Fuzzy Logic Based Formalisms for Urology Diseases Diagnosis Using Cosine Amplitude Method & Gower Coefficient

Sunil Kumar Singh[1], Dr. Megha Mishra[2], Dr. Mukesh Kumar Chandrakar[3], Dr. Mithilesh Deo Pandey[4]

[1]Assistant Professor, Department of Computer Science and Engineering CMR Engineering College Hyderabad, India, Email: sunil.sinme@gmail.com
[2]Associate Professor, Department of Computer Science and Engineering, SSTC, Bhilai CG, India, Email: megha16shukla@gmail.com
[3]Assistant Professor, Department of Electronics and Telecommunications, Bhilai Institute of Technology, Durg, CG, India, Email: mukesh.chandrakar@bitdurg.ac.in
[4]Associate Professor, Department of Applied Mathematics, Bhilai Institute of Technology, Durg,CG, India, Email: md.pandey@bitdurg.ac.in

Received: 14.07.2024 Revised: 13.08.2024 Accepted: 09.09.2024

ABSTRACT
This research investigates the diagnostic use of the Gower Coefficient and the Cosine Amplitude Method (CAM) for urological disorders, with a particular emphasis on prostate cancer. The effectiveness of these techniques was evaluated using a large dataset that was obtained from Kaggle and included clinical and diagnostic information for more than 100 patients. The dataset contains a number of attributes, including diagnostic outcomes, perimeter, texture, and radius. Preparing the dataset for analysis via cleaning, normalization, encoding, and feature selection was known as data preparation. Patient feature vector similarity was calculated using the CAM, and patient similarity was handled and computed using the Gower Coefficient for a variety of data formats. The results of the model performance assessment showed that the Cosine Similarity model performed better than the others, showing the best classification capabilities and the fewest misclassifications with an accuracy of 0.98, precision of 0.9531, recall of 0.98, and an F1-Score of 0.9683. The Naive Bayes model trailed after with somewhat worse metrics, but it was still rather effective. As an example, the models for Support Vector Machine (SVM) and Kappa Coefficient performed somewhat worse, with the Kappa Coefficient model obtaining the lowest metrics. The results highlight how effective the Cosine Amplitude Method is in making precise diagnosis with little misclassification of urological illnesses. The research sheds light on the use of CAM and the Gower Coefficient in the diagnosis of urological diseases and emphasizes the need of further investigation into these techniques in clinical settings.

Keywords: Cosine Similarity, Urology Diseases Diagnosis, Gower Coefficient, Prostate Cancer

INTRODUCTION

Global studies indicate that by 2025, more than 36 million individuals will be at high risk of developing chronic kidney disease. One healthy individual out of ten has a chronic renal condition [1]. When the kidneys suffer long-term damage and the body is unable to maintain proper fluid, electrolyte, and metabolic balances, chronic renal failure results. When the kidneys sustain long-term damage and the body is unable to maintain proper fluid, electrolyte, and metabolic balances, chronic renal failure results [2]. Five factors have caused chronic kidney disease to become more well-known worldwide: the illness's increasing incidence, its hidden pandemic, its high cost, its substantial influence on cardiovascular disease, and the need for efficient therapies [3] [4]. Regrettably, chronic renal insufficiency is usually without symptoms and the true number of patients affected is unknown. Approximately 1.5 million individuals globally are undergoing kidney transplantation or hemodialysis. This figure might potentially triple within a few years if both the government and the public disregard it. Iran has the highest incidence of renal illness worldwide, far surpassing even its neighbouring nations [5]. Global incidence of kidney disease is expected to increase by 15–20%. Many are unaware of the staggering statistic that indicates that more than 7 million Americans are afflicted with renal failure. Patients with severe renal failure undergoing therapy for end-stage renal disease (ESRD) would increase fourfold to 40,000 in the next five

years. Among them, 2% will have paclitaxel dialysis, 50% will get haemodialysis, and 48% will undergo kidney transplantation [6].

Diabetes and hypertension are the primary etiologies of chronic kidney disease (CKD), but the accurate incidence of additional causes, such as urological disorders, in adults remains uncertain [7]. These disorders may be either congenital or acquired. The most common conditions are vesicoureteral reflux (VUR), which can lead to reflux nephropathy, recurrent urinary tract infections (UTIs) causing pyelonephritis, urinary tract obstruction caused by anatomical and functional abnormalities such as ureteropelvic junction syndrome, bladder neck stricture, congenital urethral valves, urethral stenosis, nephrolithiasis, malignancies, and benign prostatic hyperplasia (BPH), as well as an overactive bladder, particularly in adult females [8]. These pathologies may have a gradual and subtle beginning, making them challenging to recognize and characterize apart from their widely recognized consequences. Hypertension, proteinuria, urine concentration abnormalities, hyperkalemia, metabolic acidosis, focal and segmental glomerulosclerosis, and chronic kidney disease (CKD) are the predominant consequences that greatly affect the long-term renal and cardiovascular prognosis [9][10]. The objective of this research is to assess the frequency of urological disorders in the nephrology department and their potential correlation with renal function, diabetes, hypertension, recurring infections, and proteinuria. Withhold enrollment from patients who declined to provide permission or patients with incomplete data. In addition to undergoing uroflowmetry, the patients completed two questionnaires: the International Prostate Symptom Score (I-PSS) [11][12], and the Incontinence Questionnaire-Short Form (ICIQ-SF) [13]. The estimated glomerular filtration rate (eGFR) was determined using the abbreviated Modification of Diet in Renal Disease method, and shown in mL/min/1.73m2 [14].

