Predicting the Side Effects of Medical Drugs Using Sentimental and Classification Mining Algorithms: A Case Study of Hydroxychloroquine and Azithromycin Drugs

ALI-OKOROSOCHIMA GODSON[1], OLEBARA COMFORT CHINAZA[2], OGUOMAIKECHUKWU STANLEY[3]

[1] *Department of Computer Science, Imo State University Owerri, Nigeria*

[2, 3] *Department of Computer Science, Faculty of Physical Science, Imo State University Owerri, Nigeria*

Abstract- ***Identification of a problem before it gets worst is a very important aspect of protection more especially when it involves human lives. Drugs are taken for two main reasons, either for cure or for prevention. The motivation behind the study is because of high increase of complains received from most hospitals and clinics within Nigerian medical field, some drug side effects are so disturbing that its results causes abnormal disorder on patients when used for medical treatment more precisely when used on pregnant women. More so, the lack of dataset on most drugs from peoples view or opinion on the drugs which could make it easier for academic and scientific research to be conducted. Furthermore, this study is aimed at predicting the side effects of Hydroxychloroquine and Azithromycin drugs using sentimental and classification data mining algorithms. The study also provides an online platform which could enable medical professionals and the general public to have access to provide report or feedback on the consumed drugs. The online platform was developed with HTML,CSS, PHP and MYSQL while the dataset was sourced from UCI machine learning repository and analyzed using R language and RStudio. SEMMA which stands for Sample Explore Modify Model Access was applied as the methodology while aspect-base sentimental analysis and Classification decision tree Model algorithms was adopted. The result after the experiment created a model that was able to predict patient's opinion and reviews on consumption of the drugs base on their various dosage, age and number of patients while the online interface of the drug effects feedback reporting was developed and deployed in the cloud for fast and easy access for users remotely.***

Indexed Terms- ***Machine Learning, Drug Effects, Classification Model, Sentimental analysis, Data mining and online Drug Effect Platform***

I. INTRODUCTION

Application of machine learning and data mining algorithm in the prediction of side effects of drugs on humans in Nigeria is an intelligent model that could enable medical practitioners in Nigeria effectively identify or predict the presents of drugs effects and reactions of patients and also identify drugs with such effects in Nigeria. It is also very much difficult to have access to an online platform for the reporting and easy collections of data base of drugs effects and to encourage researchers provide accurate result on the analysis of those feedbacks from patients.

The rate at which there is increase in the number of patient hospitalized as a result of the adverse reaction of drugs is quite alarming. Medical drugs are said to have adverse reaction if the outcome is other than the intended therapeutic effect, whether beneficial, neutral or harmful. The harmful reaction of the drugs contributes an average of the 10-25% death recorded per annum and this is based on the negligence of the scheme (yellow card) provided by the government. Mostly, when a drug outperform its uses, patients fails to connect it to the side effect of the drug rather they prefer to discontinue the drug in order to minimize the sensation that occurred [1]Side effect

also one of the most delay/failure in the drugs development process, as the tenderness to overcome proves to be indomitable. Medical drugs are design in the sense that before being post marketed, a series of testing are carried out in various stages in order to eradicate any harmful effect that the drugs might contain.

The various stages range from the drug production to the preclinical test on animals and then, clinical tests on human volunteer. In the process, the reactions associated with the specified medical drugs are detected and it is check to know the degree of the effect, as some of the effects are classified as mild examples are doziness, headache, nausea and itching. Due to the high increase of Hydroxychloroquine and Azithromycin by Nigerians as it is an antibiotic drug and Nigerian are used to self-medication therapy, it was observed that its deficiency and effects are increasing on a daily basis, there is need to identify its effects on time.

Nevertheless, the use of data mining into data discovery, prediction and extraction has helped researcher and scientist in numerous ways to uncover hidden facts both in physical and biological sciences. Data mining (DM) is one of the hidden extractions of predictive tool in progression from a very big databases, is now a new great technology by means of large potential to provide huge support to companies looking into the direction of large data discovery from knowledge warehouses. It is now clear that the introduction of DM tool in data processing has help to predict the future trends and behaviors, enabling various business owners create practical knowledge-driven selections in such that it could help to answer business queries that factually remained too time irresistible to resolve [2]. On that aspect, opinion of consumer's means so much to organizations, as it assists in the improvement and production of new products. For an organization to grow there is need for them to listen to the reviews (customers' feedback) of their customers if neglected improvement and growth of the organization will be static and hence affect the growth of the organization. In Nigerian health sector, the responds of patient is so much paramount important to the physician as it enable good examination to be carried out in providing good treatment to the patent, that is to say inclusion of all the treatments given such as drugs, test, physical experiments and cancelling. There are big pharmaceutical companies in UK and other countries in Africa, manufacturing drugs for the treatment of various illnesses, it is noted that every drug has side effects, but there are some that are very risky to be administered to patients and if such occur there should be a feedback call to the manufacturers of such drug to art immediately especially on the side effect of Hydroxychloroquine and Azithromycin as they are used as an antibiotic for the treatment of some strong bacteria. When there is no system to address such issue from reoccurring again it becomes a problem which might cause the patient's life. That is to say there should be an intelligent model to always review patient's opinion in various hospitals on the side effects of drugs administered on a routine basis for good decision making in drug production and to identify areas this case in much maybe because of the geographical location or any other cause. Therefore, this study calls to close this gap by providing a predicting model that can analyze patient's opinion in various hospitals on the side effects of drugs administered to them by also providing an online platform for such actions to be collected and handled as fast as possible by drug production companies. General Introduction to some concept in both sentimental analysis and classification machine learning algorithm, aim and objectives of the research, Reviewed Literatures, Adopted methodology, components of the dataset, while Results experiment of on the dataset on R by using RStudio, findings (results), conclusion, recommendations and appendix.

