Jurnal Informatika Ekonomi Bisnis

http://www.infeb.org

2024 Vol. 6 Iss. 1 Hal: 251-255 e-ISSN: 2714-8491

Efficacy of Machine Learning Techniques in Diagnosing Urinary Tract Infections: A Study Utilizing a Philippine Clinical Dataset

Gregorius Airlangga[1]✉

[1]Atma Jaya Catholic University of Indonesia

gregorius.airlangga@atmajaya.ac.id

Abstract

This research delves into the potential of machine learning models, namely Support Vector Machine (SVM), XGBoost, and LightGBM, to enhance the diagnosis of Urinary Tract Infections (UTIs) based on a comprehensive dataset collected from a local clinic in Northern Mindanao, Philippines, spanning from April 2020 to January 2023. The study integrates clinical variables such as age, gender, and various urine test results including color, transparency, and the presence of substances like glucose, protein, and cells, to determine the most accurate diagnostic model. The dataset presented unique preprocessing challenges, such as converting infant ages into decimal numbers. The SVM with a linear kernel showed remarkable test accuracy of 98.25%, indicating its robustness in handling linear separability in the data. Meanwhile, XGBoost and LightGBM, both with optimal hyperparameter configurations, achieved comparable accuracies of 97.95%. These results underscore the significance of machine learning in medical diagnostics, particularly in settings where swift and reliable decision-making is crucial. Our findings suggest that while ensemble methods like XGBoost and LightGBM are powerful tools for complex datasets, a well-tuned SVM can provide superior accuracy, thus advocating for a data-centric approach in model selection.

Keywords: Machine Learning, Urinary Tract Infections, Support Vector Machine, XGBoost, LightGBM.

1. Introduction

The Urinary Tract Infections (UTIs) are a prevalent condition worldwide, with significant impacts on public health and healthcare systems [1], [2], [3]. Characterized by a wide range of symptoms, UTIs can affect any part of the urinary system, with the majority involving the lower urinary tract [4], [5], [6]. The traditional diagnostic process relies heavily on clinical assessment and laboratory urinalysis, with the final diagnosis often hinging on the cultivation of urinary pathogens [7], [8], [9]. However, this approach is fraught with challenges, including time-consuming culture methods and the potential for ambiguous or misleading results due to contamination or asymptomatic bacteriuria [10], [11], [12]. Moreover, the increasing prevalence of antibiotic-resistant strains of UTI-causing bacteria accentuates the need for precise and timely diagnosis to guide effective treatment strategies [13], [14], [15].

The application of machine learning (ML) in medical diagnostics has emerged as a powerful tool to augment traditional methodologies, offering new avenues to improve accuracy, efficiency, and patient outcomes [16], [17], [18]. In the realm of UTI diagnostics, ML algorithms hold the promise of leveraging complex clinical and laboratory data to predict infection presence and guide therapeutic decisions more accurately [19], [20], [21]. This potential comes against the backdrop of an ever-growing digitalization of healthcare data and advancements in computational power and algorithmic sophistication [22], [23], [24].

A survey of the literature reveals a burgeoning interest in applying ML to various aspects of UTI management, from identifying risk factors and predicting infection susceptibility to automating urinalysis interpretation [25], [26], [27]. Despite these advances, the integration of ML into routine clinical practice for UTI diagnosis remains limited [28], [29], [30]. Several studies have highlighted the capabilities of ML models to outperform traditional statistical approaches in predicting UTIs, underscoring the feasibility and potential benefits of such technologies [31], [32], [33]. However, these studies also point to a critical gap: the need for comprehensive, real-world evaluations of ML models to establish their practical utility and operationalize their integration into clinical workflows [34], [35], [36].

The urgency of refining UTI diagnostic processes cannot be overstated. The misdiagnosis of UTIs leads to inappropriate treatment, including unnecessary antibiotic use, contributing to the global crisis of antibiotic resistance [37], [38], [39]. Moreover, delayed or missed diagnoses can result in patient morbidity, increased healthcare cost, and a higher burden on healthcare resources [40], [41]. In this context, the promise of ML to enhance diagnostic accuracy is not merely an academic pursuit but a pressing public health imperative.

