UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

COMPARING DIFFERENT CLASSIFICATION ALGORITHMS TO PREDICT THE ADHERENCE TO TUBERCULOSIS TREATMENT FOR NEW CASES IN KENYA

MUCHUNKU KAWIRA WENDY BRENDA

P52/11231/2018

SUPERVISOR: PROF. ROBERT OBOKO

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTERS IN COMPUTATIONAL INTELLIGENCE

July, 2020

DECLARATION

I declare that this research dissertation is my original work and has not been presented for a degree in any other university or for any other award.

Signature…………………………….. Date……………………….

Muchunku Kawira Wendy Brenda

P52/11231/2018

APPROVAL

The research dissertation has been submitted for examination with our approval as university supervisors.

Signature……………………………. Date…………………..……..

Prof. Robert Oboko

Director, School of Computing

University of Nairobi

DEDICATION

I want to thank God for getting me this far. I dedicate this work to my dad, Charles Muchunku. I would like to thank my family and friends for their moral support especially my mother Catherine Muchunku who always encourages me and my best friend Zainabu Marucha who always prays for me. May God bless you.

ACKNOWLEDGEMENT

I am grateful to the Almighty God for this opportunity to do this research study.

I would like to thank Professor Robert Oboko for his guidance through this research study.

I would also like to thank Kipng'etich Victor Lang'at for his guidance in the coding environment preparation.

I would like to thank Chuka County Referral Hospital, Magutuni Sub County Hospital, PCEA Chogoria Hospital, Mbaragu Level 4 Hospital, Marimanti Level 4 Hospital, Muthambi Health Centre and Kathwana Health Centre for making data collection possible.

LIST OF ABBREVIATIONS

TB- Tuberculosis

WHO- World Health Organization

SDG- Sustainable Development Goal

DOT- Directly Observed Treatment

DAT- Digital adherence technologies

HCP- Health Care Providers

MDR- Multi drug resistant tuberculosis

NTP- National Tuberculosis Control Programme

ABSTRACT

This study determines factors that are associated to non-adherence to tuberculosis treatment in Kenya. In the African Region, over 25% of the tuberculosis deaths occur. Kenya is among the 30 high burden countries accounting for more than 80% of tuberculosis cases in the world. In Kenya, TB is the number five killer. Due to the high cases of TB, WHO established a global plan called End TB Strategy that was aimed at reducing the tuberculosis deaths by 95%. Adherence to TB treatment is a key element to ensuring a successful control TB program, however, not every patient adheres to TB treatment. Non adherence to TB treatment results in the increase in number of deaths, drug resistance by patients, length of illness and disease transmission, which have economic consequences for patients and their families due to loss of income and cost of the health system. A system that tells if a patient will adhere to the tuberculosis treatment or not can help to curb the non-adherence rates.

Table of Contents

CHAPTER ONE: INTRODUCTION

1.1: Background

Tuberculosis is one of the causes of death in the world (World Health Organization, 2018). It is a communicable disease caused by bacteria called Mycobacterium tuberculosis, is a life-threatening infectious disease that attacks the lungs and also harms other body parts (Poling et al., 2011).

According to WHO(2018), the report shows that TB deaths that occur in the African Region are over 25%, with Kenya being among the 30 high burden countries accounting for more than 80% of tuberculosis cases worldwide. In Kenya, TB is the number five killer (Kenya National Bureau of statistics, 2016). Tuberculosis has been and still is one of the most serious challenges in the global public health, despite emergency measures in the last decade.

In 2009, the largest number of tuberculosis cases and deaths that have ever been reported in history in the world were recorded, with the incident cases at 9.4 million and deaths at 1.68 million deaths (World Health Organization, 2010). Millions of people continue to fall sick with TB globally each year, and the estimation was 10 million people in 2018 (World Health Organization, 2019).

The 3rd Sustainable Development Goal (SDG) emphasizes good health and well-being for everyone in all ages. However, the goal has several targets, the aimed target is the 3rd which is aimed at ending TB by stopping TB deaths and decrease in new cases by 2030 (WHO, 2018). To meet this target, WHO established a global plan called End TB Strategy aimed at reducing deaths associated with TB by 95%, incidence rates by 90%, and to ensure that the affected families sat free from the disaster costs (WHO, 2014).

Adherence to tuberculosis treatment will ensure a TB control program that is successful, and this is the foundation for most national and international strategies and also guidelines (WHO, 2011). Non-adherence to TB treatment is often as a result to the patient-related factors but it can also be as a result of delivery of services (Hopewell P., 2006), such as TB medicines being out of stock.

Supply of quality assured tuberculosis medicines that is sustained and without interruption is essential to attaining outcomes of a program that is successful.

Non adherence of tuberculosis treatment leads to rise in length of illness, severity and transmission of the disease, resistance to drugs and deaths which have economic consequences for patients and families of TB patients due to loss of income and also costs associated with the health system (WHO, 2000). Non-adherence to tuberculosis treatment is often as a result of the treatment interruption that may be caused by short irregular intake of medicine periods such as days or for longer periods such as weeks or even months that can lead to discontinuation of treatment completely.

To stop interruption of TB treatment measures are aimed at both health care providers and the TB patients. On the provider side, measures include supplying good quality medicines, ensuring that the practices of prescribing medicine are right, and ensuring that medicine is always in stock. On the patient side, TB treatment measures include encouraging the patients to carry on with treatment as directed till they finish the dose even when they feel that they have gotten better and the elimination of barriers, such as cost of transport to collect medicine. Though the savings of the economy are not well defined, these actions are thought of as a good investment.

Systems like Directly Observed Treatment (DOTS), a given name to the TB control strategy that is recommended by the WHO, have been put in place to enhance the rise of adherence to tuberculosis treatment. Curing tuberculosis is the most cost-effective method to end the increase in tuberculosis in communities that have high number of cases. DOTS is a proven TB treatment technique which is cost-effective (WHO, 1997).

With the technical and managerial components combined, the infectious cases are rapidly made non-infectious by DOTS and transmission pattern breaks. By the use of DOTS the growth of resistance to medicine strains of TB which are mostly fatal and close to 100 times more costly to heal are prevented. The "entirely supervised administration of medicines", idea developed first in the 1950s by Wallace Fox, is now known as DOT. Directly Observed Treatment was first adopted in the 1960s in TB drug trials in Madras and Hong Kong. It is now commonly recommended to control TB. One of a range of measures recommended by WHO to promote adherence to TB treatment is DOT (WHO, 1997).

DOT means much more than "supervised swallowing". Different projects in nations where the prevalence of TB is high show that doing away with socioeconomic barriers to DOT that patients face increases adherence and the rates of cure (Farmer P et al., 1991). In a country like the United States where the TB prevalence is not high, DOT programs are composite with several components that include social support, legal actions and food are highly worthwhile.

In Kenya, there has been intervention on ICT to enhance the adherence to tuberculosis treatment such as Keheala that utilizes disease management tools designed with positive community pressures to increase drug adherence for Tuberculosis patients. Keheala is centered on the patient and desired by the clinician. It is designed to ensure that adherence is maximized and there is motivation to undergoing the TB treatment (Keheala, 2018). Keheala's health platform that is digital increases the reach of the healthcare system by empowering TB patients directly. Keheala represents an advancement to better the adherence to treatment by patients. The platform strengthens the network of social support for the patients at risk, advancing the quality of life and also assists in curing the disease. With the Keheala digital platform, patients and clinicians are able to access tools needed to control the social causes of not adhering to treatment. Contrast to DOTS which is a desired protocol by WHO and the TB treatment status quo, Keheala represents a solution that is cost-effective and scalable, that saves money for the health providers while improving life’s quality and outcomes of health for the patients.

1.2: Problem Statement

Tuberculosis has a global prevalence with devastating morbidity and massive mortality (WHO, 2013). In nations with high burden like Kenya, the prevalence of HIV is high, thus this problem of tuberculosis has compounded, further leading to TB resurgence (Cegielski et al., 2014). In the Tuberculosis (TB) control, adhering to treatment leads to rates of cure being successful. It’s anticipatedd that if tuberculosis treatment success is to be attained the patient must take anti-tuberculosis drugs without fail continuously for six months. Not adhering to treatment leads to mortality rates increasing in high rates, TB treatment costs being high and cases of Multi drug resistant tuberculosis (MDR). The emergence of MDR TB strains accompanied by ineffective diagnosis is a vital challenge to TB management that leads to TB becoming among the vital public health issues in countries like Kenya that have poor resources (WHO, 2013; Cegielski et al., 2014). Systems and strategies such as DOTS have been set to ensure that TB treatment has been adhered to, yet treatment is still not adhered to by every patient. Late detection on who is not adhering to TB treatment might be a challenge when trying to ensure that the patient adheres yet they are used to not adhering to drugs. Prediction of new TB cases to detect early their probability of adhering to drugs can help in finding ways to deal with the specific cases that are not probable to adhering to the treatment hence leading to them adhering to the TB treatment. The prediction on patients under TB treatment to tell which patient is likely to adhere to treatment is still a gap that needs to be addressed.

