

# Early Detection and Diagnosis of Prostate Cancer using Artificial Intelligence Concept

Onuiri Ernest E.

Department of Computer Science  
Babcock University  
Ilishan-Remo Ogun State, Nigeria.

Awodele Oludele

Department of Computer Science  
Babcock University  
Ilishan-Remo Ogun State, Nigeria.

Ebiesuwa Oluwaseun

Department of Computer Science  
Babcock University  
Ilishan-Remo Ogun State, Nigeria.

## ABSTRACT

Contemporary research has shown that prostate cancer is the most common cancer amongst older men which can lead to death. However, it has been proven that early detection for most cancers is known to help in making therapeutic decisions in urologic medicine and oncology. Current methods of screening for prostate cancer carried out through blood PSA tests (presence of high Prostate Specific Antigen in the blood) and digital rectal examinations due to their morphological nature lead to a high percentage of False Positive Test Results (FPTR). The occurrence of this FPTR can be reduced by employing Artificial Intelligence (AI) techniques such as Artificial Neural Network (ANN) in evaluating the need for a patient to undergo biopsy. The goal of this research entails the proposition of a model for the early detection of prostate cancer to enhance early detection so as to enable early commencement of intervention to increase the chances of successful treatment and reduction in mortality rates. This research was predicated on the review and analysis of current AI paradigms tailored towards effective detection of prostate cancer. Research findings showed that the PSA test and Digital Rectal Examination (DRE) are not adequate enough to detect prostate cancer but the deployment of assay specific ANN model, which is a machine learning technique reduces the amount of unnecessary biopsies.

## Keywords

Prostate Cancer, Artificial Intelligence, Prostate Specific Antigen, Digital Rectal Examination, Biopsy.

## 1. INTRODUCTION

Cancer as a chronic disease has continued to attract global attention partly because of its deadly impact and the amount of resources in terms of time, money and man-power dedicated to finding a lasting solution to the scourging evil. Various medical trials for medicines which were presumed to be able to cure the disease have failed at the last stage of testing, probably because the initial data was not quite as encompassing as proponents thought, thus the need for a wider scope that captures all the necessary variables. Be that as it may, several studies have shown that the ever increasing number of victims is not decidedly due to a single risk factor, but rather a range of factors which have been considered to increase an individual's susceptibility to the disease; such as race, heredity, exposure to ultraviolet rays, nature of job, diet and Body Mass Index (BMI). The term "cancer" originates from the Latin word for crab because cancers are often very irregularly shaped, and like the crab, "grab on and don't not let go." The term cancer specifically refers to a new growth which has the ability to invade surrounding tissue(s), metastasize (spread to other organs) and which may eventually lead to the patient's death [1]. The most important question however, is that which is associated with

ascertaining the impact of early detection and diagnosis on patient survival.

Cancers usually start as being primary to a specific organ in the body and later metastasize to other parts. One of such is prostate cancer, which is also medically referred to as carcinoma of the prostate [2]. The prostate being an organ in the male reproductive system, is responsible for secreting the fluid that protects and nourishes the sperm cells in the seminal fluid (semen), and is located below the urinary bladder and right in front of the rectum [3]. Also, the urethra which carries urine from the bladder through the penis and out of the body, passes through the center of the prostate. Due to the apparent positioning of the prostate, the symptoms of the presence of a cancer include but are not limited to urinary dysfunction, difficulty with erection and poor ejaculation [4]. Statistics from research conducted worldwide have also shown that this type of cancer is very uncommon in men below 45 years of age and a lot of men above 75 are more likely to die with the disease than from the disease [5].

Prostate cancer is often misinterpreted as Benign Prostatic Hyperplasia (BNH), which occurs when the prostate begins to press against the urethra as a result of growth, thereby causing urinary difficulty. BNH may present similar symptoms but does not evolve into a cancer; it is also not as threatening but can pose a serious problem for some men. Treatment of BNH may be through the use of medicine or transurethral surgery. Proper testing using techniques such as biopsies (a medical test involving sampling tissues or cells for examination), prostate imaging, tumor markers, serum proteomic profiling and prostatic specific antigen testing, can help differentiate this from other diseases of the male reproductive system.

The need for accurate techniques of detection and diagnosis of the malignant potential (how fast is it killing the victim?) of any given man's prostate cancer is constantly increasing. Studies have shown that this is important not only for identification of the extent and location of the malignancy but also, perhaps more importantly, for determination of malignant potential. Various AI concepts are being explored to aid in this pursuit such as data mining and ANNs [6].

