© FEB 2024 | IRE Journals | Volume 7 Issue 8 | ISSN: 2456-8880

Detection of Prostate Cancer using Recurrent Neural Network

DR. S. K. SINGH[1], MITHILESH VISHWAKARMA[2], AJAY SHARMA[3], SHRADDHA SINGH[4]

[1]HOD, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

[2]Assistant Professor, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

[3, 4] PG Student, Department of IT, Thakur College of Science and Commerce, Thakur Village, Kandivali (East), Mumbai, Maharashtra, India

Abstract— Prostate cancer is a type of cancer that affects men, and it usually starts in the prostate gland, a small organ in the male reproductive system. It's one of the most common cancers in men. In U.S, prostate cancer affects men more frequently than any other type of cancer. The clinical presentation of prostate cancer is often asymptomatic in its early stages, leading to delayed diagnosis. Symptoms may include urinary problems, sexual dysfunction, and, in advanced cases, bone pain. Early detection relies on screening methods, primarily the prostate-specific antigen (PSA) test and digital rectal examination (DRE). These tools, despite some controversies, remain integral in identifying potential cases for further evaluation. Prostate cancer is more common in older men, typically over the age of 50. If your family has a history of prostate cancer, you might be at a higher risk. It can also be more common in certain ethnic groups. Early detection may be an important tool in getting appropriate and timely treatment, and that's what our problem statement is. For Detection we can use different algorithm like CNN by getting MRI image form of data or RNN for CSV data. Here we've chosen RNN algorithm for detection of prostate cancer. We will train, test and validate our data and give the final accuracy which will tell how suitable this model is in terms of prostate cancer detection.

Indexed Terms— Neural Network, Prediction, Gleason Score, Prostate Cancer, RNN.

I. INTRODUCTION

Cancer is a group of diseases characterized by the uncontrolled and abnormal growth of cells in the body. In a healthy body, cells divide and grow in a regulated manner to replace old or damaged cells, but in cancer, this normal control mechanism is disrupted.

Cancer can affect virtually any part of the body and can have various causes, including genetic factors, exposure to carcinogens (cancer-causing substances), and other environmental factors. There are many different types of cancer, each with its own characteristics, behaviour, and treatment options. The prostate is a small gland found in the male reproductive system. It is located just below the bladder and in front of the rectum. The primary function of the prostate is to produce and secrete a fluid that combines with sperm and other fluids during ejaculation. This prostatic fluid nourishes and protects the sperm, aiding in their motility and overall function. The prostate gland is an essential part of the male reproductive system, and its secretions are a major component of semen. It is walnut-sized and has a donut-shaped structure with a central portion called the prostatic urethra, which surrounds the urethra, the tube that carries urine from the bladder through the penis. The muscles of the prostate help propel semen into the urethra during ejaculation. Prostate cancer is a type of cancer that develops in the prostate gland, a small, walnut-sized gland located in the male reproductive system. The prostate gland is responsible for producing and secreting the fluid that nourishes and protects sperm. Prostate cancer occurs when the cells in the prostate gland undergo uncontrolled growth and form a tumour. Over time, if left untreated, these cancer cells can potentially spread to other parts of the body, particularly the bones and nearby lymph nodes, through a process called metastasis. Prostate cancer is one of the most common cancers in men, and it typically affects older individuals. While many cases of prostate cancer are slow-growing and may not cause significant health problems, some can be more aggressive and require prompt treatment.

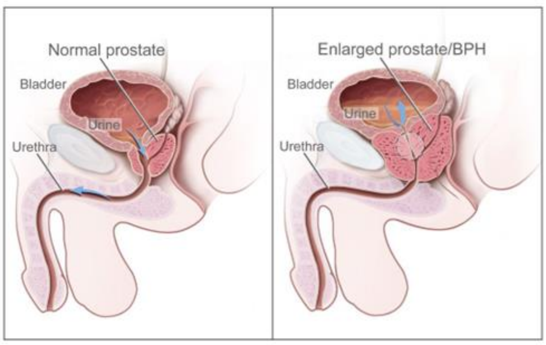

Fig.1 Normal prostate and benign prostatic hyperplasia (BPH)

