

A TWO TIER NEURAL INTER-NETWORK BASED APPROACH TO MEDICAL DIAGNOSIS USING K-NEAREST NEIGHBOR CLASSIFICATION FOR DIAGNOSIS PRUNING

J. B. Siddharth Jonathan* and K.N. Shruthi**

*Department of Computer Science, Stanford University, USA.

**Department of Biomedical Engineering, University of Southern California, USA.

siddharth.jonathan@cs.stanford.edu, sknaraya@usc.edu

Abstract: Artificial Intelligence has always lent a helping hand to the practitioners of medicine for improving medical diagnosis and treatment. In this paper, we propose the design of a two tier Neural Inter-network based Medical Diagnosis System (NIMD) that uses k-Nearest Neighbor Classification for *Diagnosis pruning*. The system (fig a.) is essentially two tiered with the first tier handling what we term, *diagnosis pruning*. Each disease is often characterized by a group of symptoms and *Diagnosis pruning* which is done in the first tier, is the classification of diseases based on the groups of symptoms observed. The second tier consists of separate modules for each disease that handles the actual detection of the disease based on the intensities of the various symptoms reported by the patient. The disease detection modules comprise different classifiers like Neural Networks, Decision Trees, Bayesian Networks etc. depending on the size of the input vector and the characteristics of the training set. NIMD thus provides diagnosis and probability of presence for a number of possible diseases at the same time by checking for each of them individually in tier 2. An analysis of the performance of our hybrid system reveals superior performance and utility compared to other existing approaches.

Index terms- Artificial Intelligence, Medical Diagnosis, Neural Networks, Decision Trees, Bayesian Networks.

Introduction

Artificial Intelligence has always lent a helping hand to the practitioners of medicine for improving medical diagnosis and treatment. A.I in medical diagnosis has been particularly relevant because studies have revealed that humans do not resemble statistical computers but pattern recognition systems.

Humans can recognize patterns or objects very easily but fail when probabilities have to be assigned to observations.

Rule based expert systems have been used for medical diagnosis but they have their share of problems. If the system consists of several thousand rules, it takes a very powerful control program to produce any conclusions in a reasonable amount of time. Another problem with rule based systems is that as the number of rules increases, the conflict set also becomes large so a good conflict resolving algorithm is needed if the system is to be usable.

In this paper, we propose the design of a two tier Neural Inter-network based Medical Diagnosis System (NIMD) that uses k-Nearest Neighbor Classification for *Diagnosis pruning*.

The rest of the paper is organized as follows. We first describe give an overview of the previous work in this field. The proposed system and its design are presented in the next section. Detailed descriptions of tier 1 and tier 2 are then presented. Next, the performance analysis of NIMD is presented. Finally, we summarize the paper and give insight into possible further development and future directions of ongoing work.

Related Works

Neural Network based Medical Diagnosis Systems: Medical Diagnosis using Artificial Neural Networks (ANN) [1] is currently a very active research area in medicine and it is believed that it will be more widely used in biomedical systems in the next few years. This is primarily because the solution is not restricted to linear form.

Neural Networks are ideal in recognizing diseases using scans since there is no need to provide a specific algorithm on how to identify the disease. Neural networks learn by example so the details of how to recognize the disease are not needed. What is

needed is a group of examples that are representative of all the variations of the disease. The quantity of examples is not as important as the 'quality'. The examples need to be selected very carefully if the system is to perform consistently and efficiently.

As cited in [5], in the detection of breast cancer by supervised learning, the system is trained from a set of cases, using the leave-one-out method. The omitted subset computes whatever error criterion is of interest. The system initially learns from digitized screen mammograms. The eight input nodes represent features of calcifications, areas in breast tissue where tiny calcium deposits build up and might indicate the presence of cancer. The AI system feeds these nodes into the neural network to provide a statistical indication of the possibility that a group of calcifications is malignant.

Instant Physician: Yet another model that uses AI in medical diagnosis is the "instant physician" [6]. Here, an auto associative memory Neural Network is trained to store a large number of medical records, each of which includes information on symptoms, diagnosis, and treatment for a particular case. After training, the net can be presented with input consisting of a set of symptoms; it will then find the full stored pattern that represents the "best" diagnosis and treatment.

Rule Based Medical Diagnosis Systems: MYCIN was an expert system developed to diagnose blood infections. MYCIN used about 450 rules and was able to perform as well as some experts and considerably better than junior doctors. No general theoretical model existed from which rules were deduced. They had to be acquired from extensive interviewing of experts who in turn acquired them from direct experience of cases. The rules also had to reflect the uncertainty associated with medical knowledge. To this end, MYCIN incorporated a calculus of certainty factors which seemed to fit well with how doctors assess the impact of evidence on the diagnosis.

NIMD-Overall Architecture

Many diseases are characterized by a common group of symptoms and although a number of diseases may have a similar group of symptoms, they differ in the relative intensity of the symptoms in the group. Common Cold and Respiratory Bronchitis are diseases that justify this fact. Diagnosis pruning is a step that identifies the diagnosis class that a disease belongs to. A diagnosis class is a subset of diseases that possess the same group of symptoms. After

diagnosis pruning, there is a separate disease detection classifier for each disease in the diagnosis class that computes the actual probability of presence of the disease based on the intensities of the symptoms in the group.

