International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145

e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

Survey on Machine Learning Models to Analyze Urinary Tract Infection Data

Impana K. P.[1], Akanksh Saya[2], Bhagyashri Shetty H.[3], Chagi Manaswini[4], Charudisha A. Ashjay[5]

[1]Assistant Professor, Computer Science and Engineering, JSS Academy of Technical Education, Bangalore, Karnataka, India.

[2,3,4,5]UG, Computer Science and Engineering, JSS Academy of Technical Education, Bangalore, Karnataka, India.

Emails: *impanaraj@jssateb.ac.in[1], akankshs.01@gmail.com[2], bhagyashrishettyh@gmail.com[3], manucse2021@gmail.com[4], charudisha357@gmail.com[5]*

Abstract

A urinary tract infection, or UTI, is caused when bacteria get into the urinary tract- kidneys, bladder, or urethra. UTIs cause more than 8.1 million visits to healthcare providers each year. About 60% of women and 12% of men get infected with UTI during their lifetime, therefore being more prominent in females. UTIs can be found by analyzing a urine sample. The urine is examined under a microscope for bacteria or white blood cells that show infection. Healthcare providers may also take a urine culture. This test examines urine to detect and identify bacteria and yeast, which may be causing a UTI. Several models have been proposed to predict urine culture positivity based on urinalysis. Applications of machine learning (ML) methods have been used extensively to solve various complex challenges in recent years in various application areas. ML methods are characterized by their ability to examine data and discover exciting relationships, provide interpretation, and identify patterns. ML can help enhance the reliability, performance, predictability, and accuracy of diagnostic systems for many diseases. This survey provides a comprehensive review of the use of ML in the diagnosis of Urinary Tract infection in human beings. To provide a reliable classification of results assistance of 27 algorithms was tested. Algorithms applied included Logistic regression, Decision tree, Random Forest, and Support Vector Machine. Each model was evaluated by F1-score, AUC-ROC, accuracy, sensitivity, and specificity. Baseline epidemiological factors, previous antimicrobial consumption, medical history, and previous culture results were included as features. Machine learning models such as the Artificial neural network have been used as well for the prediction of the presence of urinary infection. Some specific parameters have been selected with the help of the Analysis of variance technique which gave high accuracy. This survey provides a comprehensive review of the use of ML in the medical field highlighting standard technologies and how they affect medical diagnosis. It provides valuable references and guidance for researchers, practitioners, and decision-makers framing future research and development directions. It is found that Machine Learning models can improve the early prediction of urine culture positivity and UTI by combining automated urinalysis with other clinical information. Clinical utilization of the models can reduce the risk of delayed antimicrobial therapy in patients with nonspecific symptoms of UTI and classify patients with UTI who require further treatment and close monitoring. In conclusion, the paper provides a survey on the machine learning models used with the highest accuracy to detect UTI potential patients.

Keywords: *Artificial Neural Network, Machine Learning (ML), Random Forest, Support Vector Machine, Urinary Tract Infection (UTI).*

1. Introduction

An infection is the invasion and growth of germs in the body. The germs may be bacteria, viruses, yeast, fungi, or other microorganisms. Infections can begin anywhere in the body and may spread all through it [4]. It can cause fever and other health problems, depending on where it occurs in the body. When the body's immune system is strong, it can often fight the germs and cure an infection. The urinary system (or urinary tract) includes the kidneys, ureters, bladder, and the urethra. It works as your body's filtration system. When your urinary system removes toxins and wastes from your body, it comes out as pee (urine). To be able to urinate, your body must pass this waste through a series of organs, ducts, and tubes. It eliminates extra water and salt, toxins, and other waste products. Different parts of the urinary system perform tasks, including: Filtering blood, separating the toxins you don't need from the nutrients you do need, and storing and carrying pee out of your body [1, 2, 3]. Many conditions can affect the ureters, kidneys, bladder and urethra. Infections, diseases or problems can appear at birth or develop as one gets older. Urinary Tract Infections can cause issues with the kidneys, urethra or bladder. These infections occur when bacteria enter your urinary tract through your urethra [4, 5, 7]. Urinary tract infections (UTIs) are some of the most common bacterial infections, affecting 150 million people each year worldwide. Urinary tract infections (UTIs) are the most prevalent outpatient illnesses and affect about 50% of the population at some point in their lives [9, 10]. The incidence of UTIs is increasing with age, and close to 10% of postmenopausal women indicate that they had a UTI in the previous year [11]. UTIs are a significant cause of morbidity in infant boys, older men and females of all ages. Serious sequelae include frequent recurrences, pyelonephritis with sepsis, renal damage in young children, preterm birth and complications caused by frequent antimicrobial use, such as high-level antibiotic resistance. Clinically, UTIs are categorized as uncomplicated or complicated. Uncomplicated UTIs typically affect individuals who are otherwise healthy and have no structural or neurological urinary tract abnormalities; these infections are differentiated into lower UTIs (cystitis) and upper UTIs (pyelonephritis). Several risk factors are associated with cystitis, including female gender, a prior UTI, sexual activity, vaginal infection, diabetes, obesity and genetic susceptibility. Complicated UTIs are defined as UTIs associated with factors that compromise the urinary tract or host defense, including urinary obstruction, urinary retention caused by neurological disease, immunosuppression, renal failure, renal transplantation, pregnancy and the presence of foreign bodies such as calculi, indwelling catheters or other drainage devices [16, 17]. UTI is linked to a variety of diseases that can cause serious health issues or even a loss of life. Medical laboratories and hospitals are outfitted with high-tech devices and equipment to provide curative healthcare. However, in a hospital-centric world, providing healthcare services to all patients becomes difficult with such a large population. In this case, tracking of infected patients in an IoT-enabled home-centric environment has opened new doors for the healthcare industry [7]. The majority of these machines are ready to utilize and can be quickly displaced in a home's toilet system. Furthermore, limited human interference increases the system's overall performance. The paper represents the understanding of cause and effects of the Urinary Tract Infection and analyzing the requirements of forecasting the disease. Thus, the aim is to study the methodology driven by machine learning-based techniques for predicting urinary tract infections with considerable accuracy.

