

近紅外線光譜的波長選擇對水稻品種鑑別的影響

劉昌群¹，蕭介宗²，彭敬益³，洪梅珠⁴，沈明來⁵

1. 國立台灣大學生物產業機電工程學系研究生
2. 國立台灣大學生物產業機電工程學系教授，本文通訊作者
3. 前研究助理，國立台灣大學生物產業機電工程學系
4. 台中區農業改良場研究員
5. 國立台灣大學農藝系教授

摘 要

本研究探討五種水稻的近紅外線吸收光譜波長的選擇以該波長所對應之吸收光譜值做為變數對品種鑑別的影響。所使用的三種變數選擇方法分別為逐步排除法、變數之間的相關矩陣法，以及變數在主成分軸上的 Loading 值法。以全部 351 個變數建立判別分析以及倒傳遞類神經網路鑑別模式，平均鑑別率分別為 99.69% 及 97.69%。以逐步排除法選取 62 個變數建立判別分析以及倒傳遞類神經網路鑑別模式，平均鑑別率分別為 98.0% 及 92.76%。以變數之間的相關性選取 62 個變數建立判別分析以及倒傳遞類神經網路鑑別模式，平均鑑別率分別為 90.15% 以及 84.26%。以變數在主成分上的 Loading 值選取 62 個變數建立判別分析以及倒傳遞類神經網路鑑別模式，平均鑑別率分別為 89.38% 及 82.25%。在波長選擇方法中，以逐步排除法挑選的變數其鑑別率優於相關矩陣法及 loading 值法所選取的變數，且具顯著差異。所建立的鑑別模式，不僅能減少變數的數目，同時鑑別率仍可達到使用全部變數的準確性。使用相同的變數時，判別分析法的鑑別能力優於類神經網路法，且具顯著差異。

關鍵詞：近紅外線、類神經網路、水稻、變數選擇、鑑別

THE EFFECT OF WAVELENGTH SELECTION OF NEAR INFRARED SPECTRA ON CLASSIFYING PADDY RICE

Chang-Chun Liu¹, Jai-Tsung Shaw², Keen-Yik Poong³
Mei-Chu Hong⁴, Ming-Lai Shen⁵

1. Graduate Student, Department of Bio-Industrial Mechatronics Engineering, National Taiwan University.
2. Professor, Department of Bio-Industrial Mechatronics Engineering, National Taiwan University, Corresponding Author.
3. Former Research Assistant, Department of Bio-Industrial Mechatronics Engineering, National Taiwan University.
4. Agronomist, Taichung District Agricultural Improvement Station.
5. Professor, Department of Agronomy, National Taiwan University.

ABSTRACT

Five varieties of paddy rice were examined using the reflectance spectra corresponding to a selected wavelength from 1100 to 2500 nm in 3-nm steps to determine the classification rate effect. Three hundred fifty-one variables were used to develop the discriminant analysis and neural network models. The average classification rates were 99.69% and 97.69%, respectively. Sixty-two variables were selected using stepwise discrimination to develop the discriminant analysis and neural network models. The average classification rates were 98.0% and 92.76%, respectively. Sixty-two variables were selected using the correlation matrix to develop the discriminant analysis and neural network models. The average classification rates were 90.15% and 84.26%, respectively. Sixty-two variables were selected by loading the first and second principal components to develop the discriminant analysis and neural network models. The average classification rates were 89.38% and 82.25%, respectively. The stepwise discrimination method was more effective in classifying the five varieties of paddy rice using near infrared spectra.

Keywords: Near infrared, Artificial neural network, Paddy rice, Variable selection, Classification

一、前言

不同的水稻品種除了外觀的差異之外、內部成分亦有所不同。近紅外線光譜已廣泛應用於農產品的品質檢測，對於水稻的主要成分，澱粉和蛋白質的濃度測定亦有相當高的準確性（劉與蕭，1995）、（李與蕭，1996）。以此為基礎藉由測定不同水稻之間主要成分的濃度差異與近紅外線光譜之關係，建立品種辨識的模式，可以迅速鑑別水稻品種，可應用於計價收購作業系統之用。張等人（2000）曾以 1997 及 1998 年產於台中區農業改良場的台農秈 20 號、台中秈 10 號、台農 67 號、台農 70 號、台梗 8 號以及台梗 9 號共 6 種水稻品種，以水稻的近紅外線光譜及類神經網路進行品種鑑別，所得結果為 6 種水稻品種的平均鑑別率為 95.8%，鑑別標準差為 5.3%。李（1998）亦曾使用相同水稻品種的近紅外線光譜對水稻進行品種辨識，所得結果為 6 種水稻品種的平均鑑別率為 88.3%，鑑別標準差為 10.8%。

