International Journal of Multidisciplinary Studies | Vol 6, No. 1, Dec 2024
ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

PREDICTING URINARY TRACT INFECTION BASED ON URINALYSIS TEST RESULTS OF PATIENTS USING BINARY LOGISTIC REGRESSION ANALYSIS

Junel Mirontos[1], Mark Jhon Oxillo[1] and John Vianne Murcia[1,2]

[1]College of Development Management, University of Southeastern Philippines, Davao City
[2]Jose Maria College Foundation, Inc., Davao City

Received: June 05, 2024; Revised: June 27, 2024
Accepted: July 20, 2024; Published: Dec 05, 2024

ABSTRACT

This research aimed to create a model that could predict how likely clinic patients are diagnosed with Urinary Tract Infections (UTIs) using logistic regression analysis. A secondary data acquired from Kaggle, an online platform of datasets, with 1,436 instances from patients of a local clinic, was used in this research. The majority of these patients were female and belonged to the 0 to 9.21 age group. The logistic regression analysis found significant predictors of urinary tract infections (UTIs), such as urine specific gravity, urine transparency, and the presence (moderate to plenty) of bacteria in the urine. The prediction model was able to correctly categorize 95.1% of the UTI diagnoses of patients at a local clinic, whether positive or negative for UTI. It also noted that the urine specific gravity is an extreme predictor of UTI, indicating a further study along with other urinalysis test results. The findings of this study contribute to the wider concept of UTI prediction. Furthermore, it suggested to enhance the monitoring and assessment of significant predictors and utilization of advanced technologies for a more accurate UTI diagnosis.

Keywords: *urinary tract infection, local clinic patients, binary logistic regression analysis, Kaggle*

INTRODUCTION

Urinary Tract Infection (UTI) is a global health concern, and among the most prevalent bacterial infections, affecting millions of individuals. In fact, it accounts for 1 to 6 percent of medical referrals in a year, and around 150 million people succumb to death due to UTIs and related infections every year

ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

(Azami et al., 2019). UTI is most common among those with diabetes (Mama et al., 2019), has a higher incidence among women (Vasudevan, 2014), and is significantly more common to elderly men than in younger men (Schaeffer & Nicolle, 2016). UTIs, if untreated, can result in complications (Geerlings, 2016). Hence, prompt and accurate prediction of it is essential for timely intervention to prevent complications, reduce costs in healthcare, and improve patient outcomes (Schaeffer & Nicolle, 2016).

To correctly manage and treat UTIs, conducting rapid and accurate diagnosis is imperative (Arienzo et al., 2020). Laboratory tests like urinalysis are one of the oldest and most commonly used methods for determining the existence, severity, and course of infections of the kidney and urinary tract (Fogazzi & Garigali, 2008). Urinalysis examines a urine sample and, if properly collected, can produce substantial information about possible infections (Brunzel, 2021). Physical, chemical, and microscopic analyses comprise a complete urinalysis, which must be examined within two hours of midstream clean collection (Simerville et al., 2005). Nevertheless, while urinalysis is the commonly performed test for UTI, interpreting the results can be challenging and subjective, which often requires additional comprehensive diagnostic confirmation (Hurlbut & Littenberg, 1991; Chu & Lowder, 2018; Onyango et al, 2024), thus, making it extremely difficult for long-term care (LTC) providers to diagnose UTIs and decide when to begin antimicrobial medication (Nace et al, 2014). Furthermore, the introduction of machine/deep learning classification techniques like logistic regression analysis offers viability in enhancing the accuracy and reliability of urinalysis test results (Zeb et al., 2020). These approaches can help determine patterns and correlations that are helpful for early detection and management of UTIs.

However, although there are extensive researches about urinalysis and UTI that provide significant impacts, there remains a notable research gap comprehensively looking at several factors affecting timely and accurate diagnoses of UTI, particularly within local clinic settings. This study seeks to bridge this gap by exploring the complex relationship between demographics of patients and urinalysis test results and developing a predictive model for the occurrence of UTIs using logistic regression analysis of urinary test results. Logistic regression, a well-known statistical tool, is the best approach to employ to model the relationship between two dependent variables (Park, 2013; Peng et al., 2002; Wiest et al., 2015) — in this case, the presence or absence of UTI — and a data set of independent variables, which is the urinalysis test results. By utilizing this analysis, the researchers will be able to evaluate the overall accuracy of UTI diagnoses within the local clinic and determine significant predictors influencing the likelihood of having a UTI, aiming to help healthcare providers make informed decisions based on the urinalysis data.

This study will help provide valuable insights that will enrich the existing body of knowledge and encourage additional research in this field. The findings of this research intend to guide the development of targeted strategies and interventions that can improve the accuracy and timeliness of UTI diagnoses in local clinics, which will help improve patient outcomes and lower healthcare costs. Addressing this challenge is essential for establishing more effective healthcare practices in the region.

