International Science and Technology Journal
المجلة الدولية للعلوم والتقنية
Volume 39 العدد
Part 2 المجلد
ISTJ

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

Received	**2026/10/20**	**تم استلام الورقة العلمية في**
Accepted	**2026/11/09**	**تم قبول الورقة العلمية في**
Published	**2026/11/10**	**تم نشر الورقة العلمية في**

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

Youssef Alsadek Fhil Alboum[1], *Samira Alshafah[2]

[1] Higher Institute of Marine Sciences Techniques Sabratha. - Libya
[2] Computer Department, Faculty of Education
University of Zawia, Libya.
*Corresponding author email: samira_7_8@yahoo.com
* ORCID NUMBER 0009-0005-1325-6608

Abstract:

Amebic infection caused by the parasite (Entamoeba histolytica) remains a public health concern, particularly in settings where food and water contamination may facilitate transmission. Early detection, improved hygiene and sanitation, and safe food and water are important preventive measures. Confirmatory testing for species identification is also recommended because E. histolytica may be microscopically similar to non-pathogenic species. This study aimed to describe the proportion of laboratory samples recorded as positive for Entamoeba, examine the distribution of positive records by month and sex, and explore decision-tree classification. A retrospective analysis included 264 stool examination records from patients attending the biological laboratory at Al-Zawiya Hospital, Libya, from January 2020 to August 2021. The dataset comprised 104 records from 2020 and 160 from 2021. Recorded variables included year, month, sex, age, parasite test result, fecal mucus, and stool color. Of the 264 records, 90 (34.09%) were labelled positive and 174 (65.91%) negatives. The largest monthly count was recorded in July (18 cases). Of the 90 positive records, 53 were male (38.69%) and 37 were female (29.13%). The decision-tree analysis

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

was exploratory. The reported accuracy was 65.9%, precision for the positive class was 50.0%, sensitivity was 12.2%, and the F1-score was 19.6%. The low sensitivity indicates that the model missed most positive records and should not be used for clinical diagnosis or screening.

Keywords: E. histolytica; retrospective analysis; data mining; decision tree.

دراسة حول مدى انتشار طفيل الأميبا الحالة للنسج بين المرضى المراجعين لمستشفى الزاوية، باستخدام التحليل الإحصائي الوصفي وتقنيات استخراج البيانات

يوسف الصادق حسن فحيل البوم، سميرة الشفح

[1] المعهد العالي لتقنيات علوم البحار بصبراتة – ليبيا

[2] قسم الحاسوب، كلية التربية، جامعة الزاوية – ليبيا

samira_7_8@yahoo.com

الملخص:

لا تزال العدوى الأميبية التي يسببها طفيل (Entamoeba histolytica) تشكل مصدر قلق للصحة العامة، لا سيما في البيئات التي قد يُسهّل فيها تلوث الغذاء والماء انتقال العدوى. ويُعد الكشف المبكر، وتحسين النظافة والصرف الصحي، وتوفير المياه والغذاء الآمنين أمورًا أساسية للوقاية. كما يُوصى بإجراء اختبارات تأكيدية لتحديد نوع الطفيل، نظرًا لتشابه E. histolytica مجهريًا مع أنواع غير ممرضة. هدفت هذه الدراسة إلى وصف نسبة عينات المختبر التي أظهرت نتائج إيجابية لـ Entamoeba، ودراسة توزيع الحالات الإيجابية حسب الشهر والجنس، واستكشاف تطبيق تقنيات استخراج البيانات لتصنيف نتائج المختبر. أُجري تحليل استرجاعي باستخدام 264 عينة فحص براز من مرضى يراجعون المختبر البيولوجي في مستشفى الزاوية، ليبيا، في الفترة من يناير 2020 إلى أغسطس 2021. شملت مجموعة البيانات 104 عينات من عام 2020 و160 عينة

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

من عام 2021. تضمنت المتغيرات المسجلة السنة، والشهر، والجنس، والعمر، ونتائج اختبار الطفيل، ومحتوى المخاط في البراز، ولون البراز. من بين 264 عينة، صُنفت 90 عينة (34.09%) على أنها إيجابية، بينما صُنفت 174 عينة (65.91%) على أنها سلبية. سُجل أعلى عدد من الحالات الإيجابية في يوليو (18 حالة)، بينما سُجل أدنى عدد في ديسمبر (حالتان). من بين الحالات الإيجابية التسعين، كان 53 (38.69%) من الذكور و37 (29.13%) من الإناث. تم استكشاف نموذج شجرة القرار كنهج استكشافي لاستخراج البيانات لتصنيف نتائج المختبر. حقق النموذج دقة بلغت 65.9%، ودقة 50.0% للفئة الإيجابية، واستدعاء (حساسية) 12.2% للفئة الإيجابية، ودرجة F1 بلغت 19.6% للفئة الإيجابية.

