International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

Urinary Tract Infection Analysis using Machine Learning based Classification and ANN- A Study of Prediction

Anisha[1], Munish Sabharwal[2], Rohit Tripathi[3]
[1]School of Computing Science and Engineering
Galgotias University
Greater Noida, India
anishanagpal@outlook.com
[2]School of Computing Science and Engineering
Galgotias University
Greater Noida, India
mscheckmail@yahoo.com
[3]Department of Electronics Engineering
JC Bose University of Science and Technology, YMCA
Faridabad, India
rohittripathi30.iitd@gmail.com

Abstract—Urinary tract infection is the most frequently diagnosed infection among humans. A urinary tract infection (UTI) affects the areas of urinary system which includes the ureters, bladder, kidneys and urethra. The primary infected area of urinary system involves the lower tract i.e. bladder and urethra. The infection in bladder is painful as well as uncomfortable but if it spreads to kidneys, it can have severe consequences. Women are more susceptible to urinary infection in comparison to men due to their physiology. This paper aims to study and assess the impact and causes of urinary tract infection in human beings and evaluate the machine learning approach for urinary disease forecasting. The paper also proposed machine learning based methodology for the prediction of the urinary infection and estimating the outcomes of the designed procedures over real-time data and validating the same. The paper focuses to get high prediction accuracy of UTI using confusion matrix by Machine Based Classification and ANN technique. Some specific parameters have been selected with the help of Analysis of variance technique. The naive bayes classifier, J48 decision tree algorithm, and Artificial neural network have been used for the prediction of presence of urinary infection. The accuracy achieved by the proposed model is 95.5% approximately.

Keywords-Urinary tract infection, Machine learning, Artificial neural network, Urine Infection, Bacteria, Diseases.

I. INTRODUCTION

An infection refers to the incursion and growth of microorganisms that are not normal to the body. The multiplication of these microorganisms is fast and can occur anywhere in the body causing several health problems. Urinary Tract Infection (UTI) occurs in the Urinary System which consists of kidneys, ureters, bladder and urethra. Mostly infection appears in the lower tract i.e. bladder and urethra but if it get severe, it can appear in upper tract which includes kidneys and ureters [1, 2]. UTI is among a most prevalent diagnosed infection that affects both men and women and it is more common in women because of their physiology. In basic terms, it is a disease that approximately all women will face at some point during their life. When the UTI begins with the lower tract, it is called as cystitis (bladder infection) and when it moves to upper tract, the infection is called as Pyelonephritis (kidney infection) [3, 4].

Urine infection (UI) is a highly common infection, and if it left untreated it can cause several diseases related to kidneys, bladder and liver. In fact, UI affects more than 150 million people all around the world each year [5]. Also, the World Health Organization (WHO) released a survey study showing that there were over 1 million hospitalizations in the United States in 2011. Global health spending is projected to reach $8.7 trillion by 2020, due to the rising global population, disease insecurity, & the mobile healthcare industry [6].

The pathogens such as bacteria, fungi, and virus are the most common cause of infection in human beings. The symptoms of having infection in bladder and kidney can vary; in cystitis it causes painful and regular urination, and Pyelonephritis causes high fever and flank pain. The infection's prevalence among children and the elderly is still unknown, and it is currently being investigated. [7].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

UI 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 equipments 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. 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 [8, 9].

The paper represents the understanding of cause and effects of the Urinary Tract Infection and analyzing the requirements of forecasting the disease. The aim is to use the methodology driven from artificial neural networks and machine learning based techniques for predicting the urinary infection with considerable accuracy. Therefore the objectives of the study are as follows:

i. To study and evaluate Urinary infection for causes and effects.

ii. To study and evaluate the requirement of the Machine Learning process for the forecast of urinary disease.

iii. To present a Machine Learning technique for the prediction of the urinary infection.

iv. To evaluate the outcomes of designed techniques over real-time data and to validate the same using the comparison study via learning/training and validation data generated at phases of the process.

The challenging work is to correctly predict the disease on the basis of the symptoms of the patient. Also, accurate predictions and early patient care is possible due to the growth in amount of data in medical and healthcare sectors. This paper proposes a general prediction of UTI using ANN and Machine Learning algorithms.

A. Artificial Neural Networks (ANNs)

An artificial neural network is a computational system that is designed to simulate the parallel structure of human brain. An ANN is a parallel processing network made up of strongly interconnected processing elements (neurons) which is inspired from the biological nervous system. The network function is mainly determined through the connections among elements and every connection among two neurons consists of weights attached to it. The architecture of a neural network consists of group distinct interconnected layers. A layer in the network is a subgroup of the processing components. The first layer in artificial neural networks contains nodes which take initial data termed as input layer and the layer which provides the result depending upon the input is termed as output layer. Additional layers of units, referred to as intermediate layer or hidden layers which exist between the input & an output layer is responsible for all the computations.

The standard neural network is depicted in Figure 1. By altering the connections' (weights) values between components, a neural network can be trained to perform a specific function. In Medicine, Artificial Neural Network for Medical Diagnosis is currently having significant amount of coverage and it is expected to become more commonly utilized in biomedical systems in the coming years. The primary reason is that there is no constraint for the solution to be in linear form. Neural Networks are suitable for identifying diseases by utilization of scans and there is no requirement to include a complex algorithm for identification of disease. Neural networks learn by training examples, they do not need to know the particulars of how to identify a disease [10, 11].

