JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

Infectious Disease Prediction, Testing Suggestionfor Better Operational Health Careusing Machine Learning Algorithms

Dr.Bandaru Srinivasa Rao
Department of Management Studies
Vignan's Foundation for Science Technology and Research
drbsrprofessormmt10672@gmail.com

***Dr. Sujatha Kamepalli**
Department of Information Technology
Vignan's Foundation for Science Technology and Research
sujatha_kamepalli@yahoo.com

Dr. V. MadhuLatha
Department of Business Management
VR Siddhartha Engineering College
madhulathav@vrsiddhartha.ac.in

*Corresponding Author

ABSTRACT
People are suffering from different types of infectious diseases such as malaria, dengue, corona, chikungunya,etc. "Pathogenic microorganisms such as bacteria, viruses, parasites, and fungi are the primary cause of infectious diseases".For all these infectious diseases the main symptom is fever, and people suffer from other symptoms such as Dehydration, Chills, Joint Pain, Muscle Pain, Head Ache, Joint Swelling, Rash, Vomiting, Cough, Shortness of Breathing, Chest Pain, Runny and Watery Nose, Loss of Appetite, etc. Medical practitioners are suggesting various tests to confirm the type of disease. If people are aware of which type of test is necessary to confirm the disease typedepending on the symptoms or signs of the patient and domain knowledge, then the testing expenditures can be reduced for the patients. So, in this paper,a hybrid model thatcombines classification and association analysis was developed that predicts the type of infection based on symptoms of patients and also suggests the test to be performed to confirm the disease. Naïve Bayes, SMO, and Apriori algorithms were implemented in the hybrid model. All models performed equally well with 92% of prediction accuracy. The model finds various associations among symptoms, disease, and tests performed which can be used for the testing suggestion. Further, the model can be extended to implement it in real-time by developing a web application.

Key Words: Infectious diseases, Machine learning algorithms, Apriori, Naïve Bayes, Medical Test.

Introduction:
In the present scenarios, the people are suffering from various infective diseases such as chikungunya, corona, dengue, malaria, rheumatic fever, swine flu, typhoid, zika, and viral fever, etc. Infectious diseases are "diseases caused by pathogenic microorganisms such as bacteria, viruses, parasites, and fungi that can transfer from one person to another either directly or indirectly, according to the World Health Organization".Infectious sickness occurs when a person or animal is infected with a pathogen from another person or animal. It causes harm not only to individuals, but also to society as a whole, and is thus classified as a social issue[1].Fever is a common symptom of a variety of illnesses, particularly infectious disorders. "Dengue fever, typhoid fever, diarrhea, gastroenteritis, measles, pneumonia, pharyngitis, and bronchitis" are just a few of the diseases that start with a fever[2]. In addition to fever, a lot of symptoms are observed in identifying these diseases. In this paper, we analyzedthe symptoms with which patients are suffering from various infectious diseases.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

"Machine learning" can be defined as a branch of science that studies algorithms that can learn from data. Algorithms are a critical component of machine learning. Representative data is an important part of machine learning that must be produced. To put it another way, machine learning only evaluates a certain input data set that has previously been created as parameters for completing a task [4]. The medical data like symptoms of a disease, and laboratorytest data can be taken as input by the systems, and using machine learning approaches generates some results. Based on the accuracy of the result, the system decides on training and testing data[5].Machine learning algorithms[6][7] were used in analyzing the symptoms and predicting the disease. All infectious diseases cannot be confirmed with the same type of test. Based on the predicted disease the system recommends the required testing.

In the first section of the paper, we give an introduction to infectious diseases. In the section, we explained the advanced models used in infectious disease prediction. The third section gives the objectives of the paper. In the fourth section, the developed hybrid model is explained. In the fifth section, the experimental results were explored. Finally, in the last section, we concluded the paper.

