Experiments with Neural Networks as Prognostics Engines for Patient Physiological System Health Management

Peter K.Ghavami[1], Kailash C. Kapur[2]

[1,2]Department of Industrial & Systems Engineering, University of Washington, Seattle, WA 98195, USA

pghavami@uw.edu
kkapur@uw.edu

ABSTRACT

Prognostics and prediction of patients' short term physiological health status is of critical importance in medicine because it affords medical interventions that prevent escalating medical complications. Accurate prediction of the patients' health status offers many benefits including faster recovery, lower medical costs and better clinical outcomes. This study proposes a prognostics engine to predict patient physiological status. The prognostics engine builds models from historical clinical data using neural network as its computational kernel. This study compared accuracy of various neural network models. Given the diversity of clinical data and disease conditions, no single model is ideal for all medical cases. Certain algorithms are more accurate than others depending on the type, amount and diversity of possible outcomes. Thus multiple neural network algorithms are necessary to build a generalizable prognostics engine. The study proposes using an oracle, an overseer program to select the most accurate predictive model that is most suited for a particular medical prediction among several neural network options.

1. INTRODUCTION

Prognostics and Health Management (PHM) is an engineering discipline that links studies of failure mechanisms to system lifecycle management (Uckun, Goebel, & Lucas, 2008). Other definitions of PHM describe it as a method that permits the assessment of the reliability of a system under its actual application conditions, to determine the advent of failure, and mitigate system risks (Pecht, 2008). A system can be broadly defined as an integrated set of elements that accomplish a defined objective (International Council on Systems Engineering Systems Engineering Handbook, 2000). The human body is a biological system that functions as a collection of interrelated systems. The question that we wish to answer is how PHM can be applied to human biological systems as a methodology to predict and prevent adverse medical conditions in patients.

The term "diagnostics" pertains to the detection and isolation of faults or failures. "Prognostics" is the process of predicting a future state (of reliability) based on current and historic conditions (Vichare & Pecht, 2008).

The emphasis of this study is on prognostics (prediction) of the individual's short term future health condition and a rule-based prognostics engine that makes such predictions possible. Short term is defined as a time frame that spans from a few seconds to several days from any given moment. The prognostics engine is a computational component that can analyze vast amounts of historical and current physiological data and predict future health of an individual. Predictions are continuous over time as new, real time data are gathered from multiple physiological systems including warnings, alerts, events and precautions.

Admittedly developing mathematical models that make accurate predictions in biology and medicine is challenging but researchers suggest that soon such mathematical models will become a useful adjunct to laboratory experiment (and even clinical trials), and the provision of 'in silico' models will become routine (Smye & Clayton, 2002).

Advances in vital-signs monitoring software/hardware, sensor technology, miniaturization, wireless technology and storage allow recording and analysis of large physiological data in a timely fashion (Yu, Liu, McKenna, Reisner, Reifman, 2006). This provides both a challenge and an opportunity. The challenge is that the medical decision maker must sift through vast amount of data to make the appropriate treatment plan decisions. The opportunity is to analyze this large amount of data in real time to provide forecasts about the near term health state of the patient and assist with clinical decisions.

The application of PHM to human systems promises to deliver several benefits such as:

- Continuously assess the physiological and biological predictions to provide advance warning of clinical complications
- Minimize the frequency of reactive and emergency medical response by predicting and applying preventative medical interventions
- Enhance the quality of life and improve remaining useful life (RUL) of patients
- Manage the patient healthcare life cycle more effectively to improve patient care outcome and reduce medical care costs

Continuous and periodic monitoring of the individual's physiological systems involve collecting data from more than ten distinct human physiological sub-systems ranging from circulatory to respiratory to immune system. The collected data is used by the prognostics engine to predict future health of the individual.

The input data may consist of a wide array of clinically relevant information, medical history, allergies, medications, clinical procedures, genetic disposition and current physiological monitored data.

Medical science is grounded in scientific evidence, prior research, experiments and studies that have produced a body of medical knowledge based on generalizations and meta-analysis of research data. Such generalizations explain the causal relationships between risk factors, diseases and diagnosis. There are however gray areas in medical prognostics where many health treatment and screening decisions have no single 'best' choice or because there is scientific uncertainty about causes of certain diseases, or the clinical evidence is insufficient (O'Connor, Bennett, Stacey, Barry, Col, Eden, Entwistle & Fiset, 2009).

