**DEVELOPMENT OF A CLASSIFICATION MODEL FOR SIGNIFICANT BACTERIURIA IN CHILDREN WITH SICKLE CELL ANAEMIA**

**BY**

**OLORUNDARE RUTH OMOLADE**

**MATRIC NUMBER: CSC/2016/090**

**BEING SUBMITTED TO**

**THE DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING,**

**FACULTY OF TECHNOLOGY,**

**OBAFEMI AWOLOWO UNIVERSITY, ILE-IFE, OSUN STATE**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF BACHELOR OF SCIENCE (HONOURS) DEGRREE IN COMPUTER SCIENCE WITH MATHEMATICS**

**MAY, 2023.**

# CERTIFICATION

This is to certify that this thesis “Development of a Classification Model for Significant Bacteriuria in Children with Sickle Cell Anaemia” was done by OLORUNDARE Ruth Omolade of the Department of Computer Science and Engineering, Faculty of Technology, Obafemi Awolowo University, Ile-Ife, in partial fulfilment of the award of Bachelor of Science (Honours) Degree in Computer Science with Mathematics.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. P. A. Idowu Date

Supervisor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. A.O. Oluwatope Date

Head of Department

# DEDICATION

This project work is dedicated to the Almighty God for his sustenance and guidance throughout the entire duration of my course of study.

# ACKNOWLEDGEMENT

My sincere appreciation goes to God Almighty for his grace and protection, and for the successful completion of this programme.

My heartfelt gratitude goes to my parents, Mr and Mrs Olorundare, for their unflinching love, support and encouragement. God will continue to strengthen you and supply all your needs. I am also grateful to my siblings: Eniola and Esther. May you continue to excel and progress.

My profound gratitude goes to my project supervisor, Prof. P. A. Idowu, for his help, advice and guidance to make this project a success. May God bless you and your family abundantly.

I extend my gratitude to [to be known later] whose help and response made this study a reality.

I also wish to appreciate [friend].

I also acknowledge all the lecturers that have taught me up to this point. May God enrich and uphold you.

Finally, I am grateful to everyone who has helped me in one way or the other. I appreciate you all.

# ABSTRACT

This study identifies the risk factors of significant bacteriuria in children with sickle cell anaemia, which in turn are the variables used in developing a classification model for predicting significant bacteriuria. This was achieved by reviewing existing literature and conducting interviews with a paediatrician. Data was also collected from [to be known].

After the identification of important variables and preprocessing of data, a classification model was formulated using Python ensemble learning model and simulated in Jupyter Notebook.

The research concluded that there is need for further investigation into the concept employed in this study, optimistic of a positive and more accurate output in future work.

This study aims to develop a classification model for significant bacteriuria in children with sickle cell anaemia. Significant bacteriuria refers to the presence of a significant number of bacteria in urine, which indicates a urinary tract infection. This study will identify the risk factors of significant bacteriuria in children with sickle cell anaemia, by reviewing existing literature and conducting an interview with a paediatrician. These risk factors will be used as features for developing a classification model for predicting the occurrence of significant bacteriuria.

The proposed model will use supervised machine-learning algorithms to analyze patient data and predict the presence or absence of significant bacteriuria. The model will be developed using a dataset of sickle cell patients and their corresponding bacterial culture results, which will be obtained from hospital records, after taking the necessary steps to obtain approval for accessing these records.The dataset will be pre-processed, and features will be selected, based on their relevance to the classification task. Various classification algorithms, including logistic regression, decision tree, k-nearest neighbours, and ensemble learning, will be applied and evaluated using various performance metrics.

The proposed classification model has the potential to aid clinicians in the diagnosis and treatment of significant bacteriuria in children with sickle cell anaemia, leading to reduced complications and improved patient outcomes.

# Table of Contents

Title Page … i

# List of Figures

Yupppp

# List of Tables

Figures

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background to the Study

Bacteriuria is the occurrence of bacteria in urine (Zulfiqar *et al*., 2018). Bacteriuria may result from the presence of bacteria in the bladder urine or from contamination of urine during or after micturition. Urine culture is the criterion standard for the diagnosis of bacteriuria (Rahn *et al*., 2005).

The term ‘significant bacteriuria’ was developed by Edward Kass in mid-1950s to provide a means of differentiating between contamination in the voided specimen and true urinary infection. It was based on the reasonable assumption that the common pathogens of the urinary tract multiply in the urine and, therefore, when bacteria are deposited in the urine, they tend to multiply to a very large number, usually exceeding 106 colonies per millilitre. There is no specific bacterial number for use in the detection of bacteriuria, but, rather, a degree of probability that a given colony count signifies either bacteriuria or contamination in a voided specimen (Hinman, 1963). Thus, it is widely accepted that significant bacteriuria, in either case of symptomatic or asymptomatic, is defined as the presence of more than 105 colony-forming units (CFU)/mL of urine (or 108/L) in a urine sample using a standard calibrated bacteriological loop (Kehinde *et al*., 2011).

The Infectious Diseases Society of America defines asymptomatic bacteriuria (ASB) as the isolation of a specified quantitative count of bacteria in an appropriately collected urine specimen obtained from a person without symptoms or signs referable to urinary infection (Nicolle *et al*., 2005).

On the other hand, bacteriuria accompanied by symptoms attributable to the urinary tract and no alternate source is symptomatic and otherwise referred to as a urinary tract infection (UTI) (Colgan *et al*., 2010).

Susceptibility to bacterial infections is one of the symptoms of young sickle cell patients. In fact, one of the clearest precipitants of pain episodes and mortality among sickle cell patients is bacterial infection (Kaur *et al*., 2013).

Sickle cell anaemia (SCA) is a hereditary red blood cell disorder which is transferable from parent carriers [(AS father) and (AS mother)] to their offspring(s). The red blood cell of a sickle cell patient does not possess the physiological properties of an ideal red blood cell, leading to a cascade of problems that may give rise to several crises and complications. (Nwabuko *et al*., 2022).

The 63rd session of the United Nations (UN) General Assembly in December 2008 adopted a resolution on the ‘recognition of sickle cell anaemia as a public health problem’, and encouraged Member States to support health systems and primary health-care delivery, including efforts to improve the management of sickle cell anaemia (United Nations, 2009).

According to the World Health Organisation (WHO), about 300 000 babies are born every year with sickle cell anaemia. Seventy-five percent of this figure is said to be found in Sub-Saharan Africa (Zajac, 2022). The American Centre for Disease Control and Prevention (CDC) states that Nigeria carries 66% of the burden in the region (Alabi and Adebowale-Tambe, 2021).

A research conducted by Mava *et al*. (2012)studied the association between UTI and other common febrile illnesses of childhood among children with sickle cell anaemia. The study consisted of 180 children aged 6 months to 15 years with SCA, and 180 children with haemoglobin genotype AA served as the control group. It was observed that a UTI is common during febrile illnesses in children in both groups, and is difficult to differentiate from other common childhood diseases. This is a confirmation that signs and symptoms of urinary tract infection are non-specific and usually not referable to the urinary tract.

A study was carried out by Kondapaneni *et al*. (2012)on a total number of 200 children to screen for significant bacteriuria. The study group comprised of 140 girls and 60 boys. Out of the 200 children, 33 (16.5%) showed significant bacteriuria with female preponderance over male. The prominent bacteria isolates were Escherichia coli 09 (27.27%), Klebsiella pneumoniae 08 (24.24%), Proteus spp. 07 (21.21%), Staphylococcus aureus 05 (15%), and Pseudomonas aeruginosa 04 (12%).

In a comparative study to evaluate the prevalence of significant bacteriuria among children with sickle cell anaemia, 272 children with SCA in steady state and in crises, aged 6 months to 15 years, had their urine samples screened for significant bacteriuria. The prevalence of significant bacteriuria was found to be higher in SCA subjects in crisis (20.7%) than among those in steady state (2.2%) (Yauba *et al.*, 2014).

The rapid advent of information technology (IT) has revolutionized several sectors and areas, including health. Information technology presents several opportunities for improving and transforming healthcare, which includes reducing human errors, improving clinical outcomes, facilitating care coordination, and tracking data over time (Atarodi and Atarodi, 2019). According to Emmerich (2009), the technological evolutions, if not revolutions, prompted by the medical sciences and the modern contexts of healthcare delivery continue to accelerate.

Classification modelling, one of the offshoots of information technology, involves predicting a class label for a given example of input data. It is a task that requires the use of machine learning algorithms that learn how to assign a class label to examples from the problem domain. (Brownlee, 2020). In view of the foregoing, a classification model can be used to categorize children with sickle cell anaemia based on whether they have significant bacteriuria or not.

Paediatric patients with sickle cell anaemia are especially at risk for opportunistic infections. Prompt diagnosis and treatment of infections are, therefore, vital to their survival. Thus, this study is based on the need to create a classification model for the early detection and treatment of significant bacteriuria in children with sickle cell anaemia.

## 1.2 Statement of the Problem

Sickle cell anaemia is a common genetic blood disorder that affects millions of people worldwide. Children with sickle cell anaemia are at increased risk of developing significant bacteriuria due to their impaired immunological state. This can lead to serious complications such as kidney damage and sepsis (Mava *et al*., 2011). The accurate and timely diagnosis of significant bacteriuria in children with sickle cell anaemia is crucial for preventing these complications. However, current diagnostic approaches rely on urine culture, which is time-consuming. Therefore, the need to develop a classification model for the accurate and timely detection and treatment of significant bacteriuria in sickle cell patients, to improve clinical outcomes, arises, and hence this study.

## 1.3 Aim and Objectives of the Study

The aim of this study is to develop a classification model for significant bacteriuria in children with sickle cell anaemia.

The specific research objectives are to:

1. elicit the risk factors causing significant bacteriuria,
2. obtain data to train the classification model,
3. formulate the significant bacteriuria classification model based on the factors identified in (a), and using the data obtained in (b),
4. simulate the model, and
5. validate the model.

## 1.4 Methodology

In order to meet the research objectives of the development of a classification model for significant bacteriuria in this study, the following methods was followed:

1. The variables causing significant bacteriuria was elicited through the review of related works on the areas of significant bacteriuria and classification modelling, followed by a structured interview with a paediatrician.
2. The classification model’s training data was obtained from the patient records of a hospital, after receiving the necessary approval.
3. The model for classifying significant bacteriuria was formulated using a Python Machine Learning Package (scikit-learn). To enhance accuracy, an Ensemble Learning classification algorithm was implemented, which combined various base learners such as k-Nearest Neighbours, Logistic Regression, and Decision Tree.
4. The model was simulated using an open-source web application called Jupyter Notebook, and deployed with Streamlit.
5. The simulated model was validated and evaluated using performance metrics such as accuracy, precision, and error rates.

