Faculty of Computational Sciences & Informatics - Academic City University College, Accra, Ghana
Society for Multidisciplinary & Advanced Research Techniques (SMART)
Trinity University, Lagos, Nigeria
SMART Scientific Projects & Research Consortium (SMART SPaRC)
Sekinah-Hope Foundation for Female STEM Education
Harmarth Global Educational Services
ICT University Foundations USA
IEEE Computer Society Nigeria Chapter

Proceedings of the 36th iSTEAMS Accra Bespoke Multidisciplinary Innovations Conference

Towards The Development of a Self-Diagnosing System For Ailments Using Predictive Modelling Machine Learning

Amegadzie Julius & Fiagbor Harriet
Faculty of Computational Sciences & Informatics
Academic City University College
Accra, Ghana
E-mail: harriet.fiagbor@acity.edu.gh; julius.amegadzie@acity.edu.gh

ABSTRACT

This study aims on Right to Privacy on the Internet. The paper will unfold as follows. Influences on consumer privacy online, online consumer tracking; that is when online scenarios meet privacy expectations or complied with a privacy notice, and the importance of privacy notices in managing privacy online. This paper will also highlight consumer online privacy specifying ''the youth, parents, and online privacy' 'and the regulations in place to shape such policies. Will also unfold privacy in the digital age or the internet. This paper will again unfold UN general assembly on the right to privacy on the internet. The paper unfolds the potential consequences of revealing certain information online and analyzes if there are any differences between the motivations and attitudes of young people. Will again highlight on National Security Agency (NSA) surveillance which demands that Internet carriers be more forthcoming about their handling of personal information which must be intensified. Responding to this concern, this report evaluates the data privacy transparency of forty-three Internet carriers serving the public. This paper is to investigate the relationship between individual and societal determinants of online privacy concern (OPC) and behavioral intention of internet users. The study also aims to assess the degree of reciprocity between consumers' perceived benefits of using the internet and their OPC in the context of their decision-making process in the online environment.

Keywords: Self Diagnosing, System, Ailments, Predictive Modelling, Machine Learning

Proceedings Citation Format
Amegadzie Julius & Fiagbor Harriet (2023): Towards The Development of a Self-Diagnosing System for Ailments Using Predictive Modelling Machine Learning. Proceedings of the 36th iSTEAMS Accra Bespoke Multidisciplinary Innovations Conference. University of Ghana/Academic City University College, Accra, Ghana. 31st May – 2nd June, 2023.
Pp 175-182. https://www.isteams.net/ghanabespoke2023. dx.doi.org/10.22624/AIMS/ACCRABESPOKE2023P17

1. BACKGROUND TO THE STUDY

Although there are disadvantages to self-diagnosis, it is a habit that cannot be completely abandoned. Hospitals would be overcrowded if everyone went to a clinic for every ailment, resulting in high costs and long travel time. As a result, others, make it a habit to self-diagnose and avoid going to clinics.

If the practice of self-diagnosis becomes common, it can lead to addiction, overdose, and other health complications. Self-diagnosis enabled by Predictive Modeling is a health-prediction system that serves as an end-user support and online consultation project to address the issue of health facility dependency and avoid sole reliance on over-the-counter (OTC) stores. In this paper, we propose a system that allows users to get real-time advice on their health issues via an online intelligent health care system. Various symptoms associated with each disease/illness are fed into the system. It then examines the user's symptoms to predict the corresponding disease/illness that may be present. In determining the best prediction model, four data mining techniques or classifiers (Naive Bayes, Random Forest, Decision Trees, and Multi-layer Perceptron Neural Network) will be employed. The intention is to chose the most robust classifier based on robustness in terms of performance and the ability to minimize overfitting of the data.

1.1 Background and Justification of the Study

Self-medication is a human practice in which a person administers medications based on the ailments they may be experiencing. This usually does not necessitate the use of a medical prescription or instruction [1]. These are known as "over-the-counter" or "non-prescription" drugs. Self-medication can also be used to treat stress, anxiety, depression, and psychological trauma [2]. Self-medication is as a result of self-diagnosis. For both males and females, it is age-independent [3]. Age, gender, income and expenditure, self-care attitude, education level, medical knowledge, satisfaction, and perception of ailments are all characteristics that influence self-medication trends in diverse groups [4]. This is a popular technique for so-called "minor diseases," such as headaches, coughs, flu, and other mild maladies. In these instances, self-medication is resorted to when the discomfort is not so great as to require an appointment with the doctor.