Delwiche et al. (1995) 以 1100 nm 至 2498 nm 的近紅外線光譜對全粒的硬紅冬麥(hard red winter wheat)及硬紅春麥(hard red spring wheat)作品種分類，並分別以多重線性回歸(MLR)、主成分分析

(PCA)、部分最小平方法(PLSR)及類神經網路(ANNs)對兩種小麥品種做鑑別，這四種方法的平均鑑別率分別為 95.3%、93.0%、98.0% 及 98.3%。Song et al. (1995) 使用穿透式近紅外線光譜 850nm 至 1049nm 的 200 個波長的光譜值為變數，對全粒的兩種小麥的平均鑑別率介於 97% 與 100% 之間，對六種小麥的平均鑑別率為 94.7%。本研究的主要目的在探討近紅外線光譜波長的選擇對全粒的水稻品種辨識的影響，建立非破壞性、無污染的品种鑑別方法，作為稻穀計價收購之用。

二、材料與方法

樣品為產自台灣中部、東部和南部，收穫於 1997、1998、1999 以及 2000 年夏天的台農秈 20 號、台中秈 10 號、台農 67 號、台梗 8 號、台梗 9 號共 5 種水稻品種的一期稻作。各品種的樣本先以風選機風選數次以去除空稻殼、稻穗等雜物，然後將水稻樣本置於 25°C、RH 70% 的恆溫恆濕箱內約 4 天（100 小時），以 Shizudka Seiki CTR-800E 檢測其含水率，調質成 13% 的含水率。然後以 1100nm 至 2500nm 的波長每隔 3nm 對全粒的樣本進行近紅外光照射，每一個樣本共取得

351 個吸收光譜值做為變數，照射後之樣本都不重複使用。再將各年各品種原始光譜資料的界外點，即馬氏距離大於 99.7%信賴區外的樣本加以剔除(Vellwman and Welsch, 1981)。各年度所使用的樣本數如表 1 所示，所有樣本的光譜如圖 1 所示。然後將同一品種的樣本隨機選取約 2/3 的樣本為訓練組，約 1/3 的樣本為鑑別組。台農 20 號、台中 10 號、台農 67 號、台梗 8 號、台梗

9 號訓練組的樣本數依序為 261、259、258、263、261 共計 1302 個，鑑別組的樣本數依序為 130、129、129、131、130 共計 649 個。

由於樣本裡的某些變數對該品種並不具代表性，這樣的變數不僅不具品種區分的效能，且會增加取樣與分析的時間，對於這些不具鑑別能力的變數，擷取變數時事前又無從得知，因此找出真正對品種鑑別具有效能的變數是重要且有意義

表 1 水稻品種及樣本數
Table 1 Varieties of paddy rice and sample sizes

年份 Time	1997	1998	1999	2000	總計 Total
品種 Varieties					
台農 20 號 Tainung Sen 20	98	96	97	100	391
台中 10 號 Taichung Sen 10	98	98	94	98	388
台農 67 號 Tainung 67	96	95	97	99	387
台梗 8 號 Taikeng 8	98	98	100	98	394
台梗 9 號 Taikeng 9	99	97	97	98	391
總計 Total	489	484	485	493	1951

