ARTIFICIAL INTELLIGENCE FOR PHARMACOVIGILANCE IN NIGERIAN SOCIAL MEDIA TEXT

Conference Paper · May 2021

Wuraola Fisayo Oyewusi
Data Science Nigeria

Vitus Onuigwe
Data Science Nigeria

ARTIFICIAL INTELLIGENCE FOR PHARMACOVIGILANCE IN NIGERIAN SOCIAL MEDIA TEXT

Wuraola Fisayo Oyewusi, Olubayo Adekanmbi, Ifeoma Okoh,Mary Idera Salami

Opeyemi Osakuade,Sharon Ibejih, Vitus Onuigwe

Data Science Nigeria
Lagos,Nigeria
{wuraola,olubayo,ifeoma,mary,osakuade,sharon,vitus}@datasciencenigeria.ai

ABSTRACT

Pharmacovigilance has been employed in several countries including Nigeria as a means of monitoring and detecting adverse drug reactions at an earlier stage, thus preventing harm from occurring in the larger population. Under-reporting of suspected adverse drug reactions (ADRs) to public health facilities is one of the challenges facing pharmacovigilance in Nigeria. The advent of social media has opened up a new wave of communication where people can openly discuss a range of topics and they can choose to use pseudonyms. Social media conversations are known to be casual, informal and open. This paper proposes a framework that leverages the named entity recognition (NER), one of the fundamental natural language processing methods for information retrieval and extraction, to identify adverse drug reactions and other medical entities in Nigerian social media conversations.

1 INTRODUCTION

Pharmacovigilance (PV) according to WHO, detects, assesses, and prevents Adverse Events (AEs) and other drug-related problems by collecting, evaluating, and acting upon AEs (Organization). In 2004, Nigeria was admitted into the WHO International Drug Monitoring Programme (Ogunleye et al., 2016), under the leadership of the National Pharmacovigilance Centre (NPC), an offspring of the National Agency for Food and Drug Administration and Control (NAFDAC) (Avong et al., 2018). There are more than 90% unreported AEs volume of individual case safety reports (ICSRs) (Mockute et al., 2019).

The issue of under-reporting of Adverse Drug Reactions, ADRs (Akarowhe, 2020) is one of the challenges facing pharmacovigilance in Nigeria. According to NAFDAC, only 16,500 ICSRs out of 80,000 ADR forms distributed nation-wide for 12 years (2004 to 2016), were submitted back to NAFDAC (Olowofela, 2018). This may be due to poor quality of ICSRs, lack of sufficient funds, lack of skilled manpower, lack of communication among others discussed in (Aguilar et al., 2019). Although there are pharmacovigilance systems in Nigeria teaching hospitals, they function sub-optimally (Udoye et al., 2018).

This paper aims to explore the use of artificial intelligence methods such as natural language processing (NLP) for medical information extraction from open social media data and build machine learning models as a tool for pharmacovigilance from the casual and informal client/patient perspective in Nigerian social media data. The trained medical entity recognition model is robust enough to accommodate both formal English description of medical entities and informal descriptions such as slang. The focus is on using Nigerian social media as a tool to extract pharmacovigilance related information, which clients may not report at points of care, but talk about openly on social media platforms.

1.1 PHARMACOVIGILANCE IN NIGERIA

Individual safety forms are usually filled by health professionals and submitted to the regulatory body (NAFDAC) when patients report adverse drug effects but there are challenges related to this. Prior work has addressed the challenges of the national pharmacovigilance centre in Nigeria, which ranges from lack of communication among patients and pharmaceutical agents, lack of sufficient funding to improve the assessment and detection of drug adverse effects among patients (Akarowhe, 2020). The occurrence of ADR reported to NAFDAC pharmacovigilance was also examined by analysing adverse drug reaction cases to determine the relationship between ADR with suspected drugs and detect signals from reported ADRs (Awodele et al., 2018)

Leveraging social media data for public health is not unique to Nigeria, the nature of health communication is changing globally as more people are relying on the internet for health information (Gallant et al., 2011). Some researchers also argue that web-based communication tool development that engages e-patients can better guide effective healthcare strategies and intervention and promote participatory medicine. However, in Nigeria, health communication is only evolving and it is not clear the extent to which it can be argued that hospitals are taking advantage of the internet and related platforms (e.g. social media) to influence health outcomes or impact health promotion, disease prevention or health literacy generally. It is instructive to note that just as social media are utilised to forge social relationships that sometimes lead to harmful health consequences, they may equally be used to promote social relationships that pay significant attention to health promotion and disease prevention (Batta & Iwokwagh, 2015).