الكلمات المفتاحية: الأميبا الحالة للنسج (Entamoeba histolytica)؛ الأوليات المعوية؛ استخراج البيانات؛ شجرة القرار.

1. Introduction

Intestinal protozoal infections remain an important public-health concern, particularly in communities where access to safe drinking water, adequate sanitation, and hygienic food handling is limited. Entamoeba histolytica is the protozoan responsible for amoebiasis, an infection that may remain asymptomatic or cause intestinal disease, including amoebic colitis and dysentery; in a smaller proportion of cases, it may spread beyond the intestine and cause extraintestinal complications such as amoebic liver abscess. The clinical and public-health importance of amoebiasis, together with its uneven distribution across populations, has been discussed in major clinical and epidemiological reviews [1,2,3].

Transmission occurs mainly through ingestion of mature cysts in food, water, or on hands contaminated with faecal material. However, the burden measured in a laboratory-based study depends not only on local transmission, but also on the population tested, referral patterns, symptoms, sampling strategy, and diagnostic method. An important diagnostic limitation is that cysts and trophozoites of E. histolytica may be morphologically indistinguishable from those of E. dispar and E. moshkovskii. Consequently, microscopy-only findings should be interpreted

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques
http://www.doi.org/10.62341/istj-vol39-2-ys34

carefully and should not treated as species-confirmed E. histolytica infection [3,4,5,6]. Antigen detection and species-specific molecular assays, including PCR-based approaches, can improve differentiation where these facilities are available [7]. In Libya, published evidence has documented considerable variation in reported Entamoeba histolytica/dispar positivity across locations, populations, and study methods. A review of studies published from 2000 to 2015 reported a wide range of estimates and emphasized that microscopy-based studies may have included non-pathogenic Entamoeba species [8]. More recent work from western Libya also provides a useful regional point of comparison, although differences in sampling frame, time period, and diagnostic approach must be considered before comparing percentages directly [9]. Local, clearly documented data remain useful for describing laboratory caseloads and identifying where improved diagnostic confirmation and surveillance may be needed.

In addition to descriptive epidemiology, data-mining methods can be used to explore patterns in routinely collected samples. Decision trees are attractive because their rules can be interpreted more easily than some complex predictive models; nevertheless, their performance depends on data quality, transparent preprocessing, proper separation of training and testing data, and validation [10,11,12]. The present study therefore describes 264 laboratory samples summarizes the recorded positive cases, and reviews the reported decision-tree output. Its results are interpreted as findings from the supplied laboratory dataset rather than as a population-wide estimate or proof of causal risk factors.

2. literature review

Entamoeba histolytica is transmitted primarily through the faecal–oral route by ingesting mature cysts in contaminated food or water or through contact with contaminated hands. Clinical manifestations range from asymptomatic intestinal carriage to invasive colitis and extraintestinal disease. Previous reviews have highlighted the association between amoebiasis and sanitation, food and water hygiene, and socioeconomic conditions. However, reported

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques
http://www.doi.org/10.62341/istj-vol39-2-ys34

infection rates vary considerably across geographical regions and study populations [1,2,3,13,14]. These variations emphasize the importance of local epidemiological surveillance. Nevertheless, the proportion of positive cases among patients attending a laboratory should not be considered representative of the general population unless a suitable population-based sampling design is used.

The diagnostic method is an important consideration when interpreting studies reporting the occurrence of amoebiasis. Although microscopy remains widely used in routine laboratories, its sensitivity is limited, particularly when examination is based on a single stool specimen. Furthermore, morphological examination alone generally cannot reliably distinguish E. histolytica from the morphologically similar species E. dispar and E. moshkovskii [3–5,20]. Antigen-detection assays can support the differentiation of pathogenic E. histolytica from non-pathogenic species, whereas species-specific polymerase chain reaction (PCR) methods offer a more definitive approach to species identification [3,6,7,15,16]. Therefore, when results are based exclusively on microscopy, the findings should be described as microscopy-positive for Entamoeba-like organisms rather than confirmed E. histolytica infections.