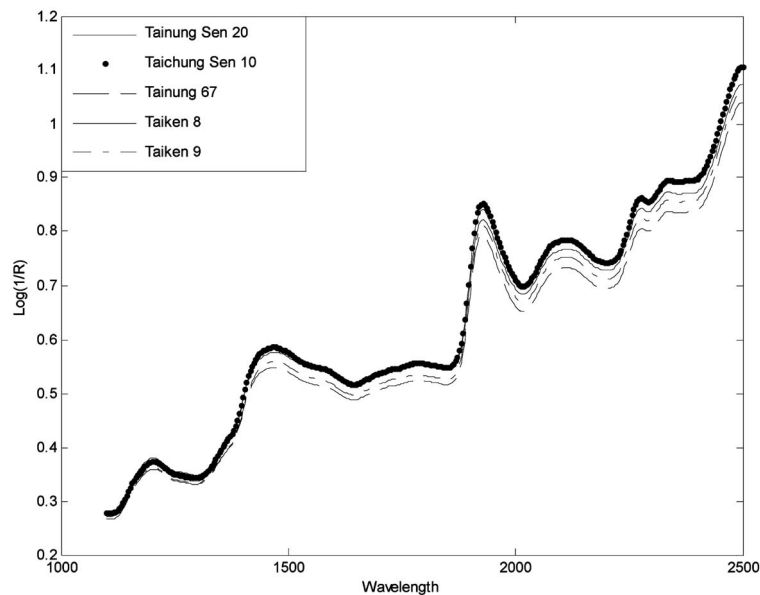


圖 1 五種水稻之光譜
Fig.1 Spectra of five varieties of paddy rice

的。本實驗除了使用全部 351 個吸收光譜值做為變數，並以逐步排除法(stepwise discrimination)、變數的相關矩陣(correlation matrix)法、變數在主成分軸上的 Loading 值等三種方法挑選變數，並比較這三種挑選變數方法的鑑別率，這三種變數選擇的方法分述如下：

(一) 逐步排除法：

先就所有的 351 個變數以 SAS 統計軟體撰寫逐步排除(Stepwise)判別程式，以有進有出的方式選出真正具有品種區分能力的變數。變數選擇的過程如下：首先讀入樣本數據以及品種類別，此時鑑別模式中不含任何變數。然後採取順向選擇法，根據威爾克斯(Wilk's)的 Lambda 值挑選變數進入鑑別模式中，然後被納入鑑別模式的變數同樣的必須再經過威爾克斯(Wilk's)的 Lambda 值衡量該變數對模式鑑別力的貢獻度大小，並將貢獻度最小的變數剔除，如此反覆進行直到沒有新的變數可以加入模式為止(彭，2000)。

(二) 以變數在主成分軸上的 Loading 值挑選變數法：

依據本實驗所使用的主成分分析軟體 Unscrambler 7.6 使用手冊，在主成分分析法中，每一個變數對每一個主成分軸都具有一個 Loading 值，其幾何意義為某一變數在某一主成分軸的 Loading 值表示該變數在該主成分軸的投影值，所以某一變數 Loading 值的大小意指該變數對某一主成分軸的影響度大小。本實驗以 Loading 值挑選樣本時先將所有的樣本數據作主成分分析，主成分分析結果如圖 2 所示。由於第一及第二主成分軸上所攜帶的資料量合計已達全部資料量的 99.7%，所以只取各變數在第一以及第二主成分軸上的 Loading 值(介於-1 和 1 之間)取絕對值之後相加，由大至小排序，然後再依據排序後的 Loading 值大小選出相對應的波長，品種鑑別時依序選擇 Loading 值較大的波長。

(三) 以變數的相關矩陣挑選變數法：

根據 Paliwal et al. (2003)，以此法選擇變數時，首先求出鑑別組樣本數據的相關矩陣，然後將相關矩陣取絕對值，再求出每一行(column)的平均

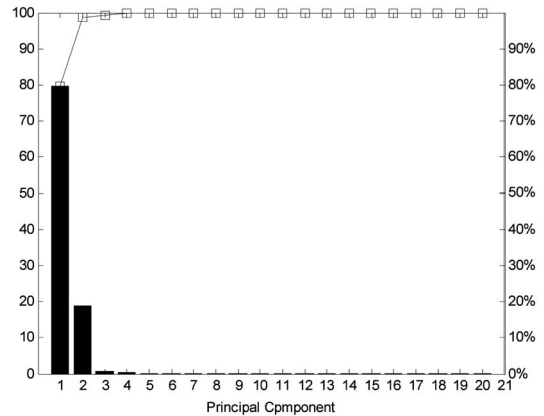


圖 2 主成分及累計資料變異量
Fig.2 Variance explained of principal component

值，此時相關矩陣即由二維陣列化為一維列向量，此列向量內的任一數值即代表該任一變數與其他變數的相關度。再據此將每一個變數與其他變數之間相關性高低由小到大排序，並選出相對應的波長，品種鑑別時依序使用相關度較小的波長，以上過程皆以 MATLAB 6.5 軟體撰寫程式完成。