There is an under-reporting of adverse drug reactions in Nigeria. Our research proposes the use of AI that leverages natural language processing to extract useful medical entities for pharmacovigilance from social media posts, which can be used to determine the safety of drugs, get information about medication use that clients may typically not share with health professionals. Drug manufacturers, regulatory agencies, researchers and stakeholders who typically track information related to Pharmacovigilance can build on our methodology.

1.2 MEDICAL NAMED ENTITY RECOGNITION (MNER)

Named Entity Recognition (NER) is the process of identifying and categorizing named entities in a given text. Examples of categories are organizations, locations, time, names, money, and rate. In clinical studies NER is used to extract entities of interest (e.g. disease names, medication names and lab tests) from clinical narratives, thus to support clinical and translational research (Wu et al., 2017). Medical named entity recognition has the same principle as general NER, the difference is the entity of interest.

2 METHOD

2.1 RESEARCH FRAMEWORK

Figure 1 shows a simple framework for extracting medical-related entities such as ADRs, symptoms, medication names from open social media health discussions, that can be leveraged for pharmacovigilance in Nigeria.

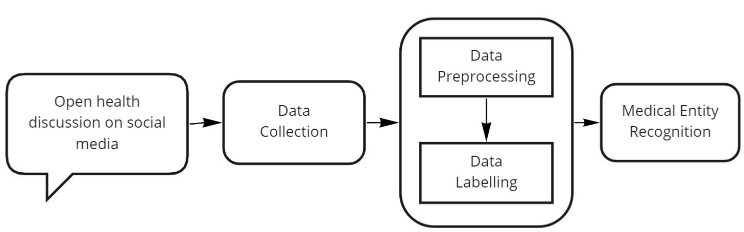

Figure 1: Framework of AI for pharmacovigilance in Nigerian social media text.

Data Collection: Data was scraped from a Nigeria public forum, where health matters are shared openly but under pseudonyms. These discussions are made by patients and health experts concerning sicknesses, drugs, doctor's prescriptions as well as some adverse drug reactions.

Data Preprocessing: The data was open and under pseudonyms, but to preserve privacy a text cleaning process to remove phone numbers, URLs and email addresses in the scraped reports used.

Data Labelling: TagEditor (v1.5), an annotation tool compatible with the Spacy NLP library for named entities was used for the annotation by an experienced health professional. Table 1 shows the entities labelled for and to give proper context, common expressions, medication brands and slangs for health terminologies were included while labelling.

Table 1: List of medical related entities labelled and their descriptions

Entity	Description
PERSON	Any Human referred to in the text. E.g Doctor, Nurse
SYMPTOM	The symptom of any disease e.g itching,
MEDICAL FIELD	Medical speciality e.g Radiology
DRUG	Medicinal products e.g drug name, medication brand name
FOOD	Edible and source of Nutrients
DOSAGE	Dosage of Medication
BODY PART	Part of the body
PLACE	Location, Town, City
MEDICAL PROCEDURE	Medical Procedures and processes
DISEASE	Illnesses
ORGANISM	Causative organism or disease vector
INJURY	Breakage in skin continuity
PHYSIOLOGIC PROCESS	Biological Processes
ADVERSE REACTION	Unintended consequences of medication or food

Medical Named Entity Recognition (MNER) Model with spaCy: As our study focuses on developing a tool for medical entity recognition in patients' reports, the annotated data was converted to the spaCy gold standard. We had 10610 training docs and 1419 development docs. The spaCy standard data recommends that each labelled entity must have at least 50 examples, which we met up to, except for the Dosage entity that had 26 examples. The model was trained using 30 iterations and had the best model saved based on the evaluation metrics.