In many areas of medical science, the causal relationships are still incompletely understood and controversial. There are environmental, situational, cultural and unique factors that provide unique clinical data about a disease or groups of patients. Although this data is inadequate for making scientific generalizations and clinical evidence, it can provide valuable information to make assessments of individual's health status. The research hypothesis is that such data can be employed to make early predictions about the future health status of individuals and allow doctors apply medical interventions that prevent diseases or adverse medical events.

2. MARKERS AND PREDICTORS: THE CANARY AND THE DOG

The use of canaries for predicting failures has been discussed in literature as an illustrative example of prognostics (Vichare & Pecht , 2006). The analogy comes from an old mining practice. Canaries are more susceptible to dangerous gases than humans. Since gases are not easily detected by humans, miners carried canaries to mines as an early indicator of dangerous gases. When the canary died, it signaled presence of dangerous gases and miners got out. The canary is an example of what in medicine is referred to as a marker (Souter, 2011).

For years, dogs have been trained to assist patients for medical purposes. Studies have shown that medical response dogs can be trained to predict and alert their owners of seizures before they occur (Brown & Strong, 1999, 2001, and Kirton, Winter, Wirrell & Snead, 2008). Other anecdotal studies claim that certain response dogs have been able to detect presence of melanoma cancer (Williams & Pembroke, 1989). A dog's ability to alert its owner of a pending seizure is an example of a prediction in medicine.

A prediction is a form of speculation about a state in the future. A *prediction* is foretelling a medical event or disease when the ingredients for that medical event are in place but have not combined to affect their significance in form of a disease yet. *Predictors* are variables that offer predictions about a disease. A *marker* is the recognition that the ingredients for a medical event are in place and have indeed combined to result in form of a disease but in lower and milder yet measurable doses (Souter, 2011).

The precursor to a disease is known as risk factors in medicine. Thus, a timeline of medical predictions starts with risk factors, leading to predictors (pre-disease state), and then on to markers (disease is in place but in low, mild state) and finally to the occurrence (onset) of the disease or medical event itself. Figure 1 illustrates the chronology of events and progression of the individual's health status from risk factors leading to the final disease or medical event manifestation. The distance between time ticks are arbitrary and vary between individuals.

Traditional medical prediction models rely on risk factors to make crude speculations about the patient's future health status. This research attempts to predict medical health problems in a more accurate and timely manner using real time physiological measurements collected from patients.

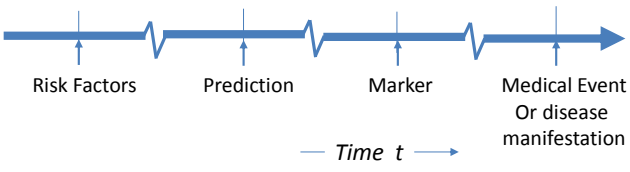

Figure 1. The progression of patient health condition

3. THE PROPOSED MODEL

The model proposed by this research considers medical treatment plan as input to the patient's physiological system.

Represented by $u(t)$, medical treatment plan involves some set of medications, procedures and care protocols prescribed by the physician. The patient's physiology is the process that produces a clinical outcome at time t, represented by $y(t)$. The patient's clinical outcome is the output or the response variable. The outcome is a vector of a single or multiple states of health for that patient. The input and response variable can be shown as:

$$U(t) = (u_1(t), u_2(t), \ldots, u_q(t)) \quad (1)$$

$$Y(t) = (y_1(t), y_2(t), \ldots, y_m(t)) \quad (2)$$

The internal physiological measurements consisting of clinical and vital sign data such as lab results and monitored data are represented by $x(t)$:

$$X(t) = (x_1(t), x_2(t), \ldots, x_k(t)) \quad (3)$$

The model includes a prognostics engine that consists of prediction rules R, and uses specific model M to predict specific outcome for time $(t + t_1)$. The prognostics engine collects vital clinical data from the patient's physiological system and makes a prediction for t_1 minutes in advance, for time $(t + t_1)$. The prognostics engine delivers a prediction that can be used to modify the medical treatment plan $u(t)$. The prediction rules are based on prior evidence and formed from retrospective collection of past patient data. The set of rules can be defined by:

$$R: = (r_1, r_2, \ldots, r_p) \quad (4)$$

The model is shown in Figure 2.

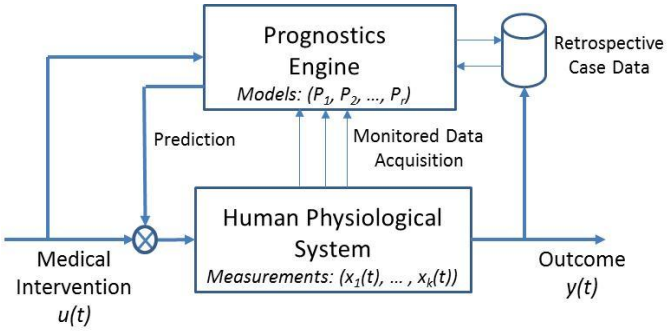

Figure 2. The Medical Prognostic Model

The prognostics engine works continuously by monitoring real time patient data and applying mathematical algorithms that can discern data patterns and make predictions about propensity of certain disease or adverse events occurring in the near future.