Literature Survey:

[1]This research optimizes the parameters of deep learning algorithms while taking into account large data, including social media data, to anticipate infectious diseases. When predicting three infectious diseases one week in the future, the performance of the deep neural network (DNN) and long-short term memory (LSTM) learning models were compared to the autoregressive integrated moving average (ARIMA). The DNN and LSTM models outperform ARIMA, according to the results.[2]One classification technique, Fuzzy K-Nearest Neighbour, evaluates the distance between training and testing data before grouping them into a fuzzy set. Based on textual medical record data, this study created a categorization system for children's disorders with fever using Fuzzy K-Nearest Neighbour. The classification system's test results indicated an accuracy of 83.3 percent in dengue fever and pneumonia data, with an 80:20 ratio of training to testing data, a K value of 10, and an M value of 2. As a result, it can be inferred that the Fuzzy K-Nearest Neighbour classification technique can be utilized to classify childhood disorders that cause fever.[8]This paper aims to use healthcare big data analysis combined with deep learning technology to provide patients with potential diseases that are often overlooked due to a lack of professional knowledge, allowing patients to conduct targeted medical examinations and prevent their health conditions from worsening. This study offers a novel deep-learning-based hybrid recommendation algorithm dubbed medical-history-based prospective disease prediction algorithm, which is inspired by existing recommendation methods. The algorithm forecasts the patient's likely ailment based on their medical history, giving patients and clinicians a point of reference to avoid postponing treatment owing to a symptom's vague description or a lack of professional understanding. The results of the experiments suggest that our method improves the accuracy of predicting probable diseases.[9]Using the machine's ability to process data, we proposed a model for patient symptom similarity analysis in this work. The program extracted crucial information from patient accounts of symptoms to achieve early prediction and intervention. As a result, the effectiveness of disease prediction is largely determined by the accuracy of the similarity analysis model.[10]To prevent misdiagnosis, this study employs a clinical decision support system (CDSS) based on medical record data and categorization approaches. This research focuses on the diagnosis of typhoid fever disease (TFD), which has a high rate of antibiotic resistance and is a difficult-to-diagnose endemic disease in Indonesia. This system employs three supervisory categorization methods: Nave Bayes (NB), k-nearest neighbor (kNN), and Support Vector Machine (SVM) (SVM). The system was then compared and assessed. The findings of the three approaches revealed that KNN had an accuracy of 88.7%, which was higher than previous typhoid fever disease classification methods.[11]For effective disease prediction, the authors employ the K-Nearest Neighbor (KNN) and Convolutional Neural Network (CNN) machine learning algorithms. A collection of disease symptoms is required for disease prediction. The person's living habits and checkup information are taken into account in this general disease prognosis. The CNN algorithm has an accuracy of 84.5 percent in general disease prediction, which is higher than the KNN approach. In addition, KNN has a higher time and memory need than CNN. After predicting general disease, this system may calculate the risk of general disease, indicating whether the risk is lower or higher.[12]The goal of this study is to find the best Backpropagation Algorithm to solve the problem by analyzingthe early detection of DHF and optimizing Multi-Layer Perceptron (MLP) through five different back-propagation training algorithms: Gradient Descent (GD), BFGS Quasi-Newton (BQN), Conjugate Gradient Descent - Powel (CGD), Resilient Backpropagation (RB), and Levenberg Marquardt (LM). The early symptoms of DHF patients were included as variables in this study, which included up to ten variables. All research data was gathered from a total of 76 medical records at RSUP Dr.Karyadi

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

Semarang. The Levenberg Marquardt algorithm performed best in the detection of dengue illness, according to the findings.[13]In the medical field, several classification algorithms are used to forecast various diseases such as diabetes, TB, and so on. Diagnosis of diabetes can be determined by comparing the patient's blood sugar levels to normal levels, as well as blood pressure, BMI, skin thickness, and other factors. For diabetes, several classification systems have been used. The major goal of this research is to develop a statistical model for diabetes data to improve classification accuracy. The principal component analysis is used to project extracted characteristics from diabetic data into a new space, which is then analyzed using the linear regression approach on these newly generated attributes. This method has an accuracy of 82.1 percent for predicting diabetes, which is superior to other existing classification methods.

Objectives:

The following are the key goals of this paper:

- To study the advanced models in predicting infectious diseases.
- To develop a hybrid model to predict the type of disease based on symptoms and testing suggestions.
- To implement the proposed model on the collected datasets.
- To investigate the performance of the proposed hybrid model.

Model for Disease Prediction and Testing Suggestion:

Here we developed a hybrid model to predict the disease and suggest the test to be performed depending on symptoms. Figure 1 represents the flowchart of the proposed hybrid model.