Complications that are associated with the prostate are more common in men who are 30 years old or older. The majority of men who are diagnosed with prostate diseases are in their middle years or older. PCa is the cancerous proliferation of the cells, and there is a possibility of infiltrating to the surrounding structures. The most common disease associated with the gland is benign prostatic hyperplasia (BPH), which is the benign enlargement of the prostate tissue without any associated malignant cell proliferation and angiogenesis. BPH is the most common disease associated with the gland. calcifications are the presence of smaller stones in the gland, and they can be caused secondary to ongoing pathologies such as benign prostatic hyperplasia (BPH) and prostate cancer (PCa). Prostatitis, both acute and chronic, is the inflammatory reaction of the prostate in response to bacterial or viral invasion. Antibiotics can be prescribed to alleviate the problem [15].

It is estimated that prostate cancer is one of the most common forms of malignant tumors seen in men across the globe [16]. Compared to other regions of Asia, the number of men diagnosed with prostate cancer in China is much greater. This disease is mostly diagnosed in men who are middle-aged or older, and it is seeing a progressive increase from year to year [17]. The treatment entails the manual implantation of (80–120) I235 radioactive seeds in the lesion region using twenty tiny needles. This process is carried out by a medical professional. Nevertheless, the therapeutic outcomes of this method are affected by the uncertainties that are inherent in manual operation, which include problems with positioning precision, inappropriate implantation of the radioactive seeds, and substantial soft tissue damage [18].

According to the current clinical paradigm, multicore random biopsies are necessary for risk stratification in men who have increased prostate-specific antigen (PSA) and a positive digital rectal exam (DRE). The usefulness of prostate PSA as a screening test for prostate cancer is still up for debate. It has been shown by two recent large randomized clinical studies [19] that there is a considerable risk of overdiagnosis for prostate cancer with PSA screening, with an estimated 50% of screened men receiving a prostate cancer diagnosis. This results in uncomfortable needle biopsies and maybe excessive therapy thereafter [20]. Furthermore, it is becoming more and more evident that performing prostate biopsies increases the likelihood of hospital admission for infectious problems, often leading to pain and potential sexual dysfunction as well as the possibility that the needle would miss malignant cells [21]. However, long-term follow-ups have shown that PSA testing lowers prostate cancer mortality by 20–30% [22]. Consequently, PSA testing continues to be a valuable indicator for the diagnosis of clinically relevant prostate malignancies. How to prevent the overdiagnosis of clinically unimportant tumors while improving prostate cancer diagnosis remains a concern.

LITERATURE REVIEW [23]

Continuous kidney disease, often known as CKD, is a disorder that is quite common. CKD is recognized to be caused by urologic illnesses; however, these conditions often go undetected and are underestimated due to their gradual development and subtle beginning. We wanted to use uroflowmetry to determine the extent to which individuals with chronic kidney disease had urological conditions that are not well known

about.
[24] chronic kidney disease, often known as CKD, is a condition that is not easily recognized. During the early stages of chronic kidney disease (CKD), the majority of individuals are often uninformed of their illness. This presents a difficulty for medical personnel who are attempting to initiate therapy or begin preventive efforts. One of the challenges associated with diagnosing chronic kidney disease (CKD) is that in the majority of regions throughout the globe, the diagnosis is still made based on measurements of serum creatinine and the concomitant estimates of eGFR. There are disagreements over the present staging system, particularly with regard to the approach that is used to diagnose and forecast chronic kidney disease (CKD). The purpose of this review is to investigate research that focused on the many kinds of samples that have the potential to act as a good and promising biomarker for the early diagnosis of chronic kidney disease (CKD) or for the detection of rapidly diminishing renal function among patients with CKD.
[25] Diagnosing prostate cancer (PCa) using multi-parametric magnetic resonance imaging (MRI) calls for highly skilled radiologists. Imaging artifacts, elastic soft-tissue deformations, and patient movement may all result in misalignments between the MRI sequences. They make the work even more difficult, which forces radiologists to interpret the pictures. Lately, computer-aided diagnostic (CAD) instruments have shown promise in diagnosing prostate cancer (PCa), usually requiring intricate co-registration of the input modalities. Research groups, however, cannot agree on whether CAD systems benefit from registration. Furthermore, no other approaches to dealing with multi-modal misalignments have been investigated up to this point. In this work, several approaches to managing picture misalignments are presented, compared, and their direct impact on PCa diagnosis accuracy is assessed. To improve CAD resilience, we suggest "misalignment augmentation" as a notion in addition to well-known registration procedures. The findings show that misalignment augmentations may enhance performance on a separate test set in addition to making up for a total absence of registration when used in combination with registration.
[26] Recently, there has been a lot of interest in compressive sensing (CS) in various real-world applications. Its benefit stems from its capacity to make use of the data's sparse representation. A sensing matrix with certain attributes will be used to code the signal in order to identify a sparse signal. Although a lot of sensing matrices have been developed in this sector, no one matrix works well for all applications. It is dependent upon the coding duration, data dimensions, application domain, and, of course, recovery accuracy. This paper compares and applies several sensing matrices for the categorization of prostate cancer. Because the characteristics in the prostate cancer data are so many, CS technique would be ideal for these kinds of applications. The findings demonstrate that optimizing the margin, which regulates the classifier's performance, is critically dependent on the sensing matrix.
[27] The most prevalent kind of cancer and the second largest contributor to cancer-related deaths in North America is prostate cancer. Early prostate cancer identification is critical to patient survival rates, and auto-detection of prostate cancer may be a big part of this process. Even though multi-parametric magnetic resonance imaging (MP-MRI) has shown promise in the diagnosis of prostate cancer, the wealth of data available in MP-MRI is not used by the auto-detection algorithms currently in use to increase detection accuracy. The purpose of this study was to use MP-MRI data to develop a radiomics-based auto-detection system for prostate cancer.

BACKGROUND

Gower Coefficient

The Gower coefficient, first proposed by Gower in 1971, was first offered as a measure of similarity. For HCA purposes, however, it is often represented as a dissimilarity measure. Let X be the data matrix defined as $\left[X_{ic}\right]$, where i and)n(is the total number of objects c is the . A formula is used to represent the dissimilarity of the objects)v(total number of variables $X_i = \left[X_{i1}, X_{i2}, \ldots, X_{iv}\right]$ and $X_j = \left[X_{j1}, X_{j2}, \ldots, X_{jv}\right]$, which are defined by values of mixed-type variables.

$$d_g\left(X_i, X_j\right) = \sum_{c=1}^{v} d_{ijc} \qquad (1)$$

where d_{ijc} is a metric of dissimilarity between the i- th and j-th items, measured by the c- th variable. The formula necessitates a dataset that includes omitted observations that have missing values.
A nominal or alternative variable at the c- th position is considered to have a dissimilarity of zero for

matches of categories and one otherwise between two categories x_{ic} and x_{jc}
For a quantitative c– th variable, dissimilarity is represented by a specific formula.

$$d_{ijc} = \frac{|x_{ic} - x_{jc}|}{\max(x_c) - \min(x_c)}. \qquad (2)$$

Provided that the c- th variable is ordinal, all categories undergo transformation according to the given formula.

$$X_{ic} = \frac{r_{ic} - 1}{R_c - 1}, \qquad (3)$$

where r_{ic} represents the rank number of the ordinal category with i order $(r = 1, \ldots, R_c)$ and R_c represents the highest rank number of the variable with c orders. Following this conversion, the resultant values may be used in equation (2) for numerical variables.

Cosine Similarity

We initially describe the CSE [28] process in technique before proposing the Multivariate Multiscale Cosine Similarity Entropy (MMCSE) technique. The CSE estimate that uses angular distance instead of amplitude distance shows less susceptibility to outliers than MSE, such as those originating from abrupt changes in amplitude. In fact, regardless of the variance of the time series, the upper limit of angular distance of 2π uses CSE to provide a well-defined measurement with a range between 0 and 1, improving stability while working with very dynamic signals. Additionally, we demonstrate that in situations when there are significant variance disparities across data sets from several channels a situation in which MMSE is insufficient the consistent estimate offered by CSE is crucial for multivariate analysis.
Coarse Graining Process (CGP) is used in the first phase for each data channel via a non- overlapping window with a scale factor, τ, in order to extend CSE to multi-variate multi-scale entropy [29]. The signal was then scaled and rebuilt as a Composite Delay Vector (CDV). Lastly, Algorithm 1 applies CDV after Steps 2–6 to calculate the Multivariate Multiscale Cosine Similarity Entropy (MMCSE) rather than using an embedding vector as in MSampEn. Because MMCSE depends on origin-coordination in order to project similarity within tolerance angle, take note that it is sensitive to DC offset. As a result, the DC offset or a long-term trend must be eliminated by filtering the MMCSE or by removing their median.

Algorithm 1. Cosine similarity Entropy (CSE)

Let $\{x(i)\}_{i=1}^{n}$ be a univariate data collection of length n. The parameters are the embedding dimension (m), tolerance (r), and time delay (l).