Studies conducted in Libya provide important context for interpreting the findings of the present investigation. A review of studies published between 2000 and 2015 reported substantial variation in estimates of E. histolytica/dispar occurrence among populations with and without gastroenteritis [8]. The review also highlighted the limitations of microscopy-based identification, which may result in the misclassification of E. dispar or E. moshkovskii as E. histolytica. In Bani Waleed, a study conducted between January and June 2023 identified 55 positive samples among 180 examined, corresponding to 30.56%, with the highest number of positive cases recorded in June [18]. Another study conducted in western Libya examined 600 stool samples collected in 2024 from Al-Zawiya, Surman, and Sabratha and reported an overall positivity rate of 29.2% [9]. Although these findings provide relevant regional evidence, differences in study periods,

populations, and sampling settings limit direct comparisons with the present investigation. Accordingly, the reported percentages should be interpreted as descriptive findings rather than evidence of statistically significant differences between studies.

Decision-tree algorithms are widely used in classification because they divide observations into groups according to predictor variables and generate explicit decision rules that can be readily interpreted. Foundational studies have established the principles of tree-based classification [10,11], while applied data-mining literature emphasizes the importance of defining the target variable, ensuring reliable class labels, separating training and testing datasets, and applying appropriate validation procedures [12,17]. Model performance should not be assessed using accuracy alone, as this measure may provide an incomplete picture of classification effectiveness. Instead, accuracy, precision, recall or sensitivity, specificity, and the F1-score should be considered together. In the present study, the decision-tree results require cautious interpretation because the terminal-node counts do not reconcile with the overall sample totals, and the accuracy displayed in the figure differs from the value calculated from the confusion matrix. These inconsistencies highlight the need to verify the model output against the original dataset before drawing conclusions about its predictive performance.

The existing literature highlights several issues relevant to the present study, including variations in reported Entamoeba positivity across Libyan settings, limitations of microscopy in establishing species identity, and the need for reproducible analysis and validation when applying predictive models to health data. Against this background, the present investigation describes laboratory records from Al-Zawiya Hospital covering the study period and explores the potential application of a decision-tree classifier to examine associations between recorded patient and stool-sample characteristics and test outcomes. The findings are intended to provide local descriptive evidence and an exploratory assessment of data-mining methods rather than definitive estimates of population prevalence or validated predictive performance. Further research

would benefit from species-specific diagnostic confirmation, clearly documented denominators for age, sex, and monthly distributions, and reanalysis of the original patient-level data using a reproducible analytical workflow.

3. Materials and Methods

Sample Collection: Samples were collected from patients attending Al-Zawiya Hospital, Libya, during 2020 and 2021 to detect the Entamoeba histolytica parasite, which causes amoebic dysentery, and to determine the percentage of positive results among the tested samples. Laboratory tests were performed on the samples, and the results were recorded in a database that included the year, month, sex, age, Entamoeba histolytica status, presence of mucus, and stool color. The database included 264 samples, 104 from 2020 and 160 from 2021. The number of samples recorded varied from month to month based on available data. During 2020, the samples were distributed across the months of January, February, and March, and then from May to December. No samples were recorded for April. For 2021, data covered the period from January to August, with no samples available for September to December. This distribution is presented based on the available samples.

Laboratory Examination: Stool samples collected from patients underwent laboratory testing, and the results were recorded in the database to determine the presence or absence of Entamoeba histolytica, along with certain morphological characteristics of the sample. This was done in two ways:

First: Gross Examination Gross examination involves observing the morphological characteristics of the stool sample, such as color, consistency, and the presence of mucus or blood, if present. These characteristics are recorded in the laboratory data to be used in describing the samples and comparing their properties with the results of microscopic examination. In the current database, stool color and the presence of mucus are recorded as variables associated with each sample. This data can be used to analyze the descriptive relationship between phenotypic characteristics and Entomobacillus

histolytica detection results, without considering these characteristics alone as conclusive evidence of infection.

Second: Stool Microscopic Examination

Stool microscopic examination is used to detect observable parasitic stages in the sample, including cysts and feeding stages, according to the laboratory's preparation and examination methods. Entomobacillus histolytica test results were recorded in the database under the Entomobacillus histolytica status variable. For data organization and statistical analysis, the value NO was considered a negative result, while other values recorded in the Entomobacillus histolytica column were considered positive results according to the data definition adopted in the study. Accordingly, the results were classified into two categories: negative and positive.

Statistical and Computational Analysis: The recorded data were used for descriptive analysis, including calculating frequencies and percentages of positive and negative outcomes, and distributing the records by year, month, sex, and age group. The results can be presented using tables and graphs to illustrate the temporal and demographic distribution of the recorded cases.

Data mining techniques were also employed, applying a decision tree algorithm. The Entomobacillus histolytica condition was used as the target variable, and age, sex, month, presence of mucus, and stool color could be used as explanatory variables. The software used, the data segmentation method, and the model performance evaluation procedures should then be documented. The results should be interpreted as exploratory taxonomic patterns and not as evidence of causal relationships.