## 1.5 Scope and Limitation of the Study

The scope of this study covers the development of a classification model for significant bacteriuria in children with sickle cell anaemia, using Obafemi Awolowo University Teaching Hospital Complex (OAU THC) as a case study. This study is limited to the development of a model that groups paediatric patients with sickle cell anaemia into two categories, based on the occurrence of significant bacteriuria.

## 1.6 Justification of the Study

Despite the high incidence of significant bacteriuria in children with sickle cell anaemia, there is currently no standardized classification model for the diagnosis of significant bacteriuria in this population. This can hinder effective clinical decision-making, potentially resulting in delayed treatment, complications and poorer patient outcomes for affected patients. Therefore, this study provides a discourse on the contributory factors of significant bacteriuria in Nigeria, and a classification model using these parameters, to improve the accuracy and timeliness of the diagnosis and treatment of significant bacteriuria in children with sickle cell anaemia.

## 1.7 Arrangement of the Thesis

The thesis is divided into five chapters, to give details on the classification model for significant bacteriuria. The thesis is presented as follows:

Chapter one is the introductory section which has been presented in this section and it consists of the background to the study, the statement of the problem, the aim and objectives of the study, the methodology of the study, the scope and justification of the study.

Chapter two is the literature review section which contains a review of related works. It includes information about significant bacteriuria (causes, risk factors), sickle cell anaemia and classification modelling.

Chapter three is the research methodology which presents details on the identification of the variables needed, methods of data collection on significant bacteriuria, and the classification model technique that was used to formulate the model.

Chapter four is the results of the study which discusses the classification model formulated and the simulation program that was used in assessing the performance of the classification model developed for categorizing significant bacteriuria in paediatric patients who have sickle cell anaemia.

Chapter five presents the summary of the work done, the conclusions drawn and the recommendations of the study.

# CHAPTER TWO

# LITERATURE REVIEW

## 2.1 Significant Bacteriuria

Bacteriuria is defined as the presence of growth of one or more bacterial species in urine culture, irrespective of the presence of pyuria (Jordano, 2022). The term ‘significant bacteriuria’ is used to differentiate laboratory evidence of bacteriuria in the bladder from bacteriuria that probably results from contamination during micturition. Significant bacteriuria is generally accepted as detection of more than 100 000 colony forming units of a single type of bacterium per millilitre of urine. Contamination with normal flora typically results in lower numbers of bacteria per millilitre and/or mixed bacterial species. However, the criterion of more than 100 000 colony forming units per millilitre of urine has limitations, in that numbers as low as 1000 colony forming units per millilitre may be clinically significant in patients with symptoms of urinary tract infection (Cormican *et al*., 2011).

Normal urine is sterile and has no bacteria in it. However, bacteria cover the skin and are found in large numbers in the rectal area and in stools. Bacteria may, at times, travel up the urethra into the bladder. When this happens, the bacteria multiply and may cause infection, unless the body gets rid of the bacteria. (Urology Care Foundation, 2023). It has been reported that the long-term sequelae of untreated bacteriuria in children can result in renal scar formation, leading to a reduction in kidney function, particularly when the bacteriuria is associated with vesicoureteral reflux (Shaikh and Craig, 2014).

Significant bacteriuria can be classified as symptomatic or asymptomatic. A patient with asymptomatic bacteriuria (ASB) is defined as having colonization with one or more organisms in a urine specimen without symptoms suggestive of urinary tract infection. Symptomatic bacteriuria is associated with an infection in the lower or upper urinary tract, usually by a single organism (Crader *et al*., 2022). Symptoms include urinary frequency and urgency, dysuria (painful or difficult urination), lower abdominal pain, and flank pain. Urinary tract infections (UTIs) can involve the urethra, ureters, bladder, or kidneys (Imam, 2021).

The proper management of urinary tract infections depends on the accurate and rapid detection of bacteriuria. The methods employed for the detection of bacteria in urine include microscopic, chemical, cultural, and automated techniques (Pezzlo, 1983). Bacteriuria can be confirmed by a urine culture, in which bacteria from a urine sample are grown in a laboratory to identify the numbers and type of bacteria (Imam, 2022).

Urinary tract infection is the most common bacterial infection in childhood and up to 30% of infants and children experience recurrent infections during the first 6–12 months after initial UTI (Stein *et al*., 2015). In the United States, there are about 1.5 million paediatric ambulatory visits annually for UTIs (Silva *et al*., 2020). Another study estimated that 150 million people worldwide are affected by UTIs annually, resulting in more than 6 billion dollars in direct health care expenditures. In 2019, more than 404.6 million individuals had UTIs globally and nearly 240 000 people died of UTIs (Zeng *et al*., 2019).

In a retrospective study to determine the bacterial profile and prevalence of antibiotic resistance patterns of uropathogens causing urinary tract infections in children from age 0 to 12 in Al-Amiri Hospital, Kuwait, Escherichia coli accounted for 67.3% of the 1342 samples that yielded significant bacteriuria. Further, a significantly higher incidence of UTI was observed in females compared with males. Of the 1342 samples, 1020 (76%) were from female patients and 322 (24%) were from male patients (Al Benwan and Jamal, 2022).

### 2.1.1 Significant Bacteriuria in Africa

Symptomatic bacteriuria, also known as a urinary tract infection, is by far the most common cause of serious bacterial infection (SBI), accounting for the majority (85%) of SBI in febrile infants less than 90 days. The risk of invasive bacterial infection (bacteraemia and bacterial meningitis) is even higher among febrile infants with UTI compared to those without UTI (Manuel *et al*., 2022).

A cross-sectional research study of 71 sickle cell disease (SCD) patients aged one to fifty years in Ghana found the prevalence of significant bacteriuria among SCD patients to be most common in females and SS genotypes. Significant bacteriuria was more in females (66.7%) than in males (33.3%) and higher in SS genotype (83.3%) than in SC genotype (16.7%). Escherichia coli was the predominant isolate (Adjato *et al*., 2019).

Mitiku *et al*. (2022) carried out an analytical study on 422 UTI-suspected participants in Ethiopia. Among the 422 urine samples, 129 cultures were found to have significant bacteriuria, and a total of 131 Gram-negative bacteria were isolated. During analysis, the following factors had a statistically significant association with Gram-negative bacterial uropathogens: gender, previous episodes of UTI, history of antibiotic usage in the previous six months, presence of chronic underlying diseases, and hospitalization.

One hundred samples which produced bacterial counts equivalent to significant bacteriuria (105 CFU/mL) were examined to determine the antibiotic resistance pattern of isolates. The bacterial uropathogens displayed varying resistance rates against different antibiotics (Brown *et al*., 2022).

In Ghana, a cross-sectional study involving 110 sickle cell disease patients and 110 age and sex matched healthy controls revealed that SCD patients have a two-fold risk of developing significant bacteriuria compared to healthy people, and the risk is particularly high among female SCD patients. The overall bacteriuria prevalence was significantly higher among SCD patients (17.2%) than among the control group (8.2%), and the microbial aetiology of bacteriuria in the SCD patients were more diverse compared to healthy people (Donkor *et al*., 2017).

A research conducted in Bushenyi District, Uganda revealed that the prevalence of bacterial UTI among patients attending hospitals was 86/267 (32.2%). Out of this bacterial UTI prevalence, symptomatic and asymptomatic patients contributed to 46/86 (53.5%) and 40/86 (46.5%), respectively. Almost half of the patients having significant bacteriuria were asymptomatic, and this situation is of utmost concern since asymptomatic bacteriuria is a strong predictor of ensuing symptomatic UTIs (Odoki *et al*., 2019).

### 2.1.2 Significant Bacteriuria in Nigeria

Bacteriuria has to be considered as an abnormal laboratory finding which should always alert physicians to the presence of UTI. It causes cystitis, frequency of urination, dysuria, suprapubic pain, sometimes haematuria, and usually pyuria, pyelonephritis, loin pain, rigors, fever and often bacteraemia. Some of the symptoms of symptomatic bacteriuria are very severe lower abdominal pain mainly on the affected side, high fever and splitting headache; frequent and difficult voiding of urine. Some UTI cases become chronic, with bloody pus discharge that may eventually lead to pelvic inflammatory disease (Dada and Akharaiyi, 2012).

In the paediatric age group, the diagnosis of UTI is often missed because the clinical features are seldom overt and, in most cases, not referable to the urinary tract (Okafor *et al*., 1993). The endemic nature of malaria, acute febrile viral illnesses and other parasitic infections and infestations in the tropics has also masked the clinical diagnostic precision of UTI in children by clinicians practicing in the region (Jombo *et al*., 2010).

A study by Asinobi *et al*. (2003) to determine the prevalence of bacteriuria and bacterial isolates in the urine of febrile children with sickle cell anaemia (SCA) was carried out at University College Hospital, Ibadan on 171 febrile children (aged 1-15 years) with SCA and 171 age- and sex-matched controls. The prevalence of bacteriuria in children with SCA was 21.6% compared with 15.8% in the controls. In the SCA group, significant bacteriuria also occurred with other conditions such as pneumonia and osteomyelitis. Urinary tract infection is common in children with SCA, therefore routine screening for it is recommended during febrile illnesses. Children with fever from other overt causes, however, should not be exempted from the urine screening procedure in case there might be concomitant UTI.

Nosocomial significant bacteriuria (NSB) appears to be quite common among hospitalized patients; it has been reported to account for 20.6% - 28.2% of all nosocomial infections. Significant bacteriuria, whether nosocomial or community-acquired, remains one of the principal causes of morbidity and mortality in infants and young children, especially when the kidneys are involved. Pyelonephritis (infection of the kidney) may result in severe renal scarring with serious complications such as hypertension and renal failure in later life. From the foregoing, it is possible that severe renal damage may occur should significant bacteriuria complicate non-infective urinary tract disorders (NUTDs) namely, acute glomerulonephritis, nephrotic syndrome, renal failure and congenital urinary tract anomalies. Such a complication may lead to either death or chronic ill health in survivors. These potential dangers underscore the need to diagnose and manage NSB early before clinical deterioration sets in (Olowu and Oyetunji, 2003).

