Dialogue Social Science Review (DSSR)
www.thedssr.com

ISSN Online: 3007-3154
ISSN Print: 3007-3146

DIALOGUE SOCIAL SCIENCE REVIEW

Vol. 3 No. 8 (August) (2025)

# Risk Factors and Prevalence of Malaria in District Malakand: A Statistical Analysis Using Logistic Regression and K-Means Clustering

**Fazal Shakoor**
Department of Statistics, University of Peshawar. fazalstate47@gmail.com

**Muhammad Ismail**
Department of Statistics, University of Peshawar. Ismailfromdir@gmail.com

**Kifayat Ullah**
Department of Statistics, Quaid-i-Azam University Islamabad. kifayatshangra@gmail.com

**Syed Ishtiaq Ahmad**
In charge of the Intensive Care Unit, MERF NGO Hospital, Kabul, Afghanistan. syedishtiaq447@gmail.com

**Sehran Hassan**
Department of Statistics, University of Peshawar. sehranhassan9078@gmail.com

**Zahid Khan**
Lecturer, Department of Statistics, University of Malakand. zahid.uom1@gmail.com

## Abstract

Malaria continues to be a significant public health burden in tropical regions such as District Malakand, Pakistan, where *Plasmodium falciparum* and *P. vivax* are endemic. This study employs binary logistic regression and K-means clustering to identify and analyze key risk factors associated with malaria occurrence. Logistic regression results reveal that variables such as gender ($p = 0.016$, OR = 5.03), type of malaria ($p = 0.010$, OR = 56.15), symptoms ($p = 0.041$, OR = 34.19), first appearance of malaria ($p = 0.011$, OR = 121.75), and route of transmission ($p = 0.001$, OR = 32.62) are statistically significant predictors. Additionally, environmental factors ($p = 0.001$, OR = 32.62) and the presence of water and toilet facilities ($p = 0.017$, OR = 0.002) play a crucial role in disease prevalence. Results from a separate model also indicate a significant negative relationship between age and malaria risk ($p = 0.048$), with older individuals being less susceptible, while recurrence of malaria is highly predictive of future cases ($p < 0.001$). To uncover underlying patterns in the dataset, K-means clustering was applied. The optimal number of clusters was determined using the Elbow method, Silhouette analysis, and the Gap Statistic, which collectively assessed within-cluster compactness and between-cluster separation. These methods confirmed distinct groupings based on key risk attributes, supporting the identification of high-risk populations. Overall, the study underscores the need for targeted interventions, improved public health infrastructure, and enhanced awareness programs to mitigate malaria transmission in District Malakand.

**Key Words**: Malaria prevalence, Risk factors, Binary logistic regression, Odds Ratio, K-Mean Clustering.

## Introduction

Malaria is a mosquito-borne infectious disease affecting humans and other animals caused by parasitic protozoan belonging to the Plasmodium type [1]. It is transmitted to humans through the bite of the Anopheles mosquito, which is a vector borne contagious disease that is wide spread in tropical and sub-tropical area's parasites the blood of infected are taken through the stomach of the mosquito as it feeds. Once an infected mosquito bites a human, the parasites multiply in the host's liver before infecting and destroying red blood cells. Malaria was eliminated from the U.S. in the early 1950s, but between 1,500 and 2,000 cases still occur each year, mostly in those who have recently traveled to malaria-endemic areas [2]. There are five parasite species (Plasmodium falciparum, P. vivax, P. Ovale, P. Malariae and as recently discovered, P. knowlesi) are responsible to cause malaria in human beings [3,4,5].

The most famous symptoms of malaria are the hyperthermia, RBCs loss, and if left unsorted turn to cerebral malaria [6]. Most cases (80%) and deaths (90%) reveal in Africa, and almost deaths (77%) were in children of five years of age. It is continuous to be a danger in the developing countries like Pakistan with more than 3000 deaths every day. It adopts an estimated 300 million people and causes more than a million deaths per annum world-wide [7]. Actually, if the prevalence of malaria delay on its present up ward course. The death rate may be double in next 21 year. Malaria is more prevalent in rural areas compared to cities. However, in Africa, the disease is found in both rural and urban areas, although the risk is generally lower in larger cities [8]. The aim of this study was to determine the prevalence of malaria in the human population of District Malakand. The two most common species identified in this region were Plasmodium falciparum and Plasmodium vivax.

## Historical Background

Malaria is one of the best-known epidemics of 20th century. After World War II, spread like a blast, resulted in great mortalities and morbidities. Currently, WHO is keenly taken the problem and start annual reporting of annual approximate mortality and incidence rates all over the world.

According to latest WHO annual report of malaria [9] globally, an estimated 3.4 billion people are at risk of malaria. The term malaria comes from Italian word such as mala it means bad; aria means air so this disease was formerly called is “argue" or marsh fever due to its association with swamp [10]. Malaria infected humans being over 50,000 years old and is a human pathogen. Close relatives of human malarial parasites remain common in chimpanzee [11]. A unique periodic fever of malaria is found throughout recorded history, beginning in 2700 B.C in China during “xia dynasty” [12]. Scientific studies in 1880 on malaria made their first significant advance, “working when French army Doctor named as “Charles Louis alps house Lavern” working in Algeria, he observed parasites RBC’S of infected people, and so he proposed that malaria is caused by a protozoon called as Plasmodium.

## Lifecycle of the Malaria-Causing Parasite

Mosquito bites, it transmits the infection by releasing sporozoites into the bloodstream. These sporozoites travel to the liver, where they invade liver cells and multiply, eventually transforming into merozoites. The liver cells then burst, releasing the merozoites back into the bloodstream. The merozoites attack red blood cells, undergoing stages of development as ring forms, trophozoites, and schizonts, which generate additional merozoites. Sexual forms of the parasite also emerge, and if a mosquito feeds on an infected individual, these forms are taken up, continuing the parasite's life cycle within the mosquito [13].

Normally, mosquito life is about from two to four week. Male mosquito dies quickly. Female mosquito lays 100 to 250 eggs at a time. And these eggs convert to pupa within two days and then converted to larva in five to seven days after this become adult mosquito [14].

## Types of Malarial Parasites

Five types of Plasmodium parasite can infect humans. They are found on different parts of the world. Plasmodium falciparum is the most deadly and widespread species responsible for the majority of malaria-related deaths globally. Plasmodium vivax follows as the second most significant species, while Plasmodium Malariae accounts for only a small percentage of infections. Plasmodium Ovale poses additional challenges due to its ability to remain dormant in the liver and reactivate without a mosquito bite. Plasmodium knowlesi, typically a parasite of primates, can also infect humans, though the transmission method is not fully understood. Overall, Plasmodium falciparum remains the most critical, causing the majority of infections and nearly 90% of malaria-related deaths annually [15].