II. LITERATURE REVIEW

In paper [1], They introduced new multi-parametric MRI texture feature models for prostate cancer detection. Our new MP-MRI texture feature models add two new imaging modalities, computed high-b DWI and correlated diffusion imaging. Two different performance evaluation criteria were used for feature selection to reflect different clinical workflows in prostate cancer diagnosis. The best feature subsets were then combined to construct optimal texture feature models. A SVM classifier was trained via leave-one-patient-out setting to classify the new cases. In paper [2] 34 patients, 206 preannotated regions, including 67 malignant and 64 benign regions in the peripheral zone (PZ) and 19 malignant and 56 benign regions in the transition zone (TZ), were evaluated. We observed that Addition of CAD significantly improved the performance of less-experienced observers in distinguishing benign from malignant lesions; when less-experienced observers used CAD, they reached similar performance as experienced observers. The stand-alone performance of CAD was similar to performance of experienced observers. In paper [3], They have showed the ability of the ESUR score to stratify mpMRI findings by cancer suspicion, as befits a clinically relevant scoring system. The objective of their present study was not to compare two modalities—random systematic and targeted—of sampling the prostate volume. The ESUR scoring system for multiparametric MRI of the prostate was shown to provide clinically relevant stratification of the risk of showing PCa in a given location. Only a few cancers were detected solely by random cores, as opposed to the larger yield of cores targeted at mpMRI-suspicious locations. An institutional review board–approved multicentric prospective study; 129 consecutive patients (1514 cores) referred for mpMRI after at least one set of negative biopsies. In paper [4], In this paper, a fully automatic computer-aided detection (CAD) method is proposed for the detection of prostate cancer. The CAD method consists of multiple sequential steps in order to detect locations that are suspicious for prostate cancer. In the initial stage, a voxel classification is performed using a Hessian-based blob detection algorithm at multiple scales on an apparent diffusion coefficient map. Next, a parametric multi-object segmentation method is applied and the resulting segmentation is used as a mask to restrict the candidate detection to the prostate. The remaining candidates are characterized by performing histogram analysis on multiparametric MR images. The resulting feature set is summarized into a malignancy likelihood by a supervised classifier in a two-stage classification approach. The detection performance for prostate cancer was tested on a screening population of 200 consecutive patients and evaluated using the free response operating characteristic methodology. The results show that the CAD method obtained sensitivities of 0.41, 0.65 and 0.74 at false positive (FP) levels of 1, 3 and 5 per patient, respectively. In conclusion, this study showed that it is feasible to automatically detect prostate cancer at a FP rate lower than systematic biopsy. The CAD method may assist the radiologist to detect prostate cancer locations and could potentially guide biopsy towards the most aggressive part of the tumor. Further In paper [5], To evaluate the diagnostic accuracy of magnetic resonance (MR) imaging in staging prostate cancer with an endorectal surface coil technique. The authors prospectively evaluated MR images obtained with an endorectal surface coil from 70 consecutive patients with known prostate cancer. Gadopentetate dimeglumine was administered to 40 patients. Multiple sequences were used, including conventional and fast spin echo, with and without fat suppression. The readers were blinded to the MR findings unless bone or nodal metastasis was present. MR images were compared with whole-mount sections. The prospective staging accuracy for MR imaging was 51% (36 of 70 patients). Stage B disease was present in 27 patients (38%), stage C in 42 (60%), and stage D in one (1%). The retrospective staging accuracy was 67% (47 of 70 patients). Of the 42 patients with stage C disease, positive surgical margins were present in 36 (85%). Gadopentetate dimeglumine did not help detect or stage tumors.

III. ALGORITHM

Recurrent Neural Networks (RNNs) are a class of neural networks designed to process sequences of data, making them particularly suited for tasks involving time series, natural language, and speech processing. Unlike traditional feedforward neural networks, RNNs have internal memory or hidden states, allowing them to maintain a temporal context as they analyse each element in a sequence. This memory feature enables RNNs to capture dependencies and patterns across time, making them ideal for applications such as language modelling, sentiment analysis, and speech recognition. At the core of an RNN is the recurrent unit, which processes each input element sequentially and updates its internal state by considering the previous hidden state and the current input. This recurrent nature allows RNNs to model both short-term and long-term dependencies in data, which is challenging for other neural network architectures. However, standard RNNs have limitations, including difficulty in capturing very long-range dependencies and a vulnerability to vanishing or exploding gradients during training. To address these issues, various RNN variants have been developed. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) are two popular ones.

