A Case-Based Adaptation Model for Thyroid Cancer Diagnosis Using Neural Networks

Abdel-Badeeh M. Salem, Bassant M. El Bagoury

Faculty of Computer & Information Sciences
Ain Shams University, Cairo, Egypt
Email: absalem@ asunet.shams.eun.eg
bassantai@ yahoo.com

Abstract

In this paper, a new hybrid adaptation model for cancer diagnosis has been developed. It combines transformational and hierarchical adaptation techniques with artificial neural networks (ANN's) and certainty factors (CF's). The model consists of a hierarchy of three phases, which simulates the expert doctor phases of cancer diagnosis. Each phase uses a single ANN to learn the adaptation knowledge to perform the main adaptation task. The model has been tested with 820 thyroid cancered patient cases. Cross-validation test has shown a very high diagnosis performance rate that reaches 99.47%. The model is described in a context of a prototype expert system namely Cancer-C.

1. Introduction

CBR is a reasoning methodology that simulates human reasoning by using past experiences to solve new problems [1,2]. The most crucial CBR tasks involve; case indexing, representation, retrieval and adaptation. In the retrieval task a set of similar cases are retrieved from a case memory of previously solved cases. In the adaptation task the solution of the retrieved similar case is adapted to fit the new case requirements by looking for prominent differences between the two cases and then applying rules to suggest a solution for the new case. Despite of the importance of the adaptation task, it has received a little attention in the research field and only conventional adaptation techniques have been used [2,3]. Recently, artificial neural networks (ANN's) have been used [4,5].

CBR has long been applied in medicine [6]. The adaptation task has failed in most medical CBR based diagnostic systems [7,8]. This is because these medical CBR systems need a huge amount of adaptation knowledge [3] in order to solve all their medical problems.

In this paper, we propose a hybrid case-based adaptation model for thyroid cancer diagnosis. It combines the transformational and hierarchical adaptation techniques [3] with the ANN's [9] and CF's [10]. The model is described in the context of our Cancer-C expert system prototype [11]. Our motivation in this work is to develop an adaptation model for cancer diagnosis domain.

2. The Medical Problem

Cancer is a group of more than 200 diseases one of which we apply in our medical CBR system is thyroid cancer [12]. Figure 1 shows the main types of this disease and the main phases of cancer diagnosis. The main types of thyroid cancer disease are considered definite types, as there are no possible overlapping between them. As shown, the first phase is the *Definite-Phase*, where the doctor has no suspicion feeling of malignancy. In this phase, thyroid cancer is the main disease that is classified into two definite diseases, which are benign and malignant. The second phase is the *Type-Phase*. In this phase, the benign disease is classified into two different types, which are hyperthyroidism (hyper) and grave diseases, while the malignant disease is classified into three different diseases types, which are papillary, medullary and anaplastic. If the doctor found that the patient has signs of benign disease in the *Definite-Phase* then he/she will examine him with benign scans and lab-tests to find the benign disease type in the *Type-Phase*. Otherwise, if the patient has signs of a malignant disease then the doctor will examine him with malignant scans and lab-tests to find the malignant disease type in the *Type-Phase*. The third phase is the *Stage-Phase*. In this phase, every malignant disease type can be classified into one of four stages, for example the stages of the medullary disease are medullary stage-I, medullary stage-II, medullary stage-III and medullary stage-IV. If the doctor found that the patient has a malignant disease type in the *Type-phase* then he/she will do further examinations and scans to determine the stage of the malignant disease invasion in the *Stage-Phase*.

Although, these are the main phases of cancer diagnosis for any new patient case, the diagnosis process is not so easy. There are many patient cases that combine different (indefinite) disease types of thyroid cancer (e.g. papillary with grave stage I, medullary with hyper stage II…). These cases are considered indefinite cases and require the experience of expert doctors. The expert doctor reasoning phases of cancer diagnosis consists of three phases. The first phase is the *Suspicion-Phase*

