds of their legal situation" [21]. Furthermore, they mentioned that for the purpose of combating epidemics like COVID-19, improving prison health care is essential and requires proactiveness. Infectious infections are significantly more likely to affect inmates than the general population. The factors that contribute to highly contagious prison conditions include overcrowding, subpar medical care, high-risk behavior, security against public health concerns, and a lack of sympathy for the convicts [8]; [16]; [21]. Moreover, special procedures are needed to deliver information about public health in jails because inmates have restricted access to outside news (and typically no connection to the internet inside of cells) [20].

According to [6] review of [12], mentioned in one COVID-19 trial, repeated testing and symptom screening (i.e., temperature screening and oxygen saturation assessments) were employed to reduce the spread of SARS-CoV-2 infection. In order to categorize instances as asymptomatic, pre-symptomatic, or symptomatic during the investigation period, people in the affected buildings were tested on three different occasions [6]. This combined approach improves the detection of SARS-CoV-2 infected individuals who are pre-symptomatic or asymptomatic and could otherwise go undetected if symptom screening alone were used [5]. The related impediments connected with reporting illness among people in prisons, such as avoiding medical isolation, and, in some situations, the costs of

medical care, can also render symptom screening inefficient [19].

Due to stigma, a lack of confidence in medical secrecy in prisons, and the potential effects of additional constraints on persons whose freedom has already been violated, people in prison may be more prone to conceal symptoms. Nevertheless, COVID-19 differs significantly from the outbreaks mentioned in earlier research, all of which were self-contained [6]. For testing, tracing, and isolation of suspected and confirmed inmates' cases to be implemented as part of this plan, a focus on staff health and safety will be especially important. Further thought should be given to the adoption of tracking mobile applications for contact tracing, but it is unlikely that such a strategy could be easily put into practice because people in prison frequently lack access to their mobile phones and perhaps other portable network devices. Additionally, following previously jailed individuals after their release would be problematic because many of them wouldn't have phones or would get new ones after their release, and it would also create additional potentially significant legal and human rights issues.

Prediction modelling help to determine the probabilities of future events based on historical data. Predictive analytics uses statistical data analysis algorithms and machine learning approaches. While it cannot predict the future with any degree of certainty, it can reasonably predict what might occur in the future [13]. Furthermore, [15] mentioned that to find patterns in data and gauge the likelihood that certain outcomes will occur, predictive modelling employs data mining, machine learning, and statistics. Developers first identify the issue and gather information before creating an accurate predictive model. Clinical data and claims are combined to achieve the majority of analytical objectives. However, to carry out complex tasks, a prediction model might either use a straightforward linear mathematical function or a sophisticated neural network. Data analysts test and evaluate several models to address the specified issue. The model with the best score is validated, tested, and applied to a dataset from the actual world after the evaluation [15]. Take for instance, a patient with a newly diagnosed disease who fits a specific demographic of race, age, health history, and genetics shows up at a hospital. In order to create a treatment strategy customized to this patient, the predictive model takes into account the new illness, all the patient-specific information discussed above, and the population cohort with similar features. The application of big data in precision medicine is then demonstrated when this plan is subsequently given to the attending physician for approval.

Generally, to ensure accurate learning, sample of the populations, and impartial predictions, the data collected is typically divided into two groups throughout the creation of AI algorithms: a training data set and a test data set. The training set, as its name implies, consists of sets of describing data points (features) and accompanying predictions and is used to train algorithms (in the case of supervised learning). The testing data set, which is brand-new to the algorithm, is used only to evaluate the algorithm's performance.

This step is done to get rid of biases introduced by the training dataset when testing the algorithm [3]. Algorithms are applied in healthcare settings whenever they successfully complete a testing and training phase with respectable outcomes.

Machine learning uses a variety of algorithmic models and statistical techniques to address issues without the need for professional programming [9]; [14]. Due to the fact that certain machine learning models are single-layered, significant portions of feature extraction and data processing are carried out before the data is introduced into the algorithm [4]. These machine learning methods need extensive data preparation in the absence of the additional layers in order to make accurate predictions and prevent over- or under-fitting the training dataset [9].