Screening refers to testing to find a disease whether present or not [3]. Some prostate cancer screening tests include; testing the level of Prostate Specific Antigen (PSA) in the blood, Digital Rectal Exam (DRE) in which the doctor puts a gloved finger into the rectum and feels the prostate for any abnormalities. If these tests come out positive, further screening is conducted. PSA screening is controversial, primarily because of the absence of randomized trials documenting that early detection and aggressive treatment of prostate cancer can reduce patient mortality. Accurate staging of prostatic carcinoma on an individual patient basis remains

problematic still. Various models and techniques are being explored to rectify this such as; Color Doppler Imaging for detection of cancer; pelvic lymphadenectomy as a diagnostic tool; and a new, AI—based model to predict survival [7].

It is a well-known fact that the use of AI-concepts will improve the accuracy of solutions to problems in many fields, not excluding Oncology. In this study, the review captures the effect of using machine learning techniques to analyze data to find complex patterns in information, also, discussing the impact of ANNs in urologic oncology using training data to predict the outcome of a prostate biopsy. Currently, six validated predictive models have been published using pre-biopsy parameters to predict prostate biopsy outcome. Five of the six are ANN models, and one is based on logistic regression [8].

## **2. PROBLEM STATEMENT**

Suffice it to say that prostate cancer is the second leading cause of mortality in men, exceeded only by lung cancer. The causes remain to be construed, but factors such as diet, heredity, and environmental factors that affect male hormones (androgens) have been implicated in epidemiological studies [9]. Early stage detection and diagnosis of prostate cancer is difficult with the conventional techniques as they rely on the physical and morphological characteristics of the tumor [10], and can therefore give false positive tests results. Hence, the need to organize the large pool of data available from various screenings whether blood, tissue or otherwise. This task can be easily achieved with AI-techniques such as; machine learning and data mining, the data can be easily evaluated thus providing an impressive array of test outcomes. The impact of application of the aforementioned concepts in this process is immeasurable, as it aids oncologists in determining the kind of treatment for a patient whether palliative or curative, if an individual truly has the cancer. Various models are being explored and in retrospect, easy and immediate physician access to these models will be imperative, if their full potential is to be realized [8].

## **3. AIM AND OBJECTIVES**

The aim of this study is the identification of artificial intelligence models for early detection and diagnosis of prostate cancer, which will eradicate the need of more tissue sampling exercises and increase the life expectancy of patients. Specific objectives of this study are to;

- i. Identify artificial intelligence techniques for improved decision making in urologic oncology.
- ii. Propose an ANN model for identifying patterns in prostate tissue and patient blood sampling.
- iii. Identify variables to serve as input in the ANN model to improve test accuracy and reduce number of unnecessary biopsies.

## **4. REVIEW OF RELATED WORKS**

Various studies have been carried out on the use of artificial intelligence techniques such as data mining in serum proteomic marking and artificial neural network models on early detection of prostate cancer and what their impact will be on the rate of unnecessary biopsies. Some of these works are discussed.

Following a model developed by Snow et al, a neural network algorithm was designed by Babbian and his colleagues, a paradigm that predicts prostate biopsy outcome in men with a low PSA level (2.5-4.0 ng/mL) based on retrospective data of

151 men of PSA levels who underwent a biopsy. Using input units such as age, free PSA, creatinine ketase and total PSA. The ANN consisted of three individually trained networks that were developed with data from three institutions. Cancer was detected in 24.5% of the men. On comparing this algorithm with other PSA parameters, it showed a considerable shift. It was better in specificity, but remained constant in terms of ROC. It was however commented that this discrepancy might have been as a result of the difficulty in using ROC kit to monotonically map the generated index. The authors thus concluded that there will be a 39% reduction of unnecessary biopsies if ANN index is used instead of free/total PSA. In a similar work, variables such as %PSA, PSA transfer zone and fPSA (free PSA) for men with levels between 2.5ng/ml and 4.0 ng/ml. They observed that novel and accurate ANNs represent a significant enhancement in propensity to detect prostate cancer early (high specificity) and avoid unnecessary biopsies (high sensitivity) [11] [12].

### **4.1. Combining Artificial Neural Networks and Trans-Rectal Ultrasound in the Diagnosis of Prostate Cancer**

In this study, it was observed that more than one million Trans-Rectal Ultra-Sound (TRUS) - guided prostate needle biopsies are carried out every year in the United States, which leads to the detection of 200,000 new cases annually. Sadly, irrespective of this fact, the urologist's ability to diagnose prostate cancer has not kept pace with therapeutic advances; in recent times, a lot of men are facing the need for prostate biopsy with the tendency of an inconclusive result. This research focused on the tools available to help the clinician in predicting the outcome of the prostate needle biopsy such as the use of machine learning models in the form of ANNs, to predict prostate biopsy outcomes using pre-biopsy variables such as age and free PSA. The role of ANNs in providing valuable predictive models to be used in conjunction with TRUS looks promising. It was observed that in the few studies that have compared machine learning to traditional statistical methods, ANN and logistic regression appear to function equivalently when predicting biopsy outcome. In conclusion they stated that with the introduction of more complex pre-biopsy variables, ANNs are in a commanding position for use in predictive models [8].