- After subtracting the offset, the embedding vectors, $x_m(i)$, are formed. These are obtained from the zero-median signal $\{u(i)\}_{i=1}^{n}$ by $u(i) = x(i) - median\left(\{x(i)\}_{i=1}^{n}\right)$, where $x_m(i) = \left(u(i), u(i+l), \ldots, u(i+(m-1)l)\right)$.
- Determine the angular separation between the paired embedding vectors $x_m(i)$ and $x_m(j)$ using Cosine Similarity. This may be expressed as follows:
$$d_m(i,j) = \frac{1}{\pi}\cos^{-1}\left(\frac{x_m(i)\cdot x_m(j)}{|x_m(i)||x_m(j)|}\right), i \neq j.$$
- Determine how many comparable patterns—defined as similar pairs $b_m^R(i)$ meet the requirement that $D_m(i,j) \leq R$.
- Compute the local probability of $b_m^R(i)$ using the formula $c_m^r(i) = \frac{b_m^R(i)}{N-n-1}$, where $n = (m-1)l$.
- Calculate the global probability using the formula $\varphi_m^R = \frac{\sum_{i=1}^{N-n} c_m^R(i)}{N-n}$.
- The formula for calculating Cosine Similarity Entropy (CSE) $(m,l,R,n) = -\left[\varphi_m^R \log_2 \varphi_m^R + \left(1-\varphi_m^R\right)\log_2\left(1-\varphi_m^R\right)\right]$.

RESEARCH GAP

Prior studies have identified a number of difficulties in the diagnosis of urological disorders connected to prostate cancer and chronic kidney disease (CKD). Due to limited sensitivity of conventional diagnostic methods such as uroflowmetry, CKD patients may not be aware of some urological diseases. Because early diagnosis of CKD is still difficult because it depends on serum creatinine and eGFR, and because of the shortcomings of the existing staging system. Due to image distortions and misalignments, multi-parametric MRI presents challenges for the detection of prostate cancer and makes the use of computer-aided diagnostic (CAD) systems more challenging. Furthermore, different application demands make it difficult to optimize sensing matrices in compressive sensing for prostate cancer. By improving the detection of subtle patterns in uroflowmetry data, addressing MRI misalignments in prostate cancer, improving early CKD diagnosis through a wider range of biomarkers, and optimizing feature selection in compressive sensing, the Cosine Similarity model offers novel solutions. various developments may result in diagnostic techniques for various urological disorders that are more precise, sensitive, and effective.

METHODOLOGY

Data Collection

Prostate cancer datasets from Kaggle are used in the data collection process for the research on urological illnesses diagnosis using the Cosine Amplitude Method and Gower Coefficient. Patients' demographics, test results, and imaging data are all included in this extensive dataset, which offers clinical and diagnostic data related to prostate cancer. Applying the Gower Coefficient to assess the separation between various diagnostic features and the Cosine Amplitude Method to analyze similarity measures both benefit greatly from the dataset. Researchers may evaluate and use suggested diagnostic approaches with greater efficacy by using this dataset, which guarantees a reliable and accurate assessment of urological disorders.

Dataset

The prostate cancer dataset utilized in this research was sourced from Prostate Cancer (kaggle.com). The medical records of more than one hundred different people, each of whom has ten different factors, are included in this collection. The following is a list of the variables:

- Id
- Radius
- Texture
- Perimeter
- Area
- Smoothness
- Compactness
- Diagnosis result
- Symmetry
- Fractal dimension

Data preprocessing

Preprocessing the prostate cancer dataset for Cosine Amplitude Method and Gower Coefficient urology diagnosis requires multiple procedures. First, data cleaning deals with missing values and discrepancies using imputation or data correction to create a reliable dataset. Normalization and standardization are necessary to standardize feature magnitudes for the Cosine Amplitude Method. Use one-hot or label encoding to convert categorical variables to numbers. Using dimensionality reduction, feature selection and extraction determine the most important characteristics for analysis. To evaluate the model, the dataset should be divided into training and testing sets, usually 70% for training and 30% for testing. Due to imbalance, oversampling or undersampling should be used to balance class distribution. These procedures optimise the dataset for prostate cancer and other urological illness diagnosis using the Cosine Amplitude Method and Gower Coefficient.

Cosine Amplitude Method (CAM)

Based on the feature vectors of two data points (patients, for example), the Cosine Amplitude Method is used to determine how similar the two are.

Representing Feature Vectors: Use the feature vector $X_i = \left[x_{i1}, x_{i2}, \ldots, x_{in} \right]$ to represent each patient i.

Where $X_i = [x_{i1}, x_{i2}, \ldots, x_{in}]$ is the value of the j-th characteristic for patient x_{ij}.

$$\textbf{Equation: } (X_i, X_j) = \frac{X_i \cdot X_j}{\| X_i \| \| X_j \|}$$

where

The dot product of the feature vectors is $X_i \cdot X_j$.

The vectors' magnitudes (Euclidean norms) are $\| X_i \| and \| X_j \|$

Gower Coefficient

The Gower Coefficient is a measure that computes patient similarity and may handle diverse data types, such as continuous and categorical data.

$$\textbf{Equation: } (X_i, X_j) = \frac{1}{n} \sum_{k=1}^{n} s_{ijk}$$

where:

The number of features is n .

