Machine Learning Algorithms For Diagnose Prostate Cancer

Raj Kumar, Ranjeev Kumar Chopra,
RIMT University, Mandi Gobindgarh, Punjab
Email id- raj.kumar@rimt.ac.in, ranjeevk.chopra@rimt.ac.in

ABSTRACT: ***Machine learning is a type of data analysis that automates the creation of analytical models. It's a subset of artificial intelligence predicated on the notion that machines can learn from data, recognize patterns, and make judgments with little or no human interaction Prostate cancer screening with prostate-specific antigen (PSA) has been widely used, however its accuracy is lacking. Building an efficient statistical model using machine learning methods (MLMs) is a potential strategy for improving accuracy. Prostate cancer is the second leading cause of cancer-related mortality among men.. The rate of mortality caused by prostate cancer can be successfully reduced if disease is detected early. Prostate cancer requires adequate diagnostic techniques and instruments due to the high and multi resolution of Magnetic Resonance Imaging (MRIs). Researchers have previously created computer-aided diagnostic (CAD) systems to assist radiologists in detecting problems. In this study, we used unique machine learning techniques to diagnose prostate cancer, including the Bayesian approach, Support vector machine (SVM) kernels: polynomial, radial basis function (RBF), and Gaussian, and Decision Tree. SVM Gaussian Kernel, based on single feature extracting techniques, has the greatest accuracy of 98.34 percent and an AUC of 0.999. SVM Gaussian kernel with texture + morphological, and EFDs + morphological features produce the maximum accuracy of 99.71 percent and AUC of 1.00 when utilizing a combination of feature extraction algorithms.***

KEYWORDS: ***Decision Tree, Morphological, Support Vector Machine (SVM), Scale Invariant Feature Transform (SIFT), Texture.***

1. INTRODUCTION

In the previous few decades, medical imaging has grown in relevance, particularly in the analysis of various body components [8]. Researchers created a variety of clinical diagnostic methods, including the digital rectal examination (DRE), prostate specific antigen (PSA), transrectal ultrasound (TRUS), and biopsy tests, which are extensively used for detecting prostate cancer regardless of accuracy. PSA was used by Schröder et al. to diagnose prostate cancer, which reduced the death rate by 20%. However, the benefit was linked to overtreatment and over diagnosis. PSA test 44 was found to be ineffective in predicting cancer aggressiveness. Machine learning is a method of data analysis that automates analytical model building.

It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Due to the limited accuracy of Due to its better soft tissue imaging capabilities, magnetic resonance imaging (MRI) has been recommended as a complement to TRUS. Many studies have shown that MRI has a higher resolution than TRUS and can assist detect smaller amounts of prostate cancer with greater precision, making it a potential tool for prostate cancer localisation. The use of a prostate MRI as an additional procedure in the identification of prostate cancer is becoming increasingly widespread[1].

To categorize tissues based on textural features presented, the support vector machine (SVM) is utilized. The Hybrid Morphological-Textural Model, which combines various textural and morphological derived characteristics from MRI for classification, yielded better results in terms of specificity and sensitivity. To perform pixel-wise Bayesian classification at each location, we used extracted texture features, First-Order Statistics, Co-occurrence Features, and Wavelet Features. To get comparable probability scenes, scale the image. Some of the characteristics retrieved are texture, morphological, scale-invariant feature transformations (SIFT), elliptic Fourier descriptors (EFDs), and Entropy base features. The recovered features (single and in various combinations) are used as input by machine learning (ML) classifiers such as SVM with polynomial, RBF, and Gaussian kernels; Bayesian classifiers, and Decision Tree classifiers. Finally, the training and target data were divided using Jackknife 10-fold cross validation to categorize the Prostate and Brach therapy sub groups. This study proposes a variety of feature extraction strategies, including texture, morphological, sample entropy, wavelet entropy, SIFT, and EFDs, to extract useful information from prostate cancer MRIs, which is then fed into novel Machine Learning classifiers, such as the Support Vector Machine

(SVM) and its kernels, Decision Trees, and Bayesian approach[2]. The morphological texture change is seen in Figure 1.

