Improve diagnostic of typhoid through open science

Proposed model

By

Antima Dwivedi

Abstract

WHO estimates the global typhoid fever disease burden at 11-20 million cases annually, resulting in about 128 000–161 000 deaths per year. Typhoid risk is higher in populations that lack access to safe water and adequate sanitation. Poor communities and vulnerable groups including children are at the highest risk. The main aim of the Project is to improve diagnostics of typhoid through Open-science. Combining AI systems with an irreplaceable human clinician can advance better diagnosis.

Proposed Solution

1. The basic idea is to develop an application(android or web) and use the features of a mobile phone for computation and reducing human power.
2. Typhoid fever will be tested in two phases.

2.1 Basic testing:

According to the reference [1] there are 18 symptoms of enteric fever. Firstly application will produce the results according to the given rough set algorithm mentioned in the paper and if tested positive patient will go through the next stage which is widal testing.

2.1.1The methodology for Basic Testing:

There are eighteen (18) conditional attributes (symptoms) and one (1) decision attribute (level of severity), shown in table 1 and table 2 below respectively. All the attributes were discretized.

RoughSet Representation	Symbol	Attribute (Symptom)	Attribute Type
A1	FVR	Fever	Discrete
A2	ABP	Abdominal Pain	Discrete
A3	COH	Cough	Discrete
A4	DIA	Diarrhoea	Discrete
A5	CON	Constipation	Discrete

A6	RPT	Rose spot	Discrete
A7	MWK	Muscle Weakness	Discrete
A8	ANR	Anorexia	Discrete
A9	HDH	Headache	Discrete
A10	SKR	Skin rash	Discrete
A11	WTL	Weightless	Discrete
A12	SMD	Stomach Distension	Discrete
A13	MAL	Malaise	Discrete
A14	OBS	Occult blood in the stool	Discrete
A15	HMR	Haemorrhages	Discrete
A16	DEM	Derilium	Discrete
A17	ABR	Abdominal rigidity	Discrete
A18	EPS	Epistaxis (Bloody nose)	Discrete

Table 1: Conditional Attributes of Typhoid Fever

Rough Set Representation	**Symbol**	**Attribute (Symptom)**	**Attribute type**
Dec	Dec	Typhoid Fever diagnosed	Discrete

Table 2: Decision Attribute of Typhoid Fever

2.1.2 Experimental result:

There are eighteen conditional attributes(symbols) A1 to A18 and one decision attribute(Typhoid Fever diagnosed) Dec. Each conditional attribute can take a value from High, Low, or Default, depending on the patient's feeling. Default exists for symptoms not perceived. For the decision attribute, there are five classes- Very Low, Low, Moderate, High, and Very High. To make the programming easier and the program more efficient, these values were converted to integers since it is easier to work around numbers. The conditional variable

Low was converted to 1 and the conditional variable High was converted to 2 while default takes the value 0. For example, A1 = 2 means Fever is High and A3 = 1 means Cough is Low. For the decision attribute, Very Low, Low, High, Moderate, High, and Very High were converted to 1,2,3,4 and 5 respectively. In this case, Dec =1 means typhoid fever diagnosed is Very Low while Dec = 5, means typhoid fever diagnosed is Very High.