The medical intervention, retrospective case information and monitored data can be mathematically described as sets of variables. We can write prediction as a function of multiple variables including the input clinical data and medical intervention. Prediction is a mapping between new input data and an outcome from a set of retrospective cases.

$$Prediction\ (t+t_1) = F(X, U, R) \quad (5)$$

where physiological data set collected from the patient is represented by vector X; medical treatment plans are selected from a set of treatment plans shown as U; and retrospective cases are denoted by R as the set of prior relationships established between physiological data and outcome

The goal of this research is to identify the appropriate mathematical model $F(X, U, R)$ that selects the appropriate prediction from a set of possible outcomes. The model is a mapping function developed based on historical data patterns that maps the input data to a specific outcome.

3.1. Mathematical Model

A large volume of literature concerning mathematical models to predict biological and medical conditions has been published. But only a few of such works in predictive mathematical tools have found their way into mainstream clinical applications and medical practice. Several reasons are cited for the low adoption of predictive tools: either important biological processes were unrecognized or crucial parameters were not known, or that the mathematical intricacies of predictive models were not understood (Swierniak, Kimmel & Smieja, 2009).

The properties of an appropriate mathematical model for predicting medical health condition include: accuracy, prediction, economy, well-posedness and utility (Smye & Clayton, 2002). Among constructs used in prior research, several distinct mathematical models can be found, such as: multivariate regression analysis, Markov chains and stochastic processes, Bayesian networks, fuzzy logic, control theory, discrete event simulation, dynamic programming and Neural Networks.

There are three evolving philosophies pertaining to biological and medical prediction: one is grounded in control theory. Decay of human physiology and adverse medical conditions such as Intra-cranial Pressure (ICP), or carcinogenesis can be viewed as a result of loss of body's control over its critical mechanisms. For example, loss of control over blood flow regulation leads to irregular intracranial pressure; or loss of control over cell cycle causes altered function of a certain cell population that leads to cancer. Medical intervention is viewed as a control action for which the human body is the system. This approach requires a deep understanding of the internal causal models between control mechanisms and human physiology.

The second approach follows the Markov chain model as it considers the disease cycle as a sequence of phases traversed by each physiological subsystem from birth to expiration. For example, a patient that develops pneumonia starts from a healthy normal state and then deteriorates through four stages of Congestion, Red hepatization, Gray hepatization, Resolution (recovery).

The third approach considers the human body as a black box. Since we don't have perfect knowledge about each individual's physiology, environmental, genetic and cultural information and in the areas of medicine where our knowledge of clinical evidence is uncertain, we can only rely on predictive models that take data from physiological sensors and laboratory results and apply certain models and rules developed through retrospective studies to make predictions.

These models have considered both deterministic and probabilistic approaches. Other mathematics constructs consider the asynchronous nature of biology and thus their approach uses simulation models. For example, one study applied simulation and statistical process control to estimate occurrence of hospital-acquired infections and to identify medical interventions to prevent transmission of such infections (Limaye, Mastrangelo, Zerr & Jeffries, 2008).

Other predictive models in cancer therapy have used stochastic process to predict drug resistance of cancer cells and variability in cell lifetimes (Kimmel, Axelrod, 2002). The most successful predictive methods in literature are model-free approaches using neural networks and fuzzy sets (Kodell, Pearce, Baek, Moon & Ahn, 2009) (Arthi & Tamilarasi 2008).

A vast majority of mathematical models in medicine are developed for diagnosis. A survey of literature from 1970's to present, reveals that more attention has been given to decision support and diagnoses models than to prediction. The development of prognostics models to predict short term medical health condition of individuals has been under explored.

Given the non-linear aspect of relationships between physiological measurements, medical outcome and medical treatment plan, a mathematical method that best models non-linear relationships is needed for the prognostics engine. It has been established that neural networks are among the most effective methods to discern patterns and non-linear relationships between data.

3.2. Neural Networks

Neural networks have been successfully applied to classify patterns based on learning from prior examples. Different neural network models use different learning rules, but in general they determine pattern statistics from a set of training examples and then classify new data according to the trained rules. Stated differently, a trained neural network model classifies (or maps) a set of input data to a specific disease from a set of diseases.

