Combining Artificial Neural Networks and Transrectal Ultrasound in the Diagnosis of Prostate Cancer

Review Article [1] | October 01, 2003
By Christopher R. Porter, MD [2] and E. David Crawford, MD [3]

Arguably the most important step in the prognosis of prostate cancer is early diagnosis. More than 1 million transrectal ultrasound (TRUS)-guided prostate needle biopsies are performed annually in the United States, resulting in the detection of 200,000 new cases per year. Unfortunately, the urologist's ability to diagnose prostate cancer has not kept pace with therapeutic advances; currently, many men are facing the need for prostate biopsy with the likelihood that the result will be inconclusive. This paper will focus on the tools available to assist the clinician in predicting the outcome of the prostate needle biopsy. We will examine the use of "machine learning" models (artificial intelligence), in the form of artificial neural networks (ANNs), to predict prostate biopsy outcomes using prebiopsy variables. Currently, six validated predictive models are available. Of these, five are machine learning models, and one is based on logistic regression. The role of ANNs in providing valuable predictive models to be used in conjunction with TRUS appears promising. 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. With the introduction of more complex prebiopsy variables, ANNs are in a commanding position for use in predictive models. Easy and immediate physician access to these models will be imperative if their full potential is to be realized.

Traditional (gray-scale) transrectal ultrasound (TRUS) is the most widely used and possibly the most important imaging modality in the diagnosis of prostate cancer. Virtually all urologists, whether working in a tertiary care center or in community practice, have immediate access to an ultrasound unit. Arguably the most important step in the treatment of prostate cancer lies in its early diagnosis. More than 1 million TRUS-guided prostate needle biopsies are performed annually in the United States, resulting in the detection of 200,000 new cases per year.[1] Refinements in treatment modalities for prostate cancer including anatomic nerve-sparing radical prostatectomy, three-dimensional (3D) conformal radiation, and brachytherapy have improved patient outcomes. Unfortunately, the urologist's ability to diagnose prostate cancer has not kept pace with therapeutic advances. Currently, many men are facing the need for prostate biopsy with the better-than-average likelihood that the result will be inconclusive. In routine urologic practice, clinicians must decide whether to perform a prostate biopsy on the basis of a few parameters. Prostate-specific antigen (PSA) levels, the results of a digital rectal exam (DRE), and the patient's age are the parameters most often employed. The clinician's prior experience, a compilation of personal results and "rules of thumb" learned during training, may influence the decision to perform a biopsy. Unfortunately, personal predictions are often subject to inherent biases, weakened by an inability to memorize the complete dataset.[2] Predictive modeling tools are available to assist the clinician in the decision-making process. The most widely known of these is the Partin nomogram, which can be employed by the
Combining Artificial Neural Networks and Transrectal Ultrasound in the Diagnosis of Prostate Cancer

Published on Physicians Practice (http://www.physicianspractice.com)

physician during pretherapy discussions about prostate cancer with the patient to predict final pathologic stage of disease.[3] This paper will focus on the tools available to the clinician to assist in the prediction of the prostate needle biopsy outcome. We will examine the use of "machine learning" models (artificial intelligence) in the form of artificial neural networks (ANNs) using prebiopsy variables.

Early Diagnosis Dilemmas

Early detection of prostate cancer relies on the judgment of the physician coupled with the application of common clinical variables. The decision to perform an ultrasound-guided prostate biopsy rests on the assessment of appropriate clinical data, including the results of a DRE, PSA level, patient's age, and ultrasound findings. Unfortunately, when assessed individually, these variables have limited efficacy in terms of accurately guiding the physician and patient in the decision to undergo biopsy. For example, PSA is associated with high false-negative and false-positive rates, 20%-40% and 21%, respectively (positive predictive value: ~32%).[4] Similarly, DRE and overall clinical judgment are associated with low positive predictive values of 21% and 33%, respectively. That said, an estimated 25% of men undergoing prostate biopsy with a PSA level between 4 and 10 ng/mL are found to be harboring cancer.[5] Predicting the outcome of prostate needle biopsy based on a few clinical variables results in an increased risk of unnecessary biopsies. Additional prebiopsy markers are currently being evaluated, including PSA density, percent-free PSA, transition-zone PSA, PSA velocity, presence of prostatic intraepithelial neoplasia (PIN) and atypical acinar proliferation. The addition of multiple new markers, while potentially improving the prediction of biopsy outcomes, makes it considerably more difficult for the practicing urologist to accurately assess the vast array of clinical data and apply appropriate judgment.