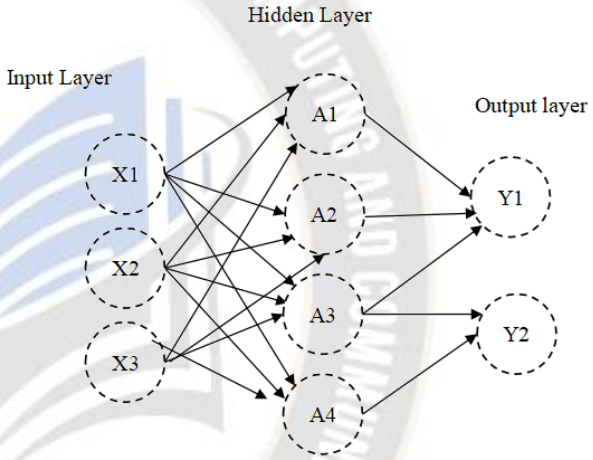

Figure 1: Artificial Neural Network [10]

B. Machine Learning (ML)

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). Complex models and medical diagnosis can be derived from machine learning, therefore, revealing new ideas to clinicians and specialists [12]. 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 prescription 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

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

to other traditional approaches, these models are commonly utilized to examine diagnostic analysis [14].

Therefore, taken 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.

II. RELATED WORK

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 solution which is too complex to find. Neural Networks is widely used in medical domain to solve all the problems related to clinical diagnosis, prediction, image processing & interpretation, drug development and pattern recognition [15].

Heckerling et al. [16] 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. The suggested work evolved five variable set that can classify the UTI cases and non-infection areas ranging from 0.853 to 0.792. The results of network influence analysis shows few variables predicted UTI in an unexpected way and also have an interaction with other variables in forming prediction.

Huang and Chen [17] investigated a computer-aided vector-based perineal ultrasound method for diagnosing urodynamic stress incontinence (USI). The proposed method uses the morphology and function of lower tract of urinary system as a feature for multilayer perception neural network. The proposed CAD system uses K-fold cross validation for performance estimation and achieves the accuracy of 91.7%.

Pérez et al. proposed a system based on a multiagent system model in which every neuronal center corresponds to an agent [18]. 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 by dysfunctions in neuronal centers as a minimum.

Moallem and Monadjemi [19] explored the use of Artificial neural network in the diagnosis of typical diseases. The actual procedure for medical diagnosis, as used by doctors, was examined, and translated into a machine implementable form. The findings of the experiments, as well as the benefits of utilizing a fuzzy method, were also considered. The proposed system achieves the accuracy of 97.5 % in detecting abnormal cases.

Gil et al. [20] looked at how well certain ANN models worked as methods for assisting in the medical diagnosis of urological dysfunctions. They generated 1 type of supervised neural network and 2 types of unsupervised neural networks. Since neurological dysfunctions are complicated to diagnose, this scheme is intended to assist urologists in getting a diagnosis for complicated multi-variable diseases and to minimize expensive & painful medical procedures. Health registers of patients with urological dysfunctions were used in clinical research.

Altunay et al. [21] examined the uroflowmetric data and helped physicians with their diagnosis. They developed an expert pre-diagnosis device that evaluates potential symptoms from uroflow signals automatically. ANN was used to generate a pre-diagnostic result. ANN's outputs are divided into 3 categories: "healthy", "pathologic", and "possible pathologic". The ANN is trained through using the back-propagation process, and the extracted features are ANN's inputs, which are selected centered on urology specialists' recommendations. A dataset of patients that have previously been diagnosed by specialists is utilized to train and test the proposed method.

Taylor et al. [22] introduced a model for prediction of urinary tract infection for the patients in emergency department. The aim of the work is to have a comparison of all machine learning models after their training and validation with a large dataset of ED patients. The results present in the paper clearly shows that XGBoost is the leading algorithm for diagnosing positive urine culture accurately.

Almansour et al. [23] seek to aid in Chronic Kidney Disease (CKD) prevention through using machine learning methods to early diagnosis of CKD. Kidney diseases are illnesses that affect the kidney's ability to function normally. They concentrate their research on applying various classification algorithms to a dataset of 400 patients with 24 attributes linked to chronic kidney disease diagnosis. ANN & Support Vector Machine (SVM) was used as classification techniques in this research (SVM). To conduct the experiments, the mean of the corresponding attributes was used to substitute all missing values in the dataset. Then, after tuning the parameters and running multiple experiments, the optimal parameters for the ANN & SVM methods were calculated.

Enshaeifar et al. [24] uses Internet of Things with machine learning to keep a check on person's well being and health. He proposes an algorithm for UI detection and tracking the changes in activity patterns for early detection of cognitive or health impairment and provide preventative and individualized care. Also, an Isolation Forest (iForest) technique is used to create a comprehensive perspective of daily activity patterns. The proposed work summarize the algorithms and examines the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

work's evaluation using huge amount of real-world data from a trial with dementia patients and caregivers.

Nyman & Jesper [25] 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 which is shown in Table 1.