2.2 Evaluation

Quantitative evaluation: In NER, evaluation assesses a model's ability in finding boundaries of names and their correct types. However, in some cases, the exact boundary detection is not so important, as long as the major part of the name has been identified (Jiang et al., 2016). Foreach prediction,the system counts the True Positive/(TP),False Positive/(FP) and False Negative/(FN). Precision(P) = TP / (TP + FP) ,Recall (R) = TP / (TP + FN) and F1 score is the harmonic mean of precision and recall.

Qualitative evaluation: The model was as well tested on new random samples of patients' reports, and health practitioners accessed its performance in identifying the medical entities present.

3 Result

Table 2 shows the evaluation metrics for individual entities and then the whole MNER model. The simplest explanation for the metrics is to measure the extent to trust the prediction of the MNER model used in this work.

Table 2: Precision, Recall and F1 Scores for Entity Types in MNER Model

ENTITY	PRECISION	RECALL	F1 SCORE
Whole model	71.711	72.458	72.082
Person	81.907	78.056	79.935
Medical field	5.882	12.500	8.00
Food	66.666	58.333	62.222
Medical procedure	80.392	60.294	68.907
Disease	65.100	78.861	71.323
Adverse reaction	87.500	87.500	87.500
Symptom	61.016	57.600	59.259
Drug	74.766	65.573	69.868
Body Part	72.368	83.333	77.464
Place	83.050	76.562	79.674
Organism	83.720	92.307	87.804
Dosage	88.888	57.142	69.565
Physiological process	8.571	18.750	11.764

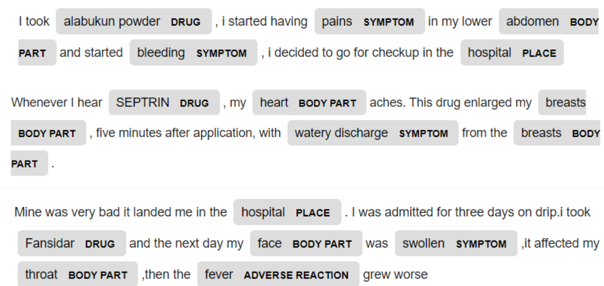

Figure 2: Examples of medical entities in social media texts recognized by the MNER model

To test the efficiency of the MNER model, data from some health conversations found on Nairaland - a social media platform in Nigeria (users) - was used. To demonstrate this, Table 3 shows vital details such as reaction/symptom, name of medicine, and current illness identified by the MNER model applied on the sample text. The table outline is similar to a NAFDAC ADR report form [1] Figure 2 displays entities recognized by the MNER model in the test data.

Table 3: Experimentation of the medical entity recognition on patients' social media reports.

REPORT	ILLNESS	REACTION/ SYMPTOM	DRUG	DOSAGE	BODY PART	PLACE
"I took **alabukun powder**, i started having **pains** in my lower **abdomen** and started **bleeding**, i decided to go for checkup in the **hospital**"	-	pains, bleeding	alabukun powder		abdomen	hospital
"Whenever I hear **SEPTRIN**, my **heart** aches. This drug enlarged my **breasts**, five minutes after application, with **watery discharge** from the **breasts**."	-	heart, watery discharge	SEPTRIN	-	-	-
"Mine was very bad it landed me in the **hospital**. I was admitted for three days on drip. i took **Fansidar** and the next day my **face** was **swollen**,it affected my **throat**,then the **fever**grew worse"	-	swollen, fever	Fansidar	face, throat	-	hospital

3.1 Discussion

The F1 score of the entire MNER model was 71.040%. Individual entities with the least F1 scores during this work are Medical Field with 8% and Physiologic Process with 11.764%, entities with the highest F1 scores are Adverse Reactions with 87.5% and Organism 87.804%. Entities with high F1 score had more representation in the training data, an increase in training data with better-labelled entity distribution can improve the model performance. Our approach of leveraging NER, an AI method for pharmacovigilance on social media text has its limitations which includes but not limited to; the proportion of the population who have access to the internet (as of 2019, only 42% of the total population have access [2]), duplication of report and technicalities of continuous deployment. It is important to be aware of these limitations and how to mitigate them.