2. Literature Review

Huge works have been done in the field of disease diagnosis using various techniques. The recent works have been discussed here with their advantages and disadvantages. In [34], they have proposed a system that could predict 41 diseases based on 130 symptoms using Naïve Bayes, decision tree classifier and random

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

forest classifier with an accuracy of 93% on train data and 95% on test data. In [35], they had implemented a disease inference question answer system in which the user question is passed, and the system would use data mining techniques to provide the answer. In [36], they have proposed a work to find hidden patterns in the disease data, for predicting the disease based on their symptoms using CNN and K Nearest Neighbor (KNN). For monitoring and prediction of UTIs, the existing paper has suggested an IoT-inspired framework in a home-centric environment. By using ID (Infection Degree), probabilistic classification (Bayesian Probabilistic Model) of UTI has been performed in accordance with both infection and non-infection categories. Temporal-Artificial Neural Network (t-ANN) technique, which incorporates IIM (Infection Index Measure), has been used for the detection and prediction of UI.

Neural Networks have the advantage over traditional programming in terms of solving problems for which there is no algorithmic solution. Also, it helps in finding solutions which are too complex to find. Neural Networks are widely used in the medical domain to solve all the problems related to clinical diagnosis, prediction, image processing & interpretation, drug development and pattern recognition. Heckerling et al. proposed a model using artificial neural networks in combination with genetic algorithms that evolved combinations related to clinical variables that are optimized to predict UTI [27]. The suggested work evolved five variable sets that can classify the UTI cases and non-infection areas ranging from 0.853 to 0.792. The results of network influence analysis show few variables predicted UTI in an unexpected way and also have an interaction with other variables in forming prediction [28]. Pérez et al. proposed a system based on a multiagent system model in which every neuronal center corresponds to an agent [29]. This scheme improves its robustness by including a heuristic in the presence of possible inconsistencies. A neural network is used as the heuristic (orthogonal associative memory). The system incorporates knowledge through instruction, utilizing appropriate urinary tract behavior patterns and behavior patterns resulting from dysfunctions in neuronal centers as a minimum. Nyman & Jesper [30] in their research, aims to investigate how different screening methods perform when applied before culturing. To predict UTI, the screening methods use flow cytometry analysis (FCA) and some general characteristics. Different screening approaches were compared using machine learning algorithms. A sensitivity adjustment was used to adjust the methods so that the sensitivity was greater than 95%. The output was evaluated using real-world data from 1316 samples and cross-validation. Random forest yielded the best results in terms of cost savings. It was able to reduce the load on the culturing process by up to 46% while maintaining a sensitivity of 95.15 percent. The specificity rate was 72%. Even though the data set collected was too limited to accurately declare real results, the savings appear to be quite promising. Several parameters are taken by several authors for Urine Infection (UI) with associated disease and their unsafe values [30]. Patient-level data were collected, including demographics, underlying comorbidities with age-adjusted Charlson comorbidity index, date of urine sample collection, and commercial analyzers on which urinalysis was performed. To extract the worst values within 24 h of urine culture sampling, both maximum and minimum values of laboratory tests and vital signs were obtained. In addition, the use of vasopressors, antimicrobial agents, and mechanical ventilation was investigated, and the worst Glasgow Coma Scale and a Sequential Organ Failure Assessment (SOFA) score were also calculated. The results of test strip analysis and urine sediment analysis using the electronic medical record collection programs of the patients at each institution were retrieved.

3. Machine Learning-Based Algorithms

Machine learning is a branch of artificial intelligence that enables computers to learn from data and make predictions or decisions without being explicitly

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

programmed. By analyzing large and complex datasets of health records, imaging, genetics, environmental factors, and other variables, they can provide personalized and preventive recommendations for patients and clinicians, such as optimal treatment plans, lifestyle interventions, and follow-up actions. In general, Machine Learning is classified as either supervised learning or unsupervised learning. In supervised learning, for the input variables, output variables are predicted and in unsupervised learning, for input variables, output variables are not predicted. (i.e., deals with clustering of distinct groups for a specific intervention) [11]. Complex models and medical diagnosis can be derived from machine learning, therefore, revealing new ideas to clinicians and specialists. In clinical practice, machine learning predictive models will illustrate better guidelines for decision making in individual patient care. These are also able to self-diagnose a variety of diseases in accordance with clinical guidelines. According to [13], integrating these models into drug prescriptions will save healthcare worker's time and open new medical possibilities in pathology detection. Medical data quality can be improved, patient rates can be reduced, and medical costs can be reduced using Machine Learning models. Consequently, when opposed to other traditional approaches, these models are commonly utilized to examine diagnostic analysis. Therefore, taking these benefits into account the artificial neural networks and machine learning techniques are used in evaluation and prediction of urinary tract infection among men and women.