經過以上三種變數挑選方法所挑選的變數，分別再以 MATLAB 6.5 程式語言自行撰寫的判別分析(discriminant analysis)程式以及帶入倒傳遞類神經網路進行品種鑑別，本實驗所使用的倒傳遞類神經網路為 Neural Works Professional II / PLUS 所提供，使用的變數皆經過正規化，化為 0 與 1 之間的數值([0,1])再帶入網路，並以 Delta Rule 學習法則訓練網路，以 Sigmoid ($y = 1 / (1 + \exp(-x))$) 為非線性轉換函數，網路訓練時具有兩層隱藏層，第一與第二隱藏層內分別具有 40 及 20 個節點，訓練次數為 50,000 次，並以 MATLAB 6.5 撰寫程式以統計及分析鑑別結果。

三、結果與討論

(一) 波長選擇的結果

如表 2 所示，Loading 值法以及相關矩陣法這兩個方法所選出的 62 個波長，除了順序些微不同之外，有 59 個波長共同為這兩個方法所選取，所不同者為 1188、1230、1236nm 為 Loading 法所

選取而未為相關矩陣法所選取，2424、2428、2432nm 為相關矩陣法所選取而未為 Loading 值法所選取。究其原因，因為主成分分析是選取線性獨立的變數組成新而獨立的變數，捨棄線性相依的變數。其目的與相關矩陣法完全相同，所以兩個方法所挑選的結果不僅所選擇的波長具高度的相似性，順序亦僅有少許差異。然而變數對品種鑑別的能力，主要在於變數在品種之間的差異，

變數具較高的線性獨立性質，並不等於具有較佳的品種鑑別能力，上述的兩種變數選擇方法，僅考慮變數之間線性相依的程度，而未考慮品種因素。逐步排除法在挑選變數時即以品種鑑別能力為依據，所挑選出的 62 個變數僅有 5 個變數和 Loading 值法以及相關矩陣法這兩個方法所選出的變數相同，這 5 個變數為 1284、2436、2456、2492、2496nm。

表 2 波長選擇的結果
Table 2 Results of wavelength selection

順序 Order	逐步排除法 Stepwise	Loading 值法 Loading value	相關矩陣法 Correlation matrix	順序 Order	逐步排除法 Stepwise	Loading 值法 Loading value	相關矩陣法 Correlation matrix
1	1248	2496	2500	32	1596	2444	2436
2	1844	2492	2496	33	1600	1168	1160
3	1980	2500	2492	34	1688	1172	2432
4	1944	2488	2488	35	1960	2440	1164
5	1876	2484	2484	36	1680	1176	1168
6	1972	2480	2480	37	1676	1180	2428
7	1864	2476	2476	38	1540	1296	1172
8	1904	2472	2472	39	1572	1300	1176
9	1964	2468	2468	40	1900	1292	1296
10	2280	1108	2464	41	1888	1288	1292
11	1912	1112	2460	42	1928	1304	1288
12	1848	1116	1100	43	1552	2436	1300
13	1548	1104	1104	44	1400	1284	1284
14	1568	1100	1108	45	1588	1308	1304
15	2316	1120	1112	46	1852	1280	1280
16	1580	1124	1116	47	1836	1276	2424
17	1940	2464	1120	48	2244	1260	1276
18	1932	1128	2456	49	1612	1264	1272
19	2428	1132	1124	50	1616	1268	1308
20	1872	1136	1128	51	1956	1272	1268
21	1792	2460	1132	52	2432	1256	1264
22	1812	1140	2452	53	2424	1312	1260
23	1584	1144	1136	54	1892	1252	1256
24	2436	2456	1140	55	2220	1248	1180
25	2492	1148	2448	56	2456	1184	1252
26	1948	1152	1144	57	1648	1316	1312
27	1952	2452	2444	58	1880	1244	1248
28	1968	1156	1148	59	1804	1240	1244
29	1576	2448	2440	60	1644	1320	1316
30	2496	1160	1152	61	2208	1188	1240
31	1668	1164	1156	62	2384	1236	1184

(二) 品種鑑別結果

1. 使用全部 351 個變數之鑑別結果

表 3 為使用全部 351 個變數以判別分析法 (discriminant analysis)，作品種鑑別的結果。在鑑別組的全部 649 個樣本中，只有 1 個台農 67 號和 1 個台稈 8 號樣本被相互誤判，平均鑑別率為 99.69%，標準差為 0.42%。

表 4 為將全部 351 個變數帶入倒傳遞類神經網路做品種鑑別的結果，網路的訓練誤差 (RMS error)=0.08，相關係數(r)=0.92，五個品種的平均鑑別率為 97.69%，標準差為 1.82%。經 z 分數檢定 (沈，1999)， $z=3.17 > 1.96$ 顯示在 95% 的信賴區間內，使用全部 351 個變數時，判別分析法的鑑別能力優於類神經網路法，並有顯著差異。