Rule No	**Details of the rule**
Rule 1	(A7 = 1) & (A8 = 1) & (A9 = 1) & (A10 = 1) & (A11 = 1) & (A12 = 1) & (A15 = 1) & (A16 = 1) => (Dec = 1)
Rule 2	(A8 = 1) & (A10 = 1) & (A11 = 2) & (A15 = 1) => (Dec = 2)
Rule 3	(A4 = 1) & (A5 = 1) & (A7 = 1) & (A9 = 2) & (A13 = 1) => (Dec = 2)
Rule 4	(A4 = 1) & (A7 = 1) & (A8 = 2) & (A9 = 1) & (A11 = 1) & (A18 = 1) => (Dec = 2)
Rule 5	(A7 = 1) & (A8 = 1) & (A9 = 2) & (A10 = 1) & (A13 = 1) & (A14 = 1) => (Dec = 2)
Rule 6	(A7 = 2) & (A14 = 1) & (A15 = 1) & (A18 = 1) => (Dec = 3)
Rule 7	(A4 = 1) & (A9 = 1) & (A12 = 2) => (Dec = 3)
Rule 8	(A10 = 2) & (A14 = 1) & (A15 = 1) & (A18 = 1) => (Dec = 3)
Rule 9	(A13 = 2) & (A14 = 1) & (A15 = 1) & (A18 = 1) => (Dec = 3)
Rule 10	(A4 = 2) & (A8 = 2) & (A15 = 1) => (Dec = 3)
Rule 11	(A7 = 2) & (A15 = 2) & (A16 = 1) => (Dec = 4)
Rule 12	(A9 = 2) & (A15 = 2) & (A16 = 1) => (Dec = 4)
Rule 13	(A5 = 1) & (A14 = 2) & (A16 = 1) => (Dec = 4)
Rule 14	(A7 = 1) & (A8 = 1) & (A15 = 2) => (Dec = 4)
Rule 15	(A16 = 2) => (Dec = 5)
Rule 16	(A4 = 1) & (A7 = 2) & (A10 = 2) & (A11 = 1) & (A15 = 1) => (Dec = 5)
Rule 17	(A5 = 2) & (A7 = 1) & (A8 = 2) & (A9 = 2) & (A15 = 1) => (Dec = 2) OR (Dec = 3)
Rule 18	(A5 = 1) & (A7 = 1) & (A8 = 2) & (A9 = 1) & (A11 = 1) & (A16 = 1) => (Dec = 4) OR (Dec = 5);

Table 3: Rules Generated for the five cases of Typhoid Fever

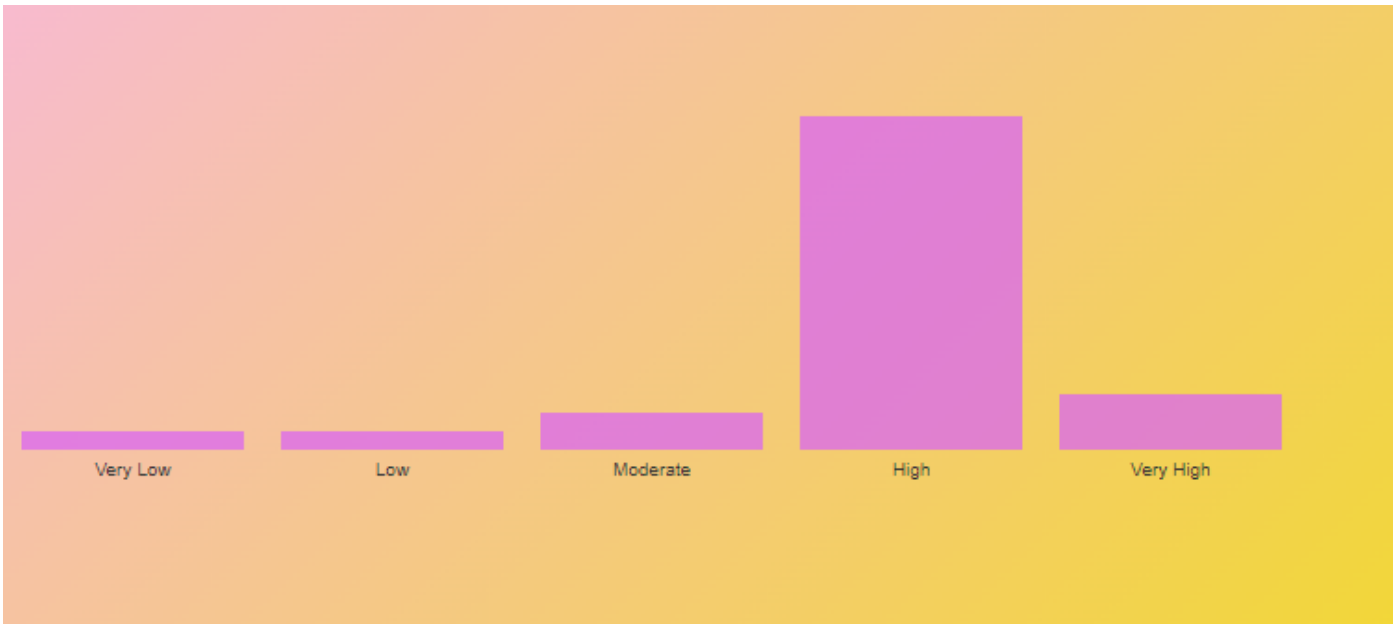


Image 1:Test Results

2.2Widal testing:

Enteric fever is mainly caused by **Salmonella Typhi and Salmonella Paratyphi A & B**.Antibodies specific Flaggellar Antigen 'H' and Somatic Antigen 'O' of Salmonella species usually become detectable in the blood after 7 days from onset of infection in individual infected with typhoid. This test facilitates quantitative estimation of antibodies to salmonella antigens in human serum by slide.

2.2.1Principal of the test:

When the colored, smooth, attenuated antigen suspensions are mixed and incubated with patient serum, Anti-salmonella antibodies present in the patient serum react with Antigen suspensions and give agglutination. Agglutination can be observed microscopically for one minute.

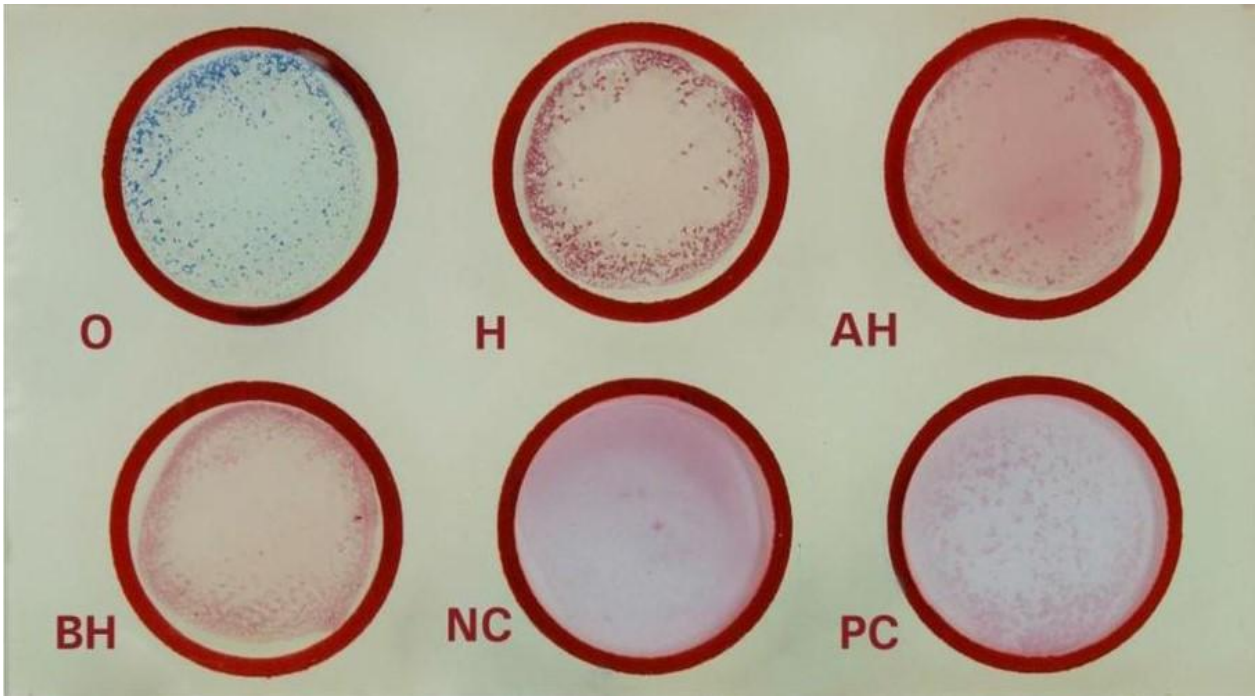


Image 2:Test Result

2.2.2 Interpretation of result

Observe agglutination or formation of clouds in the wells. Compare it with the positive control. Agglutination is the positive test results that indicate the presence of corresponding antibodies in the patients' serums. The absence of agglutination is a negative test result and indicates that corresponding antibodies are not present in the patient's serum.

