0022-5347/00/1631-0300/0
THE JOURNAL OF UROLOGY®
Vol. 163, 300–305, January 2000
Printed in U.S.A.

PREDICTION OF BLADDER OUTLET OBSTRUCTION IN MEN WITH LOWER URINARY TRACT SYMPTOMS USING ARTIFICIAL NEURAL NETWORKS

GABE S. SONKE,* TOM HESKES, ANDRÉ L. M. VERBEEK, JEAN J. M. C. H. DE LA ROSETTE AND LAMBERTUS A. L. M. KIEMENEY

From the Department of Epidemiology, the Department of Urology, and the Department of Medical Physics and Biophysics, University of Nijmegen, The Netherlands

ABSTRACT

Purpose: To evaluate the performance of a backpropagation artificial neural network (ANN) in the diagnosis of men with lower urinary tract symptoms (LUTS) and to compare its performance to that of a traditional linear regression model.

Materials and Methods: 1903 LUTS patients referred to the University Hospital Nijmegen between 1992 and 1998 received routine investigation, consisting of transrectal ultrasonography of the prostate, serum PSA measurement, assessment of symptoms and quality of life by the International Prostate Symptom Score (IPSS), urinary flowmetry with determination of maximum flow rate (Qmax), voided volume and post-void residual urine and full pressure flow studies (PFS). Using a three-layered backpropagation ANN with three hidden nodes, the outcome of PFS, quantified by the Abrams-Griffiths number (AG-number), was estimated based on all available non-invasive diagnostic test results plus patient age. The performance of the network was quantified using sensitivity, specificity and the area under the ROC-curve (AUC). The results of the neural network approach were compared to those of a linear regression analysis.

Results: Prostate volume, Qmax, voided volume and post void residual urine showed substantial predictive value concerning the outcome of PFS. Patient age, PSA-level, IPSS and Quality of life did not add to that prediction. Using a cut-off value in predicted and true AG-numbers of 40 cm. H_2O, the neural network approach yielded sensitivity and specificity of 71% and 69%, respectively. The AUC of the network was 0.75 (standard error = 0.01). A linear regression model produced identical results.

Conclusions: This study shows that at an individual level, the outcome of PFS cannot be predicted accurately by the available non-invasive tests. The use of ANNs, which are better able than traditional regression models to identify non-linear relations and complex interactions between variables, did not improve the prediction of BOO. Thus, if precise urodynamic information is considered important in the diagnosis of men with LUTS, PFS must be carried out. Both neural networks and regression analysis appear promising to identify patients who should undergo PFS, and those in whom PFS can safely be omitted. Furthermore, the ability of ANNs and regression models to predict treatment result should be evaluated.

KEY WORDS: neural networks, prostatic hyperplasia, discriminant analysis, diagnosis

Lower urinary tract symptoms (LUTS) are a major cause of discomfort for many elderly men. Symptoms include poluria, nocturia, weak stream and dribbling. Many of these patients suffer from bladder outlet obstruction (BOO) due to benign prostatic hyperplasia (BPH). A considerable group of patients, however, is hampered by detrusor instability or hypocontractility instead of BOO. Pressure-flow studies (PFS) are the best available method to distinguish BOO from bladder pathology.[1] However, PFS are often considered too invasive, time-consuming and expensive to be routinely utilized.[2,3] It will therefore be meaningful to predict the outcome from PFS without using PFS itself.

Previous studies used traditional regression models to combine the information from several non-invasive diagnostic tests, such as free uroflowmetry, prostatic volume measurement and symptom-score lists, to estimate PFS outcome.[4,5] Regression models have difficulties in identifying non-linear relations and complex interactions between various tests. Therefore, these studies may not have yielded optimal predictions. Artificial neural networks (ANNs) are a relative new mathematical technique and are well able to identify complex associations between variables. We studied the performance of an ANN in predicting the outcome of PFS in men with LUTS.