表 3 判別分析法使用 351 個變數的鑑別率

Table 3 Classification rates of discriminant analysis using 351 variables

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稈 8 號 Taikeng 8	台稈 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	130 100%	0 0%	0 0%	0 0%	0 0%
台中私 10 號 Taichung Sen 10	0 0%	129 100%	0 0%	0 0%	0 0%
台農 67 號 Tainung 67	0 0%	0 0%	128 99.22%	1 0.08%	0 0%
台稈 8 號 Taikeng 8	0 0%	0 0%	0 0%	131 100%	0 0%
台稈 9 號 Taikeng 9	0 0%	0 0%	0 0%	0 0%	130 100%

平均鑑別率 (average classification rate)=99.69%、標準差 (classification standard deviation)=0.42%

表 4 類神經網路法使用 351 個變數的鑑別率

Table 4 Classification rates of neural network using 351 variables

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稈 8 號 Taikeng 8	台稈 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	129 99.23%	1 0.77%	0 0%	0 0%	0 0%
台中私 10 號 Taichung Sen 10	4 3.10%	123 95.35%	2 1.55%	0 0%	0 0%
台農 67 號 Tainung 67	0 0%	0 0%	128 99.22%	1 0.78%	0 0%
台稈 8 號 Taikeng 8	0 0%	0 0%	1 0.76%	129 98.47%	1 0.76%
台稈 9 號 Taikeng 9	0 0%	0 0%	1 0.77%	4 3.08%	125 96.15%

平均鑑別率 (average classification rate)=97.69%、標準差 (classification standard deviation)=1.82%

2. 以逐步排除法選出 62 個變數的鑑別結果

表 5 為以 SAS 軟體所撰寫的逐步排除 (stepwise discrimination) 程式，所選取的 62 個具鑑別能力的變數，然後使用這 62 個變數以判別分析法作品種鑑別，平均鑑別率為 98.0%，標準差為 1.59%。此結果與以類神經網路法使用全部 351 個變數的鑑別率做比較，經 z 分數

檢定， $z=0.38 < 1.96$ 顯示在 95% 的信賴區間內，兩者無顯著差異。此一結果顯示以逐步排除所選取的 62 個變數，然後以判別分析法進行品種鑑別，其鑑別能力與類神經網路法使用 351 個變數相當，使用變數的數目卻只需類神經網路法的 17.66%。

表 6 為以逐步排除法所選出的 62 個變數

表 5 判別分析法使用 62 個由逐步排除所選的變數的鑑別率

Table 5 Classification rates of discriminant analysis using 62 variables selected by stepwise discrimination

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稈 8 號 Taikeng 8	台稈 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	130 100%	0 0%	0 0%	0 0%	0 0%
台中私 10 號 Taichung Sen 10	3 2.33%	126 97.67%	0 0%	0 0%	0 0%
台農 67 號 Tainung 67	0 0%	0 0%	128 99.22%	1 0.78%	0 0%
台稈 8 號 Taikeng 8	0 0%	0 0%	0 0%	127 96.95%	4 3.05%
台稈 9 號 Taikeng 9	0 0%	0 0%	0 0%	5 3.85%	125 96.15%
平均鑑別率 (average classification rate)=98.0%、標準差 (classification standard deviation)=1.59%					

表 6 類神經網路法使用 62 個由逐步排除所選的變數的鑑別率

Table 6 Classification rates of neural network using 62 variables selected by stepwise discrimination

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稈 8 號 Taikeng 8	台稈 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	124 95.38%	6 4.62%	0 0%	0 0%	0 0%
台中私 10 號 Taichung Sen 10	7 5.43%	120 93.02%	2 1.55%	0 0%	0 0%
台農 67 號 Tainung 67	0 0%	2 1.55%	120 93.02%	5 3.88%	2 1.55%
台稈 8 號 Taikeng 8	0 0%	0 0%	0 0%	123 93.89%	8 6.11%
台稈 9 號 Taikeng 9	1 0.77%	0 0%	7 5.38%	7 5.38%	115 88.46%
平均鑑別率 (average classification rate)=92.76%、標準差 (classification standard deviation)=2.59%					

帶入倒傳遞類神經網路做訓練所得的鑑別結果，網路的訓練誤差(RMS error)=0.15，相關係數(r)=0.67，平均鑑別率為 98.0%、標準差為 1.59%。比較表 5 與表 6 的結果，並經 z 分數檢定， $z=4.50>1.96$ 顯示在 95%的信賴區間內在使用較少的變數時，判別分析法的鑑別效果優於類神經網路法並有顯著差異。

