

Classification and grouping of corona patients based on disease severity using Machine Learning algorithm

K Sri Sai Swaroop
Computer Science and Engineering
SRMIST
Kattankulathur, India

J Akarsh
Computer Science and Engineering
SRMIST
Kattankulathur, India

Ganesan R
Assistant Professor
School of Computing, SRMIST
Kattankulathur, India

Abstract—The project suggests a computer vision-based system that have ability to classify and additionally organize the corona patients supported illness severity. The symptoms of body imply the seriousness of the illness and advocate the optimum approach to modify the illness. It's together needed to diagnose the illness correctly with before nice injury by providing proper treatment. Moreover, estimation of severity of illness is needed for applying correct treatment to avoid the illness to induce bit with alternative person and additionally to create others notice the severity of disease that they're facing off. Here we tend to use multi category Deep learning with for classification of diseases and degree of disease severity.

Keywords—Classification, Grouping, Disease, Unsupervised algorithm, Deep learning

I. INTRODUCTION

As of Nov 10, 2020, 8.96M cases and 132K deaths thanks to coronavirus illness 2019 are reportable globally. The pandemic began in China, at intervals towns of city Hubei. However, on March 25, 2020, the range is higher in corona cases became larger within the remainder of the globe than in China. Additional Cases area unit occurring in Italy (1.27M cases), the USA (11.6M cases) and enlargement of the pandemic continues. A strange of studies of hospitalized patients have often shown that symptom associated with infection, additional unremarkable fever, cough, cold, bitter thought. In step with the report from the WHO on COVID-19, 85% of the 84586 patients with confirmed COVID-19 in China to March 27, 2020, had mild-to-moderate cases, while 12.9% developed severe illness and 7.4% developed to a vital stage requiring medical aid.

In a few reference of progression in 35986 effected persons, 4 those at advanced stage for severe illness and death enclosed folks over the aged person of 60 years and folks with critical situations. The designation of COVID-19 extremely depends on the options, characteristics, imaging, and thermal screening, etc. The concluded result by these technologies was it's taking longer. Multiple varieties of knowledge were collection of a patient with coronavirus infection which they were integrated by the doctors to make the designation choices. The transmission model was additionally investigate however the coronavirus transmission regionally and global wise. Machine learning algorithms were wide want to implement the medical specialty knowledge parts for this designation call.

This study investigated the detection of severely sick patients with COVID-19 from those with gentle symptoms victimization the info sets. The info sets consisted of Chest

X Rays of different patients. Associative in Nursing correct austereness detection model of the patients with COVID-19 supported those options on top of the could improve the standard of this illness in massive scale practices. The sections can foremost describe the data assortment and modelling strategies, thus used the popular machine learning algorithms to make the foremost effective austereness detection model.

This paper consists of two sections:

- 1) Classification of Disease.
- 2) Severity measurement of disease.

1) Classification of Disease

Classification of Disease is inspiration for the knowing the health issues and stages globally, and therefore the international commonplace for news illness and health issues. It's the place for classifying the disease for all hospitals and analysis functions. Classification of Disease defines the affected persons and alternative connected health conditions and it will classify the disease that was affected for the patient. The main contribution for classifying the disease is :

- It is mainly used for analyzing the each and every disease of the patient.
- It is the term used in all hospitals for better treatment and it can be easily stores the data of the patient.
- It is the type of algorithm used for comparing the data in a given time period.

Based on knowledge provided by researchers, ICD has classified a number of the stages into the following:

- Monitoring of the problem and interference of illness.
- Causes of death.
- External causes of unwellness.
- Primary treatment and precaution live have taken by ICD to classify the illness.
- Some of the problems area unit taken seriously for diagnosis the patient.

- Taking medical care of the people that area unit full of illness.
- Recording each trace of the illness.
- Certain pointers are enforced.