TABLE 1: UI PARAMETERS, SAFE AND UNSAFE VALUES

Attribute	Borderline	Safe	Unsafe	Disease Associated	References
PH Value	4, 8	4.5-8	< 4 or >9	Diabetes, dehydration	[26]
Specific Gravity	1.025	1.005-1.030	1.035-1.040	Cystitis	[26]
Leukocytes	4	0-5	>5	Kidney abnormalities	[27]
Ketones	11	<10 mg/dl-Negative	>40 mg/dl	Diabetes	[28]
Nitrite	-	Negative	Positive	Urethritis, Cystitis	[26]
Glucose	0.7-0.8	0-0.8	>0.8 mmol/l	Diabetes mellitus or renal glycosuria	[29]
Bilirubin	1.1-1.2	0.3-1.2 mg/dl	>1.2 mg/dl	Gallstone, Liver disorders, hepatitis	[30]
urobilinogen	1.7-1.8 eu/dl	0.1-1.8 eu/dl	>2 eu/dl	hemolytic jaundice, Liver disease	[31]
Blood	2	0-3 RBCs	>3 RBCs	kidney tumors, Inflammation	[30]
Protein	19md/dl	0-20 md/dl	>20md/dl	Kidney infection or disorders	[26]

Innovations in information & communication technology (ICT) have become a vital foundation for delivering cost-effective solutions in a range of industries, including healthcare, logistics, and agriculture. Furthermore, with the IoT as the main driver of ICT technologies, the healthcare industry has been driven down a road of optimum resource usage and ubiquitous medical service provisioning. IoT technology consists of a network of lightweight, internet-connected sensors that can collect and transmit data to a remote location in a timely manner [32].

Bhatia et al. [33] proposed a novel framework for monitoring, diagnosing, and predicting urine infection in home-centric environment by utilizing internet of things. The framework uses a model consisting of several layers which includes perception layer, analysis layer, extraction layer, prediction layer and visualization layer for diagnosing urine infection. The model is used for prediction of urine infection using t-ANN and attains the accuracy of 93.69%.

Bhatia et al. [34] presents a predictive system for urine based diabetes. The system consists of 4 layers such as data acquisition, data classification, mining and extraction and prediction and decision making for monitoring and predicting diabetes oriented infection. The proposed system uses recurrent neural networks for prediction and achieves the accuracy of 98.9%. WEKA tool is used for the implementation of the methods for comparative analysis.

Gupta et al. [35] proposed a fog computing-based framework using XGBoost algorithm that uses data collected from IoT based sensors and Infection risk factor for computation. The proposed model achieves the accuracy of 91.45% and considerably improved the performance level of the novel framework in comparison with the other baseline strategies.

III. RESEARCH METHODOLOGY

A. Problem Formulation

A rise in patient infected from UTI over the years suggested immense interest in healthcare providers around the globe. In addition, in a small variety of medical tools, integrating machine learning in combination with sophisticated data analysis techniques in this area has become important. The actual state of work for the estimation of urinary infection revolves around the form of data gathered and similar predictions.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

In this paper, the research approach is having certain steps and the major contributions lie with the learning process and data collection. As for the learning process, data collection is considered from pathology, online Health Forum such as MedHelp. The collected data has all the necessary parameters to enhance the prediction process. Also, some real-time data which is considered for model validation.

In the methodology, the data obtained from data collection step is pre-processed and then graded to get the exact degree or create the actual description of the infection. Also, some prominent features are selected from whole database using analysis of variance. Data generated in the above step is filtered & modeled for deciding the infection and non-infection class. At last, ANN is utilized for decision making in the process of prediction. The data for the phase-I is taken from a nearby pathology and an online forum named "MedHelp".

The Urine Culture reports of patients taken from the pathology (with consent) consist of all the parameters and their values which are needed to detect Urine Infection. Also, the data taken from the online forum consists of responses of patients & doctors that cover various warnings, precautions, and information on the chronic diseases. Both patient and doctor comments are processed in the cloud regarding various diseases and disorders. There is a dataset of comments received in 2018, 2019, and 2020 to conduct research over the disease. Then these statements are stored in a local archive. Until documents are entered, the texts which can be processed into records need to be pre-processed.

Tokenization in pre-processing is a method where a paragraph is divided into sentences, each of which is a token. Words are linked with stop words within a sentence and these stop words usually do not contribute to a sentence's context. To delete stop words, the pre-processing stage uses a stop word dictionary or pre-built library depending upon the requirement. The absolute rating of a word in a text or corpus is referred to as its term frequency, and the rating of a term in the entire corpus is referred to as its inverse document frequency.

Here is the expression for frequency-inverse document frequency estimation. This term is designated with the TF value of the "t" term in a "d" document, which is similar to the number of times "t" appears shared through the number of terms in "d" as described in equation (1). The inverse document frequency (IDF) value for a word is dependent on a collection of records or half a set of records, taking any word occurrence in each equation (2). The Term frequency-inverse document frequency (TF IDF) meaning is just the multiplication of term frequency (TF) and Inverse document frequency (IDF) variables, the formula representing in Equation (3). The algorithm for text preprocessing is explained in Algorithm 1.

$$\mathrm{TF}(t, d) = (f(t)/\ \text{number of words in d}) \quad (1)$$

$$\mathrm{IDF}(t) = \ln(\text{ndocuments}/(\text{ndocuments containing t}) \quad (2)$$

$$\mathrm{TF} - \mathrm{IDF}(tk, dj) = \mathrm{TF}(\text{word}) * \mathrm{IDF}(\text{word}) \quad (3)$$

Algorithm 1: Text Preprocessing

Input: Dataset will be gathered from the defined sources.