3. If the patient is tested positive, It is also observed that there will be changes in some hematological parameters. According to the reference[2] significant reduction in the values for **PCV, WBC, ESR, and HAE** concentration in typhoid positive males in comparison to typhoid negative males was observed.

Parameters	Positive Male	Control Male
PCV	40.70 ± 4.36^a	44.19 ± 2.96^b
WBC	5.65 ± 1.81^a	6.47 ± 2.52^b
NEUT	65.27 ± 16.69^a	63.11 ± 15.32^a
LYMP	27.07 ± 15.67^a	28.94 ± 13.85^a
MONO	7.69 ± 3.52^a	7.67 ± 2.07^a
ESR	193.09 ± 64.25^a	246.37 ± 53.92^b
HAE Conc.	13.37 ± 1.48^a	14.62 ± 1.16^b
PLT	10.09 ± 7.04^a	4.97 ± 5.27^a

Image-3

In females, a significant decrease was observed in values for PCV, ESR, HAE concentration, and PLT in typhoid positive females when compared to typhoid negative females.

Parameters	Positive Female	Control Female
PCV	35.47 ± 4.25[a]	38.14 ± 2.72[b]
WBC	6.42 ± 2.29[a]	7.01 ± 2.82[a]
NEUT	60.48 ± 15.10[a]	60.12 ± 15.84[a]
LYMP	31.92 ± 13.99[a]	32.33 ± 14.91[a]
MONO	7.41 ± 2.49[a]	7.62 ± 2.20[a]
ESR	225.67 ± 92.66[a]	270.64 ± 51.06[b]
HAE Conc.	11.78 ± 1.46[a]	12.68 ± 0.91[b]
PLT	12.35 ± 6.58[a]	6.30 ± 7.96[b]

Image-4

4. To give the patients better treatment it is required that patients go through the CBC(Complete Blood Count tests).According to the reference[3]. In this research, a deep learning-based blood cell counting method has been proposed. Authors employ a deep learning-based object detection method to detect different blood cells. YOLO uses a single neural network to predict bounding boxes and class probabilities directly from the full image in one evaluation. They retrain the YOLO framework to automatically identify and count RBCs, WBCs, and platelets from blood smear images. To improve the counting accuracy, a verification method has been developed to avoid repeated counting by the framework. Also, the trained model has been tested with images from another dataset to observe the generalization of the method. Below mentioned image shows the proposed deep learning-based blood cell identification and counting system.

