

Iliad 醫學專家系統診斷可靠性之分析研究

張顯洋 黃明和* 蔣台舟** 莊玉瑛 黃淑娟*

大葉工學院資訊管理學系 * 彰化秀傳醫院 ** 板橋亞東醫院

摘 要

醫學專家系統的診斷可靠性，必須藉助實際的病歷來測試與修正，達到一定的可信度之後才能被國內的醫生接受採用。因此，本研究計劃乃是針對上述的動機，在引進 Iliad 這套醫療診斷專家系統 (Medical Diagnosis Expert System) 之前，利用國內醫院的病人病歷資料為測試樣本，以評估 Iliad 專家系統所做診斷的可靠性，及比較國內醫生所做診斷的差異性 (Difference)。根據測試的結果，Iliad 專家系統針對心臟科疾病的診斷可靠性高達 86.7%，可見其知識庫及推論邏輯的診斷方式，仍然與國內醫生非常接近，唯一不同的是罹病可能比率 70.2% 則有偏低的趨勢；此現象乃是 Iliad 專家系統的知識庫中所儲存的框架，其統計數據受到地域、種族、時間、氣候等環境因素的影響，及 Iliad 的診斷技巧與國內醫生的診斷習慣稍有不同所致。因此在使用 Iliad 專家系統之前，必須先調整知識庫中的各種症狀及統計數據，才能獲得可靠的診斷結果。

關鍵詞：人工智慧，專家系統，知識庫，推論邏輯

Abstract

Note that the accuracy and reliability of knowledge base are major concerns for the developers and users of expert system. The purpose of this research is to evaluate whether Iliad is appropriate to be used in Taiwan area. Specifically, we used real patient cases in a local hospital to evaluate Iliad's diagnostic performance. The set of cases used were selected at random from the patient medical record room. In addition, during the evaluation process, we compared the diagnostic differences between the performances of Iliad and doctors to decide if Iliad's knowledge representation is applicable for us to establish the knowledge base. In the light of test results, the reliability of Iliad as applied to diagnosis of heart diseases is as high as 86.7%. Hence such diagnoses by means of knowledge base and inference logic are very close to those made by doctors in the selected hospital in Taiwan. The only difference is that the disease posterior probability is considerably lower. This should be attributed to the frames in the knowledge base of Iliad with statistics affected by geography, races, time, weather, etc. and to the Iliad diagnostic skills, different from those of doctors in our country. Therefore, before using Iliad, it is suggested to adjust various symptoms and related statistics in knowledge base to generate more reliable diagnostic results.

Keywords: Artificial Intelligence, Expert System, Knowledge Base, Inference Logic

表一：急性心肌梗塞 (Acute MI) 的或然率框架

Title: Acute MI		
Prevalence: 0.037	P(F D)	1-P(NF ND)*
Serial ECG shows evolving patterns consistent with acute MI	.83	.001
or		
ECG compatible with acute MI	.70	.02
.Myocardial enzyme release	.95	.015
or		
Serum LDH increased 2 to 8 days after onset of chest pain	.86	.10
or		
Serum CPK increased 6 hours to 3 days after onset of chest pain	.90	.15
.Infarction chest pain	.85	.03
or		
.Unstable anginal pain	.40	.05
or		
.Typical anginal pain	.20	.02
.Risk of coronary artery disease	.85	.16
.Vasovagal reaction	.40	.10
or		
.Sympathetic reaction to stress	.40	.10
.Left-sided heart failure	.40	.11
General information: age is == years		
>= 0	.01	.19
>= 45	.30	.20
>= 65	.68	.51

*P(F|D)= 某一症狀的敏感性比率

P(NF|ND)= 某一症狀的特異性比率

壹、研究動機及目的

一、Iliad 醫學專家系統

Iliad 是由美國猶他大學醫學資訊系 (Department of Medical Informatics) 所發展出來的醫療診斷專家系統 (Medical Diagnosis Expert System)，它具有醫療諮詢 (Consultation Mode) 和模擬診斷 (Simulation Mode) 兩種基本的功能 [Warner, 1988]。

