UTI Detection and Rule-based Advisory System for Filipino Patients

A Hybrid Machine Learning and Knowledge-Based Reasoning Project

Collaborative Final Project

CSST101 – Machine Learning

CSST102 – Knowledge Representation and Reasoning

Submitted by:

Group Name: Three Musketeers

Group Members:

- Aeron Jhed Lachano
- Kean Salvahan
- Prince Genel Umali

Instructor: Mr. Mark P. Bernardino

PROJECT OVERVIEW

Urinary tract infections (UTIs) are very common in the Philippines, but many people struggle to checked early because clinics are often crowded or too far away. This project builds a model system that helps health workers quickly spot possible UTIs and other diseases using standard urine test results. It combines Machine learning prediction with logical medical rules to give an assessment and simple health advice, helping patients get care faster.

OBJECTIVES

General Objective: To build a hybrid system that uses both Machine Learning and medical rules to detect UTIs and provide easy to understand advice.

Specific Objectives:

1. Build a prediction model: Create a machine learning model that learns from urinalysis data to spot UTI patterns

2. Add medical logic: Design a rule-based system that explains why the model made a prediction and gives heath tips and the rule-based system spots other diseases from urinalysis data.

3. Make it easy to use: Create a simple tool for health workers like in barangays to input test results and get an instant answer

4. Help in limited settings: Support areas with few resources by providing a fast, free screening tool

SYSTEM ARCHITECTURE

User Input (Urine Test Results) → Machine Learning Model (Prediction) → KRR Rules (Logic Check) → Final Risk Level → Recommendations

MACHINE LEARNING COMPONENT (CSST101)

Algorithm Used: We used a “Grand Ensemble” (Voting Classifier). This combines three powerful models, which are Random Forest, XGBoost, and LightGBM to make a single, stronger prediction.

Dataset Size: 1,423 patient records

Model Accuracy: 97% accuracy

MACHINE LEARNING PIPELINE

Data Collection: We used a public dataset containing urinalysis results from kaggle

Data Preprocessing: The dataset of urinalysis results are messy, so we cleaned it up.

- Fixed text values (like changing “Negative” to 0)
- Handled ranges (turning ”1-3” into the average “2”)
- Filled in missing information using the most common values
- We converted categories like Color and Gender into numbers so that the system can understand.

Model Training: We train our “Grand Ensemble” model using 80% (1,128 records) of the data. We used special techniques to make sure the models learn to the rare positive UTI cases.

Model Evaluation: We tested the model on the remaining 20% (285 records) of the data. It achieved excellent results, correctly identifying almost all negative cases and doing a good job spotting the positive ones

Model deployment: We saved the trained model and the cleaning rules into files so they can be used in Website.

DATASET DESCRIPTION

Dataset Type: Clinical Urinalysis Data

Number of records: 1,423 rows

Target Variables: Diagnosis (Positive or Negative for UTI)

KNOWLEDGE REPRESENTATION & REASONING (CSST102)

Rule 1: IF WBC >= 10 THEN Reason = “Severe Infection Risk”

Rule 2: IF Glucose >=1 AND pH < 6.0 THEN Reason= “Diabetes Risk”

Rule 3: IF Color is RED OR RBC > 2 THEN Reason= “Hematuria Risk”

Rule 4: IF Epithelial Cells >= 3 AND Patient is MALE THEN Reason = “Sample Contamination”

Rule 5: IF Prediction is 1 THEN Reason = “Positive UTI”

HYBRID DECISION LOGIC

The Machine Learning model looks at the data patterns to make a prediction. If a UTI is positive. The advisory rule then intreprets this prediction and gives a advice of the user. Then, the Rule-Based System scans for other hidden diseases using KRR. Even if you don’t have a UTI, the system might catch early signs of Diabetes, or Kidney Stones, ensuring every patient gets useful health advice.

SYSTEM FEATURES

☒ Wellness risk prediction

☒ Rule-based recommendations

☒ Web interface / API

☒ Google Colab deployment

TESTING AND EVALUATION

Test Case	Input Summary	Expected Output
Patient A	Age: 23 Gender: Female Color: Straw Transparency: Turbid Glucose: Negative Protein: Negative pH: 5.0 S.G.: 1.015 WBC: 20-25 RBC: 0-2 Epithelial: Moderate Mucous: None Amorphous Urates: None Bacteria: Moderate	Positive For UTI
Patient B	Age: 15 Gender: Male Color: Yellow Transparency: Cloudy Glucose: Negative Protein: Trace pH: 7.0 S.G.: 1.015 WBC: 30-35 RBC: 0-2 Epithelial: Rare Mucous: None Amorphous Urates: Few Bacteria: Moderate	Positive For UTI
Patient C	Age: 3 Gender: Male Color: Light Yellow Transparency: Clear Glucose: Negative Protein: Negative pH: 7.0 S.G.:1.005 WBC: 0-3 RBC: 0-2 Epithelial: Rare Mucous: Rare Amorphous Urates: None Bacteria: Rare	Negative For UTI
Patient D	Age: 47 Gender: Female Color: Dark Yellow Transparency: Clear Glucose: 3+ Protein: Trace pH: 6.0 S.G.: 1.030 WBC: 0-1	Diabetes

	RBC: 0-2 Epithelial: Rare Mucous: None Amorphous Urates: None Bacteria: None	
Patient E	Age: 5 Gender: Female Color: Light Yellow Transparency: Clear Glucose: Negative Protein: Negative pH: 6.5 S.G.: 1.010 WBC: 0-2 RBC: 18-25 Epithelial: None Mucous: Rare Amorphous Urates: None Bacteria: None	Hematuria

CONCLUSION

We Successfully built a hybrid system that detects UTIs with 97% accuracy. By combining Machine Learning with medical rules. We created a tool that is both accurate and easy to understand. This can help health workers in the Philippines provide better, faster care.

GROUP CONTRIBUTION

Member Name | Contribution

Aeron Jhed Lachano – Website Design, Documentation, Assist in Making KRR (Non-UTI Conditions)

Kean Salvahan – Documentation, Evaluates UTI Model, Assist in Making KRR (UTI Advisory)

Prince Genel Umali – UTI Model, KRR, Documentation

REFERENCES

Datasets Avarice02. (2023). Urinalysis Test Results [Data set]. Kaggle. https://www.kaggle.com/datasets/avarice02/urinalysis-test-results