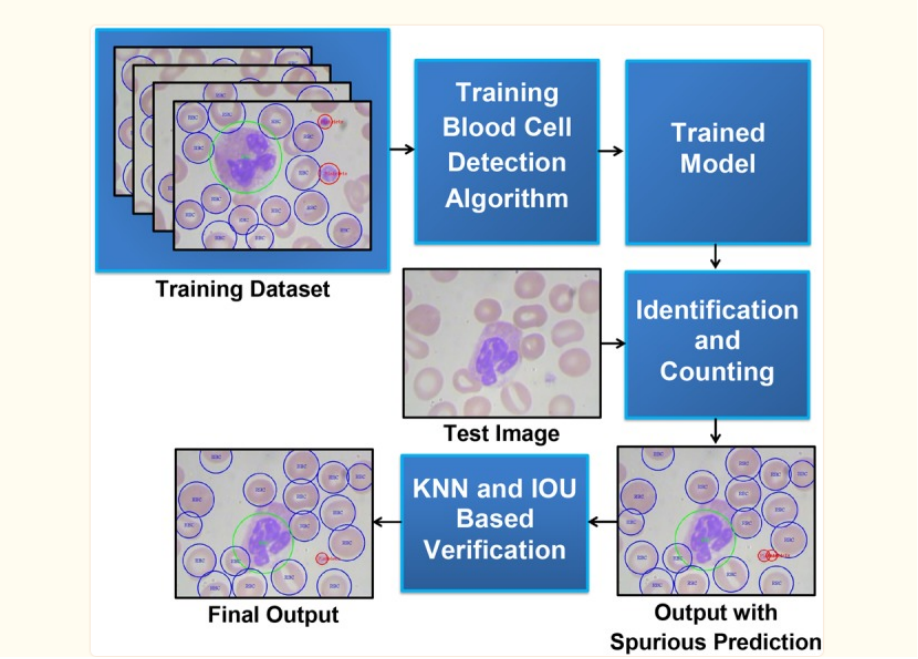


Image-5

Tasks:(All the tasks mentioned here are according to my proposed solution and not the actual tasks of the internship)

1. Collect the dataset on a large scale in which 18 conditional attributes(Symptoms) and 1 decisional attribute(Result) and, some blood smear images will be presented.

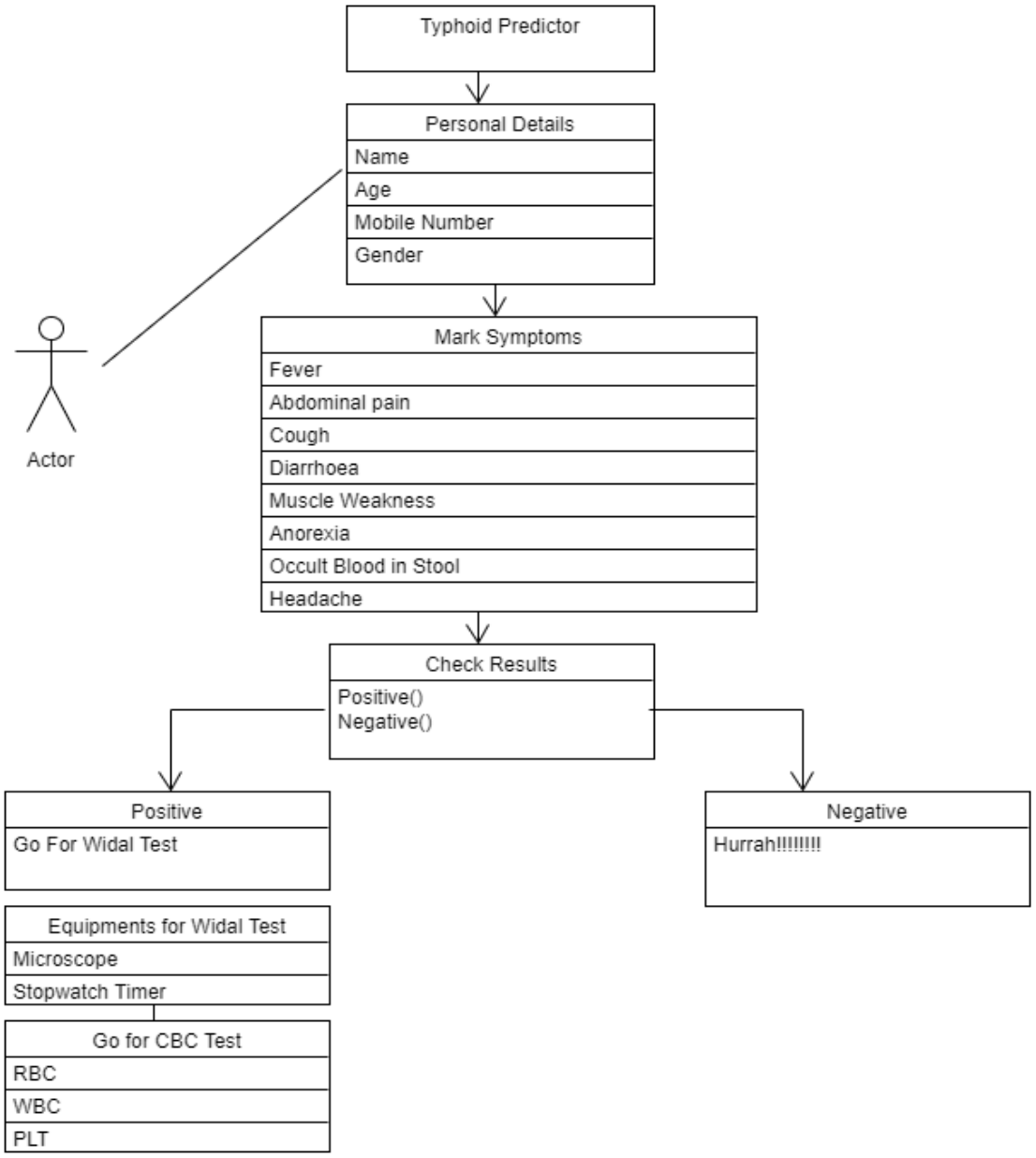
2. By Using the Rough set algorithm predict the output by taking into consideration of symptoms.