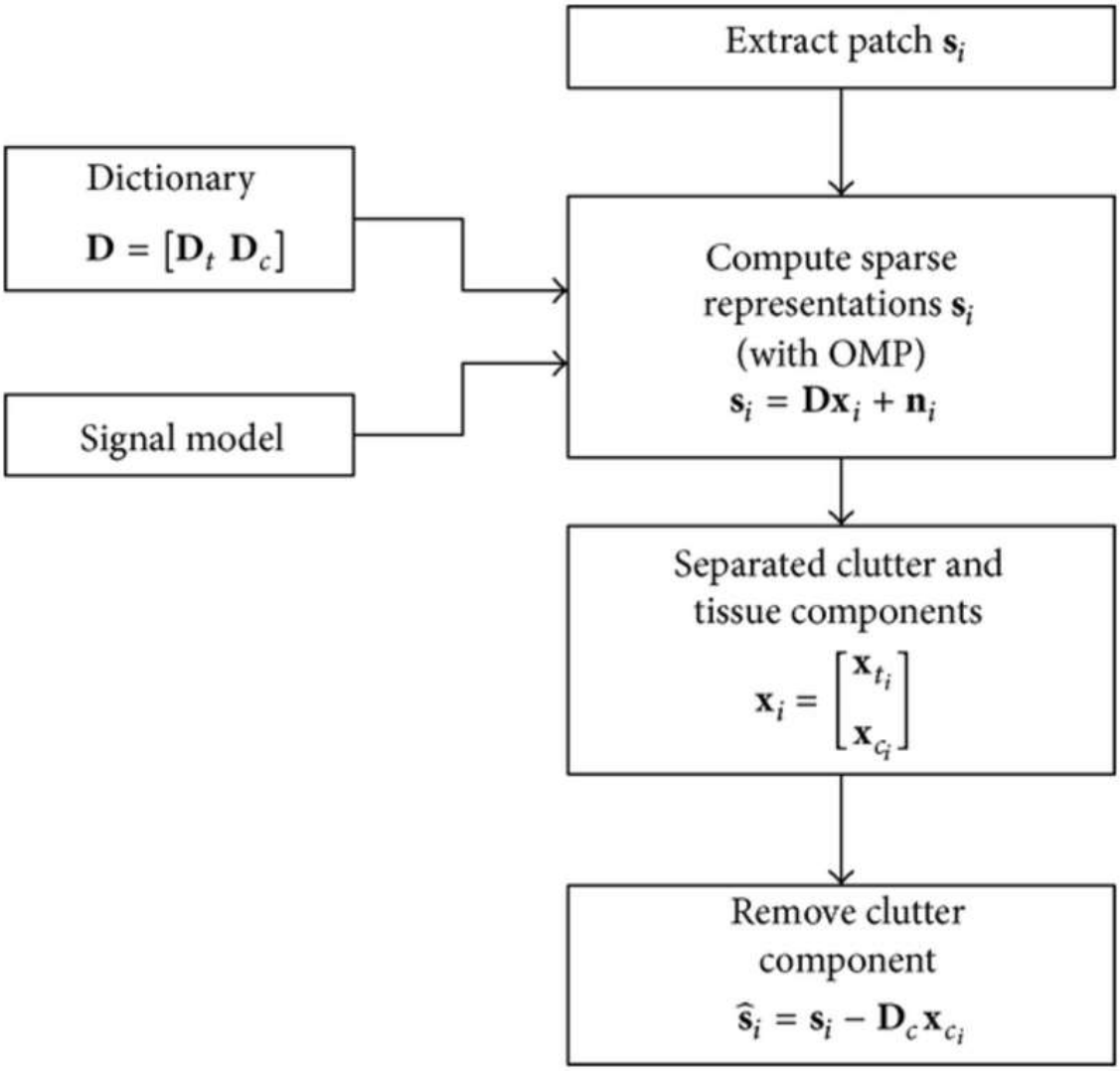

Figure 1: Morphological texture shift.

It extracts texture, morphological, SIFT (scale-invariant feature transformations), EFDs (elliptic Fourier descriptors), and Entropy base features. SVM with polynomial, RBF, and Gaussian kernels; Bayesian classifiers; and Decision Tree classifiers all employ the gathered features (single and in different combinations) as input. Finally, the training and target data were separated using Jackknife 10-fold cross validation to categorize the Prostate and Brach therapy participants[3].

The first step in solving any problem is to extract and identify the relevant data.vant aspects based on the problem's nature and attributes Researchers have already extracted many features for classification and detection applications. To detect and forecast colon cancer rather retrieved geometric and hybrid features [3] .

2. LITERATURE REVIEW

Henry j et al. in their case study suggested that Prostate cancer can pose a modest or high health risk to the patient. The current PSA-based screening method has a high proportion of false positives and false negatives, both of which have severe consequences. Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial at the National Cancer Institute provided us with a dataset of 35,875 patients. We divided the data into three groups: those who did not have prostate cancer, those who had low-risk prostate cancer, and those who had high-risk prostate cancer. We created a pipeline to cope with unbalanced data and offered preparation strategies for such datasets. We looked at the accuracy of the data.[4].