1.3: Study Objectives

1.3.1: Broad objective

To find out factors associated to non-adherence of tuberculosis treatment among patients in Kenya.

1.3.2: Specific objectives

1. To determine the patient related factors influencing non-adherence to tuberculosis treatment

2. To determine the socioeconomic related factors influencing non-adherence to tuberculosis treatment

3. To determine the health care providers related factors influencing non-adherence to tuberculosis treatment

4. To determine the relationship between reliability of the health care service delivery and non-adherence to tuberculosis treatment

5. To build a model which compares different classification algorithms to determine the most accurate algorithm on the dataset created

6. To build a prediction system with the most accurate classification algorithm.

1.4: Research questions

1. What are the patient related factors influencing non-adherence to tuberculosis treatment?

2. What are the socioeconomic related factors influencing non-adherence to tuberculosis treatment?

3. What are health care providers related factors influencing non-adherence to tuberculosis treatment?

4. What is the relationship between health care service delivery and non-adherence to tuberculosis treatment?

5. What are the different classification algorithms used and how is the performance of each?

6. How is the performance of the prediction model built to promoting the adherence to the TB treatment?

1.5: Justification

Globally TB infects a third of the population in the world with new infections happening in 1% of the population each year. Reports indicate that TB is linked to 8.6 million morbidity and 1.3 million mortality with 80% of TB burden being borne by sub-Saharan countries and Asia (WHO, 2013). For example, Kenya is among the 30 high burden countries accounting for more than 80% of tuberculosis cases in the world (World Health Organization, 2019).

The resurgence of TB cases due to misuse of TB drugs or lack of adherence to proper dosages has led to the emergence Multidrug resistant TB. More patients follow the prescription given to take TB medicine if they are well enlightened and the treatment costs are reduced (Bagoes et al., 2009). They deduced that quality of treatment and negative image that is developed towards the healthcare providers results to non-adherence to TB. Munro et al. (2007) in his study recommended that the personal character of patients and healthcare providers, religion and abuse of drugs have an impact on TB treatment adherence. Female patients compared to the male patients adhered most in spite of traditional customs to ask for permission to undergo treatment from their partners.

TB patients' non-adherence has led to rates of new cases, MDR, treatment costs and mortality related to TB increasing. The motive of this study is to discover the factors that are related to TB patients defaulting treatment in Kenya. The findings drawn from the study and the recommendations will be useful to building a sample prediction model that will promote TB treatment adherence. They will also be useful to the ministry of health, TB control program and other TB prevention and control stake holders nationally and internationally.

CHAPTER TWO: LITERATURE REVIEW

2.1: Introduction

This chapter discusses about different studies that have been reviewed studies in relation to the treatment of tuberculosis to defaulting in various settings. Strategies that have been put in place to guarantee that TB treatment is adhered to are also discussed.

The word adherence, also known as concordance or compliance is well-defined by various writers and international organizations as the degree in which patients follow the instructions of medication given to them by their health care providers (Bell, Airaksinen, Lyles, Chen, & Aslani, 2007; Dick, Jaramillo, Maher, & Volmink, 2003; Elliott RA, 2009; Haynes, 1979; National Institute for Health and Clinical Excellence, 2008; Osterberg & Blaschke, 2005; Tilson, 2004).

Adherence to tuberculosis treatment has had key challenges that have been acknowledged by researchers. These challenges include patient-related factors health system-related factors and contextual factors (Ailinger et al., 2010; Armijos et al., 2008; Ayisi et al., 2011; Clark et al., 2007; Gough & Kaufman, 2011; Naidoo & Mwaba, 2010; Qian, Smith, Tuohong, Shenglan, & Garner, 2011).

Patient-centered care is intensely recommended to be the first management approach. This approach should include a plan on adherence that on stresses DOTS, where patients are directly observed in the course of taking the anti-tuberculosis medicine. Programs that utilize DOTS as an essential component in a patient-centered approach that is comprehensive will maximize the probability of completing treatment (Bock et al., 2002; Horsburgh, Feldman, & Ridzon, 2000; Peloquin, 2003; Zierski, 1976).

2.2: Adherence to tuberculosis treatment

The TB treatment overall objective is to treat the patient and to decrease the tuberculosis transmission to other individuals. Therefore, TB treatment that is successful benefits both the patient and the community where the patient lives (Dick et al., 2003; Peloquin, 2003; WHO, 2008). Adherence to treatment seems better in the first, critical stage of the disease when patients are more symptomatic. The patients that continue adhering to treatment are optimistic that they will get well since they see that the drugs are effective. Patients can also be motivated to follow their treatments if there is treatment success of family members who had tuberculosis before. Many patients report that treatment of TB healed their family members and friends who had TB (Naidoo, Dick, and Cooper, 2009). Treatment complexity and side-effects incidences are stated to be linked with stopping or interruption of the treatment by the patients (Hand & Bradley, 1996; Haynes, 1979; Liam, Lim, Wong, & Tang, 1999; Sockrider & Wolle, 1996).

Patients experiencing the side effects of medication and realize that there is no improvement in their health even after taking medicine are more probable to decide on taking other options of treatment, like the traditional doctor. It is worse where the health services do not exist in their communities, and traditional doctors are more reachable to them (Armijos et al., 2008; Coreil, Lazardo, & Heurtelou, 2004; Jaramillo, 1998; Liefooghe, Baliddawa, Kipruto, Vermeire, & De Munynck, 1997). TB care adherence in a Canadian Aboriginal population literature review found out that adherence to the treatment of TB being poor, was the utmost common reason for the early failure of treatment and illness relapse (Orr, 2011). Thus, such failure led to levels of disease transmission that are high and development of multi-drug resistance chances being high (Baylan, 2011; Qian et al., 2011; Tahir et al., 2006).

Adherence to tuberculosis is a common concern in the world (Orr, 2011). To facilitate possible interventions, decrease obstacles and enhance adherence, understanding is further needed in the personal and social causes, health service aspects of the behavior of adherence (Garner, Smith, Munro, & Volmink, 2007). 79% of people in the globe that have tuberculosis don't have access to DOTS, which is the internationally recognized approach by WHO to the control of TB (WHO, 2015). Up to 50 million is the number of people estimated to be infected with TB that is drug-resistant.

In making sure that treatments are effective, the most cost-effective approach is DOTS and it is recommended in the globe (WHO, 2003). The strategy of DOTS needs every patient to be classified according to their treatment history of TB and illnesses, so that the drugs to be administered can be appropriate to patients during diagnosis. Patients with tuberculosis in forms that are severe are assigned in one category, and the patients with treatments that are interrupted are assigned to a different category with those newly diagnosed (Ministry of Health Timor Leste, 2008; WHO, 2003).

In South Africa, a qualitative study examined TB patient's willingness to adhere to regimens of treatment and established that many factors were linked to adherence and non-adherence to the strategy of DOTS (Naidoo & Mwaba, 2010). When the disease, economic and social related factors like poverty are influenced, consumption of alcohol, smoking tobacco and having co-morbid health conditions were treatment non-adherence significant predictors (Naidoo et al., 2013).

Key improvement in the control of TB in the globe, followed the DOTS strategy widespread implementation in nations with TB burden that is high. Yet, statistics in the globe have shown that DOTS only is not adequate to attain the elimination and control of TB (Behera, 2009). For the strategy of DOTS to be implemented successfully, the components of the strategy and its approaches to implementation is vital to know (Ministry of Health Timor Leste, 2011). Therefore, addressing the problem to control TB needs good resources and sustained efforts, physical and human infrastructures and also strong commitments from every government, which is emphasized clearly in the Stop TB strategy (WHO, 2006).

2.3: Use of Systems to ensure adherence to tuberculosis treatment

2.3.1: Supervised Treatment

Supervised treatment is assisting patients to take their tuberculosis prescription regularly and to replete the medication. Also it is meant to assure that the healthcare professionals give care that is proper and are easy to notice medication interruption. One example of treatment supervision is documenting each dose of anti-TB medicine on the medication card of the patient.

Directly Observed Therapy (DOT), an advocated method of administration, should be identified as a section of a backing package which addresses the needs of the patient. This package should assist to establish that DOT is keen and complementary to the needs of the patient. A medical supporter monitoring single dose intake, establishes that the tuberculosis patient consumes the correct medicine for anti-TB, in the doses that are correct with intervals that are correct. Regular administration and support help to maintain constant communication amid the patient and a caregiver or medication observer; this gives more freedom for TB education, recognition and decision of hardship to medication, and early recognition of non-adherence, granting interventions to restore the patient to the recommended medication. The prompt exposure and management of negative reactions of drug and clinical deterioration of tuberculosis is granted by regular supervision.

WHO recommends DOT as a form of supervision (World Health Organization, 2007). Direct monitoring of each dose of medicine is most demanding in the intensive aspect, when occasional medication issued during either aspect, and in the medication of, for instance, psychologically handicapped victims, prison subjects, or patients acquiring second-line anti-TB medicine.

Treatment supervision should be done in a specific context and a manner that is friendly to the patient. In how DOT is applied, flexibility must be present, with transformation to contrast settings that are beneficial to the patient. The whole determination of medication monitoring would be conquered if it were to curb approach to care, turn victims away from medication, or add to their suffering.