Iliad 專家系統的架構與一般的專家系統相同，主要是以一個醫學知識庫 (Medical Knowledge Base) 和一個邏輯推論機 (Logic Inference Machine) 所構成的診斷系統。知識庫中目前含有大約 1500 個關於心臟、肺臟、腸胃、腎臟、血液、神經、內分泌和傳染病等疾病的診斷方法 [2]。

Iliad 專家系統知識庫中的醫學知識表示法 (Medical Knowledge Representation)，是以人工智慧中常用的框架法 (Frame) 來表現各種疾病的診斷技巧，其中一種稱為或然率框架法 (Probability Frame)，另外一種稱為

布林框架法 (Boolean Frame) [Yu, 1988]。在每一個或然率框架中除了含有某種疾病的盛行率 (Prevalence of Disease) 之外，並列出與這種疾病有關的診斷症狀 (Symptom)，及每一個症狀與這種疾病之間的關係，也就是敏感性 (Sensitivity) 和特異性 (Specificity) 這兩種比率。例如表一就是一個診斷急性心肌梗塞 (Acute MI) 的或然率框架。

在或然率框架中，各個症狀之間的條件必須是獨立的 (Conditionally Independent)，否則要將這些條件相關 (Conditionally Dependent) 的症狀直接以 “or” 的方式並列，或者是另外列在布林框架中，而以邏輯運算符號 (Logical Operator) 來表示各個症狀之間的關係。例如表二就是一個診斷梗塞性胸痛 (Infarction Chest Pain) 的布林框架，從表二中所列症狀，可見這些症狀在病理生理 (Pathophysiology) 上一般都具有共同的診斷特徵 [Lincoln, 1988]。布林框架都是以單一症狀的方式列於或然率框架中，並以 “.” 的

符號來區別，例如表一中的 Infarction chest pain 和 Typical anginal pain 等等以“.”開頭的症狀都是。

二、Iliad 專家系統的診斷推論

二種框架中的診斷技巧與統計數據 (Statistics) 主要是來自醫院的病歷資料庫 (Patient Data Base)、醫學文獻 (Medical Literature) 和專科醫生的診斷經驗 (Expert Opinion) [Haug 1989]。而 Iliad 專家系統邏輯推論機的推論方式，則是以統計學上非常實用的貝氏定理 (Bayes' Theorem) 為推論基礎，利用或然率框架中每一種疾病的盛行率，和引起該疾病的各個症狀之敏感性和特異性這二種比率，來連續推算在出現某些症狀之後，病人罹患該疾病的可能性 (Posterior Probability)。貝氏定理的公式列於表三。

以急性心肌梗塞 (Acute MI) 為例來說明貝氏公式的用法。假設心肌梗塞這個疾病的盛行率為 3.7%，而典型心絞痛 (Typical anginal pain) 這個症狀與心肌梗塞的敏感性和特異性分別為 20% 和 2%。將這些比率分別代入貝氏公式計算之後為 26.3%，也就是具有典型心絞

痛這個症狀的病人，罹患急性心肌梗塞的比率會從原來的 3.7% 增高為 26.3%。

另外，年齡超過 45 歲的病人與急性心肌梗塞的敏感性和特異性分別為 70% 和 20%，如果要推算年齡超過 45 歲又具有心絞痛的病人，罹患心肌梗塞的比率，就必須以典型心絞痛的罹病率 26.3%，當做新的盛行率代入貝氏公式計算之後變成 55.5%，也就是年齡超過 45 歲又具有典型心絞痛的病人，罹患急性心肌梗塞的比率又從 26.3% 提高為 55.5%。如果病人還有其它與急性心肌梗塞相關的症狀，那麼就依此類推，繼續代入貝氏公式中即可算出最後的罹病比率。