II. LITERATURE REVIEW

According to [3] worked on adverse drug reactions (ADRs) on patients; the study presented a unique literature-mining background the process of improving the predictions of both drug-drug interactions (DDIs) and (ADR) using a random forest classifier approach. The result of the research found that F-score of 0.87 across the ADRs classification by the uses of the only the DDI features was used to check the performance of the DGIs with (F-score = 0.90), and applied the classification model trained with the DDI corpus to identify the drugs that might interact with the drugs for cutaneous diseases. The

study successfully predicted that the known ADRs of the drugs prescribed to be cutaneous disease and was able to detect auspicious new ADRs.

[4] Worked on the adverse drug reactions on patients using 6,100 concepts and 4,700 adverse drug reactions (ADRs) for training. The analysis was done on 2,100 concepts and 1,600 ADR relations, and obtained 0.93 precision and 0.85 recalls using the Random Forest algorithm.

[5] Worked on a study to identify the adverse drug effects worked using 3,012 Japanese discharge summaries, the study explained 1,045 drugs and 3,601 possible adverse drug effects. The study identified about 7.7% of the adopted discharge summaries of ADE. The approach adopted by the scholar was support vector machine (SVM) and pattern matching algorithms also in a similar study by [6], they worked on a study using UMLs and Naïve Bayes classifier to identify the adverse drug effects concepts in millions of biomedical articles. The Naïve Bayes classifier algorithm was used for the train dataset obtaining a 0.25 precision and 1.00 recalls after the analysis.

2.1 Sentimental Analysis

[7] Is of the view that mining of the Opinion are classified centered on users choice for the specific subject which is gotten from the analyses. Sentimental mining is one which categorizes the criticism wording as negative opinion or positive opinion. Online reviews or evaluations can be improved as stated [8] that it will be complicated for various customers might distinguish the essential evaluations which could be gotten from the valuable ones. The prognostic and descriptive are the two groups that can be divided in the data mining technique. The Data mining is one of the processes to find out the encouraging information, such as irregularities, designs changes, important constructions and relations from a large quantity of information kept in the database [9].

[10] Defined sentimental analysis data mining as a method typically used by scholars is bag-of-words method acquired from a virtual review or social media. The method, instead of viewing into the entire sentence or somewhat section for the analysis, it will only ponder individual's words and their count as the feature vectors. This scholar's research ensemble method of the experiment outperformed the old-style bag-of-words method with enough difference and other approaches used; ensemble approach showed that tremendously randomized trees classification presentation was far improved than others. The adoption of lexicon and linguistic structure was done by [11] for the analysis of a dataset while [12] implemented theirs using the traditional bag-of-words approach after the learning from an external data. Another scholar [13] adopted the association rule-based system using aspect identification sentiment while Secondly [14], brought out outlined four dissimilar events that can be applicable in data extraction using sentimental analysis. These events are: Supervised learning and topic models, target relation, frequent phrases, and opinion mining and conditional random field's aspect sentimental mining was used by [15] to carry out their research on a dataset.

III. METHODOLOGY

The study adopted SEMMA which stands for Sample Explore Modify Model Access which focuses on the main modeling tasks in the project without venturing into the business understanding and deployment according to oreilly.com. SEMMA was introduced by SAS, and it is logical organization of the functional tool set of SAS Enterprise Miner for carrying out the core tasks of data mining.

3.1 Classification Algorithms

Classification algorithm is stated by [16] as a supervised learning method were a computers are programmed to learn from set inputs as data and then uses the learning approach to classify new observations. The researcher further explained that the observed data set could be bi-class. Sometime, the class might be multi-class [16].Classification modeling can also be applied in areas like, speech recognition, handwriting identification, bio-metric identification or verification process, document classification [16]. The scholar outlined some of the few classification algorithms in machine learning as follows: Linear Classifiers: Logistic Regression, Naive Bayes Classifier, Nearest Neighbor, Support

Vector Machines, Decision Trees, Boosted Trees, Random Forest and Neural Networks.

3.1.1 Adopted Algorithms for the Study

The study applied two different algorithms namely: Decision tree model and aspect base sentimental analysis algorithm.

1. Decision Trees

Decision tree [17] could be seen as a type of tree structure typically in a form of flowchart design. These tree structures are used to carry out classification and prediction modeling of objects in a class in a form of nodes and internodes. Both root and the internal nodes are taken as the test cases in the modeling process which in terms used as a separator with different features [17]. According to [16], this decision tree uses a classification or regression models to form a tree structure. The structure breaks down a particular data set into various smaller and smaller subsets as the associated decision tree development id in progress. The researcher further noted that the decision tree is build up with the nodes and leaf nodes, where the decision nodes has two or more different branches while leaf nodes shows the classification or decision results [16] stated. Figure 1: Illustrate the structure of a decision tree

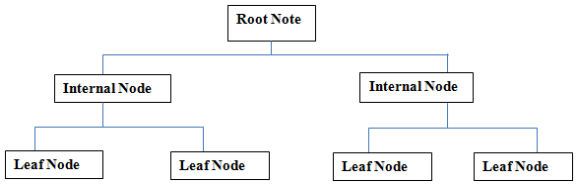


Figure 1: Exemplify the organization of a decision tree [18].