3.1. K-Nearest Neighbours Algorithm

One special feature of the non-linear classifier K-nearest-neighbors is that it doesn't need to be trained. Rather, it groups fresh samples according to how well they match existing data. The general method involves classifying based on response by examining the k most comparable known samples.[10] There are two steps involved in this. For the first, a distance metric is necessary. To calculate similarities, this metric is employed. The minkowski distance will be the one employed in this investigation [14].

$$W_i = |(X - X_i)^p|^{1/p} , (1)$$

If Xi is a known sample, p is an adjustable parameter, and X is the sample that will be classified. To obtain the most comparable known samples, one can iterate equation 1 for each sample i ∈ [1: n] and then look at the k smallest wi. For an example using random data, k = 3 and p = 2, see Figure 1. The new sample is categorized in the second step using the neighbors that were discovered. A function that determines a score value based on neighbors' responses and corresponding wi can be used to do this. The following function will be utilized in this analysis:

$$\text{Score} = \frac{\sum_{i=1}^{k} y^i \frac{1}{w_i}}{\sum_{i=1}^{k} \frac{1}{w_i}} , (2)$$

where wi is the distance to the neighbor I and yi is the neighbor I's response. Depending on a threshold value, the new sample can be classified as one or zero based on this score value [14].

Algorithm 1: K-Nearest Neighbor

Input: dataset df, x_{train}, x_{test} – independent variable, y_{train}, test target variable

Input: dataset df, x_{train}, x_{test} – independent variable, y_{train}, test target variable

Output: target words with tract infection of UTX_DIAGONSIS
begin
for each y belong df do
calculate the Dist D (x, y)bet y and x
end
subset(N) from the data frame (df)
N contains m training samples which are the knn of the test samples x
cal categorical variables of x:

$y_x = \arg_{y\epsilon Y} \max \sum_{y \epsilon Y} I\ (y = \text{class}(\text{categorical}_{variables}))$

End [4]

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

3.2. Artificial Neural Network Algorithm

Artificial neural networks are computing systems that mimic the parallel structure of the human brain. The network's function is mostly defined by connections between elements, with weights assigned to each connection between two neurons. In artificial neural networks, the input layer contains nodes that receive initial data, while the output layer offers results based on the input [14]. Additional layers of units, referred to as intermediate layers or hidden layers which exist between the input & an output layer, are responsible for all the computations. An artificial neural network (ANN) comprises linked neurons. This magnitude can then be utilized to calculate the score value. In the ANN, each neuron belongs to a distinct layer [13]. An example of an ANN using three features can be seen in Figure 2. The steps for creating the ANN architecture are as follows:

a) Layer 1 corresponds to the input vector containing input variables. Specified as I1 – I3.
b) In Layer 2, it is important to determine the number of neurons in the hidden layer within the ANN structure. As a result of the analysis realized, the number of neurons in the hidden layer that gave the optimal result was determined as 12.
c) Layer 3 is the output layer and has two outputs, cystitis, and nonspecific urethritis. These output parameters were also specified as O1[1-3].

It shows that the input vector for neurons in the input layer is the feature vector, with each neuron taking one feature. The input vector for subsequent levels is the output scores from the previous layer, with each neuron taking all of them. The final layer consists of a single neuron, and its score is used for categorization. Each neuron has a weight vector W and a bias term w0 that must be determined. This can be achieved through backpropagation (BP) [13].

Algorithm 2: MANN-AM (Modified Artificial Neural Network with Attention Mechanism)

Input: Input : dataset df, x_{train}, x_{test} - independent variable, y_{train}, y_{test} target variable
Output: target y_{preds} with tract infection of UTX_DIAGONSIS(0, 1)
scale the datasets using // input layer
divide x_{train} x_{test} and y_{train}, y_{test} //input layer
W is calculated based on dist using //hidden layer
set value for parameter k
Estimating the distance between train and test
Sorting dist in an ascending pattern select the best k_{neigh}
Repeating steps 2–4 until the algorithm is over
W matrix is saved as result Simulation involves using //hidden layer
Saving results
Retrieving ANN with attention model
Setting values for number of input, output and hidden layers
Primary weighing of existing neurons in input, output and hidden layers
Calculating the output (y) for each neuron in output layer
Updating ANN parameters
Repeating steps 3–4 until the algorithm is over
Saving results
End of hybrid model
Displaying results
End [4]

3.3. Support Vector Machine

Support vector machines are linear classifiers that aim to find the best hyperplane and margin for separating classes in data D. The margin is the distance between the hyperplane and the close feature vector, or support vector. SVM is a sophisticated supervised technique that performs well on smaller but complex datasets [14]. Support Vector Machines (SVMs) can be used for regression and classification tasks, however they perform better in classification issues. They were quite popular when they were first developed in the 1990s, and they are now the go-to method for a high-performing algorithm with minimal adjustment. It is a supervised machine learning problem in which we attempt to identify a hyperplane that best separates the two classes. Don't mix SVM with logistic regression.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

Both the algorithms strive to discover the optimal hyperplane, however the fundamental distinction is logistic regression is a probabilistic approach whilst support vector machine is based on statistical approaches. Figures 3 and 4 describes SVM.