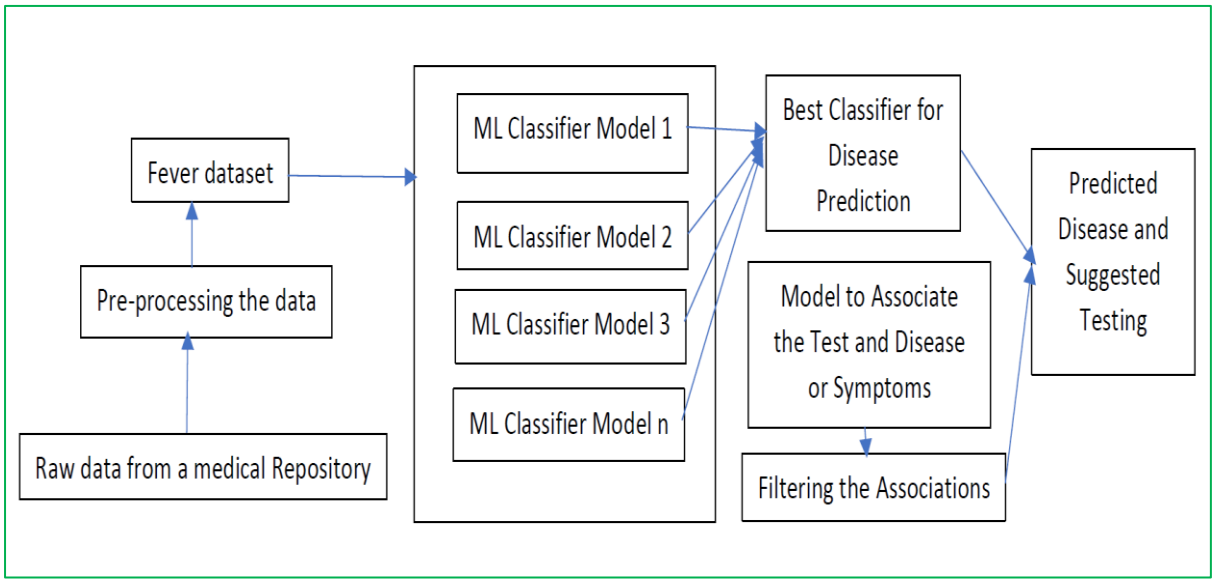

Fig. 1 Flow Chart of Proposed Model
Source: Proposed Model

We collected data from a medical repository, the data was pre-processed to avoid noise. The dataset is divided into two parts 1. The first dataset contains various symptoms and corresponding diseases. 2. the Second dataset contains symptoms, disease, and tests performed. By applying Machine Learning classifier models tothe first dataset we can predict the disease based on symptoms. The second dataset is used to find the associations between symptoms, disease, and tests performed.

Experimentation:

The experiments were conducted using the simulation tool. We used Naïve Bayes, SMO, and J48 Machine Learning Classifier Models in predicting the disease based on symptoms. Also, we used the Apriori algorithm to find the associations between symptoms, disease, and testsperformed.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

Naive Bayes Classifier: The Bayes classifier is "a probabilistic machine learning model for classification. The Bayes theorem lies at the heart of the classifier."Mathematically the Bayes theorem can be represented as below.

$$P\left(\frac{A}{B}\right) = \frac{P(B/A)P(A)}{P(B)}$$

Using Bayes' theorem, we may determine the odds of A occurring if B has previously occurred. The hypothesis is A, and the proof is B. In this situation, the predictors/features are supposed to be independent. To put it another way, the presence of one attribute does not influence the presence of the other. As a result, it's called naive.

SMO:SMO stands for Sequential Minimal Optimization and refers to the unique efficient optimization approach employed within the SVM implementation. For training, a support vector classifier implements John Platt's sequential minimal optimization approach. The Support Vector Machine (SVM) is "a supervised machine learning technique that can solve classification and regression problems". It is, however, mostly used to tackle categorization problems. Each data item is represented as a point in n-dimensional space (where n is the number of features), with the value of each feature being the SVM algorithm's value for a certain coordinate. Then we locate the hyper-plane that separates the two classes to complete categorization. (look at the below snapshot).