Adoption of decision tree for this study did not just come but it was adopted because of its powerful technique for classification and prediction ability on a particular data set. Hence the identification and prediction of drug effect on Hydroxychloroquine and Azithromycinwill have a very significant outcome after the analysis of the data set has been concluded and presented for future use.

2. Aspect based sentimental analysis

Aspect based sentimental analysis, was adopted for this research because it is one of the best approach for analyzing written text or report by customers on a particular service or product. It makes use of (positive and negative perspective). The data set was gotten from (https://https://www.kaggle.com/ and was saved as XLS (Microsoft Excel Format). From the excel file.

3.1.2 Exploration of Dataset

The dataset provides patient reviews on specific drugs along with related conditions and a 10 star patient rating reflecting overall patient satisfaction. The data was obtained by crawling online pharmaceutical review sites on the responses of users on the Hydroxychloroquine and Azithromycin drug.

The intention was to study:

1. Carryout a sentiment analysis of drug experience over multiple facets, i.e. sentiments learned on specific aspects such as effectiveness and side effects and its deficiency,
2. To study the transferability of models among domains.

3.1.2.1Attributes of the Dataset

1. Patient (Number in patients)
2. Age_in_years (Patient Years in Age)
3. Sex (Patient gender)
4. Clinical status (Patients Clinical status on the use of drug)
5. Time_between_on (Time between onset of symptoms and inclusion in days)
6. Hydroxychloroq (Hydroxychloroquine treatment)
7. Hydroxychloroq (Hydroxychloroquine_serum_concentration_Î¼g/ml_in_day_of_dosage)
8. Azithromycin_tr (Azithromycin treatment)

3.2 Analysis of the Existing System

This study looked at the existing effects of drugs on patients in Nigeria especially during the COVID-19 infectious outbreak and how the use of drugs like Hydroxychloroquine and Azithromycin which are some of the drugs used to fight the bacteria was used. Azithromycin is an antibiotic drug which are used to fight some infectious illness like acute bacteria of exacerbations of chronic obstructive pulmonary disease, skin infections, urethritis and cervicitis,

trachomatis, genital ulcer, influenza, pharyngitis or tonsillitis etc. The study discovered that there have been comments or feedback from consumers, (patients) on the side effects of the drugs such that only 5% reported effect on diarrhea, 3% report on the effect from abdominal pain, nausea 3% and that of vomiting 1%.

Base on the existing dataset on the effects of these drugs, the study reanalysis the response such that more updated result could be provided for more proper production which in turn solves the weaknesses identified such as: increase of fake drugs in Nigerian market, inability to diagnose and provide quick treatment of side effects of drugs by Nigerian Doctors and increase of drug consumption in Nigeria without adequate means of finding their side effects

3.2.1 Analysis of the Proposed Model

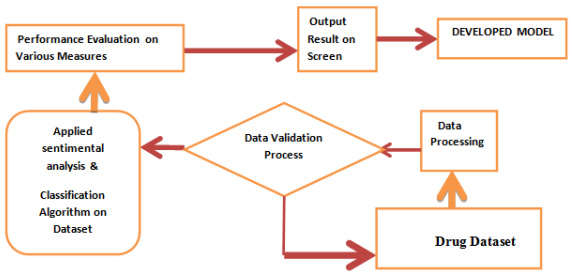


Figure 2: Analysis of the proposed Model

The proposed system was achieved by using R language and the R-Studio platform. The dataset of the Hydroxychloroquine and Azithromycincontains 1036 observations and 8 variables which were first processed, and then the data validation process begins after which the application of sentimental analysis algorithm and classification algorithm was applied on the dataset one after the other. After the application process of the two different algorithms, which helps for more clearer decision making on both usage of the drugs and the various customers dosage and age indication shown in figure (3) and age bracket of the patient shown in figure (4) respectively leads to the performance evaluation stage then to output of the result on the new model developed.

3.3. High Level Model of Proposed System

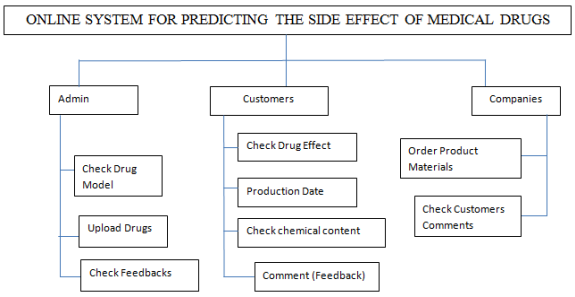


Figure 3: high level model of proposed system

The above diagram shown in figure 5 presents the hierarchical structure on the developed online base platform for fast review, collection and provision of feedback by customers (patients) on the side effect of produced drugsin other to ensure that more accurate update of Hydroxychloroquine and Azithromycin and other drugs dataset are collected for improvement. The diagram figure (3) contains three different users which includes admin user, customers (patients user), companies users. The functions of each user are shown in the use case diagram in figure 4, 5 and 6 respectively.

1. Customers (patients)User Requirement Module

Customers from various part of the world can Check Drug Effect of drug, view drug production date, check chemical content of drug, and give comment based on feedback from their observation and suggestion on the drug. Figure (4) below shows the use case diagram of the customers module.