An epidemiological survey of significant bacteriuria was conducted on patients attending the sickle cell clinics of Lagos State University Teaching Hospital, Ikeja. Significant bacteriuria was found in one‑tenth of sickle cell patients, and nearly half of the participants who had significant growth had asymptomatic bacteriuria (Akinbami *et al*., 2014).

### 2.1.3 Significant Bacteriuria in Sickle Cell Patients

Sickle cell disease is a group of inherited red blood cell disorders that affect haemoglobin, the protein that carries oxygen through the body. The condition affects more than 20 million people worldwide. Normally, red blood cells are disc-shaped and flexible enough to move easily through the blood vessels in order to deliver oxygen throughout the body. The red blood cells of a patient with sickle cell disease are crescent- or sickle-shaped. These cells are not flexible and can cause blockages that slow or stop the flow of blood (NHLBI, 2022).

The sickle cells die early, which causes a constant shortage of red blood cells. In addition, when they travel through small blood vessels, they get stuck and clog the blood flow. This can cause attacks of sudden severe pain, called pain crises, and other serious complications such as infection, acute chest syndrome and stroke (CDC, 2022).

The greatest burden of sickle cell disease is in sub-Saharan Africa, where access to medical care and public health strategies to decrease mortality and morbidity are not uniformly available (DeBaun and Galadanci, 2022). Nigeria currently has the highest burden of sickle cell disease in the whole world ahead of Democratic Republic of Congo and India, with an estimated 25% of her adult population being carriers of defective S-gene. About 150 000 babies are born annually in Nigeria with SCD and 50-80% die before age five (5) (Ezigbo, 2021).

Children with sickle cell anaemia have increased tendency to develop frequent and severe infections. They are often susceptible to infections of bones, lungs and urinary tract. These are thought to be due to abnormalities of the body defence mechanisms (Mava *et al*., 2011). Due to the subnormal immunity resulting from reduced or absent splenic function, children with SCD are at a higher risk for bacterial infections (Patel *et al*., 2020).

According to a study undertaken in Jamaica to determine the prevalence of significant bacteriuria in patients with sickle cell disease, bacteriuria is a significant problem in individuals with SCD and may be the source of pathogens in UTI. (Ali *et al*., 2006)

A cross-sectional study was conducted on 320 children with sickle cell anaemia aged 6 months to 12 years in Tanzania. The results showed a high prevalence of blood stream infections (bacteraemia) and urinary tract infections (bacteriuria). Out of the 320 participants, 28 had blood stream infections and 62 had urinary tract infections. Factors found to independently predict urinary tract infections were female sex, age between 2-5 years, and living in rural areas (Munaku *et al*., 2022).

Significant bacteriuria has been found to be more prevalent among children with SCA than their counterparts with normal haemoglobin (Adegoke and Adegun, 2013). Children with sickle cell anaemia may have compromised kidney function arising from repeated vaso-occlusive episodes and recurrent symptomatic or asymptomatic UTI (Chukwu *et al*., 2011).

### 2.1.4 Significant Bacteriuria Risk Factors

Bacteriuria is common in childhood and appropriate management of the acute episode as well as prevention is required to minimize the long-term sequelae of UTIs, such as renal scarring, hypertension, and renal failure. Sound knowledge of the factors associated with bacteriuria may allow prompt and appropriate intervention that can easily bring the disease under control. The following are the factors that predispose to significant bacteriuria:

1. **Age**

Young age is a risk factor for bacteriuria, with children less than 1 year of age being at greatest risk (Bachur, 2004)**.** Urinary tract infections have now emerged as the most common serious bacterial infection in childhood, with approximately 7–8% of girls and 2% of boys affected during the first 8 years of life. Because the clinical presentation of UTI in children can be nonspecific, any child with an unexplained fever, especially if younger than 2 years, should be tested for UTI (Bitsori and Galanakis, 2012).

1. **Gender**

Studies of the prevalence of bacteriuria among children presenting to an emergency department with fever found that girls were more than twice as likely as boys to have UTIs (Zorc *et al*., 2005). Several reports have indicated that females are more predisposed to significant bacteriuria than their male counterparts and this has been related in part to the anatomy of the female genitourinary system, as well as hormonal, and some behavioural factors (Boye *et al*., 2016).

1. **Malnourishment**

Severe acute malnutrition is associated with immune deficiency, which expectedly renders affected children more vulnerable to severe infections (Uwaezuoke, 2016). Tripathi *et al*. (2003), in their study, found that the incidence of bacteriuria was significantly higher among children with moderate and severe malnutrition. Results also showed that the risk of bacteriuria increased significantly with increasing severity of malnutrition. The incidence of bacteriuria was threefold higher in patients with severe malnutrition in comparison to those with moderate malnutrition.

1. **Nocturnal Enuresis or Bed Wetting History**

A study carried out in Nigeria revealed a strong relationship between significant bacteriuria and enuresis. Among the participants, those with enuresis and significant bacteriuria were 42%, while those without enuresis but with significant bacteriuria were 3.1% (Inyang-Etoh *et al*., 2021). Nocturnal enuresis is often associated with lower urinary tract symptoms like dysuria, urgency, suprapubic pain, and daytime incontinence (Shah *et al*., 2018).

1. **Diabetes**

Infection is one of the most frequent causes of morbidity and mortality in diabetic patients. Diabetics are more likely to have asymptomatic and symptomatic bacteriuria (Bebars *et al*., 2022). Agrawal *et al*. (2018) conducted a study with the aim of evaluating the prevalence and aetiology of significant bacteriuria in diabetics as well as non-diabetics. Of the total 78 subjects who had significant bacteriuria, 56 were diabetics and 22 were non-diabetics, depicting a higher rate of significant bacteriuria in patients with diabetes mellitus.

1. **Postoperative Phase**

In a retrospective analysis of all patients who underwent endo-urological procedures at a urology unit, the overall post‐operative bacteriuria rate was 17.85% (Seenappa *et al*., 2022). Based on the findings of Dahiya and Goldman (2018), bacteriuria is common after renal transplant, affecting nearly half of transplant recipients within the first year. Thus, they suggest that in the unique population of children who have received renal transplants or who are undergoing invasive procedures involving the urogenital tract, treatment of significant bacteriuria should be considered.

1. **History of Urinary Infection**

History of symptomatic urinary tract infection is a potential predictor of significant bacteriuria. (Adegoke and Adegun, 2013). Approximately 5% to 10% of children have at least one episode of urinary tract infection, of whom 30% to 50% will have one or more recurrences (Roupakia *et al*., 2013).

These risk factors can be used as features to build machine learning models in order to predict bacteriuria.

## 2.2 Machine Learning in Healthcare

Artificial Intelligence (AI) is a broad discipline that aims to understand and design systems that display properties of intelligence – emblematic of which is the ability to learn: to derive knowledge from data. Machine Learning (ML) is a sub-discipline of AI, where algorithms learn associations of predictive power from examples in data. Artificial intelligence and machine learning have the potential to be the catalyst for transformation of health systems to improve efficiency and effectiveness, create headroom for universal health coverage and improve outcomes. Machine learning has the potential to improve hypothesis generation and hypothesis testing tasks within a health system by revealing previously hidden trends in data, and thus has the potential for substantial impact at both the individual patient and system level (Panch *et al*., 2018).

There are huge amounts of data generated regularly by health department, which is very complex, and large to process and analyse by traditional methods (Singh and Chaurasia, 2020). Increased use of information systems in health services and digitization of patient information generates large volumes of data. Data is stored in many systems in different formats. The combination of large amounts of varied data, increased computing power and better methods enables fast and automated production of machine learning algorithms that are able to analyse complex data with accurate results (Makhlysheva *et al*., 2018).

Health care field generates big data about clinical assessment, report regarding patient, cure, follow-up, medication, etc. The analysis of this high-dimensional and multimodal bio-medical data promises the improved accuracy of perception, prediction and diagnosis of disease (Meherwar and Maruf, 2017).

When the tremendous growth in data is combined with the increasing ability of computers to process and use the data to formulate various machine learning algorithms, there is an opportunity to use machine learning to help humans make decisions by taking into account a significant amount of contextual information. Any machine-learning model depends on high-quality data that are representative of the population to which the model’s results are to be generalized. Thus, formulating effective data management at all levels is an essential requirement for integrating ML models into health care. In addition, pipelines for data processing and ML with user-friendly front-ends for the products must be formulated. These pipelines can transform the raw data into datasets that can be used to train various ML models. (Alanazi, 2022).

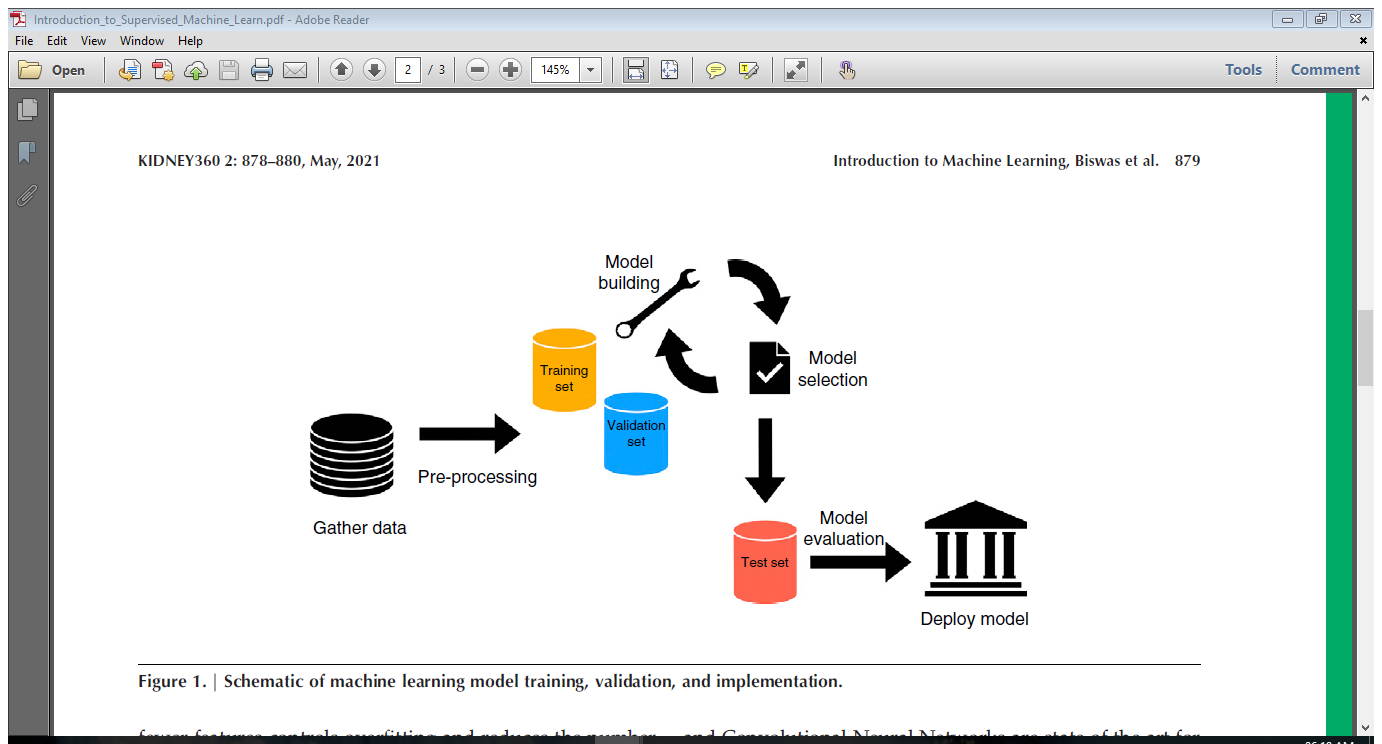
### 2.2.1 Data Mining and Machine Learning

Data mining is the process of extracting useful information and patterns from large amounts of data (Ramageri, 2010). Data mining is an interdisciplinary field, the confluence of a set of disciplines, including database systems, statistics, machine learning, visualization, and information science (Han and Kamber, 2006).