METHOD

Research Design

The research employed a quantitative research design, as it is fitting when collecting data that will later undergo statistical treatments (Creswell, 1994) – logistic regression analysis in the case of this research study. Furthermore, quantitative tools are commonly used when working with prediction or when the research aims to predict something in the future (Swanson & Holton, 2005). This study applies the said design with the intention of predicting likelihood of urinary tract infections based on certain variables present in the urinalysis test results of patients.

Dataset

The dataset used in this research is the *Urinalysis Test Results* from a local clinic in Northern Mindanao, Philippines, which was acquired from *Kaggle*, a popular online platform for datasets. This dataset is composed of 1,436 instances with 16 attributes where one attribute (Diagnosis) is defined as the class attribute. This class attribute pertains to UTI prediction with two possible outcomes: *Positive*, which indicates a positive UTI Diagnosis, and *Negative*, which indicates a negative UTI diagnosis (Ozkan, Koklu & Sert, 2018).

Table 1. *List of attributes of the UTI prediction dataset*

Attribute	Type	Description
Age	numeric	The age of the patient
Gender	nominal	The gender of the patient
Color	nominal	The urine color
Transparency	nominal	The urine transparency
Glucose	nominal	The type of sugar and its presence in the urine which is an important indicator of certain health conditions
Protein	nominal	The presence of protein in the urine which can be the parameters examined to assess kidney function and detect potential
pH	numeric	The measurement of the acidity or alkalinity level of urine
Specific Gravity	numeric	The measure of the concentration of particles in urine compared to water

International Journal of Multidisciplinary Studies | Vol 6, No. 1, Dec 2024
ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

WBC	nominal	White blood cells (WBC), also known as leukocytes, a crucial part of the immune system
RBC	nominal	Red Blood Cells (RBC), responsible for carrying oxygen throughout the body
Epithelial Cells	nominal	The cells that line the surfaces and cavities of the body, including the urinary tract
Mucous Threads	nominal	The strands of mucus that can be present in urine
Amorphous Urates	nominal	The non-crystalline formations in the urine that consist of uric acid
Bacteria	nominal	The presence of bacteria in the urine
Diagnosis	nominal	Classification whether the UTI diagnosis is Positive or Negative

.

Data Analysis

The researchers used descriptive statistics to measure the central tendency and the variation of scale variables (Age, pH, Specific Gravity). At the same time, the categorical indicators were accounted for through the determination of their relative frequency. To help identify whether these variables significantly predict a positive or negative UTI Diagnosis, the researchers employed Binary Logistic Regression. Moreover, the odds ratio was also determined and is presented as exponentiated betas in the logistic regression table with a 95% confidence level. All data in this research were analyzed using Jamovi version 2.4.11, a free-to-download statistical software providing up-to-date developments in statistical methodology (Jamovi Project, 2023).

RESULTS

Table 2 summarizes the demographic distribution of patients at the local clinic, providing insights into gender and age segmentation. The data reveals that the majority of the patients are female, accounting for 56.8% (n = 816) of the total sample (N = 1436), while males represent 43.2% (n = 620). This indicates a slight predominance of female patients in the clinic's clientele.

In terms of age, the largest proportion of patients falls within the 0 to 9.21 age group, comprising 35% (n = 499) of the sample, suggesting that a significant number of clinic visits are pediatric-related. The second-largest group, aged 18.41 to 27.61 years, accounts for 14% (n = 201), followed by patients aged 27.61 to 36.81 years at 10% (n = 147). Other age groups show a gradual decline in representation, with the 36.81 to 46.00 and 46.00 to 55.20 brackets each constituting 8% of the total. The smallest segment consists of

ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

patients aged 82.80 to 92.00 years, who represent only 1% (n = 12).

Table 2. *Distribution of patients according to demographic segments (N=1436)*

Demographic Segment	Counts (N)	% of Total
Gender		
Female	816	56.8%
Male	620	43.2%
Age		
0 to 9.21	499	35%
9.21 to 18.41	136	9%
18.41 to 27.61	201	14%
27.61 to 36.81	147	10%
36.81 to 46.00	108	8%
46.00 to 55.20	109	8%
55.20 to 64.40	99	7%
64.40 to 73.60	81	6%
73.60 to 82.80	44	3%
82.80 to 92.00	12	1%

Table 3 presents the frequency distribution of urine color among 1,435 patients at a local clinic, offering insights into common variations. The data highlights that the majority of patients (49.5%, n = 710) exhibited yellow-colored urine, which aligns with normal hydration levels and is typically indicative of general health. The second most prevalent category was light yellow, accounting for 23.8% (n = 341) of the sample. Together, these categories comprise a substantial majority (73.3%), suggesting that most patients demonstrated urine colors within a normal or hydrated range.

Table 3. *Frequency distribution of urine color (N=1435)*

Color	Counts (N)	% of Total
amber	15	1.0 %
brown	1	0.1 %
dark yellow	248	17.3 %
light red	1	0.1 %
light yellow	341	23.8 %
red	1	0.1 %
reddish	1	0.1 %
reddish yellow	1	0.1 %
straw	116	8.1 %
yellow	710	49.5 %

Dark yellow urine was observed in 17.3% (n = 248) of patients, potentially indicating mild dehydration or concentrated urine, which could warrant further attention in a clinical context. Other colors, such as straw (8.1%, n = 116), were less common but also considered variations within normal ranges. Uncommon colors, including amber (1.0%, n = 15), brown (0.1%, n = 1), light red (0.1%, n = 1), red (0.1%, n = 1), reddish (0.1%, n = 1), and reddish yellow (0.1%, n = 1), were rare occurrences and might suggest underlying conditions requiring further medical evaluation.