NIMD is a 2-tiered system. The first tier performs *diagnosis pruning* as described earlier. The second tier performs the actual disease diagnosis, making use of the various independent modules of classifiers.

The basic operation of the system is as follows. A universal symptom set is established for all the diseases that NIMD is set up to diagnose.

The patient is then asked to list the symptoms he has. We use a simple binary encoding scheme wherein the presence of a symptom is represented by a 1 in the input vector and 0 otherwise. We call this the *symptom vector*. The patient is also asked to rate each symptom on a scale of 1 to 10, with 1 indicating minimal and 10 indicating severe incidence of the symptom. The latter information is for the second tier and is called the *intensity vector*. The symptom vector is fed to the K-Nearest Neighbor Classifier, which performs diagnosis pruning by mapping this to a diagnosis class. Under the particular diagnosis class identified, there can be more than one disease. In the second tier, for each disease there is a separate detection unit. All units associated with a particular diagnosis class are activated when the diagnosis class is selected. Now, the input to the second tier is the intensity vector. The second tier detection units map the relative intensity of the symptoms to a probability of presence of the disease. The second tier units have been trained previously with comprehensive data sets for this purpose.

The need for having multiple detection units operating simultaneously under a single diagnosis class is because although multiple diseases might fall under a single diagnosis class (similar group of symptoms), diseases have varying intensity levels associated with individual symptoms.

Each module is independent of the others. For example, we can have a Neural Network for a disease which involves a lot of symptoms and a decision tree for a disease with relatively less number of symptoms. The advantages are evident from our comparative performance analysis.

Tier 1: K – Nearest Neighbor Classifier

We now look at the working of tier 1 of NIMD in detail. The k- Nearest Neighbor Classifier is a standard classifier that utilizes a form of memory based learning where the past experiences are

explicitly stored in a large memory of correctly classified input-output examples:

$$\{(x_i, d_i)\} \text{ for } i=1 \text{ to } N$$

where x_i denotes an input vector and d_i denotes the corresponding desired response. The input vector is the entire list of possible symptoms for all the diseases that we diagnose. The output is the correct classification, which in our case is the diagnosis class i.e. the logical grouping of related symptoms. The k-Nearest Neighbor Classifier proceeds by first identifying the k classified patterns that lie nearest to the test vector x_{test} for some integer k. Distance here refers to the Euclidean distance between the vectors x_i and x_{test} . The Euclidean distance between a pair of m-by-1 vectors x_i and x_j is defined by,

$$d(x_i, x_j) = \|x_i - x_j\| = \left[\sum_{k=1 \text{ to } m} (x_{ik} - x_{jk})^2 \right]^{1/2}$$

where x_{ik} and x_{jk} are the k^{th} elements of the input vectors x_i and x_j , respectively.

We then assign x_{test} to the class that is most frequently represented in the k nearest neighbors to x_{test} . The classification thus performs *diagnosis pruning* and provides us the class of symptoms that need to be handled.

Tier 2: Disease Detection Modules

The second tier consists of an array of Neural Networks, Decision Trees etc, each trained with training sets for the disease it is supposed to detect. Because of the highly modular architecture of our system, each Neural Network can be of a different model. For example, we can have Back Propagation Networks, Kohonen Networks, Boltzmann Machines etc.

Another standard classifier that can be used is a Decision Tree. A Decision tree, in general takes as input, an object or situation described by a set of properties and outputs a yes/no decision. Here the input is the set of symptoms under the classified diagnosis class and the output is a yes/no decision indicating the presence or absence of the disease.

The choice of the type of module can be decided based on the size of the input vector, the characteristics of the training set and after analyzing the performance of the module in detecting the disease. The input vector to the modules is the intensity vector based on the diagnosis class that it was classified under. This input vector contains just

the symptoms relevant to the detection of this particular disease with their stated intensities. Thus training is quicker and testing is more efficient because there are no unnecessary symptoms that the network needs to take into account during testing and training.

Performance Analysis

NIMD as described above uses a different classifier for each disease. The choice of classifier can be made based on the size of the input vector and the characteristics of the training set. Because we use a single classifier per disease, to diagnose the probability of presence of a single disease, it results in improved accuracy, lower error rates and faster training as compared to using the same classifier to detect multiple diseases.

In figures 1 and 2, we present the comparison of the performance of our disease detection unit, which diagnoses a particular disease with a generic neural network that tries to check for the presence of more than 1 disease. The analysis shows that our neural network has a lower net error and the training proceeds at a faster rate.