2) Severity Measurement of Disease

We've collected a number of the results concerning the patients from our legendary doctors who have died from illness and for cases outside of the China. These results were used to estimate the time interval between input and output of disease. These are the individual results. And next we obtained the age based results like in which age group members are attacking with disease and who are not. Using data of WHO organization that they have provided, we obtained age-based results like in which age group disease is attacking the most. Furthermore, data on age-based severity are used to identify the cases that are present at one location to another location.

II. LITERATURE SURVEY

From the study of different base papers, we understand how the diseases and outbreaks are prevented by using machine learning methods[3]. Scientist developed a machine learning model that can predict Ebola and other filovirus outbreaks by looking at the last spillover[1]. In current situation, the corona virus is the disease that is famous in all type of situations like on newspaper, articles, books and etc[5]. So, we think about the situation that can be cure the disease easily by taking few references that already discussed about the different diseases[2].

By using ML methods which is one of the field of AI, it is possible to analyze data from human health to investigate bat species with a high probability of Ebola and other viruses[6]. Amongst all the existing analytical tools in health care AI has been recognized as the most promising and powerful for mankind[9]. Today AI and its components becomes more accurate tool not only in identifying diseases, but also in providing recommendations, patient adherence and engagement and drug development[7].

In few references, they conclude as a single layer artificial neural network, SVM classifier and decision tree predictors that could be applied to different combinations of Ebola data for predictions[10]. Machine learning based model that predict Ebola and other filovirus outbreaks[8]. And also it was taken as a dataset as input and produce a model that is capable of handling new data[12]. Adopting such methodologies has always proven to enhance the accuracy of the identifying diseases[13].

We have discovered that in few studies that using machine learning and artificial intelligence for image processing can efficiently identify early signs of deadly disease, where conventional tools could not discover early sign of these diseases[15]. In this, it was explained as each and every healthcare worker will prepare the recovery models to survive[16]. So, the vaccination could be easily given to the patients in regular basis[18].

III. EXPERIMENTATION

In this paper, we are using VGG16 in TensorFlow and Keras. Here, we will take datasets of patients suffering with

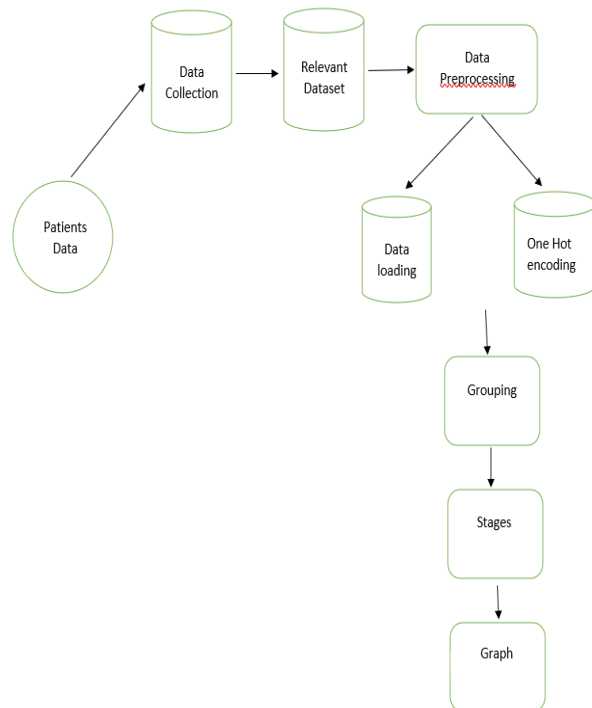
corona disease. The datasets include images of Chest X Rays.

We are using Deep learning algorithm and CNN for classification and grouping purpose. By taking this all data we will keep the patients in the stages named as Asymptomatic (mild) cases, Symptomatic cases, Severe cases, Deaths. And then proceed with the treatment.

A. Abbreviations and Acronyms

We have used abbreviations in our context. The abbreviations that have used in our context are ICD meant for Classification of disease, VGG16 meant for Visual Geometry Group from Oxford, FC meant for fully connected.