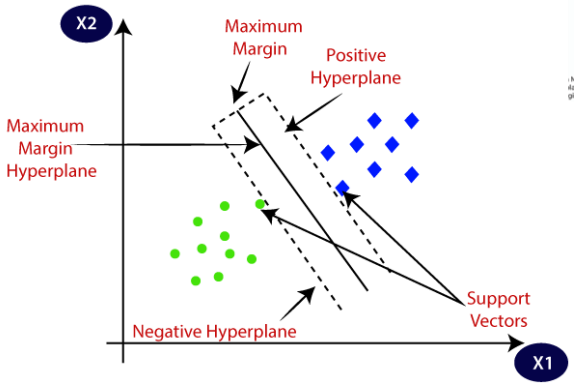

Fig 2: Snapshot of Support Vector Machines
Source: https://static.javatpoint.com/tutorial/machine-learning/images/support-vector-machine-algorithm.png
Simply put, support vectors are the coordinates of each observation. The SVM classifier is a frontier that separates the two classes (hyper-plane/line) the most effective.

J48 Pruned Tree: The J48 algorithm is "one of the greatest machine learning algorithms for categorizing and continuously examining data. C4.5 (J48) is a decision tree-generation method created by Ross Quinlan, as previously reported. Quinlan's earlier ID3 algorithm is expanded in C4.5. C4.5 generates decision trees that can be used for classification, which is why it's commonly referred to as a statistical classifier".

Apriori: The apriori algorithm is a set of procedures for determining the most often occurring item set in a database. The join and prune methods are repeated until the most often occurring item set is discovered using this data mining technique. The problem specifies or the user assumes a minimum support threshold.

Datasets Description:
The raw data collected from a medical repository is divided into two datasets.The first dataset contains various symptoms identified in the patient and the disease with which that particular patient is suffering. The following figure depicts the sample dataset1.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

P ID	DEHYDRA	CHILLS	FEVER	JOINT PAI	MUSCLE P	HEAD ACH	JOINT SW	RASH	VOMITIM	COUGH	SHARTNE	CHEST PAI	BLUE LIPS	INTERNAL	BLOODY S	PROFUSE	ABDOMIN	SWALLEN	SWALLEN	RED SPOT	DIFFICULT	RUNNY AN	LOSS OF A	CLASS
1	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CHICKEN GUNYA
2	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	CORONA
3	1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	DENGUE
4	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	MALARIA
5	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	Rheumatic fever
6	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	SWINE FLU
7	1	0	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	TYPHOID
8	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	VIRAL FEVER
9	1	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ZIKA
10	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CHICKEN GUNYA
11	1	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	CORONA
12	1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	DENGUE
13	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	MALARIA
14	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	Rheumatic fever
15	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	SWINE FLU
16	1	0	1	1	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	TYPHOID
17	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	VIRAL FEVER
18	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ZIKA
19	1	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CHICKEN GUNYA
20	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	CORONA
21	1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	DENGUE
22	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	MALARIA
23	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	Rheumatic fever
24	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	SWINE FLU
25	1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	TYPHOID
26	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	VIRAL FEVER
27	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ZIKA
28	1	0	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	TYPHOID
29	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	VIRAL FEVER
30	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	ZIKA
31	1	0	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	CHICKEN GUNYA
32	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	CORONA

Fig 3. Sample Dataset1
Source: From a Medical Repository (Details are not revealed for privacy reasons)

The succeedinggraphshows the visualization of various attributes and their values

Fig. 4 Visualization of Attributes in Dataset1
Source: Experimental setup
The following are various attributes and their range of values in dataset1.
Table 1. Attribute description of dataset1
Source: From a Medical Repository (Details are not revealed for privacy reasons)

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

S. No.	Attribute	Range of Values
1	P Id	Takes unique values
2	Dehydration	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
3	Chills	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
4	Fever	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
5	Joint Pain	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
6	Muscle Pain	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
7	Head Ache	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
8	Joint Swelling	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
9	Rash	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
10	Vomitimg	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
11	Cough	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
12	Shartness of Breathing	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
13	Chest Pain	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
14	Blue Lips and Eyes	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
15	Internal Bleeding	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
16	Bloody Stool	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
17	Profuse Sweating	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
18	Abdominal Pain	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
19	Swallen Lymph Nodes	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
20	Swallen and Red Tansils	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
21	Red Spots on Mouth Roof	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
22	Difficulty in Swallowing	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
23	Runny and Watery Nose	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
24	Loss of Apatite	0 or 1(0 indicates absence of the symptom and 1 indicates presence of the symptom)
25	Class (Disease)	Takes 9 values (chickengunya, corona, dengue, malaria, rheumatic fever, swine flu, typhoid, zika, viral fever)

The following figure depicts the sample dataset2.