MATERIALS AND METHODS

We evaluated men with LUTS, suggestive of BOO or bladder dysfunction, who were referred by their general practitioner or by urologists from community hospitals, to the Nijmegen University Hospital between January 1992 and September 1998. All patients received routine investigation, consisting of medical history taking, digital rectal examination and transrectal ultrasonography of the prostate, measurement of serum PSA-level, assessment of symptoms and quality of life as listed on the International Prostate Symptom Score (IPSS), urinary flowmetry including determination of voided and post-void residual volume (PVR) and full urodynamic investigations. Patients who were unable to void

Accepted for publication June 14, 1999.
* Requests for reprints: Department of Epidemiology and Urology, University Hospital Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands.

during PFS were excluded, as were patients who previously underwent prostatic treatment. Missing values were replaced using maximum likelihood estimation.[6]

As recommended by the International Continence Society, the Abrams-Griffiths number (AG-number), which is equal to [detrusor pressure at Qmax − 2 * Qmax], was calculated to quantify the level of urethral obstruction.[7] AG-numbers ≥ 40 cm. H_2O classify patients as obstructed. Patients with AG-numbers < 40 cm. H_2O are either equivocal or not obstructed. All other abovementioned diagnostic test results plus age of the patient were assessed as possible predictors of AG-number.

Traditional regression models, which can discriminate between diseased and non-diseased persons, require the shape of the relation between variables (diagnostic tests) to be entered in the model. These relations may for instance be linear, quadratic or logistic. Thus, the investigator must learn the shape of such relations. A visual approach is usually suitable if less than three different diagnostic tests are involved. In more complex situations, as are often present in medical problems, a visual approach is difficult and linear relations are frequently assumed. This may yield imperfect results. Neural networks provide a framework for automatically selecting the appropriate relations between variables.

In medical research, the most commonly applied ANNs use a backpropagation learning algorithm.[8] This type of network consists of several interconnected nodes, positioned in layers (figure 1). There is one input layer, consisting of all diagnostic test results, one output layer representing the gold standard variable (PFS), and typically one or more hidden layers used for computational purposes. The investigator determines the number of hidden layers as well as the number of nodes per hidden layer. The value of each input node is multiplied by a weight factor and sent through to the adjacent nodes. Weights, analogous to biological synapses, are numerical values representing the strength of connections between nodes (neurons). Each node computes a weighted sum of all received input and converts it by a free-to-choose, usually non-linear, transformation function. The value of this function is sent through to the next layer if it exceeds a certain threshold. This resembles a biological neuron, which fires an action potential if enough neurotransmitter from adjacent synaptic junctions has accumulated. The predicted outcome of the network for a particular patient consists of the weighted and transformed input to the output layer.

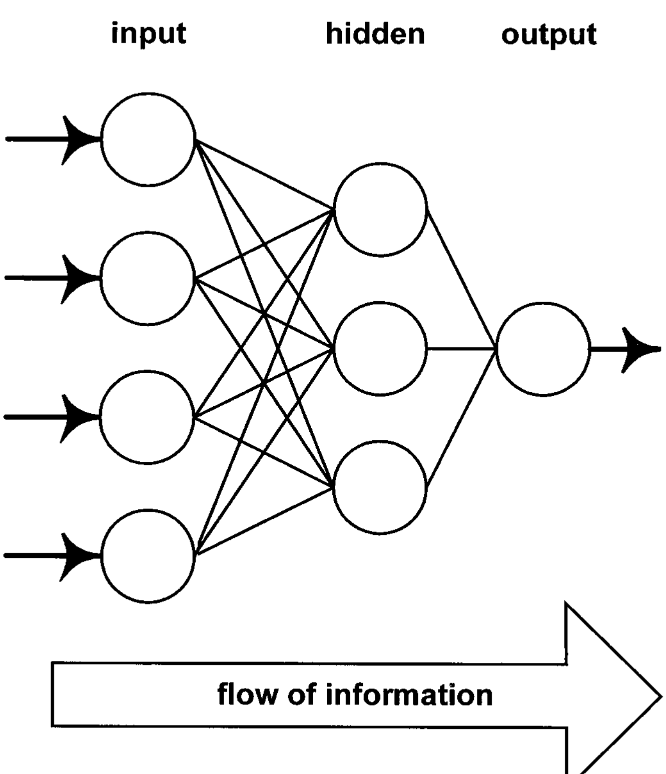


FIG. 1. Lay-out of three layered neural network. Input nodes are all relevant diagnostic test results; hidden nodes are used for calculation purposes; output node represents gold standard diagnostic test result.