Supervision can be undertaken at a health facility, in the workplace, in the community or at home depending on the local conditions. Each TB patient should have a treatment supporter, someone they are comfortable with. For patients who live close to a health facility, the ideal treatment supporter will be one of the health care providers, if convenient to the patient. Such health care providers are identified from other programs that are integrated with TB management, to help observe treatment.

For patients that reside distant from a healthcare center, the medication observer can be a community member who is trained or a health worker in the community. With effective and efficient monitoring and training, HIV/AIDS care givers in the community can act as the treatment observers .Fully recovered TB patients can be providers of DOT that are successful, same to friends, traditional doctors, members of the family, co-workers, religious leaders, neighbors, etc. (Pungrassami P, et al., 2002) stated that a treatment supporter can be any person who is willing, a patient accepts and can be answerable to the system of the health.

The National Tuberculosis Control Programme trains and monitors observers ot medication that are non-medical. All observers are directly accountable to the NTP staff. Patient's Confidentiality must be maintained. Once a patient accepts to have supervised treatment, the supporter keeps the drugs and gives to him or her only at the ingestion time.

Incentives for members of the community to become observers of tuberculosis treatment can be introduced, as well as incentives for patients and volunteers, bearing in mind the pros and cons of incentive programs (Beith A, Eichler R, & Weil D, 2007).

The WHO's training modules for health care providers contains detailed instructions for informing the patient and family about TB and its treatment as well as arranging supervised treatment (including identifying and preparing a community TB treatment supporter (World Health Organization, 2003).

2.3.2: Using a patient-centered approach to care and treatment delivery

Measures that are appropriate locally should be used to identify and address financial, cultural, physical and social problems to gain access to services on the treatment of tuberculosis (Munro SA, et al., 2007). The underprivileged and most weak groups should be given specific care. Clear efforts are also useful to handle issues on gender, improve attitudes of staff, and communication enhancement. These approaches are essential based on principles that are ethical concerning the privileges, responsibilities, wants and abilities of patients, their families and communities.

Measures to maintain the adherence to treatment consistency and completion in addition to DOT include:

- A regular drugs supply
- Tuberculosis care that is accessible, quality that is high, constant ambulatory (if it is healthcare center-based treatment)
- Action that is positive to eliminate obstacles to care and treatment
- Hospitalization availability

2.3.3: Digital technologies

The reaching of the new global End for TB targets (WHO, 2015) can be reached by the use of digital health technologies. There are three digital technologies that have been recommended by WHO to be used for guidelines for drug-susceptible TB and they include the use of short message service for example the use of SMS or a text messaging service, video-observed treatment also referred to as (VOT) and electronic medication monitor, that is (EMM) (WHO, 2017).

a) Short Message Service (SMS)

The is a standard and a default built-in function that is native to all the cell phones in the whole world, cheap when it comes to cost and very easy when it comes to how it is used. This makes it the most widely used way of communication with the outpatients. This can take the form of regular and automated messages to remind them to take their medications, supply the information about their health condition which is unidirectional communication, or providing some interaction to a patient about their care (bidirectional).

The Researchers Bekker LG, Nglazi MD, Hussey GD, Wood R, & Wiysonge CS (2013) indicated that there has been wide spread studies and evidences in reviews in areas that include TB treatment which is an impact of use of phone texting is such areas in the public health that include cessation in smoking, antiretroviral treatment and other chronic illnesses.

In the three randomized controlled trials also referred to as (RCTs) taken the variant geographical areas: Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients (Liu X, Lewis JJ, Zhang H, Lu W, Zhang S, & Zheng G et al., 2015), Impact of a daily SMS medication reminder system on tuberculosis treatment outcomes (Mohammed S, Glennerster R, & Khan AJ, 2016) and SMS reminders to improve the tuberculosis cure rate in developing countries (Bediang G, Stoll B, Elia N, Abena JL, Nolna D, & Chastonay P et al.,2014) failed to confirm that use of SMS to remind the patience of their medication could lead to an improved treatment adherence in TB considering the care standard provided. Here, the studies were made up of various Directly Observed Treatment, also referred to as (DOT), components within their constitute control groups which had a very high levels overall when it comes to adherence.

b) Electronic Medication Monitor (EMM)

EMM is a means that provide more flexibility to a patient doing a follow up to their medication; for example supporting a patient with dosing in their medication and following up of refill instructions and reminders; and to compiling on patient-specific dosing history to help in the differentiated care and counseling. EMM is part of the two main categories that include the sleeves fitting a blister package and the electronic medication boxes. EMM boxes are made up of electronic devices that are automated used to record and inform health-care provider on the regularity on how the medicine container has to be opened. For the Older devices, they recorded the usage on medication container, whereas the mobile services currently allow patient alerts and the reminders to the caregivers when the boxes used for medication have remained unopened for several days. A programmatic setting in China has provided for a study setting for the scalable and affordable medication boxes which are types of EMM for TB infected patients. EMM accuracy in dosing has been the focus of other studies where the correlation between actual ingestion of the medication by the various patients and the EMM devices (Chin J Anti tuberculosis, 2012), and the acceptability of EMM by both patients and providers (Int J Environ Res Public Health, 2017).

c) Video-Observed Treatment (VOT)

To reduce the inconvenience brought about by the need to travel during the visits as a result of healthcare can be reduced by use of communicating remotely using videos. This also help in reducing the chances of infecting other community individuals with the TB while it is at its infectious stage. This also help in saving a lot of the resources for the health providers and the patient (Clin Infect Dis., 2001). Compared to the earlier means that made good use of the landline telephone means for the videophone connection, internet based tablets and the smartphones which are readily and widespread available with equipped video communicating applications has heightened the real-time (synchronous) and the previously recorder (asynchronous) interactions (Int J Tuberc Lung Dis., 2015).

There is a need to do a proper study to provide an evaluation on comparison between VOT and the usual care standards of care which are the self-treatment administration methods used.

2.3.4: Sistema Integral de Administración al Tratamiento de la Tuberculosis (SIAT-TB)

The Sistema Integral de Administración al Tratamiento de la Tuberculosis (SIAT-TB) is a system that helps to achieve treatment management and focusing more on adherence to treatment, (The Union, 2017). This system makes good of a unique software that was developed by ASPAT, which has an ability to allow real-time monitoring of people who are affected by TB and are receiving their treatment in the various healthcare centers throughout Lima.

The testing phase for system started in January 2017 through various 30 centers. This period lead to a fall in 'loss of adherence' rate from an 11 percent to a lower 6 percent. Each patient going through the treatment for TB is source of the data saved in and register in an electronic database. The patient medication and the dosage are recorder in an electronic treatment system and a patient registers their clinic attendance made possible by a biometric scanner. The data that is collected through the process are recorded in a real-time basis and text messaging are provided by the system. Such alerts acts as a reminder to the patients on when to make the next attendance to the health center. It also reminds the staff when a patient fails to make an attendance to the clinic. Constant motivational messages are sent to the patient when they miss to make a scheduled attendance instead of using a kind of language that shifts blame to the patient. For example "You can still make it to the healthcare center to take your treatment. This system does not only reduce stigma among the patients, it also help the cases not to languish or disappear after a while for various weeks since all non-adherence by patience are picked almost immediately.

2.4: Prediction systems to promote adherence to tuberculosis treatment

2.4.1: Predicting and Explaining Default and Post-Default Sputum Smear and Drug Susceptibility Results

In urban Morocco, the treatment of defaulters amongst patients with tuberculosis had the aim to explore tuberculosis treatment defaulting risk factors and develop a tool for prediction, and to evaluate the defaulting consequences, precisely transmission risk or resistance of drug development (PLoS ONE,2014). 186 controls and 91 cases were registered to the system.

Risk factors that are independent to defaulting treatment included smoking cigarettes, work that , interfers with adherence, daily DOT, retreatment, quick resolution of symptom, drugs side effects and one's unknown duration of treatment. Above 50 years of age, a patient that does not smoke and friends knowing the diagnosis of a patient (their friend) were protective. A simple tool for scoring that incorporated these factors was sensitive by 82.4% and specific by 87.6% for this population default prediction.

Further contributors to defaulting were described by patients and clinicians and they proposed intervention targets that are relevant locally. Pulmonary TB cases were 89. Among the cases, the smear positive sputum for tuberculosis was 71%. Resistance to drug was rare. The TB treatment default causes through quantitative and qualitative data synthesis were explored from the patients and professionals of health. To predict default, a tool for scoring that has high specificity and sensitivity was developed.
They found out that the TB risk of transmission is probably high from patients defaulting the treatment to others.

2.5: Algorithms used in classification

2.5.1: Naive Bayes Classifier

It is Bayes classifier using robust (naive) independence suppositions. Bayes theorem displays in what way one conditional probability (for instance the probability of a hypothesis given monitored proof) relies on its inverse (in this instance, the likelihood of that proof given the hypothesis) in probability theory. Naive Bayes classifier presume existence of a certain attribute in a class is unconnected to the existence of any other attribute. Even if these attributes rely on each other or upon the presence of the other attributes, with all properties aiding to the probability independently. The theorem articulates the latter probability (for example; after proof E is monitored) of a hypothesis H in relation to the previous probabilities of H and E, and the probability of E given H. It suggests that before being viewed, proof was more improbable and it has a stronger asserting consequence.