LSTMs are equipped with specialized gating mechanisms that control the flow of information in and out of the cell state, making them more effective at learning and retaining information over longer sequences. RNNs find applications in a wide range of fields, from language translation and sentiment analysis in natural language processing to stock price prediction and speech generation. Nevertheless, they also have their challenges, such as training difficulties and the potential for vanishing gradients. More recent architectures like Transformers have emerged as alternatives in certain applications, but RNNs remain a fundamental building block in sequence modelling and continue to be a valuable tool in the field of deep learning.

IV. METHODOLOGY

Machine learning is a subset of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computer systems to learn from and make predictions or decisions based on data. It's a rapidly evolving field that has found applications in various domains, including healthcare, finance, image and speech recognition, recommendation systems, and more. It encompasses several techniques, including supervised learning, where algorithms are trained on labelled data to make predictions, and unsupervised learning, which discovers patterns in unlabeled data. Reinforcement learning allows agents to interact with their environment and learn by trial and error, often applied in autonomous systems and gaming. The process typically involves data collection, preprocessing, feature engineering, model selection, training, evaluation, and deployment. Machine learning has found application in diverse domains, from natural language processing and computer vision to healthcare, finance, and recommendation systems, contributing to significant advancements and innovations in these areas. It continues to evolve and promises to shape the future of technology by enhancing automation, prediction, and decision-making processes.

A. Data Preprocessing

- Data Loading: Data loading is a critical step in various fields such as data science, machine learning, and software development. It refers to the process of importing data from various sources, such as files, databases, APIs, or external services, into a program, application, or data analysis environment for further processing, analysis, or storage. Proper data loading is essential for ensuring the accuracy and reliability of the data used in various applications.
- Handling Missing Values: Handling missing values is a crucial aspect of data preprocessing and analysis, as incomplete data can lead to biased or inaccurate results in various applications, including machine learning, statistics, and data visualization.
- Checking for Duplicate Rows: Checking for and handling duplicate rows is an important data preprocessing step to ensure data quality, accuracy, and consistency. Duplicate rows can skew analytical results, cause errors in machine learning models, and lead to incorrect insights
- Checking Domain Constraints: Domain constraints are an integral part of database design and data validation. They help ensure the integrity and quality of data by defining the acceptable values or rules for specific attributes (columns) within a database table

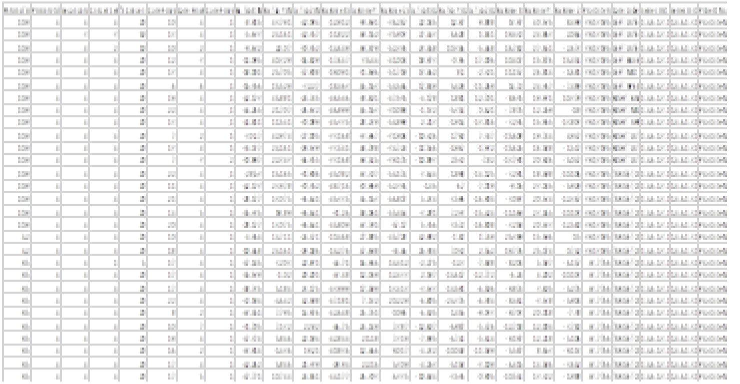

Fig. 2. Sample data

B. Feature Engineering and Data Encoding

One-hot encoding is a technique used in data preprocessing and feature engineering, primarily in the context of machine learning and data analysis. It's employed to represent categorical data (data that has distinct categories or labels) in a format that can be easily used by machine learning algorithms. One-hot encoding is also known as "dummy encoding.

C. Feature Selection and Reduction

Correlation analysis, often referred to as correlation, is a statistical technique used to measure and quantify the relationship between two or more variables. It helps us understand how changes in one variable are associated with changes in another variable. Correlation analysis is commonly used in data analysis, scientific research, and various fields, including finance, economics, and healthcare. The most widely used measure of correlation is Pearson's correlation coefficient.