When constructing a neural network, the complete dataset is split up into three different sets: a training set, a validation set and a test set. Subdivisions between training and validation sets can be generated using bootstrapping.[9] A bootstrap sample is a collection of n patients drawn with replacement from all patients, with n the total number of patients. The bootstrap sample is taken to be the training set, which is used to estimate the weights of the connections. All weights are initially designated small random numbers. Following, the information of each patient is repeatedly presented to the network, and the predicted outcome is compared to the true (gold standard) outcome. The difference between predicted and true outcome (the error) is propagated back to the network and the weights are adjusted to minimize the mean squared error (MSE) over all patients in the training set.

Patients that do not occur in the training set constitute the validation set. This set prevents the network from overfitting the data. Using the weights obtained from the training set, a prediction is also made for the patients in the validation set. If the MSE for these patients starts to increase, the network is most likely overfitting the data in the training set and training is therefore stopped at this stage. As the resulting network depends on the random subdivision in training and validation set, this process is repeated 100 times, yielding 100 different networks.

Even before subdividing into training and validation set, a random sample of the original data is set aside to be used as a test set. The weights of the training set producing the smallest MSE in this test set are eventually used to predict AG-numbers for all available patients.

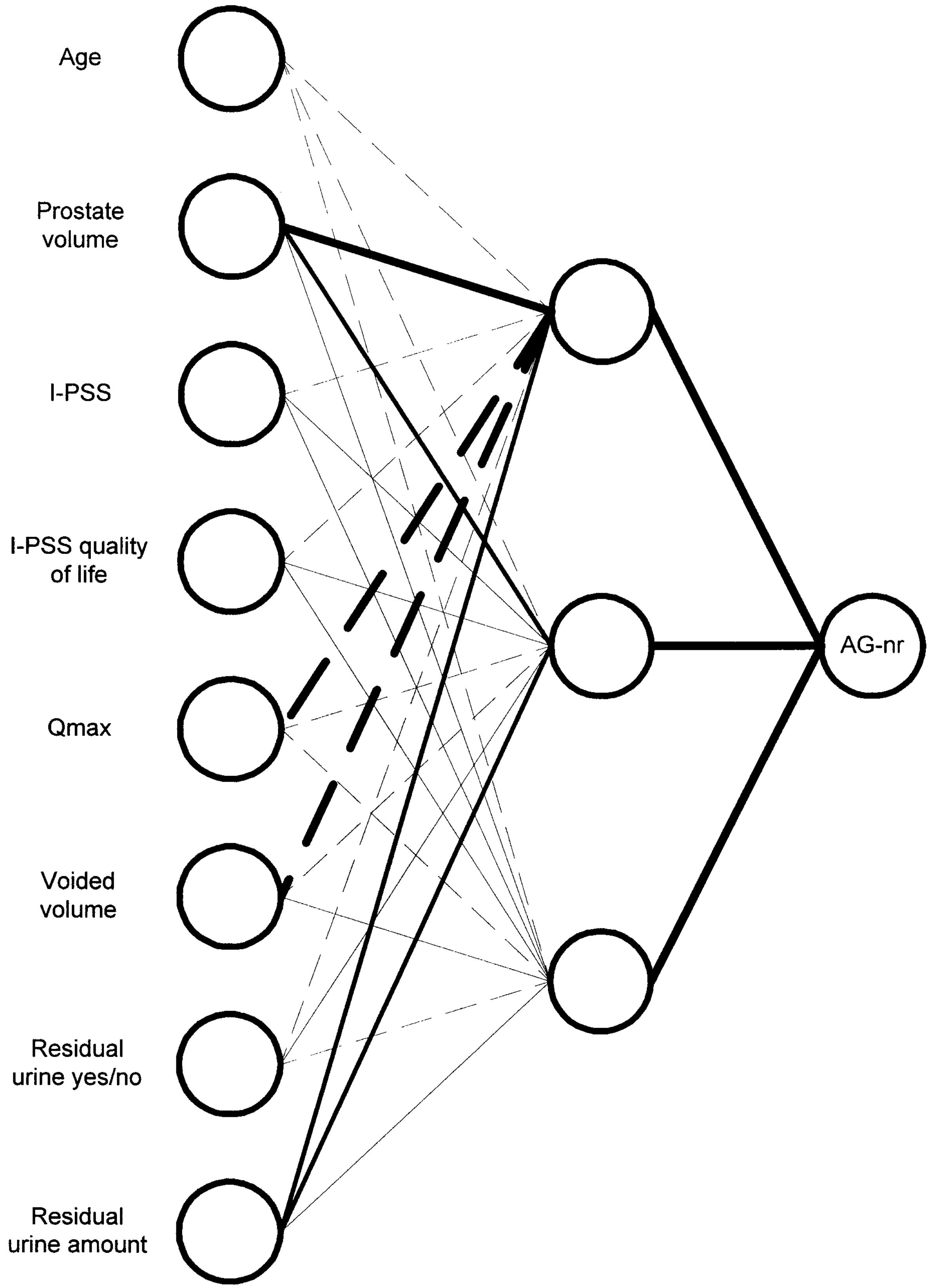
Since ANNs perform better on variables that are normally distributed, the diagnostic test results were transformed when necessary, using a natural logarithm or square root function and normalized with zero mean and standard deviation of one. The post-void residual was divided into two variables: one indicating if there was any residual urine and another indicating the actual amount of residual.