Despite some observational evidence of extremely contagious illness outbreaks in prison, the particular epidemiology of COVID-19 provides a challenge to prisons that is unlike any other. If the proper public health precautions are not put in place, jails could turn into reservoirs for COVID-19 infection in the community, unlike previous epidemics of other illnesses that self-limited. All prisons should think about isolating all new inmates, conducting contact tracing, providing a highly responsive testing regimen to all inmates and prison staff, prioritizing the early adoption and implementation of diagnostic assays and tests, and communicating concise and regular health information updates to inmates and staff. These necessitate this research work.

This research study focused on developing a predictive model which will predict the outbreak of infectious disease (COVID) using machine learning algorithm (Decision Tree). The specific research objectives are to: obtain dataset, formulate predictive model based on dataset collected using decision tree; simulate the predictive model formulated with various evaluation methods; and validate the model. The remaining sessions are arranged as follow: session 2 (literature review), session 3 (method and materials), session 4 (result of the model), and session 5 (conclusion).

## II. LITERATURE REVIEW

Several studies have been carried out over the years regarding COVID outbreaks and health management system of people in prison. This session presents some review related literatures. Using machine learning and data analysis algorithms, predictive analytics determines the probability of future events based on statistical information. While it cannot predict the future, it can make predictions about potential future events with a respectable degree of accuracy [7]; [11].

In the study carried out by [7], where prediction model for the classification risk of COVID-19 using data mining algorithms. Dataset used was collected via questionnaire and was used for the formulation of the model along with data mining algorithms, which include decision tree and multilayer perceptron. Decision tree algorithm was used for feature selection

and categorization, to classify the risk of having COVID-19. Performance comparison was carried out and result shows that multilayer perceptron perform better than decision tree. However, both algorithms used were able to correctly classify the risks as 'no, low, mild and high' with good percentage of accuracy. But the limitation of the study includes that there is no target audience or organization, which this study intends to address.

Investigations conducted by [6], which were majorly based on developed countries and known outbreaks of tuberculosis, adenovirus, varicella, measles, influenza (types A and B), mumps, and COVID-19. Their report shows some common phenomenon, which include the public health implications of communicable disease outbreaks in jail, and the role of interagency collaboration, health communication, screening for contagious diseases, restriction, isolation and quarantine, contact tracing, immunization programs, epidemiological surveillance and prison-specific guidelines in addressing any outbreaks. They further asserted that prisons are high-risk settings for the transmission of contagious diseases and there are considerable challenges in managing outbreaks in them. A public health approach to managing COVID-19 in prisons is required.

[1] concentrated on the comorbidities and risk factors for intensive care unit admission in COVID-19 patients at NGHHA in Riyadh, Saudi Arabia. They discovered that 234 of the 385 patients brought to the intensive care unit with a positive COVID-19 test were male, making up 60.8% of the entire study population. This result was in line with other research of a similar nature [2]. Regarding age, 257 patients or 66.8% of the entire study group were older than 61 at the time of intensive care unit admission. As additional studies support similar findings, this makes age and male gender the most prevalent features among patients admitted to the intensive care unit [10]. The study categorizes the risk factors as well as comorbidities, such as cardiac diseases, respiratory diseases, gastrointestinal diseases, liver diseases, hypertension, cancer, diabetes mellitus, and kidney diseases.

### III. METHODS AND MATERIALS

This session focused on the methods and materials adopted for this study. This includes the methods of data collection and identification, data pre-processing (i.e. data cleaning, data normalization), model formulations, and validation of the models. This study concentrates on developing a predictive model, which predicts the outbreak of infectious disease (COVID) with the aid of machine learning approaches.

#### A. Data

The dataset used for this study was collected from secondary source. However, the data was initially obtained as CSV file, which comprises of nine (9) attributes and two thousand and thirty-nine (2639) observations, regarding some information of prisons and other rehabilitation centres in United States, US.

The next sub-session discussed various pre-processes adopted used for this study.

#### B. Data Pre-processing

After the collection of datasets form a secondary source, various steps were carried out for data pre-processing of the collected data, which includes cleaning of dataset (removal of incomplete and duplicate rows), and normalization of data.