2.5.2: k-Nearest Neighbors

The k-nearest neighbors algorithm (k-NN) is a technique for categorizing items based on nearby training examples in the feature area. It studies on how to label points by taking and using cluster points that are labeled points. For a new point to be tagged, it looks at points that are labeled nearest to the point that's new (neighbors that are adjacent), and the neighbors vote, so the label that most of the neighbors have is the new point label (neighbors number it checks is "k").

k-NN is a lazy learning or learning based on instance, where estimation of a function is done only locally and until classification, all computation is delayed.

2.5.3: Logistic Regression (Predictive Learning Model)

Logistic Regression is a statistical technique for examining a set of data in which there are one or more variables that are independent and they influence a result. The result is calculated with a variable that is dichotomous (wherein the probable outcomes are only two). The aim of this model is find the best appropriate model to illustrate the connection between a set of autonomous variables (descriptive or predictor) and the dichotomous characteristic of interest (dependent variable = outcome or response variable). This is better compared to other binary classification like nearest neighbor as it explains the components leading to classification quantitatively.

2.5.4: Decision Trees

Decision tree creates a tree structure shape for models in regression or classification models. A set of data is broken into lesser and lesser subsections as a linked decision tree is developed incrementally at the same time. A tree with leaf nodes and decision nodes is the ultimate outcome. A decision node with a leaf node and two or more branches symbolizes a decision or classification. In a tree, the root node is the uppermost decision node that matches the predictor that is best. Decision trees can manage data that is both numerical and categorical.

2.5.5: Random Forest

Leo-braiman developed an algorithm for generating a random forest. The name 'Random Forest' came from random decision forests, a name that was initially suggested in 1995 by Tin Kam Ho of Bell Labs. It is an ensemble classifier that comprises of many decision trees and by individual trees the class that is the mode of the output of the class is produced. It functions by constructing an assortment of decision trees at the time of training and the class of the individual trees that is the mode of the mean prediction or classes is produced.

2.5.6: Neural Network

A neural network entails of units (neurons) that are organized in layers that switch an input vector into some output. Each neuron takes an input and hands the output to the subsequent layer by applying a function to it. The networks are outlined to be feed-forward generally: a unit feeds its output is fed by the unit on the subsequent layer to all the units, but to the prior layer, there isn't any reply. From one unit to another, weightings to the passing signals are applied and the weightings are tuned to adapting a neural network to the specific challenge in the training stage.

2.5.7: Support Vector Machine (SVM)

Vladimir Vapnik created the initial SVM algorithm. A group of input data is taken by the standard SVM, and for each given input it forecasts, which of two probable classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier. For classification, regression or other duties, SVM makes a hyperplane or set of hyperplanes in a dimensional space that is infinite or high. Instinctively, a good partition is attained by the hyperplane with the longest distance to the adjacent data points for training of any class, because the lower the generalization inaccuracy the larger the margin of the classifier.

2.6: Models used to ensure adherence of tuberculosis treatment

2.6.1: Directly observed therapy (DOT) Model

The World Health Organization developed the DOT model as a key element of increasing the treatment of tuberculosis (WHO, 2012). The DOT model is applicable mainly in settings where

resources are limited, where it offers a strategy that is formalized with low cost so as to stress and monitor adherence in a health system that is constrained. In nations with limited resources, the tuberculosis DOT success is documented properly, and when applied appropriately, it has been shown that the spread and the development of resistance to tuberculosis treatment, is limited. Additionally, numerous studies have found out that tuberculosis DOT is more effective in cost than the therapy of self-administered (Int J Tuberc Lung Dis., 2003)
DOT was developed to ensure that TB patients take their medication. This system uses an 'observer' who is acceptable to both the patient and health care system to ensure that a patient takes every dose of their medication. The observer puts that into record and the health system will then monitor.

The recommended treatment time to achieve cure for people with TB according to the World Health Organization is at least 6 months. This long treatment duration can be overwhelming and difficult to complete especially when a person starts feeling well. However, when TB patients don't complete treatment it can lead to re-occurrence that can even cause death. Also, it also affects the public health care system since there will be increased transmission of this disease and some individuals may develop drug resistance.

Rethinking the Directly observed therapy (DOT) model

The idea of observing the patient as they take their TB medication emerged in the 1950s and 1060s in Hong Kong and India (Bayer R, Wilkinson D, 1995). It was discovered that direct observation had any benefits including close monitoring to ensure that the patient completed their treatment, a face to face interaction between patients and HCPs and also accurate documentation of the treatment record.

The World Health Assembly embraced the short course (DOTS) strategy 'directly observed therapy, in 1991. This strategy involved a multipronged intervention where DOT is among its components. It also included a six month short course therapy, usage of smear microscopy to diagnosis TB and reporting outcomes of treatment systematically. Although the treatment improved worldwide under the DOTS system, it was still unclear whether more improvements could be attributed to the direct therapy or other components of this system.

Reviews however suggest that DOT compared to Self-Administered Therapy (SAT) achieves less superior results in several outcomes like completion of treatment, microbiological cure and failure, re-occurrence of the disease and acquired drug resistance (Pasipanodya JG, Gumbo T, 2013).

Interviews on TB patients showed that DOT may result to a decreased autonomy, lack of adequate confidence as well as stigma. One study by Yellapa V, Lefevre P, Battaglioli T, et, al., conducted in patients from South India on how to cope with tuberculosis and directly observed treatment in settings using facility based DOT revealed that visits to the health facility frequently can lead to loss of income and employment.

Digital Adherence Technologies (DAT) model was developed to improve the shortcomings of DOT. This model boost the potential of DOT to offer patient-centric approaches to observe that pills are taken, reduce financial burdens on patients who visit the health facility regularly and easily identify patients who don’t adhere.

2.6.2: Digital adherence technologies (DAT) Model

Digital adherence technologies (DATs) was developed to cater for the DOT Model limitations. DAT include technologies for feature and smart phones, electronic pillboxes and sensors which are ingestible. They have the ability to enable more approaches that are centered to the patient for TB adherence medication monitoring than the DOT model that exists.

DATs can serve various functions in the care of tuberculosis, including compiling the history patient’s dosage, facilitation of observing a patient take pills digitally. Reminder to patients to take medications and triaging patients depending on their adherence level, which can facilitate giving care individually.

DATs should be incorporated to healthcare facilities policies to recognize and address causes related to structural, medical, psychosocial and health system of non-adherence medication. Else, DATs implementation can get into the risk of concentrating on ‘observation’ excessively and duplicating DOT models aspects of existing patterns.

Depending on the exact technology, patients can be reminded to take medications by DATs, facilitate observation of taking pills digitally, compiling the history patient's dosage and triaging patients depending on their adherence level, which can facilitate giving care individually by tuberculosis programs to patients with varied risk levels.

With access to cellular and mobile phones expansion, including in nations in Africa with tuberculosis burden that is high (Communication Lifeline, 2015), Latin America and Asia, digital adherence technologies (DATs) can facilitate approaches that are alternative to adherence improvement. These technologies vary from mobile phone texts, to electronic pillboxes, to sensors that are ingestible.

To accomplish a range of functions, DATs uses cellular communication and other innovations that include patients being reminded to take drugs, observing taken doses digitally and histories of dosing compilation that healthcare providers (HCPs) can use to identify and intervene on those not adhering.

DATs may be relevant for delivery of tuberculosis care rethinking for reasons that are few. First, benefits of public health for improving the adherence of tuberculosis medication, for example disease relapse rates that are reduced, acquired resistance of drugs and infection transmission (Int J Tuberc Lung Dis 2005). Second, currently several tuberculosis programs in the globe use DOT for monitoring unlike other sicknesses in which the therapy for self-administered is the standard's care, (PLoS One, 2008).

While DATs may be viewed as a challenge to their existing DOT models by some programs in TB, in various contexts, DATs can offer an alternative for 'observing' adherence in medication, possibly making them acceptable more by HCPs and TB patients than they are for other sicknesses.

Ngwatu BK, Nsengiyumva NP, & Oxlade O, *et al*. (2018) found out that strategies based on SMS do not improve the rates of treatment completion in settings with suboptimal results at baseline. The review found out that treatment completion rates were high when comparing video DOT monitoring of treatment and in-person DOT in a nation settings with high income. Additionally, the review also found two studies that suggested that electronic pillboxes may increase the cure probability and missed doses reduced. Though the review found that the use of DATs is limited for TB, the authors propose that evidence that is more robust is required so as to understand how patients and health systems may be impacted these technologies.

2.6.3: Deep Learning Model to Prescribe Interventions for Tuberculosis Patients Using Digital Adherence Data

Data usage from a phone-call-based TB DAT deployed in India for treatment, the foundation for studying was laid from data in the real world, including evading the unnoticed interventions effects method in the data for training for machine learning usage included. To target better and improve care for the patient, there was construction of a deep learning model and training was done in different setups of clinics. Presentation of a case study was done that aimed at achieving quality solution better by 15%. The presentation was demonstrated by training the model in an end-to-end setting that focused on decision learning. To leverage the proxy screening intervention, the model was built for the task of prediction.