3.4. Random Forest Algorithm

Random forest is also a nonlinear classifier. It is based on decision trees. A decision tree follows a flowchart layout with several nodes. The nodes are directed, starting from a mutual root and ending with a leaf. Each node represents a single feature using true-false statements. The final leaf determines the categorization. In this scenario, the leaves are made up of ones and zeros, and the assertions are based on a numerical threshold value[1-4]. The following stages describe the working Random Forest Algorithm:

Step 1: Take random samples from a specified data or training set.
Step 2: For each set of training data, this algorithm will create a decision tree.
Step 3: The choice tree will be averaged to determine the vote.
Step 4: Finally, choose the predicted outcome with the most votes as the final result. Figure 5 demonstrates the working.

Algorithm 3: Random Forest
Inpu : dataset df, x_{train}, x_{test} − independent variable, y_{train}, y_{test} target variable
Output: target y_{preds} with tract infection of UTX_DIAGONSIS
A is training set TS: = (a_1 , b_1 . . . . . . ,(a_n b_n)
training set(TS) = images
F − feature,
N(t) = nooftrees
Forest = E
begin
function RandomForest(H, R)
I ← Ø
for i€1, ...E do
$H^{(i)}$ ← A bootstrap sample from TS
a_i ← Randomized Tree Learn ($H^{(i)}$, R)
I ← I ∪ (a_i)
end for
return I
end function
function RandomizedTreeLearn(H, R)
At each node:
F ← very small subset of R
split on best feature in F
Return the learned tree
end function
End for
End [4]

3.5.Logistic Regression

Logistic regression is a simpler and more efficient approach to binary and linear classification issues. It is a classification model that is simple to implement and provides excellent performance with linearly separable classes [14]. Logistic regression is best thought of as linear regression applied to classification challenges. Logistic regression models a binary output variable using the logistic function. The major distinction between linear regression and logistic regression is that logistic regression's range is limited to 0 and 1. Furthermore, unlike linear regression, logistic regression does not assume a linear relationship between input and output variables [10].

$$Logistic\ function = \frac{1}{1+e^{-x}}$$

The function is represented in Figures 6 and 7.

4. Machine Learning in Detecting Urinary Tract Infection

4.1. Decision Trees

Decision trees were employed as part of the analysis for urinary tract infection (UTI) detection, aiding in the classification of relevant features and facilitating the identification of patterns indicative of UTI presence. Using straightforward decision-making procedures, decision trees—a decision structure like a tree—split vast volumes of material into manageable

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

groupings. They are made up of leaves and branches, each of which represents an exam that needs to be completed. The test that has to be run is indicated by the decision node, and the tree branches without losing any data. A decision node is produced at the end of a branch if the classification operation cannot be completed there. One of the classes to be decided based on the data is the leaf that forms at the end of a branch if a specific class is generated there. Starting with the root node, the decision tree process moves down the nodes in order until it reaches the leaf [24]. Using wearable IoT devices, researchers have created a human collapse detector to track the actions of the elderly. The detector detects human collapses in four directions using dynamic time tracking (DT) and machine learning. We present an effective technique for mechanically identifying natural activities with five triaxial wireless accelerometers and a wireless heart scale director. 30 natural gymnasium activities performed by 21 individuals were used in two labs to test the method. Based on dependent and independent training, the identification accuracy rates were 94.6% and 56.3%, respectively, according to the results. As the amount of heart scale data grew, the accuracy rose by 1.2% and 2.1%. When determining the type of activity without considering intensity levels, the independent performance was attained at 80.6% [12].

4.2.Support Vector Machines

Originally created for the purpose of classifying two-class linear data, Support Vector Machines (SVM) is a statistical learning-based classification technique that has since been expanded to identify multi-class and non-linear data. The SVM's basic method is to use the hyperplane specification to determine which decision function will best divide the two classes. Finding a linear discriminant function that faithfully captures the essence of the training data and aligns with statistical learning theory is the goal of the SVM regression technique. The nonlinear states in the regression are processed using kernel functions. SVMs are utilized to analyze and classify various features extracted from patient data, such as clinical symptoms, laboratory test results, and medical histories, to identify patterns indicative of UTI presence [24]. Non-linear classifiers can be used in place of linear classifiers when data cannot be split linearly. In this instance, SVM converts to a high-dimensional feature space that is readily classified using the linear method from the original input space. Research on supervised machine learning (SVM) classifiers for classification and regression problems is covered in this area. It was suggested to use an active-learning data selection accelerator (ALDSA) and an SVM accelerator (SVMA) in a biomedical processor with programmable parameters for machine learning accelerators. The ECG-based arrhythmia detection and the EEG-based seizure detection systems used the least amount of energy possible. A probabilistic modeling technique for categorizing abnormal gaits utilizing SVM classifier and HMM data was reported by Mannini et al. In terms of energy consumption, obtrusiveness, identification accuracy, and flexibility, Lara and Labrador assessed 28 systems. Using feature selection and SVM, Shan and Yuan created a prototype fall detector that can identify potential falls before they happen. Sun et al. examined participant physiological data prior to and following mental stress in order to develop a physiological sensor-based mental stress detection method [12]. By leveraging SVM's ability to handle non-linear relationships and high-dimensional feature spaces, researchers and clinicians can develop accurate and reliable UTI detection systems, aiding in the prompt diagnosis and management of this common medical condition.