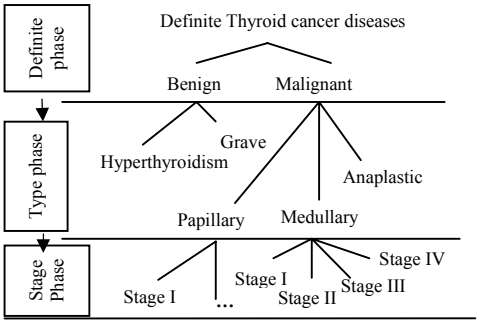

Figure 1. Definite disease types of thyroid cancer and the main phases of cancer diagnosis

where the expert doctor has suspicion feeling of malignancy. In this phase, there are three indefinite diseases, which are benign with symptoms of malignant or malignant with symptoms of benign and malignant disease with benign disease. The expert doctor always have a suspicion feeling of malignancy in the *Suspicion-phase*, even if the patient has no signs of malignancy for this reason, the expert doctor will always move on to the second phase, which is the *To_Be_Sure-Phase.* In this phase, he examines the patient with the essential lab-tests and scans of both benign and malignant diseases to be sure of the indefinite disease types, such as papillary with hyper, papillary with grave or medullary with grave. The third phase is the *Stage-Phase*, where the expert doctor makes further examinations to determine the stage of the malignant disease invasion. If a malignant disease type is found in the *To_Be_Sure-Phase* then he will do further examinations and scans to the patient to determine the stage of his malignant disease invasion in the *Stage-Phase*.

3. Cancer-C Architecture

Cancer-C architecture is shown in figure 2. It consists of three main modules, which are decomposition, retrieval and adaptation. It also includes a case-memory of patients' cases and a rule-base of transformational rules.

3.1 The Case-Memory

Expert doctors in the National Cancer Institute of Egypt supplied our system case-memory with 820 real patient cases and a detailed analysis of thyroid cancer diseases. This is besides other cancer resources from the Internet [12]. As explained by our expert doctors, a typical patient case consists of 44 features, which are critical for the diagnosis. These features can be divided into groups of features. The first group contains 18 features of the initial symptoms of the disease. The second group contains 15 features of the lab-tests and scans results. The third group contains 11 stage features of the malignant disease, if exists. These groups appear to be mutually exclusive, so we decompose each case in the case-memory into sub-cases, which are the *Symptoms-Sub-case*, the *Scans-Sub-case* and the *Stage-Sub-case*. The *Symptoms-Sub-case* is assigned the first group of features, which contains the features of the initial symptoms of the disease. The *Scans-Sub-case* is assigned the second group of features, which contains the features of the lab-tests and scans results. The *Stage-Sub-case* is assigned the third group of features, which contains the stage features of the malignant disease. The *Stage-Sub-case* is formed only if a malignant disease is found. A frame scheme is used to represent each case, where each sub-case is represented by sub-frame as attribute-value pairs.

3.2 Case Decomposition and Retrieval

Given a new case, it will be decomposed into three sub-cases, which are the *new Symptoms-Sub-case*, the *new Scans-Sub-case* and the *new Stage-Sub-case*. For each new sub-case, the *Nearest-Neighbor algorithm* [13] is used to retrieve its most similar sub-case from the case-memory. It computes the similarities between two sub-cases as the weighted average of their local similarities of features. The global similarity function is computed as follows:

$$\text{Similarity}(U,R) = \sum_{i=1}^{n} \text{sim}(U_i, R_i) \times W_i \Big/ \sum_{i=1}^{n} W_i$$

This function takes any value in the similarity range [0%-100%], where 0% is total dissimilarity and 100% is an exact match. $\boldsymbol{U}$ is the new sub-case and $\boldsymbol{R}$ is its corresponding retrieved sub-case. $\boldsymbol{n}$ is the number of features in the sub-case, $\boldsymbol{U_i}$ is the value of feature $\boldsymbol{i}$ in the new sub-case and $\boldsymbol{R_i}$ is the value of the corresponding feature $\boldsymbol{i}$ in the retrieved sub-case, where $\boldsymbol{U_i}, \boldsymbol{R_i} \in \{0,1\}$. $\boldsymbol{W_i}$ is the weight of feature $\boldsymbol{i}$ that reflects its importance on the diagnosis decision of the current new sub-case. ***Sim*** is the local similarity Boolean function computed as:

$$\text{sim}(U_i, R_i) = \left|1 - (U_i + R_i)\right|$$

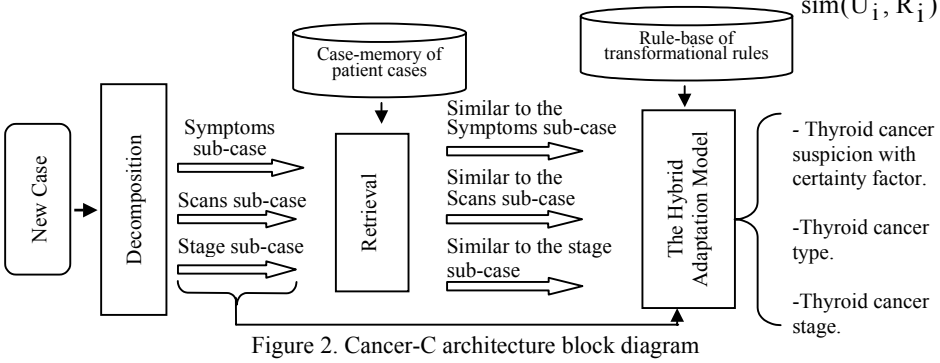

Figure 2. Cancer-C architecture block diagram

In our system, features weights are assigned by our expert doctors and by making analysis to the cases in the case-memory. All weights are normalized in the range [0-10]. All the critical features of each sub-case are used as indexes to speed up the retrieval process.

3.3 The Algorithm of the Hybrid Adaptation Model

The three decomposed sub-cases of the new patient case are diagnosed by using our hybrid adaptation model, which combines transformational and hierarchical adaptation techniques [3] with ANN's [9] and CF's [10]. Case adaptation is performed in a top-down fashion using a hierarchy of three phases, which simulates the expert doctor phases of cancer diagnosis, the *Suspicion-phase* is for diagnosing cancer suspicion, the *To-Be-Sure-phase* is for diagnosing cancer type and the *Stage-phase* is for diagnosing cancer stage. All the three phases are similar in their structure but they are different in their inputs and outputs. Each phase uses a single ANN to learn the adaptation knowledge to find the adapted diagnosis for the new sub-case. The final diagnosis of the new patient case is composed from the adapted sub-cases diagnoses of the three phases, all of which are then evaluated by the expert doctor. The main adaptation algorithm used at each phase is shown in figure 3. The algorithm of the hybrid adaptation model is shown in figure 4. The coming sub-sections explain this algorithm in more details.

3.3.1 The Main Adaptation Algorithm

The steps of the main adaptation algorithm used at each phase are described as follows:

i) Extract the Adaptation Knowledge. Adaptation knowledge [3] means the knowledge that describes how the differences between the features of the new patient case and the features of the retrieved patient case affect the differences in their diagnoses. Adaptation knowledge is extracted from the expert doctors for reliability of diagnosis decisions. An informal important example of our experts' explanations is:

> *"Even if all the initial examinations to a new patient at the Suspicion-phase suggest a benign disease, we will always have at least a 50% suspicion degree of malignancy. So, if a similar patient has very similar initial symptoms and scans to the new patient but his diagnosis was benign disease then adapt his diagnosis so that the new patient diagnosis is benign with 50% certainty and malignant with 50% certainty."*

ii) Apply Transformational Rules. The transformational rules [3] are IF-THEN rules, which use the adaptation knowledge to adapt (transform) the diagnosis of a retrieved *Symptoms-sub-case* in order to find the diagnosis of a new *Symptoms-sub-case.* An informal example of one of our transformational rules is:

> IF the symptom feature lump exists in the new and in the retrieved sub-cases and the cold nodule feature exists in the new sub-case and it doesn't exist in the retrieved sub-case and the warm nodule feature doesn't exist in the new sub-case and it exists in the retrieved sub-case… AND the diagnosis of the retrieved sub-case is benign. THEN the diagnosis of the new sub-case is 40% benign and 60% malignant.