Table 4 illustrates the frequency distribution of urine transparency among 1,436 patients at a local clinic, providing valuable insights into hydration and potential health conditions. The majority of patients (78.3%, n = 1,124) had clear urine, which is often indicative of good hydration and normal kidney function. This finding underscores the general health status of the sample population, as clear urine is a positive marker for overall wellness. The second most reported category was slightly hazy urine, observed in 12.0% (n = 172) of the patients. Slight haziness can occur due to mild dehydration or the presence of natural sediment and is typically not a cause for concern without additional symptoms. Hazy urine, recorded in 7.2% (n = 104) of the cases, may suggest slightly elevated levels of proteins or other substances, warranting closer monitoring in clinical assessments.

Less common transparency levels include turbid urine (1.4%, n = 20) and cloudy urine (1.1%, n = 16). These findings, though rare, could indicate the presence of urinary tract infections, sediment, or other underlying medical conditions requiring further diagnostic evaluation.

.

Table 4. *Frequency distribution of urine transparency (N=1436)*

Transparency	Counts (N)	% of Total
clear	1124	78.3 %
cloudy	16	1.1 %
hazy	104	7.2 %
slightly hazy	172	12.0 %
turbid	20	1.4 %

Table 5 presents the frequency distribution of glucose presence in urine samples from 1,436 patients at a local clinic. The results indicate that the vast majority of patients (93.9%, n = 1,349) tested negative for glucose, suggesting normal glucose metabolism and kidney function in most individuals within the dataset. This finding aligns with typical population health norms, where glucose is not expected to appear in urine under normal physiological conditions.

A small proportion of patients exhibited positive glucose levels, distributed across varying concentrations: 1+ (1.0%, n = 15), 2+ (1.7%, n = 24), 3+ (1.6%, n = 23), and 4+ (0.8%, n = 12). The trace level of glucose was identified in 0.9% (n = 13) of the samples. These findings, though minor in frequency, warrant attention as the presence of glucose in urine (glycosuria) may indicate underlying conditions such as hyperglycemia, diabetes mellitus, or kidney impairment. Higher glucose concentrations (3+ and 4+) could signify more advanced metabolic disturbances requiring further diagnostic evaluation.

Table 5. *Frequency distribution of presence of glucose (N=1436)*

Glucose	Counts (N)	% of Total
1+	15	1.0 %
2+	24	1.7 %
3+	23	1.6 %
4+	12	0.8 %
Negative	1349	93.9 %
Trace	13	0.9 %

Table 6 presents the frequency distribution of protein presence in urine samples from 1,436 patients at a local clinic. The results show that over half of the samples (56.0%, n = 804) tested negative for protein, indicating that most individuals in the dataset have normal kidney function, as proteinuria is not typically observed in healthy individuals. Additionally, a substantial proportion of samples (34.3%, n = 492) exhibited trace levels of protein, which may be considered within normal limits under certain physiological conditions, such as exercise or dehydration.

A smaller subset of patients had elevated protein levels: 1+ (6.5%, n = 94), 2+ (2.9%, n = 41), and 3+ (0.3%, n = 5). These findings suggest varying degrees of proteinuria, which can be indicative of kidney dysfunction or other underlying health conditions, such as hypertension, diabetes, or urinary tract infections. The higher levels (2+ and 3+) may require further clinical investigation to assess the potential severity and causes of proteinuria (Cravedi & Remuzzi, 2013).

Table 6. *Frequency distribution of presence of protein (N=1436)*

Protein	Counts (N)	% of Total
1+	94	6.5 %
2+	41	2.9 %
3+	5	0.3 %
Negative	804	56.0 %
Trace	492	34.3 %

It is noteworthy that attributes such as white blood cells (WBC) and red blood cells (RBC) were excluded from the analysis due to their potential to introduce perfect collinearity. According to Dormann et al. (2013), removing highly correlated variables is necessary to reduce collinearity and ensure robust analytical models. This approach enhances the clarity and reliability of the findings.

Table 7 shows the frequency distribution of epithelial cells in urine samples from 1,436 patients at a local clinic. The results indicate that a rare presence of epithelial cells was observed in the majority of the samples (51.7%, n = 742), which is typically considered normal and suggests minimal urinary tract shedding or contamination. This finding aligns with standard urinalysis interpretations, where low levels of epithelial cells are expected in healthy individuals (Markus et al., 2021).

Other categories revealed varying degrees of epithelial cell presence. A smaller portion of samples indicated few epithelial cells (24.2%, n = 347), moderate levels (13.1%, n = 188), and plenty (8.4%, n = 121). Occasional (1.3%, n = 19) and loaded epithelial cells (0.2%, n = 3) were less frequently observed, suggesting the possibility of inflammation, infection, or contamination in some cases. Notably, a negligible number of samples (1.1%, n = 16) showed no epithelial cells, which could reflect optimal sample collection without contamination from the urinary tract lining.