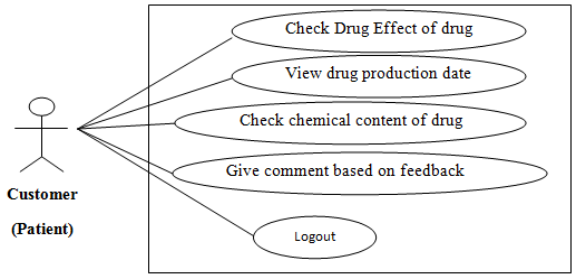


Figure 4: Customer User Module Use Case Diagram

2. Companies User Requirement Module

The requirement analysis for companies Requirement module can be transformed into the use case diagram as shown Figure 5 below:

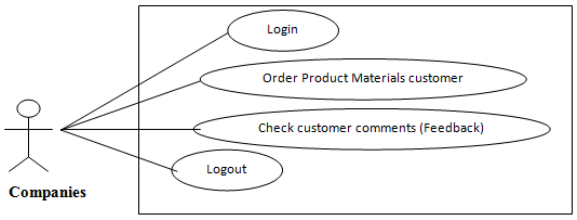


Figure 5: Companies User Module Use Case Diagram

3. ManagerUser Requirement Module

In every organization, the administrator oversees the entire activities within the required department. Therefore this module must be handled by the manager (admin) to perform the outlined task as shown in figure (6) of the use case diagram below;

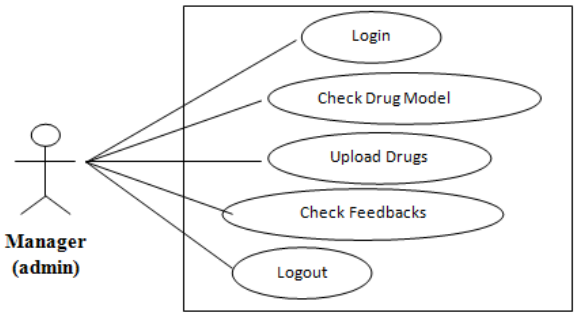


Figure 6: admin (manager) User Module Use Case Diagram

IV. RESULTS

- Experiments on the Dataset Using R

The first process was launching of the RStudio IDE after a successive launching, the following steps were done to design the model.

Step1: Loading packages to be used (that is libraries)

Step 2: Loading My Dataset to R Data frame

Step 3: Exploring the data, at this stage, skimr::skim(DataIHU) was used

Step 4:Checking the head(dataIHU) and tail(dataIHU)

Step 5:View (dataIHU) this is the stage to check all the dataset and its variables and columns which contain data on the two drugs.

Step 6:Plot Observations (Graphs between ages of patients and number of patients that reported on the effects of the drugs)

- Model Building

Step 7: The data set was Split into two with the percentage of (75% = training and 25%=testing) respectively

Step 8: applying decision tree algorithm

Step 9:applying sentimental analysis on the dataset

Step 10:On this stage, Prediction was made from Model built with the Giniindex Model

Step 11: Generating Frequency Table to Create Tabular Results of Categorical Variable

Tabular results enable the researcher to know if the result produce from the prediction by the computer is correct when compared with the original one on the dataset.

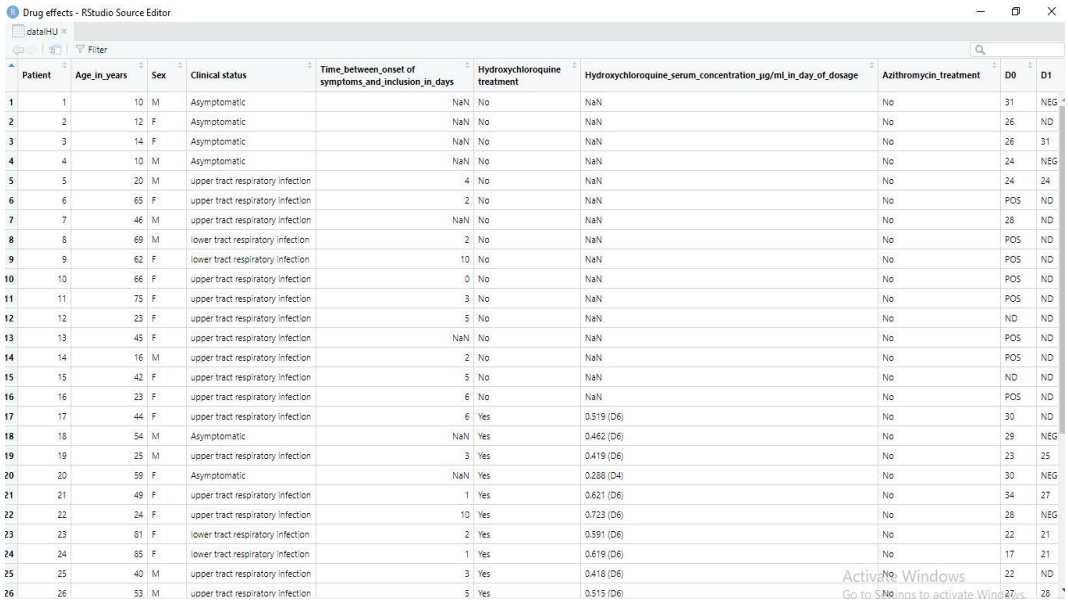
- EXPERIMENT OUTPUT

```
> tail(dataIHU)
# A tibble: 6 x 15
  Patient Age_in_years Sex   `Clinical status` `Time_between_on~ `Hydroxychloroq~ `Hydroxychloroqu~ Azithromycin_tr~ D0
    <dbl>       <dbl> <chr> <chr>                         <dbl> <chr>           <chr>              <chr>            <chr>
1      31          53 F     lower tract resp~                 7 Yes             1.076 (D6)         Yes              28
2      32          48 M     upper tract resp~                 2 Yes             0.57 (D6)          Yes              23
3      33          50 F     lower tract resp~                 5 Yes             0.827 (D6)         Yes              30
4      34          20 M     upper tract resp~                 2 Yes             0.381 (D6)         Yes              27
5      35          54 M     lower tract resp~                 6 Yes             0.366 (D4)         Yes              24
6      36          60 M     lower tract resp~                 4 Yes             0.319 (D4)         Yes              29
# ... with 6 more variables: D1 <chr>, D2 <chr>, D3 <chr>, D4 <chr>, D5 <chr>, D6 <chr>
>
```