S. Jridi et al. in their case study suggested that Data mining algorithms play a critical role in the early detection of breast cancer. We present a strategy in this research for improving the accuracy and performance of three different classifiers: Decision Tree (J48), Nave Bayes (NB), and Sequential Minimal Optimization (SMO) (SMO). We also use two benchmark datasets to validate and compare the classifiers: Wisconsin Breast Cancer (WBC) and Breast Cancer dataset. Because the chance of examples belonging to the majority class is relatively high, algorithms are far more likely to categorise new observations to the majority class in the classification phase. We deal with such issues[5].

Radhika P.R. et al. in their study suggested that lung cancer is the proliferation of malignant cells in the lungs. Because of the rising rate of cancer incidence, both men and women's mortality rates have increased. Lung cancer is a disorder in which the cells in the lungs grow out of control. Although lung cancer cannot be prevented, the risk of developing it can be lowered. As a result, early identification of lung cancer is critical for patient survival. The number of people diagnosed with lung cancer is directly proportionate to the number of chain smokers. Classification techniques such as Naive Bayes, SVM, Decision Tree, and Logistic Regression were used to analyse lung cancer prediction[6].

Lal Hussain et al. in their study suggested that this Breast cancer is the most commonly diagnosed cancer in women and the second leading cause of cancer death. Radiologists are unable to correctly diagnose breast cancer due to the intricate nature of microcalcification and masses. Researchers have previously developed computer-aided diagnostic (CAD) systems that aid radiologists in quickly detecting anomalies. To distinguish cancer mammography from normal individuals, we used strong Machine learning classification approaches such as Support vector machine (SVM) kernels and Decision Trees. Texture, morphological entropy-based, scale invariant feature transform (SIFT)[7].

3. METHODOLOGY

3.1 Design:

Prostate cancer can pose a modest or high health risk to the patient. The current PSA-based screening method has a high proportion of false positives and false negatives, both of which have severe consequences. Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial at the National Cancer Institute provided us with a dataset of 35,875 patients. We divided the data into three groups: those who did not have prostate cancer, those who had low-risk prostate cancer, and those who had high-risk prostate cancer. We created a channel to cope with unbalanced data and offered preparation strategies for such datasets. We looked at the accuracy of the data[8]. Figure 2 shows that: Base system of machine learning model.

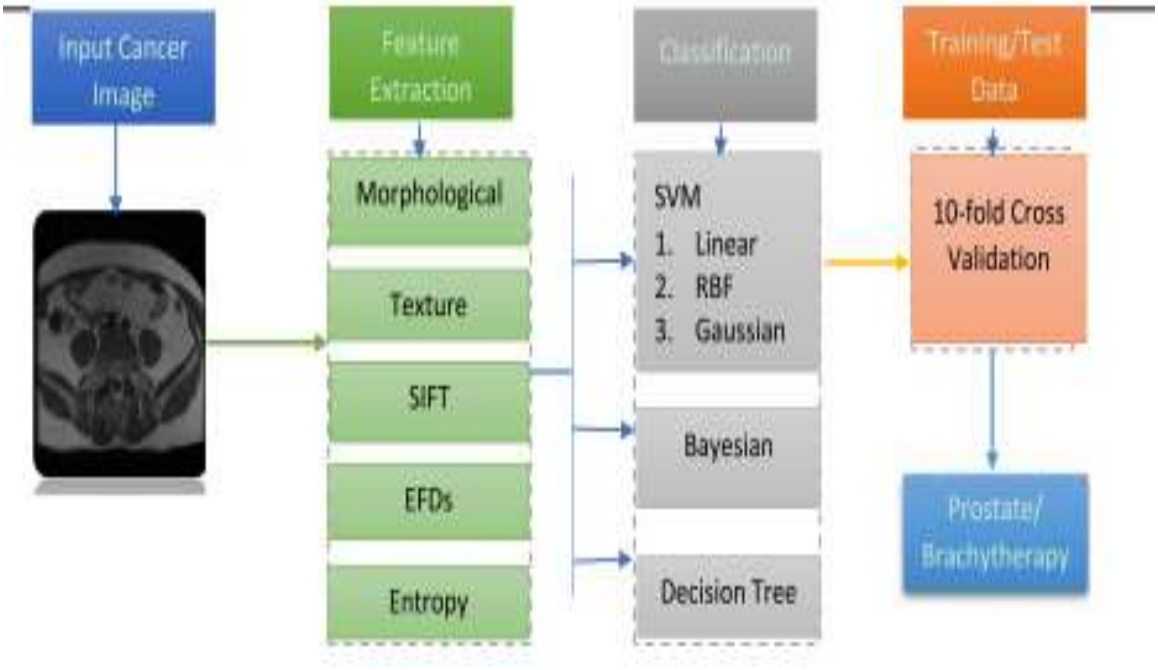

Figure 2: Base system of machine learning model.