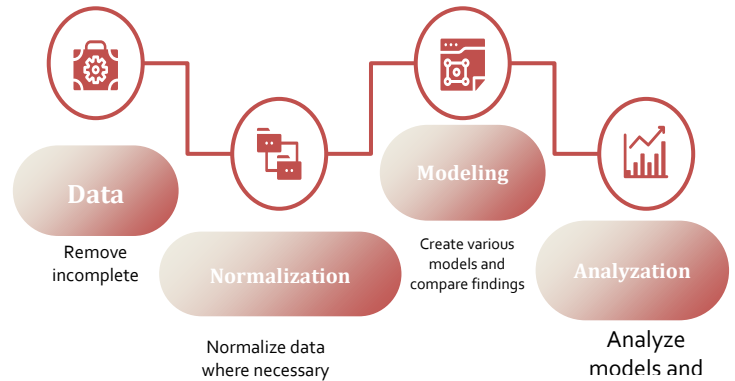


Figure 1: Model formulation flow

#### C. Data Cleaning

As part of the data cleaning process, there were many rows with incomplete data that are not unusable and thus were removed from the entire dataset. Aside from the facility\_type and facility\_state column, the latter of which was used to join another dataset, all nominal and facility identifying columns were removed. Inmate\_tests\_in\_state and total\_inmate\_cases\_in\_state was also added from a secondary dataset from the same source. Furthermore, the dataset was classified into HIGH, MODERATE, and LOW rates of COVID, and made these classifications by creating an inmate\_cases\_over\_population column and dividing its values into 3 buckets.

#### D. Data Normalization

Having cleaned the data, the dataset is normalized using statistical z-score to identify most correlated attributes for model formulation. Testing several different combinations of columns, the best results were found with the following seven columns, viz: inmate\_tests\_in\_state\_norm, total\_inmate\_cases\_in\_state\_norm, latest\_inmate\_population\_norm, max\_inmate\_population\_2020\_norm, total\_inmate\_cases\_norm, prison\_risk\_factor\_norm, pop\_over\_maxpop\_norm.

#### E. Attribute Selection and Model Formulation Using J48 Decision Tree

Decision tree is a machine learning technique that often use classification approach, like natural tree with two main parts (i.e. root node and branches node). Root node being starting point of the trees alongside the branches often called edges, which connect each consecutive node. On the other hands leaf nodes or

terminal nodes are those nodes that do not have child nodes and represent a possible value of the target variable (COVID-19 risk) given the variables represented by the path from the root node. Rules can then be induced from the trees taking paths created from the root node all the way to their respective leaf. The variables identified by the decision trees algorithm were used to formulate predictive models for the risk of COVID-19 based data collected

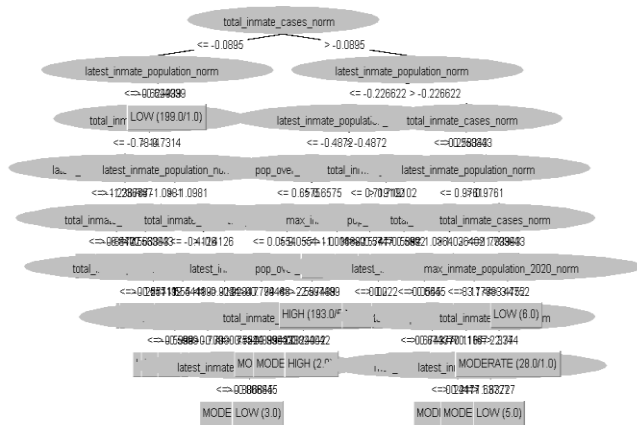


Figure 2: Tree classification COVID-19 risk in the data

The J48 decision tree classification of model for COVID-19 risk as shown in figure 2 has a root node (*total\_inmate\_cases\_norm*), which is the total confirmed cases of COVID-19 identified by the data. However, other attributes are regarded as branch nodes. Thus, each instance of the dataset was classified as 'low, moderate and high'; indicating the leaf or final node. Furthermore, the 10-fold cross validation was used along with decision tree to evaluate the model using the collected dataset.

#### F. Simulation Processes

Following the identification of machine learning algorithms adopted in formulating the prediction algorithm that predicts cases of infectious diseases (COVID-19) among people in prison; the COVID-19 data simulation was used to test the formulated model; various simulation or evaluation methods were adopted for this study. This includes holdout, k-fold cross validation and 3-way split. The simulation tool for the creation of the prediction model was the WEKA open-source software, which is a collection of machine learning techniques.