The partial similarity between patients i and j for characteristic k is denoted by s_{ijk}

Performance Evaluation

Accuracy: How frequently the classifier generates accurate predictions is easiest to assess using accuracy. Another interpretation is the fraction of accurate predictions to all estimates.

$$Accuracy = \frac{TP + TN}{S}$$

Precision: Compared to this ratio, which is derived by subtracting one for it, i.e., (1 – precise), which represents the fraction of false negatives, recall is obtained by dividing precision by one.

$$Precision = \frac{TP}{TP + FP}$$

Recall: On other hand there are called false negatives in relation with True Negatives.

$$\text{Re } call = \frac{TP}{TP + FN}$$

F1-Score: It is calculated by taking the accuracy and recall scores and squaring them. For this.

$$F1 = \frac{2 * \text{Pr } ecision * \text{Re } call}{\text{Pr } ecision + \text{Re } call}$$

RESULTS & DISCUSSION

Results

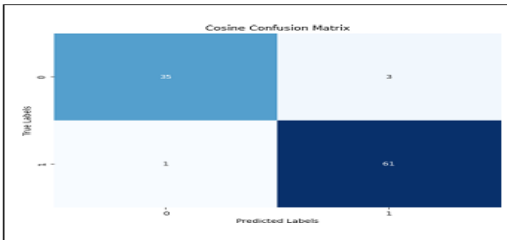

Figure 1 Confusion Matrix of Cosine Similarity Model

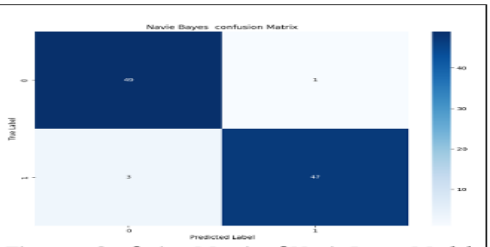

Figure 2 Confusion Matrix of Navie Bayes Model

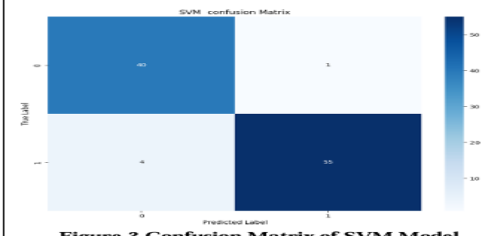

Figure 3 Confusion Matrix of SVM Model

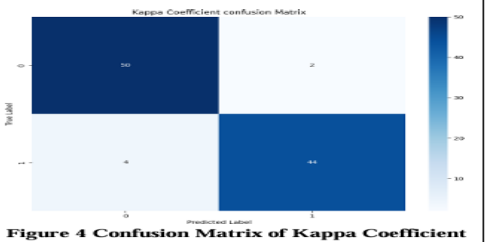

Figure 4 Confusion Matrix of Kappa Coefficient Model

The four confusion matrices show how well the Cosine Similarity, Naive Bayes, SVM, and Kappa Coefficient models perform in categorizing cases into two groups. With just 4 incorrect predictions out of 100, the Cosine Similarity model performed well, correctly categorizing 35 cases as class 0 and 61 examples as class 1. Comparably, the Naive Bayes model did well, accurately classifying 47 cases as class 1 and 49 as class 0, with just 4 misclassifications. With five misclassifications, the SVM model performed comparably, accurately categorizing 49 cases as class 0 and 55 as class 1. The Kappa Coefficient model successfully classified 50 occurrences as class 0 and 44 as class 1, with somewhat more mistakes (6 misclassifications) but still a respectable degree of accuracy. The most accurate model in this comparison is the Cosine Similarity model, which performs the best out of all the models with the fewest misclassifications. While the Kappa Coefficient model offers a strong performance while being somewhat less precise than the Naive Bayes and SVM models, they nevertheless perform extremely well with very near accuracy.