1. The output is a preprocessed dataset for the disease. At the time when all words in the document in a file are exhausted.

   i. Digital transactions

   ii. Stemming the words.

   iii. Removing the punctuation.

   iv. Stop the word removal.

2. Calculate the TF value.

   i. Use the inverse document frequency formula to determine how many times your document is viewed.

   ii. Use the TF–IDF on the matrix and set a minimum threshold value.

If the TF-IDF score is higher than the threshold i.e., 0.53, append the word into the document. To achieve high-performance accuracy, the two classification techniques that are Naïve Bayes and J48 are used.

- Naive Bayes Classifier

Bayes Theorem is like a calculation that assigns various classes and characteristics to a dataset's unknown attributes. It decides the highest probability result. This classifier assumes that the features are not contingent and while the inclusion of one variable does not mean the existence of any other component, a feature must be found if it exists. The potential of property can come into play as all features are weighed, even without including other features. The diagrammatic representation of the Naive Bayes classifier is shown in Figure 2 in which all the features are linked with the label.

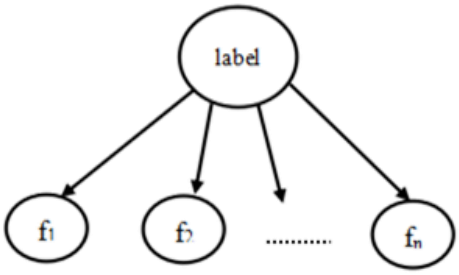

Figure 2: Naive Bayes Classifier [36]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

Nowadays, Naive Bayes technology is mainly utilized for big datasets. It performs very well to achieve excellent results. The Bayes hypothesis operates on the world state's probability, considering the evidence that have seen. The probability of an event being real, as something else has already occurred at this moment. To measure unbiased coin, toss conditional chance, the formula will be:

$$P(Hyp|Evi) = P(EvilHyp) * P(Hyp)/P(Evi) \quad (4)$$

As a probability, where P(Hyp), is hypothesis possibility h being true. There is the possibility that P(Evi) is the evidence (unrelated to the hypothesis). P(Evi|A) is the true hypothesis possibility when it refers to the evidence provided. P(Hyp|Evi) is the hypothesis probability when there is evidence in favor of it.

- J48 Decision Tree

The J48 algorithm is utilized to categorize several applications and generate correct classification results. Quinlan's C4.5 approach implements J48 to generate a trimmed C4.5 decision tree. In order to form a decision each feature of the data set is divided into tiny subsets. J48 investigates the standardized data gain, which is genuinely the outcome of partitioning the information by selecting an attribute. If a subset has a place with a comparable class in all cases, the split tactics come to an end. J48 creates a decision node based on the class's predicted estimations. J48 decision tree can manage specific features, lost or incomplete attribute estimations, and variable attribute prices. Pruning can be used to increase accuracy in this case [37].

One of the main advantages of J48 algorithm is, J48 can help with not only creating accurate predictions from data but also interpreting its patterns. It refers to problems like missing data, numeric attributes, pruning, decision tree induction difficulty, predicting error rates, and rule generation from trees [38].

The whole research methodology is illustrated in Figure 3 which consists of data collection, data classification, data representation and learning process for prediction.

B. Step by Step detailing of the Proposed Methodology

Step 1: Data Collection

In this *phase of the work, the data is collected for the learning process, for which online sources like applications, doctors/diagnostics consultations, patient records/data, etc are considered.

Step 2: Data Pre-processing

The datasets are pre-processed utilizing NLP methods like punctuation removal, stop word removal, and tokenization. The TF–IDF measure is used for the collection of especially important features through gathered pre-processed data.

Step 3: Data Classification

The pre-processed data is then classified using ML classifiers [39] as naïve bayes and J48 decision tree, and the results generated after the classification are hereafter used for the prediction process and getting infection and non-infection class.

Step 4: Feature Extraction

In this phase, the real-time data is collected from the samples and with the help of SPSS tool and analysis of variance we have extracted some prominent features form the data base containing data from a pathology and online forums.

Step 5: Data Representation

The gathered data is then filtered for more processing and the produced data is aggregated in the format that the similar can be further utilized for prediction & comparison processing through ANN. From the aggregated data different degrees of the infection are computed and are grouped into the infection classes based on the data collected in the learning phase.

Step 6: Prediction Process

The data submitted at the learning phase and real-time data generated for the prediction of the urine infection is then considered for the prediction process. For the prediction ANN is being used.

Step 7: Confusion Matrix

A Confusion matrix [40] is an N x N matrix. It is utilized to assess a classification model's efficiency, where 'N' is a number of target groups. The matrix compares actual objective values to the predictions of the machine learning model. A 2 x 2 matrix is utilizing with 4 values for a binary classification query, as shown in Figure 4, to get a detailed picture of how properly our classification model is doing and find the accuracy using these actual and predicted values.