## Causes of Malaria

Malaria parasites belong to the genus' Plasmodium. In humans, malaria is caused by P-falciparum, P-Malariae, P-Ovale, P-vivax and P-Knowlesi. Among those infected, P. falciparum is the most common species identified 75% followed by P. vivax 20%. Other fewer common routes of infection are through blood transfusion or maternal-fetal transmission [16].

## Habits of Malarial Parasites

Only female mosquito is the ability to bite. Normally, a female mosquito can bite in evening as well as night. They live mostly in shady place. Anopheles like to lay an egg in clean water. It can travel from one place to another place about 11 KM. Mosquito can be transformed from one country to another by plane. Mosquito cannot live in hot or cold season. It can live about from 34 to 38 days. Anopheles these mosquitos had the ability to spread malaria [17].

## Controlling Measurement Adult Mosquito

To finish the larva, we should finish the old places and water log area. To get rid from larva, we should spray kerosene oil and after chemical such as Paris's grin, about, due to this chemical finish the larva marathon, and laden sprays are used to control adult mosquito. Dichloride diphenyl tricolor ethane is also used against mosquito, but it is an environmental hazard so it is internationally banned by the World Health Organization [18].

## Drugs and Skin Protection of Malaria

Several drugs are effective in treating malaria. Chloroquine is both affordable and highly effective, with Amodiaquine serving as an alternative. Quinine is typically administered in 500 mg doses three times daily for five days. Fansidar is given as a single-dose treatment consisting of three tablets. For preventive measures, Doxycycline is recommended, with one capsule taken two days before travel. Mefloquine is taken once a week and continued for 30 days after returning, while Malarone is taken daily for seven days post-trip. Additionally, using products containing DEET on the skin and sleeping under protective bed nets are recommended for disease prevention [19].

## Worldwide Risk of Malaria

Malaria is one of the biggest killers in the whole world. It infects up to 250 Million and kills 800,000 people every year. Forty percent of the world's population lives in endemic areas. Epidemics have devastated large populations, and malaria poses a serious barrier to economic progress in many developing countries. There are an estimated 300-500 million cases of clinical disease per annum with 1.5-2.7 million deaths [20].

## Malaria in Pakistan

Malaria is a major public health problem in Pakistan. According to 1998 censuses, the Population of Pakistan is 184.5 million. 1.6 million cases occur per annum in Pakistan. Pakistan is categorized in group three (3). Malaria is unstable in Pakistan almost all age; group are at risk of acquiring the infection. However, children under 5 years of age and pregnant women are at high risk of infection. In Pakistan, malaria is seasonal diseases, with peak around September for vivax and for falciparum malaria. Two types are very common in Pakistan the first one Plasmodium vivax, which found 70 %, and the 2nd is Plasmodium falciparum, which found 20% in Pakistani population [21].

## Malaria in Khyber Pakhtunkhwa

In 2023, the population of Khyber Pakhtunkhwa was approximately 40.85 million. Annually, around 0.5 million malaria cases were examined in the region. In 2022, there were 99,301 confirmed malaria cases, with 57,402 testing positive for Plasmodium vivax and 5,804 cases associated with the more severe Plasmodium falciparum, which poses a higher risk of infection. Children under five, those over five, and pregnant women are particularly vulnerable to the disease. Malaria in Khyber Pakhtunkhwa is seasonal, with Plasmodium vivax cases peaking around September, similar to the pattern observed for Plasmodium falciparum [22].

## Population of District Malakand

The 2023 census reports that the district of Malakand has an updated population and spans roughly 952 square kilometers. The male population is 4,505,112, while the female population is 4,433,415. This study aims to evaluate the malaria risk in Malakand, examine the differences in malaria incidence across various areas within the district, formulate effective malaria prevention and control policies, and raise public awareness on preventing the disease [22].

## Literature Review

Yasinzai et al conducted a study on malaria prevalence in District Ziarat, Sanjawa, from July 2014 to June 2016. They tested 3,765 blood samples, finding that 26.7% were positive for malaria. Among the cases, Plasmodium falciparum was responsible for 69.5%, while Plasmodium vivax accounted for 30.2%. The researchers used two methods: Active Case Detection (ACD), where medical specialists conducted door-to-door visits to collect blood samples, and Passive Case Detection (PCD), which focused on individuals with malaria symptoms. The study included participants aged one to 21 years and older from four areas in Ziarat, noting variations in malaria prevalence due to different environmental and hygienic conditions [23].

Jennifer et al conducted a study on the prevalence of malaria parasites in the interpersonal relationships between adults and pregnant women in Nigeria. The study included seven pregnant women and involved collecting blood samples from malaria patients to compare data on sex, age, and malaria species. Out of 1,126 participants, 42.7% were infected with malaria. The infection rates were 42.7% for Plasmodium falciparum and 0.9% for Plasmodium malariae, highlighting a significant difference between these parasites. The study also found that 77.8% of those infected did not use cell phones, and 34.3% of the infected were women, compared to men [24].

The study selected well-organized hospitals with trained staff to participate, representing most government hospitals in the regions. Conducted from February 16, 2002, to February 15, 2003, with an additional site added from August 15, 2002, the research focused on nine hospitals. At the busiest locations, research staff used Giemsa stains to examine blood slides and count Plasmodium falciparum parasites. Slides were deemed negative if no parasites were detected after scanning 100 high-power fields. Discordant results were reviewed a third time, with the majority result determining positivity or negativity and geometric mean density used for parasite quantification. Analyses were based on the research slide results unless otherwise noted [25].

Microscopy revealed that 46.1% of those treated for malaria were positive for Plasmodium falciparum, with positivity rates decreasing with age and altitude. Younger children in lower-altitude, high-transmission areas had higher positivity rates. Among those who tested negative, 7.6% died, with a higher fatality rate in negative cases (12.1%) compared to positive ones (6.9%). Respiratory distress and altered consciousness were key mortality predictors across all cases [26].

Ahmed et al. examined the social behavior and control programs related to malaria in Bangladesh using a two-stage cluster sampling technique. They randomly selected 30 samples from each district and collected data through face-to-face interviews conducted by trained interviewers. The study found a prevalence rate of 3.1% across 13 districts, with 2.73% of cases due to Plasmodium falciparum, 0.16% due to Plasmodium vivax, and 0.19% mixed infections. The research highlighted a lack of comprehensive knowledge about malaria, which is influenced by education levels and poverty [27].