This is common practice in many parts of Africa. When people are unwell, they tend to take their medication and dismiss it as a minor problem. Most of the time, there are no medical specialists available to provide guidance. People do it because it saves them money, time and travelling costs. Some folks do it incorrectly. These drugs may have unfavorable effects on the body because there are no rules or proper prescriptions for them. One of the biggest issues with self-medication is the intake of wrong dosage and underestimating its effects. It is tough to eliminate self-medication due to the numerous benefits it provides. In actuality, self-medication has been historically used in different cultures to provide a means of healing. In recent times, the associated problems of self-medication can be overcome by technology. Through technology and research, a lot of progress has been made in the medical field. Using emerging technologies such as Artificial intelligence to alleviate societal challenges including self-diagnosis and medication would be a good counter-argument [5]–[7]. It would be a safer idea to have a guide that helps people to self-diagnose their symptoms and provide appropriate prescription for whatever ailments they are experiencing. Although it is not the only possible solution that could put an end to these challenges, it could be a step in the right direction.

1.1 Introduction

Machine learning, as the name implies, is the process of machines learning without being explicitly programmed or without the need for direct human participation. This machine learning method begins with providing them with high-quality data, followed by training them by creating multiple machine learning models based on the data and various algorithms. The algorithms we use are determined by the type of data we have and the task we are attempting to automate. As for the formal definition of machine learning, we can say that a machine learning algorithm learns from experience E with respect to some type of task T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E [14].

If a machine learning algorithm is used to play chess, the experience E is gathered from playing many games of chess, the task T is to play chess with a large number of players, and the performance measure P is the probability that the algorithm will win the game of chess. Supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning are the different categories (methods) of machine learning. This is a project based on supervised learning. The method uses a training dataset to create predictions, which are then compared to the actual output values. If the predictions are not accurate, the algorithm is tweaked until it is perfect. The algorithm will continue to learn until it reaches the needed level of performance. Then, for any additional inputs, it can deliver the desired output values.

As indicated in Fig. 1, there are several ways that this research will pursue as a general machine learning problem.

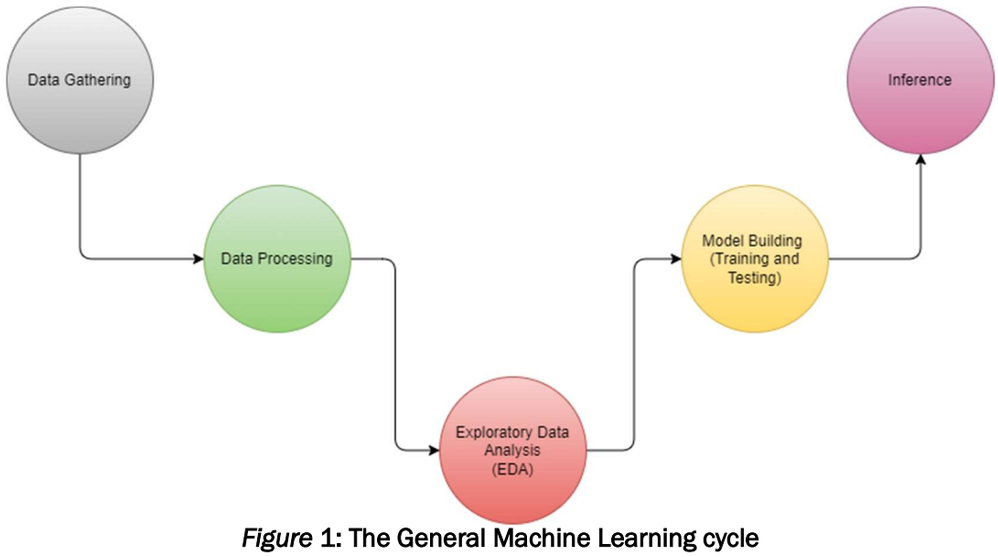

Figure 1: The General Machine Learning cycle

1. Gathering of Data

The first phase in the machine learning life cycle is data collection. It is one of the most crucial stages of the life cycle. The purpose of this step is to identify and collect all data-related issues. Firstly, identify the numerous data sources, as data can be obtained from a variety of places, including files, databases, the internet, and mobile devices. The output's efficiency will be determined by the quantity and quality of the data collected. The more data you have, the more precise your prediction will be.