Jassim and Abdulwahid (2021) describe data mining as a collection of techniques applied to large and complex databases to eradicate randomness and discover concealed patterns. It is a complex procedure involving processes such as data cleaning, data integration, data selection, data transformation, data pattern analysis and evaluation, and knowledge representation.

Machine learning is a collection of algorithms and techniques used to design systems that learn from data. These systems are then able to perform predictions or deduce patterns from the supplied data. Machine learning algorithms fall into two broad categories: supervised learning algorithms which are trained with labeled data, and unsupervised learning algorithms which are used on data with no labels (Lee, 2019).

The original dataset is split into training and validation subsets. In this manner, we can experiment with different models (and variables within models) to determine what generalizes best in the validation subset. Figure 2.1 depicts the process of machine learning.



**Figure 2.1 Schematic of Machine Learning Model Building**

Source: (Biswas *et al*., 2021)

### 2.2.2 Supervised Machine Learning

In supervised learning, the machine learning algorithms are trained using large collections of examples that are labelled with the correct answer or category. The algorithm learns how to properly identify or classify a given input by processing large, labelled datasets until it can do so with an elevated level of confidence. Supervised learning primarily deals with problems of prediction and forecasting. Supervised learning algorithms look at existing data to make an inference about new instances or predict the outcome of a yet-to-be-seen example. Thus, supervised learning algorithms can have powerful predictive and discriminatory capabilities, provided the availability of enough quality training data that is also sufficiently representative of the targeted population (Shute *et al*., 2023).

Supervised learning algorithms generate a function that maps inputs to desired outputs. In supervised algorithms, the classes are predetermined. The task of the machine-learning algorithm is to find patterns and construct mathematical models. There are two main supervised models: classification models (classifiers) and regression models. Regression models map the input space into a real-value domain. The classifiers map the input space into predefined classes (Nasteski, 2018).

## 2.3 Classification Modelling

Classification is a supervised learning concept, which categorizes a given set of data into predefined classes. The classes are often referred to as target, label or categories. Classification modelling is the task of approximating the mapping function from input variables to discrete output variables. The main goal is to identify which class/category the new data will fall into (Waseem, 2022).

The classifier is required to learn in order to approximate the behaviour of a function, which maps a vector into one of several classes by looking at several input-output examples of the function. (Osisanwo *et al*., 2017). After inferring a function from a given set of data (inputs with their respective outputs) with which the computer is trained, the performance of the resulting function should be measured on a test set that is separate from the training set (Dasgupta and Nath, 2016).

Classification models represent knowledge about the relationship between the classes to which data instances are assigned and their attribute values, and therefore can be used to predict the unknown class of arbitrary new instances from the same domain, or simply to achieve a better understanding of the available data (Cichosz, 2011).

Classification Algorithms include Logistic regression, Naive Bayes, K-Nearest Neighbours, Decision Tree, Random Forest and Support Vector Machine.

### 2.3.1 K-nearest Neighbours

Nearest neighbour classification, also known as k-nearest neighbours (KNN), is based on the idea that the nearest patterns to a target pattern , for which we seek the label, deliver useful label information. KNN assigns the class label of the majority of the k-nearest patterns in data space. For this sake, we have to be able to define a similarity measure in data space to find the k closest training points, that is, small according to some metric, viz. Euclidean, Manhattan, Hamming, etc. (Kramer, 2013).

K-nearest neighbour classification finds groups of k objects in the training set that are closest to the test object, and bases the assignment of a label on the predominance of a particular class in this neighbourhood. There are several key elements of this approach: (i) the set of labelled objects to be used in evaluating a test object’s class, (ii) a distance or similarity metric that can be used to compute the closeness of objects, (iii) the value of k, the number of nearest neighbours, and (iv) the method used to determine the class of the target object based on the classes and distances of the k-nearest neighbours (Steinbach and Tan, 2009).

KNN is a lazy learning technique that does not learn until the test example is given. It is memory-based; there is no explicit training or model. Whenever we have a new data to classify, we find its k-nearest neighbours from the training data (Kozma, 2008).

### 2.3.2 Logistic Regression

Logistic regression is used to model a dependent or response variable based on one or more other variables called predictors. It is particularly valuable in that the predictions made from a fitted model are probabilities, constrained to be within the range of values 0–1. For a logistic regression model to fit the data well, it is assumed that the predictors are uncorrelated with one another but significantly related to the response, and that the observations or data elements of a model are also uncorrelated (Hilbe, 2015).

Logistic regression, sometimes called the logistic model or logit model, analyses the relationship between multiple independent variables and a categorical dependent variable, and estimates the probability of occurrence of an event by fitting data to a logistic curve. There are two models of logistic regression, binary logistic regression and multinomial logistic regression. Binary logistic regression is used when the dependent variable is dichotomous. When the dependent variable is not dichotomous and is comprised of more than two categories, multinomial logistic regression can be employed (Park, 2013).

Since logistic regression calculates the probability of an event occurring over the probability of an event not occurring, the impact of independent variables is usually explained in terms of odds. The odds are simply the ratio of the proportions for the two possible outcomes. When response variable is just 1 or 0 (success or failure), the mean is the probability of a success. Logistic regression models the mean in terms of an explanatory variable . We might try to relate and as in simple linear regression: . However, extreme values of will give values of that fall outside the range of possible values of ,.

The logistic regression solution to this difficulty is to transform the odds using the natural logarithm. The term log odds or logit is used for this transformation:

, where is a binomial proportion and is the explanatory variable. The parameters of the logistic regression model are (Horton, 2020).

The result is the impact of each variable on the odds ratio of the observed event of interest. The main advantage is to avoid confounding effects by analyzing the association of all variables together (Sperandei, 2014).

### 2.3.3 Decision Tree

A decision tree is a classifier expressed as a recursive partition of the instance space. The decision tree consists of nodes that form a rooted or directed tree. It has a root node that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. Each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. All other nodes are called leaves or terminal nodes. A leaf node represents a class node and has exactly one incoming node and no outgoing node. Instances are classified by navigating them from the root of the tree down to a leaf according to the outcome of the tests along the path (Rokach and Maimon, 2015).

A decision tree is a flowchart-like tree structure. Internal node denotes an attribute; Branch represents the values of the node attribute; Leaf nodes represent class labels or class distribution (Wasilewska, 2022).

The following are to be considered while growing a tree: Features to choose, Conditions for splitting, stopping criteria and pruning. The decision to make a strategic split heavily affects the accuracy of the tree. Entropy/Information gain or Gini Index can be used for choosing the best split, and pruning can be used to remove redundant comparisons or subtrees while growing the tree or when the tree is fully grown up (Krishnan, 2021).

Entropy is a measure of disorder, impurity, or uncertainty in the given dataset. The value of entropy always lies between 0 and 1; the closer its value to 0, the better. The higher the entropy, the harder it is to draw any conclusion. The entropy of the classification of set 𝑆 with respect to 𝑐 states can be mathematically represented as , where is the probability of randomly choosing data from class (Jijo and Abdulazeez, 2021).

The Information Gain measures the expected reduction in entropy, and consequently, the reduction in impurity in the data. It tells us how much information a feature gives us about the class. The feature that has minimum impurity will be considered as the root node. Information gain of a parent node can be calculated as the entropy of the parent node minus entropy of the weighted average of the child node. Assume that using attribute A, a set S will be partitioned into sets S1, S2,…,Sv where v is number of values of the attribute A. If Si contains pi examples of P and ni examples of N, the encoding information that would be gained by branching on A is (Gopalan, 2020).

Decision trees provide a pictorial description of a well-defined decision problem. They are useful because they provide a clear, documentable and discussable model of either how the decision was made or how it will be made (Eriksen *et al*., 2001).

### 2.3.4 Ensemble Learning

An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way, typically by weighted or unweighted voting, to classify new examples (Dietterich, 2000). Ensemble learning can be broken down into two tasks: developing a population of base learners from the training data, and then combining them to form the composite predictor (Friedman *et al*., 2009).

Classification ensemble learning techniques have demonstrated powerful capacities to improve upon the classification accuracy of a base learning algorithm (Webb and Zheng, 2004). It has been discovered that ensembles are often much more accurate than the individual classifiers which make them up, as the ensemble methods, also known as committee-based learning or multiple classifier systems, train multiple hypotheses to solve the same problem. The three main classes of ensemble learning methods are bagging, boosting, and stacking (Choudhury, 2019).

Bootstrap aggregating, often abbreviated as bagging, is an ensemble technique that combines multiple learners trained on randomly drawn subsamples of the original data. To build a bagging model, we generate multiple datasets by bootstrapping the training data and then develop models based on the individual datasets and make predictions using these models. All the predictions are then combined to produce a representative value, depending on the problem to be solved. Since an individual learner is often sensitive to noise in the training data, bagging, by aggregating multiple results in a single prediction, should provide stable and improved results with decreased variance (Zhang *et al*., 2022).