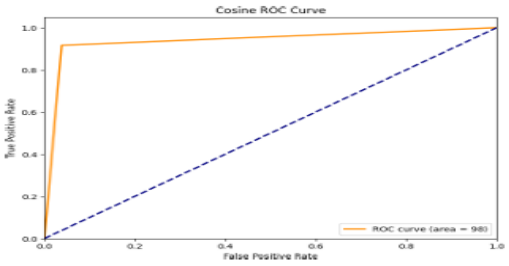

Figure 5 ROC Curve of Cosine Similarity Model

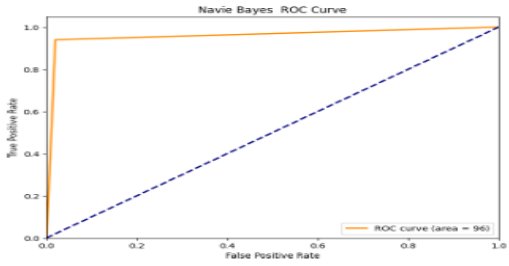

Figure 6 ROC Curve of Navie Bayes Model

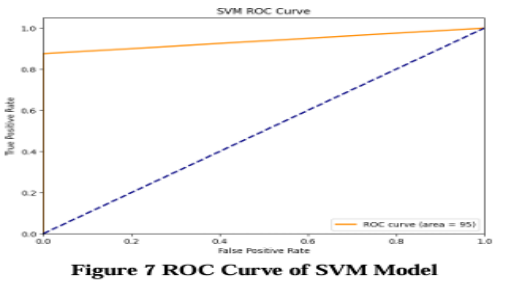

Figure 7 ROC Curve of SVM Model

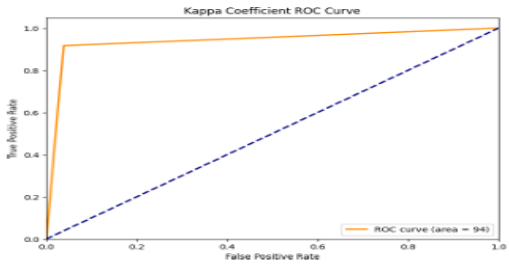

Figure 8 ROC Curve of Kappa Coefficient Model

Four models' categorization performance is thoroughly compared using the ROC curves shown in the figures. With a ROC curve that is closest to the upper left corner and an AUC of 0.98, showing excellent classification performance, the Cosine Similarity model shows the maximum efficacy. The Naive Bayes model performs well but somewhat less effectively than the Cosine Similarity model, following closely behind with a curve that also approaches the upper left corner with an AUC of 0.96. With an AUC of 0.95, the Support Vector Machine (SVM) model performs well, however its ROC curve is not as near the upper left corner as those of the previous two models. Lastly, while it is somewhat less successful than the others, the Kappa Coefficient model has an AUC of 0.94 and a curve that shows a decent balance between sensitivity and specificity. All things considered, the Cosine Similarity model is the most successful, closely followed by the Naive Bayes model. The SVM and Kappa Coefficient models also show strong classification capabilities.

Performance Metrics

Table 1 Evaluation Metrics

Models	Accuracy	Precision	Recall	F1-Score
Cosine similarity	0.98	0.9531	0.98	0.9683
Naive Bayes	0.96	0.9607	0.96	0.9599
SVM	0.95	0.9538	0.95	0.9493
Kappa Coefficient	0.94	0.9406	0.94	0.9399

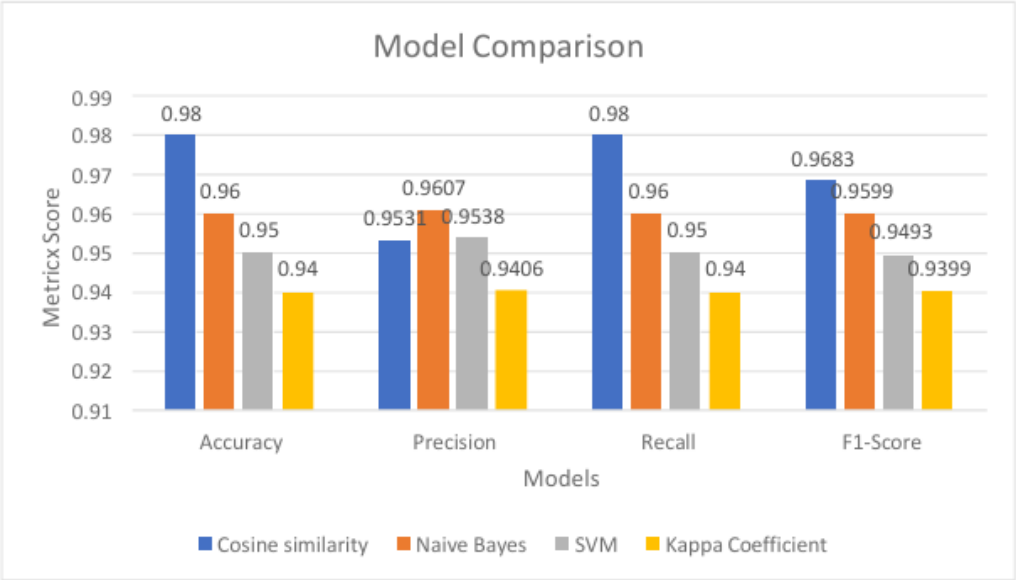

Figure 9. Model Comparison

A comparison of the accuracy, precision, recall, and F1-score of four different models is shown in the table. These models are the Cosine Similarity, Naive Bayes, Support Vector Machine (SVM), and Kappa Coefficient. With an accuracy of 0.98, a precision of 0.9531, a recall of 0.98, and an F1-Score of 0.9683, the Cosine Similarity algorithm emerges as the best performer. This indicates that there is a solid balance between precision and recall characteristics. Following closely after is the Naive Bayes algorithm, which demonstrates a well-rounded performance with an accuracy of 0.96, precision of 0.9607, recall of 0.96, and an F1-Score of 0.9599. An accuracy of 0.95, precision of 0.9538, recall of 0.95, and an F1-score of 0.9493 come from the support vector machine (SVM), which demonstrates a somewhat lesser performance. Finally, the Kappa Coefficient model has the worst performance out of the four. It has an accuracy of 0.94, a precision of 0.9406, a recall of 0.94, and an F1-Score of 0.9399, which indicates that it has a performance that is less balanced in contrast to the other models. While the Kappa Coefficient is the least effective model among all of the metrics that were tested, the Cosine Similarity model is the one that stands out as the most successful overall.