To illustrate the classification model for this case study, a simple example is described below and in Figure 3. We can classify input data about patients into two categories of predictions: Disease-True and Disease-False, by looking at prior patient data. The objective of the single-layer perceptron is to determine a linear boundary that classifies the patients on either side of the linear boundary. As shown in Figure 3, we wish to classify patients into two categories separating by a boundary called a decision boundary line. A linear set of equations define this boundary. The region where the linear equation is >0 is one class (Disease-True), and the region where the linear equation is <0 is the other class (Disease-False). The line is defined as:

$$W_0X_0 + W_1X_1 + W_2 = 0 \quad (6)$$

We can apply a threshold function to classify patients based on the following threshold function:

$$f(X_0, X_1) = \begin{cases} 1 & if\ W_0X_0 + W_1X_1 + W_2 \geq 0 \\ -1 & if\ W_0X_0 + W_1X_1 + W_2 < 0 \end{cases} \quad (7)$$

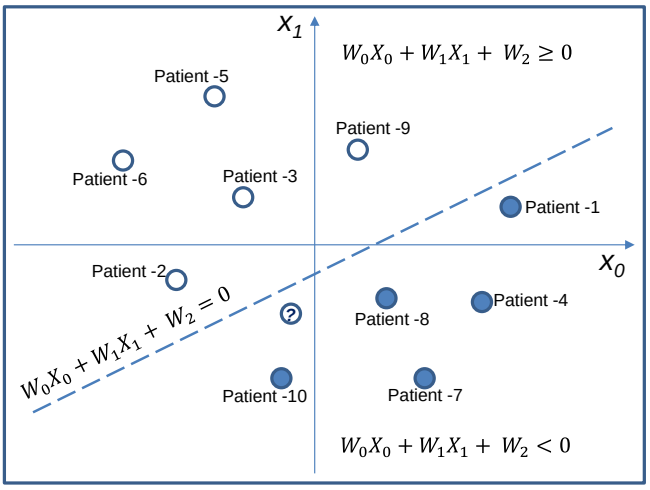

Figure 3. Classification using single-layer perceptron

Suppose we're considering classifying patients by only four input variables, Glucose (G), Body mass (M), Systolic Blood pressure (S) and White blood cell count (B). The threshold function would be computed as follows:

$$f(X_0, X_1, X_2, X_3) = \begin{cases} 1 & if\ W_4 + \sum_{i=0}^{3} W_iX_i \geq 0 \\ -1 & if\ W_4 + \sum_{i=0}^{3} W_iX_i < 0 \end{cases} \quad (8)$$

The appropriate predictive mathematical model must offer accuracy and simplicity to learn from prior cases and easily be extensible to apply new data to make predictions about a patient's health condition. It has been established that the most accurate neural network models for prediction are as follows:

1) PNN - Probabilistic Neural Networks are four layer networks. They classify data in a non-parametric method and are less sensitive to outlier data. It's been demonstrated that probabilistic neural networks using only four layers of input, pattern, summation and output perceptron can provide accurate and relatively faster classifications than the back-propagation neural networks.

2) SVM – Support Vector Machine networks. SVM performs classification by constructing a two-layer network that defines a hyperplane that separates data into multiple classifications. The SVM is a non-probabilistic binary linear classifier. It takes a set of input data and determines which of possible classes the input is a member of.

3) Generalized Feedforward Multi-Layer Perceptron (MLP) trained with LM – A feedforward neural network consists of one or more layers of nodes where the information flows in only one direction, forward from the input nodes and there are no cycles or loops in the network. In the multi-layer model, each node has direct connection to the nodes in the subsequent layer. The sum of products of the weights and the inputs are calculated in each node (Haykin 1999).

4) MLP trained with LM – Muli-layer perceptron, a method similar to gradient descent approach with variable step modification. Several variations of this model have been proposed, including the Levenberg-Marquardt model (Wilamowski & Chen, 1999) which is known to be among the most efficient algorithms.

This study applied and compared prediction results from all four neural network models. These models were compared based on their accuracy.

4. CLINICAL CASE STUDY

The clinical case study consisted of 468 patient cases who were admitted to a hospital for various treatments. The patient data consisted of 21 independent variables and one dependent variable. The input data included various relevant physical and vital sign data ranging from blood pressure to heart rate and blood lab test results. The input variables consisted of both continuous and dichotomous variables. The dependent variable was a dichotomous variable that represented the clinical outcome, the occurrence or absence of a disease. In this study, the output was defined by a marker called Deep Vein Thrombosis (DVT). Of the patient population in this study, 89 were positively diagnosed with DVT.