B. Architecture Diagram



IV. METHODS AND METHADOLOGIES

1) Data Collection

Data Collection is defines as collection of huge data and showing them in a particular condition. This is the data given for related hospitals depending on their possibility. Some studies recruited some cases and given to the different hospitals. We have collected data from patients who are hospitalized from January to March 2021. The datasets consists of all type of cases such as mild cases, moderate cases and extreme severe cases. A research worker will value their hypothesis on the concept of collected knowledge. In most cases, data assortment is that the first and most important step for analysis, despite the world of analysis. The approach of data assortment is completely different for varied fields of study, wishing on the desired data. The foremost vital objective of data collection is guaranteeing that information-rich and reliable data is

collected for applied mathematics analysis so as that data-driven choices are usually created.

2) Data Pre-processing

The information that was taken is processed in an exceedingly means of analyzing each parameter. In any Machine Learning method, Data Preprocessing is that step throughout that the information gets remodeled, or Encoded, to bring it to such a state that currently the machine will simply take apart it. In different words, the options of the information will currently be simply taken by the formula.

The data pre processing is a initial process in our paper. It will be processed as the data that is existing will be taken into place. That will be meant as disease, patients, Chest X rays, CNN, deep learning, severity of disease, classification, loading and unloading of the data. This is the basic preprocessing objects that is used in our paper.

3) Feature Extraction

A feature may be a non-public measurable property or characteristic of a development being ascertained. As an example, colour, mileage and power are usually thought about as options of a automobile. There are differing kinds of options that we'll encounter once we affect data. Feature Extraction means it will combine variables into features so that it will reduce the amount of data that must be processed and it also gives accurate and completely describing the original data set. It is the component in which each and every feature will be extracted and put them in the stages. Here we are going to use CNN.

The feature extraction is a process of extracting the each and every feature in a given data. And it is a term used for extracting all the information into different parts. So that it can be differentiated in a easy manner. In this the quantity of information will be given best analysis. The building up of each and every parameter will gives the best analysis. The speed of executing and building up the feature will make the speed of execution.

4) Algorithms

a) Deep learning

Deep learning is a algorithm that was unsupervised which allows the data that can be partitioned as a different aspects. So this is the algorithm which can be unstructured so we can use in a efficient way. In our paper it will be processed as the patient chest X rays will be taken and it will be differentiated according to the risk factor that is processed through coronavirus.

It's a algorithm used for medical image analysis and additionally they need made results reminiscent of and in some cases surpassing human skilled performance and additionally it's galvanized by informatics and distributed communication nodes in biological systems.

It's a algorithm chiefly expected for predicted the analysis of complete data given. It'll observe the actual symptoms in an exceedingly vast data. It'll be used image recognition application for sleuthing the malady. In deep learning, every level learns to remodel its computerfile into a rather a lot of abstract and composite illustration. It will divide as a layers and then perform the functioning of data. In this, it will hide some important layers and then it will execute.

5) Comparison of Different Algorithms

Two algorithms were evaluated for detection performances exploitation our default parameters on all the options of the COVID-19 respiratory illness patients. One in all the algorithm is employed for Classification and another algorithm is used for grouping. Both are related to analyzing of data.

Deep learning is for classification. It's an unsupervised learning problem where we recognize the category for a collection of a training data points and need to propose the category for the other given data point. CNN is for feature extraction. It's an unsupervised learning problem where we're doing not grasp the class or labels for any data and we're making an attempt to urge or come back up with the clusters while not this data. We may would like the simplest way to assign the cluster for the other given datum here also. The means of corroboratory, deciphering and benchmarking the results are fully completely different for supervised and unsupervised learning problems and we cannot compare them.