Our general form of the transformational rules is:

IF $(N_1, R_1, \ldots, N_i, R_i, \ldots, N_n, R_n)$
SATISFY $(condition_1, \ldots, condition_n)$
AND **Retrieved - diagnosis**
THEN **New - diagnosis**

Where,
N_i is the malignant feature ***i*** in the new *Symptoms-sub-case.*
R_i is the corresponding malignant features ***i*** of the retrieved *Symptoms-sub-case.*
The conditions ***(condition1 … condition$_n$)*** check which features of the new sub-case and of the retrieved sub-case exist and which features don't exist to adapt the retrieved diagnosis, such that:

$$N_i, R_i \in \{0,1\}, i = 1 \ldots n$$

For each phase do
i) Extract the Adaptation knowledge.
ii) Apply the IF-THEN transformational rules.
iii) Map transformational rules into numerical or binary vectors.
iv) Train and Adjust the ANN on the mapped transformational rules.
v) Find the new sub-case diagnosis.

Figure 3. The main adaptation algorithm

At the ***Suspicion-phase*** do
1. Use the main adaptation algorithm to adjust the *Suspicion-ANN*.
2. - Input the features of the new *Symptoms-sub-case* to the adjusted *Suspicion-ANN*.
   - Input the features of the retrieved most similar *Symptoms-Sub-case* with its $Diagnosis_1$ to the adjusted *Suspicion-ANN*.
3. Output *Suspicion-diagnosis* to diagnose thyroid cancer suspicion.

At the ***To-Be-Sure-phase*** do
4. Use the main adaptation algorithm to adjust the *Type-ANN*.
5. - Input the features of the *new Scans-sub-case* to the adjusted *Type-ANN*.
   - Input the features of the retrieved most similar *Scans-Sub-case* with its $Diagnosis_2$ to the adjusted *Type-ANN*.
6. Output *Type-diagnosis* to diagnose thyroid cancer type.

At the ***Stage-phase*** do
7. Use the main adaptation algorithm to adjust the Stage-ANN.
8. - Input the features of the new *Stage-sub-case* to the adjusted *Stage-ANN*.
   - Input the features of the retrieved most similar *Stage-Sub-case* with it $Diagnosis_3$ to the adjusted *Stage-ANN*.
9. Output the *Stage-diagnosis* to diagnose thyroid cancer stage.
10. Composite the New Case Diagnosis.
11. Evaluate by the Expert doctor.

Figure 4. The algorithm of the hybrid adaptation model

Retrieved-diagnosis is the diagnosis of the *retrieved sub-case* at a specific phase.
New-diagnosis is the diagnosis of the *new sub-case* at a specific phase adapted from *Retrieved-diagnosis* based on the conditions *(condition$_1$ ... condition$_n$)*.

-At the Suspicion-Phase

Retrieved-diagnosis $\in$ {benign, malignant, benign with malignant}.
New-diagnosis = $\mathbf{CF_b}$ and $\mathbf{CF_m}$ such that:

CF_b, CF_m are the assigned certainty factors [11]. Certainty factors range is [-1,1]. However, in our domain they are modified to reflect the expert doctors' feeling of thyroid cancer suspicion, such that $CF_b \in$ [0%-50%], $CF_m \in$ [50%-100%].

At To-Be-Sure-Phase

Retrieved-diagnosis, ***New-diagnosis*** $\in$ {papillary, medullary, anaplastic, hyper, grave, papillary with hyper, papillary with grave, medullary with hyper, medullary with grave, anaplastic with hyper and anaplastic with grave}.

-At the Stage-Phase

Retrieved-diagnosis, New-diagnosis $\in$ {Stage-I, Stage-II, Stage-III, Stage-IV}.

iii) Map Transformational Rules. The transformational rules of each phase are mapped into numeric or binary vectors. Our general representation form is:

$$[\mathbf{N_1, R_1, \ldots, N_i, R_i, \ldots, N_n, R_n}$$
$$\mathbf{Retrieved\text{-}diagnosis, New\text{-}diagnosis}]$$

where: ***Ni, Ri*** and ***Retrieved-diagnosis*** $\in$ {0,1}, while ***New-diagnosis*** $\in$ {0,1, CF_b and CF_m}.