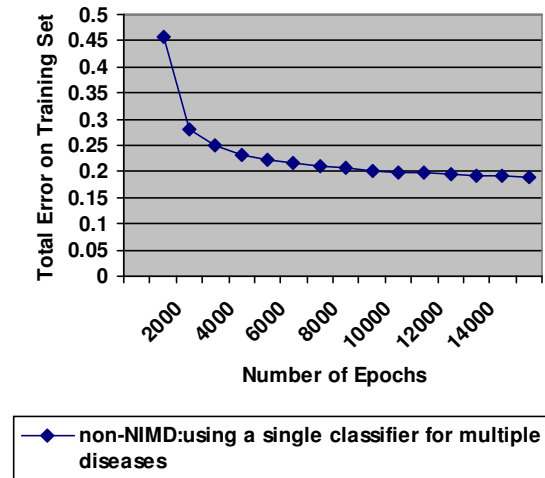


Figure 1. Training curve showing the gradual reduction in error as weights are modified over several epochs for a system that uses a single classifier for multiple diseases.

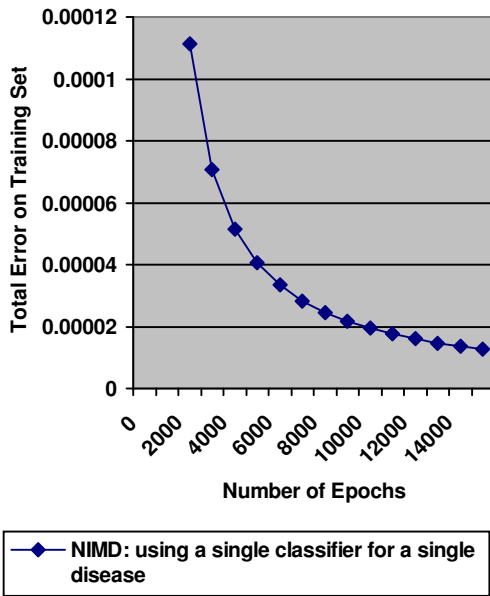


Figure 2. Training curve showing the gradual reduction in error as weights are modified over several epochs for NIMD that uses a single classifier for a single disease.

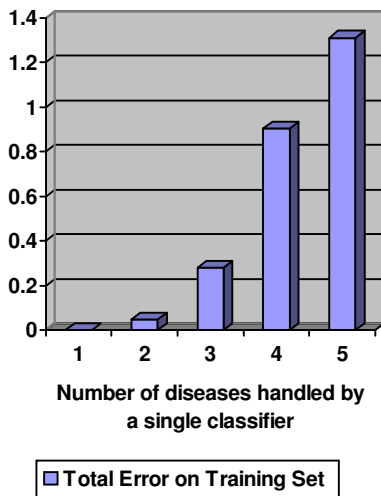


Figure 3. The total error on the training set increases as the number of diseases classified by the classifier increases.

In figure 3, we show how the total error in the classification increases as we attempt to classify multiple classes, i.e attempt to diagnose more than one disease. This happens as class boundaries are difficult to determine. Thus, in NIMD we use a different classifier unit for each disease.

Concluding Remarks

With regard to the hardware, we propose the use of analog CMOS circuits that perform in-circuit learning using Gilbert multipliers as the primary circuit component. Such circuits already exist. However, instead of using capacitors for the storage of weights, we propose the use of Flash memory, which has the advantage of being a non-volatile element thus eliminating the need to interleave learning cycles with classification cycles. It is also comparatively less expensive than an EEPROM.

We are currently exploring other alternatives for the first tier by analyzing the performance of other classifiers in performing *diagnosis pruning*. We are also working on optimizing the hardware for NIMD.

References

- [1] H. Kordylewski., D. Graupe, Kai Liu, “ A novel large-memory neural network as an aid in medical diagnosis applications” IEEE Transactions on Information Technology in Biomedicine , Sept. 2001 V 5#3.
- [2] P. Meesad, G.G.Yen, “A hybrid intelligent system for medical diagnosis” Proceedings of International Joint Conference on Neural Networks, 2001. IJCNN '01, July 2001, pp. 2558 - 2563 V 4.
- [3] Jeong-Yon Shim, “Three dimensional knowledge learning memory with dynamic Bayesian associative matrix for the medical diagnosis” Proceedings of International Joint Conference on Neural Networks, 2001, IJCNN '01, July 2001, pp. 2388 – 2392 V 4.
- [4] Hong Zhang, F. C. Lin, “Medical diagnosis by the virtual physician” Proceedings of the 12th IEEE Symposium on Computer-Based Medical Systems, 1999, June 1999 pp. 296 – 302.
- [5] Danna Voth, “Using AI to detect breast cancer” IEEE Intelligent Systems, Jan./Feb. 2005.
- [6] Christos Stergiou, Dimitrios Siganos, “Instant physician – Neural Networks”.
- [7] Wen-Jia Kuo, Ruey-Feng Chang, Woo Kyung Moon, Cheng Chun Lee, Dar-Ren Chen, “Computer-aided Diagnosis of Breast Tumors with Different US Systems”.
- [8] Wen-Jia Kuo, Ruey-Feng Chang, Woo Kyung Moon, Cheng Chun Lee, Dar-Ren Chen, “Computer-aided Diagnosis of Breast Tumors with Different US Systems”.