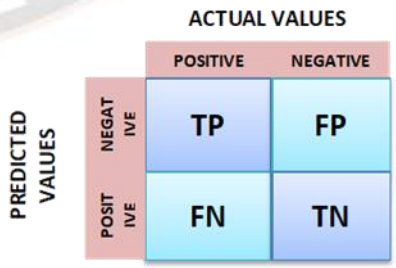

Figure 4: 2 x 2 Matrix with Four Values

i. Columns represent actual values of the target variable.

ii. Rows represent predicted values of the target variable.

iii. The target variable has 2 values: Negative or Positive.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

TP (True Positive)

i. The predicted value is the same as the actual value.

ii. The actual value was positive, and the value predicted by the model is also a positive value.

TN (True Negative)

i. The actual value was negative, and the value predicted by the model, is a negative value.

ii. The predicted value is the same as the actual value.

FP (False Positive) – Type 1 error

i. FP is also recognized as a Type 1 error.

ii. The actual value was negative, but the value predicted by the model is a positive value.

iii. The predicted value was falsely predicted.

FN (False Negative) – Type 2 error

i. FN is also recognized as a Type 2 error.

ii. The predicted value was predicted falsely.

iii. The actual value was positive, but the value predicted by the model is a negative value.

The accuracy of the prediction can be obtained by the confusion matrix values. The below equation 5 is used to calculate the accuracy [34]:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \quad (5)$$

IV. Implementation

A. *Classification*

In this paper, two classifiers are used for data classification for the gathered data i.e. naïve bayes and J48 decision tree in order to get infection and non-infection classes. The confusion matrix is generated for both the classifiers, and this confusion matrix provides the values of several parameters i.e., sensitivity, accuracy, precision, specificity, f score, and recall are calculated. Confusion matrix of the tested datasets using naïve bayes classifier and J48 is shown in Figure 5.

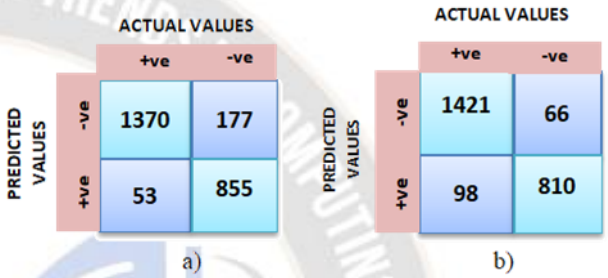

Figure 5: confusion matrix a) naïve bayes b) J48

Also, the values of the parameters obtained by the confusion matrix i.e. sensitivity, accuracy, precision, specificity, f score, and recall are shown in the Table 2.

TABLE 2: VALUES OBTAINED BY NAÏVE BAYES AND J48

Classifiers	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Score
Naïve Bayes	92.90	96.43	94.16	92.13	96.27	94.15
J48	93.15	95.56	89.20	95.56	93.54	94.54

The graphical representation of all the values of parameters obtained by both the classifiers is shown in Figure 6.

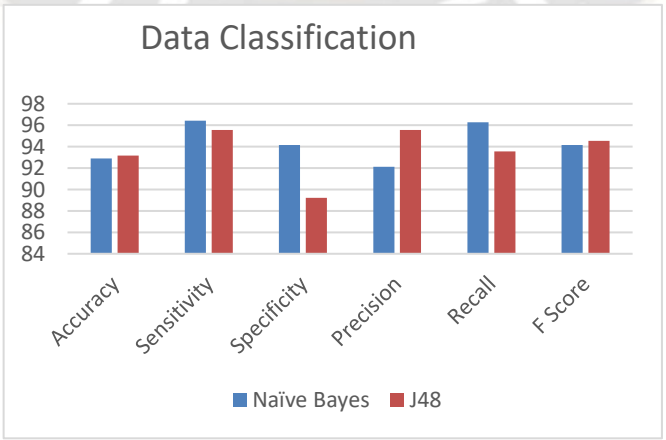

Figure 6: Data Classification using Naïve Bayes and J48

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

B. Prediction using Artificial Neural Network (ANN)

A prediction in context to analyse the UTI-affected people that has been made using artificial neural network based on certain parameters like bacteria, nitrite, epi, clarity, ketones, leukocytes, and ph. Firstly, 2000 datasets have been involved: secondly, 4000 datasets and then 6000 datasets, and in the final prediction, 9578 datasets are considered. From the datasets, prediction is made and find out the errors for the predicted diseases. The errors are obtained including mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE).

i. In the case of 2000 datasets, the values of MAE, MSE, and RMSE are 0.371355, 0.191397 and 0.437490, respectively.

ii. In the case of 4000 datasets, the values of MAE, MSE, and RMSE were 0.222511, 0.115725 and 0.340184, respectively.

iii. In the case of 6000 datasets, the values of MAE, MSE, and RMSE were 0.141594, 0.069941 and 0.264464 respectively.

iv. In the case in which 9578 datasets have been taken, the values of MAE, MSE, and RMSE were 0.100519, 0.044916 and 0.211935, respectively.

All the errors for different datasets (2000, 4000, 6000 & 9578) are demonstrated in Table 3.

