International Journal of Computer Applications Technology and Research
Volume 13–Issue 12, 19 – 25, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1312.1004

Statistical Analysis of Key Factors Influencing Leptospirosis Disease in Sri Lanka

AGR Sandeepa
Department of Computer Engineering
Faculty of Computing
General Sir John Kotelawala Defence University
Sri Lanka

S Dayarathne
IIT School of Computing
Faculty of IT
Informatics Institute of Technology
Sri Lanka

Abstract: In Sri Lanka, Leptospirosis is among the leading public health concerns due to its high morbidity and wide spectrum of clinical manifestations. This study uses data from hospitals all over Sri Lanka, from 2016 to 2019, to investigate the main determinants influencing the outcome of Leptospirosis. Statistical methods, such as logistic regression and chi-squared tests, have been used in order to examine the relationship between disease outcome and various clinical and demographic variables. The findings show a remarkable association between certain symptoms, diagnostic variables, and outcomes. Neck stiffness, leukocyte count, and diagnostic methods were the critical predictors established through logistic regression analysis. Although the predictive model has moderate accuracy 43.6% and an AUC of 0.546, it promises individually targeted interventions and better disease management in endemic areas. The present study underscores the importance of evidence-based approaches for the reduction of the leptospirosis burden in Sri Lanka.

Keywords: Leptospirosis, Chi-squared test, Logistic regression, Clinical variables, Statistical methods, Predictive model

1. INTRODUCTION

Leptospirosis frequently referred to as "rat fever," is a zoonotic bacterial infection caused by the Leptospira species. The usual mode of transmission in humans is through infection via either direct contact with contaminated water bodies or indirectly through the excreta of infected animals, especially rats. This condition is widespread in tropical and subtropical regions, for example, Sri Lanka, due to environmental conditions that, together with occupational practices, favor transmission. This disease ranges from mild, flu-like symptoms to severe presentations, such as Weil's disease and pulmonary bleeding, which may result in death if untreated. Leptospirosis has complicated clinical presentations.

Sri Lanka reports a high incidence of Leptospirosis because of the country's dependence on agriculture and periodic monsoon floods that create ideal environments for the spread of Leptospira. The World Health Organization says Leptospirosis is one of the major contributors to the morbidity of rural Sri Lanka, and that epidemics have affected thousands of people each year. Despite public health interventions, this disease remains a challenge for diagnosis and treatment since its symptoms mimic those of other febrile diseases, especially dengue fever and typhoid.

The impact of clinical and demographic variables on the outcome of leptospirosis has been widely studied in international literature. It is well known that the outcome and prognosis of any disease can be influenced by a number of factors such as age, gender, comorbid conditions, and early detection of symptoms, to name a few. However, research tailored to the specific needs of the Sri Lankan context given its peculiar epidemiological and environmental setting is scanty.

Hence, this research gap can be filled by analyzing the hospital-based data collected from 2016 to 2019 on leptospirosis, with a view toward ascertaining the major determinants of its outcomes in Sri Lanka.

The research objectives of the study are as follows.

1. To analyze the relationship of clinical and demographic variables, including symptoms and diagnostic markers, with the outcomes of Leptospirosis.
2. To employ statistical methodologies, logistic regression and chi-squared testing to determine the predictive importance of the above factors.
3. To provide legislators and medical professionals with practical recommendations for enhancing leptospirosis diagnosis and treatment protocols.

This work shows the potential of data-driven public health initiatives while furthering our understanding of the epidemiology of leptospirosis in Sri Lanka using statistical tools. It explores how particular strategies to mitigate the burden of leptospirosis in an endemic region may be informed by factors such as diagnostic accuracy, symptom prevalence, and demographic characteristics.

2. LITERATURE REVIEW

2.1 The Global Context of Leptospirosis

Leptospirosis is widely recognized to be a potentially fatal zoonotic infection, particularly in tropical and subtropical regions of the world. It has been reported that severe leptospirosis may kill up to 10% of victims and that over a million people are infected every year. The disease is often diagnosed at low rates due to its broad range of clinical symptoms, which may easily resemble other feverish disorders such as typhoid, dengue, and malaria. The excessive rains, poor sanitation, and occupational exposure of farmers and fishermen are the main reasons for its spread in many countries, including Sri Lanka.