Figure 7: This shows the tail view of the drug dataset.

```
> head(dataIHU)
# A tibble: 6 x 15
  Patient Age_in_years Sex   `Clinical status` `Time_between_on~ `Hydroxychloroq~ `Hydroxychloroqu~ Azithromycin_tr~ D0
    <dbl>       <dbl> <chr> <chr>                         <dbl> <chr>           <chr>              <chr>            <chr>
1      1          10 M     Asymptomatic                    NaN No              NaN                No               31
2      2          12 F     Asymptomatic                    NaN No              NaN                No               26
3      3          14 F     Asymptomatic                    NaN No              NaN                No               26
4      4          10 M     Asymptomatic                    NaN No              NaN                No               24
5      5          20 M     upper tract resp~                 4 No              NaN                No               24
6      6          65 F     upper tract resp~                 2 No              NaN                No               POS
# ... with 6 more variables: D1 <chr>, D2 <chr>, D3 <chr>, D4 <chr>, D5 <chr>, D6 <chr>
>
```

Figure 8: This shows the head view of the drug dataset.



	Patient	Age_in_years	Sex	Clinical status	Time_between_onset of symptoms_and_inclusion_in_days	Hydroxychloroquine treatment	Hydroxychloroquine_serum_concentration_µg/ml_in_day_of_dosage	Azithromycin_treatment	D0	D1
1	1	10	M	Asymptomatic	NaN	No	NaN	No	31	NEG
2	2	12	F	Asymptomatic	NaN	No	NaN	No	26	ND
3	3	14	F	Asymptomatic	NaN	No	NaN	No	26	31
4	4	10	M	Asymptomatic	NaN	No	NaN	No	24	NEG
5	5	20	M	upper tract respiratory infection	4	No	NaN	No	24	24
6	6	65	F	upper tract respiratory infection	2	No	NaN	No	POS	ND
7	7	46	M	upper tract respiratory infection	NaN	No	NaN	No	28	ND
8	8	69	M	lower tract respiratory infection	2	No	NaN	No	POS	ND
9	9	62	F	lower tract respiratory infection	10	No	NaN	No	POS	ND
10	10	66	F	upper tract respiratory infection	0	No	NaN	No	POS	ND
11	11	75	F	upper tract respiratory infection	3	No	NaN	No	POS	ND
12	12	23	F	upper tract respiratory infection	5	No	NaN	No	ND	ND
13	13	45	F	upper tract respiratory infection	NaN	No	NaN	No	POS	ND
14	14	16	M	upper tract respiratory infection	2	No	NaN	No	POS	ND
15	15	42	F	upper tract respiratory infection	5	No	NaN	No	ND	ND
16	16	23	F	upper tract respiratory infection	6	No	NaN	No	POS	ND
17	17	44	F	upper tract respiratory infection	6	Yes	0.519 (D6)	No	30	ND
18	18	54	M	Asymptomatic	NaN	Yes	0.462 (D6)	No	29	NEG
19	19	25	M	upper tract respiratory infection	3	Yes	0.419 (D6)	No	23	25
20	20	59	F	Asymptomatic	NaN	Yes	0.288 (D4)	No	30	NEG
21	21	49	F	upper tract respiratory infection	1	Yes	0.621 (D6)	No	34	27
22	22	24	F	upper tract respiratory infection	10	Yes	0.723 (D6)	No	28	NEG
23	23	81	F	lower tract respiratory infection	2	Yes	0.591 (D6)	No	22	21
24	24	85	F	lower tract respiratory infection	1	Yes	0.619 (D6)	No	17	21
25	25	40	M	upper tract respiratory infection	3	Yes	0.418 (D6)	No	22	ND
26	26	53	M	upper tract respiratory infection	5	Yes	0.515 (D6)	No	27	28

Figure 9: Overview of the dataset, showing all the variables and observations/columns