Using the neural network, we predicted AG-numbers for all patients. These predicted numbers were then regarded as diagnostic test results with scores ≥40 indicating obstruction and compared to the true AG-numbers. Sensitivity and specificity of this diagnostic indicator were calculated. Analogously, we predicted AG-numbers using a traditional linear regression model, dichotomized the predicted numbers using a cut-off value of 40 and again calculated sensitivity and specificity. In addition, since sensitivity and specificity depend on the chosen cut-off value, we constructed ROC-curves

TABLE 1. *Mean and standard deviation (unless otherwise specified) of patient characteristics*

	(n = 1903)	
Age (years)	64.1	(9.4)
Qmax free flow (ml./s)	10.7	(5.5)
Voided volume (ml.)	249.4	(149.7)
Prostate size (cm.^{3})	43.2	(23.0)
Serum PSA (ng./ml.)*	2.3	(1.1–4.7)
Post void residual (ml.)*	20	(0–75)
IPSS-total	16.7	(7.2)
IPSS-Quality of life*	4	(3–4)
Abrams-Griffiths number	40.2	(29.6)
Obstruction†(no:equivocal:severe) (%)	26:30:44	

* Median value (25th–75th centile).
† Based on the International Continence Society provisional nomogram.[7]



AG-nr indicates Abrams-Griffiths number
Thickness of each line represents the magnitude of the contribution to the predicted outcome of each diagnostic test. Solid lines represent positive weights, dashed lines are negative.

FIG. 2. Lay-out of neural network to predict urodynamic obstruction in LUTS patients

TABLE 2. *Stratification of patients according to neural network prediction and pressure-flow outcome*

Neural Network Prediction	PFS-Outcome		Total
	Obstructed (AG-Number ≥40)	Equivocal/Not Obstructed (AG-Number < 40)	
≥40	604	328	932
<40	244	727	971
Total	848	1055	1903

AG-number indicates Abrams-Griffiths number.
PFS indicates pressure-flow study.
Sensitivity = 819/1411 = 58%.
Specificity = 379/492 = 77%.

and estimated the associated areas under the curve (AUC).[10] Using ROC-curves and the AUC, the value of a test can be assessed without having to choose cut-off values in a continuous diagnostic test result. A test or test combination with no discriminatory power will have an AUC of 0.5, while a perfect test will have an AUC of 1.0.

RESULTS

This study investigated 1903 patients with LUTS. Table 1 shows a summary of patient characteristics. Fifteen hundred randomly selected patients were used to train and validate the network, while the remaining 403 were used as a test set.