4.3.Random Forest

Random Forest, another machine learning algorithm, has been utilized in urinary tract infection (UTI) detection as well. Random Forest learns to distinguish between patterns associated with UTI-positive and UTI-negative cases by constructing multiple decision trees using different subsets of the data and features. Once trained, the Random Forest model can be applied to new patient data to predict the likelihood of UTI presence based on the observed features. This

e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

predictive capability enables early detection and intervention, facilitating timely treatment and improved patient outcomes. Ensemble classification techniques, like Breiman's Random Forest (RF) method, employ several classifiers as opposed to just one. RF is a multi-tree classifier that uses randomly generated data from real-world scenarios. An input vector is provided to every tree in the forest, and a result is generated for every tree. The class with the greatest votes is selected as the outcome using the RF algorithm. Using the best randomly chosen variable on each node among all variables, it splits each node into branches. Selected bootstrap samples and randomly chosen n estimators are used in the RF method to achieve node separation. Every decision tree is complete and has not been trimmed. Random Forest's ability to handle non-linear relationships and feature interactions, along with its robustness to overfitting, makes it a valuable tool in developing accurate and reliable UTI detection systems. Through leveraging Random Forest algorithms, researchers and clinicians can enhance UTI diagnosis processes, contributing to better patient care [24].

4.4.Artificial Neural Network (ANN)

The basic building block of the human brain is a neuron. It is a complicated system that is capable of thinking, remembering, and solving problems. Artificial neural networks (ANNs) are biological neural network-inspired information processing systems that exhibit biological neural network-like performance traits. ANNs may mimic several key aspects of the functioning of the human brain, including learning from data, generalization, and infinite variability. The artificial neuron, also known as the process element, is the smallest unit that forms the foundation of an ANN [24]. Information that comes into the cell from other cells or external surroundings is known as an input, and it is based on examples that the network is meant to learn. Weights show how a preceding layer was impacted by the input set or another processed element. ANNs learn to recognize complex relationships between these features and UTI presence through interconnected layers of neurons. Deep learning has received a great deal of attention because of their potential for wearable sensor-based human activity recognition. Techniques for effectively managing and utilizing IoT data have been developed as a result of the data's explosive rise in a variety of industries, including sports and healthcare. For example, Ravi et al. suggested a way to use DL to classify large volumes of data obtained from IoT wearable gear, integrating shallow features with attributes from inertial sensor data for better categorization. Hammerla et al. used deep, convolutional, and recurrent algorithms in deep learning to track wearables connected to the Internet of Things and record human daily activities. While Kwon et al. suggested an unsupervised learning method to identify user motions using data from smartphones, Eskofier et al. employed DL to monitor patients with Parkinson's disease [12]. The versatility of ANNs lies in their ability to adapt to non-linear relationships and intricate data structures inherent in UTI diagnostics. By leveraging ANNs' robust analytical capabilities, researchers and clinicians can refine UTI detection methodologies, enhancing diagnostic accuracy, and ultimately elevating the standard of care for patients afflicted with UTIs.