5. Results and discussion

Overall samples prevalence of the 264 samples, 90 were labelled positive and 174 negative. This is a proportion among the examined samples, not necessarily a population prevalence estimate, because the dataset consists of patients attending medical laboratories. As shown in Figure 1, approximately one-third of the samples examined were classified as positive. The samples positivity

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

proportion was 34.09% (90/264), while 65.91% were labelled negative as in table1.

Table 1. Overall distribution of samples Entamoeba histolytica results.

Result	Number of records	Percentage (%)
Positive	90	34.09%
Negative	174	65.91%
Total	264	100.00%

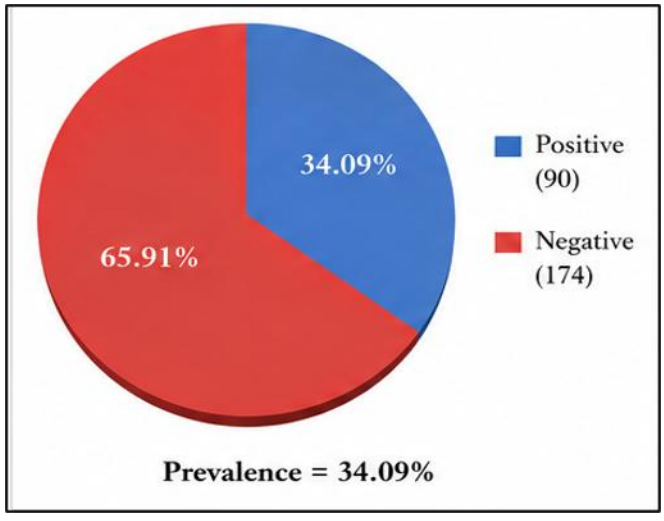


Figure 1. Figure 1, Overall distribution of positive and negative records (n = 264).

Monthly distribution of positive records. The number of positive records varied across the calendar months: January 12, February 3, March 16, April 6, May 7, June 12, July 18, August 9, September 4, October 1, November 0, and December 2. July had the largest number of positive records (18), followed by March (16), whereas October had the smallest non-zero count (1) and November had no positive records. These are counts, not monthly positivity rates. Because the number of records tested differed by month and the dataset covers January 2020–August 2021 with incomplete calendar coverage in each year, the counts alone cannot establish seasonality

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

or a higher monthly risk of infection. The monthly percentages in Table 2 are more informative for describing the proportion positive among records tested in each month, but they should still be interpreted cautiously because some monthly denominators are very small.

Table 2. Monthly distribution of samples results.

Month	Total records	Positive	Negative	Positive (%)
January	33	12	21	36.36%
February	23	3	20	13.04%
March	39	16	23	41.03%
April	16	6	10	37.50%
May	25	7	18	28.00%
June	39	12	27	30.77%
July	39	18	21	46.15%
August	32	9	23	28.12%
September	6	4	2	66.67%
October	4	1	3	25.00%
November	1	0	1	0.00%
December	7	2	5	28.57%

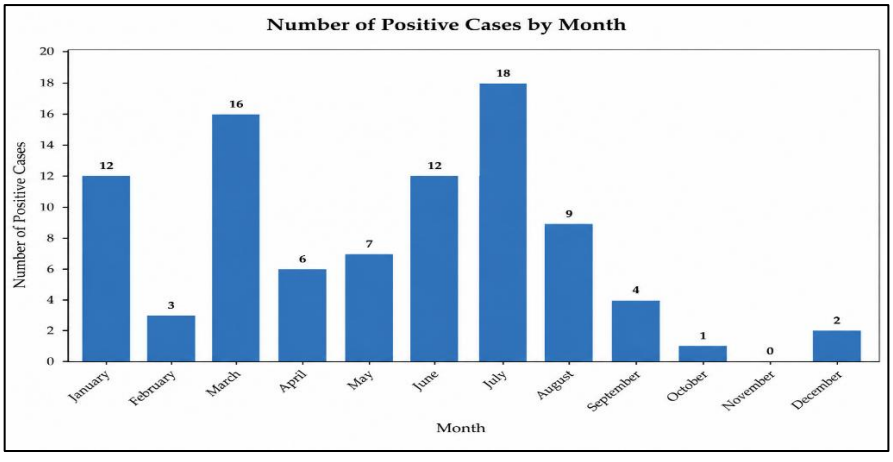


Figure 2. Monthly distribution of positive samples.

Sex distribution of the 90 records labelled positive, 53 were male and 37 were female. The table reports within-sex positivity proportions of 38.69% (53/137) for males and 29.13% (37/127) for

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

females. These figures describe the records in this hospital dataset and do not, by themselves, establish that sex is associated with positivity. No inferential comparison is reported here, and possible differences in who attended the laboratory, the reasons for testing, or other characteristics were not assessed. The results therefore support a descriptive statement about this sample only, not a conclusion that males in the wider community have a higher risk.