DISCUSSION

We used confusion matrices, ROC curves, and performance metrics to evaluate Cosine Similarity, Naive Bayes, SVM, and Kappa Coefficient. The Cosine Similarity model excelled in all evaluations. With 4 misclassifications out of 100, our model correctly identified 35 samples as class 0 and 61 as class 1. This model's ROC curve is closest to the top left corner and has a great AUC of 0.98, indicating strong classification Naive Bayes performed well with the same misclassifications as Cosine Similarity. Correctly detected 47 class 1 and 49 class 0 occurrences. Although less than the Cosine Similarity model, its ROC curve performs well with an AUC of 0.96. The successful SVM model has 5 misclassifications and 0.95 AUC. Despite being great, its ROC curve is not as close to the top left corner as Cosine Similarity and Naive Bayes. Kappa Coefficient had the greatest misclassifications (6/100), 50 in class 0 and 44 in class 1. Its ROC curve has high sensitivity and specificity with an AUC of 0.94. Though well- performing, it was the least effective model. The Cosine Similarity model provides the best precision-recall balance (0.98), accuracy (0.98), precision (0.9531), and recall (0.98). With 0.96 accuracy, 0.9607 precision, 0.96 recall, and 0.9599 F1-score, the Naive Bayes model followed closely. Kappa Coefficient had 0.94 accuracy, 0.9406 precision, 0.94 recall, and 0.9399 F1- score, whereas SVM had 0.95 accuracy, 0.9538 precision, 0.95 recall, and 0.9493 F1-score. The Cosine Similarity model is the most reliable and successful in accuracy, precision, recall, F1-score, and ROC curve performance.

CONCLUSION

In Conclusion it has been determined that the Cosine Similarity model is the most effective overall, based on the evaluation metrics and performance analysis of the Naive Bayes model, the Support Vector Machine (SVM), and the Kappa Coefficient model. The greatest performance was displayed by it, consisting of an accuracy of 0.98, a precision of 0.9531, a recall of 0.98, and an F1-Score of 0.9683. The area under the receiver operating characteristic curve (ROC) of this model is closest to the top left corner,

suggesting that it has high classification capabilities and the fewest misclassifications among the models (four out of one hundred). An accuracy of 0.96, precision of 0.9607, recall of 0.96, and an F1-Score of 0.9599 were all achieved by the Naive Bayes model, despite the fact that it was somewhat less successful than the Cosine Similarity model. The fact that its ROC curve has an area under the curve (AUC) of 0.96 is additional evidence of its excellent classification ability. Although it had slightly more misclassifications (5 out of 100) and an area under the curve (AUC) of 0.95, the SVM model did well as well. It had an accuracy of 0.95, a precision of 0.9538, a recall of 0.95, and an F1-Score of 0.9493. A total of six misclassifications and an area under the curve (AUC) of 0.94 were associated with the Kappa Coefficient model, which had the lowest performance metrics. It had an accuracy of 0.94, a precision of 0.9406, a recall of 0.94, and an F1-Score of 0.9399. Although the Naive Bayes model also provides great performance, the Cosine Similarity model is suggested because of its higher accuracy and reduced misclassification rate. In general, the Cosine Similarity model is favored. However, despite their effectiveness, the SVM and Kappa Coefficient models are not quite as good as the others.

REFERENCES

[1] D. K. Ryan, D. Banerjee, and F. Jouhra, "Management of heart failure in patients with chronic kidney disease," Eur. Cardiol. Rev., vol. 17, 2022.

[2] N. Nestler, "Nursing care and outcome in surgical patients–why do we have to care?," Innov. Surg. Sci., vol. 4, no. 4, pp. 139–143, 2019.

[3] L. E. Boulware and D. Mohottige, "The seen and the unseen: race and social inequities affecting kidney care," Clin. J. Am. Soc. Nephrol., vol. 16, no. 5, pp. 815–817, 2021.

[4] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, and C. Silva, "Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors," arXiv Prepr. arXiv1310.8002, 2013.

[5] Saber, A. N. Tahami, H. Najafipour, and J. Azmandian, "Assessment of prevalence of chronic kidney disease and its predisposing factors in Kerman city," Nephrourol. Mon., vol. 9, no. 2, 2017.

[6] V. Biniaz, E. Nemati, A. Tayebi, M. S. Shermeh, and A. Ebadi, "The effect of vitamin C on parathyroid hormone in patients on hemodialysis with secondary hyperparathyroidism: a double blind, placebo-controlled study," Nephrourol. Mon., vol. 5, no. 5, p. 962, 2013.

[7] K. Okamura, Y. Nojiri, M. Yamamoto, M. Kobayashi, Y. Okamoto, and T. Yasui, "Questionnaire survey on lower urinary tract symptoms (LUTS) for patients attending general practice clinics," Nihon Ronen Igakkai Zasshi., vol. 43, no. 4, pp. 498–504, 2006.