TABLE 3: OUTPUTS FOR OVERALL DATASET

NO. OF ENTRIES TAKEN	MAE	MSE	RMSE
2000	0.371355	0.191397	0.437490
4000	0.222511	0.115725	0.340184
6000	0.141594	0.069941	0.264464
9578	0.100519	0.044916	0.211935

A graphical representation of the results obtained for different datasets (2000, 4000, 6000 & 9578) is demonstrated in Figure 7.

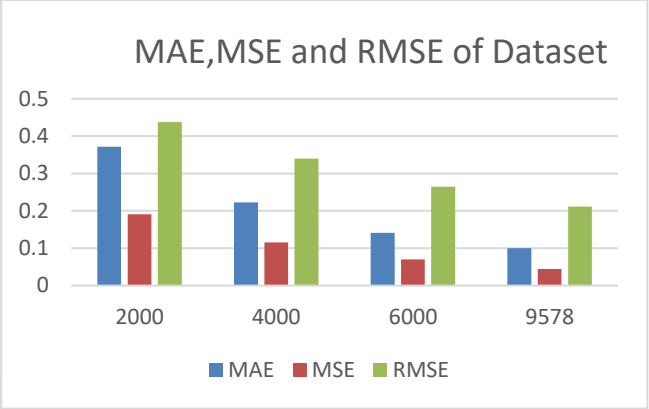

Figure 7: MAE, MSE and RMSE

The confusion matrix is also generated and with the help of this the values of several parameters i.e., accuracy, sensitivity, specificity, precision, recall, and f score are calculated. Confusion matrix of the whole dataset using ANN is shown in Figure 8.

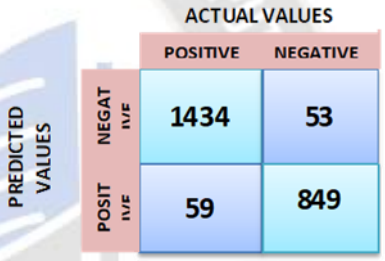

Figure 8: Confusion Matrix of the Tested Dataset Using ANN

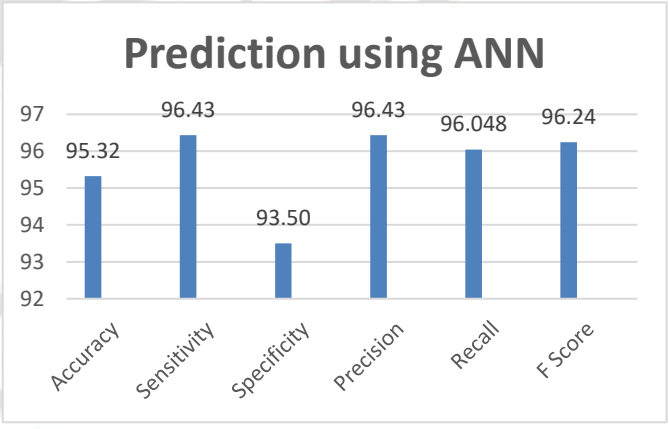

Figure 9: Values of All Obtained Parameters from Dataset using ANN

From the value of the parameter of dataset using ANN, prediction is made with the overall accuracy of 95% approximately and the errors are obtained for the predicted diseases as shown in Figure 9. The graph shown in figure 10 illustrate that the accuracy of the model at the phase of training is 94.5% approximately. The accuracy of the model at the phase of testing is 95.5% approximately.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

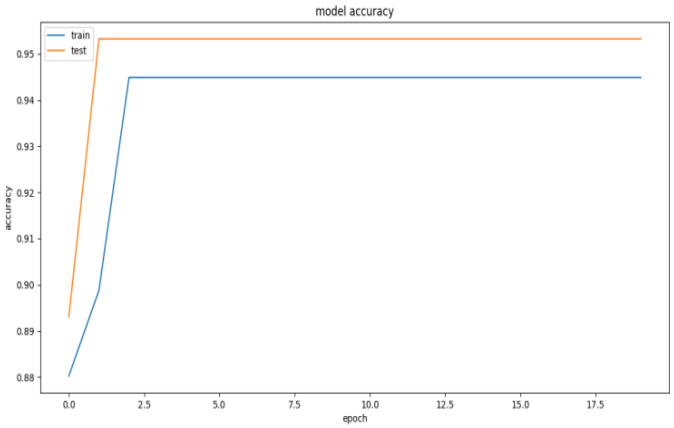

Figure 10: Model Accuracy

Figure 11 illustrates the graph of model error. The graph illustrates that the error of the model at the phase of training is 11.5% approximately. The error of the model at the phase of testing is 6.5% approximately.