Odugbemi et al investigated the increasing resistance of malaria parasites to chloroquine in Nigeria and explored the potential of indigenous medicinal plants for malaria treatment. They focused on plants used in Okeigbo, Ondo State, and identified 50 species, such as Morinda lucida (Oruwo) and Enantia chlorantha (Awopa), commonly used for malaria therapy. The study found that plant parts like barks, roots, and leaves are used in various combinations for treatment. This research highlights the potential of these plants to possess anti-malarial properties that could contribute to developing affordable malaria treatments from local resources in Nigeria [28].

Chuma et al conducted a study in impoverished areas of four malaria-endemic districts in Kenya using various data collection methods: a cross-sectional survey of 708 households, 24 focus group discussions, semi-structured interviews with 34 health workers, and patient exit interviews with 359 participants. The study found that affordability was a significant issue, with about 40% of individuals using shop-bought drugs and 42% visiting formal health facilities reporting insufficient funds for treatment. Many had to resort to borrowing money or obtaining treatment on credit. Additional factors affecting affordability included the seasonality of illness, income sources, transport costs, and unofficial costs [29].

Adimi et al studied malaria in Afghanistan, where about 60% of the population lives in endemic areas. They combined provincial malaria data from 2004-2007 with NASA satellite observations to model malaria prevalence. Surface temperature was identified as a strong predictor, while precipitation was not significantly associated with malaria incidence. The study achieved a provincial average $R^2$ of 0.845 for modeling malaria cases, with a prediction accuracy within 8.9% of actual cases. The findings support the development of a cost-effective surveillance system as part of the WHO EMRO Malaria Control and Elimination Plan [30].

Mumtaz et al. investigated the role of platelets in malaria pathogenesis, particularly their involvement in sequestrating infected red blood cells and providing innate protection against infections. The study, involving patients aged 16 to 80 years, found that 85.5% of malaria patients had low platelet counts, with 70.5% experiencing mild thrombocytopenia. Of the malaria cases, 52.5% were caused by Plasmodium falciparum, 46.5% by Plasmodium vivax, and 1% by Plasmodium Malariae. These findings align with previous studies in Pakistan, highlighting the common association of malaria with varying degrees of thrombocytopenia [31].

Rekha et al studied malaria in Ranchi, Jharkhand, India, highlighting the impact of rapid industrialization, population growth, and inadequate health infrastructure on the spread of the disease. They conducted a GIS-based retrospective study across 328 sub-centers in 14 primary health centers to analyze the spatial distribution of Plasmodium vivax and Plasmodium falciparum. Although chemical insecticides have been effective for malaria vector control, the rise of insecticide resistance has made these methods less sustainable. The study suggests considering biological control agents, such as entomopathogenic fungi, as alternatives for managing adult malaria vectors [32].

Singh et al evaluated the performance of various rapid diagnostic tests (RDTs) for malaria in tribal areas of central India, comparing them with traditional and molecular methods. They tested five commercially available RDTs alongside peripheral blood smears and polymerase chain reaction (PCR). The study found that the First Response Malaria Antigen combo test was highly sensitive (94.7%) for Plasmodium falciparum and specific (96.5%) for Plasmodium vivax, outperforming other RDTs. This suggests that RDTs, particularly the First Response test, are effective tools for malaria diagnosis and could significantly aid in malaria management [33].

## Methodology

This study was carried out in District Malakand, located in the Khyber Pakhtunkhwa province of Pakistan. District Malakand is divided into two Tehsils: Dargai and Batkhela. These Tehsils encompass 28 union councils and 26 neighborhood councils. The distribution of the union councils is detailed below.

**Table 1: List of Union Council of District Malakand**

| | | | |
|---|---|---|---|
| Middle Batkhela | Thana khas | Khar | Sakhakot Khas |
| Upper Batkhela | Thana Bandijat | Agra | Sakhakot Bandijat |
| Lower Batkhela | Thana Jaded | Wartair | Badragah |
| Totakan | Heroshah | Gu Khel | Dheri Alladand |
| Kot | Selaipattai | Dargai | Palai |
| Kopar | Julagram | Kharkai | |
| Pirkhel | Malakand | Meherdi | |

## Data Collection Method

Data was collected using a questionnaire. Primary data was gathered for the analysis, with 220 mixed questionnaires distributed across various union councils in District Malakand over the last three weeks of May to July 2023. A two-stage cluster sampling technique was employed in this cross-sectional study. A total of 220 samples were randomly selected from 10 different union councils, representing both male and female populations across different age groups (less than 60 years). The data collection was conducted by trained investigators.

## Binary Logistic Regression

Binary logistic regression is used when the response variable has two possible outcomes, typically coded as 0 or 1. Examples include pass/fail, heads/tails, or alive/dead. This method is suitable for binary outcomes because classical regression is designed for continuous variables and cannot handle categorical outcomes directly. Instead, binary logistic regression models the probability of one outcome occurring versus the other, making it ideal for analyzing scenarios where the response variable is categorical with two distinct categories.

$$ln[\pi(x)] = ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x \qquad (1)$$

In Equation (1), the natural logarithm of the odds of the response variable is used. The odds ratio is defined as the ratio of the probability of the event of interest, when the response is 1, to the probability of the event not occurring, when the response is 0. Specifically, when the response is 1, the probability of success is denoted by $\pi(x)$, and when the response is 0, the probability of failure is 1 minus $\pi(x)$. In this model, $\beta_1$ represents the regression coefficient for the predictor variable, and $\beta_0$ is the intercept term. The predictor variable x can be either continuous or categorical, while the response variable Y is categorical [34].

**Table: 2 Model Summary**

| Step | -2 Log likelihood | Cox & Snell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | $.000^a$ | .623 | 1.000 |

The model summary results indicate an extremely high fit of the model to the data. The -2 Log Likelihood value of 0.000 suggests that the model perfectly predicts the outcome, which is unusual and often indicative of overfitting. The Cox & Snell

R Square value of 0.623 shows that the model explains a significant portion of the variance in the dependent variable, though this measure is bounded below 1. The Nagelkerke R Square value of 1.000 suggests that the model accounts for all the variance in the outcome variable, further supporting the idea of a perfect or near-perfect fit. These results are atypical and may indicate potential overfitting or data issues, as such a perfect fit is rare in real-world data.

