Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006



# Prototyping Virtual Cancer Therapist (VCT): A Software Engineering Approach

S. Liu, J-L Gaudiot, and V. Cristini
The Henry Samueli School of Engineering, U.C. Irvine, U.S.A., {shaoshal, gaudiot, vcristin}@uci.edu

*Abstract*— **This paper proposes the Virtual Cancer Therapist (VCT), a scalable and robust cancer expert system that makes cancer diagnosis, recommends therapy plans, and simulates therapy plans *in silico*. This system consists of an evaluation core that makes prognosis and chemotherapy simulations, a biomedical database that supports therapy planning, and an optimizer module that makes cancer diagnosis and produces queries for the optimal therapy plans. With the support of its patient record database and simulation core, VCT can also be used to establish an *in silico* drug discovery standard that dramatically reduces the drug discovery timeline and cost. The prototype of VCT presented in this paper has not only demonstrated the capability of VCT but also identified problems that need to be addressed in the next cycle of development.**

## I. INTRODUCTION

Cancer is an extremely complex process that involves many different protein pathways. For example, two genes, BCL2 and BAX are competing in a cancer cell [13]. BCL2 is an apoptosis-suppressing gene that prevents cell death; on the other hand, BAX is an apoptosis-promoting gene that promotes cell death. When a drug is given to the patient to up-regulate the activity of BAX, it might not be effective since the competing gene BCL2 is suppressing apoptosis. Therefore, a cocktail approach, which consists of a chemical to promote the BAX gene expressions and another chemical to suppress the BCL2 gene expressions, is required to control the tumor growth. Unfortunately, since there are thousands of different chemicals and protein pathways that are associated with cancer, it is a very challenging task even for a cancer specialist to identify the optimal combination of different chemicals to cure cancer. Nevertheless, computers can handle data-intensive search operations, such as this one, much faster and more accurate than its human counterparts do.

Many efficient cancer expert systems exist, such as NED [1], RFE-SVM [2], and Padalin-DEC [3]. They have been constructed for the general analysis of only one aspect of the clinical cancer therapy process, such as cancer diagnosis or drug scheduling. However, in order to be practical, a cancer expert system must be able to address different aspects of the clinical cancer therapy process, which includes cancer diagnosis, tumor growth prognosis, drug scheduling, and therapy plan simulations.

This paper proposes the Virtual Cancer Therapist (VCT), a scalable and robust cancer expert system that diagnoses cancer, recommends anti-cancer therapies, and simulates the efficacies of the therapy plans *in silico*. With the support of its patient record database and simulation core, VCT can also be used to establish an *in silico* drug discovery standard to reduce the drug discovery timeline and cost.

## II. THE DESIGN OF VCT

As shown in Figure 1, the system architecture of VCT consists of three main modules: an integrated biomedical database (BMDB), an optimizer engine (OE), and an evaluation module (EM). After a cancer patient's data is inputted into VCT, these three modules cooperate in the following steps:

1) OE processes the input data, makes a diagnostic decision, and formulates queries to search BMDB for the optimal therapy plan.
2) OE passes the therapy plan and the patient's information to EM, which makes a tumor growth prognosis and a simulation of the therapy plan.
3) The simulation results are passed to the oncologist. After the clinical trials, clinical results are returned to OE. With machine learning approaches, the simulation and clinical results may be utilized to optimize the query engine and the simulation core. This will be addressed in future projects.
4) The simulation and clinical results are stored in BMDB.

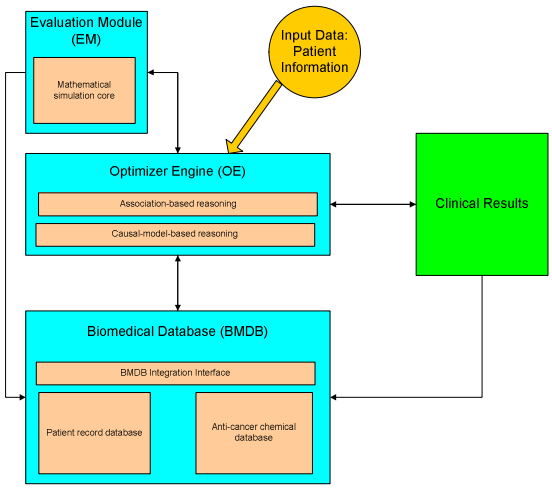


Figure 1: VCT system architecture

### A. Biomedical database (BMDB)

In order to support the cancer expert system, a highly-integrated database that stores complete and accurate biomedical data is required. In this design, BMDB consists of

two databases: the patient record database and the anti-cancer chemical database. The patient record database serves two purposes. First, it stores a patient's cancer medical information as part of the patient's overall medical record. As addressed in [4], a patient's digital medical data records are usually distributed among several hospitals and have different data formats, thus it is very hard to retrieve. To address this problem, the cancer patient records in VCT are stored in the XML format so that they can be easily transformed into other data formats. Second, this information serves as the "experiences" of VCT, which can be used to support the association-based reasoning discussed in the next section. The anti-cancer chemical database stores the anti-cancer drugs' pharmacodynamic information, such as drug toxicity and drug targets. It is used to support the therapy plan queries initiated by the optimizer engine.