Figure 2 shows the layout of the neural network, with the magnitude of the contribution of each diagnostic test represented by the thickness of the connection. Solid lines represent positive weights while dashed lines are negative. Prostate volume, Qmax and PVR (actual amount) and voided volume proved to contribute substantially to the network, while patient age, PSA-level, IPSS and quality of life score seemed less relevant. Using cut-off values in both predicted and true AG-number of 40 cm. H_2O, sensitivity and specificity of the network were 71% and 69% respectively (table 2). A linear regression model produced identical sensitivity and specificity. Figure 3 shows the ROC-curves for the neural network and the linear regression model. From both ROC-curves an AUC of 0.78 (standard error = 0.01) was estimated.

The results of the neural network and the regression analysis can also be used to preclude unnecessary PFS: figure 4 shows the distribution of predicted AG-numbers for truly obstructed and truly unobstructed patients. If, for instance, all patients with predicted values below 25 (22%) are regarded unobstructed and patients with predicted values above 55 (22%) are considered obstructed, the total number of urodynamic investigations is reduced by 44%. This diagnostic protocol yields an overall sensitivity of 94% with specificity of 91%. By moving up and down the cut-off values (25 and 55 in this example), one can easily exchange the number of PFS performed for certainty in diagnosis.

DISCUSSION

In this study, the use of an ANN did not improve the prediction of the outcome of pressure-flow studies compared to a linear regression model. Neither of these models can replace invasive urodynamics. Both models, however, appear promising to identify patients who should undergo PFS, and those in whom PFS can safely be omitted.

PFS combine information on detrusor pressure and urinary flow rate to determine urethral resistance. Urinary flow rate is measured non-invasively, but detrusor pressure is not, although non-invasive methods are being developed.[11,12] The correlation between currently used diagnostic tests (prostatic volume, IPSS, PVR) and detrusor pressure is poor, even when assessed with ANNs. Thus, at this moment accurate non-invasive prediction of PFS outcome is not possible.

A number of drawbacks must be kept in mind when using ANNs. Firstly, the way a network should ideally be designed and trained is debated: The investigator must choose the