### Diagnostic

Diagnostic refers to a set of procedures used to evaluate how well a regression model fits a particular population based on a sample of data. Various statistical tools are employed to diagnose different models, offering quantitative measures of the model's goodness of fit. Additionally, these tools provide insights into the significance of individual covariates within the model.

### Clustering Algorithm

Clustering is an unsupervised machine learning technique used to group data points into meaningful clusters based on their similarity, without the need for pre-labeled classes. Among the various clustering methods, K-means clustering is one of the most widely used due to its simplicity and efficiency. It partitions a dataset into $K$ clusters by minimizing the sum of squared distances between data points and their corresponding cluster centroids. For example, in customer segmentation, K-means can help identify distinct groups of customers based on purchasing behavior, which can then be targeted with tailored marketing strategies. This algorithm is particularly useful in exploratory data analysis where the goal is to uncover hidden patterns in data [35].

### K-Mean Clustering

K-means clustering is an unsupervised machine learning technique commonly used for partitioning a dataset into a predefined number of mutually exclusive clusters. It aims to minimize within-cluster variability while maximizing between-cluster differences. In the present study, K-means clustering was applied to classify the observations into distinct groups based on similarities in key risk factors associated with malaria prevalence. This method was selected due to its simplicity, scalability, and efficiency in handling large datasets [36]. The optimal number of clusters was determined using the Elbow method, implemented via the fviz_nbclust() function from the factoextra R package (Kassambara & Mundt, 2020), which evaluates total within-cluster sum of squares across a range of cluster solutions. The clustering results provide valuable insights into patterns of malaria risk, enabling targeted public health interventions [37].

### Results And Discussion

In this study, we analyze the malaria prevalence among individuals in Tehsil Batkhela and Dargai, presenting our findings in the following Table:3 and then through various others tables. Our study included 220 participants aged under 60, with 44 females and 176 males. The overall malaria prevalence was 19%, with a higher rate in females (27.27%) compared to males (17.04%), indicating a greater spread among the male population. Using binary logistic regression with Minitab 16, we identified significant factors such as age, first occurrence of malaria, transmission route, family history, and recurrence. Our analysis revealed that out of 176 males and 44 females, 30 males (71.4%) and 12 females (28.6%) tested positive for malaria, while 146 males (82.0%) and 23 females (18.0%) tested negative. The high prevalence among males suggests they are more affected by malaria in District Malakand. The first appearance of malaria was associated with a 54.8% prevalence rate, and a positive family history of malaria was linked to a 33.3% prevalence rate. Transmission routes included mosquitoes (73.8%), dirty water (26.0%), and no cases from polluted food. Additionally, 19 individuals (45.2%) experienced recurrent malaria. Detailed results are presented in the following table.

***Table:3 Distribution of Malaria and Non-Malaria Patients Across Various Demographic and Clinical Variables***

| **Variables** | **Factors** | **Malaria patient** | **Non-Malaria patient** | **Total** |
|---|---|---|---|---|
| Gender | Male | 30 (71.4%) | 146 (82.0%) | 176 |
| | Female | 12 (28.6%) | 32 (18.0%) | 44 |
| First appearance of malaria | Yes | 23 (54.8%) | 0 (0.0%) | 23 |
| | No | 19 (45.2%) | 178 (100%) | 197 |
| Route of transmission | Mosquitoes | 31 (73.8%) | 41 (23.0%) | 72 |
| | Dirty water | 11 (26%) | 24 (13.5%) | 35 |
| | Polluted food | 0 (0.0%) | 4 (2.2%) | 4 |
| | Do not know | 0 (0.0%) | 109 (61.2%) | 109 |
| Malaria in family | Yes | 14 (33.3%) | 10 (5.6%) | 24 |
| | No | 28 (66.7%) | 168 (94.4%) | 196 |
| Malaria again | Yes | 19 (45.2%) | 0 (0.0%) | 19 |
| | No | 23 (54.8%) | 178 (100%) | 201 |
| Age | 1 to 10 | 15 (37.5%) | 25 (62.5%) | 40 |
| | 11 to 20 | 16 (26.6%) | 44 (73.3%) | 60 |
| | 21 to 30 | 6 (9.09%) | 60 (90.90%) | 66 |
| | 31 to 40 | 3 (10.34%) | 26 (89.6%) | 29 |
| | 41 to 50 | 2 (11.11%) | 16 (88.88%) | 18 |
| | 51 to over | 0 (0.0%) | 7 (100%) | 7 |

***Table 4: Route of Transmission of Malaria***

| Route of Transmission | Malaria Patients | Non-Malaria Patients | Total |
|---|---|---|---|
| Mosquitoes | 31 (73.8%) | 41 (23.0%) | 72 |
| Dirty Water | 11 (26.2%) | 24 (13.5%) | 35 |
| Polluted Food | 0 (0.0%) | 4 (2.2%) | 4 |
| Do Not Know | 0 (0.0%) | 109 (61.2%) | 109 |

The data in Table 4 reveals that a significant majority (73.8%) of malaria patients correctly identified mosquitoes as the primary route of transmission, while a smaller proportion (26.2%) mistakenly believed that dirty water was the cause. Interestingly, a large percentage (61.2%) of non-malaria patients were unsure about how malaria is transmitted, indicating a considerable gap in public awareness. This suggests that while those affected by malaria generally have some understanding of its transmission, there is a widespread need for public education to address misconceptions and improve knowledge about the true cause-mosquito bites.

***Table: 5 Age Distribution of Malaria Patients Compared to Non-Malaria Patients"***

| Age Group | Malaria Patients | Non-Malaria Patients | Total |
|---|---|---|---|
| 1 to 10 | 15 (37.5%) | 25 (62.5%) | 40 |
| 11 to 20 | 16 (26.6%) | 44 (73.3%) | 60 |
| 21 to 30 | 6 (9.09%) | 60 (90.9%) | 66 |
| 31 to 40 | 3 (10.34%) | 26 (89.6%) | 29 |
| 41 to 50 | 2 (11.11%) | 16 (88.88%) | 18 |
| 51 and over | 0 (0.0%) | 7 (100%) | 7 |

The table: 5 shows that malaria predominantly affects younger age groups, particularly children aged 1 to 10 years, who make up 37.5% of the malaria patients. The incidence of malaria decreases with age, with only 26.6% of malaria cases occurring in the 11 to 20 age group. This trend continues in older age groups, with adults aged 21 to 30 years accounting for just 9.09% of malaria patients, and the proportion remains low in the 31 to 40 (10.34%) and 41 to 50 (11.11%) age groups. Notably, there were no malaria cases among individuals aged 51 years and over. This distribution suggests that younger individuals, especially children, are more susceptible to malaria, while older age groups appear to be at lower risk.