### *B. Optimizer engine (OE)*

Each cancer case is usually associated with genetic disorders that can be revealed from lab tests. Based on his/her experiences and knowledge, an oncologist makes diagnostic decisions according to the findings he/she gets. Following the same line of Jang's approach [5], this paper proposes to solve the cancer diagnosis problem using hybrid reasoning that makes use of association-based reasoning (experiences) in conjunction with causal-model-based reasoning (knowledge). Association-based reasoning is efficient because it makes diagnostic decisions based on past experiences. However, it is not robust since it cannot deal with previously unseen cases. On the other hand, although not efficient, causal-model-based reasoning makes diagnostic decisions based on the complicated cancer diagnosis protocols, thus it is robust. By combining these two methods, VCT has the potential of solving complex cancer problems in a robust and efficient way.

### *C. Evaluation module (EM)*

EM simulates tumor growth and chemotherapies by applying the mathematical biology approaches developed by Cristini et al. [6, 7, 8, and 9]. These models are developed by incorporating parameters obtained from *in vitro* and *in vivo* cancer invasion and drug response experiments into the mathematical equations that are used to describe the behaviors of cancer. Especially, the multi-scale, multidimensional tumor simulator [7] has the capability of showing cancer progression through the stages of diffusion-limited dormancy, vascularization and rapid growth, and tissue invasion. In addition, this model has also been successfully used to simulate the delivery of anti-cancer chemicals into tumors [8].

## III. PROTOTYPING: A SOFTWARE ENGINEERING APPROACH

Unlike other expert systems, VCT is an extremely complex system that involves many modules. As it evolves, the system will incorporate more and more functional units, such as a machine learning algorithm to optimize the query engine and EM. In order to develop VCT in a scalable and robust fashion, a rigorous software development approach, the spiral model, is applied [10].

As shown in Figure 2, at each cycle of development, the development objectives and constraints are first defined. Then, a prototype of the system is constructed and evaluated. The purpose of prototyping is to establish the system requirements and to identify potential problems for the next development cycle. This approach is scalable because at each development cycle, new system components can be added and existing components can be modified. It is also robust since the system prototype at each cycle is carefully examined and evaluated to identify any potential problems.

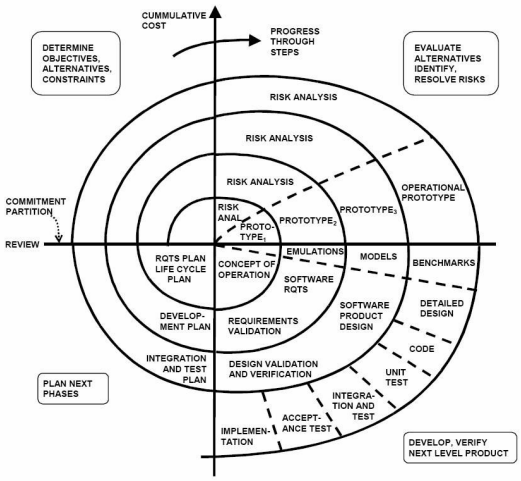


Figure 2: Spiral software development model [10]

With the objective of demonstrating the capability of VCT and identifying potential problems, a first prototype of VCT has been developed. For the evaluation module, a tumor is treated as a collective body, characterized by its size, apoptosis rate, and mitosis rate. Unlike normal cells, which tend to have equal mitosis and apoptosis rate, the mitosis rate of cancer cells is much higher than the apoptosis rate so that the tumor grows without control. In order to control tumor growth, chemotherapies are applied to either reduce the mitosis rate or increase the apoptosis rate. Using Cristini's approach [6], tumor growth and chemotherapy can be described by equation 1. In this model, U(R) is the tumor growth speed, R is the tumor size, a is the apoptosis rate, and m is the mitosis rate.

$$U(R) = \frac{dR}{dt} = -\frac{a}{3} * R + m * \left(\frac{1}{\tanh R} - \frac{1}{R}\right) \qquad (1)$$

In this prototype, a small anti-cancer drug database has been established. Each entry in this database has the following structure:

*(drug ID, drug name, drug target ID, drug target name, change of mitosis rate, change of apoptosis rate)*

To populate this database, drug information is taken from DrugBank [11] and transformed. For data transformation, the drug's mechanism of action is used to determine whether the drug decreases the mitosis rate or increases the apoptosis rate. Also, since the absorption rate and the toxicity of a drug directly affect its efficacy, these two parameters are used to decide how much change of mitosis or apoptosis rate it introduces.