iv) Train and Adjust the ANN on the mapped Transformational rules. The ANN of each phase is trained on the mapped transformational rules of its phase to learn how to make adaptation. The *ANNs'* type used is the feedforward multi-layer perceptron (MLP) with one hidden layer that is trained with the backpropagation algorithm [9]. During first experiments, the performance was very bad due to training the ANNs' on large number of transformational rules. For example, the rules at the *Suspicion-phase* may have very small dynamic ranges of CF's two similar rules may have one feature difference and a very small CF' s difference, which may be 2%. Many trials (refinements) are necessary to choose the right rules to adjust the *ANNs'*. Refinement steps are done by trial and error under the supervision of our expert doctors. After a number of trials, the topology of the three *ANNs'* at which they have better performance is described in table1.
V) Find the new Sub-case diagnosis. Given the features of a *new sub-case* and of a *retrieved* most similar one. The *ANN* adapts the *Retrieved-diagnosis* to find the *new-diagnosis.*

	Suspicion-ANN	Type-ANN	Stage-ANN
Input layer neurons	38	26	16
Hidden layer neurons	4	5	4
Output layer neurons	2	5	4
Learning rate	0.1		
Momentum	0.7		
Activation function	Tansh		
The transformational rules number trained on.	40	31	10
Root mean square error	0.00001		

Table 1. The topology description of the three ANN's

4. Diagnosis Performance

In our cross-validation test, 80 cases of the 220 definite cases are used for testing and the other 140 cases are stored in the case-memory to be used for retrieval. Also, 300 cases of the 600 indefinite cases are used for testing, while the other 300 cases are stored in the case-memory to be used for retrieval. That is, a total of 380 cases are used for testing and a total of 440 cases are stored in the case-memory to be used for retrieval. Table 2 shows the diagnosis performance (accuracy rate) of our hybrid adaptation model.
The diagnosis performance at each phase is calculated as:

$$\textbf{Phase Accuracy} = \textbf{TC / TT}$$

where ***TC*** is the total number of test sub-cases diagnosed correctly by the MLP of the phase and ***TT*** is the total number of the test sub-cases used for testing the MLP. In order to find the total number of test (new) sub-cases diagnosed correctly by each MLP, the diagnoses' outputs of each MLP are compared with the actual diagnoses and evaluated by our expert doctors. At the *Suspicion-phase*, the certainty factors values of the *Suspicion-ANN* may vary from the actual certainty factors of the expert doctors with a threshold range equals to [-10 - +10]. However, this is true because the explanation of the expert doctors is not exact. For this reason, the diagnoses' outputs of the *Suspicion-ANN* at this phase are mainly evaluated by the expert doctor. At the Stage-phase, the number of the test sub-cases is only 320 sub-cases because there are 60 cases diagnosed as benign diseases. The results show a high accuracy for all the phases.

The overall model diagnosis performance (accuracy rate) calculated as:

$$\textbf{Model Accuracy} = \textbf{MTC / MTT}$$

where ***MTC*** is the total number of test cases diagnosed correctly by the model and ***MTT*** is the total number of test cases. In our model, we consider a test case to be correctly diagnosed only if both its *Type-diagnosis* at the *To-Be-Sure* phase and its *Stage-diagnosis* at the Stage phase are diagnosed correctly. The overall accuracy rate of the model is 99.47% when it is tested with the 380 test cases.

5. Adaptation Performance

We test the adaptation performance of each phase of our model. Adaptation performance means the ability of our model to give the same diagnosis performance results at different similarity ranges, for the same test cases. Table 3 shows the similarity ranges that can be achieved at each phase. As shown, at the *Suspicion-phase*, the minimum similarity threshold between each test *Symptom-sub-case* and its corresponding retrieved *Symptoms-sub-case* is 35%, taking into consideration a variation in the certainty factors values with a threshold range equals to [-10 - +10]. At the *To-Be-Sure-phase*, the minimum similarity threshold between each test *Scans-sub-case* and its corresponding retrieved *Scans-sub-case* is 30%. At the *Stage-phase*, the minimum similarity threshold between each test *Stage-sub-case* and its corresponding retrieved *Stage-sub-case* is 55%.