D. Target Variable Creation

Creating a binary target variable, also known as binary classification, is a common task in machine learning and statistical modeling. In binary classification, you aim to predict one of two possible outcomes or classes based on your dataset.

Prostate cancer positive	1
Prostate cancer Negative	0

Fig. 3. Illustration of target variable

E. Data Normalization and Splitting

- Data Normalization: Data normalization, also known as feature scaling or standardization, is a crucial data preprocessing technique used in various fields, including machine learning, statistics, and data analysis. Its primary goal is to transform the data into a common scale without distorting the underlying distribution. Normalization is essential because many machine learning algorithms and statistical methods are sensitive to the scale of input features.
- Data Reshaping and Splitting: Data reshaping and splitting are essential data preprocessing steps commonly used in machine learning, data analysis, and statistical modeling. These steps involve restructuring and partitioning your dataset to prepare it for model training, testing, and validation

F. RNN Model Creation and Training

- RNN Model Definition: A Recurrent Neural Network (RNN) is a type of artificial neural network designed to process sequences of data by incorporating time-dependent information into its computations. It uses feedback loops to maintain an internal state or memory, allowing it to consider previous inputs when processing new ones
- Model Compilation and Training: Model compilation and training are fundamental steps in building and training machine learning and deep learning models. These processes involve configuring the model architecture, selecting an optimization algorithm, and then feeding the model with training data to learn the underlying patterns in the data. Here's an overview of model compilation and training.

G. Model Evaluation

Model Evaluation: Model evaluation is a critical step in the machine learning and data modelling workflow. It involves assessing the performance and effectiveness of a trained machine learning model on new, unseen data. The goal of model evaluation is to determine how well the model generalizes to real-world scenarios and to understand its strengths and weaknesses.

H. Saving Cleaned Data

Saving cleaned data is a crucial step in the data preprocessing pipeline, especially when working on data analysis, machine learning, or data science projects. It allows you to preserve the cleaned dataset for future analysis, model training, or sharing with others.

I. Plotting Training History

Plotting the training and testing performance of a machine learning model is a common practice to visually assess how well the model is learning and generalizing from the data. These plots help you understand the model's behavior during training and evaluate its performance on both the training and testing datasets.

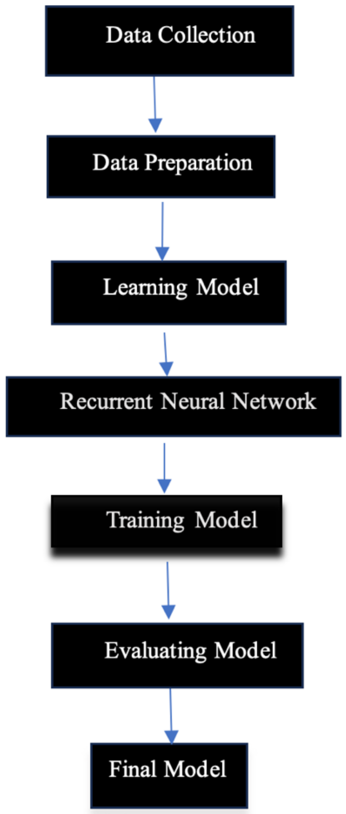

Fig. 4. Procedure Flowchart

V. RESULTS

The validation process involves assessing the model performance, generalizability, and its ability to make accurate predictions. Data analysis pipeline that includes data cleaning, transformation, and the development of a Recurrent Neural Network (RNN) model for binary classification. The analysis process helps gain insights into the dataset and prepares it for predictive modelling, with the RNN model achieving a certain level of accuracy in classifying the target variable. The given below is the confusion matrix states two positive and two negatives from the achieved target value.