This research aims to address these challenges by developing and validating an advanced ML-based diagnostic model for UTIs. Utilizing a comprehensive dataset from a local clinic in Northern Mindanao, Philippines, this study compares the performance of

Accepted: 24-03-2024 | Revision: 28-03-2024 | Publication: 31-03-2024 | doi: 10.37034/infeb.v6i1.855

several ML algorithms, including logistic regression, support vector machines, XGBoost, and LightGBM, in predicting UTI diagnoses [42]. The goals are twofold: to identify the most effective ML models for UTI diagnosis based on real-world clinical data and to integrate these models into a practical tool that can support clinicians in making more accurate and timely diagnostic decisions.

The gap this research seeks to fill is multifaceted. Firstly, it addresses the need for rigorous, comparative analyses of different ML algorithms in the context of UTI diagnostics, an area where previous research has been somewhat piecemeal. Secondly, it contributes to the methodological advancement of ML in healthcare by proposing novel approaches to data preprocessing and feature engineering, tailored to the unique challenges of UTI diagnostic data. Finally, by focusing on the operationalization of ML models in clinical settings, this study moves beyond theoretical accuracy to assess the practical viability and user acceptance of ML-assisted diagnostics.

Our contribution to the field of medical diagnostics and, more specifically, to the management of UTIs is significant. By demonstrating the superior accuracy of ML models over traditional diagnostic approaches, this research paves the way for the adoption of AI-driven diagnostic tools in clinical practice. Furthermore, the development of a practical ML-based diagnostic tool represents a concrete step toward the real-world application of AI in healthcare, potentially setting a precedent for the broader integration of ML in disease diagnosis and management. The structure of this article unfolds as follows: Section 2 details the materials and methods, including the dataset description, ML algorithms employed, data preprocessing techniques, and model evaluation criteria. Section 3 presents the results, offering a comparative analysis of the performance of different ML models and discussing the implications of these findings for clinical practice. Furthermore, we delve into the discussion, interpreting the results in the context of existing literature, addressing the study's limitations, and suggesting directions for future research. Lastly, we conclude the article by summarizing the key findings, highlighting the contributions of this research to the field, and outlining the potential impacts on the diagnosis and treatment of UTIs.

2. Research Method

In this study, we have embarked on an investigation to harness the potential of machine learning (ML) for optimizing the diagnosis of Urinary Tract Infections (UTIs). Our methodology is anchored in a comprehensive dataset, rigorous data preprocessing, and the evaluation of various ML algorithms to ascertain their diagnostic efficacy. This section elucidates the materials and methods employed in our research, detailing the dataset's origins, our preprocessing techniques, and the ML algorithms we scrutinized.

2.1. Dataset Acquisition and Description

The foundation of our study is a dataset meticulously collected from a local clinic in Northern Mindanao, Philippines, covering the period from April 2020 to January 2023 [42]. This dataset is comprised of records from patients who underwent urinalysis, a standard diagnostic test for UTIs. Each record encapsulates a wealth of information, including demographic details such as age and gender, alongside urinalysis results that span parameters like color, transparency, glucose levels, protein presence, pH, specific gravity, white blood cell count (WBC), red blood cell count (RBC), epithelial cells, mucous threads, amorphous urates, and bacteria presence. The dataset is dichotomized into two categories based on the diagnosis: UTI positive and UTI negative, providing a clear target for our ML models.

2.2. Data Preprocessing

Prior to the application of ML algorithms, our dataset underwent a series of preprocessing steps, critical for refining the data into a format conducive for analysis. This process commenced with data cleaning, where incomplete records were identified and addressed, ensuring the integrity of the dataset. Given the diverse nature of the data, especially with variables expressed in different formats and scales, normalization and encoding were paramount.

Normalization involved adjusting continuous variables like age, pH, and specific gravity to a common scale without distorting differences in the ranges of values. This was achieved using the MinMaxScaler, which transforms the data within a bounded interval. Categorical variables, such as gender and urinalysis parameters like color and transparency, were encoded to numerical values to facilitate their interpretation by the ML algorithms. This included both one-hot encoding for nominal features and ordinal encoding for ordinal features, based on their inherent order or ranking.

Furthermore, recognizing the imbalance in the dataset between UTI positive and negative diagnoses, we employed Synthetic Minority Over-sampling Technique (SMOTE). This technique generates synthetic samples from the minority class, UTI positive in this case, to create a balanced dataset, thereby enhancing the model's ability to learn from an equal representation of both classes.