#### G. K-Fold Cross Validation Technique

When data is divided into 'k' categories at random, it is often referred to as cross-validation, sometimes known as 'k-fold cross-validation'. The test set is taken from one of the groups, while the training set is taken from the remaining groups. The model is tested after being trained on the training dataset. Once every distinct group has been used as the test dataset, the process is then repeated. For this study, the cross-validation procedure was employed, which involved dividing the whole datasets into a number of folds (or

partitions) of the data. Each partition was selected for testing with the remaining  $k - 1$  partition used for training; the next partition was used for testing with the remaining  $k - 1$  partition (including the first partition used for testing) that was used for training until all  $k$  partitions had been selected for testing. The procedure adopted in this research work was a layered 10-fold cross validation techniques that requires that the entire data set split in ten different parts to evaluate the performance of J48 decision tree.

#### H. Holdout

When you divide a data into a "train" and "test" set, you are holding out. The testing data is used to evaluate how well the model works on data that has not yet been seen, whereas the training data is used to train the model. When employing the hold-out method, a typical split is to use 80% of the dataset for training and the remainder 20% for testing [17]. However, it could also be in proportion of 70% and 30% for training and test set, respectively. Often the larger proportion of the datasets are always used for training and the remaining for testing. A machine learning model is evaluated using the hold-out method to see how well it will perform on the fresh data [11]. This study applied the holdout method of evaluation by splitting the data set into two in proportion of 70% to 30%, where 70% of data was used to train the model and remaining 30% was used to test the model.

#### I. Three – way Split

The 3-way split is a validation technique that performs holdout (train-test split) twice to enable optimization of model performance. The method often refers to as 3-way holdout evaluation method. The 3-way split was carried out in this study by dividing the original dataset into two portion (70 and 30) percent, after which the 70 percent was further divided, in proportion of 70% to 30% which stands as validation set. Thus, the data was divided in training set, testing set and validation set. "3-way (train, cross validate, test) split is used to help us determine what model would work best" [18].

## IV. RESULT

The purpose of this study is to predict the output of infectious diseases outbreak using decision tree algorithm. This section presents the results of formulated model with various validation methods. The entire dataset was used to formulate model adopting J48 decision tree, which was evaluated using 10-fold cross validation.

#### A. Decision Tree Model Formulation with 10-fold Cross-Validation Result

Model formulation with J48 decision tree algorithms included in the Weka software, this comes after determining the risk factors related to COVID-19. Utilizing the historical dataset for model training, the 10-fold cross-validation procedure was performed to assess the effectiveness of the constructed prediction model for COVID-19 risk.

Correctly Classified Instances	775	93.0372 %
Incorrectly Classified Instances	58	6.9628 %

Figure 3: Classification report for J48 decision tree model with 10-fold cross validation

```

=== Confusion Matrix ===
      a  b  c  <-- classified as
    579  20  0 |  a = LOW
      24 187  4 |  b = MODERATE
       0  10  9 |  c = HIGH
    
```

Figure 4: Confusion Matrix for J48 decision tree model with 10-fold cross validation

Having formulated the prediction model using J48 decision tree with COVID-19 dataset, the classification report and confusion matrix was identified in figures 3 and 4. Utilizing 10-fold cross validation on the data, the formulated decision tree was able to correctly classified the model with about 93% accuracy measure. Confusion matrix in figure 3 and 4 shows that 775 out of 833 entire instances were classified correctly, indicating 93.04% approximately of the instances. While 58 instances were classified incorrectly, representing 6.96% of the entire instances.

**B. Decision Tree Model Formulation with Holdout Validation Result**

After formulating the model using J48 decision tree with K-fold cross validation method, the J48 decision tree was also evaluated using holdout validation method. However, this is done by dividing the entire inmates' COVID-19 data into two (train and test) in proportion of 70 percent to 30 percent. Test dataset was used to evaluate the model, which contain 250 instances representing the 30 percent of the data.

Correctly Classified Instances
Incorrectly Classified Instances

Figure 5: Classification report for J48 decision tree model with holdout validation

```

=== Confusion Matrix ===
      a  b  c  <-- classified as
    158  5  0 |  a = LOW
      10 69  1 |  b = MODERATE
       0  4  3 |  c = HIGH
    
```

Figure 6: Confusion matrix for J48 decision tree model with holdout validation

Using the holdout evaluation method as shown in figures 5 and 6 identifies that 230 instances was correctly classified out of 250 instances, while 20 instances were misclassified, representing 92% and 8% respectively in the testing dataset. Confusion matrix in figure 6 shows how the instances were correctly classified and misclassified as low, moderate and high.