DVT is the formation of blood clots in deep veins, typically in leg veins. Blood clots can dislodge and flow to lungs causing a more critical condition called Pulmonary Embolism (PE). DVT/PE is a serious medical condition that can cause serious pain and even death. In the US alone approximately 350,000 to 600,000 patients suffer from DVT and at least 100,000 deaths per year are attributed to DVT/PE (The Surgeon General's Call to Action to Prevent Deep Vein Thrombosis and Pulmonary Embolism, 2008).

Neural networks have been successfully applied to classify patterns based on learning from prior examples. Different neural network models use different learning rules, but in general they determine pattern statistics from a set of training examples and then classify new data according to the trained rules. Stated differently, a trained neural network model classifies (or maps) a set of input data to a specific disease from a set of diseases.

Four models were trained and tested in two stages: in the first stage, we used genetic neural network algorithm to identify the input variables with most predictive power. We narrowed the list of input variables from 21 down to 14 variables. In the second stage, we trained and tested all four models on the 14 input variables from stage 1. The list of the most predictive variables is given in Table 1.

Input Variable	**Data Type**	**Definition**
ADMITTED OVER 48 HRS	Dichotomous	In hospital over 48 hours?
INPATIENT	Dichotomous	Is patient admitted as inpatient?
MAX GLUCOSE	Continuous	Maximum Glucose level during patients' stay.
MAX WEIGHT	Continuous	Maximum weight during stay in Kg.
MIN PLATELET	Continuous	Minimum no. of blood platelets, tiny cells that assist in blood clotting
MIN INR	Continuous	Minimum INR (International Normalized Ratio). The standard for a healthy person is 1.
MAX INR	Continuous	Maximum INR (International Normalized Ratio).
MAX RBC	Continuous	Maximum no. of red blood cells
MIN RBC	Continuous	Minimum no. of red blood cells
MAX HEMOGLOBIN	Continuous	Maximum no. of hemoglobin, a red protein that carries oxygen in the blood.
MIN HEMOGLOBIN	Continuous	Minimum no. of hemoglobin. a red protein that carries oxygen in the blood.
MAX HCT	Continuous	Maximum hematocrit: the proportion, by volume, of red blood cells
MIN HCT	Continuous	Minimum hematocrit: the proportion, by volume, of red blood cells
MIN RDW CV	Continuous	Minimum red blood cell distribution width.
MIN RDW CV3	Continuous	Minimum red blood cell distribution width Coefficient Variation-3.
MIN RDW CV4	Continuous	Minimum red blood cell distribution width Coefficient Variation-4.

Table 1. Input variables description

4.1. Computational Method

In this study, we computed and optimized four different prediction and classification algorithms on 21 data input variables and 468 patient cases. There were 89 true positive cases in the retrospective study. We used NeuroSolutions V6.0 (NeuroDimension, Inc. 2011) tools to build and test the models.

In each computation, we trained the network using one of the four neural network methods. For all four methods, we selected the "Leave-N-out" technique. This technique is a cross training and validation method used to minimize bias due to random data selection. This approach trains the network multiple times, each time omitting a different subset of the data and using that subset for testing. The outputs from each tested subset are combined into one testing report and the model is trained one final time using all of the data.

The test results of all four models can be compared using classification measures such as number of false positives (FP), false negatives (FN), true positives (TP) and true negatives (TN), as shown in Table-2:

Model	TP	FP	TN	FN	Total
Probabilistic Neural Network	5	7	372	84	468
Support Vector Machines	30	83	296	59	468
Multi-layer Perceptron with LM	24	78	301	65	468
Generalized Feed forward with LM	19	68	311	70	468

Table 2. Model test results

4.2. Accuracy and Validation

External validity of medical prediction models is an extremely challenging task. Clinical validation is challenging not just because it involves prospective patient studies, double-blind studies and careful administration of research protocols, but for two other reasons: first, if a patient gets the treatment, could that patient have exhibited the predicted disease? Second, if a patient is not treated and the disease occurs, would the outcome been the validation of the model's prognosis had the patient been treated? We'll focus on accuracy in this paper and consign clinical validation to a future research project.