Recent international studies have emphasized the need for early diagnosis and identification of risk factors to improve therapeutic outcomes. Factors such as comorbid conditions, high bacterial load, and delayed treatment are associated with increased severity and mortality rates.

Daher et al. [1] identified that age is indeed a significant predictor of poor outcomes for leptospirosis patients who are hospitalized, with the majority of them usually being the older ones.

2.2 Leptospirosis in Sri Lanka

Coupled with recurring epidemics of monsoonal flood-associated leptospirosis, Sri Lanka ranks among the countries most severely affected by leptospirosis. Major outbreaks were reported in 2008, 2011, and 2016, each concurrent with periods of heavy rain. Because of a constant rise in exposure to contaminated water sources, males are predominantly affected, particularly those whose occupations are labor-intensive, like agriculture.

Various studies in Sri Lanka have emphasized that environmental and occupational risk factors are more or less contributing to the spread of leptospirosis. Dissanayake et al., [2] indicated that paddy field workers are the most vulnerable to acquiring leptospirosis, as more than 70% of cases occur in rural agricultural settings. In addition, the presence of Leptospira in flood-related waterlogged areas creates endemic hotspots and differentially affects rural communities [3].

2.3 Factors Influencing the Outcomes of Leptospirosis

Clinical [4], environmental, and demographic factors [5] contribute to the outcomes and prognosis of leptospirosis. The following have been observed in some relevant studies [6], [7].

Age and Gender: It is noted from research that younger individuals of less than 60 years of age have better immune responses, while the mortality rates are higher among older populations due to comorbidities and delayed diagnoses. Males are thus more prone to become infected due to occupational exposition.

Clinical Symptoms: Symptoms such as headache, nausea, and fever are often nonspecific and thus may lead to incorrect diagnoses. Acute renal impairment and hemorrhage in the lungs are complications that suggest serious cases and require immediate medical attention.

Diagnostic Techniques: Studies emphasize the need for accurate diagnostic tools, such as the polymerase chain reaction (PCR) test, for the early detection of leptospirosis. Clinical diagnosis, when relied upon exclusively, often delays appropriate treatment in Sri Lanka, contributing to poorer outcomes.

2.4 Disease Analysis Using Statistical Methods

Both have greatly benefited from statistical modeling in predicting results and understanding the dynamics of illness. [8] Chi-square tests commonly find applications when assessing correlations between categorical data, including symptoms and diseases outcomes. Chi-squared tests are commonly used in assessing symptoms and disease outcome correlations that involve categorical data [9]. Logistic regression models, by showing variable significance through odds ratios and confidence intervals, also help in determining major predictors of illness development [10].

Various statistical approaches have been applied to research in Sri Lanka to analyze data trends in leptospirosis. Among these are the following: Helen et al, [11] employed logistic regression in analyzing the relation of clinical presentations to the severity of the disease, pointing out that symptoms such as abnormal white blood cell counts, and stiff necks are important. However, detailed studies integrating clinical and demographic factors into predictive models are still relatively few.

2.5 Research Gap

Even though a great amount of research on leptospirosis has been done worldwide, little has been done locally to address the peculiarities of the Sri Lankan setting. Most studies look at clinical presentations or the incidence of illness but lack robust statistical analyses that can identify the predictive variables that influence outcomes. This study fills the gaps by applying statistical methods to investigate hospital-based data and provides insights relevant to the local endemic conditions and environment in Sri Lanka.

3. METHODOLOGY

3.1 Study Design

The study employs a quantitative analytical framework to explore the association between demographic, clinical, and diagnostic factors influencing Leptospirosis outcomes in Sri Lanka. A retrospective approach was adopted, using existing hospital data collected from 2016 to 2019.