6) Disease severity

Disease severity may be a term accustomed to characterize the impact that a disease process has on the use of resources, comorbidities, and mortality. It is often employed by funding agencies to work out what's an appropriate payment for hospitalization. Disease severity is that the share of different diseases in a body. And those diseases causing disease to humans. Severity results from the size and number of the lesions.

Disease severity is a lot of applicable in diseases like rusts, downy and powdery mildews, leaf spots and different similar disease. It tells regarding the extent of harm caused by the disease. It is a property that explains the range of the disease in a term of medical words. And also, it explains the severity of the disease. So, by considering this property, we can exactly know the complete situation of the patient. This is the type of algorithm which completely analyze the patient by giving the appropriate the patient details.

It is the parameter that will take place essential role in our paper.

7) Classification Algorithm

It will classify the data in which it will utilize input data sets for the purpose of classifying the data and it will organize the data into categories. Here we are going to use Deep learning for classification algorithm.

8) Disease Classification

In this parameter, we will classify the each and every disease. By analyzing the type of disease, we will put them in the stage. For example, if patient was having previously cancer, we will analyze the data also and then obtain them in a stage.

9) Grouping

In this step, we will group the patients in a particular stage by considering all the previous steps, we will insert in the specific stage. It is a method of grouping the patients which are belonging to the same category. That means for every patient, a score will be assigned. So, we will group the patients according to the scores.

V. IMPLEMENTATION

Firstly, we are goanna take datasets. The datasets consists of Chest X Rays of corona positive and normal patients. Now We started our project by using VGG16 in TensorFlow and Keras. VGG16 is pertained on ImageNet which is very useful dataset. Next, we have used OpenCV and some other libraries to make our project.

A. Feature Extraction

In Feature Extraction, for every file we have extracted image name. And we have transformed the image into RGB format because some images may be in some other format. So, we have used PI library to transform all images into RGB format.

A feature may be a non-public measurable property or characteristic of a development being ascertained. As an example, color, mileage and power are usually thought about as options of a automobile. There are differing kinds of options that we'll encounter once we affect data Dataset Visualization

It is nothing but to identify what is the size of the image, width, height of the image and distribution of the channel or the image. Here, we are using plot sample from dataset to take datasets and by using matplotlib lib library, we can create image from datasets.

To display images, we are using display_image and we are implementing trainX and trainY as our datasets. As TrainX consists of covid positive and trainY consists of covid negative. Firstly, we have created a list that contains all the paths to the images of the dataset.

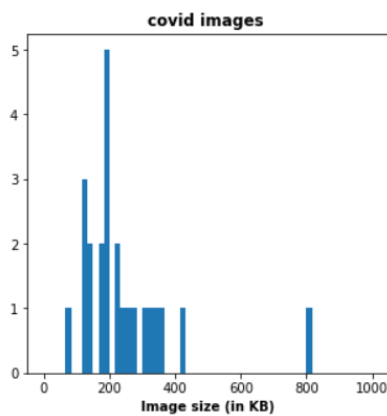


Fig. 1. Image Size of covid images

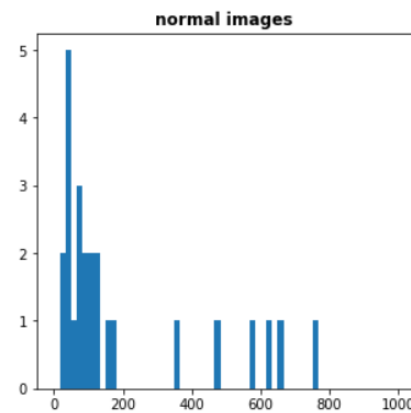


Fig. 2. Image Size of normal images.

B. Data Distribution

It is to identify the distribution of the size of the image. Because every image has different sizes. So by knowing the sizes of the images, we can implement the datasets as based on our required size.