表 7 為以 351 個變數的相關矩陣，由小到大依序選出 62 個相關性較低的變數，進行判別分析的結果。此法與逐步排除法所選出的變數，在使用同樣數目的變數之下，平均鑑別率由 98.0%降為 90.15%，同時標準偏差亦由 1.59%增為 4.62%，經 z 分數檢定， $z=6.09>1.96$ 顯示在 95%的信賴區間內，這兩種方法所選取的變數，對品種的區分能力有顯著的差異。其原因應為逐步排除法有完整的理論和步驟以找出最能代表品種間差異的變數，而相關矩陣法僅能確知所選取的變數有較低的線性相依性質，這種變數之間線性相依的程度並無法確定其與品種區分能力的關係。

表 8 為以相關矩陣法由小到大依序選出相關性較低的 62 個變數帶入倒傳遞類神經網路的鑑別結果，網路的訓練誤差 (RMS error)=

0.17、相關係數(r)= 0.61。由於相關矩陣只能解釋變數相互之間相關性的高低，並無法確定所選的變數對品種的區分能力，此結果和使用全部 351 個變數的結果比較，平均鑑別率由 97.69%降為 84.26%，顯示使用的變數過少，鑑別率也較低。

表 9 為 62 個在主成分軸上 Loading 值較大的變數以判別分析法進行品種鑑別的結果。與表 5 以逐步排除法所選取的變數的鑑別效果相比較，不僅鑑別率由 98.0%降為 89.38%，同時標準差亦由 1.59%增為 4.7%，經 z 分數檢定， $z=6.39>1.96$ 顯示在 95%的信賴區間內，這兩種方法所選取的變數，對品種的區分能力有顯著的差異。其原因應為變數在主成分軸上 Loading 值的大小，僅反應出該變數對某一主成分的影響度，並無法確定這個影響度的大小與品種間的區分能力的關係。

表 10 為以變數在主成分軸上的以 loading 值，由大到小依序選出 62 個 loading 較大的變數，帶入倒傳遞類神經網路訓練的鑑別結果，訓練誤差(RMS error)=0.18，相關係數(r)=0.61，平均鑑別率為 82.25%，此一結果和使用全部 351 個變數鑑別率做比較，鑑別率明顯較低。

表 7 判別分析法使用 62 個由相關矩陣法所選的變數的鑑別率

Table 7 Classification rates of discriminant analysis using 62 variables selected by correlation matrix

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稔 8 號 Taikeng 8	台稔 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	119 91.54%	11 8.46%	0 0%	0 0%	0 0%
台中私 10 號 Taichung Sen 10	11 8.53%	118 91.47%	0 0%	0 0%	0 0%
台農 67 號 Tainung 67	0 0%	0 0%	122 94.58%	5 3.88%	2 1.55%
台稔 8 號 Taikeng 8	0 0%	0 0%	6 4.58%	119 90.84%	6 4.58%
台稔 9 號 Taikeng 9	0 0%	0 0%	10 7.69%	13 10.0%	107 82.31%

平均鑑別率(average classification rate)=90.15%、標準差(classification standard deviation)=4.62%

表 8 類神經網路法使用 62 個由相關矩陣所選的變數的鑑別率
Table 8 Classification rates of neural network using 62 variables selected by correlation matrix

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稈 8 號 Taikeng 8	台稈 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	101 77.69%	27 20.77%	0 0%	2 1.54%	0 0%
台中私 10 號 Taichung Sen 10	9 6.98%	119 92.25%	0 0%	0 0%	1 0.78%
台農 67 號 Tainung 67	0 0%	0 0%	96 74.42%	23 17.83%	10 7.75%
台稈 8 號 Taikeng 8	1 0.76%	0 0%	0 0%	129 98.47%	1 0.76%
台稈 9 號 Taikeng 9	0 0%	7 5.38%	17 13.08%	4 3.08%	102 78.46%

平均鑑別率(average classification rate)=84.26%、標準差(classification standard deviation)=10.48 %

表 9 判別分析法使用 62 個由 loading 值所選的變數的鑑別率
Table 9 Classification rates of discriminant analysis using 62 variables selected by loading