至於布林框架的處理方式，則是先判定病人的症狀是否滿足框架中的邏輯條件？如果完全成立則按照一般貝氏公式來計算；如果部份成立，則利用表四改良的貝氏公式來計算罹病比率。最後，由 Iliad 專家系統綜合病人的年齡、性別、各種症狀與檢查結果所做的鑑別診斷 (Differential Diagnosis)，則以表五的方式顯示，供醫生參考。

表二：梗塞性胸痛 (Infarction chest Pain) 的布林框架

Title: Infarction chest pain

- A. Chest pain, substernal
- B. Chest pain lasting > 15 minutes
- C. Chest pain, heavy/pressure/aching
- D. Chest pain, relief with rest/NTG
- E. Chest pain, pleuritic or positional
- F. Chest wall tenderness

True if: A and B and C and not D or not E or not F

表三：貝氏定理的公式

$$P(D|F) = \frac{P(D) * P(F|D)}{P(D) * P(F|D) + P(ND) * P(F|ND)}$$

其中：

P(D) = 某種疾病的盛行率

P(ND) = 1 - (某種疾病的盛行率) = 1 - P(D)

P(F|D) = 某一症狀的敏感性比率

P(F|ND) = 1 - (某一症狀的特異性比率) = 1 - P(NF|ND)

P(D|F) = 具有某一症狀的病人罹患該疾病的可能性

表四：改良的貝氏公式

$$P(D|F) = \frac{P(D) * P(F|D)^a * P(NF|D)^b}{P(D) * P(F|D)^a * P(NF|D)^b + P(ND) * P(F|ND)^a * P(NF|ND)^b}$$

其中：

$P(NF|D) = 1 - (\text{某一症狀的敏感性比率}) = 1 - P(F|D)$

$P(NF|ND) = \text{某一症狀的特異性比率}$

a = 使布林框架中邏輯條件成立的比率

b = 使布林框架中邏輯條件否定的比率

表五：鑑別診斷 (Differential diagnosis) 的結果

Differential Diagnosis

- (35%) Acute MI
- (7%) Unstable anginal pain
- (3%) Pulmonary embolus
- (1%) Chest Wall Pain
- (<1%) Stroke Secondary to Cardiogenic Embolism
- (<1%) Mitral regurgitation, acute
- (<1%) Pseudomonas Pneumonia (<1%) Aortic stenosis
- (<1%) Achaasia
- (<1%) Gastric cancer
- (<1%) Acute Adrenal Crisis

三、研究目的

Iliad 專家系統使用統計學上的貝氏定理做為推論基礎，已由各種測試中認為是一種最符合診斷推論和最具信服力的醫學專家系統 [Warner, 1989]。在 Iliad 專家系統知識庫中的診斷技巧和統計數據，雖來自前述非常客觀的三種知識來源 (Source)，但是各種疾病的發病會受到地域、種族、國情、時間、氣候等環境因素的影響，因此在使用 Iliad 專家系統時，必須先測試並調整知識庫中的各種症狀及統計數據，才能獲得正確可靠的診斷結果 [Fu, 1990; Chang, 1991; Shortliffe, 1990]。

醫學專家系統的診斷可靠性，必須藉助實際的病歷來測試與修正，達到一定的可信度之後才能被國內的醫生接受採用。因此，本研究計劃乃是針對上述的動機，在引進 Iliad 這套醫療診斷專家系統之前，利用國內醫院的病人病歷資料為測試樣本，以評估 Iliad 專家系統所做診斷的可靠性，及比較國內醫生所做診斷的差異性 (Difference)。同時在測試過程中，觀察 Iliad 專家系統的知識表示方法及邏輯推論的技巧，以便將來能找出一種最適合的知識表示法及推論技巧，來做為建立本土醫學知識庫及開發自有醫學專家系統的基礎。