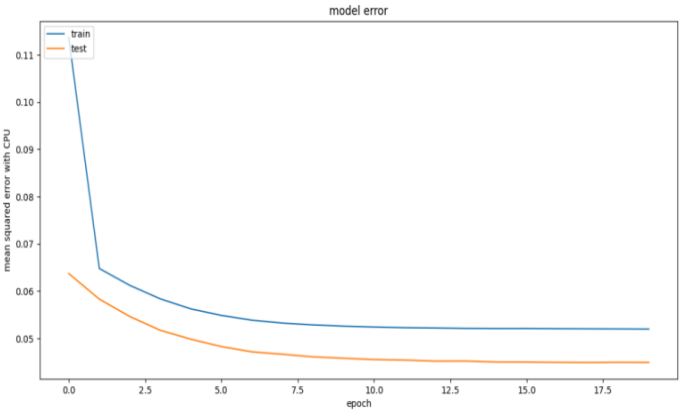

Figure 11: Model Error

V. Conclusion

Artificial neural networks and machine learning are two of the extremely recognized techniques. Both techniques have benefits and have been proven to perform very well in numerous fields related to medicine. Artificial neural network has been recommended as a new model for the prediction of UTI. A machine learning technique for predicting the urinary infection has been presented in this work, evaluation of the outcomes of designed techniques over real-time data, and validation of the same using the comparison study via learning/training and validation data has also been done. In this paper some specific parameters out of the whole dataset have been selected with the help of analysis of variance. UTI prediction done with the help of these techniques has been completed with accuracy of 95.5 % approximately. Other than that, the study and evaluation of UTI for causes and effects has been done, and the assessment of machine learning process requirement for the forecast of urinary disease has also been completed.

VI. Future Work

The future work related to predict urinary tract infection is incorporating internet of things with machine learning and forming a model which can do the prediction in real-time using the sensor technology. The combination of machine learning and IoT paves the way to consider work in the direction of prediction efficiency, energy efficiency and network durability [41, 42]. Also, the work can be further be more specific towards UTI prediction in pregnant women so that an early detection and prediction can help in reducing life threatening risks [43].

References

[1] 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.

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

[3] 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.

[4] Czajkowski, K., Broś-Konopielko, M., & Teliga-Czajkowska, J. (2021). Urinary tract infection in women. Przegląd Menopauzalny= Menopause Review, 20(1), 40.

[5] Flores-Mireles, A. L., Walker, J. N., Caparon, M., & Hultgren, S. J. (2015). Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nature reviews microbiology, 13(5), 269-284.

[6] Shahid, N., Rappon, T., & Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. PloS one, 14(2), e0212356.

[7] Vasudevan, R. (2014). Urinary tract infection: an overview of the infection and the associated risk factors. J Microbiol Exp, 1(2), 00008.

[8] Bhatia, M., & Sood, S. K. (2016). Temporal informative analysis in smart-ICU monitoring: M-HealthCare perspective. Journal of medical systems, 40(8), 1-15.

[9] Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P. M., Sundarasekar, R., & Thota, C. (2018). A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system. Future Generation Computer Systems, 82, 375-387.

[10] Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis. Journal of applied biomedicine, 11(2), 47-58.

[11] Šter, B., & Dobnikar, A. (1996, April). Neural networks in medical diagnosis: Comparison with other methods. In International conference on engineering applications of neural networks (pp. 427-30).

[12] Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC medical informatics and decision making, 19(1), 1-16.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

[13] Lo, Y. C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. Drug discovery today, 23(8), 1538-1546.

[14] Napolitano, G., Marshall, A., Hamilton, P., & Gavin, A. T. (2016). Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction. Artificial intelligence in medicine, 70, 77-83.

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] Patil, D. N. N. . (2021). Liver Tissue Based Disease Detection Using Pre-Processing and Feature Extraction Techniques. Research Journal of Computer Systems and Engineering, 2(2), 17:21. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/27

[22] Altunay, S., Telatar, Z., Erogul, O., & Aydur, E. (2009). A new approach to urinary system dynamics problems: evaluation and classification of uroflowmeter signals using artificial neural networks. Expert Systems with Applications, 36(3), 4891-4895.

[23] Taylor, R. A., Moore, C. L., Cheung, K. H., & Brandt, C. (2018). Predicting urinary tract infections in the emergency department with machine learning. PloS one, 13(3), e0194085.

[24] Almansour, N. A., Syed, H. F., Khayat, N. R., Altheeb, R. K., Juri, R. E., Alhiyafi, J., ... & Olatunji, S. O. (2019). Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. Computers in biology and medicine, 109, 101-111.

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

[26] Nyman, J. (2020). Machine learning approaches for detection of urinary tract infections.

[27] Björck, L., Christensson, B., Herwald, H., Linder, A., & Åkesson, P. (2014). U.S. Patent Application No. 13/983,224.

[28] Zimmerle, C. T., & Pugia, M. J. (2015). U.S. Patent No. 9,145,576. Washington, DC: U.S. Patent and Trademark Office.

[29] Robinson, H. L., Barrett, H. L., Foxcroft, K., Callaway, L. K., & Dekker Nitert, M. (2018). Prevalence of maternal urinary ketones in pregnancy in overweight and obese women. Obstetric Medicine, 11(2), 79-82.

[30] Khan, L. B., Read, H. M., Ritchie, S. R., & Proft, T. (2017). Artificial urine for teaching urinalysis concepts and diagnosis of urinary tract infection in the medical microbiology laboratory. Journal of microbiology & biology education, 18(2), 18-2.

[31] Penders, J., Fiers, T., & Delanghe, J. R. (2002). Quantitative evaluation of urinalysis test strips. Clinical chemistry, 48(12), 2236-2241.

[32] Arif, N., & Qadir, M. I. (2019). Relation between the presence or absence of chin dimple with urobilinogen in urine. International Journal of Advanced Research in Pharmacy and Education, 1(1), 32-34.