***Table: 6 The Significant Factors of Malaria***

| Predictors | Coefficients | Coefficient SE | P-value |
|---|---|---|---|
| Constant | 2.39301 | 1.98940 | 0.229 |
| Age | -0.0718353 | 0.0496440 | 0.048 |
| First appearance of malaria | 11.5126 | 5.54723 | 0.038 |
| Route of transmission | -1.31068 | 0.436619 | 0.003 |
| Malaria in family | 3.15210 | 1.22886 | 0.010 |

| | | | |
|---|---|---|---|
| Occurrence of malaria again | 5.86699 | 1.30902 | 0.000 |

The analysis of significant factors associated with malaria reveals that age negatively impacts malaria risk, with older individuals being less likely to contract the disease (p-value = 0.048). The first appearance of malaria and a family history of the disease both significantly increase the risk, with positive associations reflected in p-values of 0.038 and 0.010, respectively. The route of transmission, however, shows a significant negative association with malaria risk (p-value = 0.003), indicating some transmission routes might be less impactful than others. Most notably, recurrent malaria is strongly associated with a higher likelihood of future infection (p-value = 0.000), highlighting the critical role of previous malaria episodes in predicting future occurrences.

***Table: 7 Logistic Regression Analysis of Factors Influencing Malaria Occurrence***

| Variables | B | S.E. | Wald | Sig. | Exp(B) |
|---|---|---|---|---|---|
| Gender | 1.615 | 0.67 | 5.72 | 0.016 | 5.027 |
| Age | -0.80 | 1.487 | .003 | 0.862 | 0.449 |
| Union council | 0.93 | 0.2344 | 15.75 | 0.000 | 2.534 |
| Tehsil | -1.235 | 1.2957 | .000 | 1.000 | 0.290 |
| Education | 0.577 | 2.348 | 0.068 | 0.794 | 1.780 |
| Occupation | -.092 | 6.234 | 0.0217 | 0.882 | 0.389 |
| Disease | 1.056 | 1.595 | 0.956 | 0.328 | 2.874 |
| Type | 4.028 | 2.075 | 6.523 | 0.010 | 56.148 |
| Symptoms | 3.532 | 1.734 | 4.148 | 0.041 | 34.192 |
| First Appearance of Malaria | 4.802 | 1.967 | 6.35 | 0.011 | 121.753 |
| Route of Transmission | 3.485 | 1.078 | 10.451 | 0.001 | 32.622 |
| Nature of treatment | 5.737 | 4.153 | 1.908 | 0.167 | 310.132 |
| Result of treatment | 1.361 | 2.917 | 0.2176 | 0.682 | 3.900 |
| Environment | 3.485 | 1.078 | 10.451 | 0.001 | 32.622 |
| Malaria in area | -1.497 | 5.464 | 0.075 | 0.784 | 0.223 |
| Malaria in family | 2.081 | 2.897 | 0.515 | 0.472 | 8.012 |
| Malaria again | 5.585 | 3.764 | 2.201 | 0.138 | 266.400 |
| Water and toilet facility | -6.166 | 2.590 | 5.66 | 0.017 | 0.002 |
| Confined | 6.298 | 3.926 | 2.753 | 0.097 | 538.614 |

The above table-7 demonstrate the logistic regression analysis reveals several key predictors of malaria risk. Gender significantly impacts the likelihood of malaria, with males being over five times more likely to contract the disease (Exp(B) = 5.027). Union council also plays a critical role, as residing in certain areas more than doubles the odds of malaria (Exp(B) = 2.534). Specific types of malaria, the presence of symptoms, and the initial appearance of malaria are highly significant predictors, greatly increasing the risk of infection, with odds ratios of 56.148, 34.192, and 121.753, respectively. Routes of transmission and environmental factors further elevate risk significantly (Exp(B) = 32.622). Conversely, age, education, occupation, and malaria in the family show minimal to no significant effect on malaria risk. Water and toilet facilities are a significant protective factor, drastically reducing malaria risk (Exp(B) = 0.002). Factors like the nature of treatment, result of treatment, and malaria occurring again are not statistically significant predictors, suggesting limited impact on malaria risk in this analysis.

The below bar chart in Fig:1 shows the frequency distribution across various union councils, with Kot having the highest number of occurrences, indicating it as the most significant hotspot. In contrast, Totakan has the lowest frequency, suggesting it is the least affected or involved. The other union councils display varying levels of frequency, with Selai Pattai and Middle Batkhela also having relatively high counts, highlighting differences in the distribution of the measured variable across these areas.

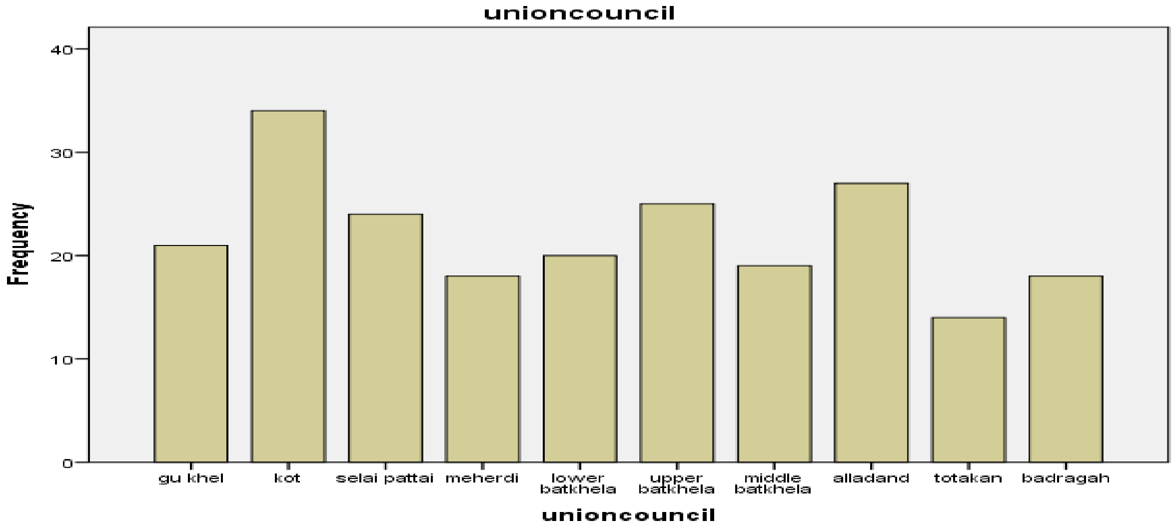


***Fig: 1 Union Council Wise Frequency***

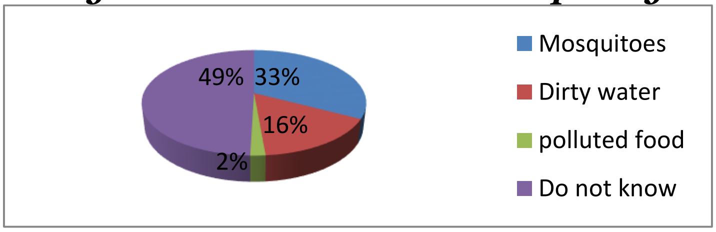
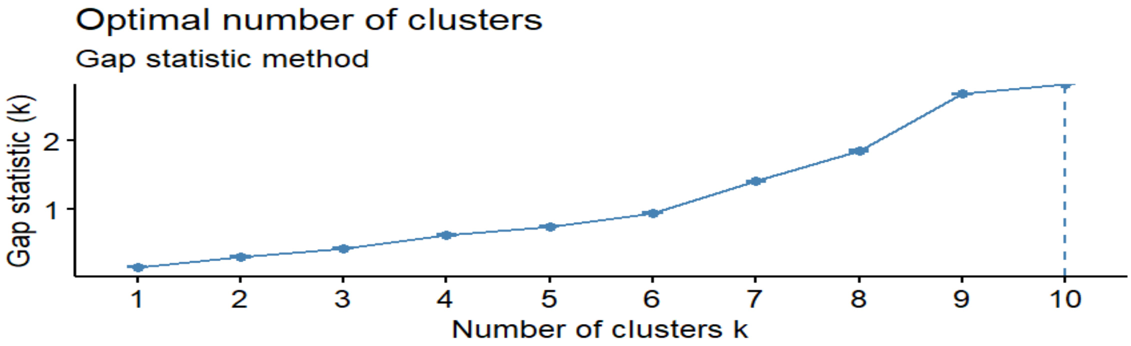


**Fig: 2 The Prevalence of Malaria Due to Route of Transmission**

The pie chart illustrates in Fig:2 the perceived routes of transmission for malaria among respondents. The majority, 49%, indicated that they "Do not know" how malaria is transmitted, which highlights a significant gap in awareness. Among those who identified specific transmission routes, 33% correctly identified mosquitoes as the primary vector for malaria. Additionally, 16% incorrectly believed that dirty water is a cause, and 2% attributed malaria to polluted food, which is also incorrect. This suggests that while some people are aware of the correct transmission route, there is still widespread misinformation or lack of knowledge regarding how malaria spreads.

### Gap Statistic

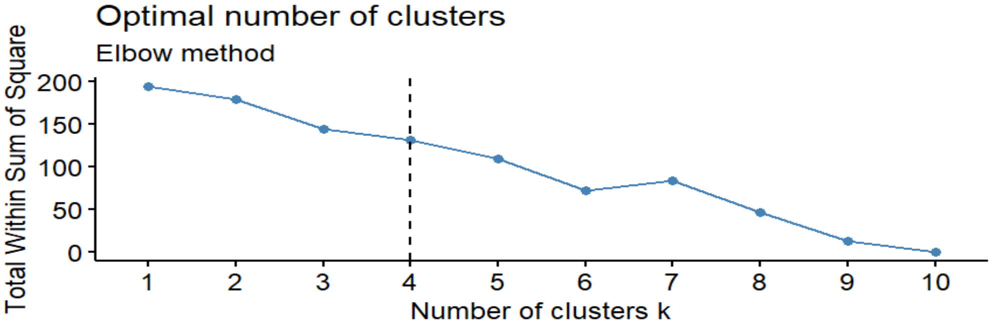
The Gap Statistic Method was applied to determine the optimal number of clusters for analyzing malaria prevalence and associated risk factors in District Malakand. The graph shows that the gap statistic increases steadily with the number of clusters, reaching its maximum value at k = 10, where a vertical dashed line is drawn. This indicates that ten clusters offer the most significant improvement over a random uniform distribution, suggesting k = 10 as the optimal number of clusters. This choice enables a more refined grouping of the population or regions based on malaria-related attributes, enhancing the accuracy of identifying high-risk areas and tailoring intervention strategies accordingly. The Fig: 3.1 below demonstrate the optimal number of clusters and risk factor in District Malakand.



***Gap Statistic Fig: 3.1***

### Elbow Method

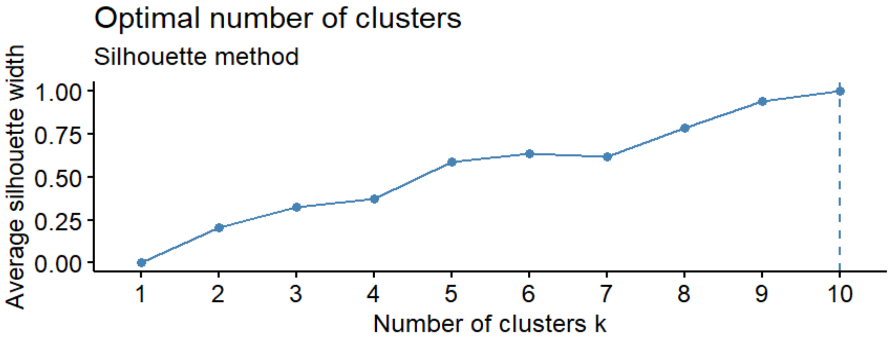
The Elbow Method plot was used to determine the optimal number of clusters for grouping malaria-related data in District Malakand. The graph shows a decreasing trend in the total within-cluster sum of squares (WSS) as the number of clusters (k) increases, indicating improved compactness of clusters. However, the rate of decrease slows significantly after **k = 4**, forming a distinct "elbow" shape. This suggests that dividing the dataset into four clusters offers the best balance between model simplicity and clustering performance. Therefore, **k = 4** was selected as the optimal number of clusters to effectively capture variations in malaria prevalence and associated risk factors across the region. Fig: 3.2 illustrates the optimal number of clusters and highlights the distribution of malaria risk factors across District Malakand.



***Elbow Method Fig: 3.2***

### Silhouette Method

The Silhouette Method was used to identify the optimal number of clusters for categorizing malaria prevalence and risk factors in District Malakand. The plot demonstrates that the average silhouette width—an indicator of how well each data point fits within its assigned cluster—increases with the number of clusters and reaches its highest value at **k = 10**, as indicated by the dashed vertical line. This suggests that ten clusters provide the best-defined and most distinct groupings in the dataset. Consequently, selecting **k = 10** allows for more precise segmentation of the population or areas, thereby enhancing the understanding of malaria distribution and aiding in the development of more targeted control strategies. Fig: 3.3 presents the ideal number of clusters and provides insights into how malaria risk factors are distributed throughout District Malakand.