	SYMPTOM	DISEASE	TESTS PERFOMED
2	chills,swea	viral fever	Blood test
3	high fever,	Dengue fe	Blood test
4	high fever,	Dengue fe	Blood test
5	Chills,Cou	swine flu	Blood test
6	Fever (but	swine flu	Chest X-rays, Blood test, CBC test
7	chills,swea	viral fever	Blood test
8	chills,swea	viral fever	Blood test
9	fever and j	chikungun	Blood test
10	high fever,	Dengue fe	Blood test
11	fever and j	chikungun	Blood test
12	a rash, itch	zika	blood test, urine tests
13	high fever,	chikungun	Blood test
14	high fever,	Dengue fe	Blood test
15	shortness	corona	blood, saliva
16	Chills,Cou	swine flu	Blood test
17	Fever (but	swine flu	Chest X-rays, Blood test, CBC test
18	a rash, itch	zika	blood test, urine tests
19	a rash, itch	zika	blood test, urine tests

Fig 5. Sample Dataset2
Source: From a Medical Repository (Details are not revealed for privacy reasons)

The following are the various attributes and their range of values in dataset2.
Table 2. Attribute description of dataset2
Source: From a Medical Repository (Details are not revealed for privacy reasons)

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

S. No.	Attribute	Range of Values
1	Symptoms	Symptoms of various patients
2	Disease	Takes 9 values (chikungunya, corona, dengue, malaria, rheumatic fever, swine flu, typhoid, zika, viral fever)
3	Test performed	Blood test, urine test, chest x-ray, saliva, CBC test, etc.

The givenfigureshows the visualization of several attributes and their values in the given dataset.

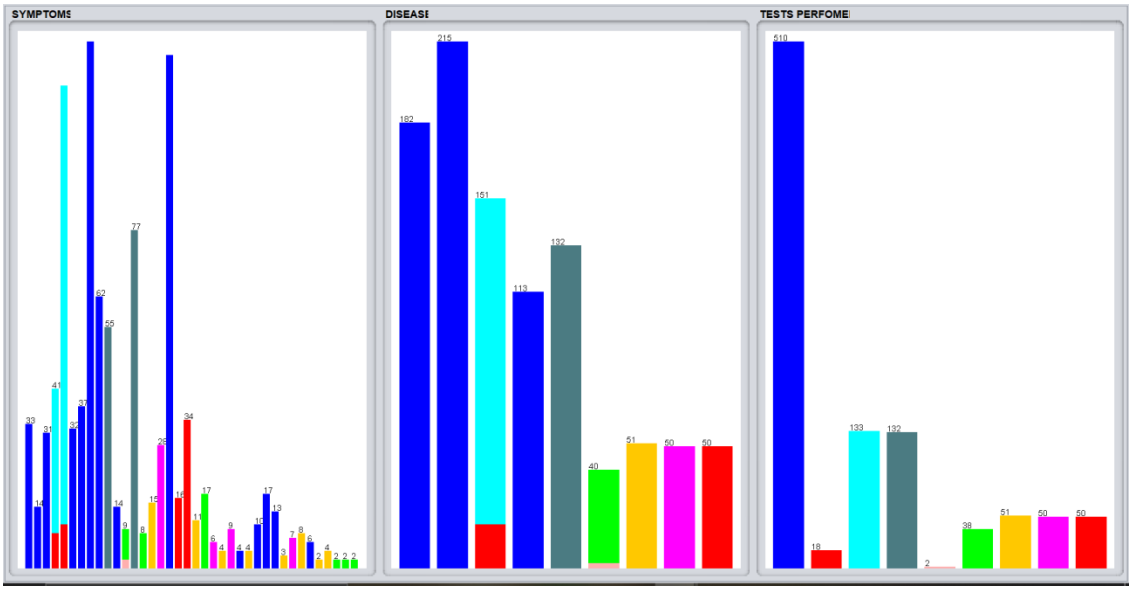

Fig. 6 Visualization of attributes in dataset2
Source: experimental setup

By applying classifier models to the first dataset, the disease can be predicted. By applying Apriorito the second dataset the test can be suggested based on symptoms and disease predicted.

Results and Discussions:
Dataset1 was divided into training and test sets, with 80% of data considered under the training set and 20% of data considered under the test set. Three classifier models Naïve Bayes, SMO, and J48 were implemented in predicting the disease based on symptoms.

