



Estimating the Approximate Necessity of Essential Medicines in Syria Using Machine Learning

Labeb Abood^(✉) 

Belgorod State University, Belgorod 308007, Russian Federation
marcabood028@gmail.com

Abstract. Essential Medicines are one of the very crucial parts of the health care needs. This article focuses on the estimation of the approximate necessity of essential medicines in the Syrian Arab Republic. Since 1977, when the idea of essential drugs or medicines was introduced by the WHO, it has been playing a vital role in society's betterment. Syria is no exception in that case. Since 2011, Syria's health care systems have been facing many problems. To make the situation better, it is necessary to focus on the modernization of the health care needs. After researching briefly about the usage, regulations, supply, and production of essential medicines in the Russian Federation, Germany, and the USA, the application of artificial intelligence seems like an ultimate necessity to make health care needs seamless and rapid. In this article, a machine learning approach is applied to predict the demand for different essential medicines in different regions of Syria. The task includes research methods, implementation, results, related work, and analysis for future progression. Random Forest Classifier is used to estimate the accuracy of the approach method.

Keywords: Essential medicine · Health care needs · Random forest classifier · Machine learning · Artificial intelligence in healthcare · Syrian healthcare system

1 Introduction

Currently, we are living in a time where there is no choice but to be frugal in every sector of life. It is necessary to choose the most up-to-date system and technology to ensure basic rights like medical care for every human being. Countries like America, Russia, and Germany have already taken their health system to a place where a patient can get almost any necessary health-related services by using just a smart phone. For many years, the primary goal of using technology in healthcare has been the rapid diagnosis and cure of diseases. With the addition of information technology, the biggest advantage in this medical field has been in data collection. Billions of people are getting sick all over the world and all their information is being stored online by internet and servers. As a result, health workers can easily store patient information for life. In Syria, the process is not that easy. Collecting data from different regions of Syria is challenging. Still, in recent times, many positive measures have been taken place related to health care needs.

The aim of this research is to predict an approximate demand of essential medicine from different regions of Syria, so that the distribution of medicine can be effective, and as per the demand of the people of that region.

The research followed several steps and is divided into two major parts. At first the proposed methodology of collecting data for essential medicine has been outlined, then the approximate estimation of essential medicine has taken place using machine learning.

Machine learning is swiftly infiltrating many areas within the healthcare industry, from diagnosis and prognosis to drug development and epidemiology, with significant potential to transform the medical landscape [1]. There is some research that has been carried out to understand the prediction of essential medicine demand using machine learning. Several countries are already getting the benefit of such an analysis.

2 Estimation of the Demand of Essential Medicines

Today, we stand on the threshold of the new medical revolution, just as big and far-reaching [2]. Earlier, either doctors or health care representatives wrote patient information on the paper for either collecting data to improve the overall health system or understanding the potential and existing health condition of a patient. And with the unimaginable advancement of digital media and information technology, the old paper document writing system has been eliminated. The combination of technology with the medical world has made it much easier to collect all these records. A highly available and well organized medical data for any country can help to update the essential medicine list in every two years by WHO more successfully. To make the research more effective, it is necessary to understand the underlying problems related to health care needs in Syria, essential medicine, and its production, supply, and availability in different regions. The article will outline these along with the usages of machine learning for predicting the demand of essential medicine in Syria.

2.1 Problem Statement

Medical data is more accessible in some countries than in others. Collecting data from Syria is comparatively difficult and time-consuming. The principal task of this research is based on the demand of the essential medicine for the Syrian health care system, and all the data collected using different methods are from various places of Syria, including Damascus, Halab (Aleppo), Hims (Homs), Tadmur (Palmyra), and so on. For collecting data, another struggle is that the patients don't feel very comfortable giving information for an independent research about what drugs they are being prescribed. But they are more comfortable to share their symptoms. The research named "Symptoms Based Disease Prediction Using Machine Learning Techniques" described the relative study of various Machine Learning algorithms for the detection of various diseases [3]. The system that is used to predict the diseases processes the symptoms provided by the user as input and gives the output as the probability of the disease [4]. Mostly, such systems used Big Data, and different Convolutional Neural Network algorithms. For type of data that have been collected, implemented using Machine Learning algorithm, i.e. K-nearest Neighbors, Decision Tree, Naïve Bayesian, random forest [4].

2.2 Research Methodology

Two major challenges were faced in the research’s method. One of these is data collection, and the other is to estimate the future demand of the essential medicine using machine learning. From the collected data first prediction of disease was done, defined in Fig. 1.