[8] M. B. Chancellor and D. A. Rivas, "American Urological Association symptom index for women with voiding symptoms: lack of index specificity for benign prostate hyperplasia," J. Urol., vol. 150, no. 5 Part 2, pp. 1706–1708, 1993.

[9] D. Noone and C. Licht, "Chronic kidney disease: a new look at pathogenetic mechanisms and treatment options," Pediatr. Nephrol., vol. 29, pp. 779–792, 2014.

[10] B. A. Warady and V. Chadha, "Chronic kidney disease in children: the global perspective," Pediatr. Nephrol., vol. 22, no. 12, pp. 1999–2009, 2007.

[11] X. Badia, M. Garcia-Losa, and R. Dal-Re, "Ten-language translation and harmonization of the International Prostate Symptom Score: developing a methodology for multinational clinical trials," Eur. Urol., vol. 31, no. 2, pp. 129–140, 1997.

[12] A. Tubaro et al., "Italian validation of the international consultation on incontinence questionnaires," BJU Int., vol. 97, no. 1, pp. 101–108, 2006.

[13] K. Avery, J. Donovan, T. J. Peters, C. Shaw, M. Gotoh, and P. Abrams, "ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence," Neurourol. Urodynamics Off. J. Int. Cont. Soc., vol. 23, no. 4, pp. 322–330, 2004.

[14] L. AS, "Chronic kidney disease epidemiology collaboration. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate," Ann Intern Med, vol. 145, pp. 247–254, 2006.

[15] R. P. Singh, S. Gupta, and U. R. Acharya, "Segmentation of prostate contours for automated diagnosis using ultrasound images: A survey," J. Comput. Sci., vol. 21, pp. 223–231, 2017.

[16] Z. Yongde, L. Yi, W. Xiaofei, and X. Yong, "Design and experimental study of joint torque balance mechanism of seed implantation articulated robot," Adv. Mech. Eng., vol. 7, no. 6, p. 1687814015589479, 2015.

[17] B. Li, Y. Zhang, L. Yuan, and X. Xi, "Study on the low velocity stability of a prostate seed implantation robot's rotatory joint," Electronics, vol. 9, no. 2, p. 284, 2020.

[18] Y. I. LIANG, D. XU, B. WANG, Y. ZHANG, and Y. XU, "Experimental study of needle insertion strategies of seed implantation articulated robot," J. Mech. Med. Biol., vol. 18, no. 03, p. 1850023, 2018.

[19] G. L. Andriole et al., "Mortality results from a randomized prostate-cancer screening trial," N. Engl. J.

Med., vol. 360, no. 13, pp. 1310–1319, 2009.
[20] A. J. Vickers et al., "Empirical estimates of prostate cancer overdiagnosis by age and prostate-specific antigen," BMC Med., vol. 12, pp. 1–7, 2014.
[21] R. K. Nam et al., "Increasing hospital admission rates for urological complications after transrectal ultrasound guided prostate biopsy," J. Urol., vol. 183, no. 3, pp. 963– 969, 2010.
[22] F. H. Schröder et al., "Prostate-cancer mortality at 11 years of follow-up," N. Engl. J. Med., vol. 366, no. 11, pp. 981–990, 2012.
[23] S. Lai et al., "Chronic kidney disease and urological disorders: Systematic use of uroflowmetry in nephropathic patients," Clin. Kidney J., vol. 12, no. 3, pp. 414–419, 2019, doi: 10.1093/ckj/sfy085.
[24] M. Z. Bidin, A. M. Shah, J. Stanslas, and C. T. S. Lim, "Blood and urine biomarkers in chronic kidney disease: An update," Clin. Chim. Acta, vol. 495, no. January, pp. 239– 250, 2019, doi: 10.1016/j.cca.2019.04.069.
[25] B. Kovacs et al., "Addressing image misalignments in multi-parametric prostate MRI for enhanced computer-aided diagnosis of prostate cancer," Sci. Rep., vol. 13, no. 1, pp. 1–12, 2023, doi: 10.1038/s41598-023-46747-z.
[26] K. Awedat, A. Essa, and I. Abdel-Qader, "Compressive Sensing Evaluation Strategy for Prostate Cancer Classification," IEEE Int. Conf. Electro Inf. Technol., vol. 2020- July, no. July 2021, pp. 423–428, 2020, doi: 10.1109/EIT48999.2020.9208262.
[27] F. Khalvati, A. Wong, and M. A. Haider, "Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models," BMC Med. Imaging, vol. 15, no. 1, pp. 1–14, 2015, doi: 10.1186/s12880-015-0069-9.
[28] T. Chanwimalueang and D. P. Mandic, "Cosine similarity entropy: Self-correlation- based complexity analysis of dynamical systems," Entropy, vol. 19, no. 12, p. 652, 2017.
[29] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," Phys. Rev. Lett., vol. 89, no. 6, p. 68102, 2002.