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稈 8 號 Taikeng 8	台稈 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	120 92.31%	10 7.69%	0 0%	0 0%	0 0%
台中私 10 號 Taichung Sen 10	11 8.53%	118 91.47%	0 0%	0 0%	0 0%
台農 67 號 Tainung 67	0 0%	0 0%	120 93.02%	6 4.65%	3 2.33%
台稈 8 號 Taikeng 8	0 0%	0 0%	6 4.58%	116 88.55%	9 6.87%
台稈 9 號 Taikeng 9	0 0%	0 0%	11 8.46%	13 10.0%	106 81.54%

平均鑑別率(average classification rate)=89.38%、標準差(classification standard deviation)=4.70%

由圖 3 可以得知，使用相同的變數時，判別分析的鑑別率優於類神經網路的鑑別率，經 z 分數檢定具顯著差異。三種變數選擇法所選取的變數，不論是使用判別分析法或是類神經網路為鑑別工具，以逐步排除法所選取的變數具有最佳的鑑別率，以相關矩陣法所

選取的變數次之，以 loading 值所選取的變數再次之，且經 z 分數檢定三者之間皆具顯著差異。以本實驗與李(1998)，張等人(2000)所作結果比較，李與張皆以同產於台中區農業改良場的台農私 20 號、台中私 10 號、台農 67 號、台農 70 號、台稈 8 號、台稈 9 號共 6 種水

表 10 類神經網路法使用 62 個由 loading 所選的變數的鑑別率
Table 10 Classification rates of neural network using 62 variables selected by loading

	台農私 20 號 Tainung Sen 20	台中私 10 號 Taichung Sen 10	台農 67 號 Tainung 67	台稔 8 號 Taikeng 8	台稔 9 號 Taikeng 9
台農私 20 號 Tainung Sen 20	103 79.23%	24 18.46%	0 0%	3 2.31%	0 0%
台中私 10 號 Taichung Sen 10	11 8.53%	117 90.70%	0 0%	0 0%	1 0.78%
台農 67 號 Tainung 67	0 0%	0 0%	92 71.32%	22 17.05%	15 11.63%
台稔 8 號 Taikeng 8	0 0%	1 0.76%	0 0%	128 97.71%	2 1.53%
台稔 9 號 Taikeng 9	0 0%	11 8.46%	14 10.77%	11 8.46%	94 72.31%

平均鑑別率 (average classification rate)=82.25%、標準差 (classification standard deviation)=11.60%

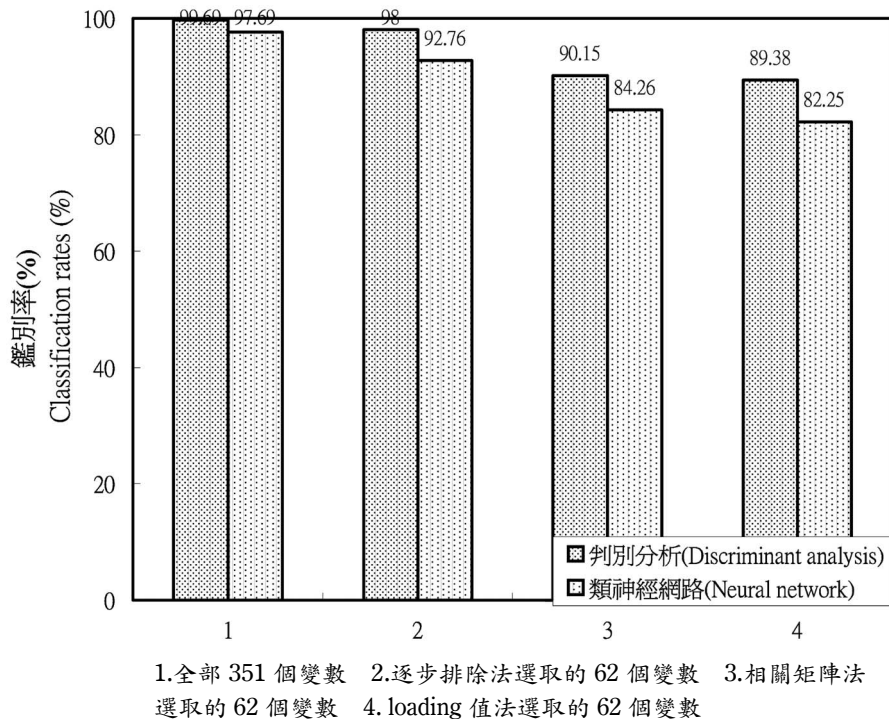


Fig.3 Classification rates with different methods of variable selection (1. total 351 variables 2. 62 variables by stepwise discrimination 3. 62 variables by correlation matrix 4. 62 variables selected by loading)