2.3. Evaluation of Machine Learning Algorithms

With a preprocessed and balanced dataset, we proceeded to the evaluation of various ML algorithms, each with its unique strengths and suitability for binary classification problems like UTI diagnosis. The algorithms selected for this study encompass a range from simple logistic regression, known for its interpretability, to more complex models like support vector machines (SVM), Random Forest, XGBoost, and LightGBM, renowned for their robustness and high performance in a multitude of predictive tasks. Additionally, we explored the efficacy of neural networks, leveraging the capabilities of deep learning for pattern recognition and classification.

For each algorithm, a detailed parameter tuning was conducted using the Optuna framework, an optimization software that automates the selection of the best hyperparameters to maximize model performance. This involved defining a search space for each parameter and employing cross-validation to evaluate the model's performance under different parameter combinations, ensuring the robustness and generalizability of our results.

2.4. Model Evaluation and Selection

The performance of each ML model was assessed using a suite of metrics, including accuracy, precision, recall, F1 score, and the area under the Receiver Operating Characteristic (ROC) curve. These metrics provided a comprehensive view of each model's diagnostic ability, considering both the precision of positive predictions and the overall rate of correct classifications. The model showcasing the best performance across these metrics was selected for further analysis and interpretation.

3. Result and Discussion

The results from the Table 1 to Table 3 indicate that the Support Vector Machine (SVM), XGBoost, and LightGBM algorithms show a range of testing accuracies across different hyperparameter settings. In this analysis, we can discuss the impact of these hyperparameters on the model's performance and draw some conclusions based on the data provided. Result for SVM test can be seen on Table 1.

Table 1. SVM Testing Accuracy Results with Hyperparameter Tuning Variations

C	Kernel	Test Accuracy
1	Linear	98.25
1	Poly	97.37
1	Poly	97.08
1	RBF	96.49
10	Linear	98.25
10	Poly	97.66
10	Poly	97.66
10	RBF	96.78
100	Linear	98.25
100	RBF	96.78

From Table 1, we observe that the SVM with a linear kernel consistently performs well across different values of the penalty parameter C, with the highest test accuracy being 98.25%. This suggests that a linear decision boundary is sufficient for the dataset used, and increasing the complexity of the model with a higher C value does not significantly impact the accuracy. The poly kernel shows varied results, which might indicate overfitting or underfitting depending on the degree of the polynomial, while the RBF kernel's performance is generally lower compared to the linear kernel, possibly due to the nature of the data distribution. Furthermore, result for XGBoost test can be seen on Table 2.

Table 2. XGBoost Testing Accuracy Results with Hyperparameter Tuning Variation

Max Depth	Learning Rate	n_estimators	Test Accuracy
3	0.1	100	97.37
4	0.1	100	97.66
5	0.1	100	97.66
3	0.01	100	96.78
3	0.1	200	97.66
4	0.01	200	96.78
5	0.05	200	97.95
3	0.05	300	97.37
4	0.05	300	97.66
5	0.01	300	96.78

Table 2 shows the performance of XGBoost with varying max depths, learning rates, and numbers of estimators. It is apparent that the model is quite robust to changes in the max depth, but there is a slight increase in accuracy as the learning rate decreases and the number of estimators increases. The best test accuracy is 97.95% with a learning rate of 0.05 and 200 estimators. This could mean that the model benefits from more gradual learning, allowing it to generalize better. Lastly, LightGBM test can be seen on Table 3.

Table 3. LightGBM Testing Accuracy Results with Hyperparameter Tuning Variation

Max Depth	LR	n_estimators	TA
-1	0.1	100	97.37
10	0.1	100	97.66
20	0.1	100	97.66
-1	0.01	100	96.49
-1	0.1	200	97.95
10	0.01	200	96.78
20	0.05	200	97.95
-1	0.05	300	97.37
10	0.05	300	97.66
20	0.01	300	96.78

Wjere LR is learning rate and TA is test accuracy. In Table 3, LightGBM's results are similar to those of XGBoost, with the highest accuracy also being 97.95% but occurring at a max depth of -1, indicating no limit, and with 200 estimators. This could suggest that LightGBM handles overfitting well even with a larger number of leaves. However, unlike XGBoost, a max depth of -1 (no limit) combined with a learning rate of 0.1 and 100 estimators did not yield the highest accuracy, possibly due to a different handling of tree complexity and growth strategies between the two frameworks.