Iliad 專家系統在美國醫學界已廣被接受，並做為學生診斷訓練的輔助教材和醫院的臨

床診斷的諮詢工具，本人在留學期間，因實際參與 Iliad 專家系統的研發與設計工作，所以希望能引進這套系統，供國內醫學界參考試用，使醫學專家系統的研發工作能在國內播種萌芽，進而帶動醫療科技與資訊科技的密切結合，後來居上也能設計一套適合國人使用的醫學專家系統，以提高醫學院的教學水準及醫生的診斷品質，造福病患。

貳、研究方法及進行步驟

在進行測試專家系統時，最重要的是如何篩選測試樣本，以便得到一個比較客觀的測試結果。尤其是在測試醫學專家系統時，受到龐大醫學領域的影響，病人病歷的篩選和診斷領域的限制，更是決定整個測試成敗的關鍵。本研究計劃是以隨機取樣的方式，在民國八十一年十月至民國八十二年九月，於彰化秀傳紀念醫院中針對心臟科取 20 個病例，再依各個病例分別篩選病人病歷 10 份，共 200 份病歷來進行測試。取樣之後，先按病歷內所記載的相關症狀和各項檢查結果，輸入 Iliad 專家系統內做診斷，然後再將輸入 Iliad 專家系統內的病歷列出，供另外二位專科醫師做診斷，以確定病歷上的診斷結果，最後以 Iliad 專家系統的診斷結果來交互比較醫生所做的診斷結果，並列出三者的診斷結果。

在進行診斷測試過程中，可能遭遇的問題是病人病歷上的診斷結果和另一位專科醫生所做的診斷結果差異過大時，甚至完全不符合時，到底要以那一位醫生的診斷結果作為比較的依據？造成這種狀況的原因可能是：

- (一) 病人病歷的診斷結果錯誤。
- (二) 專科醫生的診斷結果錯誤。
- (三) Iliad 專家系統所列出的病人病歷輸入資料有錯誤或不足，因而造成專科醫生診斷錯誤。

解決上列狀況的方式為：

- (一) 如果是原因一或原因二所造成的錯誤，可以經第三位專科醫師會診之後再做出正確的診斷結果，然後以此作為比較依據，如果會診之後仍無法得到一個正確的診斷結果，則由三位醫生共同來判別 Iliad 專家系統所做的診斷結果是否正確？
- (二) 如果是原因三所造成的錯誤，則在核對修正之後再請專科醫師重新進行診斷，以便得到正確的診斷結果。

參、研究成果

根據預定之研究方法及測試步驟，篩選之後共有 181 份（因某些病例缺乏完整住院病人病歷之故而少於 10 份，但某些病例則因住院病人病歷較豐富而多於 10 份）病歷輸入 Iliad 專家系統接受診斷，並取得各病歷的鑑別診斷結果的列表。如果專科醫生核對之後的診斷疾病，出現在列表的前五名即表示 Iliad 的診斷正確，否則即是不正確。經統計分析之後結果列於表六。

表六中的罹病可能性是表示病人罹患該病例的可能性（即 Posterior Probability），而平均罹病可能性則只計算該病例出現在鑑別診斷列表首位的 Posterior Probability 之平均值。依據表六的測試結果，Iliad 專家系統針對心臟科疾病的平均診斷可靠性高達 86.7%，可見其知識庫及推論邏輯的診斷方式，仍然與國內專科醫生非常接近，唯一不同的是平均罹病可能性比率 70.2% 則有偏低的趨勢；此現象與原先的顧慮不謀而合，也就是 Iliad 專家系統的知識庫中所儲存的框架 (Frame)，其統計數據雖來自前述非常客觀的三種知識來源，但是各種疾病的發病會受到地域、種族、國情、時間、氣候等環境因素的影響。除上述的客觀環境因素之外，根據醫護人員的測試心得報告和診斷特例說明，可見 Iliad 的疾病診斷分類及診斷技巧，與國內醫生的診斷習慣也稍有不同，造成測試過程中輸入病歷資料的困難，甚至發生 Iliad 的診斷結果雖與實際病人病歷的病名不同，其實診斷結果是一樣的情況。