Table. 3 Comparison of Classifier models in Disease Prediction on the Training set

S. No.	Classifier model	Correctly classified	Incorrectly classified	TP Rate	FP Rate	Precision	Recall
1	Naïve Bayes	710	73	0.907	0.012	0.938	0.907
2	SMO	710	73	0.907	0.012	0.938	0.907
3	J48	710	73	0.907	0.012	0.938	0.907

Table. 4 Comparison of Classifier models in Disease Prediction on the Test set

S. No.	Classifier model	Correctly classified	Incorrectly classified	TP Rate	FP Rate	Precision	Recall
1	Naïve	184	17	0.915	0.011	0.939	0.915

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

	Bayes						
2	SMO	184	17	0.915	0.011	0.939	0.915
3	J48	184	17	0.915	0.011	0.939	0.915

Three models perform equally well with 90.67695 accuracies on the training set and 91.5423% of accuracy on the test set.

From the above results, it is clear that the disease prediction can be done with an accuracy of 92%. After predicting the disease, what the test to be performed to confirm whether that particular patient is suffering from the predicted test or not. Apriori algorithm is used to know the suggested test for a particular disease. The following shows results obtained from the implementation of Apriori.

"Minimum support: 0.1 (98 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18"

Generated Sets of Large Itemsets:
Size of the set of large itemsetsL(1): 11

Large ItemsetsL(1):
The following are the frequent 1- itemsets generated from the given dataset.
"SYMPTOMS=Fever (but not always), Chills,Cough,Sore throat,Runny or stuffy nose, watery, red eyes,Body aches,Headache. 110
SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mild bleeding 120
SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. 117
DISEASE=viral fever 182
DISEASE=Dengue fever 215
DISEASE=swine flu 151
DISEASE=chikungunya 113
DISEASE=zika 132
TESTS PERFOMED=Blood test 510
TESTS PERFOMED=Chest X-rays, Blood test, CBC test 133
TESTS PERFOMED=blood test, urine tests 132"

Size of the set of large itemsetsL(2): 11
Large ItemsetsL(2):
The following are the frequent 2- itemsets generated from the given dataset.

"SYMPTOMS=Fever (but not always), Chills,Cough,Sore throat,Runny or stuffy nose, watery, red eyes,Body aches,Headache. DISEASE=swine flu 110
SYMPTOMS=Fever (but not always), Chills,Cough,Sore throat,Runny or stuffy nose, watery, red eyes,Body aches,Headache. TESTS PERFOMED=Chest X-rays, Blood test, CBC test 100
SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mild bleeding DISEASE=Dengue fever 120
SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mild bleeding TESTS PERFOMED=Blood test 120
SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. DISEASE=viral fever 117
SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. TESTS PERFOMED=Blood test 117

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 2, 2022, p. 2729 - 2739
https://publishoa.com
ISSN: 1309-3452

DISEASE=viral fever TESTS PERFOMED=Blood test 182
DISEASE=Dengue fever TESTS PERFOMED=Blood test 215
DISEASE=swine flu TESTS PERFOMED=Chest X-rays, Blood test, CBC test 133
DISEASE=chikungunya TESTS PERFOMED=Blood test 113
DISEASE=zika TESTS PERFOMED=blood test, urine tests 132"

Size of the set of large itemsetsL(3): 3
Large ItemsetsL(3):
The following are the frequent 3- itemsets generated from the given dataset.

"SYMPTOMS=Fever (but not always), Chills,Cough,Sore throat,Runny or stuffy nose,watery, red eyes,Body aches,Headache. DISEASE=swine flu TESTS PERFOMED=Chest X-rays, Blood test, CBC test 100
SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mild bleeding DISEASE=Dengue fever TESTS PERFOMED=Blood test 120
SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. DISEASE=viral fever TESTS PERFOMED=Blood test 117"

Best Association Rules Found:
The following shows the best association rules found from the given dataset.