***Average Silhouette Method Fig: 3.3***

## Segment Profiling Plot

Similarly, Figure 3.4 the Segment profiling plot the K-means clustering algorithm divided the malaria data from Tehsil Batkhela, District Malakand into four distinct clusters based on locality-specific patterns. Cluster 1, comprising 70% of the data, includes most areas such as Gu Khel, Kot, Meherdi, and others, likely representing baseline or low malaria risk. Cluster 2, dominated by Selai Pattai, suggests a distinct malaria pattern or elevated risk in that area. Cluster 3 is primarily composed of data from Lower Batkhela, indicating a localized concentration of malaria cases. Similarly, Cluster 4 is heavily represented by Totakan, which also shows a unique risk profile. This clustering highlights spatial heterogeneity in malaria prevalence and helps identify high-risk zones that may require targeted public health interventions.

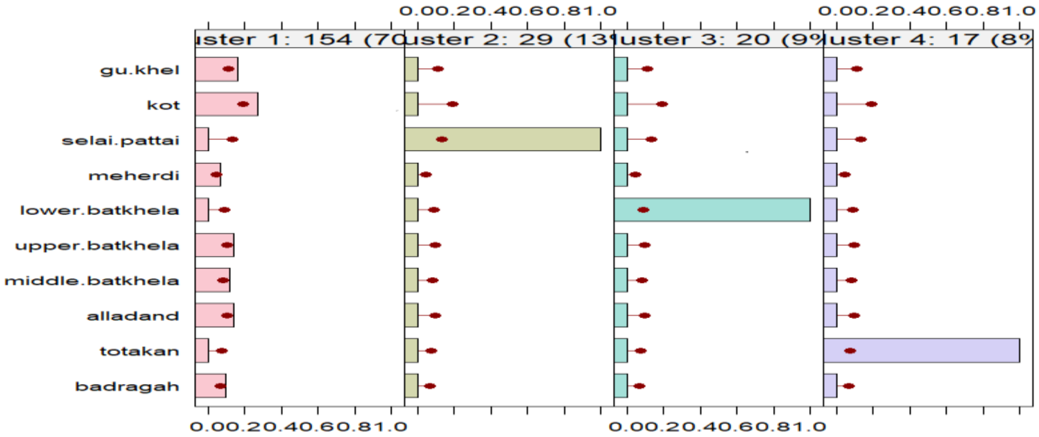


***Fig:3.4 Segment Profiling Plot***

## Conclusion

In conclusion, this study thoroughly investigated the prevalence and key risk factors associated with malaria in District Malakand, uncovering critical patterns in disease distribution and vulnerability. The analysis revealed a notably higher incidence among younger individuals, particularly males, with mosquito bites identified as the predominant mode of transmission. Using binary logistic regression, significant predictors of malaria included age, first-time occurrence, family history, and recurrent episodes, indicating a multifaceted interplay of demographic and biological factors. Furthermore, the study uncovered a considerable lack of awareness among the local population about malaria transmission, underscoring the urgent need for public health education initiatives, proper use of insecticide-treated bed nets, and focused vector control strategies. To enhance the robustness of the analysis, clustering techniques were employed to segment the population based on risk profiles. The Elbow Method indicated an optimal cluster number of k = 4, suggesting a balanced grouping structure for analyzing risk variations. In contrast, both the Gap Statistic and Silhouette Method identified k = 10 as the most effective number of clusters, offering more detailed differentiation among high- and low-risk populations. These findings emphasize the need for tailored intervention strategies that account for regional and demographic differences in malaria risk. Also, the profiling segment suggest that Cluster-1, comprising 70% of the data, includes most areas such as Gu Khel, Kot, Meherdi, and others have low malaria disease. While the others cluster selai pattai, lower batkhela and totankan have highly malaria disease found in these Cluster-2, Cluster-3, and Cluster-4 respectively. Lastly, continued research on improved diagnostic tools and treatment protocols is essential for achieving sustainable malaria control in the region.

## Recommendations

Future studies should expand geographically to cover neighboring regions, conduct longer-term research to observe seasonal malaria patterns, and explore the genetic diversity of Plasmodium species to understand drug resistance. Evaluating the effectiveness of interventions like bed nets, spraying, and education programs, along with studying public knowledge and behaviors, will guide more targeted prevention efforts. Research should focus on vulnerable groups such as children and pregnant women, while monitoring drug resistance and examining environmental factors affecting mosquito breeding will inform better control strategies. These efforts will help improve malaria prevention and management in endemic areas.

## References

1. World Health Organization. (2022). *Malaria*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/malaria
2. Vermont, G., & Tennessee C.D.C. (1997). Malaria in an immigrant and travelers. *MMWR Morb Mortal Wkly Rep*, 46(23), 536-539.
3. Cox-Singh, J., Davis, T. M., Lee, K. S., Shamsul, S. G., Matusop, A., Ratnam, S., Rahman, H. A., Conway, D. J., & Singh, B. (2008). *Plasmodium knowlesi malaria in humans is widely distributed and potentially life threatening*. Clinical Infectious Diseases, 46, 172-173.
4. Boel, M. E., Rijken, M. J., Brabin, B. J., McGready, F. N. R. (2012). The epidemiology of postpartum malaria: A systematic review. *Malaria Journal*, 11, 114.