Table 7. *Frequency distribution of presence of epithelial cells (N=1436)*

Epithelial Cells	Counts (N)	% of Total
few	347	24.2 %
loaded	3	0.2 %
moderate	188	13.1 %
none seen	16	1.1 %
occasional	19	1.3 %
plenty	121	8.4 %
rare	742	51.7 %

Table 8 presents the frequency distribution of mucous threads in the urine samples from 1,436 patients at a local clinic. The results indicate that the majority of samples (34.8%, n = 500) showed no visible mucous threads ("None Seen"), suggesting that most patients had no significant urinary tract irritation or inflammation. This finding aligns with normal urinalysis expectations, where mucous threads are typically minimal or absent in healthy individuals.

Table 8. Frequency Distribution of Presence of Mucous Threads (N=1436)

Mucous Threads	Counts (N)	% of Total
few	381	26.5 %
moderate	222	15.5 %
none seen	500	34.8 %
occasional	21	1.5 %
plenty	36	2.5 %
rare	276	19.2 %
few	381	26.5 %

Moreover, among the samples with mucous threads, "Few" was the most frequently observed category, accounting for 26.5% (n = 381) of the total. "Rare" mucous threads were present in 19.2% (n = 276), while "Moderate" levels were found in 15.5% (n = 222). Less commonly, "Plenty" (2.5%, n = 36) and "Occasional" (1.5%, n = 21) mucous threads were observed, potentially

ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

indicating mild to moderate urinary tract irritation or inflammation.

Table 9 presents the frequency distribution of amorphous urates in the urine samples from 1,436 patients at a local clinic. The analysis reveals that the vast majority of samples (89.4%, n = 1,284) exhibited no visible signs of amorphous urates ("None Seen"), indicating that the majority of the patients likely had normal urinary composition and no significant precipitation of urate salts.

Table 9. *Frequency distribution of presence of amorphous urates (N=1436)*

Amorphous Urates	Counts (N)	% of Total
few	72	5.0 %
moderate	19	1.3 %
none seen	1284	89.4 %
occasional	8	0.6 %
plenty	11	0.8 %
rare	42	2.9 %

Among the samples with amorphous urates, "few" was the most frequently observed category, accounting for 5.0% (n = 72) of the total, followed by "rare" occurrences in 2.9% (n = 42) of the samples. Other categories, such as "moderate" (1.3%, n = 19), "plenty" (0.8%, n = 11), and "occasional" (0.6%, n = 8), were observed less frequently. These findings suggest that amorphous urates are rarely present in clinically significant amounts in the studied population. The presence of amorphous urates in urine can be influenced by factors such as hydration levels, diet, and urinary pH. While small amounts of urates are typically benign and may result from transient changes in urine concentration, higher levels ("moderate" or "plenty") could indicate underlying metabolic conditions such as hyperuricosuria or dehydration (Skrajnowska & Bobrowska-Korczak, 2024).

Table 10 outlines the frequency distribution of bacterial presence in the urine samples of 1,436 patients from a local clinic. The results reveal that a significant majority (52.6%, n = 755) of the samples were categorized as having "Rare" bacterial presence, indicating minimal bacterial contamination or infection in over half of the tested population.

Table 10. *Frequency distribution of presence of bacteria (N=1436)*

Bacteria	Counts (N)	% of Total
few	434	30.2 %
loaded	4	0.3 %
moderate	158	11.0 %
occasional	8	0.6 %
plenty	77	5.4 %
rare	755	52.6 %

ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

Among the other categories, 30.2% (n = 434) of the samples showed few bacteria, representing a moderate prevalence of bacterial presence that may not yet be clinically significant. Moderate bacterial presence was observed in 11.0% (n = 158) of the samples, while plenty (5.4%, n = 77), occasional (0.6%, n = 8), and loaded (0.3%, n = 4) were less frequently recorded, suggesting that severe bacterial contamination is relatively rare in this dataset.

Table 11 presents the descriptive statistics for the scale variables included in this study: age, pH, and specific gravity. The mean age of the 1,436 patients is 27.21 years (SD = 23.46), indicating substantial variability in the age distribution. This wide age range may reflect the diverse demographic composition of the local clinic's patient population and suggests that any further analyses should account for potential age-related differences in clinical outcomes.

The average pH level of the urine samples is 6.05 (SD = 0.60), with moderate variability, suggesting that most patients' urine pH falls within the normal range of 4.5 to 8.0. This result aligns with the expected physiological pH levels for healthy individuals, although variations may warrant further examination to explore their association with potential health conditions such as urinary tract infections or metabolic disorders.

The specific gravity values average 1.02 (SD = 0.007), demonstrating low variability across the patient samples. This suggests consistent urine concentration levels, indicating adequate hydration status among the majority of the clinic's patients. These descriptive findings provide a baseline understanding of the patient cohort and serve as a foundation for subsequent analyses to identify patterns or predictors related to urinary tract health, including potential correlations with infection or other medical conditions.