5. Figures

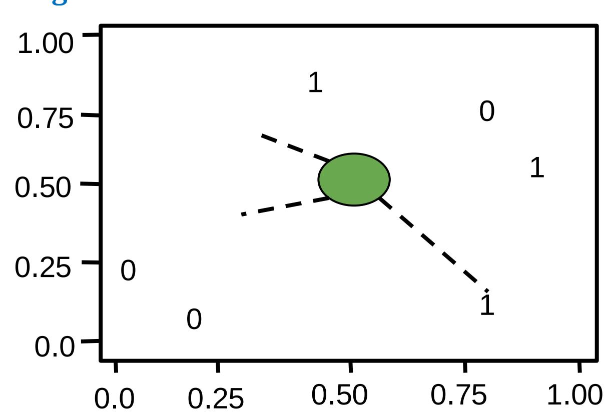

Figure 1 KNN Sample Distribution

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

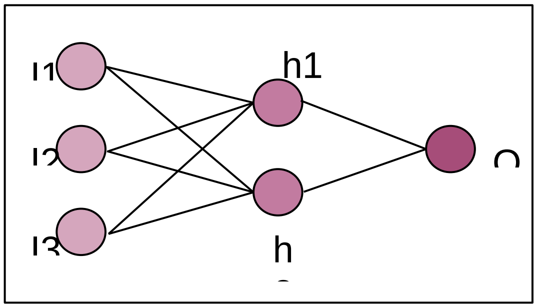

Figure 2 ANN Architecture

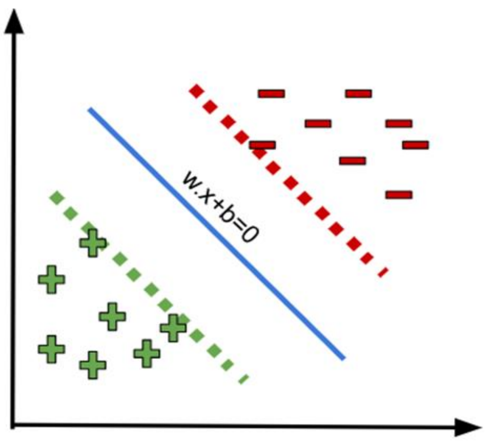

Figure 3 SVM Model

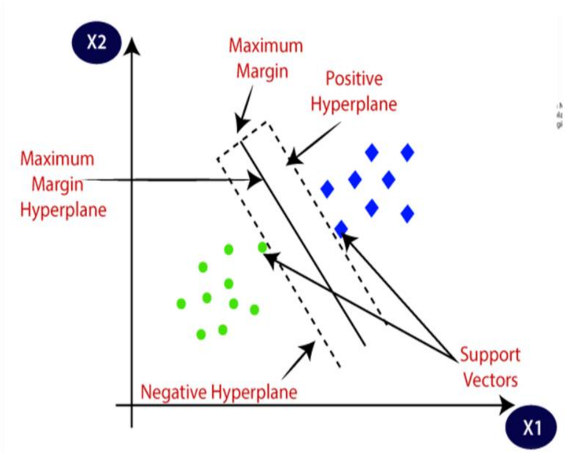

Figure 4 SVM Model

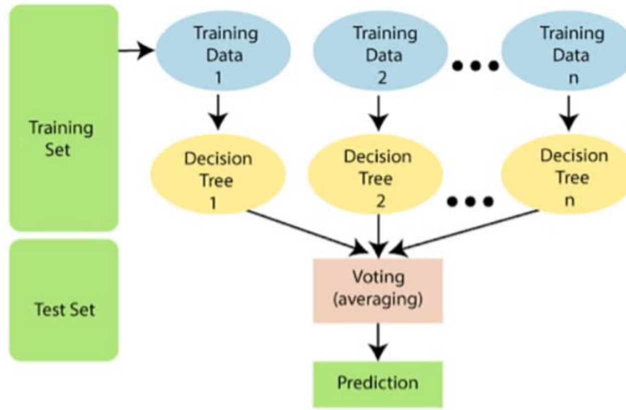

Figure 5 Random Forest Architecture

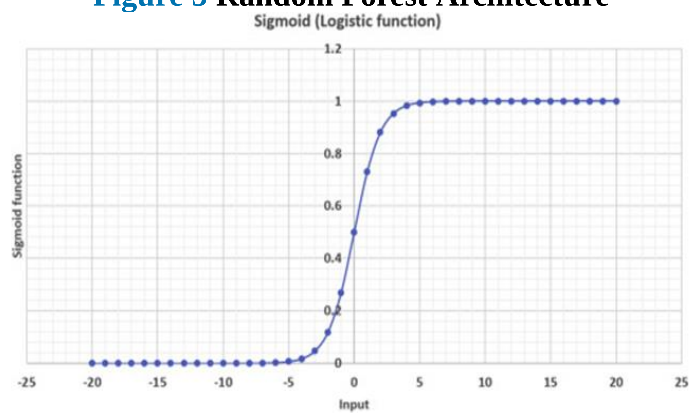

Figure 6 Logistic Function

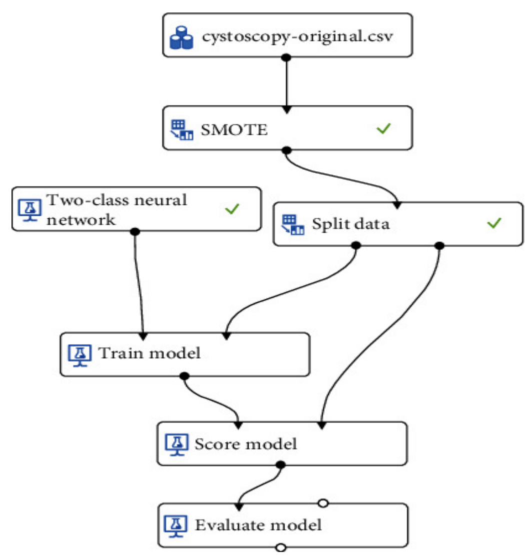

Figure 7 Flow chart of Logistic Regression

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

6. Results

Uropathogenic E. coli and P. mirabilis are pathogens of the urinary tract, a common site of bacterial infection in humans. Proteins that are required for, or contribute to, the virulence of each pathogen have been identified, and these discoveries have contributed to our understanding of the mechanisms of pathogenesis [2, 17]. Overall it is easy to see that the nonlinear classifiers performed better. However, one exception is the support vector machines (polynomials). Out of all the classifiers random forest was a better choice. Worth noting is that neural networks only used 8-fold cross-validation, thus yielding a small error in comparison [4]. The neural network could also have been altered a lot and tuned longer. it is seen that SVM, RF, and ANN classifiers have generally very high classification accuracy. If we evaluate the ANN classifier it has the highest classification accuracy value, the positive value of the obtained model for cystitis finds the probability of really having cystitis. In addition, the probability of a patient with nonspecific urethritis diagnosed as truly non-specific urethritis was determined [24, 26]. When the patient has cystitis, the positive result of the model (i.e. the ability to recognize the cystitis disease) is highly accurate. The rate of recognizing non-specific urethritis was much higher. RF algorithm has shown better performance than DT on UTI classification. In addition, the RF algorithm has very high specificity and positive predictive value along with the ANN algorithm [24].