The analysis of the results from the tables showcasing the performances of SVM, XGBoost, and LightGBM models reveals that achieving higher test accuracies is closely tied to the careful tuning of hyperparameters. This underscores the pivotal role that hyperparameter optimization plays in enhancing the efficacy of machine learning models. Particularly noteworthy is the superior performance of the SVM with a linear kernel across various configurations of the penalty parameter C, suggesting that the decision boundary within the data may not be overly complex. This could also hint at the possibility that the dimensions within the dataset are already sufficiently distinct, obviating the need for more sophisticated models.

In the case of both XGBoost and LightGBM, we observe that maintaining an equilibrium between the learning rate and the number of estimators is instrumental in achieving higher accuracy. Optimal performance is not necessarily associated with either excessively swift or unduly gradual learning rates. Intriguingly, the data reveals a potential linear separability, as indicated by the highest accuracy achieved by the linear SVM. Despite their typical aptitude for handling more intricate datasets, the ensemble methods XGBoost and LightGBM do not significantly outperform the SVM, which could reflect on the dataset's inherent characteristics. Furthermore, the consistent performance of the models under various hyperparameter conditions merits attention. The absence of any notable performance degradation suggests that the dataset in question may exhibit a degree of robustness to these parameters, or that the models are inherently resilient to overfitting. Therefore, the selection of an appropriate model and its hyperparameters is inherently contingent upon the specific traits of the dataset in question. The findings suggest that for the dataset at hand, a simpler model such as the SVM with a linear kernel is quite effective, achieving the highest level of accuracy observed. However, it is important to acknowledge that XGBoost and LightGBM also demonstrate competitive performance, and with judicious hyperparameter tuning, there is the potential to equal or even surpass the SVM's accuracy.

4. Conclusion

In summary, the empirical evaluation of SVM, XGBoost, and LightGBM models across a variety of hyperparameter configurations provides valuable insights into model selection and tuning in the context of the dataset analyzed. The empirical evidence suggests that a simpler model, SVM with a linear kernel, achieves the most impressive accuracy, indicating that the dataset may exhibit a linear separability. This finding challenges the often-held belief that more complex models invariably lead to better performance and underscores the importance of matching model complexity to dataset characteristics. While the ensemble methods, XGBoost and LightGBM, did not significantly surpass the performance of the SVM, their competitive accuracies cannot be disregarded. This demonstrates their potential effectiveness, especially when fine-tuned, which may be crucial for more complex or larger datasets where linear models might fall short.

References

[1] Yang, X., Chen, H., Zheng, Y., Qu, S., Wang, H., & Yi, F. (2022). Disease burden and long-term trends of urinary tract infections: A worldwide report. *Frontiers in public health*, *10*, 888205. https://doi.org/10.3389/fpubh.2022.888205

[2] Wagenlehner, F., Nicolle, L., Bartoletti, R., Gales, A. C., Grigoryan, L., Huang, H., ... & Lee, S. J. (2022). A global perspective on improving patient care in uncomplicated urinary tract infection: expert consensus and practical guidance. *Journal of global antimicrobial resistance*, *28*, 18-29. https://doi.org/10.1016/j.jgar.2021.11.008

[3] Öztürk, R., & Murt, A. (2020). Epidemiology of urological infections: a global burden. *World journal of urology*, *38*, 2669-2679. https://doi.org/10.1007/s00345-019-03071-4

[4] Klein, R. D., & Hultgren, S. J. (2020). Urinary tract infections: microbial pathogenesis, host–pathogen interactions and new treatment strategies. *Nature Reviews Microbiology*, *18*(4), 211-226. https://doi.org/10.1038/s41579-020-0324-0

[5] Mancuso, G., Midiri, A., Gerace, E., Marra, M., Zummo, S., & Biondo, C. (2023). Urinary tract infections: the current scenario and future prospects. *Pathogens*, *12*(4), 623. https://doi.org/10.3390/pathogens12040623

[6] Czajkowski, K., Broś-Konopielko, M., & Teliga-Czajkowska, J. (2021). Urinary tract infection in women. *Menopause Review/Przegląd Menopauzalny*, *20*(1), 40-47. https://doi.org/10.5114/pm.2021.105382

[7] McEachern, F., Harvey, E., & Merle, G. (2020). Emerging technologies for the electrochemical detection of bacteria. *Biotechnology Journal*, *15*(9), 2000140. https://doi.org/10.1002/biot.202000140