"1. DISEASE=Dengue fever 215 ==> TESTS PERFOMED=Blood test 215 <conf:(1)> lift:(1.93) lev:(0.11) [103] conv:(103.57)
2. DISEASE=viral fever 182 ==> TESTS PERFOMED=Blood test 182 <conf:(1)> lift:(1.93) lev:(0.09) [87] conv:(87.67)
3. TESTS PERFOMED=Chest X-rays, Blood test, CBC test 133 ==> DISEASE=swine flu 133 <conf:(1)> lift:(6.52) lev:(0.11) [112] conv:(112.59)
4. TESTS PERFOMED=blood test, urine tests 132 ==> DISEASE=zika 132 <conf:(1)> lift:(7.45) lev:(0.12) [114] conv:(114.29)
5. DISEASE=zika 132 ==> TESTS PERFOMED=blood test, urine tests 132 <conf:(1)> lift:(7.45) lev:(0.12) [114] conv:(114.29)
6. SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mildbleading 120 ==> DISEASE=Dengue fever 120 <conf:(1)> lift:(4.58) lev:(0.1) [93] conv:(93.78)
7. SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mildbleading 120 ==> TESTS PERFOMED=Blood test 120 <conf:(1)> lift:(1.93) lev:(0.06) [57] conv:(57.8)
8. SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mildbleading TESTS PERFOMED=Blood test 120 ==> DISEASE=Dengue fever 120 <conf:(1)> lift:(4.58) lev:(0.1) [93] conv:(93.78)
9. SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mildbleading DISEASE=Dengue fever 120 ==> TESTS PERFOMED=Blood test 120 <conf:(1)> lift:(1.93) lev:(0.06) [57] conv:(57.8)
10. SYMPTOMS=high fever, severe headache, pain behind eyes, severe joint and muscle pains, vomiting, skin rash,mildbleading 120 ==> DISEASE=Dengue fever TESTS PERFOMED=Blood test 120 <conf:(1)> lift:(4.58) lev:(0.1) [93] conv:(93.78)
11. SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. 117 ==> DISEASE=viral fever 117 <conf:(1)> lift:(5.41) lev:(0.1) [95] conv:(95.36)
12. SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. 117 ==> TESTS PERFOMED=Blood test 117 <conf:(1)> lift:(1.93) lev:(0.06) [56] conv:(56.36)
13. SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. TESTS PERFOMED=Blood test 117 ==> DISEASE=viral fever 117 <conf:(1)> lift:(5.41) lev:(0.1) [95] conv:(95.36)

14. SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. DISEASE=viral fever 117 ==> TESTS PERFOMED=Blood test 117 <conf:(1)> lift:(1.93) lev:(0.06) [56] conv:(56.36)
15. SYMPTOMS=chills,sweating, dehydration, headache, muscle aches and pains, a feeling of weakness, loss of appetite. 117 ==> DISEASE=viral fever TESTS PERFOMED=Blood test 117 <conf:(1)> lift:(5.41) lev:(0.1) [95] conv:(95.36)
16. DISEASE=chikungunya 113 ==> TESTS PERFOMED=Blood test 113 <conf:(1)> lift:(1.93) lev:(0.06) [54] conv:(54.43)
17. SYMPTOMS=Fever (but not always), Chills,Cough,Sorethroat,Runny or stuffy nose,Watery, red eyes,Bodyaches,Headache. 110 ==> DISEASE=swine flu 110 <conf:(1)> lift:(6.52) lev:(0.09) [93] conv:(93.12)
18. SYMPTOMS=Fever (but not always), Chills,Cough,Sorethroat,Runny or stuffy nose,Watery, red eyes,Bodyaches,Headache. TESTS PERFOMED=Chest X-rays, Blood test, CBC test 100 ==> DISEASE=swine flu 100 <conf:(1)> lift:(6.52) lev:(0.09) [84] conv:(84.65)
19. SYMPTOMS=Fever (but not always), Chills,Cough,Sorethroat,Runny or stuffy nose,Watery, red eyes,Bodyaches,Headache. 110 ==> TESTS PERFOMED=Chest X-rays, Blood test, CBC test 100 <conf:(0.91)> lift:(6.73) lev:(0.09) [85] conv:(8.65)
20. SYMPTOMS=Fever (but not always), Chills,Cough,Sorethroat,Runny or stuffy nose,Watery, red eyes,Bodyaches,Headache. DISEASE=swine flu 110 ==> TESTS PERFOMED=Chest X-rays, Blood test, CBC test 100 <conf:(0.91)> lift:(6.73) lev:(0.09) [85] conv:(8.65)"

Observing the association rules, based on the symptoms shows the type of disease and also predicts the type of test required to confirm the disease.