Table 11. *Descriptive statistics results for scaled variables*

Variable	Counts (N)	Mean	SD
age	1436	27.21	23.460
pH	1436	6.05	0.598
specific gravity	1436	1.02	0.007

Table 12 presents the results of the binary logistic regression analysis examining predictors of urinary tract infections (UTIs) among patients of a local clinic. The model was statistically significant, $\chi^2(47) = 184$, $p<0.05$, indicating that the set of predictors collectively contributes to predicting UTI diagnoses. The model explained between 12.0% (Cox and Snell R^2) and 31.4% (Nagelkerke R^2) of the variance in UTI outcomes and achieved an overall classification accuracy of 95.1%. It correctly predicted 19.8% of UTI-positive cases (sensitivity) and 99.6% of UTI-negative cases (specificity), highlighting its strong performance in identifying negative cases but relatively weaker sensitivity for positive cases.

International Journal of Multidisciplinary Studies | Vol 6, No. 1, Dec 2024
ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

Table 12. *Binary logistic regression showing all possible predictors of UTIs*

Predictor	Estimate	SE	Z	p	Odds ratio
Intercept	-58.95573	31.29989	-1.88358	0.06	2.49e-26
Age	0.00433	0.00676	0.64015	0.522	1.0043
pH	0.31056	0.24241	1.28115	0.2	1.3642
Specific Gravity	60.91299	29.93874	2.03459	0.042*	2.85e+26
Gender:					
Male – Female	0.53082	0.3623	1.46513	0.143	1.7003
Color:					
Brown – Amber	15.89236	3956.1806	0.00402	0.997	7.98e0+6
Dark Yellow – Amber	0.68944	1.16977	0.58939	0.556	1.9926
Light Red – Amber	14.90096	3956.1806	0.00377	0.997	2.96e0+6
Light Yellow – Amber	1.43754	1.21363	1.18449	0.236	4.2103
Red – Amber	15.86142	3956.1806	0.00401	0.997	7.74e0+6
Reddish – Amber	18.32641	3956.1806	0.00463	0.996	9.10e0+7
Reddish Yellow – Amber	-15.77472	3956.1806	-0.00399	0.997	1.41e0-7
Straw – Amber	1.51194	1.33309	1.13416	0.257	4.5355
Yellow – Amber	0.77905	1.15716	0.67325	0.501	2.1794
Transparency:					
Cloudy – Clear	-2.56079	0.83482	-3.06747	0.002*	0.0772
Hazy – Clear	-0.96121	0.46999	-2.04516	0.041*	0.3824
Slightly Hazy – Clear	-1.00855	0.39266	-2.5685	0.01*	0.3647
Turbid – Clear	-2.40557	0.71612	-3.35917	< .001*	0.0902
Glucose:					
2+ – 1+	-1.91324	1.32547	-1.44344	0.149	0.1476
3+ – 1+	-0.27518	1.63823	-0.16797	0.867	0.7594
4+ – 1+	-1.76685	1.46015	-1.21005	0.226	0.1709
Negative – 1+	-0.63653	1.13356	-0.56153	0.574	0.5291
Trace – 1+	12.96087	1060.5468	0.01222	0.99	425434.13
Protein:					
2+ – 1+	-0.09962	0.75522	-0.13191	0.895	0.9052
3+ – 1+	14.68319	1519.4536	0.00966	0.992	2.38e0+6
Negative – 1+	0.15332	0.50933	0.30102	0.763	1.1657
Trace – 1+	0.05731	0.48217	0.11887	0.905	1.059
Epithelial Cells:					
Loaded – Few	-0.65678	1.61269	-0.40726	0.684	0.5185
Moderate – Few	-0.15646	0.40551	-0.38583	0.7	0.8552
None Seen – Few	-0.81197	1.22056	-0.66525	0.506	0.444

	Occasional – Few	-0.74846	1.34263	-0.55746	0.577	0.4731
	Plenty – Few	0.24016	0.47901	0.50137	0.616	1.2715
	Rare – Few	-0.33611	0.39156	-0.85839	0.391	0.7145
Mucous Threads:						
	Moderate – Few	-0.43627	0.43706	-0.99818	0.318	0.6464
	None Seen – Few	-0.39868	0.39276	-1.01507	0.31	0.6712
	Occasional – Few	-1.52308	0.98917	-1.53976	0.124	0.218
	Plenty – Few	0.05159	0.85956	0.06002	0.952	1.0529
	Rare – Few	-0.31864	0.46143	-0.69055	0.49	0.7271
Amorphous Urates:						
	Moderate – Few	-0.43135	1.07872	-0.39987	0.689	0.6496
	None Seen – Few	-0.37702	0.50854	-0.74138	0.458	0.6859
	Occasional – Few	13.0704	1322.8774	0.00988	0.992	474679.66
	Plenty – Few	-0.52261	1.24923	-0.41835	0.676	0.593
	Rare – Few	-0.38842	0.70343	-0.55219	0.581	0.6781
Bacteria:						
	Loaded – Few	14.63487	1849.5531	0.00791	0.994	2.27e0+6
	Moderate – Few	-3.07607	0.45865	-6.70682	< .001*	0.0461
	Occasional – Few	13.96913	1322.3789	0.01056	0.992	1.17e0+6
	Plenty – Few	-2.32662	0.52075	-4.4678	< .001*	0.0976
	Rare – Few	-0.61438	0.50735	-1.21095	0.226	0.541