[8] Szabó, S., Feier, B., Capatina, D., Tertis, M., Cristea, C., & Popa, A. (2022). An overview of healthcare associated infections and their detection methods caused by pathogen bacteria in Romania and Europe. *Journal of clinical medicine*, *11*(11), 3204. https://doi.org/10.3390/jcm11113204

[9] Rogers, M. J., McManus, D. P., Muhi, S., & Gordon, C. A. (2021). Membrane technology for rapid point-of-care diagnostics for parasitic neglected tropical diseases. *Clinical microbiology reviews*, *34*(4), e00329-20. https://doi.org/10.1128/CMR.00329-20

[10] Kashouris, E. G. (2023). *Inflammatory Responses: UTI & AMR* (Doctoral dissertation, University of Sussex).

[11] Sussman, A. (2023). *Saving the Art of Medicine: Observations of a Practitioner*. FriesenPress.

[12] Siraj, N., Bwambok, D. K., Brady, P. N., Taylor, M., Baker, G. A., Bashiru, M., ... & Fakayode, S. O. (2021). Raman spectroscopy and multivariate regression analysis in biomedical research, medical diagnosis, and clinical analysis. *Applied Spectroscopy Reviews*, *56*(8-10), 615-672. https://doi.org/10.1080/05704928.2021.1913744

[13] Heffernan, J. R. (2023). *Function and Regulation of Virulence-Associated Siderophores in Uropathogenic Enterobacterales* (Doctoral dissertation, Washington University in St. Louis).

[14] Salman, H. A., Alsallameh, S. M. S., Muhamad, G., & Taha, Z. (2022). Prevalence of multi-antibiotic resistant bacteria isolated from children with urinary tract infection from Baghdad, Iraq. *Microbiology and Biotechnology Letters*, *50*(1), 147-156. https://doi.org/10.48022/mbl.2110.10011

[15] Almalki, M. A., & Varghese, R. (2020). Prevalence of catheter associated biofilm producing bacteria and their antibiotic sensitivity pattern. *Journal of King Saud University-*

Science, *32*(2), 1427-1433. https://doi.org/10.1016/j.jksus.2019.11.037

[16] Javaid, M., Haleem, A., Singh, R. P., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, *3*, 58-73. https://doi.org/10.1016/j.ijin.2022.05.002

[17] Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, *145*, 105458. https://doi.org/10.1016/j.compbiomed.2022.105458

[18] Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. *Multimedia Tools and Applications*, *82*(17), 26731-26769. https://doi.org/10.1007/s11042-022-14305-w

[19] Anahtar, M. N., Yang, J. H., & Kanjilal, S. (2021). Applications of machine learning to the problem of antimicrobial resistance: an emerging model for translational research. *Journal of clinical microbiology*, *59*(7), 10-1128. https://doi.org/10.1128/jcm.01260-20

[20] Smith, K. P., Wang, H., Durant, T. J., Mathison, B. A., Sharp, S. E., Kirby, J. E., ... & Rhoads, D. D. (2020). Applications of artificial intelligence in clinical microbiology diagnostic testing. *Clinical Microbiology Newsletter*, *42*(8), 61-70. https://doi.org/10.1016/j.clinmicnews.2020.03.006

[21] Khang, A. (Ed.). (2024). *Medical Robotics and AI-Assisted Diagnostics for a High-Tech Healthcare Industry*. IGI Global.

[22] Junaid, S. B., Imam, A. A., Balogun, A. O., De Silva, L. C., Surakat, Y. A., Kumar, G., ... & Mahamad, S. (2022, October). Recent advancements in emerging technologies for healthcare management systems: a survey. In *Healthcare* (Vol. 10, No. 10, p. 1940). MDPI. https://doi.org/10.3390/healthcare10101940

[23] Baiyewu, A. S. (2023). Overview of the Role of Data Analytics in Advancing Health Service. *Open Access Library Journal*, *10*(6), 1-19. https://doi.org/10.4236/oalib.1110207

[24] Nayak, S., Patgiri, R., Waikhom, L., & Ahmed, A. (2022). A review on edge analytics: Issues, challenges, opportunities, promises, future directions, and applications. *Digital Communications and Networks*. https://doi.org/10.1016/j.dcan.2022.10.016

[25] Allen, K. (2023). Exploration of the Implementation of an Integrated Electronic Laboratory Information Management System on Quality Diagnostics Service Indicators at a County Level Public Hospital in western Kenya.