Phase	No. of test sub-cases	No. of test sub-cases diagnosed by each MLP.	Accuracy Rate
Suspicion	380	378	99.47%
To_Be_Sure	380	379	99.73%
Stage	320	319	99.68%

Table 2. Diagnosis performance

Phase	Test subcases	Similarity Range	Accuracy Rate
Suspicion	380	[100%-35%]	99.47%
To_Be_Sur	380	[100%-30%]	99.73%
Stage	320	[100%-55%]	99.68%

Table 3. Adaptation performance

6. Conclusion and Future Work

This study illustrates a prototype hybrid expert system namely Cancer-C for cancer diagnosis, which is applied to thyroid cancer. Cancer-C is based on the case based reasoning methodology. The main aim of this research is to develop a new adaptation model, which uses much less adaptation knowledge. The model combines transformational and hierarchical adaptation techniques with ANN's. The ANN's are trained on transformational rules to learn how to make adaptation to avoid training with retrieved patient cases, which may have very similar features but completely different diagnoses. Also, to avoid the problem faced with other medical CBR systems that uses a large set of transformational rules. A high performance rate of diagnosis is achieved at different ranges of similarities between the new case and the retrieved case. The average of the similarity ranges is [40%-100%]. The model consists of a hierarchy of three phases, which simulates the expert doctors reasoning phases for cancer diagnosis. CF's are also added to reflect our expert doctors' feelings of cancer suspicion.

In future work the system will be implemented using rest of case-based reasoning techniques and fuzzy logic. The model will be tested for other learning methods that will be compared to the ANN approach. It will be also tested for other cancer diseases. More experts will be involved and more cases will be collected for better accuracy and reliability of diagnosis decisions.

Acknowledgement

Thanks especially to Dr. Reda Mostafa professor of oncology surgery in Ain Shams University, Dr. Enas M. supervisor of the statistical department in the National Cancer Institute of Egypt and Dr. Amr Ayoub a radiologist in the Cairo Center of Scans.

References

[1] Aamodt A.; and Plaza E., 1994. Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches. *AICOM,* 7(1): 39-52.

[2] Kolodner J.L., 1993. *Case-Based Reasoning*. San Francisco, Calif.: Morgan Kaufmann Publishers.

[3] Wilk W., and Bergmann R., 1998. Techniques and Knowledge used for Adaptation during Case-based Problem Solving. *In IEA-1998, 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert systems*. Springer Press.

[4] Juan Corchado and Brian Lees, 2001. A Hybrid Case-Based Model for Forecasting. *In Applied Artificial Intelligence: An International Journal* 15(2): 105-127.

[5] E.A.M. Lotfy and A.S. Mohamed, 2000. Applying Neural Networks in Case-Based Reasoning Adaptation for Cost Assessment of Steel Buildings. *Quo Vadis Computational Intelligence, New Trends and Approaches in Computational Intelligence*, 92-105, Physica-Verlag, A Springer-Verlag Company.

[6] Lothar Gierl, Mathias Bull and Rainer Schmidt, 1998. CBR in Medicine. In Mario Lenz, Brigitte Bartsch-Sporl, Hans-Dieter Burkhard, Stefan Wess (Eds.): *Case-Based Reasoning Technology, From Foundations to Applications*, *260-270,* Springer, Berlin.

[8] Luigi Portinale and Pietro Torasso, 1996. On the Usefulness of Re-using Diagnosting Solutions. In Proceedings of 12th European Conference on Artificial Intelligence (ECAI'96), 137-141. Edited by W. Walhster, John Wiley and Sons.

[9] Simon Haykin, 1999. *Neural Networks*. Prentice Hall, USA, second edition: 156-256.

[10] P.J.F. Lucas, 2001. Certainty-factor-like-structures in Bayesian belief networks. *The Knowledge-Based Systems Journal,* 14(7): 327-335. www.csd.abdn.ac.uk/~plucas/cfbayes-kbs.pdf

[11] Salem A.B.M., Roushdy M., and El-Bagoury, B. M., 2001. An Expert System for Diagnosing of Cancer Diseases. In Proceedings of the 7th International Conference on Soft Computing, MENDAL 2001, 300-305.

[12] Thyroid Disease Manager, September 2002, http: // www.thyroidmanager.org, www.cytology.com www.kfshrc.edu.sa , www.endocrineweb.com

[13] Kohavi, R., Langley, P., and Yun, Y., 1997. The utility of feature weighting in nearest-neighbor algorithms. *In Proceedings of the Ninth European Conference on Machine Learning. Prague:* Springer-Verlag.