3.2 Data Collection

The dataset was obtained from Base Hospitals and Teaching Hospital in Sri Lanka during the period 2016 to 2019, representing high and low prevalence regions for leptospirosis. It includes demographic variables (e.g., age, gender), clinical symptoms (e.g., headache, nausea), and diagnostic test results (e.g., qPCR Diagnosis). Patients diagnosed with leptospirosis during the study period form the study cohort.

3.3 Independent Variables and the Dependent Variable

Table 1. Independent Variables and the Dependent Variable

Independent Variables	Dependent Variable
Demographics: Age, gender.	Final disease status (confirmed or not detected)
Clinical symptoms: Presence of headache, nausea, neck stiffness, etc.	
Diagnostic tests: Results of qPCR Diagnosis and other blood parameters (e.g., WBC count).	

3.4 Selection of Variables

The variables selected for analysis were based on a review of existing literature on leptospirosis and clinical experience with the disease. Additionally, preliminary descriptive analyses were conducted to examine the distribution of symptoms and test results in the population.

3.5 Data Cleaning and Preprocessing

Data from patient records were cleaned to address missing or incomplete information. Missing values in the dataset, represented as '99' in some instances, were handled by

imputation methods. Categorical variables like gender and symptoms were encoded as binary variables (1 for present, 0 for absence). Continuous variables like age and WBC count were kept in their original form after handling outliers using appropriate statistical techniques such as trimming.

3.6 Descriptive Analysis

Descriptive statistics were initially employed to summarize the data. This included calculating frequencies and percentages for categorical variables (such as gender, symptoms, and diagnostic test results), and means, medians, and standard deviations for continuous variables (like age and WBC count). The objective was to understand the overall distribution of the data and identify any apparent patterns.

The final status of the patients in the dataset were graphed as shown in figure 1. The number of not-detected patients is higher than the confirmed patients as depicted in figure 1.

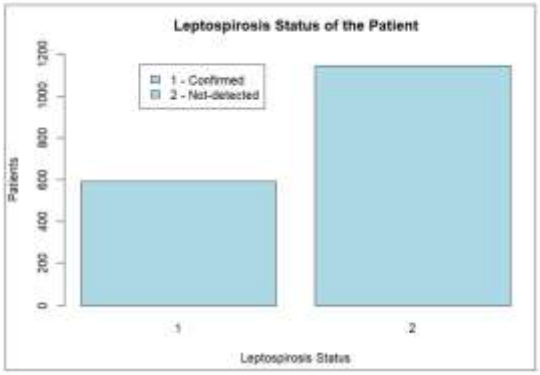

Figure 1. The Bar Plot for Final Status of Leptospirosis of Patients

The age distribution of confirmed Leptospirosis patients was plotted as shown in figure 2.

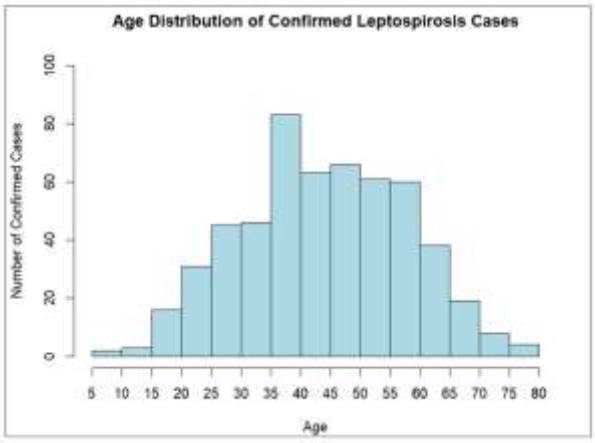

Figure 2. Age Distribution of Confirmed Leptospirosis Patients

The highest number of patients are recorded between 35-40 years. Mean, median and quartile values were also recorded and shown in the figure 3.

```
> summary(confirmed_cases$Age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   5.00   35.00   44.00   44.54   55.00   80.00
```

Figure 3. Mean, Median and Quartile Values for Confirmed Leptospirosis Cases

Third quartile value in figure 3 indicates that 75% of the confirmed patients are below the age of 55.

Similarly, the graphs and the descriptive statistics were obtained for Demographic, clinical and diagnostic characteristics.