**C. Decision Tree Model Formulation with 3-way Split Validation Result**

The 3-way split entails dividing dataset into training set, validation set, and testing set. The entire dataset was split into training, test and validation set. However, validation set was used to evaluate the J48 decision tree formulated model.

Correctly Classified Instances
Incorrectly Classified Instances

Figure 7: Classification report for J48 decision tree model with 3-way split validation

```

=== Confusion Matrix ===
      a  b  c  <-- classified as
     70  5  9 |  a = MODERATE
      8 111  1 |  b = LOW
       4  0 95 |  c = HIGH
    
```

Figure 8: Confusion matrix for J48 decision tree model with 3-way split validation

The 3-way split evaluation method as shown in figures 7 and 8 identifies that 267 instances was correctly classified out of 303 instances, while 27 instances were misclassified, representing 91.09% and 8.91% respectively in the validating dataset. Confusion matrix in figure 8 shows how the instances were correctly classified and misclassified as low, moderate and high.

The results of this study show performance of J48 decision tree with various evaluation methods. All these methods classified the prediction model with different accuracy measure.

TABLE 1: Summary of J48 decision tree accuracy measure with various evaluation methods

Evaluation Methods	ACCURACY (%)	TIME TAKEN (Seconds)
Holdout validation	92	0.02

K-Fold Cross Validation	93.04	≈ 0
3-Way Split Validation	91.09	0.01

Table 1 show the summary of model performance measure of J48 decision tree. It identifies that the model with K-fold cross validation have highest accuracy measure with 93.04 percent to outperform other evaluation methods, while 3-way split validation has the least 91.09 percent accuracy measure. Although, the model shows high performance measure in all the evaluation metric. Thus, decision tree classifier is suitable to correctly make prediction of infectious disease in the prisons.

#### V. CONCLUSION

The study concentrated on developing a predictive model which will predict the outbreak of infectious disease (COVID) using machine learning algorithm (Decision Tree). Various evaluation methods were adopted for the validation of this model, which include K-fold cross validation, holdout, and 3-way split validation. Based on the results of this study, decision tree with k-fold cross validation (10-fold cross validation) show highest performance metric with least time taken.

Furthermore, if proper public health precautions are not put in place, jails could turn into transmitters for COVID-19 spread to the community, unlike previous epidemics of other diseases that self-limited. According to [6] The isolation of all new inmates, contact tracking, and provision of a highly responsive testing regime to all inmates and prison staff, including prioritization for early adoption and execution of screening tests and tests, are all strategies that all prisons should take into consideration. However, keep in mind that many prisoners who are freed will end themselves in shelters or other problematic living conditions, necessitating planning and collaboration with public agencies. This study concluded that forecasting the spread of infectious diseases is essential and using of machine learning approach will aid proper predictions. Decision tree is found vital in predicting the possible infectious disease outbreak. Further investigation can still be made using larger dataset and other evaluation methods such as hyperparameter pruning, functionality reduction and so on, for future studies.

#### REFERENCES

[1] Alanazi, A., Alturki, A., Alosaimi, M., Alnajmi, N., Alrajeh, M., Kentab, A., & Habib, F. (2022). Clinical Characteristics and Risk Factors among Patients with Positive COVID-19 Test Admitted to ICU. *BioMed Research International*, 2022. <https://doi.org/10.1155/2022/1132399>

[2] Alghamdi, S. (2021). Clinical characteristics and treatment outcomes of severe (ICU) COVID-19 patients in Saudi Arabia: A single centre study. *Saudi Pharmaceutical Journal*, 29(10), 1096–1101. <https://doi.org/10.1016/j.jsps.2021.08.008>

[3] Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In M. W. Berry, A. Mohamed, & B. W. Yap (Eds.), *Supervised and Unsupervised Learning for Data Science* (pp. 3–21). Springer International Publishing. [https://doi.org/10.1007/978-3-030-22475-2\\_1](https://doi.org/10.1007/978-3-030-22475-2_1)

[4] Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press. <https://books.google.com/books?hl=en&lr=&id=tZnSDwAAQBAJ&oi=fnd&pg=PR7&dq=Introduction+to+machine+learning&ots=F3RWaXcwwf&sig=50DHyEjhVdDtmXlxZ0C4tXOGsdw>

[5] Arons, M. M., Hatfield, K. M., & Reddy, S. C. (2020). Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N Engl J Med* 2020;382:2081–90., 382(20), 81–90.