### **4.2. Early Diagnosis and Staging of Prostate Cancer**

This research entailed the combined modality staging of disease, color Doppler imaging for detection of cancer during biopsy since it helps to locate the exact site where the cancer is situated so as to prevent false negative biopsy result. It was stated that accurate staging of prostatic carcinoma on an individual patient basis is problematic and that staging of the cancer is important, not only for identification of the extent and location of the malignancy but also, perhaps more importantly, for determination of malignant potential. There has been significant migration in the clinical presentation of prostatic carcinoma, largely related to improved biopsy methods and, especially, to PSA testing. In recent times, a much greater percentage of patients present with more favorable malignancies. In addition, a number of these might have cancers of low malignant potential—low enough that competing causes of mortality might be a more important consideration. Thus, the need for accurate methods of ascertaining the malignant potential of any given man's cancer has never been greater than it is today [6].

### **4.3. Artificial Neural Networks and Prostate Cancer Tools for Diagnosis and Management**

ANNs are mathematical models that are based on biological neural networks and are composed of interconnected groups of artificial neurons. ANNs are used to map and predict outcomes in complex relationships between given 'inputs' and sought-after 'outputs' and can also be used to find patterns in datasets. In medicine, ANN applications have been used in cancer diagnosis, staging and recurrence prediction since the mid-1990s, when an enormous effort was initiated, especially in prostate cancer detection. The author states that modern ANNs can incorporate new biomarkers, serum markers and imaging data to improve their predictive power and can offer a number of advantages as clinical decision making tools, such as easy handling of distribution-free input parameters. Most importantly, ANNs consider non-linear relationships among input data that cannot always be recognized by conventional analyses. The author concludes that complex medical diagnostic and treatment decisions will be increasingly based on ANNs and other multivariate models in the near future [13].

### **4.4. Artificial Neural Network for Predicting Pathological Stage of Clinically Localized Prostate Cancer in a Taiwanese Population**

An ANN model was developed to predict prostate cancer staging in patient prior to when they received radical prostatectomy, as this is acclaimed to be more effective than logistic regression (LR), or the combined use of age, PSA and DRE. 299 patients undergoing retro-pubic radical prostatectomy or robotic assisted laparoscopic radical prostatectomy were subjected to evaluation. The results were meant to predict the pathological stage (T2 or T3) of the cancer. The predictive ability was compared with LR and validation of the 2007 partin tables was estimated by the areas under the ROC. Of the population, 36.45% displayed prostate cancer with extra capsular extension and 63.55% displayed organ confined disease. Logistic Regression analysis showed that only PSA and BMI were statistically significant predictors of prostate cancer with capsule invasion. ANN outperformed LR significantly  $0.795 \pm 0.023$  versus  $0.746 \pm 0.025$  respectively, with  $p = 0.016$ . In conclusion ANN was superior in predicting of prostate cancer and more accurate predictions for pathologic staging in comparison with the validation of current partin tables [14].

## **5. DISCUSSION**

Artificial neural networks (ANNs) are mathematical models that are based on the natural mammalian neural networks, and are composed of interconnected groups of artificial neurons. ANNs are used to map and predict outcomes in complex relationships among input data and output and can also be used to find patterns in datasets [13] which cannot always be recognized by conventional analyses. Like neural networks, ANNs contain nodes (sample points) of data that interact through carefully weighted connection lines. These artificial systems are trained and balanced by having previously fed data (training data), which is used in adjusting the networks' connections for the required purpose. In urologic medicine, early detection of prostate cancer is of pivotal importance, as it has great implication on the decisions to be taken.