5. Daneshvar, C., Davis, T. M., Singh, J. C., Rafa'ee, M. Z., Zakaria, S. K., Divis, P. C. S., & Singh, B. (2009). Clinical and laboratory features of human *Plasmodium knowlesi* infection. *Clinical Infectious Diseases*, 49, 852–860.
6. Monteiro, M. C., Oliveira, F. R., Oliveira, G. B., Romao, P. R. T., & Maia, C. S. F. (2014). Neurological and behavioral manifestations of cerebral malaria. *World Journal of Translational Medicine*, 3(1), 9-16.
7. Bhalli, M. A., & Samiullah. (2001). Falciparum malaria A review of 120 cases. *Journal of the College of Physicians and Surgeons Pakistan*, 11, 300–303.
8. Greenwood, B. M. (2005). *Malaria*. The Lancet, 365.
9. World Health Organization. (2013). *Eastern Mediterranean Regional Office*. Cairo: World Health Organization.
10. International Livestock Research Institute. (2001). *Investing in animal health research to alleviate poverty* (Appendix 2: Review on disease occurrence and impact). Accessed 2006.
11. Joy, D., et al. (2003). Early origin and recent expansion of *Plasmodium falciparum*. *Science*, 300(5617), 3121. https://doi.org/10.1126/science.1082211
12. Escalante, A., Freeland, R., Collins, W., & Lal, A. (2000). Evaluation of primate malaria parasites based on genes encoding cytochrome b from the linear mitochondrial genome. *Journal of Parasitology*, 86(5), 948-955.
13. Schlagenhauf-Lawlor, P. (2008). *Traveler's malaria*. PMPH-USA.
14. *Malaria – A Global Challenge*. Copyright: The Society for General Microbiology.
15. Warrell, D. A., et al. (2003). Oxford Textbook of Medicine (5th ed., 7.8.2 Malaria).
16. American Journal of Tropical Medicine and Hygiene. (2001). The neglected burden of *Plasmodium vivax* malaria. *American Journal of Tropical Medicine and Hygiene*, 64(1-2 Suppl), 97-106.
17. Mueller, I., Zimmerman, P. A., & Reeder, J. C. (2007). *Plasmodium malariae* and *Plasmodium ovale*—The "bashful" malaria parasites. *Trends in Parasitology*, 23(6), 278-283.
18. Nadjm, B., & Behrens, R. H. (2012). Malaria: An update for physicians. *Infectious Disease Clinics of North America*, 26(2), 243–259.
19. International Livestock Research Institute. (2001). *Investing in animal health research to alleviate poverty* (Appendix 2: Review on disease occurrence and impact). Accessed 2006.
20. Bear, N. A., et al. (2006). *American Journal of Tropical Medicine and Hygiene*, 75(5), 790-797.
21. Tinto, H., Rwagacondo, C., Karma, C., Mupfasoni, D., Vendor, W., Rusanganwa, E., Earhart, A., Overmeir, C. van, Merck, E. van, D'Alessandro, U. (2006). In-vitro susceptibility of *Plasmodium falciparum* to monodes ethylamodiaquine, dihydro-artemsinin, and quinine in an area of high chloroquine resistance in Rwanda. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 100(6), 509-514.
22. Gavi, the Vaccine Alliance. (2023). *Climate change causes malaria cases to triple in northwest Pakistan*. Retrieved from https://www.gavi.org/vaccineswork/climate-change-causes-malaria-cases-triple-northwest-pakistan

23. Yasinzai, M. I., & Kakarsulemankhel, J. K. (2009). Prevalence of human malaria infection in District Ziarat and Sanjavi, Pakistan. *Pakistan Journal of Zoology*, 41, 475-482.
24. Tyndall, J. A., Olumoh, S. J., Oborkhale, L., Apari, E., Nyitiba, I., & Bwala, J. A. (2012). The dilemma of falciparum malaria telecommunications in northeastern Nigeria. *American Journal of Science and Industrial Research*, 3(3), 157-165.
25. Bloland, P. B., Slutsker, L., Steketee, R. W., Wirima, J. J., Heymann, D. L., & Breman, J. G. (2003). Beyond chloroquine: Implications of drug resistance for evaluating malaria therapy efficacy and treatment policy in Africa. *Journal of Infectious Diseases*, 187(Suppl 1), S17–S23. https://doi.org/10.1086/368104
26. Berkley, J. A., Mwangi, I., Ngetsa, C. J., Mwarumba, S., Lowe, B. S., Marsh, K., & Newton, C. R. (2003). Diagnosis of acute bacterial meningitis in children at a district hospital in sub-Saharan Africa. *The Lancet*, 361(9375), 1141–1147. https://doi.org/10.1016/S0140-6736(03)12980-3
27. Ahmed, S. M., Hossain, M. A., Kabir, M., & Roy, S. K. (2009). Knowledge on the transmission, prevention, and treatment of malaria among two endemic populations of Bangladesh and their health-seeking behavior. *Malaria Journal*, 8, 173. https://doi.org/10.1186/1475-2875-8-173
28. Odugbemi, T. O., Akinsulire, O. R., Aibinu, I. E., & Fabeku, P. O. (2007). Medicinal plants useful for malaria therapy in Okeigbo, Ondo State, Southwest Nigeria. *African Journal of Traditional, Complementary and Alternative Medicines*, 4(2), 191–198. https://doi.org/10.4314/ajtcam.v4i2.31207
29. Chuma, J., Okungu, V., & Molyneux, C. (2010). Barriers to prompt and effective malaria treatment among the poorest population in Kenya. *Malaria Journal*, 9, 144. https://doi.org/10.1186/1475-2875-9-144
30. Adimi, F., Soebiyanto, R. P., Safi, N., & Kiang, R. K. (2010). Towards malaria risk prediction in Afghanistan using remote sensing. *Malaria Journal*, 9, 125. https://doi.org/10.1186/1475-2875-9-125
31. Mumtaz, A., Shaikh, S. A., Diju, I. U., & Dur-E-Yakta. (2011). Platelet count in malaria patients. *Journal of Ayub Medical College Abbottabad*, 23(1), 143–145.
32. Gandhi, M. P., Mushtaq, F., Mantoo, A. G., & Ghosh, M. (2024). Application of Geospatial Technique in Analysis of Malaria Prevalence in an Endemic Area of Ranchi, India. In *Geospatial Techniques in Health and Environmental Studies* (pp. 123–135). Springer. https://doi.org/10.1007/978-3-031-12345-6_10
33. Singh, N., Shukla, M. M., Shukla, M. K., Mehra, R. K., Sharma, S., Bharti, P. K., & Chand, G. (2010). Field and laboratory comparative evaluation of rapid malaria diagnostic tests versus traditional and molecular techniques in India. *Malaria Journal*, 9, 191. https://doi.org/10.1186/1475-2875-9-191
34. Penn State Eberly College of Science. (n.d.). Lesson 3: Logistic Regression Models. STAT 462: Applied Regression Analysis. Retrieved from https://online.stat.psu.edu/stat462/node/207
35. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8)
36. Atif, M., Ayub, G., Shakoor, F., Farooq, M., Iqbal, M., Zaman, Q., & Ilyas, M. (2024). Evolution of waterborne diseases: A case study of Khyber Pakhtunkhwa, Pakistan. *SAGE Open Medicine*, *12*, 20503121241263032.

37. Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and visualize the results of multivariate data analyses*. R package version 1.0.7. Retrieved from https://CRAN.R-project.org/package=factoextra