A model that corresponded to the healthcare staff everyday task was developed that evaluated the patient's risk adherence, using the latest call history of patients' scheduling interventions of different types. Better predictions enable healthcare providers to interact with patients before missing doses.

16,975 patients that formed the population were involved and samples of training were made from each patient. All the14 days order was considered that were sequential of the data of call, with non-overlapping first 7 days for each order. The treatment's last and first 7 days were left out so as to prevent unfairness that could result from interaction with healthcare staff during the treatment start or finish. Two steps of filtering were done. First, patient samples that had more than two doses were detached and labelled manually at the time of order of input by the provider because the patients were probably in contact outside the 99DOTS system with the provider. Secondly, patient samples that didn't miss any dose in the sequence of input were detached. The majority of data was made up by these samples but comprised of almost no positive tags that ended up distorting the training. Additionally, predictions that were positive on patients who did not miss doses are not probable for use; no intervention on limited resource could be set out broadly that the targeted patients are those recent adherence that is perfect.

This procedure produced 16,015 samples where the positive samples were 2,437. Every sample had static features and a time-series data call. The series that was timed comprised of two sequences of length 7 for all samples. A binary sequence of call data was first (for a miss, 0 was used, and for dose that was manual or a call 1 was used). An accumulative sum of every dose not taken up to that day was the second sequence, bearing in mind that the history of the patient in full was in the program. Features of the demography from the table of the patient: gender of the patient, weight-group, ID of the center of treatment and age-group were the static features included. Other features from the Call Logs of the patient were engineered and other than just a patient's adherence, captured the behavior of the patient was captured. For example, does the patient call every day irregularly or each morning at the same time? Through the variance and mean minute and hour of the call, the above was captured.

Additional features included the calls number, manual doses numbers and variance/max/mean of calls made per day and also days per call. Analogous features that used unique calls only per day were also included. 29 features that were descriptive were gotten through this process.

First, the models that were standard were tested first, which used features that were static only: a support vector machine, linear regression and a random forest. The one that performed best is random forest. LEAP (Lstm rEal-time Adherence Predictor), a deep network was built to control the time series data that takes inputs, the static feature and the time series input, that were implemented with Keras (François Chollet et al., 2015). LEAP had 2 layers of input: a LSTM that has 64 units for the input of time series that are hidden and a dense layer for the input of static feature that has 100 units. The 2 layers outputs were concatenated to feed forward into another 16 unit's dense layer, followed by an activation sigmoid unit that is single. A 128 size of batch was used and trained for 20 epochs.

All data was made random and separated with the test set, set to 25% for model evaluation. To find the model with the best parameters, 4-fold grid search was used. To handle class unevenness, SMOTE was used on the set of training to over-sample and the executed with the imblearn, a library in Python. Using SKLearn, the features were standardized as percentages which was found to do well empirically. The baseline compared to the 99DOTS platform method that existed to assess risk.

To evaluate the usefulness of the methods over the baseline, each method was considered how it could be useful in planning interventions on house-visit. This being a resource that is very

limited, the baseline with the strictest threshold was set to consider patients for this intervention; i.e. 3 missed calls. The caught missed doses to be calculate, only the doses missed that happen before the patient goes to a risk that is high are counted. 21.6% more patients were caught and those that missed doses were 76.5% more, showing that substantively other than the baseline, targeting that is more precise is needed. However, on the ground without the healthcare staff delivering prediction based interventions these achievements could not become real.
The main factor in the usefulness of the model was interpretability. For the model to be trusted, it was necessary for the healthcare staff to appreciate why the model does predictions and incorporate the model's reasoning with the professional knowledge that they have.
Using the model, the healthcare providers would then target the interventions to patients with a risk that is sooner.

2.7: Related Literature

2.7.1: Related Literature on non-adherence to tuberculosis treatment related factors

If patients are well informed and treatment costs reduced, more patients will take their tuberculosis medication according to prescription (Bagoes et al., 2009). Non adherence to treatment is the outcome of bad image developed towards treatment, health professionals and medication quality. The personal character of patients and healthcare providers, religion and drug abuse influence the adherence to treatment with female patients adhering most compared to the male patients in spite of the traditional practice of asking for permission from their husbands to seek medication (Munro et al., 2007).

Besides habits of smoking and issues related to travel, non-adherence concerns number of members in a household, period of treatment, chewing of tobacco, symptoms relief, consumption of alcohol and inadequate drugs. The study also indicates that the rate of non-compliance was 16% due to factors such as where a patient resides, status of living, stigma, time taken to travel, waiting time to be served at a healthcare facility, level of literacy, and support from family, chewing khat, employment status and patients' TB knowledge (Sathiakumar et al., 2010).

The compliance of a patient is linked to patient gender usually male, experience of treatment especially feeling sick, or past non-compliance history, usage of drugs that are recreational, receiving information that is not satisfying and poverty presence (Culqui, 2012). "This Tuberculosis is known to have a strong association with poverty" (National strategic Plan and Tuberculosis, Leprosy and Lung Health, 2015 - 2018).

Muture et al indicated the common causes for defaulting treatment mentioned by patients that didn't finish the course of treatment were; treatment compliance ignorance combined with lack of adequate knowledge on tuberculosis and travelling to areas that are outside, running out of drugs and missing appointment to the clinic. Default factors that were predictive were lack of adequate TB knowledge, HIV co-infection, use of herbal medication, abuse of alcohol, income that is low, history of defaulting previously and the male gender.

2.7.2: Related Literature on classification algorithms used in building tuberculosis treatment adherence models

Sharareh R. Niakan Kalhori, Xiao-Jun Zeng (2013) predicted the results of the treatment of tuberculosis by evaluating and comparing different methods of machine learning. This plan needed a model that would predict the DOTS therapy outcome. This tool would then determine the intensity level of providing support and services. The techniques of machine learning were applied and compared initially, so as to predict the TB therapy outcome.

Models by different algorithms were developed and validated after feature analysis. The algorithms included: radial basis function (RBF), decision tree (DT), logistic regression (LR), radial basis function (RBF), artificial neural network (ANN), support vector machine (SVM) and Bayesian networks (BN). Training data set (N = 4515) and testing data set (N = 1935) were applied and prediction accuracy (recall and F-measure) evaluated the models. Seventeen features were identified ($P <= 0.004$; 95% CI = 0.001 - 0.007) which were correlated significantly. The best algorithm found was DT (C 4.5) with 74.21% prediction accuracy in comparing with BN, RBF, LR, ANN and SVM with 57.88%, 53.74%, 57.31%, 62.06% and 51.36% respectively.

They concluded that screening patients that are risky for fail in the completion of the treatment course, the decision tree model could be used in population by use of common data collection in the general practice routine. The healthcare professionals especially in rural regions would then be assisted to evaluate the MDR-TB risks among their patients noninvasively, faster, and economically.

R.Dinesh Jackson Samuel, B.Rajesh Kanna (2019) developed a TB detection system that comprised of two subsystems that comprised of data acquisition and recognition. A motorized microscopic phase was designed and developed in the data acquisition system to automate the acquirement of the view of all the fields. After of all the fields of view were acquired, data was then handed to the recognition system. The method of transfer learning in the recognition system was implemented by modifying the Inception V3 DeepNet model.

This model learnt from the weights of Inception V3 that were pre-trained and categorized the data from the knowledge transferred using SVM. In this model, the representations of the fixed feature from the uppermost stack layer of Inception V3 DeepNet were taken and using SVM, they were categorized. This model attained 95.05% accuracy thus the dependency on technicians that are skilled decreased in the process of screening and increasing specificity and sensitivity.

Salma Jamal, Mohd Khubaib, Rishabh Gangwar, Sonam Grover, Sayed E. Hasnain (2020) used Machine Learning and Artificial Intelligence algorithms to categorize single nucleotide variations (SNVs) as susceptible or resistant and predict the resistance of novel that confers mutations. Four machine learning algorithms, support vector machine (SVM), naive Bayes (NB), artificial neural network (ANN) and k nearest neighbor (kNN) were used for the task of prediction. Various mutations that can cause resistance of drugs in *M.tb* were identified. Several features based on structure and sequence were used to capture the mutations impact for each gene target.

In addition, a method of feature selection was used for identification of the features of having the greatest role that is significant in categorizing a mutation as resistant or susceptible. Simulations studies on molecular docking and molecular dynamics were performed for mutant and wild type, predicted to be resistance causing, complexes on protein-drug to examine the result of the mutations. To generate machine learning and artificial intelligence based models the study described an approach that was integrative computational using the several features of SNVs that are structural and sequenciial in *M.tb* genes for the resistance conferring mutations prediction.

All machine learning/artificial intelligence models for the genes had good overall, approximately 70% accuracy. The algorithm that performed the best is ANN with the models accuracy being highest in the 10-fold cross validation for most genes. In the testing data that was non-redundant, mutations were categorized as resistant or susceptible with an accuracy varying between 66.66% and 100%.