The Prostate Specific Antigen (PSA) is a glycoprotein produced almost exclusively by the epithelial component of

the prostate gland. Men with prostatic disease may have high serum PSA levels because of excessive production of the antigen as well as architectural distortions in the gland that allow PSA greater access to the circulation [15]. Therefore, it is only logical that a test to determine the level of blood PSA be conducted in order to determine the possibility of carcinoma of the prostate. Most healthy men have a PSA level lower than 4.0 nanograms/milliliter (ng/ml), any rise in this is usually an indication of a cancerous presence. This however does not mean that men with a level below 4ng/ml on conducting the PSA test cannot have the cancer. Studies have in fact, show that around 15% of men with PSA < 4.0 will have cancer on biopsy. If PSA level is between 4 and 10, there is a 1 in 4 chance of having prostate cancer and a 50% chance in men with PSA greater than 10 on biopsy [3]. The information given to men that they have normal results on PSA testing may provide reassurance, although the possibility of a false negative result must be recognized leading to false expectations which might be crushed on biopsy. A false positive result has potential side effects; aggressive treatment rendered in the event of an abnormal test result and a subsequent diagnosis of cancer, the anxiety associated with a false positive result, and the burden of dealing with cancers that otherwise might never have become evident [15]. In early stage diagnosis, a needle biopsy is standard and is most accurate when done with ultrasound-guided sextant biopsy. The result of a PSA test will determine whether or not further diagnostic steps will be taken. It is however important to note that both prostate cancer and benign prostatic hyperplasia give an increase in blood PSA, hence the result of this test result is questionable and can lead to unnecessary biopsies. All this emphasize the need for a means of ensuring that the test results are not false positives or false negatives, thus the introduction of ANNs into serum PSA testing.

According to research done by various scientists, in a typical (feed-forward) ANN, there is an input layer, at least one hidden layer, and an output layer. In the case of predicting prostate biopsy outcome, the input units, for example, can be any number of these; PSA density, PSA transition zone, DRE, age, creatinine kinase, prostatic acid phosphatase, and percentage of free PSA. The dataset is randomly split into a training set and a validation set (data put aside to test the accuracy of the ANN after training). These inputs (free PSA, DRE, age) are then fed forward into the input layer, where their value is weighted to produce the desired outcome-in this case, a positive or negative biopsy result. The important point is that the outcome for the training data is known, and therefore, the neural net can be sequentially trained to achieve the perfect answer every time the training data are loaded into the system. The ability of the neural net to produce the correct answer when the outcome is theoretically unknown is then tested using the validation set. Inputs from the validation set are fed forward, and the ANN result is recorded and then compared with the known outcome from the validation set, and the two results are compared in a Receiver Operator Characteristic (ROC) curve. The area under the ROC curve is used as a measure of accuracy, with a value of 1.0 representing perfection and a value of 0.5, a 50% likelihood that the model will respond correctly [16]. The Artificial Neural Network approach has an added advantage because of its predictive accuracy. Studies have however shown that the use of complex nodes (input units); free PSA, PSA density, PSA volume, PCA-3 (Principal Component Analysis), E-PCA-2, other kallikreins and so on, increases predictive accuracy, the argument however remains over whether one or

more of this variables should be used to avoid further tissue sampling exercises [17].

## **6. FINDINGS AND PROPOSITIONS**

Percent free PSA (a ratio of how much PSA is roaming free in the blood compared to the total PSA) testing has been observed to improve the specificity of prostate cancer diagnosis. Because PSA is almost organ-specific but not cancer specific, there is an increased PSA concentration in patients with both BNH and prostate cancer. This has therefore brought about research on means of increasing the specificity of the total PSA in the blood for predictive accuracy in prostate cancer. There has been a 10-20% reduction in rate of unnecessary biopsies as a result of the use of %fPSA compared to other molecular forms or tPSA which have not experienced such clinical success [18]. It has however been purported that lower level of total PSA reduce diagnostic power of %fPSA which may explain the fact that accurate measurement of free PSA becomes difficult at low tPSA concentrations. The data of the above stated study is based on men of a tPSA < 10ng/ml. However, level of %free PSA decreases tPSA is > 10ng/ml, even though this level of concentration is already a clinical indicator of the need for a biopsy. In studies conducted, it has been observed that the accuracy of ANNs is better than the %fPSA; one of the reasons being the ability of the ANN to recognize complex non-linear relationships among input data that cannot always be recognized by conventional statistical analyses. It is important to note that so far, evaluating an ANN is only possible with limitations (group of patients it applies to). However, new possibilities to improve the accuracy of ANNs include the use of new, more complex serum markers such as kallikreins (a subgroup of serine proteases, enzymes capable of cleaving peptide bonds in protein), macrophage inhibitory cytokine 1 and pro-PSA.

Serum tumor markers are substances that are produced by cancer or by other cells in the body in response to cancer, mostly proteins [19]. Some of the serum markers identified by various researchers include but are not limited to; EPCA (-2) proteins early prostate cancer antigen; a highly specific serum marker [20], pro-PSA and kallikreins. Several of these serum markers have been proven theoretically, but the clinical values are yet to be ascertained. It can therefore be purported that the application of multiple complex serum markers together with other units such as age and race will give a greater accuracy than a mere %fPSA test and may even surpass the results of logistic regression. Using these serum markers as input data for ANN will yield a higher specificity and sensitivity than a mere PSA test.