References

[1]. Min Hyuk Choi, Dokyun Kim, Yongjun Park, Seok Hoon Jeong: Development and validation of artificial intelligence models to predict urinary tract infections and secondary bloodstream infections in adult patients. Journal of Infection and Public Health 17 (2024) 10–17, https://doi.org/10.1016/j.jiph.2023.10.021

[2]. Natasha Liou, Trina De, Adrian Urbanski, Catherine Chieng, Qingyang Kong, Anna L. David, Rajvinder Khasriya, ArturYakimovich, & Harry Horsley: A clinical microscopy dataset to develop a deep learning diagnostic test for urinary tract infection. Scientific Data | (2024) 11:155 | https://doi.org/10.1038/s41597-024-02975-0

[3]. Nithesh Naik, Ali Talyshinskii, Dasharathraj K. Shetty, B. M. Zeeshan Hameed, Rano Zhankina, Bhaskar K. Somani: Smart Diagnosis of Urinary Tract Infections: is Artificial Intelligence the Fast-Lane Solution. Current Urology Reports (2024) 25:37–47 https://doi.org/10.1007/s11934-023-01192-3

[4]. Abdullah Alqahtani, Shtwai Alsubai, Adel Binbusayyis, Mohammed Shah, Abdu Gumaei and Yu-Dong Zhang: Prediction of Urinary Tract Infection in IoT-Fog Environment for Smart Toilets Using Modified Attention-Based ANN and Machine Learning Algorithms. Appl. Sci. 2023, 13, 5860. https://doi.org/10.3390/app13105860.

[5]. Paweł Wityk, Patryk Sokołowski, Małgorzata Szczerska, Kacper Cierpiak, Beata Krawczyk, Michał J. Markuszewski, Optical method supported by machine learning for urinary tract infection detection and urosepsis risk assessment. J. Biophotonics. 2023;16:e202300095. https://doi.org/10.1002/jbio.202300095

[6]. Anirudh Sridhar, Ahmed Mawia, A L Amutha: Mobile Application Development for Disease Diagnosis based on Symptoms using Machine Learning Techniques. Procedia Computer Science 218 (2023) 2594–2603

[7]. Athul Janev , John S. Kang , Sung-Yong Park: A smartphone integrated paper (SIP)-based platform for rapid and on-site screening of urinary tract infections. Sensors and Actuators: B. Chemical 382 (2023) 133498, Department of Mechanical Engineering, San Diego State University, San Diego, CA, 92182, USA

[8]. Yanjie Zhao, Chaoyue Chen, Zhouyang Huang, Haoxiang Wang, Xin Tie, Jinhao Yang, Wenyao Cui and Jianguo Xu: Prediction of upcoming urinary tract infection after intracerebral hemorrhage: a machine learning approach based on statistics collected at multiple time points. Front. Neurol., 14 September 2023 Sec. Neurocritical and Neurohospitalist Care Volume 14 - 2023 | https://doi.org/10.3389/fneur.2023.1223680.

[9]. KurtG. Naber, FlorianWagenlehner, Michael Kresken, WendyY. Cheng, Maryaline Catillon, Mei Sheng Duh, LouiseYu, Anamika Khanal, Aruni Mulgirigama, AshishV. Joshi, Shinyoung Ju & Fanny S. Mitrani-Gold: Escherichia coli resistance, treatment patterns and clinical outcomes among females with uUTI in Germany: a retrospective physician-based chart review study. Scientific Reports | (2023) 13:12077 | https://doi.org/10.1038/s41598-023-38919-8

[10]. Tsai-Jung Chen, Yu-Huang Hsu, and Chieh-Hsiao Chen: Comparison of Neural Network and Logistic Regression Analysis to Predict the Probability of Urinary Tract Infection Caused by Cystoscopy. Hindawi, BioMed Research International, Volume 2022, Article ID 5775447, 6 pages ,https://doi.org/10.1155/2022/5775447.

[11]. Shuen-Lin Jeng, Zi-Jing Huang, Deng-ChiYang, Ching-HaoTeng, & Ming-ChengWang: Machine learning to predict the development of recurrent urinary tract infection related to single uropathogen, Escherichia coli. Scientific Reports | (2022) 12:17216. https://doi.org/10.1038/s41598-022-18920-3

[12]. Mohammad Shehab, Laith Abualigah, Qusai Shambour, Muhannad A. Abu-Hashem, Mohd Khaled Yousef Shambourf, Ahmed Izzat Alsalibi, Amir H. Gandomi: Machine learning in medical applications: A review of state-of-the-art methods. Computers in Biology and Medicine 145 (2022) 105458

[13]. Anisha, Munish Sabharwal, Rohit Tripathi: Urinary Tract Infection Analysis using Machine Learning based Classification and ANN- A Study of Prediction. International Journal on Recent and Innovation Trends in Computing and Communication. ISSN: 2321-8169 (2023) Volume: 11 Issue: 7s

[14]. Jesper Nyman: Machine learning approaches for detection of urinary tract infections. Department of Physics, Umea University, Sweden, 2020.

[15]. Grampurohit S and Sagarnal C, "Disease Prediction using Machine Learning Algorithms," In 2020 International Conference for Emerging Technology (INCET), 2020, pp. 1-7, doi: 10.1109/INCET49848.2020.9154130.

[16]. Shirin Enshaeifar, Ahmed Zoha, Severin Skillman, Andreas Markides, Sahr Thomas Acton, Tarek Elsaleh, Mark Kenny, Helen Rostill, Ramin Nilforooshan, Payam Barnaghi: Machine learning methods for detecting urinary tract infection and analyzing daily living activities in people with dementia. PLoS ONE 14(1): e0209909. (2019) https://doi.org/10.1371/journal.pone.0209909.

[17]. Mehmet Demirci, Özge Ünlü, Ayse, Istanbullu Tosun: Detection of O25b-ST131 clone, CTX-M-1 and CTX-M-15 genes via real-time PCR in Escherichia coli strains in patients with UTIs obtained from a university hospital in Istanbul. Journal of Infection and Public Health 12 (2019) 640–644, https://doi.org/10.1016/j.jiph.2019.02.017.