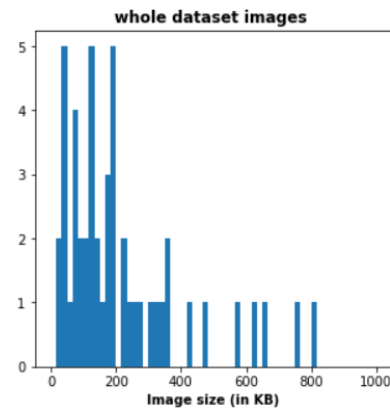


Fig. 3. Image size of whole dataset images.

C. Data loading and One -Hot encoding

After loading the data, we need to normalize. So we are dividing the data by 255.0 so that it will give perfect pixels size. We performed normalization because in order to make the conversion to the optimum solution faster.

Now, we have performed One-Hot Encoding on the labels. Now to convert string names into numbers, we have used LabelBinarizer() which is a library which we can import from sklearn pre-processing. Then, One-Hot encoding converts the binary numbers into vector format. For converting binary numbers to vector format, we have used to_categorical() which is a library from TensorFlow.

D. Splits

Now, we have partitioned the data into training and testing data in which we are using 80% of the data for training and remaining 20% for testing. We have used train_test_split for splitting the training and testing purpose.

E. Data Augmentation

Now, we have initialized the training data augmentation objects. For this we have used Image Data Generator

library. Data Augmentation is nothing but it is going to create additional images out of the existing images in order to increase number of images to be put into training.

F. Learning Rate

Now, we have initialized the initial learning rate, number of epochs to train for and batch size.

G. Transfer learning

Bringing the modularity pre treat and then we can change and train the output layers for classification problem. It is most useful in our project.

H. Loading of VGG16

Now, we loaded the VGG16 network, ensuring the head fully connected(FC) layer sets are left off.

I. Head Model

Now, we constructed the head of the model that will be placed on top of the base model.

J. Base Model

Now, we placed the head Fully Connected(FC) model on top of the base model and this will become the actual model we will train. Next, we are looping all layers in the base model and we will freeze them so they will not be updates during the first training process.

K. Compilation

Now, we have compiled our model and after compilation over, we have trained the head of the network and saved the model.

L. Predictions

The last step in our project in terms of classification is to make predictions on testing set and make predictions image by image. Now, I have used my model to predict class of the image and then compare true labels with predicted labels.

If they are correct, then the model is predicting correctly and it turns out to be true. For each image in the testing set, we need to find the index of the label with corresponding largest predicted probability.

M. Classification Report

Last step is to show a nicely formatted classification report. For this step, we have used confusion matrix. The confusion matrix allows us to tell us the performance of the model. Now, we have computed the confusion matrix and used it to derive the raw accuracy, sensitivity and specificity. Next is we plotted the training loss and accuracy and we serialized the model to disk.

N. Deep learning

Deep learning is a process that has the ability of learning unsupervised from the given data. And that data will be unstructured or unlabeled. It will be used image recognition application for detecting the disease. In deep learning, each level learns to rework its input file into a rather more abstract and composite representation. It is a AI based function which can imitates the working of the human brain in processing the data and creating patterns for use in decision making.

O. Figures and Tables

In the below table, we will be shown the entire paper in a tabular representation. In this, we will give the every stage with the percentage of the disease affected for the patient and every name of the stage will be perfectly explained.

TABLE I.

Stage Representation	Scores	Indication
Stage 1	0	No Symptoms
Stage 2	1 to 2	Starting
Stage 3	2 to 3	Asymptomatic
Stage 4	3 to 4	Symptomatic
Stage 5	4 to 5	Severe
Stage 6	5 to 6	Deaths

Fig. 4. Representation of Stages

1) Stage 1

In stage 1, the patient affected with virus can be cured by providing proper food. In this stage the score of the affected patients is 0. This stage will be indicated as No Symptoms.

2) Stage 2

In stage 2, the person will be affected with virus with having few symptoms like cough and cold. The score of the affected patients is 1 to 2. This stage will be indicated as Starting stage.