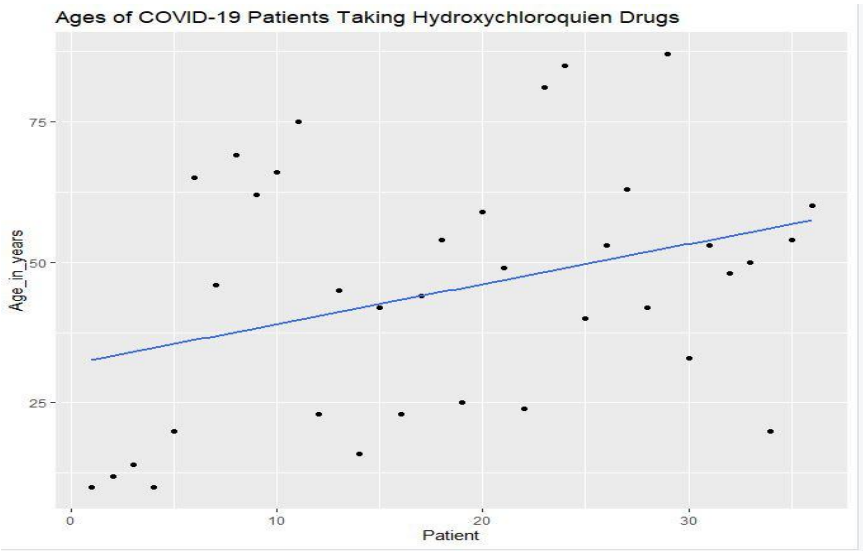


Figure 10: Ages of covid-19 patients taking hydroxychloroquien drugs

The above plot shows that ages of patients taking the drugs are from 25 years and above. Critically looking at the graph shown in figure 10 above, it shows that use of hydroxychloroquien drug for the treatment of COVID-19 effectious disease are used and effective on patient between the bracket of 25years and above meaning that its effects are low between the age bracket.

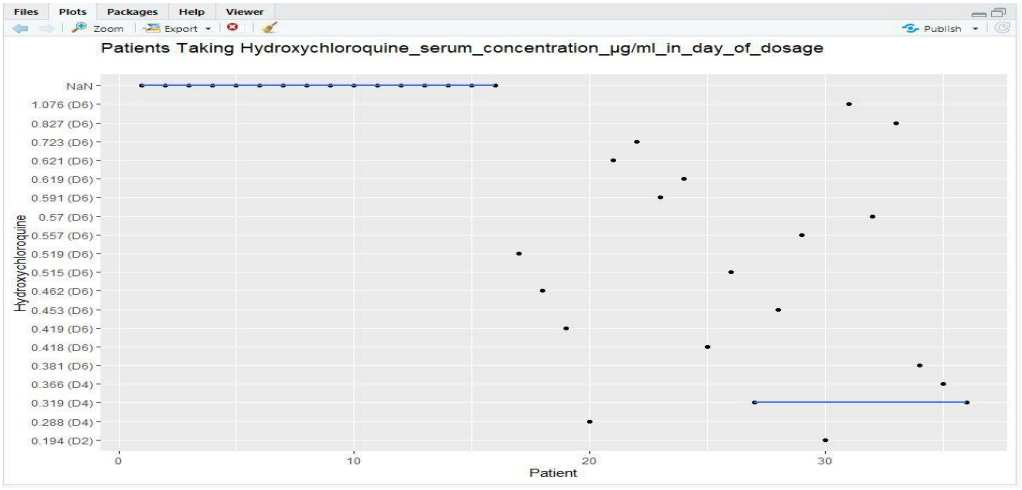


Figure 11: showing the intake of hydroxychloroquien serum concentration base on dosage and uses by patient.