稻品種的一期稻作的近紅外線光譜辨識水稻品種。就樣本來源而言，李與張所使用的樣本皆僅採自台中區農業改良場一地，本實驗所使用的樣本採自台灣的北、中、南三個地區，是故本實驗的結果應具有較佳的地域適應性。就樣本的採集年份而言，李與張所使用的水稻樣本皆為連續2年所收穫的樣本，本實驗所使用的樣本為連續4年所收穫的樣本，故本實驗應具較佳的穩定度。就分析方法而言，本實驗挑選變數所使用的方法皆能選出特定的變數，捨棄不具品種區分能力的變數，李與張分別使用6個與3個主成分軸進行水稻品種的辨識，這種方式雖亦能降低變數的數目，卻無法確知那些變數具有區分品種的能力。就鑑別效果而言，我們認知同時辨識6種品種的困難度較同時辨識5種品種為高，本實驗因無法取得台農70號的樣本，故僅以相同的品種及品種數做比較，李與張的平均鑑別率分別91.8%及95.86%標準差分別為7.43%及5.91%。本實驗使用62個由逐步排除法所選出的變數以判別分析進行品種鑑別，平均鑑別率與標準差分別為98.0%及1.59%。除了較高的鑑別率之外，較低的標準差顯示本實驗對各品種的鑑別能力較為平均。

四、結 論

以水稻的近紅外線光譜鑑別水稻品種，在波長選擇方法中，逐步排除法所挑選的變數，因挑選的過程同時考慮變數與品種的關係，逐步把對鑑別模式貢獻度最小的變數排除，故鑑別率優於相關矩陣法及loading值法所選取的變數，且具顯著差異。逐步排除法挑選的變數所建立的鑑別模式，不僅能減少變數的數目，同時鑑別率仍可達到使用全部變數的準確性。使用相同數量及相同波長為變數時，判別分析法的鑑別能力優於類神經網路法，且具顯著差異。在取得水稻的光譜之後，就分析過程而言判別分析法所需時間亦短於類神經網路法。

五、誌 謝

本實驗承蒙國科會計劃編號 NSC 89-2313-B-002-241 及 90-2313-B-002-327 之經費支援下方得以完成，謹致謝意。同時感謝在實驗理論及方法上給予諸多建議及協助的台灣大學生物產業機電工程學系周瑞仁教授及林達德教授，中興大學生物產業機電工程學系萬一怒教授及陳育仁博士 (Dr. Yud-Ren Chen, the research leader of Instrumentation and Sensing Laboratory, ARS, USDA)。

六、參考文獻

1. 沈明來。1999。生物統計學入門。3版。177-178。台北：九州出版社。
2. 李汪盛、蕭介宗。1996。資料處理對於稻穀食味主要成分之近紅外線校正線之影響。農業機械學刊 5(4):19-34。
3. 李盛銘。1998。以近紅外線與影像技術鑑別水稻品種。碩士論文。台北：台灣大學農業機械工程研究所。
4. 彭昭英。2000。SAS 與統計分析。11版。台北：儒林圖書有限公司。
5. 張鴻文、蕭介宗、李盛銘、洪梅珠。2000。以近紅外線及影像技術鑑別水稻品種。農業機械學刊 9(4):1-16。
6. 劉民卿、蕭介宗。1995。以近紅外線光譜儀感測稻米之含水率及蛋白質含量。農業機械學刊 4(3):1-14。
7. Delwiche, S. R., Y. R. Chen, and W. R. Hruschka. 1995. Differentiation of hard red wheat by near-infrared analysis of bulk samples. *Cereal Chem.* 72 (3): 243-247.
8. Song H., S. R. Delwiche, and Y. R. Chen. 1995. Neural network classification of wheat using single kernel near-infrared transmittance spectra. *Optical Engineering* 34(10): 2927-2934.
9. Paliwal, J., N. S. Visen, D. S. Jayas, and N. D. G. White. 2003. Comparison neural network and a non-parametric classifier for grain kernel

identification. *Biosystems Engineering* 85(4): 405-413.

10. Velleman, P. F. and R. E. Welsch. 1981. Efficient computing of regression diagnostics. *The Amer. Statistician* 35: 234-242.

收稿日期：2004年 6月 8日

修改日期：2004年 8月13日

接受日期：2005年 3月 1日