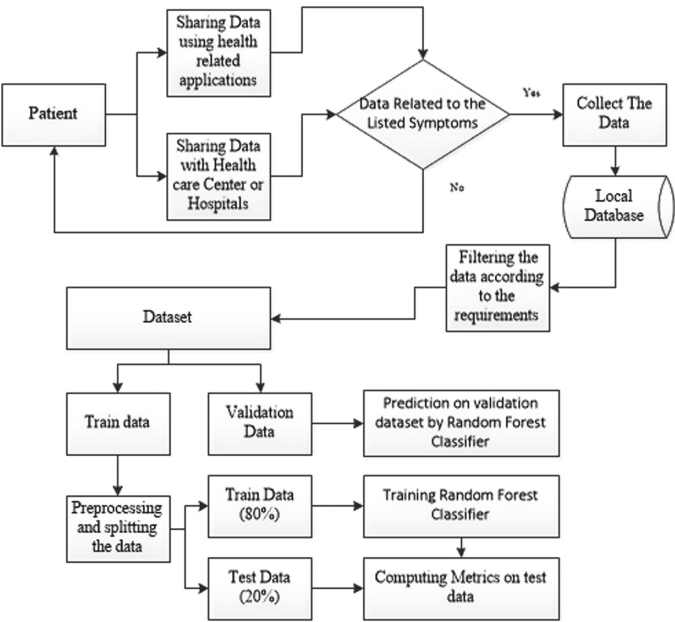


Fig. 1. The process of collecting data, and performing the data classification using random forest classifier.

Methods like document examination, observations, and questionnaires have been chosen for data collection. From the 2019 list of essential medicines by WHO, only twenty widely used drugs have been used in this study. Processes like, Data preparation, cleaning, and model building have taken place by studying various researches and existing machine learning approaches. To understand the quality of the model, a confusion matrix has been used. The prediction from the random forest classifier is applied and has been collected. The results will show the prediction of the disease, and these data will be utilized to predict the medicine for the disease defined in Fig. 2.

From that data, it is possible to predict the demand of the essential medicine by using a sorting algorithm. The proposed research methodology allows to apply the machine learning approach in different regions of Syria separately. After finding out the demand it will be compared to understand which regions demand what sort of essential medicines, and the result will show a count of maximum to minimum (Fig. 3).

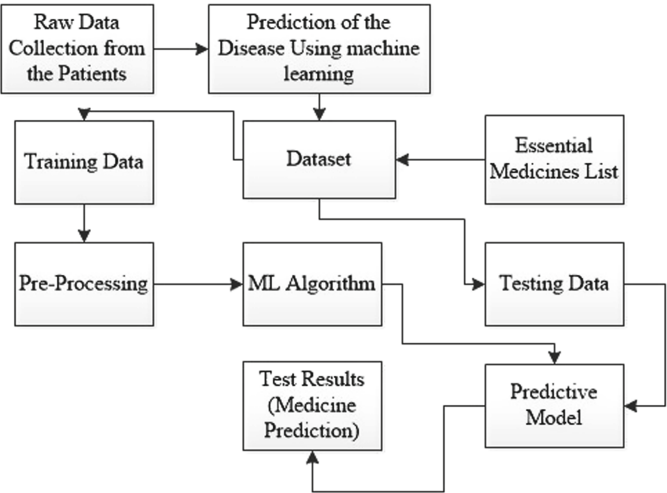


Fig. 2. The proposed framework for performing the prediction of the medicines for the diseases.

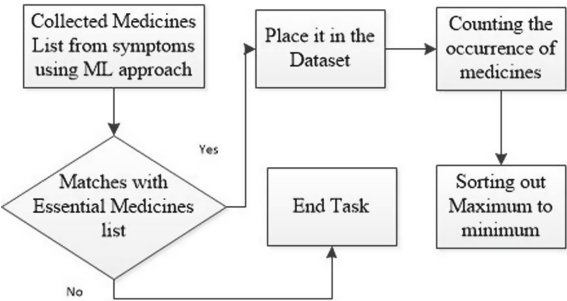


Fig. 3. The proposed approach to product the demand of essential medicines.

2.3 Understanding the Role of Essential Medicines

The World Health Organization (WHO) publishes a list of essential medicines every two years. Nearly 150 countries around the world follow the World Health Organization’s list of essential medicines. These drugs are very important for public health. Contemporary morbidity, cost, accessibility, and treatment factors are considered in compiling this list for each country. The World Health Organization has recently not only planned a list of essential medicines but also published a list of essential diagnostics. The list focuses on global public health challenges, including cancer, with an emphasis on effective and timely solutions. A list of essential medicines produced by the World Health Organization (WHO) in 2019 was followed for this study (Fig. 4).

