**About**

This dataset was collected for the fulfilment of the researchers capstone entitled: *Optimizing UTI Diagnosis with Machine Learning and Artificial Neural Network for Reducing Misdiagnoses* by Agdeppa et al. (2023). If you wish to acquire the research paper, please message me on my LinkedIn Account: [**https://tinyurl.com/54nnt2ea**](https://tinyurl.com/54nnt2ea)

**Content**

* Age (The age of the patient) **Note: Some patients are months old, so the age of these patients are preprocessed by dividing it by a hundred)** e.g, 8 MONTHS OLD, 8/100 = 0.08
* Gender (The gender of the patient) **Note: Either male or female**
* Color (urine color)
* Transparency (urine transparency)
* Glucose (glucose is a type of sugar, and its presence in the urine can be an important indicator of certain health conditions)
* Protein (the presence of protein in the urine is one of the parameters examined to assess kidney function and detect potential)
* pH (the pH level measures the acidity or alkalinity of urine)
* Specific Gravity (urine specific gravity is a measure of the concentration of particles in urine compared to water)
* WBC (White Blood Cells) **Note: Also known as leukocytes, white blood cells are a crucial part of the immune system**
* RBC (Red Blood Cells) **Note: RBC are responsible for carrying oxygen throughout the body**
* Epithelial Cells (epithelial cells are cells that line the surfaces and cavities of the body, including the urinary tract)
* Mucous Threads (mucous threads are strands of mucus that can be present in urine)
* Amorphous Urates (amorphous urates are non-crystalline formations in the urine that consist of uric acid)
* Bacteria (presence of bacteria in the urine)
* Diagnosis (UTI Diagnosis) **Note: Either NEGATIVE or POSITIVE**

**Acknowledgements**

I would like to thank Jan Laurence Virtudez for helping and assisting the researchers gather the data from the local clinic. Moreover, I would also like to thank the University of Science and Technology of Southern Philippines, Cagayan de Oro (USTP - CDO) (Ms. Angelie P. Sabanal and Mr. Romen Samuel Wabina, MSc) for the help they provided.

**Urinary Tract Infection (UTI) Diagnosis Dataset**

Urinary Tract Infections (UTIs) are among the most common bacterial infections worldwide, often leading to **misdiagnosis and improper treatment**. This dataset was collected to **optimize UTI diagnosis** using **Machine Learning and Artificial Neural Networks**, as part of a capstone project by **Agdeppa et al. (2023)**. It includes **patient demographics, urine test results, and diagnosis** to help develop **AI-driven diagnostic models** for **early and accurate detection** of UTIs.

**Table: Dataset Attributes and Description**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| **Age** | Age of the patient (Note: Infants’ ages are divided by 100, e.g., 8 months = 0.08) |
| **Gender** | Patient’s gender (Male/Female) |
| **Color** | Urine color |
| **Transparency** | Urine transparency (clear or cloudy) |
| **Glucose** | Presence of glucose in urine (indicator of diabetes or kidney issues) |
| **Protein** | Presence of protein in urine (indicator of kidney function) |
| **pH** | Acidity or alkalinity of urine |
| **Specific Gravity** | Concentration of particles in urine compared to water |
| **WBC** | White Blood Cells (indicator of infection) |
| **RBC** | Red Blood Cells (indicator of bleeding or infection) |
| **Epithelial Cells** | Cells lining the urinary tract (may indicate infection or contamination) |
| **Mucous Threads** | Strands of mucus in urine |
| **Amorphous Urates** | Non-crystalline uric acid formations in urine |
| **Bacteria** | Presence of bacteria in urine (indicator of infection) |
| **Diagnosis** | UTI Diagnosis (Positive/Negative) |

**Applications in Machine Learning & Healthcare**

* **Early detection of UTIs using AI**
* **Reducing misdiagnosis and improving patient care**
* **Developing AI-powered diagnostic models for clinics and hospitals**

This dataset is **crucial for AI-based UTI detection**, assisting healthcare professionals in **improving diagnostic accuracy and patient outcomes**. 🚑