Similarly, an artificial patient record database has been established and it has the following structure:

*(patient ID, original tumor size, mitosis rate, apoptosis rate, target genes, drug plan, prognosis, estimated final tumor size after applying the therapy plan)*

The main function of the patient records in VCT are to serve as "experiences" of the optimizer engine.

The combination of association-based reasoning (experiences) and causal-model-based reasoning (knowledge) has been implemented in this optimizer engine (OE). For each new case, OE first checks whether the patient record database contains a similar case (which has the same genetic mutations). If this is the case, then the same therapy plan is retrieved and applied. Otherwise, OE searches for the best therapy plan based on a set of predefined rules. There are two rules in this implementation:

*Rule 1: always pick the drug that induces the biggest change of mitosis or apoptosis rate.*
*Rule 2: when dealing with mutations in competing genes, always pick multiple drugs, each deals with one mutation.*

According to Equation 1, both the final tumor size and the speed of tumor growth depend on the tumor cells' mitosis rate and apoptosis rate. Chemotherapy works by either decreasing the mitosis rate or increasing the apoptosis rate. Therefore, by imposing Rule 1, the queries always search for the optimal therapy plan. Also, as discussed at the beginning [14], when competing genes are mutated, applying one drug that acts on only one gene has little effect because the other gene is still functioning. Rule 2 ensures that the optimal therapy plan, which consists of chemicals to deal with all of these mutations, can still be identified under such situations.

## IV. NUMERICAL EXPERIMENTS

In order to demonstrate the mechanism of VCT and to evaluate the performance of this prototype, three experiments are carried out. The first two cases test VCT's ability to handle different cancer cases, while the third case tests the efficiency of the association-based reasoning. All the experiment parameters and experiment results are dimensionless for calculation simplicities [6].

Experiment 1 deals with a tumor of size 20 with mutation at the RB gene. The mitosis rate of this tumor is 1 and the apoptosis rate is 0.08. After OE takes the case 1 input, it searches the drug database and obtains the following candidate drug plans:

| *Drug name* | *Target* | $\Delta mitosis$ | $\Delta apoptosis$ |
|---|---|---|---|
| *Amsacrine* | *DNA* | *0.0* | *0.05* |
| *Chlorambucil* | *DNA* | *0.0* | *0.1* |
| *Cisplatin* | *DNA* | *0.0* | *0.13* |
| *Epirubicin* | *DNA* | *0.0* | *0.13* |
| *Tretinoin* | *RB* | *-0.16* | *0.0* |

Some drugs, such as Cisplatin, increase the apoptosis rate of the tumor cells, thus controlling tumor growth by increasing the number of cell deaths. Other drugs, such as Tretinoin, decrease the mitosis rate of the tumor cells to slow down the tumor growth speed. Among these candidate drugs, Tretinon introduces the biggest change to the mitosis rate of the tumor, thus it is the best performing drug according to Rule 1. As a result, OE picks Tretinon and sends its information to EM for simulations, which return the following results:

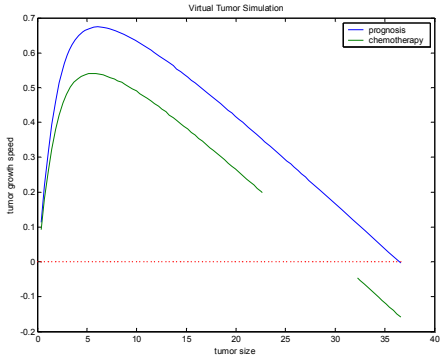


Figure 3: case 1 simulation results

In Figure 3, the prognosis (solid line) predicts that the tumor size will increase to 36 while after applying chemotherapy (dotted line) the tumor size will only increase to 30.

Unlike experiment 1, experiment 2 involves mutations in competing genes. This tumor of size 15 has mutations in both the BAX and the BCL2 genes. The mitosis rate of this tumor is 1 and the apoptosis rate is 0.03. For this case, OE returns one combination of drugs, each deals with one mutation.

| *Drug name* | *Target* | *Δmitosis* | *Δapoptosis* |
|---|---|---|---|
| *Minoxidil* | *BAX* | *-0.13* | *0.0* |
| *Paclitaxel* | *BCL2* | *0.0* | *0.18* |

Then this therapy plan is passed to EM for simulations, which return the following results:

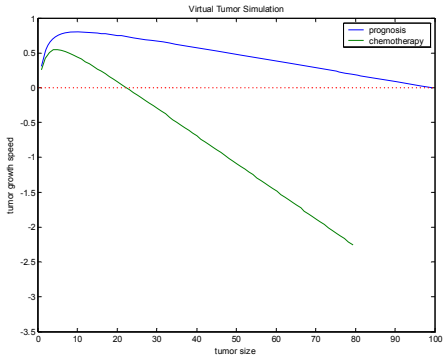


Figure 4: case 2 simulation results

For this case, the prognosis predicts that the tumor size will increase to 99 and the chemotherapy is able to control to size to within 22.