3.7 Hypothesis Testing

Several hypothesis tests were performed to assess the relationship between different factors and the final disease status (confirmed or not detected).

- **Chi-Squared Test**: A Chi-Squared test for independence was performed to assess the relationship between categorical variables (e.g., gender and final disease status). The null hypothesis (H_0) tested was that there is no association between gender and the final disease status, while the alternative hypothesis (H_1) suggested there is an association. A p-value of less than 0.05 indicated that the null hypothesis should be rejected.

- **t-Tests**: Independent sample t-tests were applied to assess differences in continuous variables, such as age and WBC count, between the confirmed and non-confirmed groups. A significance level of 0.05 was used for all statistical tests.

3.8 Logistic Regression Analysis

A logistic regression model was employed to predict the probability of leptospirosis being confirmed based on independent variables. Logistic regression was chosen because the outcome variable (final diagnosis) is binary (confirmed or not confirmed).

The logistic regression equation was constructed to include significant variables identified through preliminary analysis (such as age, gender, symptoms like headache or nausea, and diagnostic markers like WBC count). The model was fitted using maximum likelihood estimation, and the coefficients were interpreted as log-odds of being diagnosed with leptospirosis. Odds ratios (OR) were calculated to determine the relative odds of leptospirosis diagnosis for each unit increase in a predictor variable. For example, the odds ratio for age indicated how the likelihood of a confirmed diagnosis changes with each year of age.

3.9 Software Tools Used

In this study, R and Excel were the primary software tools used for data analysis and model development, each playing a crucial role in different stages of the analysis process. R is a powerful statistical computing and graphics software

environment that is widely used in data analysis and statistical modeling. It was employed for the core analysis, including logistic regression, hypothesis testing, and model evaluation. The glm() function in R was utilized to fit generalized linear models (GLMs), which were key in identifying significant predictors of leptospirosis confirmation. R's vast ecosystem of packages, such as ggplot2 for data visualization and dplyr for data manipulation, made it ideal for conducting sophisticated statistical analyses and refining models. The flexibility and extensive functionality of R allowed for in-depth exploration of the relationships between various predictors and the response variable.

Excel, on the other hand, was used primarily for data preprocessing and initial exploratory analysis. Its user-friendly interface facilitated data organization and quick computations of summary statistics. Excel was also invaluable for creating pivot tables, contingency tables, and basic visualizations, such as bar charts and pie charts, to gain initial insights into the data. Before advancing to more complex statistical modeling in R, Excel allowed for an easy and efficient way to examine trends, distributions, and potential relationships within the dataset. Together, R and Excel formed a complementary toolkit for the study. While Excel provided a straightforward platform for data organization and preliminary analysis, R enabled more sophisticated modeling, analysis, and visualization, leading to more accurate predictions and a deeper understanding of the factors influencing leptospirosis outcomes.

4. RESULTS AND DISCUSSION

Table 1 was prepared to compare the prevalence of symptoms of Age <= 60 group and Age > 60 group. p-value for each symptom was calculated using Chi-squared test, since symptoms used below are categorical variables.

Table 2. Comparison Chart for Symptoms between Age Groups

Comparison of Symptomps between Age <= 60 and Age > 60 Leptospirosis Patients on Admission			
Symptomps	Age <= 60 (Total = 476)	Age > 60 (Total = 69)	p-value
Fever	367 (77.1%)	48 (69.6%)	0.306
Headache	335 (70.4%)	37 (53.6%)	0.005
Chills	302 (63.4%)	39 (56.5%)	0.593
Rigors	200 (42.0%)	26 (37.7%)	0.734
Muscle Pain	329 (69.1%)	39 (56.5%)	0.056
Muscle Tenderness	271 (56.9%)	33 (47.8%)	0.263
Vomiting	119 (25.0%)	12 (17.4%)	0.268
Jaundice	116 (24.4%)	11 (15.9%)	0.200
Diarrhea	82 (17.2%)	8 (11.6%)	0.369
Oliguria	40 (8.4%)	6 (8.7%)	1.000
SOB	55 (11.6%)	13 (18.8%)	0.094
Nausea	198 (41.6%)	18 (26.1%)	0.025

At 95% significance level, there's a difference between Age <= 60 group and Age > 60 group regarding the prevalence of 'Headache' and 'Nausea' on admission. Age <= 60 group showed a higher prevalence for both 'Headache' and 'Nausea' compared to the elderly group. There's no significant difference in other symptoms between two groups at 95% significance level since the calculated p-values are greater than 0.05.