因此在使用 Iliad 專家系統之前，確實須先測試並調整知識庫中的各種症狀及統計數據，才能獲得正確可靠的診斷結果。當然，國人自行建立的本土醫學知識庫及開發適用的醫學專家系統，乃是最為理想及可靠的辦法。

肆、結論與討論

本研究計劃乃在於檢驗 Iliad 醫學專家系統的實際診斷結果，以評估這套系統所做診斷的可靠性及差異性，除可做為引進使用 Iliad 專家系統的參考依據和測試基礎外，也可做為培養知識工程師的搖籃，以訓練建立自有醫學知識庫的能力。

醫學診斷知識庫之建立，可說是人工智慧應用在實際醫療診斷上一個最大的瓶頸。以 Iliad 專家系統為例，其發展重心主要還是集中在診斷知識之萃取、建立、維護及更新上。這些工作都需要藉助知識工程師 (Knowledge Engineer) 與醫學領域專家 (Domain Expert)，耗費大量的時間才能完成，使得醫學診斷知識庫之建立與擴大非常的困難。

為了解決這個困難，美國醫療資訊界已積極的合作進行醫學與電腦科技之整合及標準設定的研究計劃，這些計劃依其特性可分為如下三個層次 [張顯洋，1992；Chang，1991]。

第一個層次是由基層的醫院開始進行，首先建立一個標準化的醫療資訊匯流排 (MIB: Medical Information Bus)，以便醫院中各單位所使用的醫療儀器，能夠在共同的協定 (Protocols) 之下，透過匯流排立即互通所需的資料，而不必人工介入。

第二個層次則是一個中程的目標，由美國國家醫學圖書館 (NLM: National Library of Medicine) 發展一套統一的醫學名詞系統 (UMLS: Unified Medical Language System)，提供各醫院或圖書館間能夠有一個標準的醫學用語來進行溝通或檢索的工作，以方便相關的醫療資訊或文獻之參考使用。

第三個層次則是一個終極目標，也是由美國國家醫學圖書館負責推動的一項超大型計劃，名為整合性學術資訊管理系統 (IAIMS: Integrated Academic Information Management Systems)。這個計劃的主要目的是要整合醫學與電腦科技的資源與資訊，並建立一個學術網路來支援這項整合的工作。

由上面這三項進行中的研究計劃，可知其最後的目標就是要發展一個標準化的學術網路，提供給醫護人員和研究學者一個溝通的介面，來達到資料共通、資訊共享和知識共取的整合環境。那麼，當達成這個目標之後，在建立醫學專家系統的知識庫時，知識取得的瓶頸即

表六：Iliad 的測試結果

Disease Name	Cases	Accuracy(%)	Mean of PP(%)*
CAD	10	100%	69.71%
Cardiomyopathy	16	100%	78.45%
PSVT	10	60%	26.50%
Old MI	10	100%	76.78%
H/T	10	60%	46.00%
MS	10	90%	55.67%
CHF	15	87%	72.50%
ASHD	10	70%	69.50%
HCVD	11	55%	59.50%
AF	10	100%	82.60%
Acute AMI	10	100%	78.78%
Angina Pectoris	10	100%	45.83%
RHD	10	100%	89.60%
Cor Pulmonale	3	100%	65.50%
VPC's	4	75%	62.00%
MR	8	100%	66.83%
MVP	6	100%	44.33%
Palpitation	5	80%	73.00%
S. S. S.	10	80%	78.14%
A. S.	3	100%	73.00%
Average	181	86.7%**	70.20%***