Table 3. Distribution of recorded results by sex.

Sex	Total samples	Positive	Negative	Positive within sex (%)
Male	**137**	**53**	**84**	**38.69%**
Female	**127**	**37**	**90**	**29.13%**
Total	**264**	**90**	**174**	**34.09%**

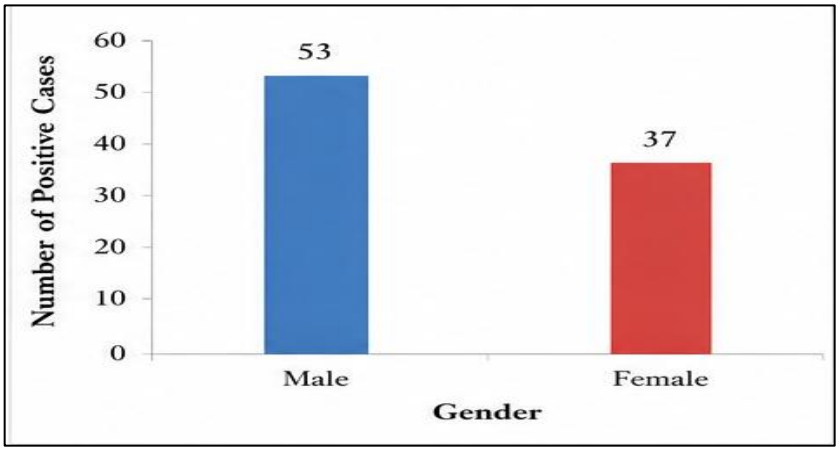


Figure 3. Gender distribution among positive records (n = 90).

Table 4 and figure 4 presents the distribution of stool examination samples according to age group and the number and percentage of (Entamoeba histolytica) positive cases. The highest positivity proportion was observed among patients aged 6–14 years, with 28 positive cases out of 50 sample (56.00%) and the lowest positivity proportion was observed among patients aged 45 years and older, with 17 positive cases out of 67 sample (25.37%)

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

Table 4. Distribution of recorded results by age group.

Age group	Total records	Positive	Negative	Positive within group (%)
<1 year	50	14	36	28.00%
1–5 years	49	13	36	26.53%
6–14 years	50	28	22	56.00%
15–44 years	48	23	25	47.9%
≥45 years	67	17	50	25.37%

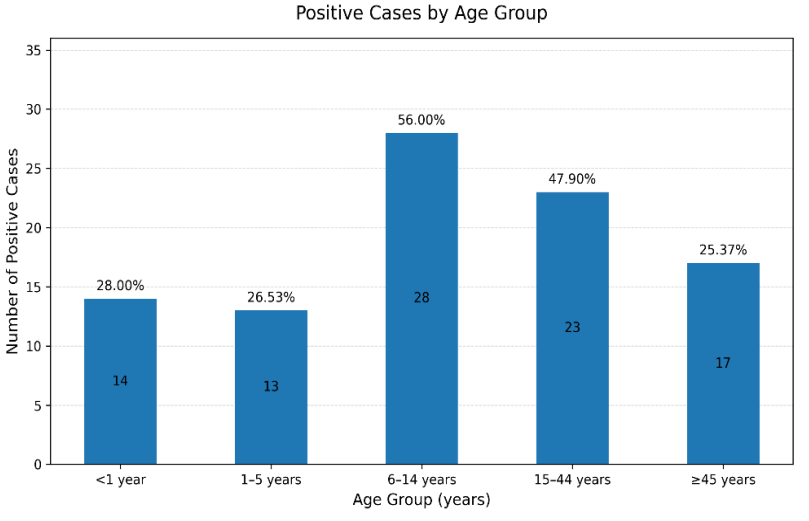


Figure 4. Percentage of positive samples within each specified age group.

Decision-tree analysis. A decision tree was explored to examine whether recorded demographic and stool-sample variables could classify the recorded laboratory outcome. The analysis is exploratory: branches represent patterns in this dataset and must not be interpreted as causal relationships. The recorded outcome was defined as negative when the parasite field contained “NO” and positive for other values, following the coding rule described in the Methods. The model also depends on the completeness and accuracy of age, month, sex, mucus, and stool-colour entries.

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

In the decision tree model Figure 5, the root node was restricted by age, and the subdivision from the dataset was selected as follows: age ≤ 5.5 versus age > 5.5 on the recorded numerical scale. Separate surface decision trees were then constructed within each age branch using mucus condition, stool color, month, and sex. The numbers displayed within each node represent the records that reached that node and their positive/negative distribution.