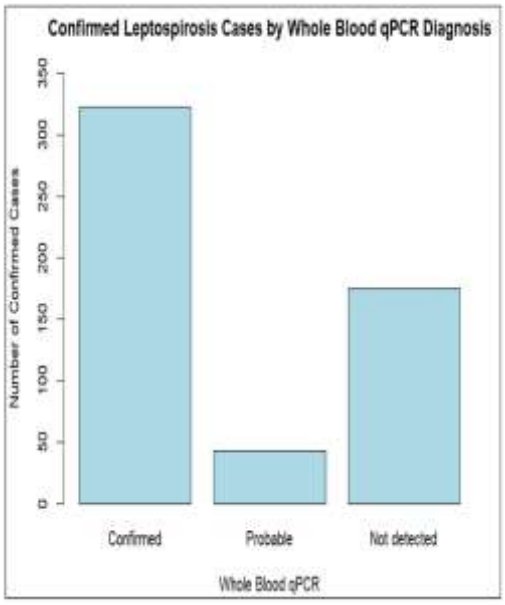

Figure 4. The Bar Plot for Confirmed Leptospirosis Cases by Whole Blood qPCR Diagnosis

Figure 4 shows that people who get confirmed for 'Whole Blood qPCR' are more likely to be confirmed for leptospirosis.

```
> lepto_table
                Final
WPqPCRDiagnosis Confirmed Not detected
   Confirmed          322            0
   Probable            43          164
   Not detected       175          751
> # Run a chi-squared test of independence
> chisq_test <- chisq.test(lepto_table)
> chisq_test

        Pearson's Chi-squared test

data:  lepto_table
X-squared = 700.93, df = 2, p-value < 2.2e-16
```

Figure 5. Contingency Table and p-Value obtained for Whole Blood qPCR

The p-value is less than 2.2e-16 indicating that there's a significant association between 'WPqPCRDiagnosis' and 'Final' variable.

International Journal of Computer Applications Technology and Research
Volume 13–Issue 12, 19 – 25, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1312.1004

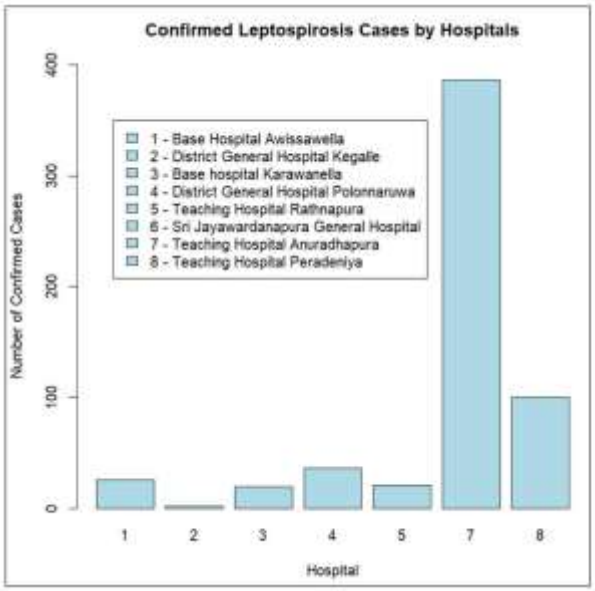

Figure 6. The Bar Plot for Confirmed Leptospirosis Cases by Hospitals

According to the figure 6, During the period of 2016 to 2019, Teaching Hospital in Anuradhapura has recorded the highest number of leptospirosis patients while District General Hospital Kegalle has recorded the least.