2. Data Processing

Data preprocessing is a data mining approach for converting unstructured data into a usable format. Data cleansing, data integration, data transformation, and data reduction are the four primary stages of this process. To ensure quality data and analysis outcomes, data cleaning will filter, detect, and treat unclean data. There may be noise in this scenario due to unrealistic and extreme values, outliers, and missing numbers. Inconsistent data, redundant attributes, and data are examples of errors. Null values in the dataset will be discovered and, if possible, correctly replaced as a first step. The most crucial phase in a machine learning project is data cleaning. The machine learning model's quality is determined by the quality of the data. As a result, data must always be cleaned before being fed to the model for training.

3. Exploratory Data Analysis

Exploratory data analysis (EDA) is used by data scientists to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.
The key objectives of EDA include:

- Identifying and extracting variables from the underlying structure.
- Detecting abnormalities and putting assumptions to the test.
- Increasing the value of data sets and selecting the best settings for factors.
- Selecting the tools required for training the data.
- Identifying new data-gathering channels

4. Model Building

The data is ready to be utilized to train a machine learning model after it has been collected and cleaned. The Decision Tree Support Vector Classifier, Multilayer Perceptron Neural Network, Naive Bayes Classifier, and Random Forest Classifier models will be used to train the clean data for this project.

5. Inference

After training the data with the selected four models, the disease for the input symptoms will be predicted by choosing the best performing model. This makes our overall prediction more robust and accurate.

1.2. Predictive Modelling

Predictive modelling, also known as predictive analytics, is a statistical method that attempts to predict future events or outcomes by studying patterns that are likely to tell future occurrences. The purpose of predictive modeling is to answer the following question: "What is most likely to happen in the future based on known past behavior?" [15], [16]. A predictive model isn't static; it's updated or validated on a regular basis to account for changes in the underlying data. Predictive models can also be customized for individual needs by incorporating user input in the form of information or assumptions. That is, this isn't a once-and-done prediction. Predictive models make assumptions based on what has already occurred and what is currently occurring. If new data arrives that reveals changes in the current situation, the impact on the expected future outcome must also be recalculated. These adjustments are made every time new data is acquired.

Machine learning uses predictive modeling to predict a variety of things, such as diseases and weather conditions [17]. Prediction is hard, and it can come roughly right most of the time, but even there, the process is never completely accurate. This project uses predictive modeling to take a patient's symptoms and predict or recognize the illness he or she may be suffering from. Predictive modeling has recently been applied to Parkinson's disease, heart cancer, hypertension, and other conditions as elaborated below.

2. REVIEW OF RELATED WORK

Much progress has been made in the field of medical diagnosis. Hospitals use database systems to help diagnose patients. These systems consist of doctors entering patients' symptoms. This is basically a "crutch" that helps doctors diagnose patients' ailments. Much technological research has been done to improve the algorithms of these systems in order to make them autonomous and easy to use [18]. Artificial intelligence has made some of the most significant advances in making these systems more intelligent and assist in the diagnosis of symptoms with greater accuracy. Machine learning has enabled many advancements in detecting cancerous cells. A typical scenario is a study by Gayathri et al.[19], who discovered that mammogram images occasionally have a risk of false detection, putting the patient's health at risk.

They used machine learning algorithms to predict and detect breast cancer in its early stages in order to find alternative and vital approaches to detect cancer. These machine learning methods are easier to implement and used with a variety of data sets, less expensive and safer, and can produce more accurate predictions. Support Vector Machine (SVM) [20], Artificial Neural Network (ANN) [21], K-Nearest Neighbor (KNN) [22], and Decision Tree (DT) [23] machine learning algorithms were used to predict breast cancer in patients in order to find the machine learning technique that produces the best and most accurate predictions.

According to the findings of these researchers, SVM is the most widely used method for cancer detection applications [19]. To improve performance, SVM was used alone or in combination with other methods. The maximum accuracy achieved by SVM (single or hybrid) was 99.8%, which can be improved to 100%. Another study by Amin Ul Haq et al. [24], used the "Cleveland heart disease dataset 2016" [25] to predict heart disease and discussed the use of machine learning algorithms to detect complex human diseases that can lead to early heart failure. In the system, popular algorithms such as Logistic Regression, K-NN, ANN, SVM, DT, and Naïve Baiyes (NB) were used to distinguish healthy people from those with heart disease. This resulted in the creation of a hybrid intelligent system framework for the prediction of heart disease.