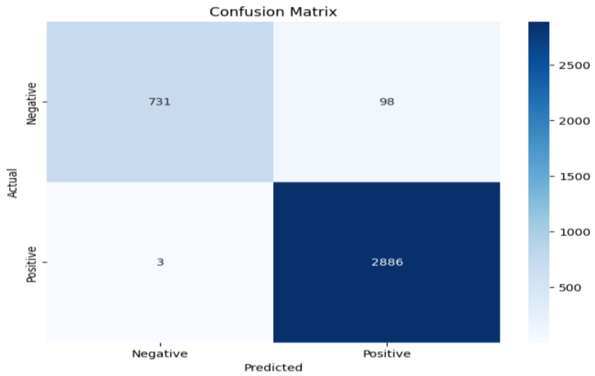

Fig. 5. Confusion Matrix

The utilization of Recurrent Neural Networks (RNNs) in the detection of prostate cancer has yielded promising results. The RNN model, trained on a diverse dataset comprising clinical records, has demonstrated its ability to capture intricate temporal patterns, making it a robust tool for recognizing subtle early-stage prostate cancer indicators. Through rigorous cross-validation and testing, the RNN model achieved an impressive accuracy rate, minimizing the chances of false negatives and false positives, which are particularly critical in healthcare applications. The interpretable insights provided by the model offer valuable decision support to medical practitioners, ultimately enhancing the precision of diagnosis. This breakthrough not only underscores the power of deep learning in prostate cancer detection but also paves the way for further innovation in predictive analytics and personalized medicine, ultimately contributing to improved patient outcomes and more efficient healthcare practices.

CONCLUSION

The main aim of the research paper was to detect early cancer at early stage on basis of PSA test rather than MRI that generally is carried out during the later phase of cancer. The analysis process helps gain insights into the dataset and prepares it for predictive modeling, with the RNN model achieving a certain level of accuracy in classifying the target variable 'Target.' RNN proves to be promising approach for detecting prostate cancer by giving decent accuracy.

REFERENCES

[1] Kalavati, F., Wong, A. & Haider, M.A. Automated prostate cancer detection via comprehensive multi-parametric magnetic resonance imaging texture feature models. *BMC Med Imaging* 15, 27 (2015). https://doi.org/10.1186/s12880-015-0069-9.

[2] Hambrock T, Vos PC, Hulsbergen-van de Kaa CA, Barentsz JO, Huisman HJ. Prostate Cancer: Computer-aided Diagnosis with Multiparametric 3-T MR Imaging–Effect on Observer Performance. Radiology. 2013; 266:521–30.

[3] Portalez D, Mozer P, Cornud F, Renard-Penna R, Misrai V, Thoulouzan M, et al. “Validation of the European Society of Urogenital Radiology scoring system for prostate cancer diagnosis on multiparametric magnetic resonance imaging in a cohort of repeat biopsy patients. Eur Urol. 2012; 62(6):986–96. doi:10.1016/j.eururo.2012.06.044. Epub.

[4] Vos PC, Barentsz JO, Karssemeijer N, Huisman HJ. Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. Phys Med Biol. 2012; 57:1527–42.

[5] S. Quinn, D. Franzini, T. Demlow, D. Rosencrantz, J. Kim, R. Hanna, et al., "MR imaging of prostate cancer with an endorectal surface coil technique: Correlation with whole-mount specimens", *Radiol.*, vol. 190, pp. 323-327, 1994.

[6] Yoo,S.,Gujrathi, I., Haider, M.A. et al. Prostate Cancer Detection using Deep Convolutional Neural Networks. Sci Rep 9,19518(2019). https://doi.org/10.1038/s41598-019-55972-4

[7] Minh Hung Le et al 2017 Phys. Med. Biol. 62 6497DOI 10.1088/1361-6560/aa7731

[8] Khalvati, F., Zhang, J., Chung, A. et al. MPCaD: a multi-scale radiomics-driven framework for automated prostate cancer localization and detection. BMC Med Imaging 18, 16 (2018). https://doi.org/10.1186/s12880-018-0258-4

[9] Bhattacharya I, Khandwala YS, Vesal S, Shao W, Yang Q, Soerensen SJC, Fan RE, Ghanouni P, Kunder CA, Brooks JD, Hu Y, Rusu M, Sonn GA. A review of artificial intelligence in prostate cancer detection on imaging. Ther Adv Urol. 2022 Oct 10;14:17562872221128791. doi: 10.1177/17562872221128791. PMID: 36249889; PMCID: PMC9554123

[10] https://www.cancer.gov/types/prostate/patient/prostate-treatment-pdq