3) Stage 3

In stage 3, the person affected with virus will be having symptoms like high fever and vomiting, this can be cured by providing proper medication. The score of the affected patients is 2 to 3. This stage will be indicated as Asymptomatic stage.

4) Stage 4

In stage 4, the person affected with virus will be having some advanced health issues like chest pain, sour throat. This person should admitted in hospital. The score of the affected patients is 3 to 4. This stage will be indicated as Symptomatic stage.

5) Stage 5

In stage 5, the person affected with virus will be considering the advanced chest related issues and also the symptoms attacked for the patient. The person must and should taken care by doctor. The score of the affected patients is 4 to 5. This stage will be indicated as Severe stage.

6) Stage 6

In stage 6, the patient attacked with virus is in advanced stage and severity of the above stage and also the patient should be given proper medicine. The score of the affected patients is 5 to 6. This stage will be indicated as Deaths.

VI. RESULT AND DISCUSSION

In this paper, we have gone through the positive datasets as 0 and normal datasets as 1. Then, after verifying this result we are gonna proceed with implementing the algorithms and then we are gonna classify and group the entire data

tested as positive by considering the X Ray of Chest. Finally, we will show the data in the form of stages. And then, we will obtain the graph according to the severity of disease.

VII. SCREENSHOT OF THE OUTPUT

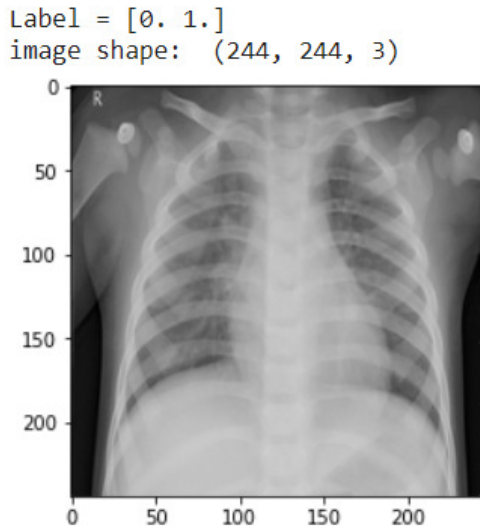


Fig. 5. Chest X Ray of covid

	precision	recall	f1-score	support
covid	1.00	0.80	0.89	5
normal	0.83	1.00	0.91	5
accuracy			0.90	10
macro avg	0.92	0.90	0.90	10
weighted avg	0.92	0.90	0.90	10

Fig. 6. Classification Report

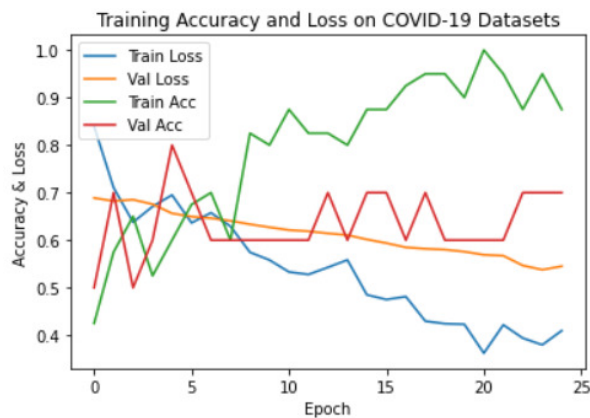


Fig. 7. Training Loss and Accuracy on COVID-19 Dataset

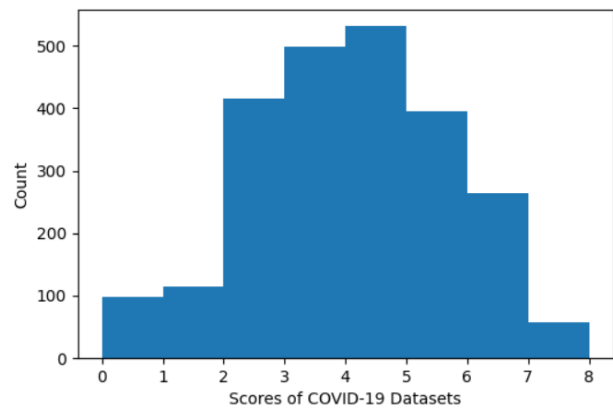


Fig. 8. Scores of Chest X Rays

Score 7.3
'Symptomatic'