[33] Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation computer systems, 25(6), 599-616.

[34] Bhatia, M., Kaur, S., & Sood, S. K. (2020). IoT-inspired smart home based urine infection prediction. Journal of Ambient Intelligence and Humanized Computing, 1-15.

[35] Bhatia, M., Kaur, S., Sood, S. K., & Behal, V. (2020). Internet of things-inspired healthcare system for urine-based diabetes prediction. Artificial Intelligence in Medicine, 107, 101913.

[36] Mr. Dharmesh Dhabliya, Ms. Ritika Dhabalia. (2014). Object Detection and Sorting using IoT. International Journal of New Practices in Management and Engineering, 3(04), 01 - 04. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/31

[37] Gupta, A., & Singh, A. (2022). Early Urine Infection Prediction Framework using XGBoost Ensemble Model in IoT-Fog Environment.

[38] Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, 403.

[39] Taylor, D., Roberts, R., Rodriguez, A., González, M., & Pérez, L. Efficient Course Scheduling in Engineering Education using Machine Learning. Kuwait Journal of Machine Learning, 1(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/121

[40] Nahar, N., & Ara, F. (2018). Liver disease prediction by using different decision tree techniques. International Journal of Data Mining & Knowledge Management Process, 8(2), 01-09.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 7s
DOI: https://doi.org/10.17762/ijritcc.v11i7s.6989
***Article Received**: 29 March 2023 **Revised**: 14 May 2023 **Accepted**: 30 May 2023*

[41] Venkatesan, E., & Velmurugan, T. (2015). Performance analysis of decision tree algorithms for breast cancer classification. Indian Journal of Science and Technology, 8(29), 1-8.

[42] Othman, M. F. B., & Yau, T. M. S. (2007). Comparison of different classification techniques using WEKA for breast cancer. In 3rd Kuala Lumpur international conference on biomedical engineering 2006 (pp. 520-523). Springer, Berlin, Heidelberg.

[43] Mahato, M. K. ., Seth, S. ., & Yadav, P. . (2023). Numerical Simulation and Design of Improved Optimized Green Advertising Framework for Sustainability through Eco-Centric Computation. International Journal of Intelligent Systems and Applications in Engineering, 11(2s), 11–17. Retrieved from https://ijisae.org

[44] Maria Navin, J. R., & Pankaja, R. (2016). Performance analysis of text classification algorithms using confusion matrix. International Journal of Engineering and Technical Research (IJETR), 6(4), 75-8.

[45] Bae, J. H., & Lee, H. K. (2018). User health information analysis with a urine and feces separable smart toilet system. Ieee Access, 6, 78751-78765.

[46] Leila Abadi, Amira Khalid, Predictive Maintenance in Renewable Energy Systems using Machine Learning , Machine Learning Applications Conference Proceedings, Vol 3 2023.

[47] Tasoglu, S. (2022). Toilet-based continuous health monitoring using urine. Nature Reviews Urology, 1-12.

[48] Nkwelle, C. E., Akoachere, J. F. T. K., Ndip, L. M., Nzang, F. A., Esemu, S. F., & Ndip, R. N. (2022). Asymptomatic Urinary Tract infection in Pregnant and Non-pregnant Women in the Limbe Health District of Cameroon: A Phenotypic and Biochemical analytic study.

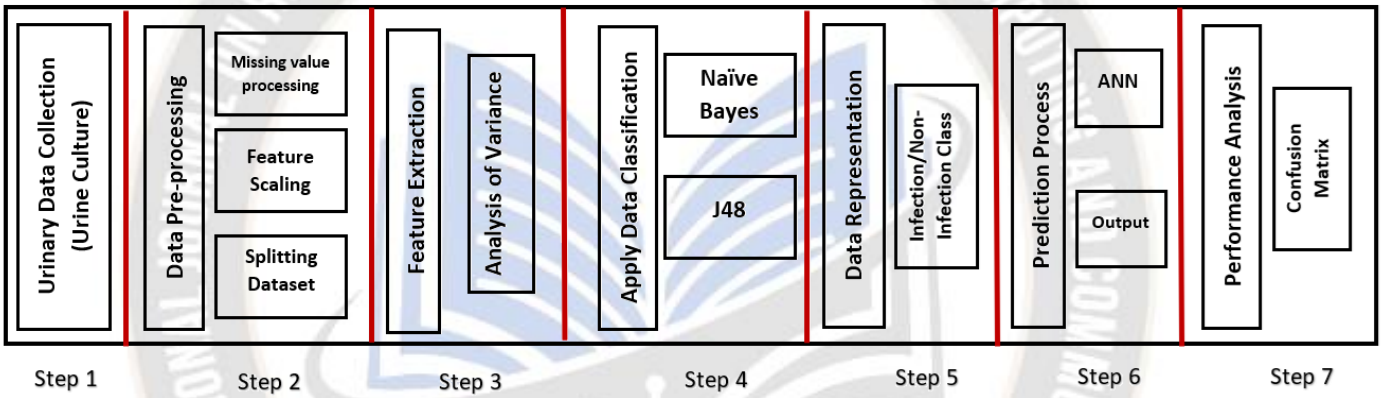

Figure 3: Research Methodology