3.2 Instruments:

We developed two machine learning models for identifying cancer, one for detecting the presence of cancer and the other for detecting the presence of cancer. There is a cancer with a significant chance of spreading. For the two machine learning models, we labeled the dataset in two ways. The first model classified occurrences without prostate cancer as negative and those with low-risk or high-risk prostate cancer as positive, resulting in the presence dataset. A kind of cancer that affects males is prostate cancer (PoPC).

3.3 Sample:

The patient's PSA levels (note that these are total PSA, which includes PSA bound to other proteins and unbounded PSA), Overall ROC, and Recent ROC, as well as the patient's PSA levels (note that these are total PSA, which includes PSA bound to other proteins and unbounded PSA) were used to develop our initial models. K-neighbors (KN), support vector machine (SVM), decision tree (DT), random forest (RF), multi-layer perceptron classifier (MLPC), adaptive boosting (ADA), and quadratic discriminant analysis were all employed on the dataset and were all acquired from scikit-learn (QD). The radial basis function was used as a kernel function in the support vector machine, which had a gamma value of 2.[9][10].

3.4 Data Analysis:

To evaluate the effect that Overall ROC and Recent ROC had on accuracy, we tested the difference in holdout AUC for the optimal classifiers in each dataset when each of the two ROC features was individually removed and when both were removed. To test the effectiveness of age and BMI, we measured the increase in AUC holdout when age alone, BMI alone, and both age and BMI were added to the model so far developed. Since 90% of the patients are white Non-Hispanic people , we tested whether filtering by race would have an effect on the accuracy by filtering out all the patients with races other than white Non-Hispanic and seeing if the models built on the remaining data have a higher accuracy.

4. RESULTS AND DISCUSSION

The ability of MRI (Magnetic resonance imaging) to identify prostate cancer has been widely researched. End rectal coil MRI gives for a clearer view of prostate zonal architecture as well as the location and extent of tumors inside the gland. The validation cohort comprised of prospective biopsies done from July 2017 to November 2019. The training cohort consisted of consecutive systematic prostate biopsies taken from January 2003 to June 2017.. Men were included if their PSA was between 0.4 and 50 ng/mL and they had information about their DRE, TRUS prostate volume, and TRUS abnormality. Gleason 3 + 4 or higher tumours were considered clinically significant PCa (csPCa). The area-under-curve (AUC) of receiver-operating characteristics (ROC) was compared between PSA, PSA density, the European Randomized Study of Screening for Prostate Cancer (ERSPC) risk calculator (ERSPC-RC), and various ML techniques using PSA, PSA density, the European Randomized Study of Screening for Prostate Cancer (ERSPC) risk calculator (ERSPC-RC), and various ML techniques using PSA, PSA density, the European Randomized Study of Screening.

The RF model outperformed other ML methods, as well as PSA, PSA density, and ERSPC-RC, in predicting PCa or csPCa in the validation population. In csPCa prediction, the AUC of PSA, PSA density, ERSPC-RC, and RF were 0.71, 0.80, 0.83, and 0.88, respectively. For csPCa at 90–95 percent sensitivity, the RF model provides a negative predictive value (NPV) of 97.5–98.0 percent, eliminating 38.3–52.2 percent unnecessary biopsies. In terms of clinical benefit, the RF model beat PSA, PSA density, and ERSPC-RC, according to decision curve analysis (DCA). ML methods outperformed ERSPC-RC or PSA density in predicting csPCa using the same clinical criteria, and might save up to 50% in unnecessary testing. Figure 3 shows the graph between true positive rate and false positive rate.