Parkinson's disease is a medical problem that is currently undergoing extensive research, and it is critical to accelerate the development of more accurate models for early detection. Parkinson's disease is a progressive nervous system disease characterized by tremor, muscular rigidity, and slow, imprecise movement that primarily affects the middle aged and elderly. It is linked to degeneration of the brain's basal ganglia and a lack of the neurotransmitter dopamine [26].

The processing of voice signals for detecting Parkinson's disease using machine learning techniques is investigated in a study by Jefferson S. Almeida et al [27]. To classify data obtained from sustained phonation and speech tasks, the method compares the use of eighteen feature extraction techniques and four machine learning methods. Phonation is concerned with the voicing of the vowel /a/, while speech is concerned with the pronunciation of a short sentence in Lithuanian. The audio tasks were recorded using two microphone channels from an acoustic cardioid and a smartphone, allowing researchers to compare the performance of various microphone types. The researchers found that phonation tasks were more effective than speech tasks in detecting disease.

Gopi Battineni et al [28] conducted a study on machine learning (ML) predictive models in the diagnosis of chronic diseases. Because each machine learning method has advantages and disadvantages, their findings suggest that there are no standard methods for determining the best approach in real-time clinical practice. Support vector machines (SVM), logistic regression (LR), and clustering methods were the most commonly used methods among those considered. These models are extremely useful in the classification and diagnosis of chronic diseases, and they are expected to play a larger role in medical practice in the near future. The goal of the study is to use machine learning techniques to detect chronic diseases like diabetes, hepatitis, and other similar conditions early on, preventing medical complications.

These medical advances have aided the medical industry in a lot of aspects.

Machine learning has made major advances to disease diagnosis, detection, and prediction in medicine, according to all of the studies reviewed. These algorithms have mostly enabled scientists to discover new diseases earlier and with greater accuracy.

3. PROBLEM STATEMENT

Self-diagnosis has drawbacks, but it is a habit that cannot be completely abandoned. If every ailment was taken to a clinic or hospital, it would be overcrowded, resulting in expensive costs and travel time. If a person goes to an emergency room for a hypochondriac's concerns, it can be a long wait for little treatment ("Visit the doctor? He will just listen to my breathing and send me away").

A person seeking help for a self-medicated condition faces the risk of an annoying assessment by physicians and filling out unnecessary paperwork [8]. On the other hand, there are those who make a habit of self-medication and do not go to clinics because of the high cost. If the practice is widespread, it can lead to addiction, over dosage, and other problems. This circumstance creates a tough social environment.

There is also the danger of mistaking major illnesses for "minor" ones, which can progress to more serious problems if not handled properly [9], [10]. There are chronic illnesses that appear to be trivial at first. It's not a headache just because it's a headache. Some "minor" illnesses are stepping stones to more significant medical problems. For example, strep throat may seem like a run-of-the-mill cold, but strep throat can also lead to rheumatic fever [11].

4. OBJECTIVES OF THE RESEARCH

The goal of this project is to create an end-user health prediction system that serves as a guide to self-diagnose and also provide online consultation for better medical advice.

Specific Objectives of the Study

The specific objectives of the study are;

- To create a smart health prediction system that serves as a guide in diagnosing and consulting doctors after diagnosis is performed.
- To enable people have access to medical advice in the fastest way possible.
- To reduce over-dependence on healthcare facilities.
- To reduce the risks generated through self-diagnosis.

5. SCOPE OF STUDY

- The health prediction system has three users; doctor, patient and admin.
- Each user of the system is authenticated by the system.
- There is a role-based access to the system.
- The system allows the patient to select symptoms which they are experiencing for disease prediction.
- The system suggests doctors for predicted diseases.
- The system allows online consultation for patients.
- The system helps the patients to consult the doctor at their convenience by sitting at home.

The following are the stages the project will go through;

1. **Data Preparation** is the process of collecting, combining, structuring, and organizing data in order to make it usable.
2. **Data Transformation** is the process of changing the format, structure, or values of data is known as data transformation.
3. **Feature Extraction** is to reduce the number of resources needed to describe a large set of data. One of the major issues with performing complex data analysis is the large number of variables involved. An analysis with a large number of variables usually necessitates a lot of memory and processing power. It may also lead to an overfitting of

a classification algorithm to training samples and poor generalization to new samples. Feature extraction is a broad term that refers to methods for constructing combinations of variables to get around these issues while still adequately describing the data [12].