Several measurements have been proposed as methods for internal validation. Some of the measurements that are commonly used to compare accuracy of classification models include: Accuracy, Sensitivity, Specificity, Area Under Receiver Operating Curve (AUROC) and Likelihood Ratio (LR). Sensitivity measures the fraction of positive cases that are classified correctly as positive. Specificity is the fraction of negative cases that are classified correctly as negative. AUROC is a good overall measure of predictive accuracy of a model. It represents a plot of sensitivity versus (1 - specificity). An AUROC close to 1.0 is a considered an excellent discrimination, but a value near 0.50 suggests no discrimination (similar to a coin flip). Positive LR is the ratio of sensitivity to (1 - specificity). (Delen 2009). The accuracy measures may be defined as:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

$$sensitivity = \frac{TP}{TP+FN} \quad (10)$$

$$specificity = \frac{TN}{TN+FP} \quad (11)$$

$$LR+ = \frac{sensitivity}{1-specificity} = \frac{\Pr\ T+\ D+)}{\Pr\ T+\ D-)} \quad (12)$$

where *(T+ | D+)* denotes a case with positive test of a disease when the disease actually exists, and (T+ | D-) denotes a case with positive test of a disease but the patient does not present with the disease.

When a model uses continuous data measurements, then different thresholds may be applied in order to decide which value is the cut-off to distinguish between patients with disease. The best model has the highest values for sensitivity and specificity. In certain situations, both may not be equally important. For example, a false-negative (FN) prediction might be more critical than a false-positive (FP) prediction. If we apply no preference to either measurement then, Youden's index (*J*) may be used to choose an appropriate cut-off, computed by (Bewick, Cheek, Ball 2004):

$$J = sensitivity + specificity - 1 \quad (13)$$

The maximum value that *J* can take is 1, when the test is perfect.

4.3. Comparison of results

All four models were optimized for classification of cases into a dichotomous dependent variable: the presence or absence of DVT.

The results showed that the SVM algorithm was most accurate followed by the MLP model and the General feed forward neural network model. All four methods are compared using the accuracy measurements in Table 3.

Measure-ment	Probabil - istic Neural Network	Support Vector Machine	Multi-Layer Percep-tron - LM	General-ized Feed forward-LM
Accuracy	0.8056	0.6966	0.6944	.7051
Sensitivity	0.0562	0.3371	0.2697	0.2135
Specificity	0.9815	0.7810	0.7942	0.8206
LR+	3.0417	1.5392	1.3103	1.1899
Youden's *J*	0.0377	0.1181	0.0639	0.0341

Table 3. Accuracy Measures of Neural network models

All four models exhibited low sensitivity measures indicating their poor ability to detect true positives. This is due to the lower number of positive DVT cases in this study (only 89 out of 468 cases had positive DVT cases).

4.4. Use of an Oracle to Select the Best Model

Since their introduction in 1960's, various neural network algorithms have been proposed and successfully implemented to classify and predict future state of output variables. Certain models are more suitable to specific class of problems based on the type and number inputs and output classifications. Typically, no single neural network model is best for all types of problem.

Given that there are many neural networks to select from, the goal is to select the most accurate model for prediction of each disease condition. Therefore, we propose that the prognostics engine utilizes several different algorithms to determine accuracy of each method and then use an oracle, an overseer program that selects the most accurate model. An oracle is defined as the medium which selects the best answer amongst a set of options. An oracle can be defined to select the best algorithm or a combined prediction from an ensemble of algorithms based on desired accuracy or characteristic of the prediction problem at hand.

Given that one model performs better in predicting true positives and another better at predicting the true negatives, we propose the oracle program to combine the predictions from models in a way that the model with higher accuracy is assigned a higher weight and the worst model still contributes to the prediction but at a smaller weight. This way, the oracle can improve the classification accuracy, sensitivity and specificity by combining the best classification characteristics from different models.

An approach that uses an ensemble of prognostic algorithms is shown to be effective in providing more accurate prediction (Hu, Youn & Wang 2010).

We produced two Oracle methods to compute the combined predictions of the ensemble. The first Oracle used conditional logic to maximize the number of TP and minimize the number of FP predictions.

The results of the oracles are shown in Table 4 and accuracy comparison in Table 5.

Model	TP	FP	TN	FN	Total
Oracle #1 – Ensemble of PNN & SVM models	35	107	272	54	468
Oracle#2 – Ensemble of all four models	46	141	238	43	468

Table 4. Results of the two oracle program

Ensemble1 essentially took the best traits from the PNN and SVM models to produce a more accurate prediction. The second Oracle combined weighted sum of predictions from each model in the ensemble. The weights were determined to maximize the number of FP predictions.