2.8: Conceptual Framework

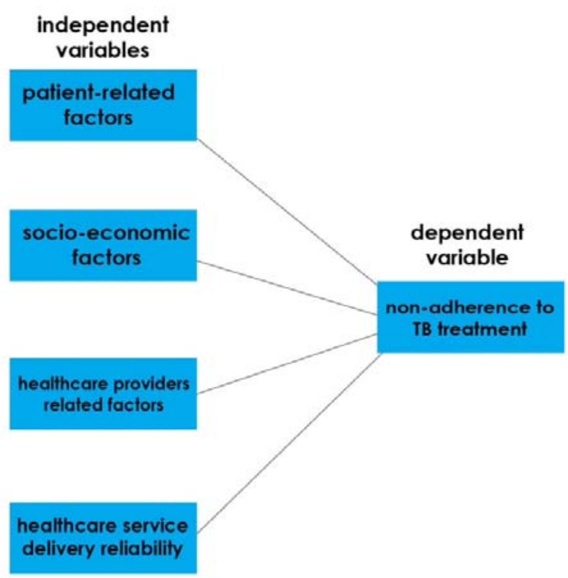

Figure 1

The literature review has revealed several causes linked to tuberculosis treatment non-adherence. The literature has been mapped to relations on the reliability of the health care service delivery and on the factors related to the patient, health providers and socio-economic.

2.8.1: Patient related factors

There are several factors related to the patient, which includes treatment, such as having other illnesses and the treatment duration, treatment progress over time, risk behaviors such as drinking alcohol and/or smoking cigarette.

Treatment will greatly influence adherence to TB treatment. A person who's been diagnosed with TB and has never adhered to treatment that exceeds 5days, would definitely not manage to adhere to the TB treatment 8 months duration treatment.

Risk behavior describes behaviors such as drinking alcohol, smoking cigarette and/or illegal use of drugs that impacts or causes harms that are significant to the health of individuals directly or indirectly. Usage of alcohol and drug problems are linked to non-adherence of treatment and outcomes of treatment that are poor. Usage of alcohol or drug increases the hospitalization periods. Missed appointments could be associated with alcohol and drug use.

2.8.2: Socio-economic factors

Socio-economic factors directly influence TB treatment adherence. One of the main contributor to adherence of treatment that is poor, failure of treatment and the diagnostic tuberculosis delay is the income of households being low. Many patients often experience various challenges financially to cater for the cost of transport or buying more drugs, and the challenges include demands that members face that are financially competing. As much as the drugs for tuberculosis are freely given, lacking transport money, unreachability of the TB services due to the geographical constraints, and the travel time needed to access a healthcare facility would cost much.

One of the supporting factors to adherence to treatment is social support. Social and family support affects behaviors related to adherence both directly and indirectly. Directly can include positive reinforcement and encouragement. Indirectly can include relief of stress, depression and anxiety reduction. Lack of social support reduces individuals' well-being and has effects that are adverse on a person's stressful events coping abilities, like suffering from a disease.

2.8.3: Health care providers related factors

Health care providers related factors like availability and accessibility influences the TB treatment adherence greatly.

A health care provider without the presence of sufficient health care workers would mean not all patients are able to be served. It is challenging for a patient to walk to a health centre to collect TB drugs, but there are no healthcare workers to provide the drugs. A health care provider without drugs would fail the patients as they cannot provide the drugs for the patients.

The accessibility of a health care provider determines if the patient is able to collect his medicine. A patient considers transport cost to access a health care provider, the distance from his home and the time they take to reach a health center to determine if they will access the health care provider or not.

2.8.4: Healthcare Service delivery reliability

The health care service delivery reliability include working hours of the healthcare workers, the patients and the healthcare providers' relationship.
There are several factors that will determine if the healthcare service delivery is reliable. Health systems or workers that do not offer services according to the standards or guidelines recommended, can't expect patients to adhere to treatment. These would include:

- Long waiting time to access medication
- Lack of dignity, respect and trust between healthcare workers and patients as they relate.
- Lack of skills and knowledge by healthcare facility staff in the drugs side-effects diagnosis and management.
- Care that is only attentive on tuberculosis but does not recognize and help patients to overcome obstacles, help patients to undergo counseling, homelessness, concerns, food and other factors linked to the patients.
- Care that seems available, but inaccessible in reality due to an operating tradition that doesn't attend to the needs of the patient such as availability of services that is limited and operation hours, barriers to transportation etc.
- Mechanisms or healthcare facilities that promote feelings of physical rejection or insecurity actively or passively.

The above factors are linked among the literature variables, for example health service centers not accessible will make patients travel to and from a health service that is available. As a result, financial support is required by patients so that such a need is met. This links the healthcare providers' related factors and the social economic factors.

The independent variables include: patient related factors, social and economic factors, health care providers related factors and the reliability of the health care service delivery. The dependent variable is non-adherence to TB treatment.

CHAPTER THREE: METHODOLOGY

3.1: Introduction

This chapter covers the methodology of the research. This includes the area of study, population, the design of the study, and size of the sample, data collection tool, comparing different classification algorithms and building the prediction system.

3.2: Study Area

The area of study was in seven health care facilities which were: Chuka County Referral Hospital, Magutuni Sub County Hospital, PCEA Chogoria Hospital, Baragu Level 4 Hospital, Marimanti Level 4 Hospital, Muthambi Health Centre and Kathwana Health Centre.

3.3: Study Population

The population that took place in the study was the tuberculosis patients registered in the seven health care centers within the six months before the start date of the research study. The doctors and the nurses in the tuberculosis department in the health care facilities were also interviewed.

3.4: Study design

In this study, the research design used was descriptive and the descriptive Method used was the Survey method. This method included answering questions through interviews. Interviews were done twice. In the first round interviews were semi structured and the interviewees were the traced treatment defaulters, the nurses and doctors in the TB department. In the second round, the interviews were structured and the interviewees were the enrolled TB patients.

3.5: Sample size

Tuberculosis patients within the six months before the start date of the research study were accessed through the Simple Random Sampling method.

The TB patients, the doctors and nurses in the TB department were interviewed through the Simple Random Sampling method.

3.6: Data collection

Interview data collection method was used to collect the data. The interviews were conducted to the Tuberculosis patients in seven health care facilities. A visit to the healthcare providers was first done to engage the heads of the hospitals to ask for permission to hold the interviews with the TB department. After being allowed to visit the TB department, there was an interaction with the nurses and doctors at the TB department hence knowing the time best time to hold the interviews.

The doctors and nurses in the TB department gave insights on the procedures the TB patients undertake as soon as they are diagnosed with the disease and how frequent they visit the hospital in the six months period. The doctors and nurses were interviewed through semi-structured interviews to find out what factors contribute to TB treatment not adhered to. They also helped in identifying the defaulters. Through the assistance of 5 community health volunteers, tuberculosis defaulters were traced and semi-structured interviews were conducted in order to know the issues that influence non adherence to the TB treatment.

Adding together this, helped in designing the questions used in the structured interviews. After understanding the factors that influence non adherence to TB treatment, structured interviews were conducted to the TB patients. The nurses and doctors in the TB department and the 5 community health volunteers helped to interview TB patients through structured interviews. These structured interviews were used to create the dataset used to building the predictive model.

3.7: Comparing different classification algorithms

After the second interview was held, the information received was entered in an excel sheet. This created the dataset that was used to build the model to compare different classification algorithms. The created dataset that was in a CSV file and it had 20 input variables and one output.

7 classification algorithms were used, which included: Random Forest, Support Vector Machine, Naives Bayes Classifier, k-Nearest Neighbors, Logistic Regression, Decision Trees, and Neural Network. 80:20 was the ratio of the train set to test set. A seed value was assigned to ensure that

the generated random numbers were the same. The most accurate algorithm was determined by comparing the 7 classification algorithms on the model built.

3.8: Building the Prediction Model

The most accurate classification algorithm was used in building a prediction model that tells if a new TB patient would adhere to the treatment or not. The dataset created was used as the training dataset to predict the output for the new input.

With the training dataset, a model was fit on the data. This was done by providing the training data to the most accurate algorithm on the dataset created. The model was then fit on the training dataset by calling the function fit ( ) and passing in the training dataset. The model was then evaluated by making predictions on the training dataset by calling the function predict ( ). The fit machine learning model takes inputs and makes predictions if a new TB case will adhere to TB treatment or not.

3.9: Data Analysis

All data captured through the structured interview was entered in an excel sheet and analyzed.

CHAPTER FOUR: RESULTS AND DISCUSSION

4.1: Introduction

This chapter discusses the outcomes of the study. This includes: the patient related factors, social-economic related factors, healthcare providers related factors and the healthcare service delivery reliability that influence the non-adherence to TB treatment.

4.2: Characteristics of the study Participants

The study conducted interview twice. In the first interview, 3doctors, 13 nurses and 19 TB defaulters were interviewed.

In the second interview, 354 TB patients were interviewed, where 50 were TB defaulters.