4. **Machine Learning Algorithms Implementation** involves implementing classification algorithms.
5. **Model Creation** is an iterative process in which data is continuously trained and machine learning models are tested to find the best one for the job.

The system preprocesses the data using data mining techniques before applying machine learning algorithms to the results. As a result, an accurate prediction of the likelihood (frequency) of a specific disease or condition can be made.

6. CONCLUDING REMARKS

Machine learning is a relatively new digital technology for solving problems in today's world [13]. This technique has grown particularly promising in tackling real-world problems due to its capacity to construct an algorithm and teach it data rather than write it down. This research is intended on using machine learning to improve the possibilities of self-diagnosis by creating a health care prediction system. This study will not prescribe pharmaceutical or medications that will be used in conjunction with the condition. To avoid keeping sensitive patient data, the system's input will exclude a patient's medical history and/or lab test results. In this case, the product will a be a prototype which is a web based app, so a mobile app will be a secondary priority. Prototyping will be done using a Django framework app. Django is a free and open-source web framework written in Python that follows the model–template–views architectural paradigm.

REFERENCES

[1] L. A. F. Souza, C. D. da Silva, G. C. Ferraz, F. A. E. F. Sousa, and L. V. Pereira, "The prevalence and characterization of self-medication for obtaining pain relief among undergraduate nursing students," Rev. Lat. Am. Enfermagem, vol. 19, pp. 245–251, Apr. 2011, doi: 10.1590/S0104-11692011000200004.

[2] K. A. Bashrahil and A. S. Baruzaig, "Self-medication: concept, prevalence & risks in Mukalla City (Yemen) 2004–2005," Andal. Stud. Res, vol. 2, pp. 1–3, 2008.

[3] A. Befekadu, "Self Medication", Accessed: Jan. 17, 2022. [Online]. Available: https://www.academia.edu/4228902/Self_Medication

[4] A. Sharrif, D. Almasdy, and A. Sherrif, "Self-Medication Practice with Nonprescription Medication among University Students: a review of the literature," Arch. Pharm. Pract., vol. 2, no. 3, p. 7, 2011.

[5] L. Thomas, The technology of medicine, vol. 285, no. 24. Mass Medical Soc, 1971, pp. 1366–1368.

[6] K. J. Petrie and S. Wessely, Modern worries, new technology, and medicine, vol. 324, no. 7339. British Medical Journal Publishing Group, 2002, pp. 690–691.

[7] S. J. Reiser and S. J. Reiser, Medicine and the Reign of Technology. Cambridge University Press, 1981.

[8] H. N. Sherwin, M. McKeown, M. F. Evans, and O. K. Bhattacharyya, "The waiting room 'wait': From annoyance to opportunity," Can. Fam. Physician, vol. 59, no. 5, pp. 479–481, 2013.

[9] S. B. Agampodi, N. D. Wickramasinghe, J. Horton, and T. C. Agampodi, "Minor ailments in pregnancy are not a minor concern for pregnant women: a morbidity assessment survey in rural Sri Lanka," PloS One, vol. 8, no. 5, p. e64214, 2013.

[10] W. C. Cockerham, M. C. Creditor, U. K. Creditor, and P. B. Imrey, "Minor ailments and illness behavior among physicians," Med. Care, pp. 164–173, 1980.

[11] "Strep Throat: All You Need to Know | CDC." https://www.cdc.gov/groupastrep/diseases-public/strep-throat.html (accessed Jun. 27, 2022).

[12] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Feature extraction: foundations and applications, vol. 207. Springer, 2008.