Boosting is a general method for improving the performance of any learning algorithm. It involves incrementally building an ensemble by training each new model instance to emphasize the training instances that previous models misclassified. Observations are weighted to be repeated more frequently. Thus, new training data is created. The reason for this is to enable the algorithm to learn better from incorrectly predicted observations. All models run consecutively, and it is decided by weighted voting which hypothesis is the most correct. Schapire introduced the first boosting algorithm in 1990 (Turan and Cengiz, 2022).

Stacking, also known as stacked generalization, is an ensemble method that seeks a diverse group of members by varying the model types fit on the training data and using a model to combine predictions. Stacking is a procedure where a learner is trained to combine the individual learners. Here, the individual learners are called the first-level learners, while the combiner is called the second-level learner or meta-learner (Brownlee, 2021).

## 2.4 Justification for the use of Ensemble Learning

Ensembling is a well-known strategy that consists of fusing several different models to achieve a new model for classification or regression tasks. The main idea of ensemble methodology is to combine a set of models, each of which performs the same original task, in order to obtain a better composite global model, with more accurate and reliable estimates or decisions than can be obtained from using a single model. Ensembling has been proven to provide superior performance in various contexts.

This type of work is a binary classification task with the aim of identifying the category of a new observation based on the set of training data containing the observed categories. The accuracy of the classification rules must be acceptable for the rules to be applied to the new data tuples. By using multiple learners, the generalization ability of an ensemble can be much better than that of a single learner, and accuracy can be improved.

## 2.6 Related Works

A research work conducted by Ozkan *et al*. (2018) aimed to develop an artificial intelligence model to support the diagnosis of UTI with complex symptoms. For this purpose, routine examination data and definitive diagnosis results for 59 UTI patients were gathered and composed as a UTI dataset. Four different artificial intelligence methods, i.e., decision tree (DT), support vector machine (SVM), random forest (RF) and artificial neural network (ANN), were used to create classification structures. Accuracy, specificity, and sensitivity statistical measurements were used to determine the performance of the created models. The comparison of individual AI methods showed that ANN had the highest accuracy result of 98.3% for UTI diagnosis. Unlike clinical-based diagnosis, this ANN model only needs the variables of pollakiuria, suprapubic pain, and erythrocyturia to receive a proper diagnosis with similar accuracy. This study was limited to the use of a single model (ANN) for the final classification result.

Burton *et al*. (2019) in their study tried to use artificial intelligence to reduce diagnostic workload without compromising the detection of UTIs. The researchers’ aim was to identify which markers in the urine samples were the most sensitive and specific in order to help diagnose UTIs without the need to culture. They retrospectively analysed 212,554 urine reports using two methods of classification: a heuristic model and a machine learning approach. All of the machine learning algorithms (Random Forest, Neural Network, Extreme Gradient Boosting) were developed using a set of 16 factors and they outperformed the heuristic model. The optimal solution was found to be three Extreme Gradient Boosting algorithms, trained independently for the classification of pregnant patients, children, and then all other patients. When combined, this system granted a relative workload reduction of 41% and a sensitivity of 95% for each of the stratified patient groups. The limitations of this study were relatively sparse clinical details and difficulty in clarifying details.

In a study by Taylor *et al*. (2018), a retrospective cohort analysis of 80,387 adults who visited the emergency department with urine culture results and UTI symptoms was performed. The authors developed models for UTI prediction with six machine learning algorithms using demographic information, vitals, laboratory results, medications, past medical history, chief complaint, and structured historical and physical exam findings. Models were developed with both the full set of 211 variables and a reduced set of 10 variables. The top performing algorithm for both the full and reduced models was extreme gradient boosting (XGBoost), which had an area under the curve of 0.904. The XGBoost full and reduced models demonstrated greatly improved specificity in comparison with the provider judgment proxy of UTI diagnosis or antibiotic administration, while also demonstrating superior sensitivity when compared with the documentation of UTI diagnosis. This study was limited to data elements available during each emergency department visits.

A research work by Heckerling *et al*. (2007) used artificial neural network (ANN) coupled with genetic algorithms to determine combinations of clinical variables optimized for predicting UTIs. Feed-forward backpropagation ANN was used to determine the relation between symptom and urinalysis input data and urine culture output data. The ANN model examined 212 women enrolled in the study aged between 19 and 84 with symptoms of UTIs. Five-variable sets were created that classified cases of urinary tract infection and non-infection with receiver operating characteristic (ROC) curve areas that ranged from 0.853 (95% CI, 0.796–0.909) to 0.792 (95% CI, 0.726–0.858). Using a network output threshold of 0.25, the accuracy was 0.764. Network influence analysis revealed that some factors (urinary frequency, foul urine odour, etc.) predicted urine infection in unexpected ways, and interacted with other variables when making predictions. This study was limited to the female gender and the use of artificial neural network models.

The study performed by Mancini *et al*. (2020) attempted to build machine-learning models for predicting multidrug resistant urinary tract infections (MDR UTI). To achieve this goal, learning methods such as Catboost, Support Vector Machine and Neural Networks were utilized to build predictive models. Moreover, the authors integrated an easy-to-use cloud platform, called DSaaS (Data Science as a Service), which is well suited for hospital structures, where healthcare operators might not have specific competences in using programming languages but need to analyse data as a continuous process. DSaaS was used on data related to a total of 1486 hospitalized patients with nosocomial urinary tract infection (UTI). Five factors viz. sex, age, age class, ward and time period, were used to predict the onset of a MDR UTI. However, potential risk factors that were unavailable from the review of medical records were not considered. Catboost exhibited the best predictive results with the highest value in every metric already implemented in DSaaS. Its accuracy and area under receiver operating characteristic curve were 0.717 and 0.739 respectively.

A retrospective cohort study by Jeng *et al*. (2022) was aimed at predicting the development of recurrent urinary tract infection (RUTI) related to E. coli using machine learning. There were two stages for development of prediction models for RUTI. The first stage was a scenario at the clinical visit for UTI with 963 patients. The second stage was a scenario after hospitalization with a complete survey of 809 patients with UTI caused by Escherichia coli. Both host and bacterial characteristics made important contribution to the development of RUTI in the prediction models in the two clinical scenarios, respectively. Three machine learning models, logistic regression (LR), decision tree (DT), and random forest (RF) were built for the RUTI prediction. The RF model had higher prediction accuracy than LR and DT (0.700, 0.604, and 0.654 in stage 1, respectively; 0.709, 0.604, and 0.635 in stage 2, respectively). It was found that the decision rules constructed by the DT model could provide high classification accuracy (up to 0.92 in stage 1 and 0.94 in stage 2) in certain subgroup patients in different scenarios. A limitation of this study is that not all important characteristics of patients and bacteria were included for analysis.

In summary, the related works discussed above focused on the prediction of symptomatic bacteriuria in mostly teenagers and adults using machine learning. This current study focuses on developing a classification model for the timely detection of significant bacteriuria in children with sickle cell anaemia using ensemble learning. This is a different approach as it centres on a specific paediatric population group and its specific risk factors.

**Table 2.1 Table of Related Works for the Classification of Significant Bacteriuria**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S/N** | **Authors** | **Title** | **Scope** | **Methods** | **Result/Strength** | **Limitations** | **Remarks** |
| 1. | Ozkan, I.A., Koklu, M., and Sert, I.U. (2018) | Diagnosis of urinary tract infection based on artificial intelligence methods | Prediction of UTI with complex symptoms | Artificial Neural Network | The model had an accuracy result of 98.3% for UTI diagnosis. | This study was limited to the use of a single model (ANN) for the final classification result. | This study focused on symptomatic bacteriuria (UTI) whereas the current study focuses on both classes of bacteriuria. |
| 2. | Burton, R.J., Albur, M., Eberl, M., and Cuff, S.M. (2019) | Using artificial intelligence to reduce diagnostic workload without compromising detection of urinary tract infections | Classification of urine samples that will yield positive culture results prior to culture | Extreme Gradient Boosting  Algorithm | A potential reduction in culture workload of around 41%, while detecting 95·2 ± 0·22% of culture positive samples successfully, was demonstrated. | The limitations of this study were relatively sparse clinical details and difficulty in clarifying details. | This study focused on reducing the number of samples needing culture while the current study focuses on identifying suspected positive bacterial culture. |
| 3. | Taylor, R.A., Moore, C.L., Cheung, K.H., and Brandt, C.  (2018) | Predicting urinary tract infections in the emergency department with machine learning | Prediction of UTI in emergency department patients | Extreme Gradient Boosting  Algorithm | XGBoost achieved an AUC of 0.904 and overall accuracy of 87.5%. | This study was limited to data elements available during each emergency department visits. | This study focused on training and comparing predictive models for  UTI in a large diverse set of ED patients whereas the current study focuses on only children with sickle cell anaemia. |

**Table 2.1 Table of Related Works for the Classification of Significant Bacteriuria (cont’d)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **S/N** | **Authors** | **Title** | **Scope** | **Methods** | **Result/Strength** | **Limitations** | **Remarks** |
| 4. | Heckerling, P.S., Canaris, G.J., Flach, S.D., Tape, T.G., Wigton, R.S., and Gerber, B.S. (2007) | Predictors of urinary tract infection based on artificial neural networks and genetic algorithms | Prediction of urinary tract infection among women with urinary complaints | Artificial Neural Network | ANN classified cases of infection and non-infection with ROC areas that ranged from 0.853 to 0.792 Using a network output threshold of 0.25, the accuracy of the model was 0.764. | This study was limited to the female gender and the use of artificial neural network models. | A generally healthy outpatient population of women was examined in this study but the current study focuses on children (male and female) with sickle cell anaemia. |
| 5. | Mancini, A., Vito, L., Marcelli, E, Piangerelli, M, De Leone, R., Pucciarelli, S.., and Merelli, E. (2020) | Machine learning models predicting  multidrug resistant urinary tract infections  using “DsaaS” | Binary classification based on whether an individual is at risk of contracting a multidrug resistant urinary tract infection or not | Catboost in DSaaS | The results of Catboost in DSaaS showed an accuracy rate of 0.717. Its area under receiver operating characteristic curve was 0.739. | Potential risk factors, which were unavailable from the review of medical records, were not considered. | This study focused on the use of classification algorithms implemented in DSaaS while the current study focuses on the use of Python ensemble classification algorithms. |
| 6. | Jeng, S.-L. , Huang, Z.-J. , Yang, D.‑C. , Teng, C.‑H. , and Wang, M‑C. (2022) | Machine learning to predict the development of recurrent urinary tract infection related to single uropathogen, Escherichia coli | Prediction of the development of recurrent urinary tract infection (RUTI) related to E. coliusing machine learning | Random Forest (RF) and Decision Tree (DT) | RF model had an accuracy of 0.700 and 0.709 in stage 1 and 2 respectively. The DT model could provide an accuracy of up to 0.92 in stage 1 and 0.94 in stage 2 in certain scenarios. | Not all important characteristics of patients and bacteria were included for analysis. | This study focused on only recurrent bacteriuria whereas the current study focuses on recurrent and non-recurrent bacteriuria. |

# CHAPTER THREE

# METHODOLOGY

## 3.1 Introduction

This study requires the development of a classification model for significant bacteriuria in children with sickle cell anaemia. The risk factors of significant bacteriuria represent the variables used for the development of the classification model, and a Python Machine Learning Package (scikit-learn) was used to formulate the classification model.