## **7. RECOMMENDATIONS AND CONCLUSION**

In view of information from various studies, it is undisputable that the PSA test and DRE are not adequate enough to detect prostate cancer for reasons that have been earlier stated. However, the use of machine learning techniques will reduce the amount of unnecessary biopsies. Data gotten so far only proves that the level of accuracy of any of the ANN models proposed by various scientists is affected by limitations that apply only to a category of patients. It is therefore suggested that an assay specific ANN model be used to decide whether a biopsy is required. The assay specific ANN model to be used should make use of complex variables such as mentioned above as nodes in the ANN in order to improve specificity and sensitivity of the test result. It is advised that an ANN which can cover many assays should be developed to make it

easy for doctors to analyze a patient's outlook. However despite the apparent advantages of using ANNs, clinicians should not dispute medical facts to confirm the test results. In the case where a biopsy ends up negative, the doctor is advised to advice the patients on a re-biopsy to ensure that the result is not a false negative. It is also recommended that a baseline serum PSA level be obtained from men between 40 – 50 years of age and testing for early detection should be offered to men with life expectancy greater than 50 years. There should also be awareness on the issue of prostate cancer as well as advice on steps a patient should take so as to reduce mortality due to ignorance.

## **8. REFERENCES**

- [1]. Hruban, R. (2012). What are tumors? Retrieved 2015, from <http://pathology.jhu.edu/>
- [2]. National Cancer Institute. (2014). Prostate cancer.
- [3]. American cancer society. (2015). Prostate cancer prevention and early detection.
- [4]. Miller, D., Hafez , K., Stewart , A., Montie JE, J., & Wei , J. (2003). "Prostate carcinoma presentation, diagnosis, and staging: an update form the National Cancer Data Base".
- [5]. Hankey, B., Feuer, E., Clegg, L., Hayes , R., & Legler, J. (1999). Cancer surveillance series: interpreting trends in prostate cancer—part I: Evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates.
- [6]. Brawer, M., Bartsch, G., D'Amico, A., Donohue, R., Siam, O., & Tewari, A. (2003). early diagnosis and stging of prostate cancer.
- [7]. Vogelzang, N., & Shore, N. (2015). upfront chemotherapy in prostate cancer. Retrieved 2015, from cancer network: <http://www.cancernetwork.com/prostate-cancer/upfront-chemotherapy-prostate-cancer?>
- [8]. Porter, C., & Crawford, D. (2003). combining artificial neural networks and transurethral ultrasound in the diagnosis of prostate cancer.
- [9]. Revett, k., Sergio Tenreiro de Magalhaes, & Santos, H. (n.d.). Data Mining a Prostate Cancer Dataset Using.
- [10]. Raza, K., & Hasan, A. (n.d.). A comprehensive evaluation of machin learning technique for cancer class prediction based on microarray data.
- [11]. Snow, P., Smith, D., & Catalona , W. (1994). Artificial neural networks in the diagnosis and prognosis of prostate cancer: A pilot study.
- [12]. Djavan, B., Remzi, M., & Zlotta, A. (2002). Novel artificial neural network for early detection of prostate cancer.
- [13]. Meyer, H. (2013). Artificial neural networks and prostate cancer-tools for diagnosis and management.
- [14]. Tsao, C., Liu, C., Cha, T., Wu, S., & Sun, G. (2014). Artificial neural network for predicting pathological stage of clinically localized prostate cancer in a Taiwanese population. *Journal of the Chinese Medical Association* 77.
- [15]. Barry, M. (2001). Prostate specific antigen testing.

- [16].Schwarzzer, G., & Schumacher, M. (2002). Artificial neural networks for diagnosis and prognosis in prostate cancer.
- [17].Anagnostou, T. (2003). Artificial neural networks for decision making in urologic oncology.
- [18].Stephan, C., Jung, K., Lein, M., & Diamandis, E. (2007). PSA nad other tissue kalikreinsfor prostate cancer detection. *Eur J Cancer*.
- [19].National Cancer Institute. (2011). Tumor marker sheet. Retrieved from [www.cancer.gov: http://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-markers-fact-sheet](http://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-markers-fact-sheet)
- [20].Leman, E., Cannon, G., Trock, B., Sokoll, L., & Chan, D. (2007). EPCA -2 a highly specific serum marker for prostate cancer. Retrieved from [www.ncbi.nlm: http://www.ncbi.nlm/pubmed](http://www.ncbi.nlm/pubmed)