Mathematical models and ANNs have been developed to assist the physician in assessing the risk of positive biopsy based on multiple parameters. Artificial Neural Networks ANNs are named after the natural mammalian neuron arrangement, in which neurons as specialized entities are interconnected, receiving signals propagated throughout the system. By receiving weighted signals from specialized systems, it is felt that the mammalian nervous system has the capacity to learn. A typical (feed-forward) ANN has an input layer, at least one hidden layer, and an output layer (Figure 1). In the case of predicting prostate biopsy outcome, the input units, for example, would be PSA, DRE, age, and percentage of free PSA. The dataset is randomly split into a training set (data used to teach the neural network) and a validation set (data put aside to test the accuracy of the ANN after training). These four inputs (PSA, free PSA, DRE, and age) are then "fed forward" into the hidden 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 where 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. This result is then compared to the...
known outcome from the validation set, and the two results are compared in a receiveroperator 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.[6] Standard statistical techniques (eg, logistic regression) rely on a linear relationship between the variable and the outcome. In biologic systems, such linearity often does not exist, and the ANN has the benefit of being able to capture complex nonlinear relationships by virtue of its architectural arrangement of neurons and their ability to weight the forward signal in a nonlinear fashion. The ANN model is, therefore, felt to have a potential advantage in terms of predictive accuracy.[2]

Validated Predictive Models

Currently, six validated predictive models have been published using prebiopsy parameters to predict prostate biopsy outcome (Table 1). Five of the six are ANN models, and one is based on logistic regression. Snow et al

In one of the first applications of ANNs to urologic oncology, Snow et al developed a model to predict biopsy outcome using data from 1,789 patients who were undergoing prostate cancer screening.[7] This model used the input variables of age, PSA, PSA velocity, and TRUS findings. The ANN model by Snow et al demonstrated a sensitivity of 0.7 and a specificity of 0.92, but unfortunately, these investigators did not report the model's accuracy as a ROC curve. The study was based on a retrospective screening population and validated against an independent patient cohort. Since the early ANN model of Snow et al, three groups have subsequently attempted to develop paradigms that predict prostate biopsy outcome in men with a low PSA level (2-4 ng/mL). Two of these have used the ANN approach. Babaian et al

Babaian et al developed an ANN model using PSA, creatinine kinase, prostatic acid phosphatase, and age to predict the likelihood of a positive biopsy.[8] This ANN was reported to have an ROC accuracy of 0.74 with appropriate validation. In this select cohort of patients (PSA: 2-4 ng/mL), it appears that this model would prevent almost 50% of unnecessary biopsies at a sensitivity of 92%. The authors found ROC accuracies of 0.74 and 0.75, respectively, for PSA density and PSA transition zone density. Thus, PSA density appears to have an accuracy equivalent to that of the ANN model used in the cohort of patients examined by Babaian et al. Djavan et al

For patients with PSA levels ranging from 2.5 to 4 ng/mL, Djavan et al similarly reported using an ANN based on PSA transition zone density, free PSA, PSA density, and prostate volume for 272 patients.[9] In this cohort, patients underwent sextant biopsies with two transition zone biopsies. The patient population comprised men in the European Prostate Cancer Detection Study, who had been referred to a urologist with lower urinary tract symptoms or for early detection of prostate cancer. It is unclear what proportion of men in this PSA range had abnormal DREs. The overall positive biopsy rate was 24%, and the ANN produced a validated ROC accuracy of 0.876. This model was compared to logistic regression models constructed using only 66% of the original data. The accuracy of the logistic regression model was 0.85. Both the Babaian and the Djavan models were constructed to evaluate patients with relatively low serum PSA values. Unfortunately, the application of these models to men in the United States presenting to a urologist for evaluation for prostate cancer may be difficult, as the majority of those with a PSA between 2.5 and 4 ng/mL have an abnormal DRE. Eastham et al

Eastham et al recently examined a similar patient cohort of men with PSA < 4 ng/mL and an abnormal DRE, and developed a logistic regression model to predict positive prostate biopsy.[10] In evaluating a diverse patient population using race, PSA, and age, Eastham et al reported an ROC accuracy of 0.75. In this distinct patient subset (PSA < 4 ng/mL), the three validated models cited above have accuracies ranging from 0.75 to 0.875. Although these results represent substantial improvements in accuracy over the use of single serum tests (eg, PSA), neither ANN model is available on the World Wide Web.
Predicting a Positive Biopsy in High-Risk Patients