## 3.2 Variable Description

In this study, the variables were identified through the review of existing literature on significant bacteriuria in paediatric patients, followed by a structured interview with a paediatrician. The detailed description of the variables is provided below. Refer to Table 3.1 for a comprehensive overview.

1. **Age:** This describes the age of the patient in years. It is recorded as a numeric value, within the range of 0 to 17, which aligns with the classification of paediatric patients.
2. **Gender:** This describes whether the patient is male or female.
3. **Malnourishment:** This indicates whether the patient is suffering from malnutrition or not. It is recorded as either yes if the patient is malnourished, or no if the patient is not malnourished.
4. **History of Urinary Infection:** This pertains to whether the patient has previously experienced a urinary infection. It is recorded as yes or no.
5. **Fever:** This indicates the presence of fever. It is recorded as yes or no.
6. **Presence of Urinary Symptoms:** This shows whether there are clinical features suggestive of a urinary tract infection. It is recorded as yes or no.

**Table 3.1: Identification of Variables Required for Developing Classification Model**

|  |  |  |  |
| --- | --- | --- | --- |
| **S/N** | **Variable** | **Labels** | |
| 1. | Age | 0…17 | |
| 2. | Gender | Male | Female |
| 3. | Malnourishment Level | Yes | No |
| 4. | History of Urinary Infection | Yes | No |
| 5. | Fever | Yes | No |
| 6. | Presence of Urinary Symptoms | Yes | No |

Following the identification of the variables that were needed for the classification of significant bacteriuria in children with sickle cell anaemia, there was a need to collect data from patients to develop and validate the classification model. Therefore, the required dataset was obtained by extracting information from 412 patient records of OAU THC, Ile-Ife, Osun State.

## 3.3 Data Collection

Approval was sought from the ethics committee of OAU THC to access patient data. After obtaining approval, a comprehensive dataset containing 412 records was collected from the patient records of paediatric sickle cell patients with suspected urinary tract infections, from age 0 to 17. All patient information was anonymized to protect patient privacy, and ensure confidentiality.

This dataset includes relevant variables for predicting significant bacteriuria such as age, gender, presence of symptoms related to urinary tract infection, medical history, and laboratory test results.

The target variable is the test results, which indicates whether significant bacteriuria is present or absent. All other variables serve as explanatory variables, which influence the outcome of the target variable.

## 3.4 Data Preparation

Data preparation plays a crucial role in ensuring that the data is accurate, and ready for the subsequent stages of data analysis and model development.

An exploratory data analysis was performed to gain insights on the collected data. This step involved summarizing and visualizing the data to understand its distribution and identify patterns.

In addition, categorical features were transformed into numeric variables, in order to utilize them in training the classification model. Categorical variables typically consist of non-numeric values, which could present difficulties for their direct inclusion in the model without undergoing this transformation.

**Furthermore, to optimize the dataset for this specific task of model development, two columns, namely 'Record\_Year' and 'Patient\_Number' were removed. These columns were deemed irrelevant to the model training objective, and were thus dropped to streamline the data and focus on the essential features for the model's training and performance.**

## 3.5 Development of Classification Model

Prior to running machine learning algorithms or any other stochastic procedures, a fixed random seed value (21) was set to ensure consistent outcomes. By using the same seed value in subsequent runs of the code, the model’s random number generator will produce the same sequence of random numbers. Consequently, the results of model training, evaluation, and other stochastic procedures will remain the same across various executions of the code.

The dataset was subsequently divided into a 70% training set and a 30% test set. The training set was used to train the classification models, and the test set was used to assess the performance of the model on unseen data.

An Ensemble Learning approach known as Stacking was used to combine multiple diverse base learners (k-Nearest Neighbours, Logistic Regression and Decision Tree) to create a more robust and precise classification model. This approach leverages the strengths of individual models, and integrates their outputs to obtain a final result for the classification task. The predictions of the base models served as inputs for the meta-model (scikit-learn Voting Classifier) which combined and weighted these predictions to make the final classification for significant bacteriuria.

All of this was done in an open-source web application called Jupyter Notebook, which provides an interactive environment to write and execute code. Its capabilities were leveraged for efficient exploration, model training, and interpretation of results.

## 3.6 Evaluation and Validation of Classification Model

This step measures how well the model generalizes to new, unseen data, and assesses its ability to correctly classify cases of significant bacteriuria. The developed classification model for significant bacteriuria in paediatric patients with sickle cell anaemia was evaluated using the following performance metrics.

* **Accuracy:** This measures the proportion of correctly identified instances out of the total number of instances in the test set.
* **Precision:** This is the proportion of true positive predictions out of the total number of positive predictions. It measures how many of the predicted positive instances are actually positive.
* **Recall:** This is also known as True Positive Rate (TPR). it is the ratio of true positive predictions to the total number of actual positive instances. It measures how many of the actual positive instances were correctly predicted as positive.
* **Specificity:** This is also known as True Negative Rate (TNR). is the ratio of true negative predictions to the total number of actual negative instances. It measures how many of the actual negative instances were correctly predicted as negative.
* **Area Under the Receiver Operating Characteristic Curve** (AUC-ROC): It represents the area under the ROC curve, which plots the true positive rate against the false positive rate at various thresholds. It measures the model's ability to discriminate between positive and negative instances. It is an overall summary of diagnostic accuracy.

## 3.7 Deployment of Classification Model

When deploying a machine learning model for classification tasks, Streamlit is employed as the framework or platform to make the model accessible to end-users. Streamlit is a Python library that simplifies the process of creating interactive web applications for data science and machine learning projects. By using Streamlit, data scientists and developers can easily showcase their classification model to others, enabling users to interact with the model, input data, and view the model's predictions in a user-friendly web interface. This deployment method allows for quick and seamless sharing of the model's capabilities without the need for users to have expertise in coding or data science.

Streamlit is utilized to deploy a classification model for significant bacteriuria.

# CHAPTER FOUR

# RESULTS AND DISCUSSION

A machine learning algorithms were applied on the pre-processed dataset, to learn the underlying patterns and relationships between the selected features and the presence or absence of significant bacteriuria. This involved feeding the model with the selected features and the corresponding target labels.

The evaluation metrics of the classification model was analysed to gain insights into the model’s strengths and weaknesses, and to assess whether the model achieves satisfactory accuracy, recall, and AUC-ROC values.

The contributions of different variables in predicting significant bacteriuria was evaluated by calculating the feature importance scores using techniques specific to the model used.

The model evaluation helped to assess the effectiveness of the developed classification model for significant bacteriuria, identify potential areas for improvement, and possible applicability in real-world clinical settings.

# REFERENCES

Adegoke, S. A., & Adegun, P. T. (2013). Asymptomatic Bacteriuria in Nigerian Children with Sickle Cell Anaemia. *Indian Journal of Nephrology, 23*(2), 103-107.

Adjato, S., Amoako, E., Abaka-Yawson, A., Agbodzakey, H., & Tawiah, P. (2019). Bacterial Profile and Antimicrobial Sensitivity Patterns in Asymptomatic Bacteriuria: A Cross-sectional Study of Sickle Cell Disease Patients in the Ho Municipality, Ghana. *Asian Journal of Medicine and Health, 15*(2), 1-7.

Agrawal, R., Srivastava, J., Srivastava, A., & Arya, A. (2018). Asymptomatic Bacteriuria. *Indian Journal of Microbiology Research, 5*(2), 184-187.

Akinbami, A., Bode‑Shojobi, I., Oshinaike, O., Adediran, A., Osikomaiya, B., & Moronke, R. (2014). Prevalence of Significant Bacteriuria among Symptomatic and Asymptomatic Homozygous Sickle Cell Disease Patients in a Tertiary Hospital in Lagos, Nigeria. *Nigerian Journal of Clinical Practice, 17*(2), 163-167.

Al Benwan, K., & Jamal, W. (2022). Etiology and Antibiotic Susceptibility Patterns of Urinary Tract Infections in Children in a General Hospital in Kuwait: A 5-Year Retrospective Study. *Medical Principles and Practice, 31*, 562–569.

Alabi, M., & Adebowale-Tambe, N. (2021). *World SCD 2021: Nigeria, Global Epicentre of Sickle Cell, Not Doing Enough to Tackle Disease*. Available from The Premium Times, Nigeria: https://www.premiumtimesng.com/news/headlines/468767-world-scd-2021-nigeria-global-epicentre-of-sickle-cell-not-doing-enough-to-tackle-disease.html [Accessed November 9, 2022]

Alanazi, A. (2022). Using Machine Learning for Healthcare Challenges and Opportunities. *Informatics in Medicine Unlocked, 30*, 1-5.

Ali, S., Reid, M., Forrester, T., Roye-Green, K., & Cumming, V. (2006). Asymptomatic Bacteriuria in Sickle Cell Disease: A Cross-sectional Study. *BMC Infectious Disease, 6*(46), 1-6.

Asinobi, A., Fatunde, O., Brown, B., Osinusi, K., & Fasina, N. (2003). Urinary Tract Infection in Febrile Children with Sickle Cell Anaemia in Ibadan, Nigeria. *Annals of Tropical Paediatrics, 23*(2), 129-134.

Atarodi, A., & Atarodi, A. (2019). The Impact of Information Technology on Health. *Journal of Research & Health, 9*(3), 193-194.

Bachur, R. (2004). Pediatric Urinary Tract Infection. *Clinical Pediatric Emergency Medicine Journal, 5*, 28-36.

Bebars, G. M., Mostafa, A. N., Moness, H. M., & Abdel-Aziz, R. A. (2022). Assessment of Early Kidney Injury Caused by Asymptomatic Bacteriuria in Children with Type 1 Diabetes. *BMC Pediatrics, 22*(643), 1-9.

Biswas, A., Saran, I., & Wilson, P. F. (2021). Introduction to Supervised Machine Learning. *Kidney360, 2*, 878-880.

Bitsori, M., & Galanakis, E. (2012). Pediatric Urinary Tract Infections: Diagnosis and Treatment. *Expert Review of Anti-Infective Therapy, 10*(10), 1153–1164.

Boye, A., Dwomoh, F., Morna, M., Ephraim, R., & Essien-baidoo, S. (2016). Prevalence of Asymptomatic Bacteriuria in a Sub-population of Tertiary Female Students: Antimicrobial Resistance Patterns and Group-specific Risk Factors. *Journal of Medical and Biological Science Research, 2*(4), 56-64.

Brown, S., Imarenezor, E., Briska, J., Cletus, H., & Zakka, F. (2022). Antibiotic Resistance Pattern of Bacterial Isolates Associated with Asymptomatic Urinary Tract Infections in Wukari Metropolis, Taraba State, Nigeria. *Medical and Health Sciences European Journal, 6*(6), 8-14.

Brownlee, J. (2020). *4 Types of Classification Tasks in Machine Learning*. Available from MachineLearningMastery: https://machinelearningmastery.com/types-of-classification-in-machine-learning/ [Accessed November 9, 2022]

Brownlee, J. (2021). *A Gentle Introduction to Ensemble Learning Algorithms*. Available from MachineLearningMastery: https://machinelearningmastery.com/tour-of-ensemble-learning-algorithms/#:~:text=The [Accessed February 2, 2023]

Burton, R., Albur, M., Eberl, M., & Cuff, S. (2019). Using Artificial Intelligence to Reduce Diagnostic Workload without Compromising Detection of Urinary Tract Infections. *BMC Medical Informatics and Decision Making, 19*, 171-181.

Centers for Disease Control and Prevention (CDC). (2022). *Sickle Cell Disease (SCD)*. Available from CDC: https://www.cdc.gov/ncbddd/sicklecell/facts.html [Accessed November 8, 2022]

Choudhury, A. (2019). *Basics Of Ensemble Learning in Classification Techniques Explained*. Available from Analytics India Mag: https://analyticsindiamag.com/basics-of-ensemble-learning-in-classification-techniques-explained/#:~:text=The [Accessed February 2, 2023]

Chukwu, B. F., Okafor, H. U., & Ikefuna, A. N. (2011). Asymptomatic Bacteriuria in Children with Sickle Cell Anemia at The University of Nigeria Teaching Hospital, Enugu, South East, Nigeria. *Italian Journal of Pediatrics, 37*(45), 1-5.

Cichosz, P. (2011). Assessing the Quality of Classification Models: Performance Measures and Evaluation Procedures. *Central European Journal of Engineering, 1*(2), 132-158.

Colgan, R., Geerlings, S. E., Rice, J. C., Saint, S., Tambayh, P. A., Bradley, S., & Cardenas, D. D. (2010). Diagnosis, Prevention, and Treatment of Catheter-Associated Urinary Tract Infection in Adults. *Clinical Infectious Diseases, 50*(5), 625–663.

Cormican, M., Murphy, A. W., & Vellinga, A. (2011). Interpreting Asymptomatic Bacteriuria. *British Medical Journal, 343*, 1-5.

Crader, M. F., Kharsa, A., & Leslie, S. W. (2022). *Bacteriuria.* Available from NCBI Bookshelf: https://www.ncbi.nlm.nih.gov/books/NBK482276/?report=printable [Accessed January 26, 2023]

Dada, E., & Akharaiyi, F. (2012). Bacteriuria among Primary and Secondary School Pupils in Akure North Local Government Area of Ondo State, Nigeria. *Acta Clinica Croatica, 51*(3), 379-385.

Dahiya, A., & Goldman, R. D. (2018). Management of Asymptomatic Bacteriuria in Children. *Child Health Update, 64*, 821-822.

Dasgupta, A., & Nath, A. (2016). Classification of Machine Learning Algorithms. *International Journal of Innovative Research in Advanced Engineering, 3*(3), 6-11.

DeBaun, M. R., & Galadanci, N. A. (2022). *Sickle Cell Disease in sub-Saharan Africa*. Available from UpToDate: https://www.uptodate.com/contents/sickle-cell-disease-in-sub-saharan-africa [Accessed January 31, 2023]

Dietterich, T. G. (2000). *Ensemble Methods in Machine Learning.* Berlin: Springer Berlin Heidelberg.

Donkor, E. S., Osei, J. A., Anim-Baidoo, I., & Darkwah, S. (2017). Risk of Asymptomatic Bacteriuria among People with Sickle Cell Disease in Accra, Ghana. *Diseases, 5*(4), 1-8.

Emmerich, N. (2009). Tracking the Impact of Health Care Technology. *Metascience, 18*, 501-504.

Eriksen, S., Huynh, C., & Keller, R. (2001). *Decision Trees.* Amsterdam: Kluwer Academic Publishers.

Ezigbo, O. (2021). *Nigeria Has Highest Prevalent Rate of Sickle Cell Worldwide, Says FG*. Available from THISDAYLIVE: https://www.thisdaylive.com/index.php/2021/06/24/nigeria-has-highest-prevalent-rate-of-sickle-cell-worldwide-says-fg/ [Accessed January 31, 2023]

Friedman, J. H., Tibshirani, R., & Hastie, T. (2009). *The Elements of Statistical Learning* (2nd ed.). Berlin: Springer Science+Business Media.

Gopalan, B. (2020). *What is Entropy and Information Gain? How are they used to construct decision trees?* Available from Numpy Ninja: https://www.numpyninja.com/post/what-is-entropy-and-information-gain-how-are-they-used-to-construct-decision-trees [Accessed February 2, 2023]

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). San Francisco: Morgan Kaufmann Publishers.

Heckerling, P., Canaris, G., Flach, S., Tape, T., Wigton, R., & Gerber, B. (2007). Predictors of Urinary Tract Infection Based on Artificial Neural Networks and Genetic Algorithms. *International Journal of Medical Informatics, 76*, 289-296.

Hilbe, J. M. (2015). *Practical Guide to Logistic Regression.* Boca Raton: CRC Press.

Hinman, F. (1963). The Meaning of "Significant Bacteriuria". *Journal of the American Medical Association, 184*(9), 727-728.

Horton, N. (2020). *IPS\_09\_Ch14.* Available from Nicholas Horton Amherst College Website: https://nhorton.people.amherst.edu/ips9/IPS\_09\_Ch14.pdf [Accessed February 10, 2023]

Imam, T. H. (2021). *Bacterial Urinary Tract Infections*. Available from MSD Manual Professional Version: https://www.msdmanuals.com/professional/genitourinary-disorders/urinary-tract-infections-utis/bacterial-urinary-tract-infections?query=Asympto%E2%80%A6 [Accessed January 26, 2023]

Imam, T. H. (2022). *Asymptomatic Bacteriuria*. Available from MSD Manual Consumer Version: https://www.msdmanuals.com/home/kidney-and-urinary-tract-disorders/urinary-tract-infections-utis/bladder-infection [Accessed January 26, 2023]

Inyang-Etoh, P., Etefia, E., & Asuquo, O. (2021). Relationship between Enuresis and Asymptomatic Bacteriuria among Children in Ikot Ene Community, Akpabuyo, Cross River State, Nigeria. *International Journal of Research in Academic World, 1*(1), 5-9.

Jassim, M. A., & Abdulwahid, S. N. (2021). Data Mining Preparation: Process, Techniques and Major Issues in Data Analysis. *IOP Conference Series: Materials Science and Engineering, 1090*, 1-8.

Jeng, S.-L., Huang, Z.-J., Yang, D.-C., Teng, C.-H., & Wang, M.-C. (2022). Machine Learning to Predict the Development of Recurrent Urinary Tract Infection Related to Single Uropathogen, Escherichia coli. *Scientific Reports, 12*, 1-11.

Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Teechnology Trends, 2*(1), 20-28.

Jombo, G., Odey, F., Ibor, S., Bolarin, D., Ejezie, G., Enenebeaku, M., & Alao, O. (2010). Subclinical Significant Bacteriuria among Pre-school Children in Calabar Municipality: A Survey. *Journal of Medicine and Medical Sciences, 1*(4), 134-140.

Jordano, Q. (2022). Bacteriuria Revisited. *Medicina Clínica, 159*, 437–439.

Kaur, M., Dangi, C., & Singh, M. (2013). An Overview on Sickle Disease Profile. *Asian Journal of Pharmaceutical and Clinical Research, 6*(1), 25-37.

Kehinde, A. O., Adedapo, K. S., Aimaikhu, C. O., Odukogbe, A. A., Olayemi, O., & Salako, B. (2011). Significant Bacteriuria among Asymptomatic Antenatal Clinic Attendees in Ibadan, Nigeria. *Tropical Medicine and Health, 39*(3), 73–76.

Kondapaneni, S. L., Surpam, R., Azaruddin, M., & Devi, G. (2012). Screening for Asymptomatic Bacteriuria in School-going Children. *Indian Journal of Public Health, 56*(2), 169-170.

Kozma, L. (2008). k Nearest Neighbors algorithm (kNN). *Special Course in Computer and Information Science*. Helsinki: Helsinki University of Technology.

Kramer, O. (2013). *Dimensionality Reduction with Unsupervised Nearest Neighbours.* Berlin: Springer Berlin Heidelberg.

Krishnan, S. (2021). *Decision Tree for Classification, Entropy, and Information Gain*. Available from Medium: https://medium.com/codex/decision-tree-for-classification-entropy-and-information-gain-cd9f99a26e0d#:~:text=Entropy [Accessed February 2, 2023]

Lee, W.-M. (2019). *Python Machine Learning* (1st ed.). Indianapolis: John Wiley & Sons, Inc.

Makhlysheva, A., Bakkevoll, P. A., Nordsletta, A. T., & Linstad, L. (2018). Artificial Intelligence and Machine Learning in Healthcare. *Norwegian Centre for E-health Research Fact Sheet, 9*, 1-2.

Mancini, A., Vito, L., Marcelli, E., Piangerelli, M., De Leone, R., Pucciarelli, S., & Merelli, E. (2020). Machine Learning Models Predicting Multidrug Resistant Urinary Tract Infections using “DsaaS”. *BMC Bioinformatics, 21*(10), 1-12.

Manuel, M., Patel, V., Filkins, L., Jason, P., & Nadeem, S. (2022). Urinary Predictors of Bacteremia in Febrile Infants with Urinary Tract Infection. *Journal of Scientific Innovation in Medicine, 5*(1), 1-9.

Mava, Y., Ambe, J. P., Bello, M., Watila, I., & Nottidge, V. A. (2011). Urinary Tract Infection in Febrile Children with Sickle Cell Anaemia. *West African Journal of Medicine, 30*(4), 268–272.

Mava, Y., Timothy, S., Baba, U., Garba, M., & Pius, S. (2012). Association Between Urinary Tract Infection and Other Common Febrile Illnesses of Childhood among Children with Sickle Cell Anaemia. *Indo-Global Research Journal of Pharmaceutical Sciences, 2*(2), 260-263.

Meherwar, F., & Maruf, P. (2017). Survey of Machine Learning Algorithms for Disease Diagnostic. *Journal of Intelligent Learning Systems and Applications, 9*, 1-16.

Mitiku, A., Aklilu, A., Tsalla, T., Woldemariam, M., Manilal, A., & Biru, M. (2022). Magnitude and Antimicrobial Susceptibility Profiles of Gram-Negative Bacterial Isolates among Patients Suspected of Urinary Tract Infections in Arba Minch General Hospital, Southern Ethiopia. *PLoS ONE, 17*(12), 1-9.

Munaku, H., Ambrose, E., Kidenya, B., Seni, J., Kayange, N., & Mshana, S. (2022). *Bacteremia, Bacteriuria and their Associated Factors among Children with Sickle Cell Anaemia at the Bugando Medical Centre, Mwanza, Tanzania.* Available from SSRN Electronic Journal: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=4076835 [Accessed January 16, 2023]

Nasteski, V. (2017). An Overview of the Supervised Machine Learning Methods. *Horizons in Computer Science Research, 15*, 1-10.

Nicolle, L. E., Bradley, S., Colgan, R., Rice, J., Schaeffer, A., & Hooton, T. M. (2005). Infectious Diseases Society of America Guidelines for the Diagnosis and Treatment of Asymptomatic Bacteriuria in Adults. *Clinical Infectious Diseases, 40*(5), 643–654.

Nwabuko, O. C., Onwuchekwa, U., & Iheji, O. (2022). An Overview of Sickle Cell Disease from the Socio-demographic Triangle - a Nigerian Single-institution Retrospective Study. *Pan African Medical Journal, 41*(161), 3-24.

Odoki, M., Bazira, J., Tibyangye, J., Maniga, J., Agwu, E., & Wampande, E. (2019). Prevalence of Bacterial Urinary Tract Infections and Associated Factors among Patients Attending Hospitals in Bushenyi District, Uganda. *International Journal of Microbiology, 2019*(1), 1-9.

Okafor, H., Okoro, B., Ibe, B., & Njoku-Obi, N. (1993). Prevalence of Asymptomatic Bacteriuria among Nursery School Children. *Nigerian Journal of Paediatrics, 20*(4), 84-88.

Olowu, W. A., & Oyetunji, T. G. (2003). Nosocomial Significant Bacteriuria: Prevalence and Pattern of Bacterial Pathogens among Children Hospitalized for Non-infective Urinary Tract Disorders. *West African Journal of Medicine, 22*(1), 72-75.

Osisanwo, F. Y., Akinsola, J. E., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology, 48*(3), 128-138.

Ozkan, I., Koklu, M., & Sert, I. (2018). Diagnosis of Urinary Tract Infection Based on Artificial Intelligence Methods. *Computer Methods and Programs in Biomedicine, 166*, 51-59.

Panch, T., Szolovits, P., & Atun, R. (2018). Artificial Intelligence, Machine Learning and Health Systems. *Viewpoints: The FEBS Journal, 8*(2), 1-8.

Park, H.-A. (2013). An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain. *Journal of Korean Academy of Nursing, 43*(2), 154-164.

Patel, N., Farooqi, A., Callaghan, M., & Sethuraman, U. (2020). Urinary Tract Infection in Febrile Children with Sickle Cell Disease Who Present to the Emergency Room with Fever. *Journal of Clinical Medicine, 9*, 1-8.

Pezzlo, M. T. (1983). Automated Methods for Detection of Bacteriuria. *The American Journal of Medicine, 75*(1), 71-77.

Rahn, D. D., Boreham, M. K., Allen, K. E., Nihira, M. A., & Schaffer, J. I. (2005). Predicting Bacteriuria in Urogynecology Patients. *American Journal of Obstetrics and Gynecology, 192*, 1376–1378.

Ramageri, B. M. (2010). Data Mining Techniques and Applications. *Indian Journal of Computer Science and Engineering, 1*(4), 301-305.

Rokach, L., & Maimon, O. (2015). *Data Mining with Decision Trees: Theory and Applications* (2nd ed.). Singapore: World Scientific Publishing.

Roupakias, S., Sinopidis, X., Karatza, A., & Varvarigou, A. (2013). Predictive Risk Factors in Childhood Urinary Tract Infection, Vesicoureteral Reflux, and Renal Scarring Management. *Clinical Pediatrics, 20*(10), 1-15.

Seenappa, N., Sinha, M., Prasad, T., & Krishnamoorthy, V. (2022). An Analysis of Bacteriuria Rates after Endourological Procedures. *International Journal of Urological Nursing, 17*(1), 45-49.

Shah, S., Jafri, R., Mobin, K., Nanji, K., Jahangir, F., Iftikhar, H., & Maqbool, H. S. (2018). Frequency and Features of Nocturnal Enuresis in Pakistani Children aged 5 to 16 years Based on ICCS Criteria: a Multi-center Cross-sectional Study from Karachi, Pakistan. *BMC Family Practice, 19*(198), 1-9.

Shaikh, N., & Craig, J. C. (2014). Identification of Children and Adolescents at Risk for Renal Scarring After a First Urinary Tract Infection. *JAMA Pediatrics, 168*(10), 893-900.

Shute, V. J., Fulwider, C. G., Liu, Z., & Rahimi, S. (2023). Machine Learning. *International Encyclopedia of Education, 6*, 83-91.

Silva, A., Oliveira, E. H., & Mak, R. (2020). Urinary Tract Infection in Pediatrics: An Overview. *Jornal de Pediatra, 96*(S1), 65-79.

Singh, N., & Chaurasia, P. K. (2020). Machine Learning in Healthcare Industry: Tools and Techniques. *Test Engineering and Management, 82*, 4146- 4150.

Sperandei, S. (2014). Understanding Logistic Regression Analysis. *Biochemia Medica, 24*(1), 12-18.

Stein, R., Dogan, H., Hoebeke, P., Kocvara, R., Nijman, R., Radmayr, C., & Tekgul, S. (2015). Urinary Tract Infections in Children: EAU/ESPU Guidelines. *European Urology, 67*, 546-558.

Steinbach, M., & Tan, P.-N. (2009). *kNN: k-Nearest Neighbors.* Milton Park: Taylor & Francis Group.

Taylor, R., Moore, C., Cheung, K., & Brandt, C. (2018). Predicting Urinary Tract Infections in the Emergency Department with Machine Learning. *PLoS ONE, 13*(3), 1-15.

The National Heart, Lung, and Blood Institute (NHLBI). (2022). *What Is Sickle Cell Disease?* Available from NHLBI, NIH: https://www.nhlbi.nih.gov/health/sickle-cell-disease [Accessed November 8, 2022]

Tripathi, P., Jatana, V., Hari, P., Kapil, A., Srivastava, R. N., Bhan, M. K., & Bagga, A. (2003). Bacteriuria and Urinary Tract Infections in Malnourished Children. *Pediatric Nephrology, 18*, 366–370.

Turan, S. C., & Cengiz, M. A. (2022). Ensemble Learning Algorithms. *Journal of Science and Arts, 22*(2), 459-470.

United Nations. (2009). Resolution Adopted by the General Assembly on 22 December 2008. *General Assembly Sixty-third Session A/RES/63/237.*

Urology Care Foundation. (2023). *Urinary Tract Infections in Children*. Available from Urology Health: https://www.urologyhealth.org/urology-a-z/u/urinary-tract-infections-in-children [Accessed January 16, 2023]

Uwaezuoke, S. N. (2016). The Prevalence of Urinary Tract Infection in Children with Severe Acute Malnutrition: A Narrative Review. *Pediatric Health, Medicine and Therapeutics, 7*, 121–127.

Waseem, M. (2022). *How to Implement Classification in Machine Learning*. Available from Edureka: https://www.edureka.co/blog/classification-in-machine-learning/ [Accessed November 8, 2022]

Wasilewska, A. (2022). *Decision Tree Classification Basic Algorithm.* Available from StonyBrook Education: https://www3.cs.stonybrook.edu/~cse352/L8DTIntro.pdf [Accessed February 2, 2023]

Webb, G. I., & Zheng, Z. (2004). Multi-Strategy Ensemble Learning: Reducing Error by Combining Ensemble Learning Techniques. *IEEE Transactions on Knowledge and Data Engineering, 16*(8), 980-991.

Yauba, M., Aikhionbare, H., Ogunrinde, G., & Bugaje, M. (2014). Significant Bacteriuria in Children with Sickle Cell Anaemia in a Nigerian Tertiary Hospital. *Nigerian Journal of Paediatrics, 41*(4), 350-353.

Zajac, M. (2022). *Sickle Cell Disease: Statistics and Facts*. Available from Sickle-Cell: https://sickle-cell.com/statistics [Accessed November 9, 2022]

Zeng, Z., Zhan, J., Kaimin, Z., Chen, H., & Cheng, S. (2019). Global, Regional, and National Burden of Urinary Tract Infections from 1990-2019: an Analysis of the Global Burden of Disease Study 2019. *World Journal of Urology, 40*(1), 1-9.

Zhang, Y., Liu, J., & Shen, W. (2022). A Review of Ensemble Learning Algorithms Used in Remote Sensing Applications. *Applied Sciences, 12*, 1-19.

Zorc, J. J., Kiddoo, D. A., & Shaw, K. N. (2005). Diagnosis and Management of Pediatric Urinary Tract Infections. *Clinical Microbiology Reviews, 18*(2), 417-422.

Zulfiqar, A., Zulfiqar, S., & Rahat, S. (2018). Asymptomatic Bacteriuria. *The Professional Medical Journal, 25*(1), 115-118.