Conclusion:

When a person is infected by a pathogen from another person or animal, an infectious disease develops.Those diseases can be identified by various symptoms. In this paper, we analyzed the symptoms with which patients are suffering from various infectious diseases and a hybrid model was developed to predict the type of disease based on symptoms and also suggesting the required test for confirming the predicted disease. We collected data from a medical repository, the data was pre-processed to avoid noise. The dataset is divided into two parts 1. The first dataset contains various symptoms and corresponding diseases. 2. the Second dataset contains symptoms, disease, and tests performed.The experiments were conducted using the simulation tool. We used Naïve Bayes, SMO, and J48 Machine Learning Classifier Models in predicting the disease based on symptoms. Also, we used the Apriori algorithm to find the associations between symptoms, disease, and tests performed. All models performed equally well with 92% of prediction accuracy. The model finds various associations among symptoms, disease and tests performed which can be used for the testing suggestion. Further, the model can be extended to implement it in real-time.

References:

[1] S. Chae, S. Kwon, and D. Lee, "Predicting infectious disease using deep learning and big data," *Int. J. Environ. Res. Public Health*, vol. 15, no. 8, pp. 1–20, 2018, doi: 10.3390/ijerph15081596.

[2] R. K. Putra and S. Mulyati, "Classification of childhood diseases with fever using fuzzy K-nearest neighbor method," in *2018 International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2018*, 2018, pp. 332–337, doi: 10.1109/ISRITI.2018.8864475.

[3] K. Sujatha, K. V. K. Kishore, B. S. Rao, G. District, and A. Pradesh, "Machine learning models for Forecasting Confirmed , Recovered and Deceased COVID-19 Cases in India," *Int. J. Control Autom.*, vol. 13, no. 4, pp. 841–854, 2020.

[4] Y. Mulyani, E. F. Rahman, Herbert, and L. S. Riza, "A new approach on prediction of fever disease by using a combination of Dempster Shafer and Naïve bayes," in *Proceeding - 2016 2nd International Conference on Science in Information Technology, ICSITech 2016: Information Science for Green Society and Environment*, 2017, pp. 367–371, doi: 10.1109/ICSITech.2016.7852664.

[5] K. Sujatha and B. Srinivasa Rao, "Recent Applications of Machine Learning : A Survey," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 6, pp. 263–267, 2019, doi: Retrieval Number: F10510486C219 /19©BEIESP.

[6] S. Kamepalli, Sujatha. Bandaru, "Symptoms and Spreading of COVID-19 in India and Predicting COVID Infected Patients Based on Symptoms," *J. Xi'an Univ. Archit. Technol.*, vol. XII, no. V, pp. 3446–3457, 2020.

[7] K. Sujatha, "Early Prediction and Analysis of Caesarian Section Delivery for Better Healthcare Operational Decisions," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 3, pp. 10789–10799, 2020.

[8] W. Hong, Z. Xiong, N. Zheng, and Y. Weng, "A Medical-History-Based Potential Disease Prediction Algorithm," *IEEE Access*, vol. 7, pp. 131094–131101, 2019, doi: 10.1109/ACCESS.2019.2940644.

[9] P. Zhang, X. Huang, and M. Li, "Disease Prediction and Early Intervention System Based on Symptom Similarity Analysis," *IEEE Access*, vol. 7, pp. 176484–176494, 2019, doi: 10.1109/ACCESS.2019.2957816.

[10] B. Andrianto, Y. K. Suprapto, I. Pratomo, and I. Irawati, "Clinical decision support system for typhoid fever disease using classification techniques," in *Proceedings - 2019 International Seminar on Intelligent Technology and Its Application, ISITIA 2019*, 2019, pp. 248–252, doi: 10.1109/ISITIA.2019.8937286.

[11] D. Dahiwade, G. Patle, and E. Meshram, "Designing disease prediction model using machine learning approach," in *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019*, 2019, no. Iccmc, pp. 1211–1215, doi: 10.1109/ICCMC.2019.8819782.

[12] P. S. Sasongko, H. A. Wibawa, F. Maulana, and N. Bahtiar, "Performance comparison of Artificial Neural Network models for dengue fever disease detection," in *Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017*, 2018, vol. 2018-Janua, pp. 183–187, doi: 10.1109/ICICOS.2017.8276359.

[13] H. Roopa and T. Asha, "A Linear Model Based on Principal Component Analysis for Disease Prediction," *IEEE Access*, vol. 7, pp. 105314–105318, 2019, doi: 10.1109/access.2019.2931956.