Item	Description	Comments
Paracetamol	Inj IM	Doctors have been advised to avoid intramuscular injection of paracetamol in patients with fever.
Paracetamol	Injection:10mg/ml, 100ml v vial	It is an analgesic (it relieves pain) and an antipyretic (it lowers fever).
Tramadol	Injection: 50mg in 2ml amp	It is used to relieve moderate to moderately severe pain, including pain after surgery
Hyoscine Hydrobromide	Injection: 400 micrograms/ml	It is taken to prevent travel sickness (motion sickness).
Epinephrine (adrenaline)	Injection: 1 mg (as hydrochloride or hydrogen tartrate) in 1ml apoule	It is used for emergency treatment of severe allergic reactions

Fig. 4. Randomly selected five medicines from Syria’s essential medicine list 2019

2.4 Dataset

The dataset that has been used is collected from a large scale online and offline survey about symptoms of various diseases in different regions of Syria. In the dataset symptoms like: abdominal pain, abnormal menstruation, acidity, acute liver failure, altered sensorium, anxiety, back pain, belly pain, blackheads, bladder discomfort, blister, etc. are used (Fig. 5).

Symptoms						Diseases
abdominal_pain	acidity	altered_sensorium	anxiety	back_pain	...	
1	0	1	0	0	Allergy	
0	0	0	1	0	Common Cold	
0	1	0	0	1	Diabetes	
0	0	1	0	0	...	

Fig. 5. Random Samples of Medical Dataset with Symptoms, and diseases.

This dataset comprises either 0 and 1. It is a clean dataset with no null values [5]. If the patients have symptoms listed here, then it will be 1, but if it is absent, then it will be 0. Diseases like allergy, common cold, diabetes, and so on, are recorded to predict. After predicting the disease, the recorded data will be used for estimating the medicine. The prediction will be done using the suggestion of medical professionals using a machine learning approach. Then the dataset will be used to count the occurrences of the medicines. All the collected data from different regions will be implemented differently, so that it can be easier to compare, and find out the demand from maximum to minimum.

2.5 Random Forest Classifier

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both classification and regression problems in Machine Learning [6]. Random Forest classifier works by constructing a multitude of decision trees at training time, and the output is the class selected by most trees [7, 8] (Fig. 6).

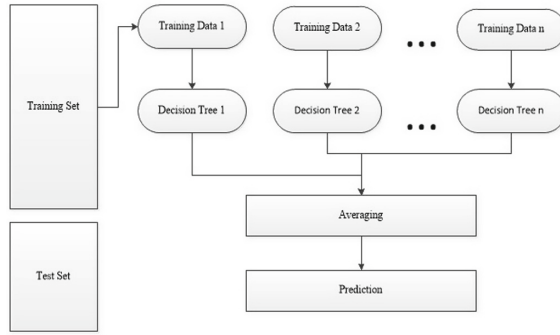


Fig. 6. A simple diagram of the random forest algorithm.

Random Forest Classifier during the time of classification often uses the Gini index. The formula is used to decide the nodes on a decision tree. It applies the class and probability to determine the Gini of each branch on a node, determining which of the branches is more likely to occur [9].

$$Gini = 1 - \sum_{i=1}^C (P_i)^2$$

Here, p_i represents the relative frequency of the class you are observing in the dataset and c represents the number of classes [9].

2.6 Implementation

Predicting future trends provides an additional value for improved healthcare system management in today's global business trend and step forward technologies [10]. The proposed methodology requires implementation of two prediction algorithms using a random forest classifier. Python programming language is used to implement them. The program starts with importing libraries like numpy, pandas, seaborn, etc. Then these steps are taken place [11]:

- Reading the dataset;
- Splitting the data;
- Training and testing the model;
- Model building (Random Forest Classifier);
- Fitting the model;
- The creation of a function for predicting the disease.

It is necessary to check the accuracy of the model, and for that, the model will be trained on the train data, using a confusion matrix. Confusion matrix or error matrix is a widespread approach in machine learning to evaluate the quality of a classifier is to cross-classify predicted and actual decision classes [12] (Fig. 7).