[26] Xiong, P., Lor, M., Park, L. S., & Jacobs, E. A. (2023). How Patients With Limited English Proficiency Make Health Care Decisions: Hmong Patients' Perspectives. *WMJ: official publication of the State Medical Society of Wisconsin*, *122*(3), 178-183.

[27] Malone-Lee, James. *Cystitis unmasked*. tfm Publishing Limited, 2021.

[28] Morado, F., & Wong, D. W. (2022). Applying diagnostic stewardship to proactively optimize the management of urinary tract infections. *Antibiotics*, *11*(3), 308. https://doi.org/10.3390/antibiotics11030308

[29] Luz, C. F., Vollmer, M., Decruyenaere, J., Nijsten, M. W., Glasner, C., & Sinha, B. (2020). Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. *Clinical Microbiology and Infection*, *26*(10), 1291-1299. https://doi.org/10.1016/j.cmi.2020.02.003

[30] Xu, R., Deebel, N., Casals, R., Dutta, R., & Mirzazadeh, M. (2021). A new gold rush: a review of current and developing diagnostic tools for urinary tract infections. *Diagnostics*, *11*(3), 479. https://doi.org/10.3390/diagnostics11030479

[31] Vakitbilir, N., Froese, L., Gomez, A., Sainbhi, A. S., Stein, K. Y., Islam, A., ... & Zeiler, F. A. (2024). Time-Series Modeling and Forecasting of Cerebral Pressure–Flow Physiology: A Scoping Systematic Review of the Human and Animal Literature. *Sensors*, *24*(5), 1453. https://doi.org/10.3390/s24051453

[32] Iriya, R. (2020). *Computer Vision Methods for Urinary Tract Infection Diagnostics* (Doctoral dissertation, Arizona State University).

[33] Ahmed, A., Xi, R., Hou, M., Shah, S. A., & Hameed, S. (2023). Harnessing big data analytics for healthcare: A comprehensive review of frameworks, implications, applications, and impacts. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3323574

[34] Lwakatare, L. E., Raj, A., Crnkovic, I., Bosch, J., & Olsson, H. H. (2020). Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Information and software technology*, *127*, 106368. https://doi.org/10.1016/j.infsof.2020.106368

[35] Weissler, E. H., Naumann, T., Andersson, T., Ranganath, R., Elemento, O., Luo, Y., ... & Ghassemi, M. (2021). The role of machine learning in clinical research: transforming the future of evidence generation. *Trials*, *22*, 1-15. https://doi.org/10.1186/s13063-021-05489-x

[36] Sandhu, S., Lin, A. L., Brajer, N., Sperling, J., Ratliff, W., Bedoya, A. D., ... & Sendak, M. P. (2020). Integrating a machine learning system into clinical workflows: qualitative study. *Journal of Medical Internet Research*, *22*(11), e22421. https://doi.org/10.2196/22421

[37] Mangal, S. L. (2020). *Patient and Family Engagement in the Prevention of Catheter-Associated Urinary Tract Infections and Antibiotic Resistance*. Columbia University.

[38] Lescure, D. L., van der Velden, A. W., Huijser van Reenen, N., Richardus, J. H., & Voeten, H. A. (2022). Perspectives of Immigrants and Native Dutch on Antibiotic Use: A Qualitative Study. *Antibiotics*, *11*(9), 1179. https://doi.org/10.3390/antibiotics11091179

[39] De Paor, M. (2022). *Diagnosis and management of sore throat in young adults* (Doctoral dissertation, Royal College of Surgeons in Ireland).

[40] Kokolakis, G., Wolk, K., Schneider-Burrus, S., Kalus, S., Barbus, S., Gomis-Kleindienst, S., & Sabat, R. (2020). Delayed diagnosis of hidradenitis suppurativa and its effect on patients and healthcare system. *Dermatology*, *236*(5), 421-430. https://doi.org/10.1159/000508787

[41] Santosh, K. C., & Gaur, L. (2022). *Artificial intelligence and machine learning in public healthcare: Opportunities and societal impact*. Springer Nature.

[42] Avarice02. (2024). Urinalysis Test Results Dataset. Kaggle Dataset. Retrieved from https://www.kaggle.com/datasets/avarice02/urinalysis-test-results