- Result of the Online Drug Effect Reporting Platform

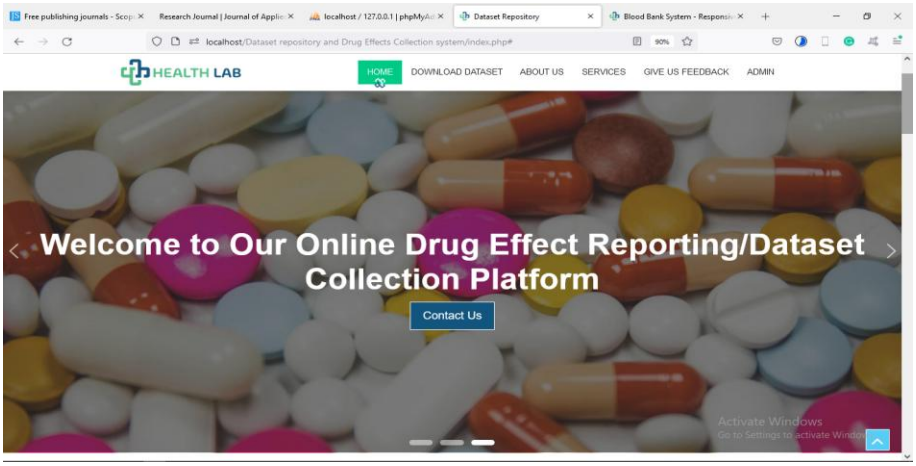


Figure 12: System Landing page 1

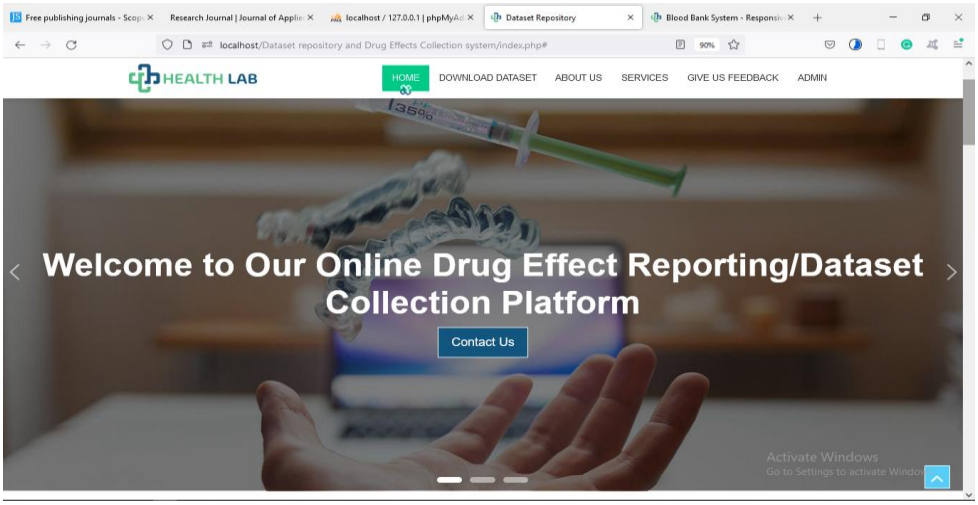


Figure 13: Landing Page 2

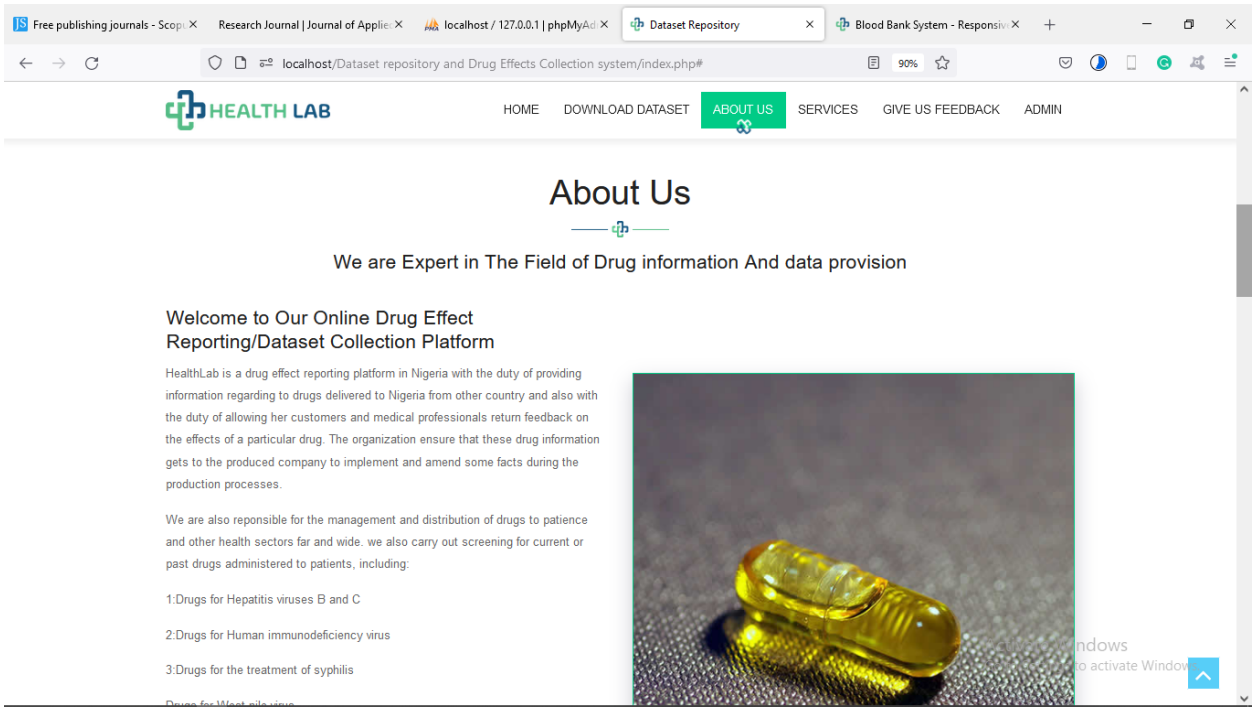


Figure 14: About Us Page: The page shows the information and detailed explanation of the drug effect reporting online platform which helps patients and the general public to know more about the organization.