4.3: First Interview

The first interview was conducted to find the factors influencing non-adherence to tuberculosis treatment. This included; patient related factors, healthcare providers related factors, social-economic related factors and the healthcare service delivery reliability.

4.3.1: Patient related factors

The following are the factors related to the patient that influence non-adherence to the treatment of tuberculosis:

i Patient’s age
ii Patient’s gender
iii Educational levels of patients
iv Smoking habits
v Drinking habits
vi History of not following medication prescribed in the past due to other illnesses.

vii Experiencing the side effects of the medicine.
viii Preferring traditional medicine instead of medicine given at the healthcare facility
ix Commitment to work other than attending clinic days
x Medicine burden for a long time
xi Taste of medicine
xii Forgetting to take medicine
xiii Ignorance on being educated on how to take the medicine given at the healthcare facility

4.3.2: Socio-economic factors

The following are the socio-economic related factors influencing non-adherence to the treatment of tuberculosis:

i The patient's income level
ii The patient's employment status
iii Financial constraints to support transport costs for the patient to access health care centers
iv Stigma from family
v Stigma from friends

4.3.3: Healthcare providers related factors

The following are the healthcare providers' related factors influencing non-adherence to the treatment of tuberculosis:

i Shortage of staff at the healthcare facilities
ii Lack of medicine at the healthcare facility
iii Distance to the healthcare facilities
iv Healthcare facilities mechanisms which promote feelings of rejection

4.3.4: Healthcare Service delivery reliability

The healthcare service delivery reliability contributes to non-adherence to the treatment of tuberculosis in the following ways:

i Long wait time to see a doctor at the health care facility
ii Not being educated on how to take the medicine given at the healthcare facility
iii Patients experience less support from the healthcare workers
iv Healthcare workers operate during specific hours
v Mistrust and disrespect between patients and the healthcare workers

4.4: Second Interview

After the first interview, the factors influencing non-adherence to the treatment of tuberculosis were used to structure questions used for the second interview. The second interview was used to create a dataset that would be used in creating the prediction model.

4.4.1: Defaulters Results

The defaulters were 74% male and 26% female.

8% of the defaulters were below the age of 18, 36% were between the age of 18 and 35, 44% were between the age of 35 and 60 and 12% were above the age of 60.

The defaulters education levels were: tertiary (8%), secondary (24%), primary (58%) and 10% never went to school.

22% were unemployed, 64% were employed while 14% were students. 60% earned a monthly income of ksh15000 and below, 18% earned between ksh15001 and ksh30000, 16% earned between ksh30001 and ksh50000, 6% earned above ksh50000.

18% of the defaulters lived alone, 58% lived with their families and 24% lived with friends. 66% have ever had a member of their family diagnosed with TB before while 34% have never had a family member diagnosed with TB. 42% had their families members with TB that experienced stigmatization, 24% of those with family members that had TB before did not face stigmatization, 34% never had a TB patient in their family before.

26% with friends that have had TB faced stigmatization from friends while 36% did not face stigmatization. 38% had never had friends that have ever been diagnosed with TB.

10% of the defaulters smoked every day, 16% smoked occasionally and 74% did not smoke at all. 14% took alcohol daily, 28% took alcohol occasionally and 58% did not take alcohol.

56% had been on medication for more than 3 months, and they completed their medicine. 30% had been on medication for more than 3 months, but they did not complete their medication.14% had never been on medication for more than 3 months. 78% forgot to take their medicine sometimes while 22% never forgot to take their medicine.

18% preferred to use traditional medicine other than medicine given at healthcare facilities. 52% preferred to take medicine given at the healthcare facilities other than the traditional medicine. 30% were okay with either of the medicine.

36% accessed the healthcare facilities by walking, 40% used either a bicycle, motorbike or car, while 24% were sometimes not able afford the transport cost to access the healthcare facility.92% were able to find time to access the healthcare facility easily to collect medicine, while 8% would work during the day and could not find time to visit the healthcare facility. 36% would have to wait for more than 3hours so as to see a health worker while 64% did not wait for more than 3 hours.
94% would always get medicine prescribed to them at the healthcare facility, while 6% could not get the medicine prescribed at the healthcare facility sometimes. 90% were educated on how to take their medicine and advised on the diet to follow, while 10% were either not educated on how to take the medicine or not advised on the diet to follow or both.
84% were contented with the offered services to them at the healthcare facility while 16% were not contented with the offered services to them at the healthcare facility.

4.4.2: TB treatment patients Results

The TB patients interviewed were 68.36% male and 31.64% were female.

11.86% of the TB patients were below the age of 18, 32.2% were between the age of 18 and 35, 46.05% were between the age of 35 and 60 and 11.86% were above the age of 60.

Their education levels were: tertiary (15.54%), secondary (29.38%), primary (45.2%) and 9.89% never went to school.

31.64% were unemployed, 59.32% were employed while 9.04% were students. 64.97% earned a monthly income of ksh15000 and below, 19.21% earned between ksh15001 and ksh30000, 11.58% earned between ksh30001 and ksh50000, 4.24% earned above ksh50000.

14.41% lived alone, 65.25% lived with their families and 20.34% lived with friends. 57.91% have ever had a member of their family diagnosed with TB before while 42.09% have never had a family member diagnosed with TB. 20.62% had their families members with TB that experienced stigmatization, 37.57% of those with family members that had TB before did not face stigmatization, 41.81% never had a TB patient in their family before.

18.93% with friends that have had TB faced stigmatization from friends while 48.59% did not face stigmatization. 32.49% had never had friends that have ever been diagnosed with TB.

8.77% of the TB patients smoked every day, 24.01% smoked occasionally and 67.23% did not smoke at all. 9.89% took alcohol daily, 25.14% took alcohol occasionally and 64.97% did not take alcohol.

57.34% had been on medication for more than 3 months, and they completed their medicine. 24.86% had been on medication for more than 3 months, but they did not complete their medication.17.80% had never been on medication for more than 3 months. 43.5% forgot to take their medicine sometimes while 56.5% never forgot to take their medicine.

13.28% preferred to use traditional medicine other than medicine given at healthcare facilities. 59.32% preferred to take medicine given at the healthcare facilities other than the traditional medicine. 27.40% were okay with either of the medicine.

44.92% accessed the healthcare facilities by walking, 39.83% used either a bicycle, motorbike or car, while 15.25% were sometimes not able afford the transport cost to access the healthcare facility. 93.22% were able to find time to access the healthcare facility easily to collect medicine, while 6.78% would work during the day and could not find time to visit the healthcare facility. 50.28% would have to wait for more than 3hours so as to see a health worker while 49.72% did not wait for more than 3 hours.
95.48% would always get medicine prescribed to them at the healthcare facility, while 4.52% could not get the medicine prescribed at the healthcare facility sometimes. 92.94% were

educated on how to take their medicine and advised on the diet to follow, while 7.06% were either not educated on how to take the medicine or not advised on the diet to follow or both. 90.4% were contented with the offered services offered to them at the healthcare facility while 9.6% were not contented with the offered services to them at the healthcare facility.
85.88% of the TB patients adhered to the TB treatment, while 14.12% did not adhere to the TB treatment.

4.5: Comparing different classification algorithms

Support Vector Machine was the most accurate algorithm with 85.8753% accuracy. Naives Bayes Classifier had 80.5030% accuracy, k-Nearest Neighbors had 85.3119% accuracy, Decision Trees had 80.2374% accuracy, Random Forest had 85.0302%, Logistic regression had 84.7445% accuracy and Neural Networks had 85.3078% accuracy.

Support Vector Machine being the most accurate algorithm, was used to create the prediction model to determine if a new TB case will adhere to treatment or not.

4.6: Prediction System

After comparing different classification algorithms, Support Vector Machine being the most accurate algorithm, it was used in building a prediction system to determine if a new TB case will adhere to treatment or not.

The system asks the user questions, and the user inputs answers. On completion of answering all the questions, the system gives an output that tells if the patient will adhere to the TB treatment or not

Two TB patients tested the system by inputting answers when asked questions by the system and received an output from the system on whether or not they would adhere to the treatment.

Since Support Vector Machine has an accuracy of 85.8753% the system may not provide 100% accurate output for a given input.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

5.1: Introduction

Chapter five covers the conclusions and recommendations of the study and system.

5.2: Conclusion

Non-adherence to tuberculosis treatment is as a result of patient related factors, socio-economic related factors, healthcare providers related factors and the reliability of the health care service delivery.