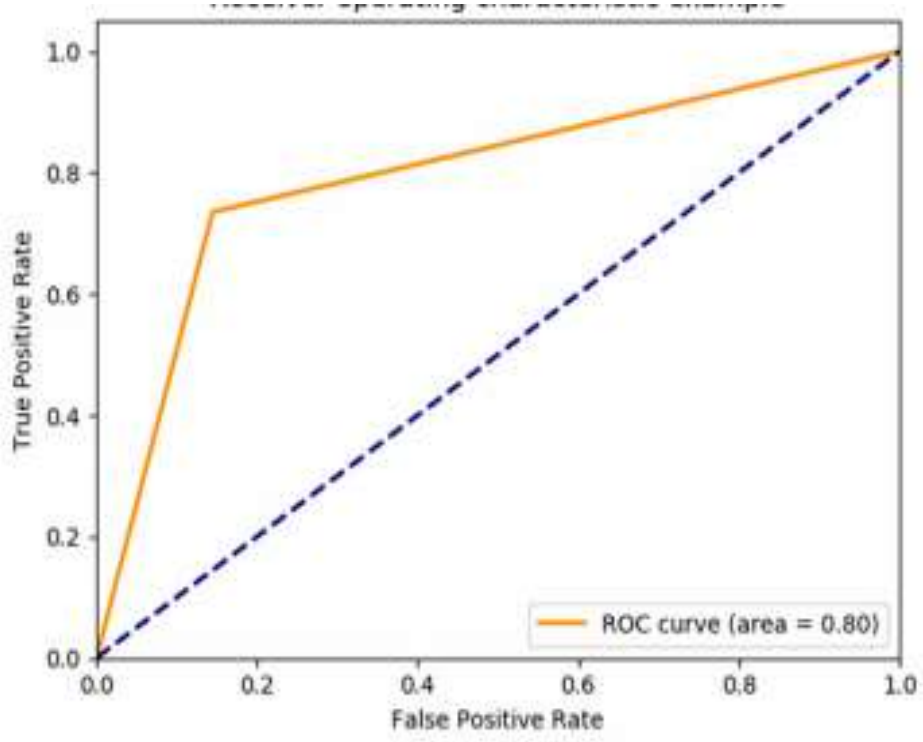

Figure 3: Graph between true positive rate and false positive rate.

5. CONCLUSIONS AND IMPLICATIONS

Breast cancer is the most prevalent malignancy in women and the second greatest cause of death from cancer. Radiologists are unable to correctly diagnose breast cancer due to the intricate nature of microcalcification and masses. Researchers have previously developed computer-aided diagnostic (CAD) systems that aid radiologists in quickly detecting anomalies. To distinguish cancer mammography from normal individuals, we used strong Machine learning classification approaches such as Support vector machine (SVM) kernels and Decision Trees. Texture, morphological entropy-based, scale invariant feature transform (SIFT), and elliptic Fourier d are some of the features presented. Correlations have been found in studies.reducing androgens and raising estrogens, there is a link between these steroids and prostate cancer.Prostate cancer is more likely to develop as a result of this. When several kinds of estrogens are found, however, the picture gets more muddled, as the activation of classical estradiol becomes more convoluted. Receptors (and) have a range of impacts on prostate cancer progression, which can be difficult to anticipate due to the sometimes complicated nature of prostate cancer. The receptors implicated influence the contradictory nature of the findings, as well as the present insufficiency of our prostate cancer models and the function of estrogens (whether they enhance or decrease cancer spread). The Gleason grade of the biopsy material is critical in defining case treatment in prostate cancer. However, Gleason grading is linked with significant interobserver variability, necessitating the use of decision support systems to improve Gleason grading's repeatability in clinical practice.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2018,” *CA. Cancer J. Clin.*, 2018.

[2] J. J. Valletta, C. Torney, M. Kings, A. Thornton, and J. Madden, “Applications of machine learning in animal behaviour studies,” *Animal Behaviour*. 2017.

[3] J. A. Cruz and D. S. Wishart, “Applications of machine learning in cancer prediction and prognosis,” *Cancer Informatics*. 2006.

[4] H. J. Adler *et al.*, “Community network for deaf scientists,” *Science*. 2017.

[5] S. Jridi, R. Azzeddine, and J. E. Bourkadi, “DRESS syndrome secondary to antituberculosis drugs: About a case,” *Pan Afr. Med. J.*, 2017.

[6] P. R. Sudhakaran, A. Radhika, and K. Saja, "Tu-P7:261 Inflammation dependant upregulation of matrix metalloproteinases in monocyte-macrophages," *Atheroscler. Suppl.*, 2006.

[7] L. Hussain, W. Aziz, S. Saeed, S. Rathore, and M. Rafique, "Automated Breast Cancer Detection Using Machine Learning Techniques by Extracting Different Feature Extracting Strategies," in *Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018*, 2018.

[8] M. P. Menden *et al.*, "Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties," *PLoS One*, 2013.

[9] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*. 2015.

[10] R. D. Clark, "Emerging applications for high K materials in VLSI technology," *Materials*. 2014.