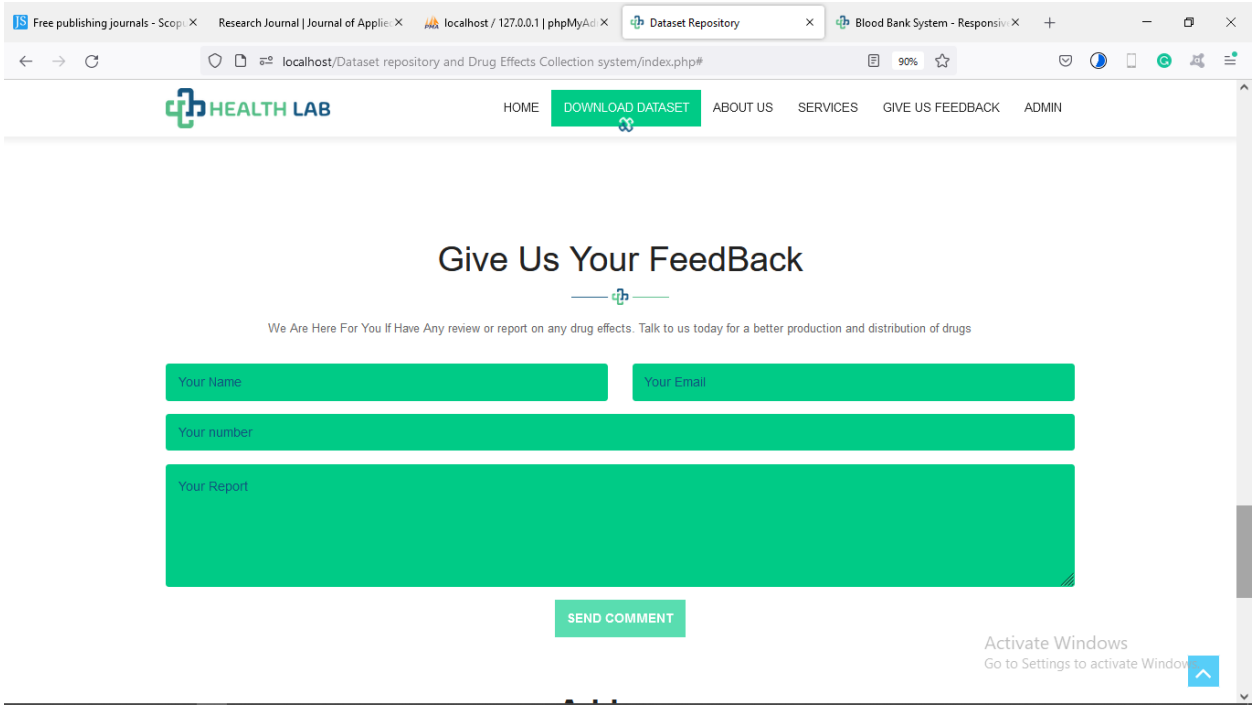


Figure 14: This page provide an interface for opinion/comment reporting about a particular drug from consumers worldwide from a remote location.

CONCLUSION

As earlier stated, that the aim of this work is to Predicting the Side Effects of Medical Drugs Using Sentimental and Classification Mining Algorithms. The studies looks at the identification and prediction on the effects of Hydroxychloroquine and Azithromycin drugs and also provide an online platform that could enable the public especially those in the medical field to easily identify the effects of a particular drug and at the same time provide a feedback on the drug after consumption from the online platform remotely.

5.1 Recommendation

The researcher therefore recommends the following:

1. The researcher therefore recommend the Nigerian health sectors and general public to adopt and always involve the use of artificial intelligence and machine learning towards the prediction and provision of health related issues for fast solving and efficient delivery.
2. The online platform for collection of peoples view or opinion towards the effects of drug in other to provide enough data for researchers.

REFERENCES

[1] Pharmacologyeducation.org. (2021). Clinical Pharmacology | Pharmacology Education Project. Retrieved 20 September 2021, from https://www.pharmacologyeducation.org/clinical-pharmacology.

[2] Pravarti Jain And Santosh Kr Vishwakarma (2017) A Case Study on Car evaluation and Prediction: Comparative Analysis using Data mining Models, International Journal of Computer applications (0975 – 8887) Volume 172 – No.9

[3] Kalpana Raja, Matthew Patrick, James T. Elder & Lam C. Tsoi (2017) machine learning workflow to enhance predictions of Adverse drug Reactions (ADRs) through drug-gene interactions: application to drugs for cutaneous diseases, retrieved from https://www.nature.com/scientificreports

[4] Sara Santiso, Alicia Pérez, KoldoGojenola, IXA Taldea, ArantzaCasillas, and maiteOronoz. (2014).Adverse Drug Event prediction combining shallow analysis and machine learning. In proceedings of the 5th International Workshop on Health Text mining and Information Analysis (Louhi)@ EACL,

[5] Aramaki, E., Miura, Y., Tonoike, M., Ohkuma, T., Masuichi, H., Waki, K., &Ohe, K. (2010). Extraction of adverse drug effects from clinical records. Studies in health technology and informatics, 160(Pt 1), 739—743.

[6] Roller, R., & Stevenson, M. (2014). Applying UMLS for Distantly Supervised Relation Detection. Proceedings Of The 5Th International Workshop On Health Text Mining And Information Analysis (Louhi). https://doi.org/10.3115/v1/w14-1112

[7] Vijay B. Raut, D.D. Londhe, (2014) Opinion Mining and Summarization of Hotel reviews, Sixth International Conference on Computational intelligence and Communication Networks, Pune, India

[8] BetulDundar, SuatOzdemir, DiyarAkay (2016)"Opinion Mining and Fuzzy Quantification in Hotel Reviews", IEEE TURKEY,

[9] GaganjotKaur, AmitChhabra (2014) "Improved J48 Classification Algorithm Github.com (no date) "gitten" available from https://github.com/yihui/knitr(CRAN

[10] Renuka S. Deshmukh, Dr. Aashish A. Bardekar(2020) EVOLUTION OF SENTIMENTAL ANALYSIS, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

[11] Kiritchenko, S., et al(2014) NRC-Canada-2014: Detecting aspects and sentiment incustomer reviews. in Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014).

[12] Castellucci, G., et al.(2014) Unitor: Aspect based sentiment analysis with structured learning. in Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). 2014.

[13] Hu, M. and B. Liu (2004) Mining opinion features in customer reviews. in AAAI.

[14] Liu, M., Wu, Y., Chen, Y., Sun, J., Zhao, Z., & Chen, X. et al. (2012). Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. Journal Of The American Medical Informatics Association, 19(e1), e28-e35. https://doi.org/10.1136/amiajnl-2011-000699

[15] Jakob and Gurevych (2010) Jakob, N. and I. Gurevych (2010) Extracting opinion targets in a single-and cross-domain setting with conditional random fields. in Proceedings of the 2010 conference on empirical methods in natural language processing. 2010. Association for Computational Linguistics.

[16] Sidana (2017) A Case Study on Car evaluation and Prediction: Comparative Analysis using Data mining Models, International Journal of Computer applications (0975 – 8887) Volume 172 – No.9

[17] Jiawei Han and MichelineKamber, (2001) "Data Mining Concepts and Techniques", Morgan Kauffman Publishers.

[18] Pilani, Dubai, BITS Pilani, Dubai and RonakSumbaly (2015) Diagnosis of Diabetes Using Classification Mining Techniques accessed from International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1)