Experiment 3 compares the efficiency of the association-based reasoning (experience) and the one of the causal-model-based reasoning (knowledge). For the association-based reasoning, OE searches the past patient records and retrieves therapy plans that have been applied on similar cancer cases. On the other hand, the causal-model-based reasoning follows a set of rules to form a query to search the drug database for the optimal therapy plan. The experiment result shows that the association-based reasoning is four times more efficient than the causal-based reasoning when searching for the same optimal therapy plan. In this prototype, both reasoning approach take less than one second. However, as the size of the database scales up, the system runtime will become a very important issue.

*Association-based reasoning: 17,437,979 ns*
*Causal-model-based reasoning: 70,629,601 ns*

## V. CONCLUSION

This paper proposes the Virtual Cancer Therapist (VCT), a scalable and robust cancer expert system. Through the first prototype of VCT, several problems have been identified. First, chemical side effects are not considered here. In the next development cycle we plan to implement the approach used in [12], which imposes two drug dose constraints, the daily drug dose and cumulative drug dose to prevent drug overdosing. Second, in order to obtain more accurate and realistic simulations, a 3D tumor simulator [9] will be incorporated into VCT in the next development cycle.

Other than being a clinical decision-support system, VCT can also be used to establish an *in silico* drug discovery standard. Currently, the average time-to-market for a new drug is 15 years and on average it costs $0.5 billion. For every 5000 compounds filed for FDA drug application, only one drug makes its way to market. This cost can be dramatically reduced if an *in silico* trial standard is established. VCT is perfectly fit for this task. As VCT operates, more and more patient records are added to the patient record database. Thus when a new drug is proposed, it can be tested against virtual patients (from the patient record database) with different phenotypes and genotypes through the VCT simulation core at a very low cost. These simulation results can then be used to decide whether the drug should go to clinical trials or not.

## ACKNOWLEDGEMENTS

This work was partly supported by the National Science Foundation under Grant No. CCF-0541403. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

1. Zhou Z., Jiang Y., Yang Y., Chen S., Lung cancer cell identification based on artificial neural network ensembles, Artificial Intelligence in Medicine 24 (2002) 25-36.
2. Huang T-E., Kecman V., Gene extraction for cancer diagnosis by support vector machines—an improvement, Artificial Intelligence in Medicine 35 (2005) 185-194.
3. Tan K-C., Khor E.F., Cai J., Heng C-M., Lee T-H., Automating the drug scheduling of cancer chemotherapy via evolutionary computation, Artificial Intelligence in Medicine 25 (2002) 169-185.
4. Riva A., Mandl K.D., Oh D-H, Nigrin D.J., Butte A., Szolovits P., Kohane I.S., The Personal Internetworked Notary and Guardian, Int. J of Med. Informatics 62:27-40 (2001)
5. Yeona Jang, HYDI: A Hybrid System with Feedback for Diagnosing Multiple Disorders, Ph.D. Thesis, MIT (1993)
6. Frieboes HB, Zheng X., Sun C-H., Tromberg B., Gatenby R., Cristini V., An integrated experimental/computational model of tumor invasion, Cancer Research 66(3):1597-1604 (2006)
7. Zheng X., Wise S.M., and V. Cristini, Nonlinear simulation of tumor progression and interaction with the vasculature via an adaptive finite-element/level-set method. Bulletin of Mathematical Biology (2004)
8. Sinek J., Frieboes H, Zheng X., Cristini V., Two-dimensional chemotherapy simulations demonstrate fundamental transport and tumor response limitations involving nanoparticles, Biomedical Microdevices 6:4 297-309, 2004.
9. Wise SM, Frieboes HB, Cristini V., Three-dimensional diffuse-interface simulation of multispecies tumor growth-I: numerical method. Bulletin of Mathematical Biology (in review)
10. Sommerville I., Software Engineering, Addison-Wesley, Reading, MA, 1998.
11. DrugBank: http://redpoll.pharmacy.ualberta.ca/drugbank/
12. Martin RB. Optimal control drug scheduling of cancer chemotherapy. Automatica 28:1113-23 (1992)
13. Bougras G., Cartron P-F, Gautier F., Martin S., LeCabellec M., Meflah, K., Gregoire M., and Vallette F.M. , Opposite role of Bax and BCL-2 in the anti-tumoral responses of the immune system, BMC Cancer 2004; 4:54