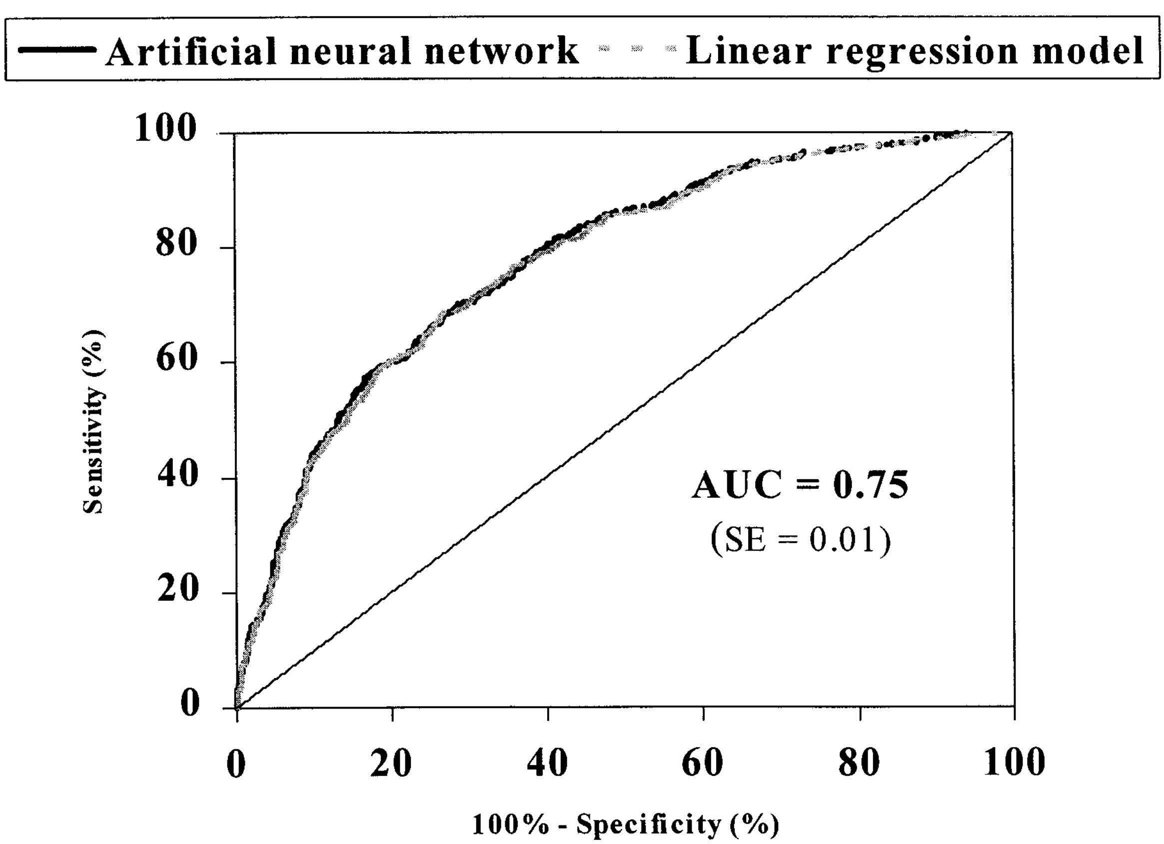


FIG. 3. ROC curves for predicting bladder outlet obstruction in LUTS patients

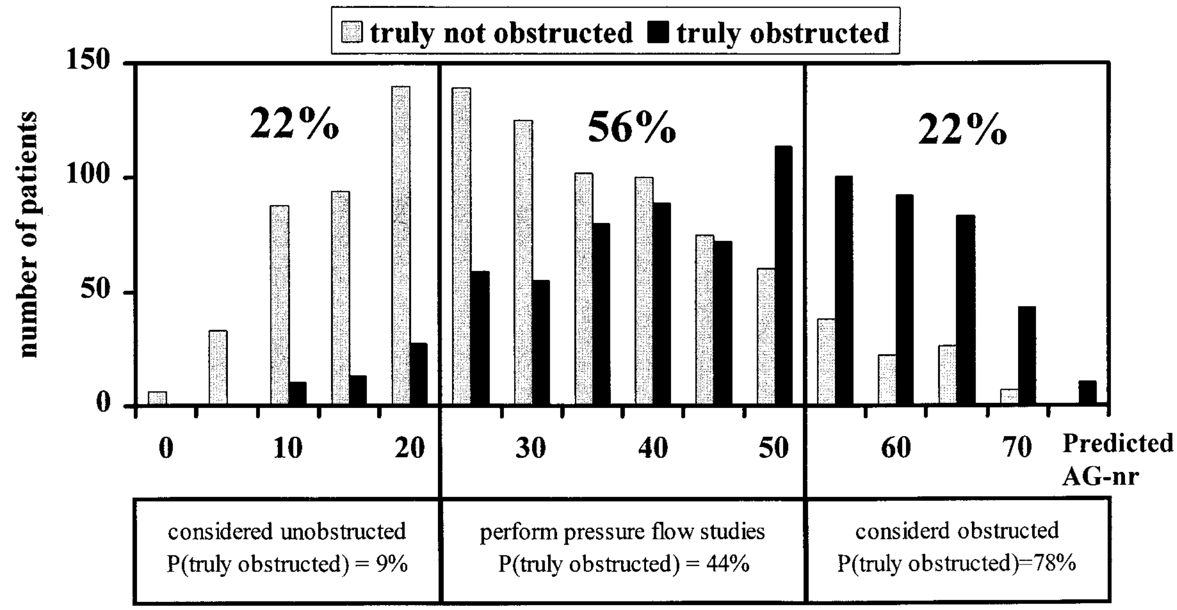


FIG. 4. Histogram of predicted AG-number, in obstructed and equivocal/not obstructed patients

number of hidden layers, the number of nodes per layer, the way to optimize the connecting weights and the shape of the transfer function. These choices may all influence the performance of the network. More research in this field is needed. Furthermore, new types of artificial intelligence are developed, among which so-called genetic adaptive models have shown promising results.[13] Secondly, in noisy problems (much unexplainable variance) neural networks have a tendency to restrict themselves to linear relations. As noisy problems are abundant in medical research, the performance of a neural network will often be equal to that of linear techniques. Thirdly, ANNs provide little understanding of the relations between variables; they merely predict outcome measures. Whereas the parameters in traditional regression models have clear interpretations, at this moment the weights of ANNs have not, unless very specific transfer functions are used.[14]