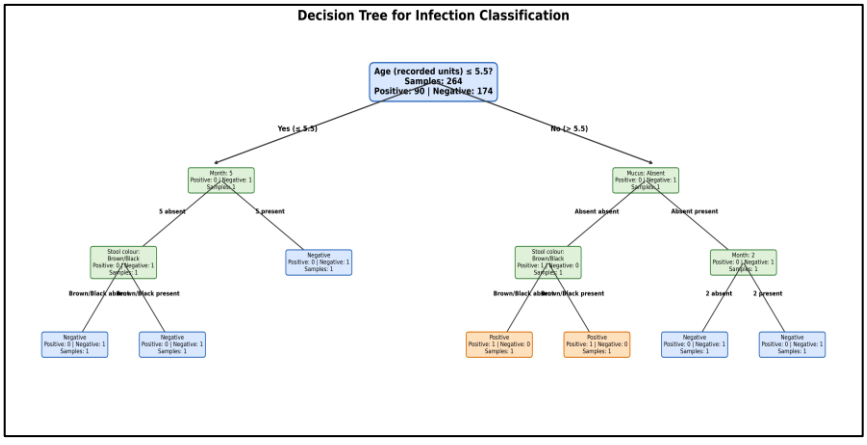


Figure 5. Decision-tree model with age as the root split.

Model performance. The reported metrics were accuracy 65.9%, precision 50.0%, sensitivity (recall) 12.2%, and F1-score 19.6%. The confusion matrix shown in Figure 6 contains 11 true positives, 79 false negatives, 11 false positives, and 163 true negatives. These counts yield an accuracy of 65.9%, precision of 50.0%, sensitivity of 12.2%, and an F1-score of approximately 19.6%. Thus, the model correctly classified many negative records but missed most records labelled positive. Accuracy alone would consequently give an incomplete impression of performance. Given the low sensitivity and large number of false negatives, this model should not be used for clinical diagnosis or screening. Its role in this study is limited to exploratory analysis, and any future predictive use would require verified labels, transparent preprocessing, and robust validation.

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques

http://www.doi.org/10.62341/istj-vol39-2-ys34

Table 5: performance metrics for the decision tree model

Metric	Result	Formula	Interpretation
Accuracy	**65.9%**	**(TP + TN) / (TP + TN + FP + FN)**	**Proportion of all test records classified correctly.**
Precision (Positive class)	**50.0%**	**TP / (TP + FP)**	**Proportion of predicted positive cases that are truly positive.**
Recall/ Sensitivity (Positive class)	**12.2%**	**TP / (TP + FN)**	**Proportion of true positive cases identified by the model.**
F1-score (Positive class)	**19.6%**	**2TP / (2TP + FP + FN)**	**Harmonic mean of precision and sensitivity.**

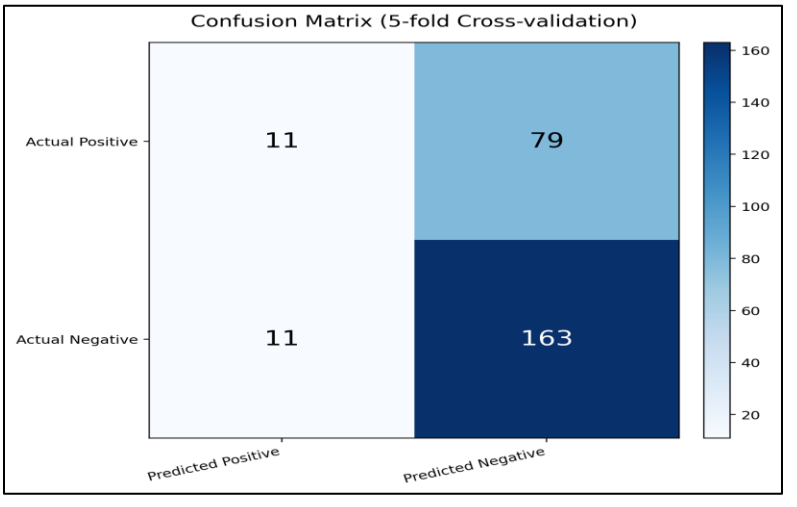


Figure 6: Confusion matrix from stratified five-fold cross-validation. Rows are actual classes and columns are predicted classes.

Confusion-matrix interpretation. With rows representing actual classes (positive, negative) and columns representing predicted classes (positive, negative), the reported matrix is [[11, 79], [11, 163]]. Therefore, 11 positive records were correctly classified, 79 positive records were classified as negative, 11 negative records were classified as positive, and 163 negative records were correctly

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques
http://www.doi.org/10.62341/istj-vol39-2-ys34

classified. This interpretation is consistent with the reported performance metrics. The manuscript should clearly state how these predictions were generated (for example, cross-validation or a held-out test set) and ensure that the same evaluation procedure is described consistently in the Methods, figure caption, and Results

6. Comparing the results with previous studies

Comparison with previous Libyan findings. In the present hospital dataset, 90 of 264 records were labelled positive (34.09%). The cited Bani Waleed study reported 55 positive samples among 180 examined (30.56%), while the cited study from Al-Zawiya, Surman, and Sabratha reported an overall positivity of 29.2% among 600 stool samples [9,18]. The numerical difference between the present estimate and the Bani Waleed estimate is 3.53 percentage points; the difference from the western Libya estimate is 4.89 percentage points. These are descriptive differences only. The manuscript does not report a statistical test comparing the studies, and the studies differ in time period, location, patient selection, sample size, and potentially diagnostic procedures. It is therefore not justified to state that the differences are statistically significant or to infer a temporal increase or decrease in infection. The present estimate should be understood as the proportion of positive records among patients tested at Al-Zawiya Hospital during the study period, not as a population-wide prevalence estimate.

Monthly patterns in context. The largest monthly count in the present dataset was recorded in July (18 positive records), followed by March (16); the smallest non-zero count was in October (1), while November had zero positive records. The Bani Waleed study reported its highest count in June (15 cases) during its January–June observation period [18]. The difference in peak month means that the two datasets do not show an identical monthly pattern. However, raw counts are influenced by the number of people tested in each month, and the studies cover different periods and settings. Consequently, these counts cannot establish a common seasonal effect or explain why positive records occurred in particular months. Monthly positivity proportions, comparable sampling, and

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques
http://www.doi.org/10.62341/istj-vol39-2-ys34

appropriate statistical analysis would be needed for a stronger comparison.

Diagnostic interpretation. The cautious interpretation adopted here is consistent with reviews describing the limitations of microscopy for distinguishing E. histolytica from morphologically similar Entamoeba species, including E. dispar and E. moshkovskii [3–5,8]. Unless species-specific antigen testing or PCR was performed and documented, the 34.09% result should be reported as the proportion of samples recorded as positive by the laboratory method used, rather than as a confirmed prevalence of E. histolytica. Differences in diagnostic methods may contribute to variation between studies, but the available data do not allow the contribution of diagnostic method to be quantified. Overall, the local records are useful for describing the hospital laboratory caseload; they do not establish causal risk factors, community prevalence, or a seasonal mechanism.

7. Conclusion

This retrospective study summarized 264 stool-examination samples from patients attending Al-Zawiya Hospital, Libya, during January 2020–August 2021. Ninety records (34.09%) were labelled positive and 174 (65.91%) negative. The largest monthly positive count was in July (18 records). The recorded data also showed positive results among males and females, while the age-group table indicated a high reported positivity proportion in children aged 6–14 years.

The decision tree model achieved an overall accuracy of 65.9%. The confusion matrix indicated that many positive cases were incorrectly classified as negative, limiting the model's usefulness in identifying positive cases. Therefore, the model should be considered exploratory in its current form.

8. Recommendations

Based on the findings and limitations of this study, the following recommendations are proposed:

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques
http://www.doi.org/10.62341/istj-vol39-2-ys34

1. Confirm parasite identification: Where feasible, antigen-detection assays or species-specific PCR should be used to distinguish (Entamoeba histolytica) from morphologically similar Entamoeba species.
2. Improve laboratory data collection: Future studies should ensure complete and consistent recording of demographic characteristics, examination dates, laboratory findings, and diagnostic results.
3. Expand epidemiological investigations: Larger studies involving multiple healthcare facilities and longer observation periods are recommended to provide a broader understanding of Entamoeba positivity in western Libya.
4. Improve decision-tree performance: Future analyses should evaluate alternative classification algorithms, appropriate class-imbalance handling, and systematic parameter optimization to improve the identification of positive cases.
5. Strengthen model validation: The dataset, preprocessing procedures, classification rules, and confusion matrix should be checked for consistency. Model performance should be assessed using stratified cross-validation and, where feasible, an independent test dataset.
6. Evaluate clinically relevant performance measures: Future studies should report sensitivity, specificity, precision, F1-score, and confusion matrices alongside accuracy, with particular attention to false-negative classifications.
7. Support preventive public-health measures: Continued health education on hand hygiene, food safety, safe drinking water, and adequate sanitation is recommended to help reduce the risk of faecal–oral transmission.

References

[1] Stanley, S. L., Jr. (2003). Amoebiasis. The Lancet, 361(9362), 1025–1034.
[2] Haque, R., Huston, C. D., Hughes, M., Houpt, E., & Petri, W. A., Jr. (2003). Amebiasis. New England Journal of Medicine, 348(16), 1565–1573. https://doi.org/10.1056/NEJMra022710

[3] Shirley, D. A. T., Farr, L., Watanabe, K., & Moonah, S. (2018). A review of the global burden, new diagnostics, and current therapeutics for amebiasis. Open Forum Infectious Diseases, 5(7), ofy161. https://doi.org/10.1093/ofid/ofy161

[4] Tanyuksel, M., & Petri, W. A., Jr. (2003). Laboratory diagnosis of amebiasis. Clinical Microbiology Reviews, 16(4), 713–729. https://doi.org/10.1128/CMR.16.4.713-729.2003

[5] Fotedar, R., Stark, D., Beebe, N., Marriott, D., Ellis, J., & Harkness, J. (2007). Laboratory diagnostic techniques for Entamoeba species. Clinical Microbiology Reviews, 20(3), 511–532. https://doi.org/10.1128/CMR.00004-07

[6] Centers for Disease Control and Prevention. (n.d.). DPDx: Amebiasis. https://www.cdc.gov/dpdx/amebiasis/

[7] Qvarnstrom, Y., James, C., Xayavong, M., Holloway, B. P., Visvesvara, G. S., Sriram, R., & da Silva, A. J. (2005). Comparison of real-time PCR protocols for differential laboratory diagnosis of amebiasis. Journal of Clinical Microbiology, 43(11), 5491–5497. https://doi.org/10.1128/JCM.43.11.5491-5497.2005

[8] Ghenghesh, K. S., Ghanghish, K., BenDarif, E. T., Shembesh, K., & Franka, E. (2016). Prevalence of Entamoeba histolytica, Giardia lamblia, and Cryptosporidium spp. in Libya: 2000–2015. Libyan Journal of Medicine, 11, 32088. https://doi.org/10.3402/ljm.v11.32088

[9] Al-Qamoudi, R. S. (2025). Study of the prevalence of Entamoeba histolytica infection in western Libya (Al-Zawiya–Surman–Sabratha). International Science and Technology Journal, 37 (July 2025). https://mail.stcrs.com.ly/istj/en/article-details.php?id=1263

[10] Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. Wadsworth.

[11] Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann.

A Study of the Prevalence of the Parasite Entamoeba histolytica Among Patients Attending Al-Zawiya Hospital, Using Descriptive Statistical Analysis and Data Mining Techniques
http://www.doi.org/10.62341/istj-vol39-2-ys34

[12] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques (4th ed.). Morgan Kaufmann.

[13] Ximénez, C., Morán, P., Rojas, L., Valadez, A., & Gómez, A. (2009). Reassessment of the epidemiology of amebiasis: State of the art. Infection, Genetics and Evolution, 9(6), 1023–1032.

[14] Servián, A., Helman, E., Iglesias, M. del R., Panti-May, J. A., Zonta, M. L., & Navone, G. T. (2022). Prevalence of human intestinal Entamoeba spp. in the Americas: A systematic review and meta-analysis, 1990–2022. Pathogens, 11(11), 1365. https://doi.org/10.3390/pathogens11111365

[15] Haque, R., Ali, I. K. M., Akther, S., & Petri, W. A., Jr. (1998). Comparison of PCR, isoenzyme analysis, and antigen detection for diagnosis of Entamoeba histolytica infection. Journal of Clinical Microbiology, 36(2), 449–452. https://doi.org/10.1128/JCM.36.2.449-452.1998

[16] Khairnar, K., & Parija, S. C. (2007). A novel nested multiplex polymerase chain reaction (PCR) assay for differential detection of Entamoeba histolytica, E. moshkovskii and E. dispar DNA in stool samples. BMC Microbiology, 7, 47. https://doi.org/10.1186/1471-2180-7-47

[17] Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

[18] Alashhab Barnous, M. J. (2023). Study of the prevalence of the parasite Entamoeba histolytica, which causes amoebic dysentery, in the city of Bani Waleed. Journal of Bani Waleed University for Humanities and Applied Sciences, (30), 28–39.

[19] Petri, W. A., Haque, R., & Lyerly, D. (2000). Estimating the impact of amoebiasis. Parasitology Today, 16(8), 320–321.

[20] Centers for Disease Control and Prevention. (n.d.). DPDx: Diagnostic procedures—Stool specimens, antigen detection. https://www.cdc.gov/dpdx/diagnosticprocedures/stool/antigendetection.html