Note. Estimates represent the log odds of "Diagnosis = NEGATIVE" vs. "Diagnosis = POSITIVE"

Log Likelihood = 439; Cox and Snell R^2 = 0.120; Nagelkerke R^2 = 0.341

Significant predictors of UTIs include specific gravity, transparency, and the presence of bacteria in urine samples. Specific gravity showed a substantial effect with an odds ratio of 2.85×10^{26} (p = .042), indicating that higher specific gravity levels dramatically increase the likelihood of a UTI. Transparency also emerged as a significant predictor, with cloudy, hazy, slightly hazy, and turbid transparency levels associated with decreased odds of a UTI, supported by odds ratios ranging from 0.0772 to 0.3824 ($p < .05$). This finding suggests that clearer urine samples are more likely to be linked to UTI-negative outcomes. Similarly, the presence of bacteria, particularly in moderate or plenty amounts, significantly increased the odds of a UTI diagnosis. The odds ratios for bacterial presence ranged from 0.0461 to 0.0976 ($p < .001$), indicating a high likelihood of UTI when moderate to heavy bacterial loads are observed in the urine.

Finally, table 13 presents the classification table from the binary logistic

International Journal of Multidisciplinary Studies | Vol 6, No. 1, Dec 2024
ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)
regression model, highlighting its accuracy in predicting urinary tract infections (UTIs) among the patients of a local clinic. The model demonstrated an overall classification accuracy of 95.1%, indicating its strong performance in predicting the binary outcome of the dependent variable (UTI diagnosis). Specifically, the model correctly predicted 19.8% of the cases that were positive for UTI, which suggests limited sensitivity in identifying actual positive cases. However, it accurately classified 99.6% of the cases that were negative for UTI. These findings suggest that while the model is effective in identifying non-UTI cases, its ability to detect UTI-positive cases requires improvement. This discrepancy between sensitivity and specificity may warrant further refinement of the model, such as the inclusion of additional predictors or adjustments in the decision threshold, to enhance its capability in identifying UTI-positive cases. Nonetheless, the high specificity and overall accuracy underscore the model's reliability in excluding UTI-negative cases, which is critical in clinical settings for minimizing unnecessary treatments or interventions.

Table 13. *Classification table showing the accuracy of the model in predicting UTIs*

Observed	**Predicted**		
	Positive	**Negative**	**% Correct**
Positive	16	65	19.8
Negative	5	1349	99.6
	Overall %		95.1

DISCUSSION

This research identified specific gravity, urine transparency, and bacterial presence as significant predictors of urinary tract infections (UTIs), offering valuable insights into UTI diagnosis. Specific gravity, which measures the concentration of particles in urine, emerged as a strong predictor of UTI. This finding aligns with studies by Shaikh et al. (2019) and Johnson (2023), which emphasize that specific gravity is essential for interpreting urinalysis components. Abnormal specific gravity levels can indicate an increased risk of infections or complications, such as urinary stones, which are closely associated with UTI severity (Chen et al., 2001; Conrad et al., 1991). While specific gravity alone cannot definitively predict UTIs (Shaikh et al., 2019), its abnormality serves as a vital diagnostic component when analyzed alongside other urinalysis factors (Chaudhari et al., 2017). The integration of specific gravity measurements into diagnostic workflows can improve early identification and management of UTI risk factors.

Urine transparency also significantly predicted UTIs, with cloudy or turbid urine linked to a higher likelihood of infection compared to clear urine. This finding aligns with studies by Phillips et al. (1992) and Fan and Bai (2020), which identified cloudy urine as a hallmark of infection. Although clear urine is typically associated with an absence of UTI (Pouchot et al., 1990), it cannot fully exclude the possibility of infection, as other conditions or excessive hydration might influence urine clarity (Reine & Langston, 2005). Thus, while transparency offers a high predictive value for ruling out infections, it must be interpreted cautiously and in conjunction with other diagnostic markers. The study underscores that abnormal urine transparency often reflects infection or inflammation, reaffirming its utility as a clinical indicator.

Finally, bacterial presence in urine emerged as a robust predictor of UTI, with moderate to heavy bacterial loads significantly increasing infection likelihood. This result corroborates previous research highlighting pathogenic bacteria as the primary causative agents of UTIs (Steenbeke et al., 2020; Suhartono et al., 2021). Studies have consistently demonstrated that a significant bacterial count (104–106 CFU/mL) in urine correlates with symptomatic UTIs (Vahlensieck et al., 2019; Pappas, 1991). However, the mere detection of bacteria does not always signify infection, as asymptomatic bacteriuria or contamination can yield false positives (Schulz et al., 2016). This study emphasizes the importance of interpreting bacterial presence within the clinical context, ensuring accurate differentiation between contamination, asymptomatic bacteriuria, and true infections.