3. If the result from step 2 is positive, Make a UI to use the mobile phone‘s high-resolution camera and timer to detect agglutination with the reactions of antigens and patients serum.

4. It is observed that enteric fever significantly reduces the PCV and PLT of human blood. To calculate the complete blood count YOLO algorithm mentioned in the reference [3] will be deployed in the Application.

5. Some basic UIs such as personal details, mark your symptoms, and test results will also be included in the application.

Unified Model of Proposed Solution



Link of an android app based on the malaria screener model:

https://drive.google.com/file/d/10gD3-HwT-tcQ_pQX1h7n-OQSYEQEAdnq/view?usp=sharing

Timeline

Period	Tasks
Week-1	I Will try to know more about the Mboalab community and try to find out different methods of connecting with mentors.
Week-2	I Will start my outreachy blog and write about my initial phase application, my contributions, and my final application so that in the upcoming outreachy internship, applicants can take help from the blog

Week-3	I will figure out which type of dataset(images or recorded symptoms data) is needed to predict typhoid fever.
Week-4	With the help of medical practitioners, figure out what will be the possible features of the dataset, what would be the form(continuous or discrete) of the predicted output of the dataset, and how it can be collected on a large scale.
Week-5	In the next step will try to clean the dataset and deal with the missing columns of the dataset followed in these ways. 1. Remove the row. 2. Remove the column. 3. Make some guesses with a high probability of correctness. 4. Calculate the mean, median, and mode of the given column and fill the missing value by this.
Week-6	Will try to find out the best possible machine learning algorithm for the dataset(eg. decision tree classifier, rough set algorithm) to train the model and predict the output.
Week-7	I Will wrote about my outreachy experiences and my work in the outreachy internship blog.
Week-8	Trained 67% of the dataset and run the designed algorithm on training data.
Week-9	Test the remaining 33% of the data with the help of the designed algorithm and check to what extent it works on test data.
Week-10	If testing is successful, use it with new data for which we are trying to find the predictions. Will check the accuracy of your model and try to gain the maximum possible accuracy on the dataset. Make the confusion matrix if needed and calculate the error rate, precision, and recall.
Week-11	I will design an android application by merging all sets of algorithms to produce the correct output.
Week-12	Document my whole work during the internship and conclude my journey of outreachy internship in the blog.

Tools needed to develop the application

- Datasets
- Java, Kotlin, SQLite, XML
- Android Studio, Java Development Kit
- GitHub for Version control.

Future Enhancement

- Other methods of Typhoid diagnosis such as the Slide Semi-Quantitative Method and Quantitative Tube Method can also be added to the application for more accuracy of test results.

- Collection of datasets on a large scale so that accuracy of algorithms can be increased.

References

[1]A Machine Learning Approach to Clinical Diagnosis of Typhoid Fever by Abiodun Oguntimilehin, Adebayo O Adetunmbi and Abiola O.B.
[2] Changes in some hematological parameters in typhoid fever patients attending Landmark University Medical Center, Omuaran-Nigeria by James A.Ndko Victor T.Dojumo, Jeremiah A.Akinwumi, Vicyor O.Fajobi Akinyomade O.Owolabi, Oludolapo Olatinsu.
[3] Machine learning approach of automatic identification and counting of blood cells by Mohammad Mahmudul Alam and Mohammad Tariqul Islam.