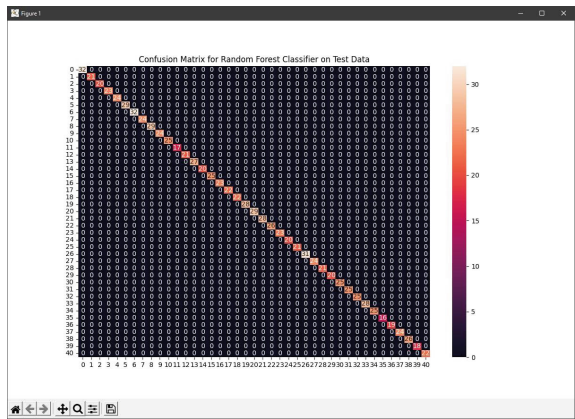


Fig. 7. Confusion matrix for random forest classifier on test data (disease prediction)

In the final part of the implementation, the function shows the prediction from the input data as 'Disease_prediction': 'Hypertension'. Then, with the collected data, prediction of medicine is implemented in a similar way (Fig. 8).

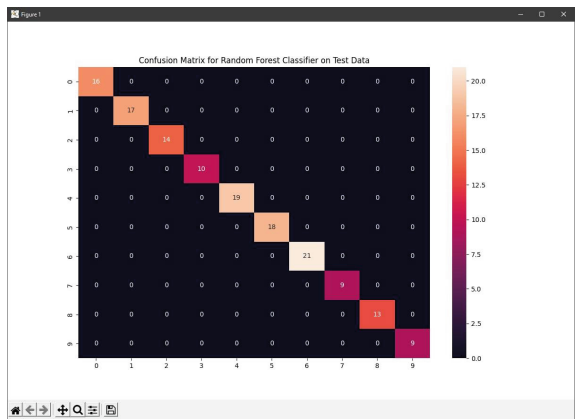


Fig. 8. Confusion matrix for random forest classifier on test data (medicine prediction)

Then, from the collected data, a function for counting medicines for different regions of Syria is implemented to find out maximum to minimum demands.

2.7 Results and Analysis

From various researches, it is evidenced that the data-driven predictive modeling with machine learning (ML) could become the cornerstone of health supply chain planning and operational management [13]. The results from the proposed models also show silver lining for those who are interested in the application of the machine learning

approaches to health care needs, including the production, demand, and necessity of essential medicines.

The 2019 list of essential medicine in Syria by WHO contains a lot of medicines. For this research, ten essential medicines were selected from the implemented models based on the outcome of the predictions. From the counting occurrences function, Paracetamol shows the highest usages among the patients. Considering this in mind, the places, and the number of patients took part in this research, the following demands are predicted for different regions (Fig. 9).

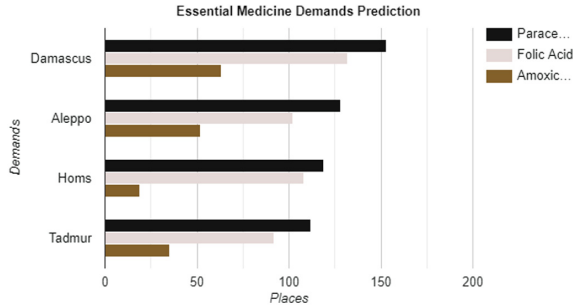


Fig. 9. Prediction of demands of some essential medicines in different regions of Syria using collected data.

To get an overall idea or predict the demand, it is necessary to be mindful about the fact that the population, weather, gender, education, and age play vital roles according to some researches. All the results and prediction shows:

- A more advanced dataset is necessary to implement the models more accurately.
- Patients should have given as much options of symptoms as possible during the time of data collection.
- More machine learning models should check the accuracy of the models, and a combined result could make the prediction additionally successful and accurate.
- The Dataset should be not only symptom based independently collected data but also with items like prescriptions, laboratory tests, details about important diseases, etc. [14].

The limitations for creating an up-to-date dataset in Syria are more complicated than in many other countries. However, there is no alternate rather than trying to make it possible for the improvement of the correctness of implementing different machine learning approaches. In a recent study, [15] there is a description of fundamental steps for preparing medical imaging data in artificial intelligence algorithm development, which explains current limitations to data correction and explores alternative approaches to address the problem of data availability. Such studies help to estimate the demand more accurately.

3 Conclusion/Future Work

When considering machine learning as a task, the task can be planned as a generalization of classical approximation tasks, since in the real-life of the applied projects, the data can be non-numeric, incomplete, and heterogeneous. [16, 17] Machine learning models are now well used in improving health care system worldwide.

The biggest hurdle in this research was data collection. If it were the case that a list of the prescriptions of patients for the past few years from different areas of Syria was available, then it would be easy to predict the demand. But the information collected was the signs and symptoms of diseases. First, it was necessary to find out the disease from the dataset, then the prediction of the drugs. From there, a dataset has been created of essential medicines, which were later used to predict the demand.