Selection of independent parameters was decided referring to past research and descriptive analysis performed above. Generalized Linear Models (GLM) were used to model the factors influencing the final status of leptospirosis cases (confirmed or not confirmed). The analysis aimed to identify significant predictors, such as demographic factors (age, gender), clinical symptoms (headache, nausea, neck stiffness), and diagnostic test results (e.g., WPqPCR). The GLM approach was specifically chosen for its flexibility in modeling different types of response variables, including binary outcomes (e.g., confirmed vs. not detected) and continuous measures.

```
# Subset the data without missing values
sub_lepto <- subset(lepto, Age != 99 & Sex != 99 & Muscletendernessad != 99
                    & Jaundicead != 99 & Neckstiffnessad != 99 & WBCcount != 99)


# Identify the columns to convert to factors
sub_lepto$Sex <- factor(sub_lepto$Sex)
sub_lepto$Muscletendernessad <- factor(sub_lepto$Muscletendernessad)
sub_lepto$Jaundicead <- factor(sub_lepto$Jaundicead)
sub_lepto$Neckstiffnessad <- factor(sub_lepto$Neckstiffnessad)
sub_lepto$Final <- factor(sub_lepto$Final)

n = nrow(sub_lepto)
n
set.seed(2000)
trainSerial = sample(1:n, n*0.8, replace=F)
train=sub_lepto[trainSerial,]
test=sub_lepto[-trainSerial,]


nrow(train)
nrow(test)

fullmodel <- glm(Final ~ Age + Sex + Muscletendernessad + Jaundicead +
                   Neckstiffnessad + WBCcount, family = "binomial", data = train)
reducedmodel <- step(fullmodel, direction = "backward")
summary(reducedmodel)
```

Figure 7. Building Logistic Regression Model Using R Software

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)    3.499e-01  2.914e-01   1.201 0.229900
Age           -8.636e-03  5.080e-03  -1.700 0.089143 .
Sex2           6.665e-01  1.994e-01   3.342 0.000831 ***
WBCcount      -2.486e-05  1.477e-05  -1.684 0.092217 .
Neckstiffnessad2 7.311e-01  1.660e-01   4.405 1.06e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 8. Results of the Logistic Regression Model

'Sex2' and 'Neckstiffnessad2' have p-values less than 0.05 and 'Age' and 'WBCcount' have p-values less than 0.1. Hence, there's a significant association between those variables and the response variable at the 90% significance level.

Equation 1. Logistic Regression Model

$$\ln\left(\frac{Pr(Leptospirosis)}{1-Pr(Leptospirosis)}\right) = 0.3499 - 0.008636 * Age + 0.6665 * Sex2 - 2.486e^{-5} * WBCcount + 0.7311 * Neckstiffnessad2$$

Interpretation of numerical variables: (E.g., Age)

When all other variables are held constant, for every one-unit increase in Age, there's a 0.008636 decrease in the log-odds of the outcome. Odds Ratio (Age) = e−0.008636 = 0.991 This implies that for every one-unit increase in Age, the odds of the outcome decrease by 0.9%, [= (1-0.991)*100%] when other variables are held constant.

Interpretation of categorical variables: (E.g., Sex)

When all other variables are held constant, Sex2 (Female) has a 0.665 increase in the log-odds of the outcome compared to Sex1 (Male). Odds Ratio (Sex2) = e0.665 = 1.944 This implies that females have 94.4% [= (1.944-1)*100%] higher odds of the outcome compared to males, when other variables are held constant.

4.1 Validation of the Logistic regression Model

The following code was implemented in R software to evaluate the metrics such as Accuracy, and Precision.

```
test$predicted <- predict(reducedmodel, newdata = test, type = "response")
test$predicted_status <- ifelse(test$predicted >= 0.5, 1, 2)


# Evaluate the model performance on the testing data
confusion_matrix <- table(test$Final, test$predicted_status)
accuracy <- sum(diag(confusion_matrix)) / sum(confusion_matrix)
precision <- diag(confusion_matrix) / colSums(confusion_matrix)
```