4 Conclusion

This work explores the practical application of artificial intelligence as a tool to support pharmacovigilance in Nigeria. The work demonstrated the effectiveness of AI models to extract medical entities from informal texts that pertain to health issues and complaints in Nigeria. Our MNER

[1] https://primaryreporting.who-umc.org/Reporting/Reporter?OrganizationID=NG
[2] https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=NG

model has precision and recall scores of 70.583 and 71.503 respectively. We understand that pharmacovigilance is multidimensional, our approach leveraging AI can be included in the tools used by stakeholders to improve monitoring in Nigeria. Our approach is feasible and improves data source for pharmacovigilance. The performance of the experimental MNER model can be improved by increasing the training data and inclusion of useful medical entities such as duration of reaction/symptom and report source in future data labelling.

REFERENCES

Gustavo Aguilar, Suraj Maharjan, Adrian Pastor López-Monroy, and Thamar Solorio. A multi-task approach for named entity recognition in social media data. *arXiv preprint arXiv:1906.04135*, 2019.

Kingsley Akarowhe. A case study of pharmacovigilance in nigeria: Challenges and solutions. *American Journal of Biomedical Science & Research*, 9:36–40, 08 2020. doi: 10.34297/AJBSR.2020.10.001470.

Yohanna Kambai Avong, Bolajoko Jatau, Ritmwa Gurumnaan, Nanfwang Danat, James Okuma, Istifanus Usman, Dennis Mordi, Blessing Ukpabi, Gbenga Ayodele Kayode, Saswata Dutt, et al. Addressing the under-reporting of adverse drug reactions in public health programs controlling hiv/aids, tuberculosis and malaria: A prospective cohort study. *PloS one*, 13(8):e0200810, 2018.

Olufunsho Awodele, Rebecca Aliu, Ibrahim Ali, Yetunde Oni, and Christianah Mojisola Adeyeye. Patterns of adverse drug reaction signals in nafdac pharmacovigilance activities from january to june 2015: safety of drug use in nigeria. *Pharmacology research & perspectives*, 6(5):e00427, 2018.

Herbert E Batta and Nicholas S Iwokwagh. Optimising the digital age health-wise: utilisation of new/social media by nigerian teaching hospitals. *Procedia-Social and Behavioral Sciences*, 176: 175–185, 2015.

Linda M Gallant, Cynthia Irizarry, Gloria Boone, and Gary L Kreps. Promoting participatory medicine with social media: new media applications on hospital websites that enhance health education and e-patients' voices. *J Participat Med*, 3:e49, 2011.

Ridong Jiang, Rafael E Banchs, and Haizhou Li. Evaluating and combining name entity recognition systems. In *Proceedings of the Sixth Named Entity Workshop*, pp. 21–27, 2016.

Ruta Mockute, Sameen Desai, Sujan Perera, Bruno Assuncao, Karolina Danysz, Niki Tetarenko, Darpan Gaddam, Danielle Abatemarco, Mark Widdowson, Sheryl Beauchamp, et al. Artificial intelligence within pharmacovigilance: a means to identify cognitive services and the framework for their validation. *Pharmaceutical medicine*, 33(2):109–120, 2019.

Olayinka O Ogunleye, Ibrahim A Oreagba, Catherine Falade, Ambrose Isah, Okezie Enwere, Sunday Olayemi, Sunday O Ogundele, Reginald Obiako, Rachel Odesanya, Peter Bassi, et al. Medication errors among health professionals in nigeria: a national survey. *International Journal of Risk & Safety in Medicine*, 28(2):77–91, 2016.

Abimbola Olowofela. *Evaluation of pharmacovigilance system performance in South-South Nigeria*. PhD thesis, Université de Bordeaux, 2018.

World Health Organization. Essential medicines and health products information portal. `https://digicollections.net/medicinedocs/#p/home`, note = [Online; accessed February,2021].

Jude A Udoye, Raymond I Ozolua, and Jude Nwokike. Assessment of the knowledge, attitude and practice of pharmacovigilance by pharmacists in two states in southeastern nigeria. *Tropical Journal of Pharmaceutical Research*, 17(5):937–945, 2018.

Forum users. Adverse drug reactions discussion forum,nairaland. `https://www.nairaland.com/285625/post-adverse-drug-reaction-experience`, note = [Online; accessed February,2021].

Yonghui Wu, Min Jiang, Jun Xu, Degui Zhi, and Hua Xu. Clinical named entity recognition using deep learning models. In *AMIA Annual Symposium Proceedings*, volume 2017, pp. 1812. American Medical Informatics Association, 2017.