[6] Beaudry, G., Zhong, S., Whiting, D., Javid, B., Frater, J., & Fazel, S. (2020). Managing outbreaks of highly contagious diseases in prisons: A systematic review. *BMJ Global Health*, 5(11), e003201. <https://doi.org/10.1136/bmjgh-2020-003201>

[7] Faluyi, S. G., Adu, M. K., Balogun, T. E., & Fapohunda, K. (2021). Model to predict the classification risk of covid-19 using data mining techniques. 3rd International Conference, Center for Research, Innovation And. [https://scholar.google.com/citations?view\\_op=view\\_citation&hl=en&user=mQdtJD4AAAAJ&citation\\_for\\_view=mQdtJD4AAAAJ:UebtZRa9Y70C](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=mQdtJD4AAAAJ&citation_for_view=mQdtJD4AAAAJ:UebtZRa9Y70C)

[8] Fazel, S., & Baillargeon, J. (2011). The health of prisoners. *The Lancet*, 377(9769), 956–965. [https://doi.org/10.1016/S0140-6736\(10\)61053-7](https://doi.org/10.1016/S0140-6736(10)61053-7)

[9] Habehh, H., & Gohel, S. (2021). Machine Learning in Healthcare. *Current Genomics*, 22(4), 291–300. <https://doi.org/10.2174/1389202922666210705124359>

[10] Kong, K. A., Jung, S., Yu, M., Park, J., & Kang, I. S. (2021). Association Between Cardiovascular Risk Factors and the Severity of Coronavirus Disease 2019: Nationwide Epidemiological Study in Korea. *Frontiers in Cardiovascular Medicine*, 8. <https://www.frontiersin.org/articles/10.3389/fcvm.2021.732518>

[11] Kumar, V., & L., M. (2018). Predictive Analytics: A Review of Trends and Techniques. *International Journal of Computer Applications*, 182, 31–37. <https://doi.org/10.5120/ijca2018917434>

[12] Njuguna, H. (2020). Serial Laboratory Testing for SARS-CoV-2 Infection Among Incarcerated and Detained Persons in a Correctional and Detention Facility—Louisiana, April–May 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69. <https://doi.org/10.15585/mmwr.mm6926e2>

[13] Oladeji, F. A., Idowu, P. A., Egejuru, N., Faluyi, S. G., & Balogun, J. A. (2019). Model for predicting the risk of kidney stone using data mining techniques. *International Journal of Computer Applications*.

[14] Samuel, A. L. (2000). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 44(1.2), 206–226. <https://doi.org/10.1147/rd.441.0206>

[15] Shugalo, I. (2022). Predictive Modeling in Healthcare: Applications & Adoption. <https://www.itransition.com/blog/predictive-modeling-in-healthcare>

- [16] Simooya, O. O. (2010). Infections in Prison in Low and Middle Income Countries: Prevalence and Prevention Strategies. *The Open Infectious Diseases Journal*, 4(1), 33–37. <https://doi.org/10.2174/1874279301004010033>
- [17] Sreeja. (2021, May 10). Hold-out vs. Cross-validation. Medium. <https://sreeja123ut.medium.com/hold-out-vs-cross-validation-75597d8410b0>
- [18] Su, L. (2019, May 18). 3-way (train, cross validate, test) split is used to help us determine what model would work best. Medium. [https://medium.com/@lily\\_su/3-way-train-cross-validate-test-split-is-used-to-help-us-determine-what-model-would-work-best-5502047d14f0](https://medium.com/@lily_su/3-way-train-cross-validate-test-split-is-used-to-help-us-determine-what-model-would-work-best-5502047d14f0)
- [19] Wallace, M. (2020). Public Health Response to COVID-19 Cases in Correctional and Detention Facilities—Louisiana, March–April 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69. <https://doi.org/10.15585/mmwr.mm6919e3>
- [20] World Health Organization. (2020). Preparedness, prevention, and control of COVID-19 in prisons and other places of detention. <https://www.euro.who.int>
- [21] Yang, H., & Thompson, J. R. (2020). Fighting COVID-19 outbreaks in prisons. *BMJ*, m1362. <https://doi.org/10.1136/bmj.m1362>