*Mean of Posterior Probability

** $p < 0.05$, *** $p < 0.005$

可突破，不但可以無限擴大而且所獲得的知識也可不受時空限制，都能隨時更新調整，適用各地。

醫學專家系統是整合醫學、統計學、心理學、社會學、圖書館學及電腦科學之最新科技，參與本研究計劃的醫生及醫護人員已獲得此方面之研究經驗及整合能力。而且本研究乃屬醫學專家系統的實際測試，其評估結果及測試模式，可做為往後相關專家系統的測試標準和依據。

證諸在全國各醫院及醫學中心所做的學術推廣，獲致相當不錯的反應，顯見透過這個研究計劃的執行，具有拋磚引玉的效果，來帶動國內醫學界對人工智慧的研究參與，尤其是在國內日漸重視醫療品質的趨勢下，醫學專家系統在醫學院的教學訓練及醫院診斷品質的提昇，確有正面的效果。

我國的醫療體系因沿用歐美的方式，所以針對醫學專家系統這個領域做重點式的推廣，不失為一條可行的方向，而且以歐美目前的研

究方向做為借鏡，他山之石可以攻錯，這也是我國在人工智慧方面的研究後來居上的一條捷徑。希望上述的個人經驗與淺見，能提供國內醫療單位及先進參考，以便在全民健保及全國醫療網實施初期，未雨綢繆儘早規劃並盼早日實現這個理想。

伍、參考文獻

1. 張顯洋，『從 Iliad 醫學專家系統看未來國內外醫療資訊的發展』，中華民國醫療資訊雜誌，第一期，1992，pp. 78-82。
2. "Iliad Users and Technical Manual." Applied Informatics, Salt Lake City, Utah, 1989.
3. FU, L. S., HUFF, S. M. SORENSON, D. K., AND WARNER, H. R. "Toward a Public Domain UMLS Patient Database." In Proceedings of the 14th Annual Symposium on Computer Application in Me-

- dical Care, Washington DC, 1990, pp. 170-174.
4. HAUG, P., CLAYTON, P. D., SHELTON, P., RICH, T., TOCINO, I., FREDERICK, P. R., CRAPO, R. O., MORRISON, W. J., AND WARNER, H. R. "Revision of Diagnostic Logic Using a Clinical Database." Medical Decision Making, Hanley & Belfus, Inc., Philadelphia, PA, 1989.
 5. LINCOLN, M. J., TURNER, C., HESSE, B., AND MILLER, R. "A Comparison of Clustered Knowledge Structures in Iliad and in Quick Medical Reference." In Proceedings of the 12th Annual Symposium on Computer Application in Medical Care, Washington DC, 1988, pp. 131-135.
 6. SHIEN-YOUNG CHANG, KANES: Knowledge Acquisition for a Neuroradiology Expert System. Dissertation for Doctor of Philosophy, University of Utah, 1991.
 7. SHORTLIFFE, E. H., PERREAULT, L. E., WIEDERHOLD, G., AND FAGAN L. M. Medical Informatics: Computer Applications in Health Care. Addison-Wesley Publishing Company, Inc., Reading, Mass, 1990, pp. 620-644.
 8. WARNER, H. R. JR. "Iliad-Moving Medical Decision-Making Into New Frontiers." International Symposium of Medical Informatics and Education, University of Victoria, B. C., Canada, May 1989, pp. 267-270.
 9. WARNER, H. R., HAUG, P., LINCOLN, M., WARNER, H. R. JR., SORENSON, D., FAN, C., BOUHADDOU, O., AND WILLIAMSON, J. "Iliad: A Computer Aided Program for Teaching Differential Diagnosis to Medical Student." In Proceedings of the 12th Annual Symposium on Computer Application in Medical Care, Washington DC, 1988, pp. 267-270.
 10. YU, H., HAUG, P., LINCOLN, M., TURNER, C., AND WARNER, H. "Clustered Knowledge Representation: Increasing the Reliability of Computerized Expert Systems." In Proceedings of the 12th Annual Symposium on Computer Application in Medical Care, Washington DC, 1988, pp. 126-130.