Despite these drawbacks and despite the fact that ANNs did not show additional value over traditional regression analysis in diagnosing men with LUTS, this relative new technique can find its place in medical science. The ability to detect non-linear relations and complex interactions between variables impose a large potential for using ANNs. ANNs have already been evaluated and found useful in diagnosing myocardial infarction,[15] predicting cancer survival[8] and evaluating pathology specimens[16] among many other applications. In addition, ANNs have recently appeared in urologic literature.[13,17,18,19]

Three previous studies assessed the correlation between PFS outcome and a combination of non-invasive tests.[4,5,20] Rosier et al constructed a clinical prostate score that correlated with PFS outcome better than any individual variable. Using logistic regression, Van Venrooij et al and Madersbacher et al calculated algorithms, which correlated moderately with degree of obstruction. In agreement with our findings, all three studies provided high degree of certainty regarding the presence or absence of obstruction in extreme categories (low Qmax, large prostate, high PVR vs. high Qmax, small prostate, no PVR). PFS may be safely omitted in patients in these two categories. The large intermediate group of patients, however, remained difficult to diagnose. Since accurate predictions of PFS outcome are not possible, it is important to minimize the number of patients in this intermediate group.

Performing PFS or predicting their outcome, as part of the diagnostic assessment of LUTS patients, aims at selecting patients for specific interventions, which may change the natural history of the disease. Ideally, a certain diagnosis explicitly points to a specific intervention. The disease that the urologist wants to demonstrate in LUTS patients is BOO. If BOO is diagnosed by means of PFS, transurethral resection of the prostate or open prostatectomy are the most effective treatments. Nevertheless, even in urodynamically obstructed patients, treatment failure occurs. Next to difficulties during operation, this may be caused by incorrect diagnosis, even if PFS are performed. Natural variation in urethral resistance exists between healthy men. Thus, regardless of any lower urinary tract pathology, some men will have relative low urethral resistance while others have high. If men with an initial high urethral resistance develop LUTS, PFS will invariably classify them as obstructed even if in fact the main problem is situated in the bladder. Although high urethral resistance is likely to be reduced by removing or reducing the size of the prostate, in these latter cases desobstructive treatment will probably be unsuccessful. Only if an increase in urethral resistance coincided with the development of LUTS, desobstructive treatment is likely to be successful. In order to be informed about the changes that occurred in urethral resistance, information on the lower urinary tract, prior to the onset of symptoms, is required. As such information is seldom available, we settle for current level of urethral resistance, and treatment decisions are based on this 'silver' standard.

Considering that PFS not always correctly predict treatment result, the assessment of LUTS patients should not focus on establishing the presence of high resistance. Instead, predicting the result of an intervention should be aimed at.[21] A patient classification based on several 'diagnostic' tests, possibly including PFS, with clear prognostic interpretation should be established in longitudinal studies.[22] It can then be clarified in which patients PFS improve the prediction of treatment success and in which patients PFS can be omitted. The use of ANNs should also be explored in such prognostic-intervention research.

CONCLUSION

The use of artificial neural networks does not improve the prediction of PFS outcome in men with LUTS based on non-invasive test results compared to traditional regression models. Prostatic volume, Qmax, voided volume and PVR help to predict the outcome of PFS, while patient age, PSA-level, IPSS and quality of life score do not. If exact prediction of PFS is considered important, other ways to measure or to predict detrusor pressure must be further developed. However, using prediction models, the number of patients having to undergo PFS can be greatly reduced. Furthermore, predicting treatment result rather than focusing on diagnosis may be a more feasible methodology. In such prognostic-intervention research as well as in diagnostic research, the value of ANNs should be further explored.