Measurement	Oracle #1	Oracle #2
Accuracy	0.6560	0.6068
Sensitivity	0.3933	0.5169
Specificity	0.7177	0.6280
LR+	1.3929	1.3893
Youden's *J*	0.1109	0.1448

Table 5. Comparison of Oracles' accuracy

One method to compare all four models and the two oracle programs is to use the Receiver Operating Curve (ROC) plot. The ROC curve is a plot of sensitivity versus (1 – specificity), and generally is considered a good accuracy measure of binary classifiers (Bourdes, Ferrieres, Amar, Amelineau, et al, 2011). Figure 4 shows a scatter plot of ROC for all models. The best prediction method would result in a point in the upper left corner of the diagram. The diagonal line depicts a random guess or prediction by a flip of coin.

The diagram illustrates two observations: The prediction results are not as accurate as one would like. This is attributed to the fact there were too few positive cases in the entire population to help train a more accurate predictive model. Furthermore, several of input variables were highly correlated such that the predictive contribution of some variables was less significant for making a more accurate prediction.

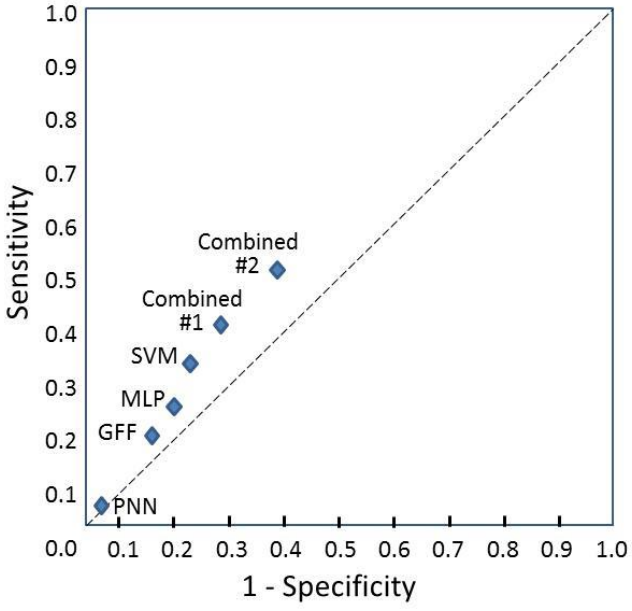

Figure 4. ROC curve for results of all models

The second observation is that the combined ensemble methods #1 and #2 were more accurate than each neural network model alone.

5. CONCLUSION

Various neural network methods can be used to identify the most accurate model to make short term predictions about patient health condition. The performance of these models varies depending on the type and volume of input and output variables. The conclusion of our study is not to say which model or method is better, but, to recognize that each model has strengths and weaknesses. By combining multiple models we can improve classification accuracy. Since no single model can be the best fit for all medical prediction problems, an oracle program is proposed to select the best weighted combination of multiple neural network models.

ACKNOWLEDGEMENT

Authors would like to thank and acknowledge insights and feedback provided by Doctors John Bramhall, M.D, PhD. and Michael Souter, M.B, Ch.B. to this research.

REFERENCES

Arthi, K., Tamilarasi, A. (2008). Prediction of autistic disorder using neuro fuzzy systems by applying ANN technique, *International Journal of Developmental Neuroscience*, 26 (2008) 699-704.

Bewick, V., Cheek, L., Ball, J., Statistics review 13: Receiver operating characteristic curves, *Critical Care*, Vol. 8, no. 6, December 2004

Bourdes, V., Ferrieres, J., Amar, J., Amelineau, E., Bonnevay, S., Berlion, M., Danchin, N., Prediction of persistence of combined evidence-based cardiovascular medications in patients with acute coronary syndrome after hospital discharge using neural networks, *Medical Biological Engineering Computing*, 49:947-955, 2011

Brown, S.W., Strong, V. (2001), The use of seizure-alert dogs, *Seizure*, 2001, 10:39-41.

Daley, M., Narayanan, N., Leffler, C. W. (2010), Model-derived assessment of cerebrovascular resistance and cerebral blood flow following traumatic brain injury, *Experimental Biology and Medicine*, Vol 235, April 2010

(Delen 2009) D. Delen, "Analysis of cancer data: a data mining approach", *Expert Systems*, February 2009, Vol. 26, No. 1

Gao, E., Young, W., Ornstein, E., Pile-Spellman, J., Qiyuan, M. (1997), A theoretical model of cerebral hemodynamics: Appliation to the study of Arteriovenous Malformations, *Journal of Cerebral Blood Flow and Metabolism*, 1997, 17, 905-918

Hahnfeldt, P., Panigraphy, D, Folkman, J., Hlatkey, L., Tumor development under angiogenic signaling: a dynamic theory of tumor growth, treatment response and postvascular dormancy, *Cancer Research* 59, 4770-4778

Haykin, S. (1998), Neural Networks, A Comprehensive Foundation, 2nd Edition, Prentice Hall, 1999

Hu, C., Youn, B.D., Wang, P., Ensemble of data-driven prognostics algorithms with Weight Optimizatino and K-Fold Cross Validation, Annual Conference of the Prognostics and Health Management (PHM) Society, Oct 10-16 2010, Portland, OR.