CONCLUSIONS

The findings of this study create noteworthy information in the context of predicting urinary tract infections (UTIs). The logistic regression analysis employed in this research yielded key insights that helped attain the objectives of this research. The study revealed that among the significant predictors, urine specific gravity is the extreme predictor of UTI, but this extremity needs to be studied further for a more comprehensive analysis. Additionally, urine transparency is a factor towards having UTI. The more transparent the urine sample is, the less likely it is to lead to a positive UTI diagnosis. The presence of bacteria in the urine also contributes to urinary tract infections, more specifically when there is a moderate to plenty amount of it in the urine. Based on the findings of this research, age, pH level, gender, urine color, presence of glucose, protein, epithelial cells, mucous threads, and amorphous urates do not significantly predict urinary tract infections.

Finally, the ability of the model to predict UTI in 95.1% of the instances highlights the importance of considering the significant predictors in the analysis of UTI diagnosis. These findings will contribute to the enhancement of accuracy as far as UTI diagnosis is concerned.

REFERENCES

Arienzo, A., Cellitti, V., Ferrante, V., Losito, F., Stalio, O., Murgia, L., & Antonini, G. (2020). A new point-of-care test for the rapid detection of urinary tract infections. *European Journal of Clinical Microbiology & Infectious Diseases, 39,* 325–332.

Azami, M., Jaafari, Z., Masoumi, M., Shohani, M., Badfar, G., Mahmudi, L., & Abbasalizadeh, S. (2019). The etiology and prevalence of urinary tract infection and asymptomatic bacteriuria in pregnant women in Iran: A systematic review and meta-analysis. *BMC Urology, 19,* 1–15.

Banda, J. M., Cletus, D., Zakka Sheyin, S. A. J., John, B., Mohammed, S. S., & Damen, J. G. (2020). Prevalence of asymptomatic bacteriuria among pregnant women attending antenatal clinic at Plateau State Specialist Hospital, Jos, Nigeria. *Archives of Microbiology & Immunology, 4*(3), 121–130.

Brunzel, N. A. (2021). *Fundamentals of urine and body fluid analysis-e-book.* Elsevier Health Sciences.

Chen, Y., Roseman, J. M., Funkhouser, E., & DeVivo, M. J. (2001). Urine specific gravity and water hardness in relation to urolithiasis in persons with spinal cord injury. *Spinal Cord, 39*(11), 571–576.

Chu, C. M., & Lowder, J. L. (2018). Diagnosis and treatment of urinary tract infections across age groups. *American Journal of Obstetrics and Gynecology, 219*(1), 40–51.

Conrad, S., Busch, R., & Huland, H. (1991). Complicated urinary tract infections. *European Urology, 19*(Suppl 1), 16–22.

Cravedi, P., & Remuzzi, G. (2013). Pathophysiology of proteinuria and its value as an outcome measure in chronic kidney disease. *British Journal of Clinical Pharmacology*, *76*(4), 516-523.

Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches.* SAGE Publications.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46.

Fan, S. L., & Bai, S. (2020). Urinalysis. In *Contemporary Practice in Clinical Chemistry* (pp. 665-680). Academic Press.

Fogazzi, G. B., Verdesca, S., & Garigali, G. (2008). Urinalysis: Core curriculum 2008. *American Journal of Kidney Diseases, 51*(6), 1052–1067.

Geerlings, S. E. (2016). Clinical presentations and epidemiology of urinary tract infections. *Microbiology Spectrum, 4*(5), 10–1128.

Hurlbut III, T. A., & Littenberg, B. (1991). The diagnostic accuracy of rapid dipstick tests to predict urinary tract infection. *American Journal of Clinical Pathology, 96*(5), 582–588.

Jamovi Project. (2023). *jamovi* (Version 2.4) [Computer software]. Retrieved from https://www.jamovi.org.

Johnson, S. (2023). What is a urine specific gravity test? *Healthline Media UK Ltd.* Retrieved from https://www.medicalnewstoday.com/articles/322125.

Mama, M., Manilal, A., Gezmu, T., Kidanewold, A., Gosa, F., & Gebresilasie, A. (2019). Prevalence and associated factors of urinary tract infections among diabetic patients in Arba Minch Hospital, Arba Minch Province, South Ethiopia. *Turkish Journal of Urology, 45*(1), 56.

Markus, H., Zhao, J., Contente-Cuomo, T., Stephens, M. D., Raupach, E., Odenheimer-Bergman, A., ... & Murtaza, M. (2021). Analysis of recurrently protected genomic regions in cell-free DNA found in urine. *Science Translational Medicine*, *13*(581), eaaz3088.

Nace, D. A., Drinka, P. J., & Crnich, C. J. (2014). Clinical uncertainties in the approach to long-term care residents with possible urinary tract infection. *Journal of the American Medical Directors Association, 15*(2), 133–139.

Nall, R. (2023). Urine specific gravity test. *Healthline Media LLC.* Retrieved from https://www.healthline.com/health/urine-specific-gravity#results.