In the future, looking forward to thoroughgoing some tasks to make the proposed approach more efficient:

- Enhancing the data collecting approach,
- Understanding other research methodologies related to this topic
- Encouraging patients in Syria to share medical reports, prescriptions, and analysis in order to improve the dataset.
- Access to collecting data from all the regions of Syria,
- Using simultaneously three or four classifier models to predict more accurately,
- Developing health care applications to collect data, and so on.

The availability or right to essential drugs is not limited to those who can afford it, but is universal. Over 150 countries around the world use the World Health Organization's list of essential medicines. From which it is proved that essential medicines are very important for public health. And countries like Syria, Yemen, and other developing, and under-developed countries need to focus on more advanced, and effective health education to make sure the ultimate demands and needs of people.

If artificial intelligence (AI) technologies have a similar impact on healthcare as in other industries, then health care can become more effective and more efficient, improving the daily lives of millions of people [18]. It is a must understand that AI is not one technology, but a collection of them [19]. This is the new future, and the earlier researchers related to medicine, or any health care system, will understand this, the quicker and better the outcome will be.

References

1. Ascent of machine learning in medicine. *Nat. Mater.* **18**, 407 (2019). <https://doi.org/10.1038/s41563-019-0360-1>
2. Ahmed, Z., Mohamed, K., Zeeshan, S., Dong, X.: Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database (Oxford)*. 2020 Jan 1;2020:baaa010. <https://doi.org/10.1093/database/baaa010>. PMID: 32185396; PMCID: PMC7078068

3. Hamsagayathri, P., Vigneshwaran, S.: Symptoms based disease prediction using machine learning techniques. In: 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pp. 747–752 2021. <https://doi.org/10.1109/ICICV50876.2021.9388603>
4. Gomathy, C.K.: The prediction of disease using machine learning. *Int. J. Sci. Res. Eng. Manage. (IJSREM)* (2021)
5. Disease prediction using machine learning. <https://www.geeksforgeeks.org/disease-prediction-using-machine-learning/>. Last accessed 08 July 2022
6. Machine learning random forest algorithm. <https://www.javatpoint.com/machine-learning-random-forest-algorithm>. Last accessed 06 Aug 2022
7. Ho, T.K.: Random Decision Forests (PDF). In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16. pp. 278–282
8. Ho, T.K.: The random subspace method for constructing decision forests (PDF). *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8):832–844. <https://doi.org/10.1109/34.709601>
9. Random forest algorithm for machine learning. <https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb>. Last accessed 03 July 2022
10. Dash, S., Shakyawar, S.K., Sharma, M., Kaushik, S.: Big data in healthcare: management, analysis and future prospects. *J. Big Data* **6**, 54 (2019)
11. Disease prediction using machine learning. <https://www.geeksforgeeks.org/disease-prediction-using-machine-learning/>. Last accessed 05 June 2022
12. Düntsch, I., Gediga, G.: Confusion matrices and rough set data analysis. In: Published Under Licence by IOP Publishing Ltd *Journal of Physics: Conference Series*, (vol. 1229), 2019 3rd International Conference on Machine Vision and Information Technology (CMVIT 2019) 22–24 February 2019
13. Mbonyinshuti, F., Nkurunziza, J., Niyobuhungiro, J., Kayitare, E.: The prediction of essential medicines demand: a machine learning approach using consumption data in Rwanda. *Processes* **10**(1), 26. <https://doi.org/10.3390/pr10010026>
14. What is health data. <https://datasaveslives.eu/what-is-health-data>. Last accessed 05 Feb 2022
15. Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L.R., Summers, R.M., Rubin, D.L., Lungren, M.P.: Preparing medical imaging data for machine learning. *Radiology*. **295**(1), 4–15 (2020). <https://doi.org/10.1148/radiol.2020192224>. Epub 2020 Feb 18. PMID: 32068507; PMCID: PMC7104701
16. Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, p 746
17. Prudius, A.A., Karpunin, A.A., Vlasov, A.I.: Analysis of machine learning methods to improve efficiency of big data processing in Industry 4.0. *J. Phys. Conf. Ser.* **1333**, 032065 (2019). <https://doi.org/10.1088/1742-6596/1333/3/032065>
18. Goldfarb, A., Teodoridis, F.: Why is AI adoption in health care lagging? Brookings 9 Mar 2022
19. Davenport, T., Kalakota, R.: The potential for artificial intelligence in healthcare. *Future Healthc J.* **6**(2), 94–98 (2019). <https://doi.org/10.7861/futurehosp.6-2-94>. PMID:31363513; PMCID:PMC6616181