[13] K. Siau and W. Wang, “Building trust in artificial intelligence, machine learning, and robotics,” Cut. Bus. Technol. J., vol. 31, no. 2, pp. 47–53, 2018.
[14] J. Alzubi, A. Nayyar, and A. Kumar, “Machine learning from theory to algorithms: an overview,” in Journal of physics: conference series, 2018, vol. 1142, no. 1, p. 012012.
[15] C. Brooks and C. Thompson, “Predictive modelling in teaching and learning,” Handb. Learn. Anal., pp. 61–68, 2017.
[16] L. Swani and P. Tyagi, “Predictive Modelling Analytics through Data Mining,” Int. Res. J. Eng. Technol. IRJET, vol. 4, no. 09, 2017.
[17] C. Mair et al., “An investigation of machine learning based prediction systems,” J. Syst. Softw., vol. 53, no. 1, pp. 23–29, 2000.
[18] P. Szolovits, R. S. Patil, and W. B. Schwartz, “Artificial intelligence in medical diagnosis,” Ann. Intern. Med., vol. 108, no. 1, pp. 80–87, 1988.
[19] B. M. Gayathri, C. P. Sumathi, and T. Santhanam, “Breast cancer diagnosis using machine learning algorithms-a survey,” Int. J. Distrib. Parallel Syst., vol. 4, no. 3, p. 105, 2013.
[20] W. S. Noble, “What is a support vector machine?,” Nat. Biotechnol., vol. 24, no. 12, pp. 1565–1567, 2006.
[21] B. Yegnanarayana, Artificial neural networks. PHI Learning Pvt. Ltd., 2009.
[22] M.-L. Zhang and Z.-H. Zhou, “ML-KNN: A lazy learning approach to multi-label learning,” Pattern Recognit., vol. 40, no. 7, pp. 2038–2048, 2007.
[23] S. R. Safavian and D. Landgrebe, “A survey of decision tree classifier methodology,” IEEE Trans. Syst. Man Cybern., vol. 21, no. 3, pp. 660–674, 1991.
[24] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, “A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms,” Mob. Inf. Syst., vol. 2018, p. e3860146, Dec. 2018, doi: 10.1155/2018/3860146.
[25] “UCI Machine Learning Repository: Heart Disease Data Set.” https://archive.ics.uci.edu/ml/datasets/heart+disease (accessed Jan. 24, 2022).
[26] A. E. Lang and A. M. Lozano, “Parkinson’s disease,” N. Engl. J. Med., vol. 339, no. 16, pp. 1130–1143, 1998.
[27] J. S. Almeida et al., “Detecting Parkinson’s disease with sustained phonation and speech signals using machine learning techniques,” Pattern Recognit. Lett., vol. 125, pp. 55–62, 2019.
[28] G. Battineni, G. G. Sagaro, N. Chinatalapudi, and F. Amenta, “Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis,” J. Pers. Med., vol. 10, no. 2, Art. no. 2, Jun. 2020, doi: 10.3390/jpm10020021.
[29] “Disease Prediction Using Machine Learning.” https://www.kaggle.com/kaushil268/disease-prediction-using-machine-learning (accessed Mar. 28, 2022).
[30] A. D’Angela, “Why Weight? The Importance of Training on Balanced Datasets,” Medium, Feb. 04, 2021. https://towardsdatascience.com/why-weight-the-importance-of-training-on-balanced-datasets-f1e54688e7df (accessed Mar. 28, 2022).
[31] “Gastroesophageal reflux disease (GERD) - Symptoms and causes,” Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/gerd/symptoms-causes/syc-20361940 (accessed May 25, 2022).
[32] “(99+) Dr. Rakshith Raj Bharadwaj | LinkedIn.” https://www.linkedin.com/in/drrakshith/ (accessed May 26, 2022).
[33] B. C. M. Keong and W. Sulaiman, “Typhoid and Malaria Co-Infection – An Interesting Finding in the Investigation of a Tropical Fever,” Malays. J. Med. Sci. MJMS, vol. 13, no. 1, pp. 74–75, Jan. 2006.
[34] M. H, “Typhoid, Malaria, Dengue Fever,” Grabbing Life By the Lapels, Jan. 27, 2016. https://penthesilearising.wordpress.com/2016/01/27/typhoid-malaria-dengue-fever/ (accessed May 27, 2022).
[35] “Malaria - Symptoms and causes - Mayo Clinic.” https://www.mayoclinic.org/diseases-conditions/malaria/symptoms-causes/syc-20351184 (accessed May 27, 2022).
[36] “Typhoid fever - Symptoms and causes,” Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/typhoid-fever/symptoms-causes/syc-20378661 (accessed May 27, 2022).
[37] “What is Natural Language Processing? | IBM.” https://www.ibm.com/cloud/learn/natural-language-processing (accessed May 27, 2022).