[18]. Enshaei Far, S., Zoha, A., Skillman, S., Markides, A., Acton, S. T., Elsaleh, T., ... & Barnaghi, P. (2019). Machine learning methods for detecting urinary tract infection and analyzing daily living activities in people with dementia. PloS one, 14(1), e0209909.

[19]. Kirk J. Wojno, David Baunoch, Natalie Luke, Michael Opel, Howard Korman, Colleen Kelly,

S. Mohammad A. Jafri, Patrick Keating, Dylan Hazelton, Stephany Hindu, Bridget Makhloouf, David Wenzler, Mansour Sabry, Frank Burks, Miguel Penaranda, David E. Smith, Andrew Korman, and Larry Sirls: Multiplex PCR Based Urinary Tract Infection (UTI) Analysis Compared to Traditional Urine Culture in Identifying Significant Pathogens in Symptomatic Patients. © 2019 Pathnostics. Published by Elsevier Inc. https://doi.org/10.1016/j.urology.2019.10.018

[20]. Sinem Diken Gür, Monireh Bakhshpour, Adil Denizli: Selective detection of Escherichia coli caused UTIs with surface imprinted plasmonic nanoscale sensors. Materials Science & Engineering C 104 (2019) 109869, https://doi.org/10.1016/j.msec.2019.109869.

[21]. Govindan Rajiv Gandhi, Muthuchamy Maruthupandy, Govindan Ramachandran, Muthu Priyanga, Natesan Manoharan: Detection of ESBL genes from ciprofloxacin resistant Gram negative bacteria isolated from urinary tract infections (UTIs). Frontiers in Laboratory Medicine 2 (2018) 5–13, https://doi.org/10.1016/j.flm.2018.01.001.

[22]. Burange T P and Chatur P N, "Analysis of Symptoms Wise Disease Inference System Using Data Mining Technique," In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1160-1165, doi: 10.1109/ICCONS.2018.8663062

[23]. GWalentowicz, M., Krzemiński, D., Kopański, Z., Liniarski, M., Tabak, J., Dyl, S., ... & Mazurek, M. (2017): Selected aspects of the urinary system anatomy and physiology. Journal of Clinical Healthcare, 89(2017_3), 01-05.

[24]. Ilker Ali Ozkana, Murat Koklu, Ibrahim Unal Sert: Diagnosis of urinary tract infection based on artificial intelligence methods. Computer Methods and Programs in Biomedicine 166 (2018) 51–59

[25]. Alessio Mancini, Leonardo Vito, Elisa Marcelli, Marco Piangerelli, Renato De Leone, Sandra Pucciarelli and Emanuela Merelli: Machine learning models predicting multidrug resistant urinary tract infections using "DsaaS". 13th Bioinformatics and Computational Biology Conference - BBCC 2018

[26]. Alfred Lok Hang Lee, Curtis Chun Kit To, Janus Siu Him Wong, Ronald Cheong Kin Chan: Artificial intelligence in UTI – can machine learning models accurately predict antimicrobial susceptibility in urinary culture isolates?.

[27]. Bono, M. J., & Reygaert, W. C. (2021). Urinary tract infection. StatPearls [Internet].

[28]. Ameen, W. A., & Hummade, S. H. (2015). Risk factors for urinary tract infection among women at productive age at Babel Technical Institute in Hilla city. Age, 16(20), 44.

[29]. Sordo, M. (2002). Introduction to neural networks in healthcare. Open Clinical: Knowledge Management for Medical Care.

[30]. Heckerling, P. S., Canaris, G. J., Flach, S. D., Tape, T. G., Wigton, R. S., & Gerber, B. S. (2007). Predictors of urinary tract infection based on artificial neural networks and genetic algorithms. International Journal of Medical Informatics, 76(4), 289-296.

[31]. Huang, Y. L., & Chen, H. Y. (2007). Computer-aided diagnosis of urodynamic stress incontinence with vector-based perineal ultrasound using neural networks. Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology, 30(7), 1002-1006.

[32]. Pérez, F. M., Chamizo, J. M. G., Payá, A. S., & Fernández, D. R. (2008). A robust model of the neuronal regulator of the lower urinary tract based on artificial neural networks. Neurocomputing, 71(4-6), 743-754.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0145
e ISSN: 2584-2854
Volume: 02
Issue: 04 April 2024
Page No: 1097-1109

[33]. Monadjemi, S. A., & Moallem, P. (2008). Automatic diagnosis of particular diseases using a fuzzy-neural approach. International Review on Computers & Software, 3(4), 406-411.
[34]. Gil, D., Johnsson, M., Chamizo, J. M. G., Paya, A. S., & Fernandez, D. R. (2009). Application of artificial neural networks in the diagnosis of urological dysfunctions. Expert systems with applications, 36(3), 5754-5760.
[35]. Ana L. Flores-Mireles, Jennifer N. Walker, Michael Caparon, and Scott J Hultgren: Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Micobiol. 2015 May; 13(5): 269-284, doi: 10.1038/nrmicro3432
[36]. Dahiwade D, Patle G and Meshram E, "Designing Disease Prediction Model Using Machine Learning Approach," In 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 2019, pp. 1211-1215, doi: 10.1109/ICCMC.2019.8819782.