Factors related to the patient that influence non-adherence to the treatment of tuberculosis include; the patient's age, patient's gender, educational levels of patients, smoking habits, drinking habits, history of not following medication prescribed in the past due to other illnesses, experiencing the side effects of the medicine, preferring traditional medicine instead of medicine given at the healthcare facility, commitment to work other than attending clinic days, medicine burden for a long time, taste of medicine, forgetting to take medicine and ignorance on being educated on how to take the medicine given at the healthcare facility

Factors related to the socio-economy that influence non-adherence to the treatment of tuberculosis include; the patient's income level,, the patient's employment status and financial constraints to support transport costs for the patient to access health care centers, stigma from family and stigma from friends

Factors related to the healthcare providers' that influence non-adherence to the treatment of tuberculosis include; shortage of staff at the healthcare facilities, lack of medicine at the healthcare facility, distance to the healthcare facilities and healthcare facilities mechanisms which promote feelings of rejection

The healthcare service delivery reliability contributes to TB treatment non-adherence in the following ways; long wait time to see a doctor at the health care facility, not being educated on how to take the medicine given at the healthcare facility, patients experience less support from the healthcare workers, healthcare workers operate during specific hours and mistrust and disrespect between patients and the healthcare workers

Non adherence leads to losing lives, new infections, incurred budget expenses and deters the economic progress of a nation

The built system which predicts if a new case of TB will adhere to treatment or not, will help in taking measures on the patients predicted not to adhere to treatment to make sure that they will adhere to the tuberculosis treatment from the start of their medication to the end.

5.3: Recommendations

Education that addresses smoking and drinking habits during treatment should be done to the patients. Patients should complete their treatment even when they feel well before the dose completion.

Family and friends of patients diagnosed with TB should encourage and support the patients instead of stigmatizing them. Mobile clinics or doorstep delivery of medicine should be conducted.

Ensure constant of supply of drugs to healthcare facilities. Health care workers should support and encourage the patients.

Health care workers should educate patients on how to take medicine.

More data should be collected to ensure that the accuracy of the predictions made is 100%

SCHEDULE

	Duration	Start	Finish	November 2019	December 2019	January 2020	February 2020	March 2020	April 2020	May 2020	June 2020	July 2020
Problem identification	8 days	01/11	08/11									
Introduction	8 days	09/11	16/11									
Literature review	46 days	12/11	27/12									
Methodology	10 days	28/12	06/01									
Documentation	10 days	07/01	16/01									
Milestone 1 presentation	1 day	17/01	17/01									
Data collection	52 days	18/01	10/03									
Data Analysis	14 days	11/03	25/03									
Building classification alg model	15 days	26/03	10/04									
Building the system	22 days	10/04	02/05									
Testing	7 days	03/05	10/05									
Conclusion and recommendations	3 days	11/05	14/05									
Documentation	19 days	15/05	03/06									
Milestone 2 presentation	1 day	03/06	03/06									
Final Documentation	43 days	04/06	17/07									
Milestone 3 presentation	1 day	17/07	17/07									

REFERENCES

1 WHO. (2010). Treatment of tuberculosis: guidelines (4th ed).

2 WHO. (2011). Guidelines for the programmatic management of drug-resistant tuberculosis – 2011 update.

3 Hopewell P. (2006). Tuberculosis Coalition for Technical Assistance. International Standards for Tuberculosis Care (ISTC).

4 WHO. (2000). The Economic Impacts of Tuberculosis. Stop TB Initiative.

5 Farmer, P., Robin, S., Ramilus, S., & Kim, J. (1991). Tuberculosis, poverty, and "compliance": lessons from rural Haiti.Semin Respir Infect, 254-260

6 WHO. (2013a). Timor Leste. Retrieved 30 November, 2015, from http://www.who.int/countries/tls/en/

7 WHO. (2013b). Tuberculosis Financing and Funding Gaps

8 Bagoes Widjanarko, Michelle Gompelman, Maartje Dijkers, et al. (2009). Factors that influence treatment adherence of tuberculosis patients living in Java, Indonesia Dovepress Journal 3

9 Munro SA, Lewin SA, Smith HJ, Engel ME, Fretheim A, Volmink J. (2007). Patient Adherence to Tuberculosis Treatment: A Systematic Review of Qualitative Research. PLoS Med

10 Bell, J., Airaksinen, M. S., Lyles, A., Chen, T. F., & Aslani, P. (2007). Concordance is not synonymous with compliance or adherence.

11 Peloquin, C. (2003). Treatment of Tuberculosis. MMWR: Morbidity & Mortality Weekly Report

12 Naidoo, Dick, and Cooper. (2009). Exploring Tuberculosis Patients' Adherence to Treatment Regimens and Prevention Programs at a Public Health Site

13 Liefooghe, R., Baliddawa, J., Kipruto, E., Vermeire, C., & De Munynck, A. (1997). From their own perspective: a Kenyan community's perception of tuberculosis. Tropical Medicine & International Health.

14 Orr, P. (2011). Adherence to tuberculosis care in Canadian Aboriginal populations, Part 1: definition, measurement, responsibility, barriers. International Journal Of Circumpolar Health,

15 Baylan, O. (2011). Extensively drug resistant and extremely drug resistant tuberculosis forms after multi-drug resistant tuberculosis: new faces of the old disease

16 Garner, P., Smith, H., Munro, S., & Volmink, J. (2007). Promoting adherence to tuberculosis treatment. Bulletin of the World Health Organization

17 World Health Organization. (2015). Implementing the End TB strategy: the essentials. Geneva. Retrieved from http://www.who.int/tb/publications/2015/The_Essentials_to_End_TB/en

18 Ministry of Health Timor Leste. (2008). Manual -National Tuberculosis Program Control Dili: Ministry of Health Timor Leste.

19 Naidoo, P., & Mwaba, K. (2010). Helplessness, Depression, and Social Support Among People Being Treated For Tuberculosis in South Africa Social Behavior & Personality

20 Naidoo, P., Peltzer, K., Louw, J., Matseke, G., McHunu, G., & Tutshana, B. (2013). Predictors of tuberculosis (TB) and antiretroviral (ARV) medication non-adherence in public primary care patients in South Africa

21 Behera, D. (2009). Implementing the WHO stop TB strategy: A handbook for national tuberculosis control programmes. Indian Journal of Medical Research, 130(1), 95- 96

22 Ministry of Health Timor Leste. (2011). Knowledge, Attitude and Practice (KAP) towards Tuberculosis and Socio-economic Impact of the Disease in Timor-Leste

23 World Health Organization. (2006).The Stop TB Strategy: building on and enhancing DOTS to meet the TB-related Millennium Development Goals. Geneva.

24 Kipp, A. M., Pungrassami, P., Nilmanat, K., Sengupta, S., Poole, C., Strauss, R. P., Van Rie, A. (2011a). Socio-demographic and AIDS-related factors associated with tuberculosis stigma in southern Thailand: a quantitative, cross-sectional study of stigma among patients with TB and healthy community members.

25 Beith A, Eichler R, & Weil D. (2007). Performance-Based Incentives for Health: A Way to Improve Tuberculosis Detection and Treatment Completion

26 Munro, S.A, et al. (2007). Patient Adherence to Tuberculosis Treatment: A Systematic Review of Qualitative Research. PLoS Med

27 LG, Wood R, Hussey GD & Wiysonge CS. (2013). Mobile phone text messaging for promoting adherence to anti-tuberculosis treatment: a systematic review.

28 Liu X, Lewis JJ, Zhang H, et al. (2015). Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. PLoS Med.

29 Mohammed S, Glennerster R & Khan AJ. (2016). Impact of a daily SMS medication reminder system on tuberculosis treatment outcomes: a randomized controlled trial.

30 Bediang G, Stoll B, Elia N, Abena JL, Nolna D, Chastonay P, et al. (2014). SMS reminders to improve the tuberculosis cure rate in developing countries (TB-SMS Cameroon): A protocol of a randomised control study.

31 J Su, G Liang, L Bao, et al. (2012). Analysis of culture and drug-resistance of Mycobacterium tuberculosis in Kunming [J]. Chin J Anti tuberculosis

32 Int J Tuberc Lung Dis. (2015). High Incidence of Latent Tuberculous Infection Among South African Health Workers: An Urgent Call for Action

33 The Union. (2017). The Union Annual Report

34 Bayer R, Wilkinson D. (1995). Directly Observed Therapy for Tuberculosis: History of an Idea

35 Yellappa V, Lefèvre P, Battaglioli T, et al. (2016). Coping With Tuberculosis and Directly Observed Treatment: A Qualitative Study Among Patients From South India

36 Thomas A, Gopi PG Santha T, et al. (2005). Predictors of relapse among pulmonary tuberculosis patients treated in a DOTS programme in South India. Int J Tuberc Lung Dis.

37 Ngwatu BK, Nsengiyumva NP, & Oxlade O, et al. (2018). The Impact of Digital Health Technologies on Tuberculosis Treatment: A Systematic Review

38 Bagoes et al. (2009). Factors that influence treatment adherence of tuberculosis patients living in Java, Indonesia

39 Munro SA, Lewin SA, Smith HJ , et al. (2007). Patient adherence to tuberculosis treatment: a systematic review of qualitative research. PLoS Med

40 National strategic Plan and Tuberculosis. (2015) . Leprosy and Lung Health.

41 Sharareh R. Niakan Kalhori, Xiao-Jun Zeng, R.Dinesh Jackson Samuel & B.Rajesh Kanna. (2019). Tuberculosis (TB) detection system using deep neural networks

42 Salma Jamal, Mohd Khubaib, Rishabh Gangwar, Sonam Grover, Sayed E. Hasnain. (2020). Artificial Intelligence and Machine learning based prediction of resistant and susceptible mutations in Mycobacterium tuberculosis