INCOSE (International Council on Systems Engineering Council) *Systems Engineering Handbook* (2000), *What is a system?*, Version 2.0, July 2000.

Kimmel, M., Axelrod, D. E. (2002), *Branching processes in biology*, Springer Verlag, New York, NY, 2002

Kirton, A., Winter, A., Wirrell, E., Snead, O.C. (2008), Seizure response dogs: Evaluation of a formal training program, *Epilepsy & Behavior*, 13 (2008) 499-504.

Kodell, R. L. , Pearce, B. A. , Baek, S., Moon, H., Ahn, H., et al. (2009), A model-free ensemble method for class prediction with application to biomedical decision making, *Artificial Intelligence in Medicine*, 46, 267-276

Limaye, S. S., Mastrangelo, C. M.,. Zerr, D. M., Jeffries, H. (2008), A statistical approach to reduce hospital-associated infections, *Quality Engineering*, 20:414-425, 2008

NeuroDimension, Inc. (2011), Gainesville, Florida, NeuroSolutions software, Version 6.0

Niu, G., Yang, B., Pecht, M., Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance, *Reliability Engineering and System Safety*, V95, n7, p786-796, 2010

O'Connor, A. M., Bennett, C. L., Stacey, D., Barry, M., Col, N. F., Eden, K. B. , Entwistle, V. A., Fiset, V., et. al. (2009), Decision aids for people facing health treatment or screening decisions (Review), *The Cochrane Collaboration*, Wiley 2009

Pecht, M. (2008), Prognostics and Health Management of Electronics, Wiley 2008

Smye, S. W., Clayton, R. H. (2002), Mathematical modeling for the new millennium: medicine by numbers, *Medical Engineering & Physics*, 24 (2002), 565-574

Souter, M. (2011), Conversations about diagnostic markers and predictors, April 7, 2011.

Strong, V., Brown, S. W., Walker, R. (1999), Seizure-alert dogs-fact or fiction?, *Seizure*. 1999; 8:26-65.

Swierniak, A., Kimmel, M., Smieja, J. (2009), Mathematical modeling as a tool for planning anticancer therapy, *European Journal of Pharmacology*, 625 (2009) 108-121

Uckun, S., Goebel, K., Lucas, P. J. F. (2008), Standardizing Research Methods for Prognostics, *2008 International Conference on Prognostics and Health Management*

Vichare, N. M., Pecht, M. (2006), Prognostics and Health Management of Electronics, *IEEE Transactions on Components and Packaging Technologies*, Vol 29, No. 1, March 2006.

Virchow, R. (1856), Virchow's triad was first formulated by the German physician Rudolf Virchow in 1856.

Williams, H., Pembroke, A. (1989), Sniffer dogs in the melanoma clinic?, *Lancet*, 1989, 1(8640);734

Wilamowski, B. M., Chen, Y. (1999), Efficient algorithm for Training Neural Networks with one Hidden Layer, IEEE International Joint Conference on Neural Networks, 1999

Yu, C., Liu, Z., McKenna, T., Reisner, A. T., Reifman, J. (2006), A Method for Automatic Identification of Reliable Heart Rates Calculated from ECG and PPG Waveforms, *Journal of the American Medical Informatics Association*, vol. 13, No. 3, May/Jun 2006.

Peter K. Ghavami Received his M.S. in Engineering Management from Portland State University and B.S. from Oregon State University. He is currently Director of Imaging Informatics at Harborview Medical Center and a graduate student at University of Washington, Dept. of Industrial and Systems Engineering. He has authored papers on software process improvement, vector processing, distributed network architectures, and software quality. He authored the book ***Lean, Agile and Six Sigma Information Technology Management*** in 2008. He is a member of IEEE Reliability Society and IEEE Life Sciences Initiative.

Kailash C. Kapur is a Professor of Industrial & Systems Engineering in the College of Engineering at the University of Washington. He was the Director of Industrial Engineering at the University of Washington from January 1993 to September 1999. He has co-authored the book ***Reliability in Engineering Design***, John Wiley & Sons, 1977. He is a *Fellow* of American Society for Quality and a fellow of the Institute of Industrial Engineers.