Fig. 9. Score of the person affected with covid and his Stage

The above output will describe the particular stage of the effected person. It will show the stage name according to accuracy of disease severity that was named as score. By taking that stage name we can exactly know the condition of the patient. so, that we can provide proper treatment. And also one more advantage in this paper is we can know the severity of disease that means how much virus is spread into the lungs part can be determined. According to the score the stage name will be denoted.

VIII. CONCLUSION

In this paper, We have concluded that it will give entire data about the patients affected with covid positive and negative. And we have classified all the datasets and grouped them into stages which can be better understandable.

REFERENCES

- [1] Sabeti P, B Fry, T Fradet, K Retzepi, I Silver, Colubri A. Machine learning and field-deployable app to predict outcome of Ebola patients and for transforming clinical data into actionable prognosis model.
- [2] G Hinton, Y LeCun, Bengio Y Deep learning. Nature 521, 436– 444 (2015).
- [3] MJ Farrell, T Huang, JP Schmidt, S Maher, JM Drake, BA Han. 2019. Mammal Exposure by using Ecological indicators to Ebolavirus. Phil. Trans.
- [4] K Simonyan, Schrittwieser J, D Silver et al. Without human knowledge, Mastering the game of Go Alpha.
- [5] S Abe, Y Ilakozaki, T Matsui, G Sun. Using a neural network and the fuzzy grouping method J Infect, developed an infectious disease/fever screening radar system that can stratifies higher risk patients within ten seconds., 70 (3) (2015), pp. 230.
- [6] WS Barclay, SM Ilaslam, B Mistry, Long JS. 2018 A virus species specificity is determined by Host and viral . 17, 81 - 67.
- [7] S.M Asch, J H Chen Machine learning and prediction in medicine— beyond the peak of inflated expectations N. Engl. J. Med., 376 (26) (2017), pp. 2509-2507.
- [8] GN Thomas, B Tonilinson, CS Cockram, Abdullah AS. Emerging Infectious Diseases by the severe acute respiratory syndrome outbreak in Hong Kong. 2017

- [9] DA Sack, AR Nelson, AL Lopez, Ali M. PLOS neglected the tropical diseases which is updated global burden of cholera in endemic countries 2015;9:e0003832.
- [10] J R Maddicott. Past and Present. "Plague in seventh century England" 156 (1): 54-7. doi:10.1093/past/156.1.7. ISSN 0031-2746.
- [11] Meiyu Duan, Nan Zhang, Haochen Yao. "Using Blood and Urine Tests, Detection of Severity of COVID19 Patients using Machine Learning algorithm" doi: 10.3389/fcell.2020.00683.
- [12] J Oon Alyssa (June 17 2019). "In Kelantan, A measles outbreak is the cause of 15 Orang Asli Deaths".
- [13] Paul E Willam (2018). Lippincott Williams and Wilkins. Fundamental immunology. ISBN 6519-0-7817-0-978.
- [14] "In South Asia, the control and eradication of small pox". www.smallpoxhistory.ucl.ac.uk.
- [15] "Global Task Force on Cholera Control in Cholera Country of Zimbabwe by World Health Organization".
- [16] "Swine flu deaths in 1895 and number of cases near to 32K mark by Press Trust of India in March 21, 2015".
- [17] News Desk (July 4 2020). "MERS coronavirus".
- [18] Oon J Alyssa (June 17 2019). "In Kelantan, A Measles outbreak is the cause of 15 Orang Asli Deaths".