Figure 9. Testing Code of the Logistic Regression Model

International Journal of Computer Applications Technology and Research
Volume 13–Issue 12, 19 – 25, 2024, ISSN:-2319–8656
DOI:10.7753/IJCATR1312.1004

```
> cat("Accuracy:", accuracy, "\n")
Accuracy: 0.4357798
> cat("Precision:", precision, "\n")
Precision: 0.4147727 0.5238095
> table(test$Final, test$predicted_status)

     1   2
  1  73  20
  2 103  22
> table(test$Final)

  1   2
 93 125
```

Figure 10. Obtained Accuracy and Precision Values and Confusion Matrix

Accuracy of the model is 43.6%. Predicting confirmed and not-detected leptospirosis are 73/93 = 0.78 and 22/125 = 0.18 respectively. The model shows moderate accuracy while having a more tendency to detect confirmed patients than not-detected patients.

AUC value is calculated as 0.546, which indicates it performs slightly better than random chance in distinguishing final status of the outcome.

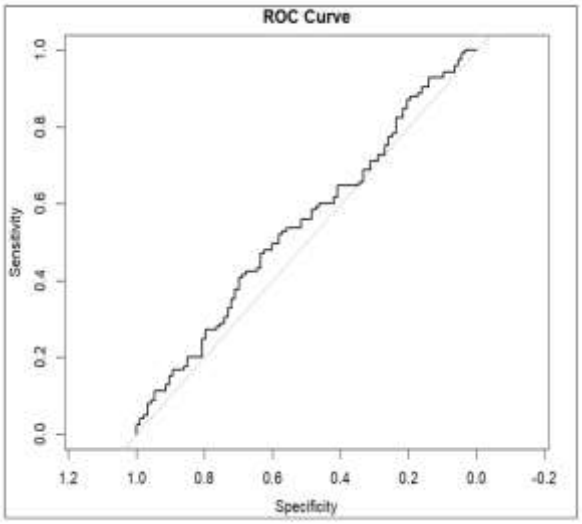

Figure 11. ROC Curve for the Model

Hosmer-Lemeshow Goodness-of-fit Test was performed under the following Hypothesis and a p-value was obtained.
H0 = Model is adequate
H1 = Model isn't adequate
Obtained p-value = 0.5589811
The obtained p-value from the Hosmer-Lemeshow goodness-of-fit test was 0.5589811. Given that this p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis. Therefore, we conclude that there is no significant evidence to suggest that the model is inadequate. In other words, the model fits the data adequately.

5. CONCLUSION

This study aimed to identify key factors influencing the confirmation of leptospirosis diagnosis through the application of generalized linear models (GLM), specifically logistic regression. The analysis revealed several significant predictors, including age, gender, neck stiffness, and the results of diagnostic tests such as Whole Blood qPCR (WPqPCR). While the model demonstrated moderate predictive performance, the accuracy was 43.5%, and the area under the curve (AUC) was 0.546. These findings suggest that the model's ability to distinguish between confirmed and non-confirmed leptospirosis cases was only slightly better than random chance. Despite this, certain predictors, such as age, gender, and neck stiffness, were identified as having significant associations with the likelihood of a confirmed diagnosis.

The recall for confirmed (positive) cases was found to be 78.4% (73 out of 93 confirmed cases correctly predicted), indicating that the model effectively identifies a large proportion of true confirmed cases. However, the recall for non-detected (negative) cases was relatively low at 17.6% (22 out of 125 non-detected cases correctly predicted), suggesting that the model struggles to correctly identify non-confirmed cases. These results highlight the model's strong performance in identifying confirmed cases of leptospirosis, but also underscore its limitation in distinguishing non-detected cases from confirmed ones. The relatively low accuracy and AUC value further emphasize the need for additional improvements in the model, including the incorporation of more relevant factors and possibly more advanced modeling techniques.

Despite these limitations, the study contributes valuable insights into the demographic and clinical factors that influence leptospirosis diagnosis. Specifically, being female, older, and exhibiting symptoms like neck stiffness were found to increase the odds of leptospirosis confirmation, which could guide initial clinical assessments. Moreover, the findings emphasize the importance of improving diagnostic capabilities and patient triaging in regions affected by leptospirosis. The model could still serve as a valuable starting point for healthcare professionals, particularly in resource-limited settings where quick access to diagnostic tools may not always be feasible. By prioritizing patients with certain demographic and clinical features, healthcare providers can make informed decisions on which patients may benefit from more immediate diagnostic testing, such as WPqPCR.

6. FUTURE WORK

While logistic regression has provided valuable insights in this study, there is significant potential for improving prediction accuracy by incorporating more advanced machine learning algorithms. Techniques such as random forests, support vector machines (SVM), and gradient boosting are capable of capturing non-linear relationships and complex interactions between variables that linear models may not effectively handle. These algorithms are particularly well-suited for datasets with intricate patterns and high-dimensional features, which is often the case in medical data like leptospirosis prediction. By utilizing these advanced methods, the model could not only improve its ability to distinguish between confirmed and non-detected cases but

also enhance its robustness and generalizability, leading to more accurate and reliable predictions in clinical practice. Incorporating these techniques would allow for better handling of the complexities inherent in the data, potentially transforming the model into a more powerful tool for early diagnosis and decision-making in leptospirosis management.

7. REFERENCES

[1] E. D. F. Daher *et al.*, "Leptospirosis in the elderly: the role of age as a predictor of poor outcomes in hospitalized patients," *Pathog. Glob. Health*, vol. 113, no. 3, pp. 117–123, May 2019, doi: 10.1080/20477724.2019.1621729.

[2] N. D. B. Ehelepola, K. Ariyaratne, and D. S. Dissanayake, "The interrelationship between meteorological parameters and leptospirosis incidence in Hambantota district, Sri Lanka 2008–2017 and practical implications," *PLOS ONE*, vol. 16, no. 1, p. e0245366, Jan. 2021, doi: 10.1371/journal.pone.0245366.

[3] "Determinants of leptospirosis in Sri Lanka: Study Protocol | BMC Infectious Diseases | Full Text." Accessed: Dec. 13, 2024. [Online]. Available: https://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-10-332

[4] P. S. Pinto, H. Libonati, and W. Lilenbaum, "A systematic review of leptospirosis on dogs, pigs, and horses in Latin America," *Trop. Anim. Health Prod.*, vol. 49, no. 2, pp. 231–238, Feb. 2017, doi: 10.1007/s11250-016-1201-8.

[5] A. F. B. Victoriano *et al.*, "Leptospirosis in the Asia Pacific region," *BMC Infect. Dis.*, vol. 9, no. 1, p. 147, Sep. 2009, doi: 10.1186/1471-2334-9-147.

[6] "Environmental and Behavioural Determinants of Leptospirosis Transmission: A Systematic Review | PLOS Neglected Tropical Diseases." Accessed: Dec. 13, 2024. [Online]. Available: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0003843

[7] "Leptospirosis in pregnancy: A systematic review | PLOS Neglected Tropical Diseases." Accessed: Dec. 13, 2024. [Online]. Available: https://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0009747

[8] "Spatial and statistical analysis of leptospirosis in Thailand from 2013 to 2015 | Geospatial Health." Accessed: Dec. 13, 2024. [Online]. Available: https://www.geospatialhealth.net/gh/article/view/739

[9] "Early identification of leptospirosis-associated pulmonary hemorrhage syndrome by use of a validated prediction model - ScienceDirect." Accessed: Dec. 13, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0163445309003892

[10] G. S. Galdino *et al.*, "Development and validation of a simple machine learning tool to predict mortality in leptospirosis," *Sci. Rep.*, vol. 13, no. 1, p. 4506, Mar. 2023, doi: 10.1038/s41598-023-31707-4.

[11] H. J. Mayfield, J. H. Lowry, C. H. Watson, M. Kama, E. J. Nilles, and C. L. Lau, "Use of geographically weighted logistic regression to quantify spatial variation in the environmental and sociodemographic drivers of leptospirosis in Fiji: a modelling study," *Lancet Planet. Health*, vol. 2, no. 5, pp. e223–e232, May 2018, doi: 10.1016/S2542-5196(18)30066-4.