REFERENCES

1. Abrams, P.: Objective evaluation of bladder outlet obstruction. Br J Urol, **76(Suppl. 1):** 11, 1995.
2. Barry, M. J., Fowler, F. J. Jr., Bin, L. and Oesterling, J. E.: A nationwide survey of practicing urologists: current management of benign prostatic hyperplasia and clinically localized prostate cancer. J Urol, **158:** 488, 1997.
3. McConnell, J. D.: Benign prostatic hyperplasia: treatment guidelines and patient classification. Br J Urol, **76(Suppl. 1):** 29, 1995.
4. Madersbacher, S., Klingler, H. C., Djavan, B., Stulnig, T., Schatzl, G., Schmidbauer, C. P. and Marberger, M.: Is obstruction predictable by clinical evaluation in patients with lower urinary tract symptoms? Br J Urol, **80:** 72, 1997.
5. Van Venrooij, G. E. and Boon, T. A.: The value of symptom score, quality of life score, maximal urinary flow rate, residual volume and prostatic size for the diagnosis of obstructive benign prostatic hyperplasia: a urodynamic analysis. J Urol, **155:** 2014, 1996.
6. Arbuckle, J. L.: Full information estimation in the presence of incomplete data. In: Advanced Structural Equation Modeling Techniques. Edited by G. A. Marcoulides, R. E. Schumacker. Mahwah: LEA, 1996.
7. Griffiths, D., Höfner, K., van Mastrigt, R., Rollema, H. J., Spangberg, A. and Gleason, D.: Standardisation of terminology of lower urinary tract function: pressure-flow studies of voiding, urethral resistance and urethral obstruction. Neurourol Urodyn, **16:** 1, 1997.
8. Burke, H. B.: Artificial neural networks for cancer research: outcome prediction. Semin Surg Oncol, **10:** 73, 1994.
9. Heskes, T.: Balancing between bagging and bumping. In: Advances in Neural Information Processing Systems 9. Edited by M. Mozer, M. Jordan and T. Petsche. Cambridge: MIT Press, pp. 466–72, 1997.
10. Hanley, J. A. and McNeil, B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, **142:** 29, 1982.
11. Griffin, J. H., Walter, J. S., Schister, T. G., Schlehahn, L., Scarpine, V., Damaser, M. S. and Wheeler, J. S. Jr.: Preliminary noninvasive back-pressure recording of bladder pressure. Tech Urol, **2:** 108, 1996.
12. Gleason, D. M., Bottaccini, M. R. and McRae, L. P.: Noninvasive urodynamics: a study of male voiding dysfunction. Neurourol Urodyn, **16:** 93, 1997.
13. Tewari, A. and Narayan, P.: Novel staging tool for localized prostate cancer: a pilot study using genetic adaptive neural networks. J Urol, **160:** 430, 1998.
14. Duh, M. S., Walker, A. M. and Ayanian, J. Z.: Epidemiologic interpretation of artificial neural networks. Am J Epidemiol, **147:** 1112, 1998.
15. Baxt, W. G.: Analysis of the clinical variables driving decision in an artificial neural network trained to identify the presence of myocardial infarction. Ann Emerg Med, **21:** 1439, 1992.
16. O'Leary, T. J., Mikel, U. V. and Becker, R. L.: Computer-assisted image interpretation: use of a neural network to differentiate tubular carcinoma from sclerosing adenosis. Mod Pathol, **5:** 402, 1992.
17. Tewari, A.: Artificial intelligence and neural networks: concept, applications and future in urology. Br J Urol, **80(Suppl. 3):** 53, 1997.
18. Pantazopoulos, D., Karakitsos, P., Iokom-Liossi, A., Pouliakis, A. and Dimopoulos, K.: Comparing neural networks in the discrimination of benign from malignant lower urinary tract lesions. Br J Urol, **81:** 574, 1998.
19. Wei, J. T., Zhang, Z., Barnhill, S. D., Madyastha, K. R., Zhang, H. and Oesterling, J. E.: Understanding artificial neural networks and exploring their potential applications for the practicing urologist. Urology, **52:** 161, 1998.
20. Rosier, P. F., de Wildt, M. J., Wijkstra, H., Debruyne, F. M. and de la Rosette, J. J.: Clinical diagnosis of bladder outlet obstruction in patients with benign prostatic enlargement and lower urinary tract symptoms. J Urol, **155:** 1649, 1996.
21. Sonke, G. S. and Kiemeney, L. A.: Diagnostic research in benign prostatic hyperplasia: from sensitivity to neural networks. Curr Opin Urol **9:** 31, 1999.
22. Vallancien, G.: The need for an international classification of benign prostatic hyperplasia. Eur Urol, **33:** 248, 1998.