Onyango, H., Sloan, D. J., Keenan, K., Kesby, M., Ngugi, C., Gitonga, H., & Hammond, R. (2024). The diagnostic accuracy of the point-of-care urine dipstick test in detecting urinary tract infection among symptomatic patients in Nairobi County, Kenya. *medRxiv, 2024–05.*

Ozkan, I. A., Koklu, M., & Sert, I. U. (2018). Diagnosis of urinary tract infection based on artificial intelligence methods. *Computer Methods and Programs in Biomedicine*, *166*, 51-59.

Pappas, P. C. (1991). Laboratory in the diagnosis and management of urinary tract infections. *Medical Clinics of North America, 75*(2), 313–325.

Park, H. A. (2013). An introduction to logistic regression: From basic concepts to interpretation with particular attention to nursing domain. *Journal of Korean Academy of Nursing, 43*(2), 154–164.

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3–14.

Phillips, G., Fleming, L. W., Khan, I., & Stewart, W. K. (1992). Urine transparency as an index of absence of infection. *British Journal of Urology, 70*(2), 191–195.

Phillips, G., Fleming, L. W., Stewart, W. K., & Duncan, I. (1993). Managing urinary tract infections. *BMJ: British Medical Journal, 306*(6878), 653.

Pouchot, J., Launay, I., Cahen, P., Boussougant, Y., Vinceneux, P., Hook, E., & Shread, P. (1990). Significance of crystal-clear urine. *The Lancet, 336*(8710), 320–321.

Reine, N. J., & Langston, C. E. (2005). Urinalysis interpretation: how to squeeze out the maximum information from a small sample. *Clinical Techniques in Small Animal Practice*, *20*(1), 2-10.

ISSN No: 2704-4327 (online)
ISSN No: 2651-768X (print)

Schaeffer, A. J., & Nicolle, L. E. (2016). Urinary tract infections in older men. *New England Journal of Medicine, 374*(6), 562–571.

Schnarr, J., & Smaill, F. (2008). Asymptomatic bacteriuria and symptomatic urinary tract infections in pregnancy. *European Journal of Clinical Investigation, 38,* 50–57.

Schulz, L., Hoffman, R. J., Pothof, J., & Fox, B. (2016). Top ten myths regarding the diagnosis and treatment of urinary tract infections. *The Journal of Emergency Medicine, 51*(1), 25–30.

Shaikh, N., Shope, M., & Kurs-Lasky, M. (2019). Urine specific gravity and the accuracy of urinalysis. *Pediatrics, 144*(5), e20190467.

Simerville, J. A., Maxted, W. C., & Pahira, J. J. (2005). Urinalysis: A comprehensive review. *American Family Physician, 71*(6), 1153–1162.

Skrajnowska, D., & Bobrowska-Korczak, B. (2024). The effects of diet, dietary supplements, drugs and exercise on physical, diagnostic values of urine characteristics. *Nutrients*, *16*(18), 3141.

Steenbeke, M., De Bruyne, S., Boelens, J., Oyaert, M., Glorieux, G., Van Biesen, W., & Speeckaert, M. M. (2020). Exploring the possibilities of infrared spectroscopy for urine sediment examination and detection of pathogenic bacteria in urinary tract infections. *Clinical Chemistry and Laboratory Medicine (CCLM), 58*(10), 1759–1767.

Suhartono, S., Mahdani, W., Hayati, Z., & Nurhalimah, N. (2021). Species distribution of Enterobacteriaceae and non-Enterobacteriaceae responsible for urinary tract infections at the Zainoel Abidin Hospital in Banda Aceh, Indonesia. *Biodiversitas Journal of Biological Diversity, 22*(8).

Swanson, R. A., & Holton, E. F. (2005). *Research in organizations: Foundations and methods in inquiry.* Berrett-Koehler Publishers.

Vahlensieck, W., Lorenz, H., Schumacher-Stimpfl, A., Fischer, R., & Naber, K. G. (2019). Effect of a herbal therapy on clinical symptoms of acute lower uncomplicated urinary tract infections in women: Secondary analysis from a randomized controlled trial. *Antibiotics, 8*(4), 256.

Vasudevan, R. (2014). Urinary tract infection: An overview of the infection and the associated risk factors. *Journal of Microbiology Experiment, 1*(2).

Wiest, M. M., Lee, K. J., & Carlin, J. B. (2015). Statistics for clinicians: An introduction to logistic regression. *Journal of Paediatrics & Child Health, 51*(7).

Zeb, B., Khan, A., Khan, Y., Masood, M. F., Tahir, I., & Asad, M. (2020). Towards the selection of the best machine learning techniques and methods for urinalysis. In *Proceedings of the 2020 12th International Conference on Machine Learning and Computing* (pp. 127–133).

How to cite this article:

Mirontos, J., Oxillo, M. J., & Murcia, J. V. (2024). Predicting urinary tract infection based on urinalysis test results of patients using binary logistic regression analysis. *International Journal of Multidisciplinary Studies, 6*(1), 102-118