The broader question of predicting a positive prostate biopsy in men presenting with either a PSA > 4 ng/mL or an abnormal DRE has been approached using ANN models. In the same report cited above, Djavan et al developed an ANN from a screening population of 974 men in the European Prostate Cancer Detection Study.[9] All men underwent sextant biopsy with two additional transition zone biopsies. If the first biopsy was negative for prostate cancer, a second identical biopsy was performed. The data from which the Djavan ANN model is developed therefore represents an extremely select group of patients: The ability of this model to predict outcome is limited to men with a PSA of 4 to 10 ng/mL who underwent repeat biopsy if the first biopsy was inconclusive. Within this paradigm, the ROC accuracy of the ANN model was 0.91 compared to the accuracy of a logistic regression model of 0.90. Although this represents a robust model, caution should be exercised in applying this result to other contemporary series. First, the model applies only to men who underwent repeat sextant biopsy; men in the United States usually do not automatically undergo an immediate repeat biopsy unless worrisome histologic markers are identified (eg, PIN). Moreover, current opinion favors 8- to 10-core biopsies with attention to the lateral most aspect of the prostate.[11] Second, models and nomograms used to predict outcome usually report a range of accuracies representing ROC results from ANN "cross-validation." Cross-validation refers to the splitting of the dataset several times into test sets and training sets. This allows assessment of the overall performance of the model. The performance of the model is then reported as a range, with average ROC accuracy noted. It is unclear from the Djavan ANN model whether cross-validation was performed.

ANN vs Logistic Regression

Methods

Recently, Porter et al published the results of their predictive models, based on both ANN and logistic regression methods, from a racially diverse prospective series of 319 patients.[12] All patients underwent a 10-core prostate needle biopsy with attention to the lateral aspect of the gland. The patient population represented men referred to the urologist with either an abnormal DRE or an elevated PSA (> 4 ng/mL; range: 0.8-367 ng/mL). Five-way cross-validation was performed, and the mean ROC accuracies of the ANN and logistic regression models were reported as 0.77 (range: 0.83-0.71) and 0.76 (range: 0.81-0.71), respectively. Although Djavan et al studied a larger patient population, the work by Porter et al may represent a more common clinical scenario. Thus, ANNs appear to be equivalent to their logistic regression counterparts in predicting prostate biopsy outcome, but no ANNs are currently available on the World Wide Web. In their report, however, Porter et al noted that they have scheduled the inclusion of an ANN on the Web at prostatecalculator.org. If ANNs are to be clinically useful, it is necessary for them to be easily accessible in either handheld computer versions or on the Web.

Advantages and Limitations of ANNs

In general, biologic systems are neither binary nor linear. Clinicians are faced with a constellation of parameters, many of which are not related to each other in a straightforward fashion. Traditional statistical methods cope with this variance by assigning cut-off points—for example, a PSA >10 ng/mL. This system has led to risk-group analysis that is easily memorized and simply applied. Artificial intelligence, in the form of machine learning or ANNs, has an advantage over traditional statistical methods in that the relationships between variables need not be linear. The ANN can theoretically learn, and therefore, weight the input variables so that the most efficient predictive
model is acquired. The ANNs listed in Table 1 are not perfect in predicting the outcome of prostate biopsy (nor, as a result, in predicting a diagnosis of prostate cancer). Their reported accuracies range from 0.75 to 0.91, and most are limited in their application to a distinct subset of patients. Nevertheless, it can be argued that a predictive accuracy of even 0.75 may be preferable to no model at all. Most studies demonstrate the superior accuracy of predictive models over human judgment,[13] although the outcome with respect to biopsy may be binary (i.e., the patient will have either a positive or negative result), counseling patients with regard to invasive testing can be assisted by relatively accurate predictive models. Predictive models, however, need to be improved. Improving the predictive accuracy of these models requires more extensive collection of data, which increases the sample size, and the application of more sophisticated modeling techniques. Increasing the sample size and, therefore, the data available to train the ANN, refers not only to the number of patients included in the database but also to the number and quality of the pretest parameters. Therefore, prospectively collected data with an emphasis on the collection of multiple variables is essential to the creation of accurate, clinically applicable predictive models. Wide-ranging clinical variables as well as serum markers (biomarkers) will likely enhance predictive accuracy. It is hoped that the addition of multiple variables, both clinical and serum based, will enhance the accuracy of models developed to predict the outcome of prostate biopsy. These models need to be able to prevent unnecessary biopsies while identifying all men with the disease. It could be argued, therefore, that the accuracy of models designed to predict the outcome of biopsy should be more accurate than models designed to predict final pathologic stage after surgery. After all, the fate of a man with organ-confined disease on one side of the prostate vs the same man with organ-confined disease on both sides of the gland is likely to be considerably different from that of a man who harbors a clinically significant cancer and fails to undergo biopsy on the "recommendation" of a predictive model. **Conclusions** In summary, the role of ANNs in providing valuable predictive models to be used in conjunction with TRUS appears promising. In the few studies that have compared ANNs to traditional logistic regression, both ANN and logistic regression appear to function equivalently when predicting the outcome of a biopsy.[14] With the introduction of more complex prebiopsy variables, ANNs appear to be in a commanding position for use in predictive models. Easy and immediate physician access to these models will be imperative if their full potential in predicting outcomes is to be realized.

**Disclosures:** The author(s) have no significant financial interest or other relationship with the manufacturers of any products or providers of any service mentioned in this article.

**References:**


Source URL:

Links: