

UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE CIENCIAS EXACTAS, FÍSICAS Y NATURALES

DOCTORADO EN CIENCIAS DE LA INGENIERÍA

TESIS DOCTORAL



AMBIENTE AVANZADO PARA
CLASIFICACIÓN ESTADÍSTICA SUPERVISADA
DE IMÁGENES SATELITALES

AUTOR: MG. SUSANA BEATRIZ FERRERO

DIRECTOR: DR. OSCAR HUMBERTO BUSTOS

FEBRERO DE 2018

AMBIENTE AVANZADO PARA CLASIFICACIÓN ESTADÍSTICA SUPERVISADA DE IMÁGENES SATELITALES

por

MG. SUSANA BEATRIZ FERRERO

DR. OSCAR HUMBERTO BUSTOS
DIRECTOR

COMISIÓN ASESORA

DR. OSCAR HUMBERTO BUSTOS
FAMAF-UNC

DR. ERNESTO GUILLERMO ABRIL
FCEFYN-UNC

DR. CARLOS ALBERTO RINALDI
FCEFYN-UNC

Esta Tesis fue enviada a la Facultad de Ciencias Exactas Físicas y Naturales de la Universidad Nacional de Córdoba para cumplimentar los requerimientos de obtención del grado académico de Doctor en Ciencias de la Ingeniería.

CÓRDOBA, ARGENTINA
FEBRERO DE 2018



UNIVERSIDAD NACIONAL DE CORDOBA

Facultad de Cs. Exactas, Físicas y Naturales

ACTA DE EXAMENES

Libro: 00001 Acta: 04176 Hoja: 01/01
LLAMADO Y CATEDRA - MESA: 27/02/2018

D1002 TESIS DOCTORADO EN CIENCIAS DE LA INGENIERIA

NUMERO	APELLIDO Y NOMBRE	DOCUMENTO	INGRESO	COND	NOTA	FIRMA
12388168	FERREIRO, Susana Beatriz	DNI 12388168	2006	T	Aprobado	

VARGAS, LAURA MONICA - PINOTTI, Lucio - ACOSTA, Gerardo - CORRAL BRIONES, Graciela - MONTORO, M

Observaciones:

Sergio Elaskar

Sergio Elaskar
DIRECTOR
Doctorado en Ciencias de
la Ingeniería
F.C.E.F. y N. - U.N.C.

Córdoba, / /

Certifico que las firmas que han sido puestas en la presente Acta pertenecen a Sergio Elaskar

1	0	1	0	1
Inscritos	Ausentes	Examinados	Reprobados	Aprobados

16/02/2018 11:13:05

*Dedicado a mis hijos Juan Gabriel, Elisa y José Ignacio,
por su apoyo,
y acompañamiento*

Agradecimientos

Deseo agradecer a la Universidad Nacional de Córdoba por darme esta oportunidad.

A la Universidad Nacional de Río Cuarto, en particular a la Facultad de Ciencias Exactas y al Departamento de Matemática, en especial a su Directora, Profesora Graciela Giubergia, por las concesiones brindadas para avanzar en mi trabajo.

Este trabajo no habría sido posible sin el apoyo y el estímulo del *Dr. Oscar Humberto Bustos* quien me acompañó en todo el proceso.

Al Dr. Alejandro Frery por sus sugerencias, siempre atinadas, durante el desarrollo del trabajo.

No puedo terminar sin agradecer a mi familia, mi mamá, mi hermana, mis sobrinos Pablo y Lucía, por *confiar en mi y por su aliento*, en particular a mis hijos a quienes dedico esta tesis.

Resumen

Palabras claves: *Clasificación Supervisada - Muestras de Entrenamiento- Reducción de Dimensión - Combinación de Clasificadores - Coeficiente Kappa.*

La problemática relacionada con la producción de *mapas temáticos* en zonas donde predominan rocas y minerales, es una tarea difícil por la disposición espacial que presentan. Además, cuando se cuenta con imágenes de Teledetección de alta dimensión espectral (multiespectrales o hiperespectrales), frecuentemente existe alta correlación entre bandas espectrales y se necesita *reducir la dimensión*, sin perder la información presente. Ninguno, de los clasificadores supervisados utilizados para la producción de los mapas, resulta mejor para todas las clases o regiones presentes en la imagen. La Clasificación Supervisada comienza, en una primera fase, con la selección de muestras de entrenamiento. El tamaño y la separabilidad de las mismas es fundamental para el éxito de la clasificación, pero al aumentar el tamaño de las muestras hay más posibilidades de error.

Esta tesis propone:

- Dos métodos de reducción de dimensión: Análisis por Factores y Análisis Discriminante Lineal. Las variables obtenidas, en estos análisis, no son utilizadas. Se plantea, con los resultados obtenidos, generar nuevas variables que serán utilizadas como entrada para los algoritmos de clasificación.
- Utilizar tres clasificadores estadísticos supervisados y un cuarto clasificador, obtenido como una combinación de estos clasificadores.
- Una metodología para la obtención de nuevas muestras de entrenamiento, ampliadas y con mayor separabilidad, a partir de la información de las muestras de entrenamiento, seleccionadas manualmente.

Toda la metodología propuesta se implementa con códigos elaborados con el software R, que están disponibles en este trabajo. Para su validación se utilizan tres subimágenes ASTER, de la misma zona al Oeste de la provincia de Córdoba, Argentina.

Los resultados se evaluaron visualmente y numéricamente. Se utilizaron muestras de referencia, construyendo la matriz de confusión, el coeficiente de concordancia Kappa, el coeficiente de

Fiabilidad por clase y el coeficiente de Error, para la evaluación numérica. Para cada subimagen, se utilizan las mismas muestras de referencia para que los resultados sean comparables.

Los resultados muestran que, para todas las situaciones, los valores de los coeficientes son más altos cuando se utilizan todas las bandas espectrales (14 bandas). Sin embargo, cuando la dimensión es reducida a 4 bandas los valores son similares a los de la dimensión completa, tanto visualmente como numéricamente, con la ventaja de la reducción de dimensión. Cuando se utiliza la metodología que reduce a 3 bandas, visualmente la Composición RGB es más nítida, pero la clasificación se ve afectada, obteniendo mapas temáticos con clases mal asignadas.

En el estudio comparativo entre los cuatro clasificadores Análisis Discriminante Lineal (ADL); Análisis Discriminante Cuadrático (ADC); Máquinas de Soporte Vectorial (SVM) y el que resulta de la combinación de ellos (COMB), ninguno es mejor para todas las situaciones, aunque COMB se mantiene con coeficientes altos en la mayoría de ellas, arrojando mejores resultados.

En casi todas las situaciones generadas, a las muestras ampliadas le corresponden valores más bajos del coeficiente Kappa.

Abstract

Keywords: *Supervised Classification - Training Samples - Dimension Reduction - Combination of classifiers - Kappa coefficient.*

The production of thematic maps in areas where rocks and minerals predominate is a difficult task because of the spatial layout they present. In addition, the high-spectral (multispectral or hyperspectral) remote sensing images used to construct such thematic maps exhibit high correlation between their spectral bands. In order to automate detection and classification of minerals from these images, a Supervised Classification method can be applied. Then, the number of samples and their separability is fundamental for the classification success. However, the probability of error, as well as the complexity of the classification function, increase with the dimensionality of the samples. Therefore, it is necessary to reduce the dimensionality of the input without losing relevant information to ensure a high accuracy. The Supervised Classification procedure begins with the selection of training samples and none of the classifiers tested by us for map production is the best for all classes or all regions of the images

This thesis proposes:

- Two methods for dimensionality reduction (or feature reduction), using: Factor Analysis and Linear Discriminant Analysis. The variables obtained with these analyses are not used to generate new input variables for classification, instead of this, the analyses provide us a way to select the most relevant bands of the image.
- Compare three supervised statistical classifiers and a fourth classifier constructed as an ensemble of these.
- A supervised methodology to expand the training set to another one with greater separability.

Finally, the codes for the proposed methodology are implemented in R, which are available in this work. For test and validation of the proposed method, three ASTER subimages are used belonging to the same area of west of Córdoba, Argentina.

Results were assessed visually and numerically. We use reference data (ground truth) to construct the confusion matrix, the Kappa coefficient, the Producer's accuracy coefficient and the error coefficient, as measures of map accuracy. For each subimage, the same ground truth samples are used to make the results comparable.

The results show that regardless the image, the values of the coefficients are higher when all the spectral bands (14 bands) are used. However, when the dimension reduction is applied to get four bands, the coefficient values were similar to the case of full dimension. Both visually and numerically results confirm the advantage of dimensionality reduction. Furthermore, when the images are reduced to 3 bands, the RGB composition is sharper, but the classification is severely affected, obtaining thematic maps with misclassified classes.

In our comparative study between Linear Discriminant Analysis (ADL), Quadratic Discriminant Analysis (ADC), Support Vector Machine (SVM) and an ensemble learning (COMB), none of these got superior performance in all scenarios. Nonetheless, COMB shows better results in most of the scenarios.

In almost all situations generated, the extended samples correspond to lower values for Kappa.

Resumo

Palavras-chave: *Classificação supervisionada - Amostras Training - Redução de dimensão - combinação de classificadores - coeficiente Kappa.*

O problema relacionado à produção de mapas temáticos em áreas onde predominam rochas e minerais é uma tarefa difícil devido ao layout espacial que apresentam. Além disso, quando as imagens de sensoriamento remoto de alto espectro (multiespectrais ou hiperespectrais) estão disponíveis, há frequentemente alta correlação entre as bandas espectrais e a dimensão precisa ser reduzida, sem perder a informação presente. Nenhum dos classificadores supervisionados usados para a produção de mapas é o melhor para todas as classes ou regiões presentes na imagem. A Classificação Supervisionada começa, numa primeira fase, com a seleção das amostras de treino. O tamanho e a separabilidade destes são fundamentais para o sucesso da classificação, mas ao aumentar o tamanho das amostras há mais possibilidade de erro.

Esta tese propõe:

- Dois métodos de redução de dimensionalidade (características de extração característica de redução): Análise Fatorial e Análise Discriminante Linear. As variáveis obtidas nestas análises não são utilizadas. Propõe-se, com os resultados obtidos, a geração de novas variáveis que serão utilizadas como entrada para os algoritmos de classificação.
- Utilizar três classificadores estatísticos supervisionados e um quarto classificador, obtido como uma combinação destes classificadores.
- Uma metodologia para a obtenção de novo conjunto de treinamento, ampliado e com maior separabilidade, a partir das informações das amostras de treinamento, selecionadas manualmente.

Toda a metodologia proposta é implementada com códigos feitos com o software R, que estão disponíveis neste trabalho. Para sua validação, são utilizadas três sub-imagens ASTER, da mesma área a oeste da província de Córdoba, Argentina.

Os resultados foram avaliados visual e numericamente. Utilizamos dados de referência (ground truth), construindo a matriz de confusão, o coeficiente Kappa, o coeficiente de precisão do Produtor e o coeficiente de erro, como medidas de precisão de mapa. Para cada subimagem, as mesmas amostras de referência são utilizadas para tornar os resultados comparáveis.

Os resultados mostram que, para todas as situações, os valores dos coeficientes são maiores quando todas as bandas espectrais (14 bandas) são utilizadas. No entanto, quando a redução de dimensão é de 4 bandas os valores são semelhantes aos da dimensão total, tanto visual como numericamente, com a vantagem de redução de dimensão. Ao usar a metodologia que reduz a 3 bandas, visualmente a Composição RGB é mais acentuada, mas para a classificação é a que é mais afetada, obtendo mapas temáticos com classes mal atribuídas.

No estudo comparativo entre os quatro classificadores Analise Discriminante Lineal (ADL) , Analise Discriminante Quadrático (ADC), Maquinas de Suporte Vectorial (SVM) eo que resulta da combinação deles (COMB), nenhum é melhor para todas as situações, embora o COMB seja mantido com altos coeficientes na maioria deles, produz melhores resultados.

Em quase todas as situações geradas, as amostras estendidas correspondem a valores mais baixos do coeficiente Kappa.

Abreviaturas

ACP : Análisis de Componentes Principales

ADC : Análisis Discriminante Cuadrático. Clasificador

ADL : Análisis Discriminante Lineal. Clasificador

ANOVA : Análisis de la Varianza

ASTER : The Advanced Sparborne Thermal Emission and Reflectance Radiometer

COMB : Clasificador Combinado

$\mathcal{D} = \{D_1, D_2, \dots, D_L\}$: Conjunto de L clasificadores (*ensemble*).

\hat{F} : Coeficiente Fiabilidad Global (overall accuracy) estimado.

FA : Análisis por Factores (Factor Analysis)

$\hat{F}C_i$: Coeficiente de Fiabilidad por Clase estimado.

$\hat{\kappa}$: Coeficiente de Concordancia Kappa estimado.

MA : Muestras Ampliadas

MC : Matriz de Confusión o tabla de contingencia

ME : Muestras de Entrenamiento

MO : Muestras Originales

$\hat{P}D_s$: Probabilidad a posteriori estimada del clasificador s , con $s = 1, 2, \dots, L$

R : Software R es un conjunto integrado de programas para cálculo y gráficos.

SVM : Clasificador Máquinas de Soporte Vectorial (Support Vectors Machine). Clasificador

UE : Unidad Experimental

Índice general

Agradecimientos	VIII
Resumen	X
Abstract	XIII
Resumo	XV
Abreviaturas	XVII
1. Introducción	1
1.1. Antecedentes	1
1.2. El Problema	2
1.3. Clasificación digital Supervisada de Imágenes	4
1.4. Área de estudio	5
1.5. Imágenes ópticas multiespectrales ASTER	5
1.6. Entorno R	7
1.7. Organización de la Propuesta	8
2. Desarrollo	9
2.1. La presentación de los datos multivariados	9
2.2. Reducción de dimensión	10
2.2.1. Análisis por factores	10
2.2.2. Análisis Discriminante Lineal	16
2.3. Clasificadores utilizados	22
2.3.1. Clasificador Análisis Discriminante Lineal	23
2.3.2. Clasificador Análisis Discriminante Cuadrático	26
2.3.3. Las Máquinas de Soporte Vectorial	27
2.3.4. Combinación de clasificadores: Clasificador COMBinado	32

2.4. Muestras Originales vs Nuevas Muestras	32
2.5. Fase de Validación	34
3. Resultados	39
3.1. Análisis para la Subimagen 1	40
3.1.1. Resultados de las Clasificaciones de la Subimagen 1 con catorce bandas espectrales	42
3.1.2. Resultados de las Clasificaciones de la Subimagen 1 con tres bandas espectrales	46
3.1.3. Resultados de las Clasificaciones de la Subimagen 1 con cuatro bandas espectrales	49
3.2. Análisis para la Subimagen 2	52
3.2.1. Resultados de las Clasificaciones de la Subimagen 2	53
3.2.2. Validación de la Clasificación para Subimagen 2	54
3.3. Análisis para la Subimagen 3	55
3.3.1. Resultados de las Clasificaciones de la Subimagen 3	56
3.3.2. Validación de la Clasificación de la Subimagen 3	57
4. Conclusiones y trabajos futuros	61
4.1. Conclusiones	61
4.1.1. Muestras Originales vs Muestras Ampliadas	61
4.1.2. Comparando los Clasificadores	62
4.1.3. Clasificador por clase espectral	62
4.1.4. Las dimensiones espectrales	62
4.2. Contribuciones originales de la tesis	63
4.3. Trabajos futuros	63
A. Anexo A	65
A.1. Subimagen 1	65
A.1.1. Subimagen 1 con 14 bandas espectrales	65
A.1.2. Subimagen 1 con tres bandas espectrales	66
A.1.3. Subimagen 1 con cuatro bandas espectrales	67
A.2. Subimagen 2	69
A.2.1. Clasificaciones de la Subimagen 2 con 3 bandas	70
A.2.2. Clasificaciones de la Subimagen 2 con 4 bandas	72
A.3. Subimagen 3	74
A.3.1. Resultados de las Clasificaciones de la Subimagen 3	74

A.3.2. Validación de la Clasificación para la Subimagen 3	75
B. CODIGOS en R	79
C. Anexo Digital	103
Bibliografía	104

Índice de figuras

1.1. Área Geográfica de estudio	6
1.2. Características de ASTER comparado con Landsat ETM+	7
2.1. Ejemplos de diagramas obtenidos con R al aplicar Análisis por factores	15
2.2. Separación lograda por la función discriminante para dos grupos	18
2.3. Diagrama de un modelo canónico de un clasificador	23
2.4. Ejemplo de dos clases, linealmente separables, con hiperplano óptimo (de margen máximo).	28
2.5. Ejemplo de dos clases no linealmente separables, en \mathfrak{R}^2 , que se separan linealmente en \mathfrak{R}^3 , por medio de Φ	31
3.1. Composición RGB439 de la Subimagen 1	40
3.2. Diagrama de Dispersión de las muestras de entrenamiento, Subimagen 1 con 14 bandas	42
3.3. Mapas temáticos obtenidos con el clasificador COMB de la Subimagen 1, con 14 bandas	43
3.4. Diagrama de líneas múltiples de los Coeficientes Kappa estimados para la Subimagen 1 con 14 bandas.	44
3.5. Coeficientes de Fiabilidad por clase para la Subimagen 1, con 14 bandas.	45
3.6. Composición RGB123 de la Subimagen 1 con 3 bandas	46
3.7. Kappa vs clasificador de la Subimagen 1 con 3 bandas	47
3.8. Coeficientes de Fiabilidad por clase, de la Subimagen 1 con 3 bandas	48
3.9. Mapas temáticos obtenidos con el clasificador COMB de la Subimagen 1 con 3 bandas	49
3.10. Mapas temáticos obtenidos con el clasificador COMB de Subimagen 1 con 4 bandas	50
3.11. Kappa vs clasificador de la Subimagen 1 con 4 bandas	51
3.12. Coeficientes de Fiabilidad por cada clase, de la Subimagen 1 con 4 bandas	52
3.13. Composición RGB249 de la Subimagen 2	52

3.14. Mapas temáticos obtenidos con el clasificador COMB para Subimagen 2	53
3.15. Coeficientes Kappa para tres dimensiones de la Subimagen 2	54
3.16. Coeficientes de Fiabilidad por clase, de la Subimagen 2 con 14 bandas	55
3.17. Coeficientes de Fiabilidad por clase, de la Subimagen 2 con 4 bandas	56
3.18. Composición RGB439 de la Subimagen 3	56
3.19. Mapas temáticos obtenidos con el clasificador COMB de la Subimagen 3	57
3.20. Coeficientes Kappa vs clasificador para las tres dimensiones de la Subimagen 3 .	58
3.21. Coeficientes de Fiabilidad por clase de la Subimagen 3 con 14 bandas.	58
3.22. Coeficientes de Fiabilidad por clase de la Subimagen 3 con 4 bandas	59
A.1. Mapas temáticos de la Subimagen 1 con 14 bandas y las muestra originales . . .	66
A.2. Histogramas y Diagramas de dispersión de la Subimagen 1 y las muestras originales	67
A.3. Mapas temáticos de la Subimagen 1 con 3 bandas y las muestras originales. . . .	68
A.4. Composición RGB123 de la Subimagen 1, con 4 bandas	68
A.5. Clasificación de la Subimagen 1 con 4 bandas y muestras originales	69
A.6. Clasificación de la Subimagen 1 con 4 bandas y las muestras ampliadas	70
A.7. Composición RGB de la Subimagen 2, con 3 bandas	71
A.8. Coeficientes de Fiabilidad por clase de la Subimagen 2, con 3 bandas	72
A.9. Mapas temáticos para ADL y ADC con las MO de la Subimagen 2	73
A.10. Muestras de entrenamiento y datos de referencia en la Subimagen 3	74
A.11. Composición RGB123 de la Subimagen 3 con 3 y 4 bandas	75

Índice de tablas

2.1. Ejemplo de los resultados del Clasificador COMBINado	33
2.2. Ejemplo de cuartiles 1 y 3, por variable, para la i –ésima clase.	33
2.3. Estructura general de una Matriz de Confusión	35
3.1. Situaciones abordadas para cada una de las subimágenes	39
3.2. Descripción de las muestras para cada región de la Subimagen 1	41
3.3. Tamaño de Muestras Ampliadas para la Subimagen 1 con 14 bandas	42
3.4. Matriz de Confusión para el clasificador COMB con las MO	43
3.5. Coeficientes estimados del Error y Kappa para la Subimagen 1 con 14 bandas	44
3.6. Coeficientes de Fiabilidad por clase (en %) para la Subimagen 1 con 14 bandas y las MO	45
3.7. Coeficientes Fiabilidad por clase (en %) para la Subimagen 1 con 14 bandas y las MA	45
3.8. Coeficientes Error y Kappa para cada clasificador con MO y MA para la Subimagen 1 con 3 bandas	47
3.9. Coeficientes Fiabilidad por clase (en %) para la Subimagen 1 con 3 bandas y las MO	48
3.10. Coeficientes Fiabilidad por clase (en %) para la Subimagen 1 con 3 bandas y las MA	48
3.11. Error y Kappa por clasificador con MO y MA para la Subimagen 1 con 4 bandas	50
3.12. Coeficientes de Fiabilidad por clase (en %) para la Subimagen 1 con 4 bandas y las MO	51
3.13. Coeficientes de Fiabilidad por clase (en %) para la Subimagen 1 con 4 bandas y las MA	51
3.14. Medidas de Fiabilidad para todas las situaciones de la Subimagen 2	54
A.1. Clase asignada por el clasificador COMB	65
A.2. Tamaño de las muestras ampliadas para la Subimagen 1 con 3 bandas	66

A.3. Tamaño de las muestras ampliadas para la Subimagen 1 con 4 bandas	68
A.4. Detalles de las muestras seleccionadas para cada región de la Subimagen 2	70
A.5. Coeficientes de Fiabilidad (en %) por clase para la Subimagen 2 con 14 bandas y las MO	71
A.6. Coeficientes de Fiabilidad por clase (en %) para la Subimagen 2 con 14 bandas y las MA	71
A.7. Coeficientes de Fiabilidad por clase para la Subimagen 2 con 3 bandas y las MO	72
A.8. Coeficientes de Fiabilidad por clase para la Subimagen 2 con 3 bandas y las MA	72
A.9. Coeficientes de Fiabilidad por clase para la Subimagen 2 con 4 bandas y las MO.	73
A.10. Coeficientes de Fiabilidad por clase para la Subimagen 2 con 4 bandas y las MA	73
A.11. Detalles de las muestras seleccionadas para cada región de la Subimagen 3	74
A.12. Tamaño de las muestras ampliadas para la Subimagen 3 con 14 bandas	75
A.13. Clase asignada con el clasificador COMB para Subimagen 3	75
A.14. Matriz de confusión para el clasificador COMB, con MO y Subimagen 3, 14 bandas	76
A.15. Medidas de confiabilidad para todas las situaciones de la Subimagen 3	76
A.16. Coeficientes de Fiabilidad por clase para la Subimagen 3 con 14 bandas y las muestras originales	76
A.17. Coeficientes de Fiabilidad por clase para la Subimagen 3 con 14 bandas y MA .	76
A.18. Coeficientes de Fiabilidad por clase para la Subimagen 3 con 3 bandas y las muestras originales	77
A.19. Coeficientes de Fiabilidad por clase para la Subimagen 3 con 3 bandas y MA . .	77
A.20. Coeficientes de Fiabilidad por clase para la Subimagen 3 con 4 bandas y MO . .	77
A.21. Coeficientes de Fiabilidad por clase para la Subimagen 3 con 4 bandas y MA . .	77

Capítulo 1

Introducción

En este capítulo, se presentan los antecedentes del tema de estudio y se realiza el planteamiento del problema que motiva este trabajo. Se muestra el área de estudio, además de exponer las características de la Imagen ASTER, de describir en qué consiste la Clasificación Digital de Imágenes y las propiedades del software **R**.

1.1. Antecedentes

La Teledetección está reconocida como fuente primaria de información para realizar cartografía, sus productos son esenciales para estadística agraria, exploración minera y estudios hídricos entre otros. Por lo tanto, una de las tareas más importantes en el procesamiento y análisis de imágenes es la producción de cartografía temática. Una de las formas más comunes de realizar esta tarea es mediante la clasificación (visual o digital) de imágenes para discriminar cubiertas en una serie de categorías homogéneas. Los datos de sensores remotos son transformados en *información*. Ellos pueden ayudar en la cartografía litológica (discriminar tipos de rocas), a escalas macroscópicas, y orientar los futuros estudios geológicos en esos lugares. Los conjuntos de datos más utilizados para la cartografía litológica son los obtenidos con el sensor ASTER (The Advanced Spaceborne Thermal Emission and Reflectance Radiometer) debido a sus múltiples bandas, una alta resolución espacial y relativamente bajo costo.

Muchos métodos se han utilizado para detectar minerales y rocas con los datos provenientes de las bandas espectrales de ASTER, las que abarcan desde el visible al infrarrojo cercano (VNIR) y de onda corta infrarroja (SWIR). Ciertos minerales con silicios y silicatos presentan características espectrales específicas en el infrarrojo térmico (TIR). Diversos estudios han demostrado el gran potencial de la región TIR para el mapeo litológico por teledetección aerotransportada (Benkhoff y the MERTIS team, 2006; Vaughan y Calvin, 2003; Watanabe y

Matsuo, 2003). El sensor ASTER es el primero a bordo de un satélite con suficiente resolución, espacial (90x90 m), espectral y radiométrica, para estudios geológicos. Para identificar y producir mapas de distintos tipos de granitos fueron propuestos cocientes entre bandas del TIR, combinaciones de bandas, cocientes entre estas combinaciones, Índices de Cuarzo (QI, por su sigla en inglés), tal como se muestra en los trabajos de (Pérez et al., 2007; Watanabe y Matsuo, 2003). Valores altos de QI ocurren para rocas ricas en cuarzo y pobres en feldespato y valores bajos para rocas ricas en yeso o feldespato potásico. Granitos ricos en Albita se detectaron como píxeles oscuros en la imagen QI. Datos ASTER del TIR son ampliamente utilizados para detectar rocas máficas, ultramáficas y ricas en cuarzo, y se han propuesto varios índices basados en las características de emisividad. Sin embargo, las bandas del TIR están altamente correlacionadas entre ellas, por lo que la información independiente procedente de diferentes bandas puede ser limitada. Además, los datos de la región del Infrarrojo Térmico contienen efecto atmosférico y de la temperatura de la superficie, interfiriendo así con la disponibilidad de estos índices previamente propuestos. En Ding et al. (2014) se propone una metodología que intenta explicar la correlación entre pares de bandas y utilizan una aproximación lineal de la función de Planck para deducir una ecuación que representa la relación de la luminosidad entre dos bandas TIR. Dos índices de rocas máficas-ultramáficas y dos de rocas ricas en cuarzo fueron propuestos a partir del Análisis de Correlación y Regresión Lineal Simple. Una comparación entre estos índices con otros ya propuestos en la literatura, usando las curvas espectrales de emisividad, muestran la mayor utilidad de estos índices. Estos métodos serán adecuados para la detección de estas rocas en regiones de alta elevación y presentan interferencia de la temperatura de la superficie y la atmósfera, por lo que su utilidad se limita para regiones muy específicas y además para calcularlos, hay que disponer de la información de la temperatura de la superficie en el momento de la adquisición de la imagen. Las características espectrales originales de rocas félsicas a menudo se entremezclan con otros objetos de la superficie, lo que se traduce en dificultad para detectar estas rocas utilizando técnicas de teledetección. La interpretación visual con composición RGB en falso color es ampliamente utilizado para detectar rocas félsicas y también se propusieron índices en el trabajo de Ding et al. (2015). Además Kalinowski y Oliver (2004) brinda un manual, donde aparecen una lista de rocas y minerales y las bandas que se deberían combinar para visualizarlos.

1.2. El Problema

De lo expuesto hasta aquí, se desprende que producir un mapa temático con los tipos de rocas presentes es una tarea engorrosa, ya que las rocas y minerales son difíciles de detectar en la superficie por la disposición espacial que presentan. En la literatura, se observa que la metodo-

logía propuesta realiza clasificación **visual** del área de interés mediante índices, que representan operaciones entre bandas o necesitan información adicional, por ejemplo, la temperatura.

La propuesta de esta Tesis es realizar clasificación **digital** de rocas y minerales de una zona del oeste de la provincia de Córdoba; no hay antecedentes en esa zona ni en otras zonas. Para esta tarea, se propone un clasificador que es la combinación de los clasificadores: Análisis Discriminante Lineal (ADL), Análisis Discriminante Cuadrático (ADC) y Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés) que llamaremos Combinado (COMB). Aunque en la literatura hay antecedentes de combinación de clasificadores (ensemble) Tulyakov et al. (2008), no de la manera que se presenta aquí.

Otro problema es la presencia de correlación entre las 14 variables (bandas del sensor). Se necesita realizar reducción de dimensión para sacar la información redundante. Esta disminución agiliza y facilita el análisis multivariado. Recientemente en Wang et al. (2016) se presenta una propuesta para reducción de dimensión, pero diferente a lo planteado en este trabajo. En este marco se plantean dos estrategias:

1. Seleccionar las variables originales (bandas) que mejor separan las clases o regiones en la imagen. Para ello se utiliza Análisis Discriminante Lineal (ADL), que genera nuevas variables llamadas *funciones discriminantes* en una dimensión menor. Se seleccionan aquellas bandas que tienen mayor coeficiente en la función, es decir, mayor contribución a las separación de los grupos, con la ventaja de utilizar la información original.
2. Aplicar el procedimiento estadístico multivariado “Análisis por Factores” (FA, por sus siglas en inglés) para determinar cómo se agrupan las variables. Lo que se propone es combinar las variables originales en cada grupo, generando una variable nueva por grupo, utilizándolas como entrada para la clasificación.

Todo este pre-procesamiento ayudará a seleccionar las mejores muestras de entrenamiento, que es la primera fase de la Clasificación digital. Además, de la selección manual en la imagen de las muestras de entrenamiento (ME), se formula una nueva técnica de selección de muestras, que reduce la intervención del operador.

En síntesis, se analiza cada imagen en tres dimensiones diferentes, la total y las dos obtenidas al aplicar la propuesta; con dos tipos de muestras de entrenamiento (una de ellas con la metodología planteada) y cuatro clasificadores (uno de ellos es combinación de los otros). Para poner a prueba lo planteado se utilizan tres sub-escenas, pertenecientes al área de estudio, las cuáles están a disposición del usuario al igual que los códigos del Software **R** desarrollados.

1.3. Clasificación digital Supervisada de Imágenes

A partir de una imagen mediante la clasificación digital se puede producir un mapa temático de la zona en estudio, discriminando las regiones homogéneas presentes o realizar un inventario de los resultados obtenidos (Chuvieco, 2002; Richards y Jia, 2006a). Entre los métodos desarrollados para clasificación digital ver, por ejemplo, los trabajos de (Duda et al., 2001; Gordon, 1999; Krzanowski y Marriott, 1995). Se analizan, en este trabajo, los que pertenecen a la clase de los métodos estadísticos supervisados por su fuerte relación con la elección de muestras.

Estos métodos de clasificación supervisada constan de tres fases fundamentales: entrenamiento, producción y validación. Para generar buenos productos, estos procedimientos deben obedecer ciertas reglas que se comentan a seguir.

Ninguna de esas tareas es trivial y se las considera particularmente difíciles para construir mapas con diferentes litologías (tipos de rocas), cuando se trabaja con datos de imágenes ASTER. Es por ello que, en esta tesis se dará particular énfasis al trabajo con este tipo de datos y se contribuirá con las tres etapas.

La primera fase es fundamental para la validez de los resultados que se obtienen en las fases siguientes. Obtener un buen producto depende críticamente de la calidad de la etapa de entrenamiento (Congalton y Green (1999)). En los trabajos de (Ferrero et al., 2005; Frery et al., 2006, 2009), se muestra la influencia que tienen los errores de la fase de entrenamiento, en la clasificación y por lo tanto en el producto final.

El entrenamiento de algoritmos de clasificación requiere:

- la identificación de clases o regiones presentes en la imagen,
- la selección de las variables (bandas) que separen mejor las clases y
- la selección de muestras representativas en cada una de ellas.

Además de las muestras de entrenamiento, es necesario recolectar otras muestras, también llamadas puntos test o datos de referencia, para evaluar el desempeño esperado de la clasificación.

La contribución que esta tesis da a la primera fase tiene por objetivo proponer, implementar y verificar la utilidad de un sistema de producción, que posibilite la aplicación de una metodología para la identificación, el tratamiento y la validación de muestras de clases, a ser usadas por algoritmos de clasificación automáticos de imágenes satelitales.

En la segunda fase donde se utilizan los algoritmos de clasificación, se presentan aquí los supervisados siendo el más usado el clasificador de Máxima Verosimilitud Gaussiana Puntual. Este método está tratado en los textos de Teledetección como los de (Chuvieco, 2002; Richards y Jia, 2006a), además de estar disponible como rutina en los paquetes computacionales específicos

para tratamiento de imágenes. En este trabajo se desarrollan, implementan en el software R y se aplican a una imagen real los clasificadores “*ADL; ADC; SVM y COMB*”.

En la tercera fase, donde se evalúa la calidad de la clasificación realizada, es decir del mapa temático obtenido, se realiza la validación utilizando los datos de referencia y calculando diferentes medidas, aconsejándose más asiduamente los coeficientes de concordancia Kappa, de Error y de Fiabilidad por clase, también llamado Fiabilidad del Productor (Congalton (1991)). Estos coeficientes se emplean además para comparar rendimientos entre los clasificadores.

De esta manera, esta tesis aporta nuevas técnicas integradas para la generación de mejores productos cartográficos a partir de imágenes satelitales. Las próximas secciones dan detalles de cada una de las contribuciones pretendidas, bien como de las plataformas a ser utilizadas.

1.4. Área de estudio

Las imágenes del sensor ASTER, que se utilizan en este trabajo, son adecuadas para la elaboración de mapas temáticos que describan los minerales presentes en el área de estudio. Aquí se analiza una zona ubicada al sur-oeste de la Provincia de Córdoba, República Argentina, en las sierras de Comenchingones, próxima a la localidad de Alpa Corral con ubicación geográfica $32^{\circ}41'26''S$, $64^{\circ}43'16''W$ y Altitud entre 800 y 2000 msnm. La imagen utilizada es *AST_L1B_003_03032003143302_03222003172450*. En la misma se selecciona una subimagen localizada en $32^{\circ}26'S$ y $64^{\circ}55'W$ para la esquina superior izquierda, que abarca las rocas del cuerpo “Batolito del Cerro Áspero”, analizándose tres subimágenes de esta área. Cualquier usuario puede obtener, de forma gratuita, la imagen de la zona utilizada aquí y también de otras regiones. La dirección donde se encuentra disponible es: <https://earthexplorer.usgs.gov/>.

1.5. Imágenes ópticas multiespectrales ASTER

El Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) es un sensor satelital de imágenes de alta resolución, transportado por el satélite TERRA, lanzado el 18 de diciembre de 1999. ASTER es un esfuerzo cooperativo entre la NASA, el Ministerio japonés de Economía, Comercio e Industria (METI) y Space Systems Japón (J-Spacesystems).

Presenta una órbita heliosincrónica a una distancia de 705 kilómetros, con un ciclo de repetición de 16 días, un ancho de barrido de 60 kilómetros y una distancia entre órbitas de 172 km. Tiene una resolución espectral de 14 bandas en diferentes regiones del espectro electromagnético; 3 en el visible e infrarrojo cercano (VNIR) con 15 metros de resolución espacial, 6 en las ondas cortas del Infrarrojo (SWIR), con 30 m de resolución, y 5 en el Infrarrojo Térmico (TIR) con 90

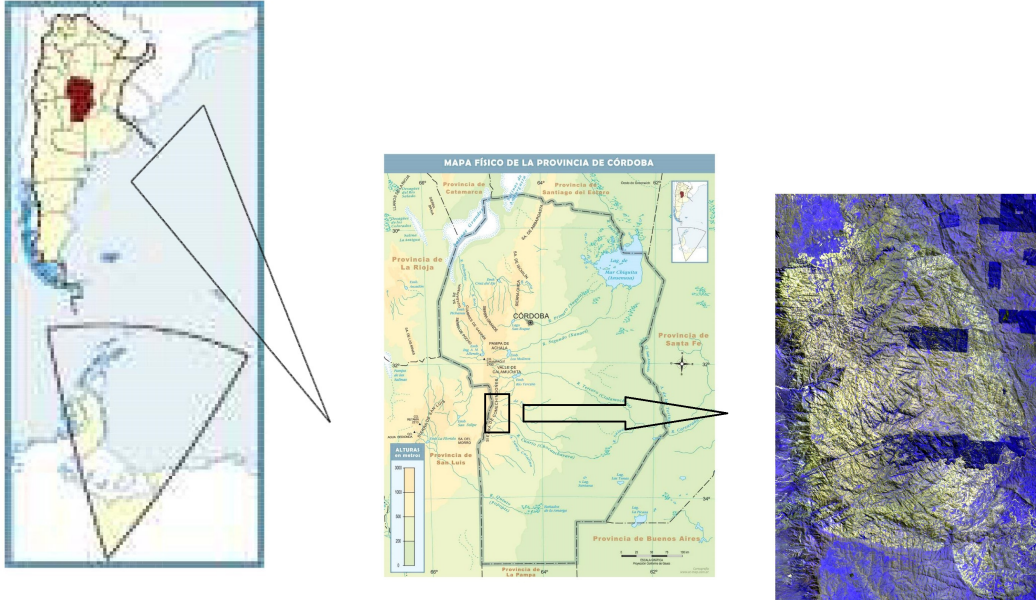


Figura 1.1: Área Geográfica de estudio

m de resolución espacial. Comparado con Landsat Enhanced Thematic Mapper Plus (ETM+), ASTER tiene mayor resolución espectral (14 bandas), resolución espacial en el VNIR (15 m) y alta resolución radiométrica (sensibilidad para detectar variación en la cantidad de energía que recibe). La resolución espacial varía con la longitud de onda y cada escena cubre 60km x60km. Las bandas son espectralmente más contiguas que los sensores LANDSAT, especialmente en el rango de longitud de onda del Infrarrojo cercano. Tiene un alto grado de exactitud en la identificación de rocas y minerales (Crosta et al. (2003)). En Zhang et al. (2007), los autores indican que este sensor comparado con otros similares es adecuado para mapeos litológicos. Datos del infrarrojo térmico ASTER (TIR), bandas 10 a 14, son ampliamente utilizados para detectar rocas ricas en cuarzo, rocas anfibolitas y otros minerales presentes en la zona de estudio, de interés en este trabajo. Varios índices se han propuesto sobre la base de las características de emisividad (proporción de radiación térmica emitida por una superficie u objeto debida a una diferencia de temperatura con su entorno). Sin embargo, las bandas de ASTER TIR de datos de radiancia están altamente correlacionadas, por lo que la información independiente procedente de diferentes bandas puede ser limitada. El interés es proponer índices que ayuden a identificar la presencia de minerales, por ejemplo zonas ricas en cuarzo, para la construcción del mapa temático de la zona. Esto se puede lograr mediante operaciones entre las bandas del TIR.

En la Figura 1.2 se muestran la ubicación de las 14 bandas de ASTER según su longitud de onda en el espectro electromagnético y se comparan con las de LANDSAT ETM+.

Se puede observar que mientras el LANDSAT ETM+ tiene una banda (la 6) en el Infra-

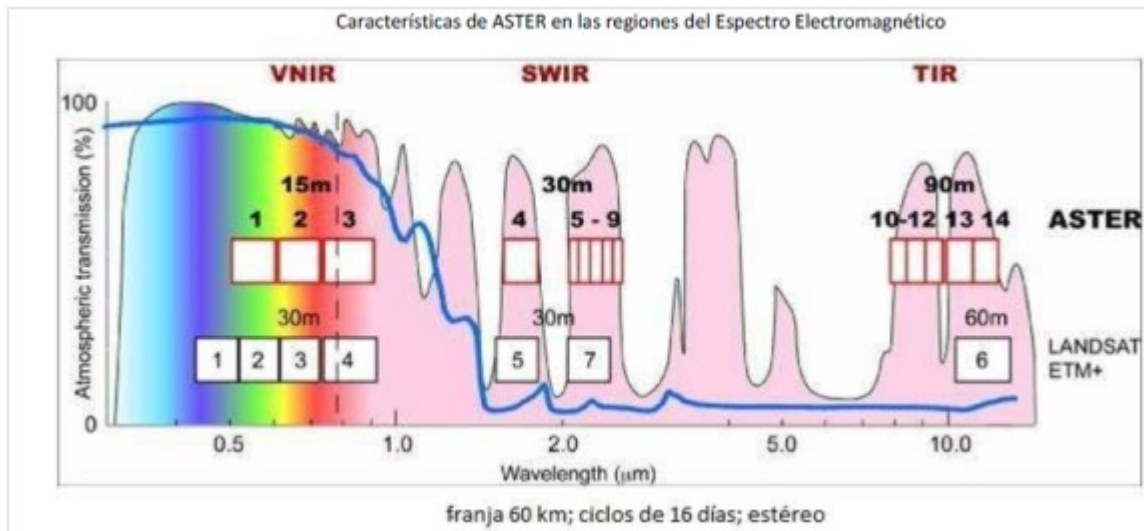


Figura 1.2: Características de ASTER comparado con Landsat ETM+

rojo Térmico, ASTER tiene 5 bandas en esa región. La página de internet para obtener más información sobre las imágenes ASTER es: <http://asterweb.jpl.nasa.gov/>

1.6. Entorno R

R es un conjunto integrado de programas para manipulación de datos, cálculo y gráficos, ver <http://www.r-project.org>. Entre otras características dispone de:

- almacenamiento y manipulación efectiva de datos,
- operadores para cálculo sobre variables indexadas (Arrays), en particular matrices,
- una amplia, coherente e integrada colección de herramientas para análisis de datos,
- posibilidades gráficas para análisis de datos, que funcionan directamente sobre pantalla o impresora, y
- un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y posibilidad de entradas y salidas.

El término “entorno” lo caracteriza como un sistema completamente diseñado y coherente, antes que como una incorporación creciente de herramientas muy específicas e inflexibles, como ocurre frecuentemente con otros programas de análisis de datos. R es en gran parte un vehículo para el desarrollo de nuevos métodos de análisis interactivo de datos. Como tal, es muy dinámico y las diferentes versiones no siempre son totalmente compatibles con las anteriores. Mientras que

algunos usuarios prefieren los cambios debido a los nuevos métodos y tecnología que los acompañan, a otros les molesta, ya que algún código anterior deja de funcionar. Aunque R puede entenderse como un lenguaje de programación, los programas escritos en R deben considerarse esencialmente efímeros. R Core Team (2017) es elegido por muchos investigadores en metodología estadística. Como R está realizado en un lenguaje computacional le permite al usuario agregarle aplicaciones definiendo nuevas funciones. R es un ambiente en el cual son implementadas las técnicas estadísticas y se destaca por la precisión numérica, respecto de otros software. La implementación de los procedimientos en el Software R tiene la ventaja de que es libre y gratuito, eliminando la dependencia de productos comerciales. R está disponible como software libre bajo los términos de Free Software Foundation's GNU General Public License forma de código fuente. Compila y corre bajo una amplia variedad de plataformas y sistemas (UNIX, FreeBSD Linux, Windows and MacOS). Cuando el software es "libre", su código fuente está disponible universalmente y se puede cambiar para satisfacer necesidades específicas. El software gratuito (freeware) se puede utilizar sin necesidad de adquirir o pagar, pero no necesariamente proporciona acceso al código fuente, por lo que no puede ser cambiado; sólo se puede usar tal como está disponible. Existen una gran variedad de paquetes desarrollados para las más diversas aplicaciones; el depósito oficial de estos paquetes, está en <http://www.cran.r-project.org>.

Se pretende integrar todas las etapas de la clasificación ofreciendo un producto que abarque todo el proceso, desde la entrada de datos, hasta la obtención del producto final, un mapa temático. Esta tarea se realiza implementando Códigos en **R** y complementando con las librerías que están disponibles en el software.

1.7. Organización de la Propuesta

La estructura de esta tesis es la siguiente:

- En el Capítulo 2 se introduce al lector en las diferentes técnicas utilizadas para el desarrollo de esta tesis. Además, describe en detalle las propuestas originales para este trabajo.
- En el Capítulo 3, se presentan los resultados obtenidos al aplicar las técnicas propuestas, en diferentes situaciones (24 en total).
- Las Conclusiones son discutidas en el Capítulo 4, donde también se plantean posibles trabajos futuros.
- En el Anexo A, se muestran algunos de los resultados para completar la información del Capítulo 3. En el Anexo B, están los Códigos en R elaborados para implementar las técnicas propuestas.

Capítulo 2

Desarrollo

En este capítulo se presentan los conceptos del Análisis Multivariado necesarios para el desarrollo de este trabajo. Se detalla la metodología de FA y ADL con el objetivo de aplicarla como técnica de *reducción de la dimensión* en este trabajo, mostrando sus características, los aspectos que son usados y sus ventajas respecto del ACP. En las secciones 2.2.1 y 2.2.2 se plantea la propuesta para la aplicación aquí realizada. Cuando se tienen grupos dentro de la población, para la reducción de dimensión se muestra una metodología diferente de ACP, usado habitualmente. Para construir mapas temáticos, se presentan las características básicas de los clasificadores que se utilizan y también se propone un clasificador que resulta de la combinación de los otros aplicados, siguiendo las ideas de Kuncheva (2014). También se describe una metodología original para ampliación de las muestras de entrenamiento reduciendo lo más posible la intervención del operador. Finalmente, se detallan los métodos empleados para la validación de los resultados: matriz de confusión, coeficiente Kappa, de Error, de Fiabilidad por clase entre otros.

2.1. La presentación de los datos multivariados

En muchas aplicaciones, donde se realizan varias medidas sobre cada unidad experimental (UE) (individuos, objetos o eventos) se genera un vector de observaciones para cada unidad. A un conjunto de datos de naturaleza multivariada se los puede arreglar como en (2.1), es decir, en una matriz donde las filas representan las unidades experimentales y las columnas a las

variables.

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1j} & \cdots & y_{1p} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ y_{i1} & y_{i2} & \cdots & y_{ij} & \cdots & y_{ip} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nj} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} Y'_1 \\ \vdots \\ Y'_i \\ \vdots \\ Y'_n \end{pmatrix} = (Y_1, Y_2, \dots, Y_j, \dots, Y_p) \quad (2.1)$$

Donde n es tamaño de la muestra (número de unidades experimentales) y p es la cantidad de variables medidas sobre cada unidad experimental, entonces y_{ij} representa el valor de la j –ésima variable para la i –ésima unidad experimental, con $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, p$. En el contexto de imágenes de teledetección el conjunto de datos representan los valores de radiancia, asociados con la posición geográfica (latitud y longitud) del área captada por el sensor. Una imagen digital es una grilla (cuadrícula) de F filas y C columnas, donde cada celda de la grilla es el valor correspondiente a un pixel (mínima unidad de información de la imagen) y se la suele representar como una matriz $F \times C$. Para cada variable, banda espectral, la matriz $F \times C$ correspondiente es transformada en un vector de dimensión $n = F \times C$ y la totalidad de datos se disponen como una matriz de n filas y p columnas; resultando como en (2.1).

2.2. Reducción de dimensión

La reducción de dimensión es un problema importante cuando se manejan gran cantidad de variables, como en el campo de reconocimiento de patrones, data mining, visión computacional, machine learning entre otros. El problema también se lo llama *extracción de características* y el interés es encontrar un *nuevo* conjunto de m variables a partir de p variables de la dimensión original, con $m < p$, manteniendo toda la información. Se proponen dos métodos, uno supervisado y otro no supervisado, ADL y FA, respectivamente, tal como se presenta en Alpaydin (2010) y en Rencher (2002).

2.2.1. Análisis por factores

El Análisis por Factores (FA: factor analysis) es un método del Análisis Multivariado que pretende explicar p variables observables como una combinación lineal de m variables no observables o latentes, denominadas *factores* ($m < p$). FA se emplea para crear nuevas variables que resuman la información disponible en las originales. El objetivo es caracterizar la dependencia entre las variables observadas por un pequeño número de factores. Se supone que hay

un conjunto de factores F_k , $k = 1, \dots, m$ ($m < p$) que, cuando actúan en combinación, generan las variables medidas, Y_1, Y_2, \dots, Y_p . Al igual que las variables originales, los factores varían de individuo a individuo; pero a diferencia de ellas, los factores no pueden ser medidos u observados. Si un grupo de variables tienen alta correlación entre sí y baja correlación con las demás, entonces puede haber un solo factor subyacente que dio lugar a estas variables. Si las otras variables también pueden ser agrupadas en subconjuntos, entonces unos pocos factores representan esos grupos de variables. FA realiza particiones de las variables en grupos, los factores, pero éstos podrían no tener significado o no existir; es un método no supervisado y el objetivo es modelar los datos en un espacio de dimensión menor, sin pérdida de información. El FA se relaciona con el ACP ya que ambos buscan una estructura más simple en un conjunto de variables, pero difieren en muchos aspectos. Las componentes principales se definen como combinaciones lineales de las variables originales mientras que en FA las variables originales se expresan como combinaciones lineales de los factores. En el ACP se explica una gran parte de la varianza total de las variables, en tanto que en FA se buscan grupos de variables analizando la correlación entre ellas.

A veces, para algunos conjuntos de datos surgen factores fácilmente interpretables, pero para otros ni el número de factores ni la interpretación es clara. Por lo que, FA en muchas aplicaciones es subjetivo y considerado polémico por algunos estadísticos, no proporcionando un ajuste satisfactorio. El FA obtiene e interpreta los factores comunes a partir de las correlaciones entre las variables, mostradas en la matriz \mathbf{R} :

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \cdots & \cdots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

Definición del Modelo y Supuestos

Sea una muestra aleatoria de tamaño n de una población común con vector de medias $\boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}$. El modelo del FA de m factores comunes considera que p variables observables Y_1, Y_2, \dots, Y_p dependen de m variables latentes, F_1, F_2, \dots, F_m , llamadas factores comunes y p factores específicos (también llamado error aleatorio) $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots, \boldsymbol{\epsilon}_p$. Cada observación se puede expresar según el modelo lineal:

$$y_{ij} - \mu_j = v_{j1}f_{1i} + v_{j2}f_{2i} + \cdots + v_{jm}f_{mi} + \epsilon_{ij}, \forall j = 1, \dots, p, \forall i = 1, \dots, n$$

$$y_{ij} - \mu_j = \sum_{k=1}^m v_{jk} f_{ki} + \varepsilon_{ij} \quad (2.2)$$

De acuerdo a este modelo se admite que las variables, en conjunto, dependen de los factores comunes, salvo una parte de su variabilidad, sólo explicada por el correspondiente factor específico. A los factores f_{ki} se los considera variables aleatorias que generan las y_{ij} . Los coeficientes v_{jk} son llamados *cargas* y sirven como pesos, mostrando la importancia del k –ésimo factor para la j –ésima variable y se los utiliza para la interpretación del factor. Estos coeficientes ayudan a la partición de las variables en grupos, correspondientes a los factores.

En notación matricial (2.2) se puede escribir como:

$$\mathbf{Y} - \mathbf{1}\boldsymbol{\mu}' = \mathbf{F}\mathbf{V}' + \boldsymbol{\Xi} \quad (2.3)$$

donde:

- \mathbf{Y} es la matriz de datos, $n \times p$;
- $\mathbf{1}$ vector de unos, $n \times 1$;
- $\boldsymbol{\mu}'$ vector de medias, $1 \times p$;
- \mathbf{F} contiene los m factores para las n observaciones, $n \times m$;
- \mathbf{V}' llamada matriz de *cargas factoriales*, $m \times p$;
- $\boldsymbol{\Xi}$ matriz de error o factores específicos, $n \times p$.

Por simplificación, se pueden expresar las p variables para una observación de la forma siguiente:

$$\mathbf{y} - \boldsymbol{\mu} = \mathbf{V}\mathbf{f} + \boldsymbol{\epsilon} \quad (2.4)$$

detallando cada elemento se tiene que

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{pmatrix}, \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1m} \\ v_{21} & v_{22} & \cdots & v_{2m} \\ \vdots & \cdots & \cdots & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pm} \end{pmatrix}, \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}, \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix}$$

Los supuestos del modelo son:

- $E(\mathbf{f}) = \mathbf{0}$, los factores comunes tienen media cero;
- $E(\boldsymbol{\epsilon}) = \mathbf{0}$, los factores específicos tienen media cero;
- $Cov(\mathbf{f}) = E(\mathbf{f}\mathbf{f}') = \mathbf{I}$, los factores comunes no están correlacionados y estandarizados;
- $Cov(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \Psi$, donde $\Psi = \text{diag}(\psi_1, \psi_2, \dots, \psi_p)$, los factores específicos no están correlacionados. Ψ se denomina matriz de especificidad;
- $Cov(\boldsymbol{\epsilon}\mathbf{f}) = E(\boldsymbol{\epsilon}\mathbf{f}') = \mathbf{0}$, ausencia de correlación entre factores comunes y específicos.

A partir de estos supuestos se obtiene una descomposición para la matriz de covarianza. Dado que $\boldsymbol{\mu}$ no afecta a las varianzas y covarianzas de \mathbf{y} , se tiene de (2.4):

$$\begin{aligned}
\boldsymbol{\Sigma} &= Cov(\mathbf{y}) = Cov(\mathbf{V}\mathbf{f} + \boldsymbol{\epsilon}) \\
&= Cov(\mathbf{V}\mathbf{f}) + Cov(\boldsymbol{\epsilon}) \\
&= \mathbf{V}Cov(\mathbf{f})\mathbf{V}' + \Psi \\
&= \mathbf{V}\mathbf{I}\mathbf{V}' + \Psi \\
&= \mathbf{V}\mathbf{V}' + \Psi
\end{aligned} \tag{2.5}$$

Si \mathbf{V} tiene pocas columnas, dos o tres, entonces $\boldsymbol{\Sigma}$ de (2.5) tiene una estructura simplificada, en la que las covarianzas son modeladas por los v_{jk} , pues Ψ es diagonal. En (2.5) se muestra una partición de la varianza de los Y_j en una componente debida a los factores comunes, llamada *comunalidad* y otra componente única, la varianza específica, llamada *unicidad*.

$$\begin{aligned}
\sigma_{jj} = Var(Y_j) &= (v_{j1}^2 + v_{j2}^2 + \dots + v_{jm}^2) + \psi_j \\
&= h_j^2 + \psi_j \\
&= \text{comunalidad} + \text{unicidad}
\end{aligned} \tag{2.6}$$

donde

$$\begin{aligned}
\text{comunalidad} &= h_j^2 = v_{j1}^2 + v_{j2}^2 + \dots + v_{jm}^2 \\
\text{unicidad} &= \psi_j
\end{aligned}$$

La comunalidad es la parte de la varianza explicada por los factores comunes y la unicidad es la variabilidad no explicada o error experimental.

A partir de (2.5) se puede mostrar que las cargas representan la covarianza de las variables con los factores: $Cov(\mathbf{y}, \mathbf{f}) = \mathbf{V}$

Si las variables están estandarizadas, (2.5) es reemplazada por $\mathbf{R} = \mathbf{V}\mathbf{V}' + \Psi$ y las *cargas* representan las correlaciones de las variables con los factores, esto es, $Corr(\mathbf{y}, \mathbf{f}) = \mathbf{V}$, es decir, para la variable j y el factor k la correlación es: $corr(Y_j, f_k) = v_{jk}$.

Estimación de las *cargas* y las *comunalidades*

Se han propuesto varios métodos para la estimación de las *cargas* y las *comunalidades*. Se muestra aquí uno de ellos llamado el método del *componente principal*. Se comienza, en primer lugar, con la matriz de covarianza muestral \mathbf{S} como estimación de Σ y se trata de encontrar una estimación para \mathbf{V} , que se aproximará a la expresión fundamental dada en (2.5) como:

$$\mathbf{S} \cong \hat{\mathbf{V}}\hat{\mathbf{V}}' + \hat{\Psi} \quad (2.7)$$

En este método, en primer lugar, se ignora $\hat{\Psi}$ tomando $\mathbf{S} = \hat{\mathbf{V}}\hat{\mathbf{V}}'$. Con el fin de factorizar \mathbf{S} , se utiliza la descomposición espectral:

$$\mathbf{S} = \mathbf{C}\mathbf{D}\mathbf{C}'$$

donde \mathbf{C} es una matriz ortogonal construida con los autovectores normalizados ($c'_i c_i = 1$) de \mathbf{S} como columnas y $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$, una matriz diagonal con los autovalores de \mathbf{S} . Dado que los autovalores de una matriz definida positiva son positivos o ceros, se puede factorizar $\mathbf{D} = \mathbf{D}^{1/2}\mathbf{D}^{1/2}$, donde $\mathbf{D}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_p})$. Con esta factorización se obtiene:

$$\mathbf{S} = \mathbf{C}\mathbf{D}\mathbf{C}' = \mathbf{C}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{C}' = (\mathbf{C}\mathbf{D}^{1/2})(\mathbf{C}\mathbf{D}^{1/2})'$$

de esta manera se escribe \mathbf{S} como una matriz por su traspuesta, pero esta matriz es de dimensión $p \times p$, no es la dimensión que se necesita. Para ello, se define $\mathbf{D}_1 = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ con los primeros m autovalores más grandes, con $m < p$; $\lambda_1 > \lambda_2 > \dots > \lambda_m$ y $\mathbf{C}_1 = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m)$ contiene los autovectores correspondientes. Luego, se tiene una nueva expresión:

$$\mathbf{S} = \hat{\mathbf{V}}\hat{\mathbf{V}}' = (\mathbf{C}_1\mathbf{D}_1^{1/2})(\mathbf{C}_1\mathbf{D}_1^{1/2})'$$

donde $\hat{\mathbf{V}} = \mathbf{C}_1\mathbf{D}_1^{1/2} = (\sqrt{\lambda_1}\mathbf{c}_1, \sqrt{\lambda_2}\mathbf{c}_2, \dots, \sqrt{\lambda_m}\mathbf{c}_m)$. Las columnas de $\hat{\mathbf{V}}$ son proporcionales a los autovectores de \mathbf{S} , esto es, las *cargas* sobre el factor k son proporcionales a los coeficientes en la k -ésima componente principal. Así, un elemento de $\hat{\mathbf{V}}$ o sea, la *carga* para un factor es:

$$\hat{v}_{jk} = \sqrt{\lambda_k}\mathbf{c}_{jk}, \quad \text{con } j = 1, 2, \dots, p \text{ y } k = 1, 2, \dots, m. \quad (2.8)$$

El j -ésimo elemento de la diagonal de $\hat{\mathbf{V}}\hat{\mathbf{V}}'$ es la suma de cuadrados de la j -ésima fila de $\hat{\mathbf{V}}$, o $\hat{\mathbf{v}}_j'\hat{\mathbf{v}}_j = \sum_{k=1}^m \hat{v}_{jk}^2$. Luego, una estimación para la comunalidad es

$$\hat{h}_j^2 = \sum_{k=1}^m \hat{v}_{jk}^2 \quad (2.9)$$

Para completar la aproximación de \mathbf{S} en (2.7) y de acuerdo a (2.6) se define

$$\hat{\psi}_j = s_{jj} - \hat{h}_j^2 = s_{jj} - \sum_{k=1}^m \hat{v}_{jk}^2 \quad (2.10)$$

En FA es más apropiado usar la matriz de correlación \mathbf{R} en lugar de \mathbf{S} y en la mayoría de las aplicaciones se obtienen mejores resultados. Se comienza con los autovalores y autovectores de \mathbf{R} , en (2.7) se cambia \mathbf{S} por \mathbf{R} , la expresión de (2.6) ahora es: $1 = h_j^2 + \psi_j$ y por lo tanto $\hat{\psi}_j = 1 - \hat{h}_j^2$.

Cuando el objetivo es la reducción de dimensión, realizando el análisis con el *software R* se obtienen los resultados que se muestran en los diagramas de la Figura 2.1, donde están los factores encontrados y los coeficientes correspondientes a las *cargas*. B_1, B_2, \dots, B_{14} son las variables medidas y MR_1, MR_2 son los factores, se observan dos grupos. En el segundo diagrama las variables $B_1, B_2, B_4, \dots, B_9$ tienen coeficientes altos con MR_2 y $B_{10}, B_{11}, \dots, B_{14}$ con MR_1 , también se observa B_3 con coeficientes bajos para los dos factores. Se puede concluir que se forman tres grupos de variables. En FA se proponen métodos de rotación para los factores que

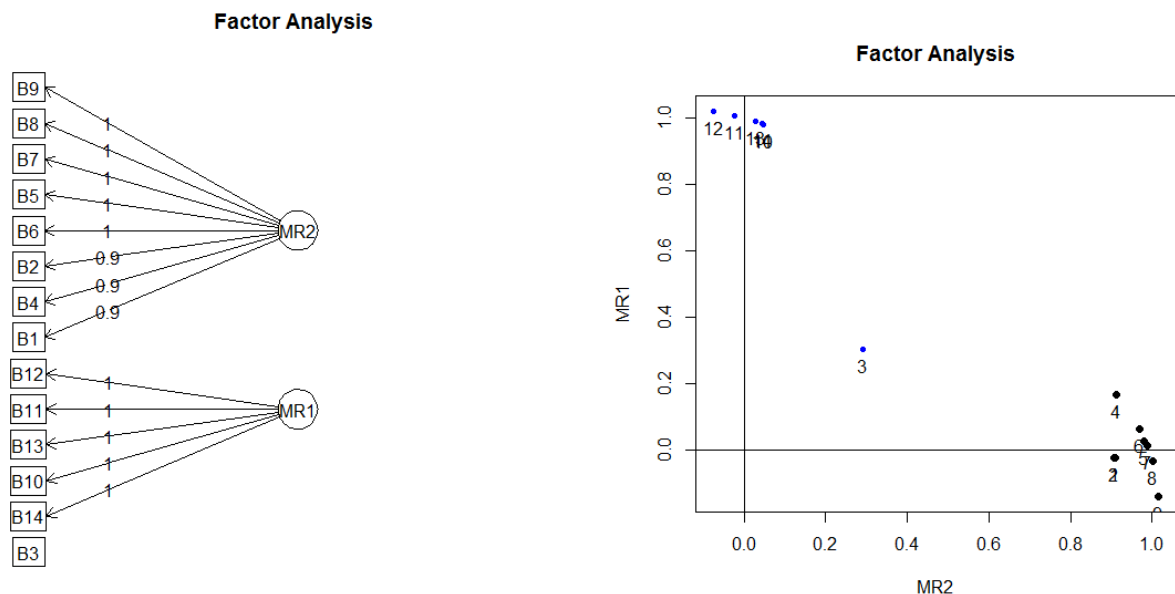


Figura 2.1: Ejemplos de diagramas obtenidos con R al aplicar Análisis por factores

ayudan a su interpretación. También se pueden estimar los factores, pero no será desarrollado aquí pues no se utilizan en este trabajo.

Metodología propuesta con FA

La Figura 2.1 muestra dos grupos de variables y una única variable, $B3$, que no está en ninguno de ellos. Una vez encontrados los grupos de variables lo que se propone en este trabajo es generar nuevas variables, basada en la idea del trabajo de Imani y Ghassemian (2014). Una vez obtenidos los grupos de variables, en cada uno se combinan las variables generando una nueva variable cuyos valores son las medias de las variables originales. Para cada vector fila (formado por las mediciones de todas las variables para una UE) se calcula la media y éstos serán los valores de la nueva variable. De acuerdo al diagrama de la Figura 2.1 se generan dos variables $NB1$, $NB2$ y la tercera variable $NB3 = B3$. Estas nuevas variables son la entrada para utilizarlas en el análisis de clasificación. Se produce una reducción de dimensión de 14 a 3, conservando la información original.

2.2.2. Análisis Discriminante Lineal

Otro método muy utilizado para la reducción de dimensión es el Análisis de Componentes Principales (ACP), suponiendo que las observaciones provienen de una sola población en estudio y el análisis se basa en la matriz de covarianzas de la población (Σ). Pero cuando se tiene conocimiento de que las UE corresponden a g poblaciones (o grupos) diferentes, lo conveniente es usar otro método para reducción de dimensión: el Análisis Discriminante Lineal (ADL), ya que tiene en cuenta la variabilidad de cada población ($\Sigma_1, \dots, \Sigma_g$). La información de los grupos o clases ayuda a revelar la estructura de los datos.

Suponiendo que se midieron p variables sobre las UE de cada población, las *funciones discriminantes* son las combinaciones lineales, de las variables originales, que separan mejor los grupos. El objetivo del ADL es identificar la contribución relativa de las p variables a la separación de los grupos y encontrar el hiperplano óptimo donde los puntos pueden ser proyectados para revelar mejor la configuración de los grupos. Las funciones lineales que contribuyen a la separación de los grupos, se conocen frecuentemente como variables canónicas o coordenadas discriminantes. Para evitar confusiones, es preferible reservar el término canónica sólo para el análisis de correlación canónica. El ADL es un método supervisado para reducción de dimensión en problemas de clasificación.

La función discriminante para dos grupos

Se supone que se desea comparar dos poblaciones, que tienen la misma matriz de covarianza $\Sigma_1 = \Sigma_2 = \Sigma$ pero distintos vectores de medias μ_1 y μ_2 . De cada población se selecciona una muestra aleatoria: $\mathbf{y}_{11}, \mathbf{y}_{12}, \dots, \mathbf{y}_{1n_1}$ y $\mathbf{y}_{21}, \mathbf{y}_{22}, \dots, \mathbf{y}_{2n_2}$, respectivamente. Cada vector \mathbf{y}_{1i} consiste en mediciones sobre las p variables, $\mathbf{y}'_{1i} = (y_{1i1}, y_{1i2}, \dots, y_{1ip})$, de la misma forma

$\mathbf{y}'_{2i} = (y_{2i1}, y_{2i2}, \dots, y_{2ip})$, con $i = 1, 2, \dots, n_1$ o $i = 1, 2, \dots, n_2$ según corresponda al grupo 1 o 2, con $n_1 + n_2 = n$. Se desea encontrar la dirección, definida por un vector $\mathbf{a}' = (a_1, a_2, \dots, a_p)$, tal que cuando los datos son proyectados sobre \mathbf{a} , las muestras de cada población estén separadas tanto como sea posible.

La combinación lineal $z = \mathbf{a}'\mathbf{y}$ es la proyección de \mathbf{y} en la dirección de \mathbf{a} , reduciendo la dimensión de p a 1 pues transforma cada vector de observaciones a un escalar:

$$z_{1i} = \mathbf{a}'\mathbf{y}_{1i} = a_1 y_{1i1} + a_2 y_{1i2} + \dots + a_p y_{1ip}, \quad i = 1, 2, \dots, n_1$$

$$z_{2i} = \mathbf{a}'\mathbf{y}_{2i} = a_1 y_{2i1} + a_2 y_{2i2} + \dots + a_p y_{2ip}, \quad i = 1, 2, \dots, n_2.$$

La función discriminante es la combinación lineal de estas p variables que maximiza la distancia entre los vectores de medias transformadas, \bar{z}_1 y \bar{z}_2 , de los dos grupos. Las medias son $\bar{z}_1 = \sum_{i=1}^{n_1} z_{1i}/n_1 = \mathbf{a}'\bar{\mathbf{y}}_1$ y $\bar{z}_2 = \mathbf{a}'\bar{\mathbf{y}}_2$, por ser una combinación lineal. Se desea encontrar el vector \mathbf{a} que maximiza la diferencia estandarizada $(\bar{z}_1 - \bar{z}_2)/s_z$. Como esta diferencia puede ser negativa se usa la distancia al cuadrado $(\bar{z}_1 - \bar{z}_2)^2/s_z^2$ la cual puede ser expresada:

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{s_z^2} = \frac{[\mathbf{a}'(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)]^2}{\mathbf{a}'\mathbf{S}_{pl}\mathbf{a}} \quad (2.11)$$

En (2.11) \mathbf{S}_{pl} es un estimador insesgado de la matriz de covarianza común de la población Σ :

$$\mathbf{S}_{pl} = \frac{1}{n_1 + n_2 - 2} (\mathbf{W}_1 + \mathbf{W}_2)$$

donde \mathbf{W}_1 y \mathbf{W}_2 son las matrices de sumas de cuadrados y productos cruzados para las dos muestras:

$$\mathbf{W}_1 = \sum_{i=1}^{n_1} (\mathbf{y}_{1i} - \bar{\mathbf{y}}_1)(\mathbf{y}_{1i} - \bar{\mathbf{y}}_1)' = (n_1 - 1)\mathbf{S}_1,$$

$$\mathbf{W}_2 = \sum_{i=1}^{n_2} (\mathbf{y}_{2i} - \bar{\mathbf{y}}_2)(\mathbf{y}_{2i} - \bar{\mathbf{y}}_2)' = (n_2 - 1)\mathbf{S}_2$$

Puesto que $(n_1 - 1)\mathbf{S}_1$ es un estimador insesgado de $(n_1 - 1)\Sigma$ y $(n_2 - 1)\mathbf{S}_2$ es un estimador insesgado de $(n_2 - 1)\Sigma$, se combinan para obtener \mathbf{S}_{pl} .

El máximo ocurre cuando $\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$, o cuando \mathbf{a} es cualquier múltiplo de $\mathbf{S}_{pl}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$.

Luego el vector \mathbf{a} no es único, sin embargo su dirección si es única; éstos es valores relativos o razones de a_1, a_2, \dots, a_p son únicos y $z = \mathbf{a}'\mathbf{y}$ proyecta a \mathbf{y} sobre la línea en que $(\bar{z}_1 - \bar{z}_2)^2/s_z^2$ se maximiza. Notar que para que \mathbf{S}_{pl}^{-1} exista, debe ser $n_1 + n_2 - 2 > p$. La dirección óptima dada por $\mathbf{a} = \mathbf{S}_{pl}^{-1}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$ es paralela a la línea que une $\bar{\mathbf{y}}_1$ con $\bar{\mathbf{y}}_2$ porque $(\bar{z}_1 - \bar{z}_2)^2/s_z^2$ es equivalente a la distancia estandarizada entre $\bar{\mathbf{y}}_1$ y $\bar{\mathbf{y}}_2$ pues sustituyendo \mathbf{a} en (2.11) obtenemos :

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{s_z^2} = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \mathbf{S}_{pl}^{-1} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) \quad (2.12)$$

La Figura 2.2 ilustra la separación de dos grupos (de dimensión $p = 2$ variables) en un espacio de dimensión uno, representado por la función discriminante z .

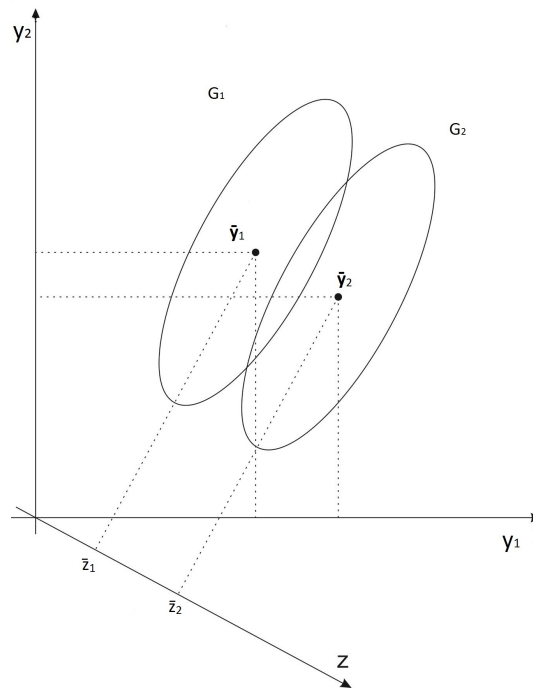


Figura 2.2: Separación lograda por la función discriminante para dos grupos

Funciones Discriminantes para varios grupos

En el análisis discriminante para varios grupos, se está interesado en la búsqueda de combinaciones lineales de las variables que separan mejor g grupos de observaciones multivariadas. El análisis discriminante para varios grupos puede servir para diversos objetivos:

- 1 Examinar la separación de los grupos en un gráfico de dos dimensiones. Cuando hay más de dos grupos, se requiere más de una función discriminante para describir la separación

entre grupos. Si se proyectan los puntos del espacio p -dimensional en uno bidimensional, representado por las dos primeras funciones discriminantes, se obtiene la mejor vista posible de la forma de separar los grupos.

- 2 Encontrar un subconjunto de las variables originales que separa a los grupos casi tan bien como el conjunto original.
- 3 Ordenar las variables en términos de su contribución relativa a la separación de los grupos.

Para g grupos (muestras) con n_k observaciones en el k -ésimo grupo, se transforma el vector \mathbf{y}_{ki} para obtener $z_{ki} = \mathbf{a}'\mathbf{y}_{ki}$ con $k = 1, 2, \dots, g$ y $i = 1, 2, \dots, n_k$ y encontrar las medias $\bar{z}_k = \mathbf{a}'\bar{\mathbf{y}}_{k\bullet}$ donde $\bar{\mathbf{y}}_{k\bullet} = \sum_{i=1}^{n_k} \mathbf{y}_{ki}/n_k$, o sea, es el vector de medias de las p variables para el grupo k , $\bar{\mathbf{y}}_{k\bullet} = (\bar{y}_{k\bullet 1}, \bar{y}_{k\bullet 2}, \dots, \bar{y}_{k\bullet p})'$. Como en el caso de dos grupos se busca el vector \mathbf{a} que maximice la separación entre las medias $\bar{z}_1, \bar{z}_2, \dots, \bar{z}_g$. Para expresar la separación entre estas medias se extiende la expresión (2.11) a g grupos. En el caso de dos grupos, retomando (2.11) y teniendo en cuenta que $\mathbf{a}'(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2) = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{a}$, se puede escribir:

$$\frac{(\bar{z}_1 - \bar{z}_2)^2}{s_z^2} = \frac{[\mathbf{a}'(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)]^2}{\mathbf{a}'\mathbf{S}_{pl}\mathbf{a}} = \frac{\mathbf{a}'(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{a}}{\mathbf{a}'\mathbf{S}_{pl}\mathbf{a}}. \quad (2.13)$$

Fisher introduce el Análisis Discriminante basado en la idea de Análisis de la Varianza (ANOVA), Fisher (1936), de tal manera que la variación entre grupos sea máxima y la variación dentro de los grupos sea lo menor posible. Entonces para extender (2.13) a g grupos se usa la matriz \mathbf{H} en lugar de $(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'$ y la matriz \mathbf{E} en lugar de \mathbf{S}_{pl} , Rencher (2002), para obtener:

$$\lambda = \frac{\mathbf{a}'\mathbf{H}\mathbf{a}}{\mathbf{a}'\mathbf{E}\mathbf{a}} \quad (2.14)$$

donde \mathbf{H} y \mathbf{E} son las matrices “entre” y “dentro”, respectivamente y están definidas como:

$$\mathbf{H} = n \sum_{k=1}^g (\bar{\mathbf{y}}_{k\bullet} - \bar{\mathbf{y}}_{\bullet\bullet})(\bar{\mathbf{y}}_{k\bullet} - \bar{\mathbf{y}}_{\bullet\bullet})' \text{ y } \mathbf{E} = \sum_{k=1}^g \sum_{i=1}^{n_k} (\mathbf{y}_{ki} - \bar{\mathbf{y}}_{k\bullet})(\mathbf{y}_{ki} - \bar{\mathbf{y}}_{k\bullet})'$$

Estas sumas de cuadrados representan lo mismo que en ANOVA, es decir, la variación entre grupos y la variación dentro de los grupos, por lo tanto λ mide la proporción de la variación entre grupos con respecto a la variación dentro. Valores grandes de λ indican alta variación entre grupos, lo que muestra la separación entre ellos. Se puede escribir (2.14) como

$$\mathbf{a}'\mathbf{H}\mathbf{a} = \lambda\mathbf{a}'\mathbf{E}\mathbf{a}$$

$$\mathbf{a}'(\mathbf{H}\mathbf{a} - \lambda\mathbf{E}\mathbf{a}) = 0. \quad (2.15)$$

Se examinan los valores de λ y de \mathbf{a} que son soluciones de (2.15), buscando valores de \mathbf{a} que produzcan un λ máximo. Para encontrar la solución no trivial se debe resolver: $\mathbf{H}\mathbf{a} - \lambda\mathbf{E}\mathbf{a} = 0$, la que se puede reescribir como:

$$(\mathbf{E}^{-1}\mathbf{H} - \lambda\mathbf{I})\mathbf{a} = 0 \quad (2.16)$$

Las soluciones de (2.16) son los autovalores $\lambda_1, \lambda_2, \dots, \lambda_s$ y autovectores asociados $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$ de $\mathbf{E}^{-1}\mathbf{H}$. Se toman los autovalores de tal manera que λ_1 es el autovalor más grande y así sucesivamente. Los s autovalores, distintos de cero, dan el rango de la matriz \mathbf{H} y s es el mínimo entre $g - 1$ y p . Así, el valor propio más grande λ_1 es el valor máximo de $\lambda = \mathbf{a}'\mathbf{H}\mathbf{a}/\mathbf{a}'\mathbf{E}\mathbf{a}$ en (2.14) y el vector de coeficientes que produce el máximo es el correspondiente autovector \mathbf{a}_1 . Por lo tanto, la función discriminante que maximiza la separación entre las medias es $z_1 = \mathbf{a}'_1\mathbf{y}$, z_1 representa la dimensión o la dirección de máxima separación de las medias.

De los s autovectores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_s$ de $\mathbf{E}^{-1}\mathbf{H}$ correspondientes a $\lambda_1, \lambda_2, \dots, \lambda_s$ obtenemos s funciones discriminantes $z_1 = \mathbf{a}'_1\mathbf{y}, z_2 = \mathbf{a}'_2\mathbf{y}, \dots, z_s = \mathbf{a}'_s\mathbf{y}$, las cuales muestran las dimensiones o direcciones de las diferencias entre $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_g$. Estas funciones discriminantes no están correlacionadas pero no son ortogonales ($\mathbf{a}'_i\mathbf{a}_j = 0$ para $i \neq j$), porque $\mathbf{E}^{-1}\mathbf{H}$ no es simétrica Rencher (1998). Notar que la numeración z_1, z_2, \dots, z_s corresponde a los autovalores y no a los g grupos.

Observaciones:

1. La matriz $\mathbf{E}^{-1}\mathbf{H}$ no es simétrica. Muchos algoritmos para el cálculo de autovalores y autovectores sólo aceptan matrices simétricas. Se puede demostrar que los autovalores de la matriz simétrica $(\mathbf{U}^{-1})'\mathbf{H}\mathbf{U}^{-1}$ son los mismos que los de $\mathbf{E}^{-1}\mathbf{H}$, donde $\mathbf{E} = \mathbf{U}'\mathbf{U}$ es la factorización de Cholesky de \mathbf{E} . Sin embargo, un ajuste es necesario para los autovectores. Si \mathbf{b} es un autovector de $(\mathbf{U}^{-1})'\mathbf{H}\mathbf{U}^{-1}$, entonces $\mathbf{a} = \mathbf{U}^{-1}\mathbf{b}$ es un autovector de $\mathbf{E}^{-1}\mathbf{H}$.
2. Los tamaños de muestras diferentes no presentan problemas, pero el más pequeño n_i debería ser mayor a p , éste no es un requerimiento matemático pero ayudará a producir funciones discriminantes más estables.

La importancia de cada función discriminante z_k se puede evaluar calculando la proporción de su autovalor sobre la suma de los autovalores:

$$\frac{\lambda_k}{\sum_{j=1}^s \lambda_j} \quad (2.17)$$

Por este criterio, dos o tres funciones discriminantes serán suficientes para describir las diferencias entre grupos; las funciones discriminantes asociadas con los autovalores más pequeños

pueden ser descartadas.

Estandarización de las funciones discriminantes

La contribución relativa de las variables Y_1, Y_2, \dots, Y_p para la separación de los grupos se puede evaluar mediante la comparación de los coeficientes a_r , $r = 1, 2, \dots, p$, en la función discriminante: $z = \mathbf{a}'\mathbf{y} = a_1Y_1 + a_2Y_2 + \dots + a_pY_p$. Sin embargo, esto sólo se puede realizar si las Y_j fueron medidas en la misma escala y con varianzas comparables. Si no es así, es necesario primero estandarizar cada variable y luego comparar los coeficientes \mathbf{a}^* , obtenidos con estas nuevas variables. Los coeficientes estandarizados reflejan, correctamente, la contribución conjunta de las variables a la función discriminante z , para separar los grupos. Estos coeficientes se pueden obtener a partir de los (a_1, \dots, a_p) obtenidos para las variables sin estandarizar. Para el caso de dos grupos: $\mathbf{a}^* = (\text{diag } \mathbf{S}_{pl})^{1/2}\mathbf{a}$. Para el caso de varios grupos: dado que el m -ésimo autovector es único, multiplicando por un escalar se puede obtener $a_{mr}^* = \sqrt{e_{rr}}a_{mr}$ con $r = 1, 2, \dots, p$, donde e_{rr} es el r -ésimo elemento de la diagonal de \mathbf{E} . Esto muestra que para realizar la *estandarización* se utiliza \mathbf{E} , que es la matriz de covarianza *dentro* de grupos.

Cada vector de coeficientes de la función discriminante $\mathbf{a}' = (a_1, a_2, \dots, a_p)$ es un autovector de $\mathbf{E}^{-1}\mathbf{H}$, y como tal, tiene en cuenta las correlaciones entre las variables de la muestra, así como la influencia de cada variable en presencia de las otras.

Los coeficientes estandarizados reflejan correctamente la contribución conjunta de las variables a la función discriminante z , para separar los grupos.

Contribución de las funciones discriminantes

Estas nuevas variables definen un nuevo espacio de dimensión menor $s = \min(p, g - 1)$ y podrían utilizarse como las variables de entrada para análisis posteriores, por ejemplo, para cualquier algoritmo de clasificación.

Otra manera para analizar la contribución es calcular la correlación entre cada variable y la función discriminante, $r_{y_i z_j}$. Este método es el más recomendado, pero es el menos útil, ya que muestra sólo cómo cada variable por sí misma separa los grupos, haciendo caso omiso de la presencia de las otras variables. Las correlaciones se refieren a menudo como cargas o coeficientes de estructura y están proporcionados como rutinas en muchos programas. Sin embargo, en Rencher (1998) se muestra que las correlaciones en cuestión revelan la contribución de cada variable en un contexto univariado, en lugar de multivariado. Por lo tanto, estas correlaciones no proporcionan ninguna información acerca de cómo las variables contribuyen conjuntamente a la separación de los grupos.

La objeción a los coeficientes estandarizados se basa en el argumento de que son “inestables”

ya que cambian cuando se eliminan algunas variables y se añaden otras. Sin embargo, en realidad lo que se desea es que se comporten de esta manera, a fin de reflejar la influencia mutua de las variables entre sí. En un análisis multivariado, el interés se centra en la actuación conjunta de la serie de variables que son de interés. Si n/p es demasiado pequeño, las variables importantes en una muestra pueden ser menos importante en otra muestra, por ello la cantidad de variables a analizar debe ir acompañada por tamaños de muestras grandes.

En consecuencia, para la interpretación de las funciones discriminantes se recomiendan los coeficientes estandarizados en lugar de las correlaciones Rencher (2002).

Metodología propuesta para ADL

En este trabajo *no* se utilizan las *funciones discriminantes* como variables de entrada para la *clasificación*. Lo que se propone es seleccionar aquellas variables originales que tengan mayor contribución a la separación de los grupos, y para ello se procede de la siguiente manera:

1. seleccionar las funciones discriminantes con mayor porcentaje de contribución a la variación entre grupos, de acuerdo al criterio,
2. examinar los coeficientes de la función discriminante estandarizados,
3. ordenar los coeficientes en valor absoluto, ya que en la determinación de la contribución, los signos son ignorados,
4. seleccionar aquellas variables originales que muestran mayores coeficientes en las funciones discriminantes seleccionadas.

2.3. Clasificadores utilizados

La clasificación consiste en la asignación de una UE (sujeto u objeto) a uno de diferentes grupos o clases predefinidos, sobre la base de los vectores Y_1, Y_2, \dots, Y_p asociados a la unidad. Si se asume que hay \mathcal{L} clases, rotuladas ω_1 a $\omega_{\mathcal{L}}$, organizadas como un conjunto de rótulos $\Omega = \{\omega_1, \dots, \omega_{\mathcal{L}}\}$ y cada objeto pertenece a una y solo una clase, un *clasificador* es cualquier función $\mathcal{D} : \mathbb{R}^p \rightarrow \Omega$. En el modelo canónico de un clasificador, se considera un conjunto de \mathcal{L} funciones discriminantes $G = \{g_1(\mathbf{y}_{i\bullet}), \dots, g_{\mathcal{L}}(\mathbf{y}_{i\bullet})\}$, $g_k : \mathbb{R}^p \rightarrow \mathbb{R}, k = 1, \dots, \mathcal{L}$, cada una da una puntuación para la respectiva clase. Normalmente, $\mathbf{y}_{i\bullet}$ se etiqueta en la clase con la puntuación más alta. Esta opción de etiquetado (diagrama de la Figura 2.3) se denomina regla

de membresía máxima, esto es,

$$\mathcal{D}(\mathbf{y}_{i\bullet}) = \omega_{k^*} \in \Omega \Leftrightarrow g_{k^*}(\mathbf{y}_{i\bullet}) = \max_{k=1,\dots,\mathcal{L}} (g_k(\mathbf{y}_{i\bullet})) \quad (2.18)$$

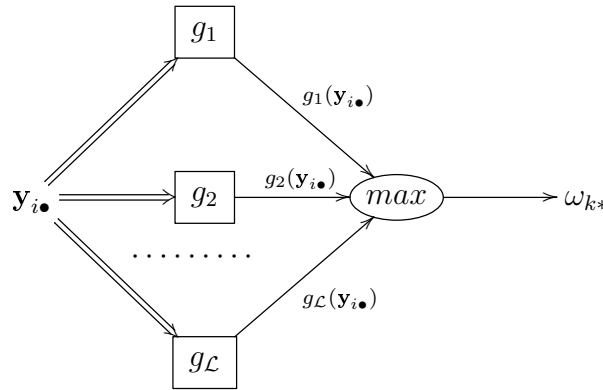


Figura 2.3: Diagrama de un modelo canónico de un clasificador

Si ocurre un empate $\mathbf{y}_{i\bullet}$ es asignado aleatoriamente a una de las clases empatadas.

Las funciones discriminantes particionan el espacio de las variables, \mathfrak{R}^p , en \mathcal{L} *regiones de decisión* o *de clasificación*, indicadas por $\mathcal{R}_1, \dots, \mathcal{R}_{\mathcal{L}}$

$$\mathcal{R}_k = \left\{ \mathbf{y}_{i\bullet} \mid \mathbf{y}_{i\bullet} \in \mathfrak{R}^p, g_{k^*}(\mathbf{y}_{i\bullet}) = \max_{k=1,\dots,\mathcal{L}} (g_k(\mathbf{y}_{i\bullet})) \right\}, k = 1, \dots, \mathcal{L}$$

La región de decisión para la clase ω_k es el conjunto de puntos para los cuales la k –ésima función discriminante tiene el puntaje más alto. De acuerdo a (2.18) todos los puntos en la región de decisión \mathcal{R}_k son asignados en la clase ω_k . Estas regiones son especificadas por el clasificador \mathcal{D} , o equivalentemente, por las funciones discriminantes G .

Los clasificadores utilizados en este trabajo son del tipo supervisado, que es cuando se cuenta con información de cada uno de los grupos, mediante muestras obtenidas de cada clase (ME). En campos de la ingeniería o de la informática, la clasificación se suele llamar *reconocimiento de patrones*.

En este trabajo se utilizan los clasificadores Análisis Discriminante Lineal (ADL), Análisis Discriminante Cuadrático (ADC), Máquinas de Soporte Vectorial (SVM) y otro que es combinación de los anteriores y se lo llama COMBinado (COMB).

2.3.1. Clasificador Análisis Discriminante Lineal

El clasificador ADL surge cuando se emplean funciones lineales, llamadas *funciones de clasificación lineal*, para asignar una UE a uno de los grupos. Los valores medidos en el vector de

observaciones para un individuo u objeto son evaluados por las funciones de clasificación para encontrar el grupo al que el individuo muy probablemente pertenece.

Cuando se tienen dos poblaciones se puede utilizar el enfoque de Fisher (1936). El principal supuesto del procedimiento de Fisher es que $(\Sigma_1 = \Sigma_2)$, las dos matrices de covarianza son iguales. No se requiere el supuesto de Normalidad. Se selecciona una muestra de cada población y se calculan $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2$ y \mathbf{S}_{pl} . Un simple procedimiento de clasificación se basa en la función de clasificación $z = \mathbf{a}'\mathbf{y} = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{S}_{pl}^{-1}\mathbf{y}$ donde \mathbf{y} corresponde a las medidas de una nueva UE que se desea clasificar en uno de dos grupos (G_1 o G_2). Determinar si \mathbf{y} está cerca de $\bar{\mathbf{y}}_1$ o de $\bar{\mathbf{y}}_2$ es equivalente a ver si z está más próximo a \bar{z}_1 o a \bar{z}_2 . Según lo que se muestra en la Figura 2.2, se determina el punto medio $\frac{1}{2}(\bar{z}_1 + \bar{z}_2)$ y se puede ver que z estará cercano a \bar{z}_2 si $z > \frac{1}{2}(\bar{z}_1 + \bar{z}_2)$. Expresando en términos de \mathbf{y} , la regla de clasificación será:

$$\begin{aligned} \text{Asignar } \mathbf{y} \text{ a } G_2 \text{ si } \mathbf{a}'\mathbf{y} = (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{S}_{pl}^{-1}\mathbf{y} > \frac{1}{2}(\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)'\mathbf{S}_{pl}^{-1}(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2) \\ \text{caso contrario se asigna a } G_1 \end{aligned}$$

Así, en el caso de dos grupos, la función discriminante también sirve como una función de clasificación lineal. Sin embargo, en el caso de varios grupos las funciones de clasificación son *diferentes* de las funciones discriminantes descriptivas, presentadas en la Sección 2.2.2. La regla de clasificación presentada es esencialmente no paramétrica, porque no se hicieron hipótesis de distribución. Sin embargo, si las dos poblaciones son normales con matrices de covarianza iguales, este método es óptimo; es decir, la probabilidad de errores de clasificación se reduce al mínimo y, se dice asintóticamente óptimo cuando se utilizan estimadores para los parámetros.

Pensando a la Teoría de la Probabilidad como el marco de estudio de la toma de decisiones bajo incertidumbre, en clasificación, la regla de Bayes se usa para calcular las probabilidades de las clases, tomando decisiones racionales minimizando el riesgo esperado. Se definen a continuación las probabilidades utilizadas.

Se supone que se tienen k clases mutuamente excluyentes y exhaustivas $G_i, i = 1, \dots, k$, y se conocen las probabilidades de las clases, llamadas *probabilidades a priori* (p_i), que satisfacen $P(G_i) \geq 0$ y $\sum_{i=1}^k P(G_i) = 1$.

La probabilidad condicional $P(\mathbf{y}|G_i)$, llamada la verosimilitud de la clase, es la probabilidad de obtener un valor observado \mathbf{y} dado que ocurre la clase G_i .

La probabilidad *a posteriori* de la clase G_i puede ser calculada, por el Teorema de Bayes,

como:

$$P(G_i|\mathbf{y}) = \frac{P(\mathbf{y}|G_i)P(G_i)}{p(\mathbf{y})} = \frac{P(\mathbf{y}|G_i)P(G_i)}{\sum_{j=1}^k P(\mathbf{y}|G_j)P(G_j)} \quad (2.19)$$

$p(\mathbf{y})$ es llamada la *evidencia* y entonces (2.19) es:

$$\text{posteriori} = \frac{\text{verosimilitud} \times \text{priori}}{\text{evidencia}}$$

Minimizando el error, el clasificador de Bayes elige la clase con la más alta probabilidad *a posteriori*; esto es,

$$\boxed{\text{se selecciona } G_i \text{ si } P(G_i|\mathbf{y}) = \max_j P(G_j|\mathbf{y})} \quad (2.20)$$

Suponiendo varianzas iguales ($\Sigma_1 = \Sigma_2 = \dots = \Sigma_k$), se estima la varianza común con

$$S_{pl} = \frac{1}{n-k} \sum_{i=1}^k (n_i - 1) \mathbf{S}_i = \frac{\mathbf{E}}{n-k}$$

donde n_i y \mathbf{S}_i son el tamaño de la muestra y la matriz de covarianza del grupo i , respectivamente. \mathbf{E} es la matriz de error como se definió antes y $n = \sum n_i$. Comparamos \mathbf{y} con cada $\bar{\mathbf{y}}_i$, $i = 1, 2, \dots, k$ mediante la función que es el cuadrado de la distancia

$$D_i^2 = (\mathbf{y} - \bar{\mathbf{y}}_i)' \mathbf{S}_{pl}^{-1} (\mathbf{y} - \bar{\mathbf{y}}_i) \quad (2.21)$$

y asigna \mathbf{y} al grupo para el cuál D_i^2 es más pequeña. Desarrollando (2.21), eliminando el término $\mathbf{y}' \mathbf{S}_{pl}^{-1} \mathbf{y}$, ya que no depende de i y multiplicando por $-\frac{1}{2}$ se obtienen las *funciones de clasificación* que se denotaran como $\mathbf{L}_i(\mathbf{y})$ y la regla de clasificación lineal queda:

Asignar \mathbf{y} al grupo para el cual

$$\mathbf{L}_i(\mathbf{y}) = \bar{\mathbf{y}}_i' \mathbf{S}_{pl}^{-1} \mathbf{y}_i - \frac{1}{2} \bar{\mathbf{y}}_i' \mathbf{S}_{pl}^{-1} \bar{\mathbf{y}}_i \quad (2.22)$$

es *máximo* (se revierte el signo al multiplicar por $-1/2$)

Para poner de relieve la linealidad de (2.22) como función de \mathbf{y} se la puede expresar como:

$$\mathbf{L}_i(\mathbf{y}) = \mathbf{c}'_i \mathbf{y} + c_{i0} = c_{i1}y_1 + c_{i2}y_2 + \dots + c_{ip}y_p + c_{i0}$$

donde $\mathbf{c}'_i = \bar{\mathbf{y}}_i' \mathbf{S}_{pl}^{-1}$ y $c_{i0} = -\frac{1}{2} \bar{\mathbf{y}}_i' \mathbf{S}_{pl}^{-1} \bar{\mathbf{y}}_i$. Para asignar \mathbf{y} a un grupo usando este procedimiento,

se calculan c_i y c_{i0} para cada uno de los k grupos o clases, evaluando $\mathbf{L}_i(\mathbf{y})$, $i = 1, 2, \dots, k$, y localizando \mathbf{y} al grupo para el cual $\mathbf{L}_i(\mathbf{y})$ es más grande. Éste será el mismo grupo para el cual $D_i^2(y)$ en (2.21) es más pequeña, es decir, el grupo cuyo vector de medias $\bar{\mathbf{y}}_i$ está más cerca de \mathbf{y} .

Si se utiliza la información de las probabilidades *a priori*, p_1, p_2, \dots, p_k , suponiendo matriz de covarianzas iguales y normalidad multivariada $\mathbf{y} \sim N_p(\mu_i, \Sigma)$ en cada uno de los i grupos, usando lo del Teorema de Bayes y aplicando logaritmo las $\mathbf{L}_i(\mathbf{y})$ ahora es:

$$\mathbf{L}'_i(\mathbf{y}) = \ln(p_i) + \bar{\mathbf{y}}'_i \mathbf{S}_{pl}^{-1} \mathbf{y}_i - \frac{1}{2} \bar{\mathbf{y}}'_i \mathbf{S}_{pl}^{-1} \bar{\mathbf{y}}_i, \quad (2.23)$$

y asigna \mathbf{y} al grupo con máximo $\mathbf{L}'_i(\mathbf{y})$. Notar que si $p_1 = p_2 = \dots, p_k$, entonces (2.23), que optimiza la tasa de clasificación para la distribución normal, se reduce a (2.22), que se basa en el enfoque heurístico de reducir al mínimo la distancia de \mathbf{y} a $\bar{\mathbf{y}}_i$.

Las funciones lineales $\mathbf{L}_i(\mathbf{y})$ definidas en (2.22) son llamadas *funciones de clasificación lineal* (muchos autores se refieren como funciones discriminantes lineales). Ellas son diferentes de las funciones discriminantes presentadas en la sección 2.2.2, cuyos coeficientes son los autovectores de $\mathbf{E}^{-1}\mathbf{H}$. De hecho, habrá k funciones de clasificación y $s = \min(p, k - 1)$ funciones discriminantes, donde k es el número de grupos y p es el número de variables. En muchos casos no se necesitan las s funciones discriminantes para describir de manera efectiva las diferencias entre grupos, mientras que todas las k funciones de clasificación deben ser utilizados en la asignación de las observaciones a los grupos.

2.3.2. Clasificador Análisis Discriminante Cuadrático

En la sección anterior se presentaron funciones de clasificación suponiendo matriz de covarianzas iguales. Si $\Sigma_1 = \Sigma_2 = \dots = \Sigma_k$ no vale, la regla de clasificación es cambiada usando \mathbf{S}_i (matriz de covarianza para el grupo i) en lugar \mathbf{S}_{pl} y queda

$$D_i^2 = (\mathbf{y} - \bar{\mathbf{y}}_i)' \mathbf{S}_i^{-1} (\mathbf{y} - \bar{\mathbf{y}}_i), \quad i = 1, 2, \dots, k \quad (2.24)$$

y asigna \mathbf{y} al grupo para el cuál D_i^2 es más pequeña. Para esta situación la regla que se obtiene es la regla de clasificación cuadrática.

Asumiendo normalidad, con matrices de covarianza desiguales y probabilidades a priori p_1, p_2, \dots, p_k la *Regla de Clasificación Cuadrática* es:

$$Q_i(\mathbf{y}) = \ln(p_i) - \frac{1}{2} \ln |\mathbf{S}_i| - \frac{1}{2} (\mathbf{y} - \bar{\mathbf{y}}_i)' \mathbf{S}_i^{-1} (\mathbf{y} - \bar{\mathbf{y}}_i) \quad (2.25)$$

y asigna \mathbf{y} al grupo con máximo $\mathbf{Q}_i(\mathbf{y})$.

Si se suponen las p_i iguales, el término $\ln(p_i)$ es eliminado. Para esta regla el tamaño de las muestras en cada grupo, n_i , tienen que ser mayores a la cantidad de variables p , debido a la cantidad de parámetros a estimar. Si eso no ocurre los estimadores se vuelven inestables. Esta restricción no se necesita en las *funciones de clasificación lineal*, ya que se estiman menos parámetros.

2.3.3. Las Máquinas de Soporte Vectorial

Las Máquinas de Soporte Vectorial (SVM) fueron propuestas por Vapnik como un sistema de aprendizaje automático basado en la teoría de aprendizaje estadístico Vapnik (1995). Este método ha captado la atención de los investigadores debido a su exitoso rendimiento en diferentes áreas como el reconocimiento de rostros, categorización textual, predicciones, recuperación de imágenes y reconocimiento de escritura manuscrita. Una de las características destacables es su excelente capacidad de generalización aún en espacios de alta dimensionalidad y utilizando conjuntos de entrenamiento pequeños.

SVM es un método de clasificación supervisada, que emplea un algoritmo de optimización para determinar la frontera óptima entre dos clases. Se puede generalizar para múltiples grupos. El caso más simple es el linealmente separable, donde existe una distancia positiva entre ambos grupos y es posible elegir un hiperplano de separación maximizando la distancia de éste a cada grupo. Para decidir a qué grupo pertenece una nueva observación, se considera el signo de la función que define al hiperplano de separación. En el caso más complejo (no separable linealmente), los grupos se superponen espacialmente haciendo imposible la separación por medio de un hiperplano. Una alternativa es cambiar la frontera lineal por una no lineal. SVM lo hace de una manera indirecta que resulta computacionalmente eficiente: se apela a una transformación no lineal del espacio de origen en otro de mayor dimensión en el cual los dos grupos transformados quedan separados linealmente. Luego, es posible proceder como en el primer caso. La transformación inversa deforma dicho hiperplano en una frontera no lineal que separa a los dos grupos en el espacio original. Existe hasta aquí un grave problema: se produce un sobre-ajuste (análogo al de regresión si se tomara un polinomio interpolatorio que ajuste todos los puntos). En efecto, la transformación que separa perfectamente ambos grupos no sólo contiene la información que distingue a un grupo del otro, sino que contiene todo el ruido como si fuese información relevante. Una manera de lograr que la frontera separe sólo por información relevante, ignorando el ruido, es tolerar que algunos puntos queden mal asignados a un grupo. Para ello, se introducen variables de “*holgura*” que cuantifican la distancia de un punto que se encuentra del lado equivocado de la frontera de separación. De este modo, ambas cuestiones se

formulan como problemas de programación cuadrática y sólo se necesita conocer el producto interno de los puntos transformados que, como función de los datos originales, se denomina *función núcleo* (Alpaydin, 2010; Vargas et al., 2012).

Clases linealmente separables

Supongamos que se tienen dos clases y se usan los rótulos -1 y $+1$ para las dos clases. Se tiene la muestra $((\mathbf{y}_{1\bullet}, r_1), (\mathbf{y}_{2\bullet}, r_2), \dots, (\mathbf{y}_{n\bullet}, r_n))$ donde $r_i = +1$ si $\mathbf{y}_{i\bullet} \in C_1$ y $r_i = -1$ si $\mathbf{y}_{i\bullet} \in C_2$. Se desea encontrar \mathbf{w} y w_0 tal que $g(\mathbf{y}) = \mathbf{w}'\mathbf{y} + w_0 \geq +1$ para $r = 1$

$$g(\mathbf{y}) = \mathbf{w}'\mathbf{y} + w_0 \leq -1 \text{ para } r = -1$$

donde \mathbf{w} es un vector normal al hiperplano y, por conveniencia en las interpretaciones geométricas, se lo supone de norma uno $(\frac{\mathbf{w}}{\|\mathbf{w}\|})$, lo cual puede reescribirse como

$$r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) \geq 1$$

No sólo se desea que los vectores de observaciones (UE) estén en el lado derecho del hiperplano, sino que también se quiere que estén a cierta distancia, para una mejor generalización. Se denomina *margen* a la distancia desde el hiperplano a los puntos más cercanos de cada lado. Existen infinitos hiperplanos para separar a los grupos, pero el óptimo es aquel que maximiza el margen, como se observa en la Figura 2.4, donde la línea de trazo grueso es el hiperplano, las

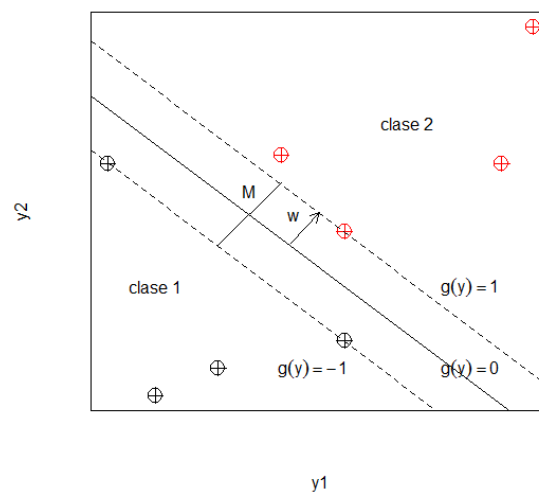


Figura 2.4: Ejemplo de dos clases, linealmente separables, con hiperplano óptimo (de margen máximo).

líneas punteadas son los bordes del margen M y los puntos en el borde son los vectores soportes.

Los coeficientes \mathbf{w} y w_0 se eligen para maximizar el margen M que mide la distancia del hiperplano al punto más cercano de cada grupo:

$$\max_{\mathbf{w}, w_0} M$$

sujeito a:

$$\|\mathbf{w}\| = 1$$

$$r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) \geq M \text{ para } i = 1, 2, \dots, n$$

Aquí, el sentido del vector \mathbf{w} indica el grupo rotulado por $r_i = 1$ y la expresión $r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0)$ mide la distancia del i -ésimo punto al plano separador; el factor r_i compensa con un menos cuando el punto $\mathbf{y}_{i\bullet}$ se encuentra del lado opuesto al que apunta \mathbf{w} . En ambos lados del hiperplano de separación óptima, los puntos están al menos $1/\|\mathbf{w}\|$ de distancia y el margen total es de $2/\|\mathbf{w}\|$. Luego maximizar el margen es minimizar $\|\mathbf{w}\|$. Pero esto último es matemáticamente equivalente a:

$$\min_{\mathbf{w}, w_0} \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{sujeito a: } r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) \geq 1 \quad (2.26)$$

Las restricciones aseguran que el clasificador de máximo margen asigne cada punto correctamente, lo cual es posible ya que se asume que los datos son linealmente separables. Este planteo corresponde a una cuestión de programación cuadrática, pues se tiene una función objetivo cuadrática con restricciones lineales. Para resolverlo, se puede reformular como un problema sin restricciones, utilizando los multiplicadores de Lagrange α_i :

$$\begin{aligned} L_P &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i [r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) - 1] \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_i \alpha_i r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) + \sum_i \alpha_i \end{aligned} \quad (2.27)$$

La solución se encuentra minimizando L_P respecto de \mathbf{w} , w_0 y maximizado para $\alpha_i \geq 0 \forall i$. Una vez que se resuelve para α_i , se ve que aunque hay n de ellos, la mayoría se anulan ($\alpha_i = 0$) y sólo un pequeño porcentaje tiene $\alpha_i > 0$. El conjunto de $\mathbf{y}_{i\bullet}$ cuyos $\alpha_i > 0$ son los *vectores de soporte*, y son justamente los puntos que quedan sobre el margen. Estos son los $\mathbf{y}_{i\bullet}$ que satisfacen: $r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) = 1$ y que yacen sobre el margen. Se puede usar este hecho para calcular w_0 desde cualquier vector de soporte como:

$$w_0 = r_i - \mathbf{w}'\mathbf{y}_{i\bullet}$$

Para la estabilidad numérica, se aconseja que esto se haga para todos los vectores de soporte y se tome una media. El discriminante así encontrado se llama *máquina de soporte vectorial*.

Clases no linealmente separables

En la práctica, existen muestras donde los grupos están superpuestos espacialmente, es decir, que no son separables por un hiperplano a menos que se introduzcan variables de *holgura* que permitan a algunos puntos permanecer en el grupo equivocado. Cada variable de *holgura*, una por punto, mide la distancia de un punto al margen (a distancia M del hiperplano separador) que corresponde al grupo al que pertenece el punto. Si el punto se encuentra del lado correcto, su variable de holgura es cero, pero si sobrepasa el margen a distancia M del hiperplano separador, su variable de holgura es positiva. Para permitir errores se reemplaza la desigualdad en (2.26) con

$$r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) \geq 1 - \xi_i \quad i = 1, 2, \dots, n$$

donde $\xi_i \geq 0$ son variables de holgura que permiten que un punto esté en el margen ($0 \leq \xi_i \leq 1$), o sea mal clasificado ($\xi_i > 1$). Dado que una observación está mal clasificada si el valor de su variable de holgura es mayor que 1, $\sum_i \xi_i$ es un límite para el número de puntos mal clasificados. Para penalizar las malas clasificaciones a la expresión (2.26) se le agrega un término y el problema de optimización se convierte en:

$$\begin{aligned} \min_{\mathbf{w}, w_0} \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i \right) & \quad (2.28) \\ \text{sujeto a : } r_i(\mathbf{w}'\mathbf{y}_{i\bullet} + w_0) \geq 1, \quad \xi_i \geq 0 & \end{aligned}$$

La constante $C > 0$ establece la importancia relativa de maximizar el margen y minimizar al mismo tiempo la totalidad de las *holguras*. Esta formulación es llamada *SVM de margen suave*. Para valores pequeños de C el margen es más ancho y a medida que C aumenta el margen disminuye y torna el procedimiento más rígido. El planteo se formula en términos de multiplicadores de Lagrange y con el fin de hacerlo más eficiente numéricamente, se computa el problema dual equivalente. El Lagrangiano Dual es función de las variables α_i :

$$L_{Dual} = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{l=1}^n \alpha_i \alpha_l r_i r_l \langle \phi(\mathbf{y}_{i\bullet})' \phi(\mathbf{y}'_{l\bullet}) \rangle \quad (2.29)$$

Esta función se maximiza sujeto a las restricciones:

$$0 \leq \alpha_i \leq C$$

$$\sum \alpha_i r_i = 0$$

La formulación dual produce una expansión del vector \mathbf{w} en términos de los vectores de entrada (ME): $\mathbf{w} = \sum_{i=1}^n \alpha_i r_i y_{i\bullet}$. Los *vectores de soporte* (para los cuales $\alpha_i > 0$) no son solamente los que están en el margen sino que también los que están dentro del margen.

La idea es que, cuando se tienen dos grupos donde el borde de separación no es lineal, se usa una transformación biyectiva no lineal Φ , del espacio \mathfrak{R}^n en otro \mathfrak{R}^m con $m > n$, que transforma los puntos $\mathbf{y}_{i\bullet}$ en $\phi(\mathbf{y}_{i\bullet})$, de manera que sean linealmente separables, tal como se muestra en la Figura 2.5.

Otra ventaja es que no se necesita explicitar la función Φ , pues el producto escalar en (2.29) se define como una función *núcleo* (*kernel*) $K(\mathbf{y}, \mathbf{y}') = \phi(\mathbf{y})' \phi(\mathbf{y}')$ que sólo depende de los vectores de entrada por lo que esta función *núcleo* puede ser calculada sin explicitar Φ , ver Ben-Hur y Weston (2010). Las funciones *núcleo* más utilizadas en SVM son lineal, polinomial, Gaussiano de base radial (RBF) y Laplace de base radial. En este trabajo se utiliza RBF.

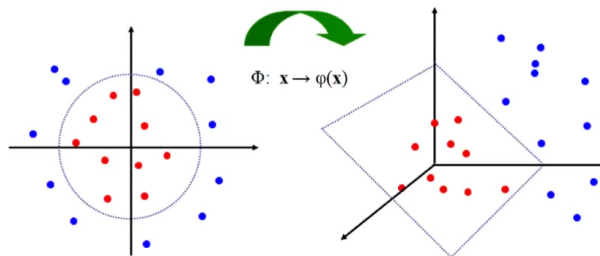


Figura 2.5: Ejemplo de dos clases no linealmente separables, en \mathfrak{R}^2 , que se separan linealmente en \mathfrak{R}^3 , por medio de Φ .

La expresión del *núcleo* Gaussiano es $K(\mathbf{y}, \mathbf{y}') = \exp(-\gamma \|\mathbf{y} - \mathbf{y}'\|^2)$, con $\gamma > 0$. A medida que se incrementa el parámetro γ se produce un aumento en la curvatura de las bandas de decisión. Comúnmente se utiliza $\gamma = 1/p$, donde p es la dimensión (cantidad de variables). Las etapas de entrenamiento y clasificación se realizan a través de esta función *núcleo*, sin necesidad de operar en el espacio no lineal.

Las SVM fueron diseñadas originalmente como clasificadores binarios. Para resolver problemas de clasificación con más de dos clases (multiclase), se han planteado estrategias para combinar las SVM binarias. Entre las más difundidas se puede mencionar *uno -contra- todos* y *uno- contra -uno*. La librería “*libsvm*” de R, utiliza la segunda, en la cual $k(k-1)/2$ clasificadores binarios son entrenados y la clase correcta es encontrada por un esquema de votación.

2.3.4. Combinación de clasificadores: Clasificador COMBINado

Además, de utilizar los clasificadores presentados, se propone otro clasificador combinando los resultados de estos clasificadores. Tal como se muestra en las secciones previas, se pueden aplicar diferentes algoritmos de clasificación (ADL, ADC, SVM) y aunque son generalmente exitosos, ninguno de ellos es el más preciso. Aquí se discute un modelo compuesto de múltiples algoritmos, que se complementan entre sí, para obtener mayor precisión. El Teorema “No Free Lunch” Wolpert y Macready (1997) afirma que no hay un solo algoritmo que sea el más exitoso para todas las clases. Este tema se desarrolla en (Alpaydin, 2010; Dietterich, 2000; Kuncheva, 2014) entre otros. Generalmente, para combinar clasificadores se utilizan clasificadores llamados *blandos* como $k - nn$ (vecino más cercano), también son presentadas técnicas como *Boosting* o *Bagging*. En este trabajo se utilizan ADL, ADC y SVM, llamados *duros* y la combinación de ellos no está implementada en el software *R*, y aquí se generan los códigos para esta tarea.

Hay dos cuestiones básicas:

- a) ¿Cómo se genera un conjunto de algoritmos de base que se complementen entre sí?
- b) ¿Cómo combinar los resultados de los algoritmos de base para obtener la máxima precisión?

Existen diferentes maneras en que los múltiples algoritmos de base se combinan para generar el resultado. En los métodos de combinación *multiexpertos* los algoritmos de base trabajan en paralelo. Para una entrada (input) cada uno genera una salida (output) y todas ellas son utilizadas para la solución final. El más simple y utilizado es el *método de votación*.

Sean $\mathbf{y} \in \mathbb{R}^p$ y $\Omega = \{\omega_1, \dots, \omega_k\}$ el conjunto de rótulos para las clases. Si se tiene una colección de L clasificadores (*ensemble*): $\mathcal{D} = \{D_1, D_2, \dots, D_L\}$, cada D_s , con $s = 1, \dots, L$, asigna una UE a la clase ω_{i^*} , que por (2.20) es la que tiene probabilidad a posteriori estimada más alta: $\hat{P}D_s$. El *método de votación* selecciona entre todas las clases obtenidas, la clase que tiene $\max \hat{P}D_s$. Para aclarar el método, la Tabla 2.1 muestra los resultados con 6 clases y 4 clasificadores para algunas UE (pixel en el contexto de este trabajo):

La última columna es la clase para COMB. Por ejemplo, el dato 23 es asignado a la clase 4, por ser la que obtuvo probabilidad más alta.

2.4. Muestras Originales vs Nuevas Muestras

La selección de las ME es la primera etapa de la clasificación supervisada, se la llama Fase de Entrenamiento y es fundamental para la validez de los resultados que se obtienen en las etapas posteriores. Esta fase depende fuertemente del operador y del conocimiento que éste

Tabla 2.1: Ejemplo de los resultados del Clasificador COMBINado

	clasif 1		clasif 2		clasif 3		clasif 4	
id	clase	$\hat{P}D_1$	clase	$\hat{P}D_2$	clase	$\hat{P}D_3$	max	clase nueva
20	2	0.99	2	1.00	2	0.80	1.0	2
21	2	0.95	2	0.85	1	0.48	0.95	2
22	2	0.93	4	0.97	2	0.70	0.97	4
23	2	0.86	4	0.98	1	0.44	0.98	4
24	2	0.84	2	0.99	1	0.56	0.99	2
25	1	0.89	2	0.99	1	0.85	0.99	2

tenga del área de estudio. Se recomienda en Richards y Jia (2006b) que el tamaño de las ME sea aproximadamente el 1 % de los píxeles de la imagen. Las muestras deben ser exclusivamente de las clases predefinidas. En el trabajo de Frery et al. (2009) se determinó que si las muestras de entrenamiento presentan errores se afecta seriamente al algoritmo de clasificación. La falta de conocimiento del área en la imagen a clasificar, zonas de difícil acceso o errores al digitalizar los píxeles, pueden producir muestras que presenten grupos mezclados.

Cuánto mayor sea el tamaño de las muestras mayor es el riesgo de tener errores, es decir elementos pertenecientes a otra clase. Por ello, se propone una metodología que permita ampliar las muestras, con mayor separabilidad entre ellas y con la mínima intervención del operador.

Si se supone que en toda la imagen se identificaron k regiones o clases C_1, C_2, \dots, C_k , los pasos del algoritmo se presentan a continuación:

1. Selección de las muestras de entrenamiento: Se identifican visualmente y se seleccionan de forma manual, marcando puntos digitalmente sobre la imagen.
2. Cálculo de los cuartiles: Para cada una de las k muestras, se calculan los cuartiles en cada una de las p variables.

Tabla 2.2: Ejemplo de cuartiles 1 y 3, por variable, para la i -ésima clase.

	B_1	B_2	\dots	B_p
1° Cuartil	q_{11}	q_{12}	\dots	q_{1p}
3° Cuartil	q_{31}	q_{32}	\dots	q_{3p}

3. Recorriendo toda la matriz de datos, \mathbf{Y} : Se selecciona cada pixel de la imagen, es un vector de tamaño p y se determina si su valor está entre los cuartiles para cada una de las clases. Esto se realiza para todos píxeles de la imagen.
4. Cada pixel en una única clase: Se determina si el pixel está entre los cuartiles de una única clase. Si cumple la condición se lo asigna a la clase correspondiente, caso contrario se lo

desestima.

5. Se generan las nuevas muestras: Todos los vectores de dimensión p que cumplan el paso 4, forman las *nuevas* muestras de entrenamiento (MA).

Lo importante es que se recorren todas las posiciones en la imagen, o sea, las filas en la matriz y además cada vector (correspondiente a un pixel) que fue asignado a una clase no será adjudicado a ninguna otra, esto ayuda a la separabilidad entre las muestras.

2.5. Fase de Validación

Es importante conocer el desempeño de un clasificador. El componente más destacado en el rendimiento del mismo es la precisión de las clasificaciones. Si se pudiera probar el clasificador en todos los posibles objetos de entrada, se sabría cuán exacto es. Lamentablemente, esto no es un escenario posible, por lo que se utiliza una estimación de la exactitud.

Validar cuantitativamente el producto obtenido, mapa temático, es la tercera fase del proceso de clasificación supervisada. Requiere de muestras de referencia, también llamadas *puntos test*, donde se conoce la verdad del terreno. Éstas se puede obtener *in situ* o a partir de fotografías aéreas, por ejemplo; en este trabajo se seleccionaron manualmente en la imagen.

Si estas muestras no son del tamaño suficiente o no contienen la verdadera cubierta del terreno, no se realizará una buena evaluación del producto obtenido. Los datos obtenidos para estas muestras no deben usarse en el algoritmo de clasificación.

El error en el mapa temático sería la discrepancia entre las clases o regiones representadas en el mapa y las que están presentes realmente en el terreno. La verificación de la exactitud de la clasificación, históricamente pasó por varias etapas (Congalton, 1991; Congalton y Green, 1999). Primero se realizaba una inspección visual de la calidad del mapa obtenido (altamente subjetiva), en una segunda etapa se calculaba aproximadamente el área que ocupaba cada clase (en km^2 o %), comparándola con información que se tenía de otras fuentes (podría coincidir en la extensión pero no en la localización), en la tercera etapa histórica se calcula el coeficiente de fiabilidad global (*overall accuracy*), que mide el porcentaje de píxeles correctamente clasificados sobre el total píxeles evaluados. Luego del cálculo de esta medida, en la cuarta etapa se perfecciona esto con la construcción de la matriz de confusión o matriz error de la cual se pueden obtener otras medidas.

En el artículo Foody (2002) se examinan, brevemente, los antecedentes y métodos de evaluación de la exactitud de la clasificación que se utilizan y recomiendan en la literatura. La comunidad de investigadores tiende a utilizar a menudo, incondicionalmente, las técnicas que se basan en la matriz de confusión. Pero la correcta aplicación e interpretación requiere que

se satisfagan ciertas condiciones, por ejemplo, que sean datos comparables, que se utilicen los mismos datos de referencia para todas las situaciones.

Las medidas para la evaluación de la exactitud del mapa temático se utilizan para comparar clasificadores, más que para estudiar la representación de la realidad que se hace con el mapa. En este trabajo se aplican varios clasificadores, en todos ellos se utiliza la misma ME y se evalúan utilizando la misma muestra de puntos test.

Las medidas utilizadas se derivan de la matriz de confusión, que se describe a continuación.

Matriz de Confusión

La matriz de confusión es uno de los métodos más comunes para expresar la exactitud de la clasificación; se trata de una tabla de contingencia, también llamada *matriz de error*. Esta matriz se forma con los datos de referencia, para los que se posee tanto la clase real, la verdad, como la deducida por la clasificación. Si se consideran k clases o regiones, la matriz de confusión se define como $MC = [X_{ij}]_{1 \leq i \leq j \leq k}$ una matriz cuadrada de dimensión $k \times k$, las columnas tienen la clase verdadera (verdad) de cada UE y las filas las categorías deducidas de la clasificación para cada unidad. Los elementos de la diagonal representan las UE correctamente asignadas, es decir aquellas UE que eran de una clase determinada y el clasificador la asignó a esa clase. La matriz de confusión ideal sería la que tiene ceros fuera de la diagonal. El total de cada columna es la cantidad de datos por clase de la muestra de referencia, mientras que, el total por fila son las UE asignadas a esa clase por el clasificador. El total general es el tamaño de la muestra de referencia. En la Tabla 2.3 se presenta la forma general de una Matriz de Confusión.

Tabla 2.3: Estructura general de una Matriz de Confusión

Verdad del terreno Clasificador	Clase 1	Clase 2	...	Clase k	Total	Usuario	Comisión
Clase 1	x_{11}	x_{12}	...	x_{1k}	$X_{1\bullet}$	$\frac{x_{11}}{X_{1\bullet}}$	$\frac{X_{1\bullet} - x_{11}}{X_{1\bullet}}$
Clase 2	x_{21}	x_{22}	...	x_{2k}	$X_{2\bullet}$	$\frac{x_{22}}{X_{2\bullet}}$	$\frac{X_{2\bullet} - x_{22}}{X_{2\bullet}}$
...
Clase k	x_{k1}	x_{k2}	...	x_{kk}	$X_{k\bullet}$	$\frac{x_{kk}}{X_{k\bullet}}$	$\frac{X_{k\bullet} - x_{kk}}{X_{k\bullet}}$
Total	$X_{\bullet 1}$	$X_{\bullet 2}$...	$X_{\bullet k}$	$X_{\bullet\bullet}$		
Productor	$\frac{x_{11}}{X_{\bullet 1}}$	$\frac{x_{22}}{X_{\bullet 2}}$...	$\frac{x_{kk}}{X_{\bullet k}}$			
Omisión	$\frac{X_{\bullet 1} - x_{11}}{X_{\bullet 1}}$	$\frac{X_{\bullet 2} - x_{22}}{X_{\bullet 2}}$...	$\frac{X_{\bullet k} - x_{kk}}{X_{\bullet k}}$			

A partir de la matriz de confusión se pueden derivar varias medidas de la calidad o fiabilidad del clasificador. Todos los coeficientes, que se presentan a continuación, están calculados a partir de la muestra de *puntos test*, por lo tanto son estimadores de los parámetros poblacionales:

Fiabilidad Global: Proporción de píxeles correctamente asignados respecto del total muestreados.

Utiliza sólo los elementos de la diagonal de la matriz:

$$\hat{F} = \frac{\sum_{i=1}^k X_{ii}}{n}. \quad (2.30)$$

Error: Proporción de valores fuera de la diagonal respecto del total de la muestra. $Error = 1 - \hat{F}$.

Errores de omisión y de comisión: Utilizando los totales por fila o por columna de la matriz se pueden calcular:

Errores de omisión: proporción de píxeles que, perteneciendo a una determinada clase, no fueron incluidos en ella.

Errores de comisión: proporción de píxeles que perteneciendo a otras clases el clasificador los asignó a la clase de interés.

Desde otro punto de vista, algunos autores Congalton (1991) se refieren a la exactitud del usuario y a la del productor, definiendo la Fiabilidad del usuario y del productor.

Fiabilidad del usuario: es la proporción de píxeles de una clase sabiendo que el clasificador le asignó esa clase.

Fiabilidad del productor o Fiabilidad por clase: es la proporción de píxeles asignados por el clasificador a una determinada clase dado que verdaderamente eran de esa clase.

$$\hat{FC}_i = X_{ii}/X_{\bullet i}$$

Coefficiente de Concordancia Kappa: Hasta ahora se ha estudiado lo que ocurre en la diagonal y en los marginales de la matriz de confusión, pero resulta de interés estudiar las relaciones múltiples entre categorías y con este objetivo se utilizan herramientas del Análisis Categórico Multivariado (Agresti, 1990; Chuvieco, 2002). Uno de los índices más usados es el coeficiente de concordancia Kappa(κ), que mide la discrepancia observada entre la “regla de clasificación” y la “verdad del terreno” y la que cabría esperar simplemente por azar. Es decir, intenta delimitar el grado de ajuste debido a la exactitud de la clasificación, prescindiendo de la que se da por efectos aleatorios. La estimación de κ se obtiene como

$$\hat{\kappa} = \frac{n \sum_{i=1}^k x_{ii} - \sum_{i=1}^k \sum_{j=1}^k x_{ij} x_{ji}}{n^2 - \sum_{i=1}^k \sum_{j=1}^k x_{ij} x_{ji}}. \quad (2.31)$$

Un valor de $\hat{\kappa}$ igual a 1 indica un acuerdo total entre el mapa y la realidad, mientras que un valor cercano a 0 sugiere que el acuerdo observado es puramente debido al azar. Si el valor es negativo supone una clasificación muy mala, peor que una obtenida por azar. Una de las

aplicaciones de este coeficiente es comparar clasificaciones realizadas por distintos métodos, con el objeto de estudiar si difieren significativamente, en cuanto a su ajuste con la realidad.

Se pueden construir intervalos de Confianza para κ utilizando su varianza asintótica, dada en Agresti (1990), y también para F , cuyo estimador está dado en la ecuación (2.30).

Existen numerosos métodos para realizar la validación, presentados, entre otros, en el libro de Kuncheva (2014). El procedimiento llamado *Hold-out* o método H, consiste en separar el conjunto de datos en dos, una parte utilizarla para entrenamiento (*training*) del algoritmo y otra reservada para evaluación del producto final (*testing*), similar a lo que se presenta aquí. En otro criterio, denominado *validación cruzada*, se determina un entero m y al conjunto total se lo separa en m subconjuntos de tamaño de tamaño n/m . Entonces se usa uno de los subconjuntos para probar el clasificador D , entrenado con la unión de los $m - 1$ subconjuntos restantes. Este procedimiento se repite m veces, seleccionando una parte diferente para validar cada vez. Para obtener el valor final de \hat{P}_D se promedian las m estimaciones. Cuando $m = n$, el método se llama *leave-one-out* o método U.

Capítulo 3

Resultados

En el presente capítulo, se muestran los resultados obtenidos utilizando los datos experimentales de tres sub-escenas de una imagen óptica multiespectral ASTER (puede ser descargada de <https://earthexplorer.usgs.gov/>), correspondientes a la zona de interés, descrita en el Capítulo 1. Se realiza la clasificación con ADL, ADC, SVM y un cuarto clasificador (COMB) que es la propuesta de este trabajo, comparando los resultados obtenidos con cada uno de ellos. Para la reducción de dimensión se utilizan las dos estrategias propuestas, obteniendo imágenes con dos dimensiones diferentes (3 y 4 bandas espectrales), comparándolas con la que tiene la dimensión completa (14 bandas). Además, se emplean dos tipos de muestras de entrenamiento, las muestras originales y las muestras ampliadas que fueron obtenidas con la metodología propuesta aquí. Así, se generan 24 situaciones diferentes para cada imagen, en la Tabla 3.1 se resumen las situaciones analizadas.

Tabla 3.1: Situaciones abordadas para cada una de las subimágenes

Subimagen	Clases	Bandas	Muestras	Clasificadores
1	6	14 3 4	MO MA	ADL ADC SVM COMB
2	4	14 3 4	MO MA	ADL ADC SVM COMB
3	5	14 3 4	MO MA	ADL ADC SVM COMB

La zona de estudio posee un relieve de sierra con caída abrupta al oeste y más suave al este. El interés está en identificar el tipo de roca presente en la zona, lo cual es una tarea difícil por la estructura espacial de las mismas, ya que se encuentran muy mezcladas. Determinar el tipo de roca permite deducir los minerales que se encuentran en la región. La dimensión espectral de la imagen ASTER es de 14 bandas, altamente correlacionadas y con valores de radiancia similares, por lo que presenta información redundante. Ésto motivó a generar imágenes con menor dimensión espectral para analizar si al reducir la dimensión se obtienen resultados similares en la clasificación. Los análisis se llevaron adelante con Códigos implementados en el Software **R**, R Core Team (2017) (se incluyen en el Anexo B). En primer lugar se muestran los resultados para la Subimagen 1 con las 14 bandas espectrales (completa), con tres y cuatro bandas. Luego se muestran los resultados para la Subimagen 2, en todas las situaciones. Finalmente, están los resultados para la Subimagen 3.

3.1. Análisis para la Subimagen 1

La Subimagen 1 es la porción superior de la Imagen completa. En la Figura 3.1 se muestra la subimagen cuya dimensión es $360 \times 739 \times 14$ (*fila* \times *columna* \times *banda*), o sea 266040 píxeles (datos) para cada una de las 14 bandas (las variables).

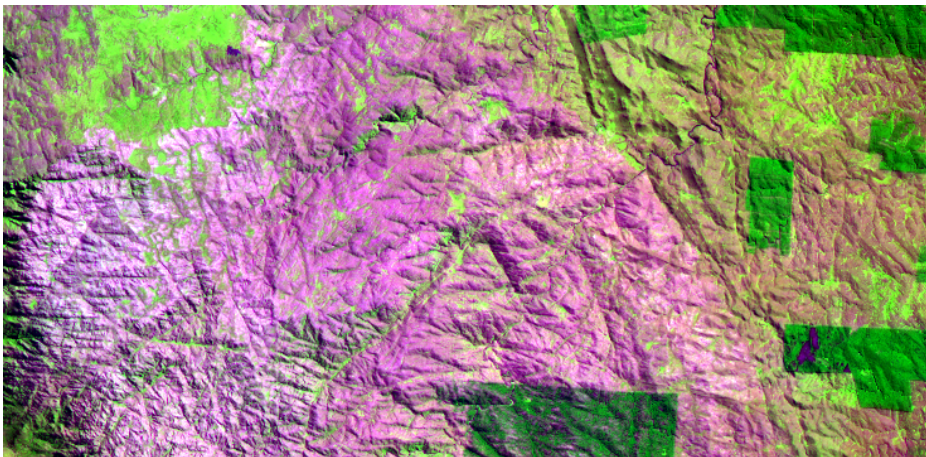


Figura 3.1: Composición RGB439 de la Subimagen 1

- Unificación del tamaño del píxel

La imagen ASTER tiene diferentes tamaños de píxel, según las bandas espectrales. Dado que las bandas tienen píxeles cuadrados de tamaños 15m, 30m y 90m de lado, se deben transformar todas al mismo tamaño de píxel. En la literatura se aconseja convertir todas a 30m ya que son 6 las bandas de ese tamaño. Esto se realizó con el software ENVI (ENVI,

2009), utilizando el método *re-muestreo del vecino más cercano*. Por ejemplo, 4 píxeles de $15 \times 15\text{m}$ se convierten en uno de $30 \times 30\text{m}$.

- Selección de las regiones de entrenamiento

Se determinaron 6 clases o regiones en la imagen. En la Figura 3.1 el pasto de altura representado en color verde claro corresponde a la clase 1, además se identificaron tres tipos de rocas: rocas metamórficas, rocas con alto contenido de sílice y rocas que forman una unidad externa, llamadas clase 2, 3 y 4 respectivamente. Se identificó una zona de falla en el sector izquierdo de la imagen en color negro, se la indica como clase 5 y los rectángulos, en verde oscuro, que representan los pinares de la zona, es la clase 6. Las muestras se identificaron visualmente y se recolectaron de forma manual, seleccionando digitalmente sobre la imagen puntos para las *muestras de entrenamiento* y para los *datos de referencia (puntos test)*. La información de las muestras seleccionadas se puede leer en la Tabla 3.2. El total de píxeles seleccionados para la muestra de entrenamiento (ME) fue 2127 y de 1689 para los datos de referencia.

Tabla 3.2: Descripción de las muestras para cada región de la Subimagen 1

Muestras de entrenamiento				Datos de referencia	
Nro.	Clases	Color	píxeles	Color	píxeles
1	pastoal	Verde	200	Magenta	175
2	roca metam	Azul	480	Marrón	239
3	altasilice	Amarillo	256	Púrpura	150
4	unidadexter	Cian	721	Coral	481
5	zona de falla	Verde Mar	213	Aguamarina	160
6	pinar	Rojo	257	Orquídea	484
Total			2127		1689

Además de las Muestras Originales (MO) también se utilizaron otras muestras llamadas “Muestras Ampliadas” (MA). La selección de las mismas se basa en el trabajo de Ferrero (2012) y la metodología se describe en el Capítulo 2, en la Sección 2.4. Con respecto a las MA, se analiza si el mayor tamaño y el aumento de separabilidad de las mismas mejora la clasificación. Brevemente, se puede decir que luego de realizado el pre-procesamiento, el algoritmo recorre toda la imagen, selecciona para cada clase aquellos píxeles cuyos valores se encuentran entre el 1° y 3° cuartil y además que no pertenezcan a ninguna otra clase. El tamaño obtenido para cada clase se muestra en la Tabla 3.3. Se puede observar que las muestras para algunas de las clases aumentaron considerablemente su tamaño pero otras son pequeñas, ello se debe a que los píxeles seleccionados corresponden únicamente a esa clase.

Tabla 3.3: Tamaño de Muestras Ampliadas para la Subimagen 1 con 14 bandas

1	2	3	4	5	6
183	2655	195	10319	95	303

Los diagramas de dispersión para las dos primeras funciones discriminantes (LD1 y LD2), obtenidos al aplicar ADL con las MO y las MA, se muestran en la Figura 3.2. Se ve claramente la mayor separabilidad obtenida con las muestras ampliadas.

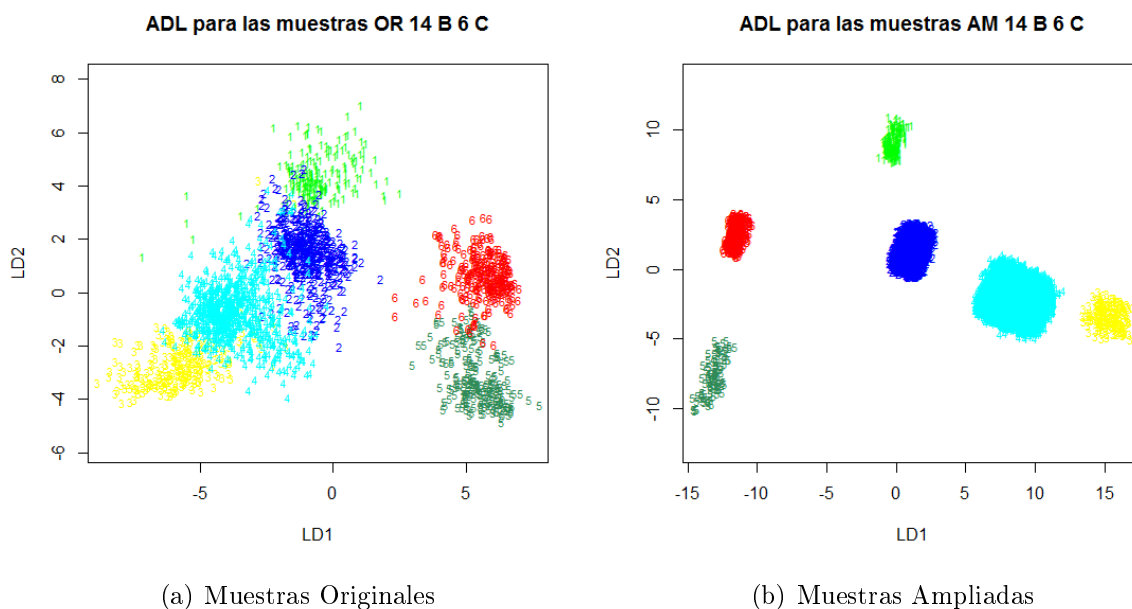


Figura 3.2: Diagrama de Dispersión de las muestras de entrenamiento, Subimagen 1 con 14 bandas

Para realizar la clasificación se utilizaron cuatro clasificadores ADL, ADC, SVM y COMB. El clasificador Combinado (COMB), como fuera explicado en el Capítulo 2 Sección 2.3.4, se genera de la siguiente manera: a cada píxel se le atribuye la clase que haya obtenido la probabilidad de asignación estimada más alta ($\hat{P}D_s$), con alguno de los otros clasificadores. En la Tabla A.1 del Anexo A se presentan los resultados para algunos píxeles.

3.1.1. Resultados de las Clasificaciones de la Subimagen 1 con catorce bandas espectrales

En la Figura 3.3 se muestran los mapas temáticos obtenidos con clasificador COMB para las MO y MA. El mapa temático (b) muestra mayor confusión entre las clases 2 (azul) y 4 (cian)

y menor área para la clase 3 (amarillo). Los mapas temáticos para los otros clasificadores y las muestras originales se presentan en la Figura A.1 del Anexo A.

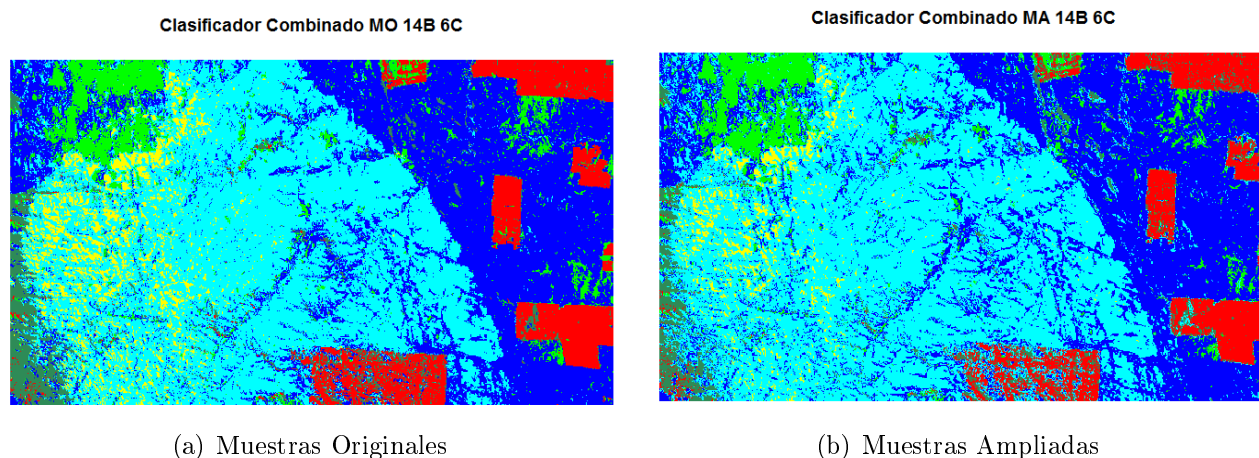


Figura 3.3: Mapas temáticos obtenidos con el clasificador COMB de la Subimagen 1, con 14 bandas

Para realizar la validación de los resultados de la clasificación, con los datos de referencia se construye la Matriz de Confusión para cada uno de los clasificadores, tal como se describe en el Capítulo 2, Sección 2.5. En la Tabla 3.4 se muestra la que corresponde al clasificador COMB con las muestras originales. Las columnas corresponden a la verdad del terreno representada por los datos de referencia. Los valores de la diagonal representan la cantidad de píxeles que fueron asignados correctamente por el clasificador, por ejemplo para la clase 3 la totalidad de los píxeles fueron clasificados en esa clase, le siguen, en cantidad de píxeles correctamente asignados, las clases 1, 5 y 6 mientras que las clases 2 y 4 son las que tienen más píxeles fuera de la diagonal, lo cual indica que el clasificador tuvo más dificultades para discriminar estas clases, tal como se indicó en los mapas temáticos de la Figura 3.3.

Tabla 3.4: Matriz de Confusión para el clasificador COMB con las MO

clasif	verdad						Sum
	1	2	3	4	5	6	
1	167.00	5.00	0.00	0.00	0.00	0.00	172.00
2	8.00	225.00	0.00	14.00	1.00	1.00	249.00
3	0.00	0.00	150.00	20.00	0.00	0.00	170.00
4	0.00	9.00	0.00	447.00	0.00	1.00	457.00
5	0.00	0.00	0.00	0.00	158.00	1.00	159.00
6	0.00	0.00	0.00	0.00	1.00	481.00	482.00
Sum	175.00	239.00	150.00	481.00	160.00	484.00	1689.00

Se puede resumir numéricamente, la información de la matriz de confusión, calculando los

coeficientes Error y Kappa. Como se expuso en el Capítulo 2, Sección 2.5, el Error es la proporción de píxeles fuera de la diagonal, es decir, los incorrectamente clasificados y Kappa es la proporción de clasificación correcta, no sólo teniendo en cuenta la diagonal de la tabla sino también los totales marginales. Los resultados se presentan en la Tabla 3.5. Se obtuvieron valores bajos para el Error y altos para Kappa, siendo el Error más bajo (Kappa más alto) para COMB con MO.

Tabla 3.5: Coeficientes estimados del Error y Kappa para la Subimagen 1 con 14 bandas

Clasificador	Error (%)		Kappa(%)	
	MO	MA	MO	MA
ADL	5.4	5.2	93.2	93.5
ADC	4.1	5.0	94.9	93.7
SVM	4.2	6.0	94.7	92.4
COMB	3.6	3.9	95.4	95.1

Los valores de Kappa en la Tabla 3.5 se presentan gráficamente en la Figura 3.4. Analizando para cada clasificador, se observa que Kappa tiene un valor levemente superior para las MA con ADL, los otros clasificadores tienen valores de Kappa más altos para las MO y para el clasificador COMB, Kappa tiene valores más altos, tanto para MO como para MA. El valor más bajo de Kappa corresponde al clasificador SVM con las MA. Debido a que no hay un clasificador que sea

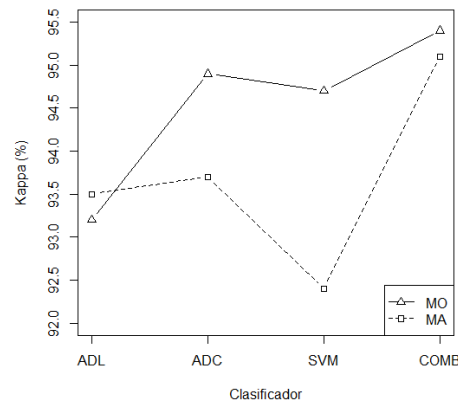


Figura 3.4: Diagrama de líneas múltiples de los Coeficientes Kappa estimados para la Subimagen 1 con 14 bandas.

el mejor para todas las clases, para cada uno de los clasificadores, se calcularon los coeficientes de Fiabilidad por clase ($\hat{F}C_i$, ésto es los porcentajes de píxeles correctamente clasificados por clase (Fiabilidad del productor, Capítulo 2, Sección 2.5).

Los resultados se presentan en la Tabla 3.6 para las MO y en la Tabla 3.7 para las MA.

Tabla 3.6: Coeficientes de Fiabilidad por clase (en %) para la Subimagen 1 con 14 bandas y las MO

Clasificador	Clases					
	1	2	3	4	5	6
ADL	95.4	92.9	96.7	88.2	99.4	99.4
ADC	96.0	94.6	100	91.3	96.9	99.6
SVM	95.4	92.5	98.7	91.7	99.4	99.6
COMB	95.4	94.1	100	92.9	98.8	99.4

Tabla 3.7: Coeficientes Fiabilidad por clase (en %) para la Subimagen 1 con 14 bandas y las MA

Clasificador	Clases					
	1	2	3	4	5	6
ADL	93.7	96.2	98.0	87.7	99.4	99.2
ADC	95.4	84.5	94.0	99.2	99.4	94.8
SVM	92.0	79.9	92.0	96.9	95.6	99.0
COMB	94.3	92.1	97.3	96.1	100.0	97.1

La Figura 3.5 muestra los resultados de los coeficientes de Fiabilidad y se observa que, con las MO, para la clase 4 se obtienen los valores más bajos de Fiabilidad con todos los clasificadores, por ser la clase más difícil de discriminar. Para las MA, los valores más bajos se obtienen en

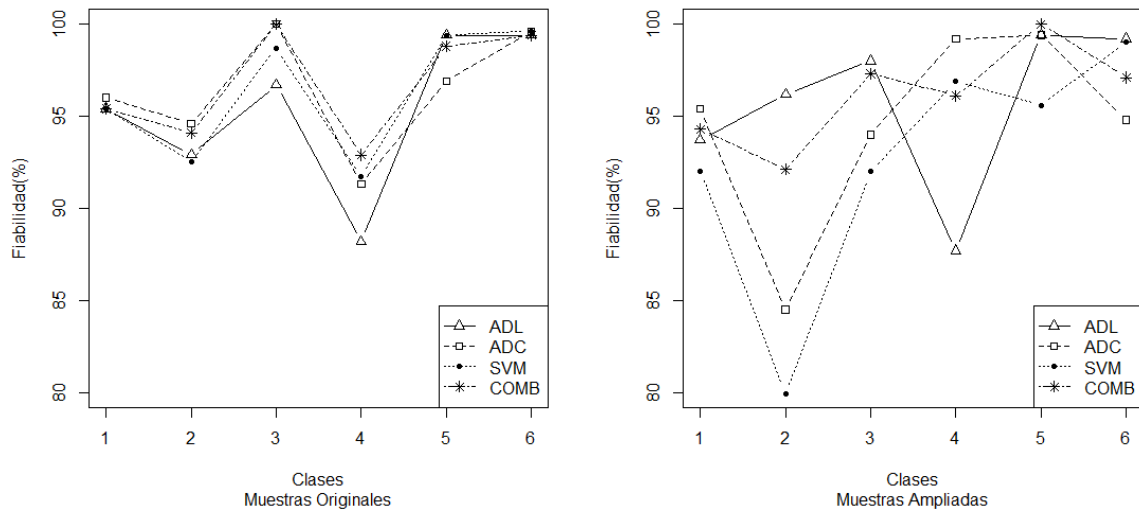


Figura 3.5: Coeficientes de Fiabilidad por clase para la Subimagen 1, con 14 bandas.

la clase 2, ADL tiene un valor similar al obtenido con las MO para la clase 4. Las clases 5 y 6 tienen los coeficientes más altos, con valores entre el 95 y 100 %.

Por otra parte, el clasificador COMB se mantiene entre los valores más altos para todas las clases, aún en los casos donde algunos clasificadores bajan.

3.1.2. Resultados de las Clasificaciones de la Subimagen 1 con tres bandas espectrales

En esta sección se presenta el análisis para la Subimagen 1, pero reduciendo la dimensión de la imagen completa de 14 bandas espectrales a menor cantidad de bandas. En la Figura A.2 del Anexo A se presentan los histogramas de cada banda y los diagramas de dispersión entre las bandas donde, en la mayoría de ellos, se observa asociación lineal alta. Ésto muestra la presencia de información redundante, por lo que se utilizará un método para reducir la dimensión.

La metodología utilizada se describe en el Capítulo 2 Sección 2.2.1. Brevemente, se puede decir que para seleccionar grupos de variables se aplica “Análisis por Factores”. De acuerdo a los resultados obtenidos, resumidos en el diagrama presentado en la Figura 2.1 del referido capítulo, se seleccionan las nuevas variables, tres en este caso. Este diagrama y todo el análisis fue realizado con *programas* desarrollados en R (los códigos se presentan en el anexo B).

La propuesta en este trabajo es: luego de decidir los grupos de variables presentes, se calculan las medias por grupo y éstas son los valores de las nuevas variables, para todos los píxeles de la imagen. Éstas nuevas variables se utilizan como entrada para cada algoritmo de clasificación.

En la Figura 3.6 se presenta la composición RGB de la nueva imagen generada, cuya dimensión es $360 \times 739 \times 3$. Lo que se observa es que con este método se mejora la visualización, destacando detalles en la imagen, es decir, la imagen es más nítida.

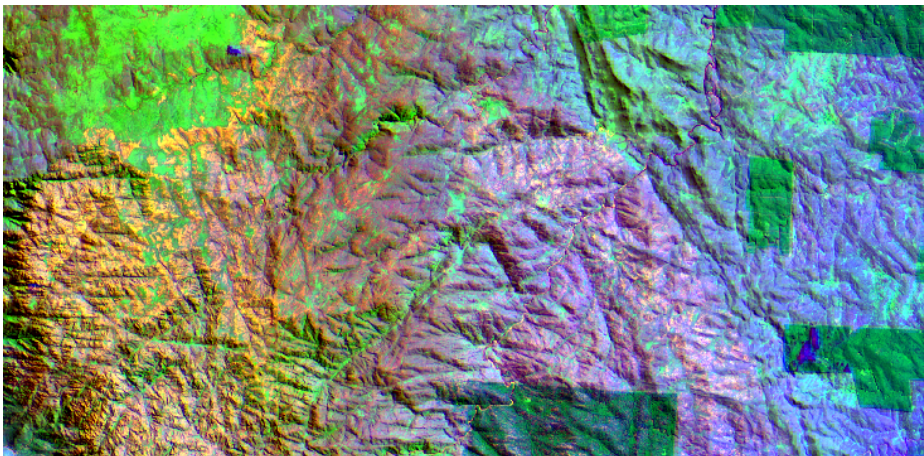


Figura 3.6: Composición RGB123 de la Subimagen 1 con 3 bandas

Una vez obtenida la nueva imagen se realiza la clasificación, con los cuatro clasificadores, para las MO y para las MA. En la Tabla A.2 del Anexo A se muestra el tamaño de las MA.

Para la validación de la clasificación, se genera la matriz de confusión, para cada clasificador y se calculan los coeficientes del Error y Kappa, cuyos valores se exhiben en la Tabla 3.8. Para el Error, los valores son mayores a los obtenidos con la imagen de las 14 bandas y los Kappa algo menores. Los valores más bajos para el Error se obtienen con COMB, tanto para MO como para MA y para la misma situación los valores más altos son para Kappa, alrededor del 91%.

Tabla 3.8: Coeficientes Error y Kappa para cada clasificador con MO y MA para la Subimagen 1 con 3 bandas

Clasificador	Error (%)		Kappa(%)	
	MO	MA	MO	MA
ADL	7.6	8.1	90.4	89.8
ADC	7.5	8.5	90.6	89.1
SVM	7.2	7.2	90.9	90.9
COMB	6.7	7.0	91.5	91.1

En la Figura 3.7 se observa que los valores de Kappa más altos se alcanzan con el clasificador COMB tanto para MO como para MA. El valor más bajo es para ADC con MA. Todos los clasificadores tienen mejores resultados para las MO.

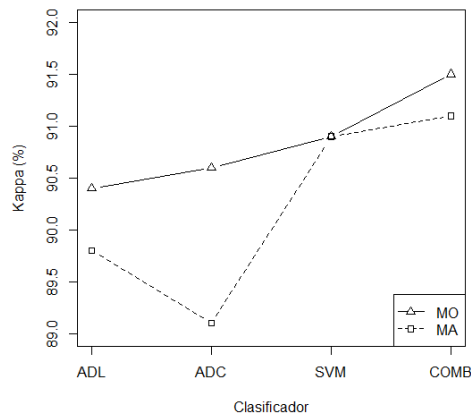


Figura 3.7: Kappa vs clasificador de la Subimagen 1 con 3 bandas

El comportamiento de los clasificadores en cada clase, se resume en las tablas 3.9 y 3.10 mostrando los coeficientes de Fiabilidad por clase (%) para MO y para MA, respectivamente.

La Figura 3.8 muestra gráficamente la información. Del gráfico se desprende que, con las MO la clase 3 fue asignada correctamente por todos los clasificadores (Fiabilidad 100%), las clases 2 y 4 son las que tienen valores más bajos. El clasificador COMB se mantiene entre los valores más altos, salvo para la clase 2. Los valores de Fiabilidad son altos tanto para la clase 5 como para la 6.

Tabla 3.9: Coeficientes Fiabilidad por clase (en %) para la Subimagen 1 con 3 bandas y las MO

Clasificador	Clases					
	1	2	3	4	5	6
ADL	89.1	84.1	100	88.2	97.5	97.9
ADC	94.3	87.0	100	82.5	100	99.8
SVM	85.1	75.7	100	93.4	99.4	99.2
COMB	92.0	81.2	100	89.8	98.8	99.4

Tabla 3.10: Coeficientes Fiabilidad por clase (en %) para la Subimagen 1 con 3 bandas y las MA

Clasificador	Clases					
	1	2	3	4	5	6
ADL	88.6	91.6	100	84.6	92.5	97.7
ADC	88.6	63.6	94.7	98.3	100	95.7
SVM	88.6	81.2	94.7	97.7	97.5	93.2
COMB	88.6	74.9	97.3	96.3	98.8	97.1

Con las MA, la clase 1 presenta los mismos valores de Fiabilidad para todos los clasificadores. ADL se mantiene en valores altos para las clases 1, 2 y 3, aunque baja para 4 y 5. La Fiabilidad es alta para todos los clasificadores en la clase 6. El clasificador COMB tiene Fiabilidad alta para todas las clases, salvo en la 2.

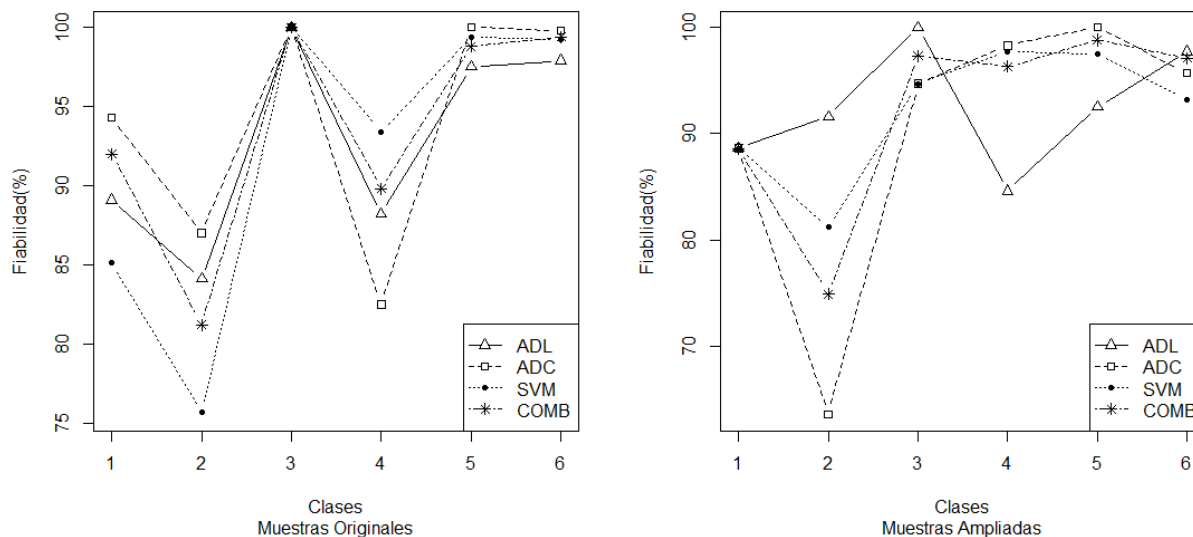


Figura 3.8: Coeficientes de Fiabilidad por clase, de la Subimagen 1 con 3 bandas

En la Figura 3.9 se presentan los mapas temáticos obtenidos con el clasificador COMB con MO y con las MA, para la Subimagen 1 con 3 bandas.

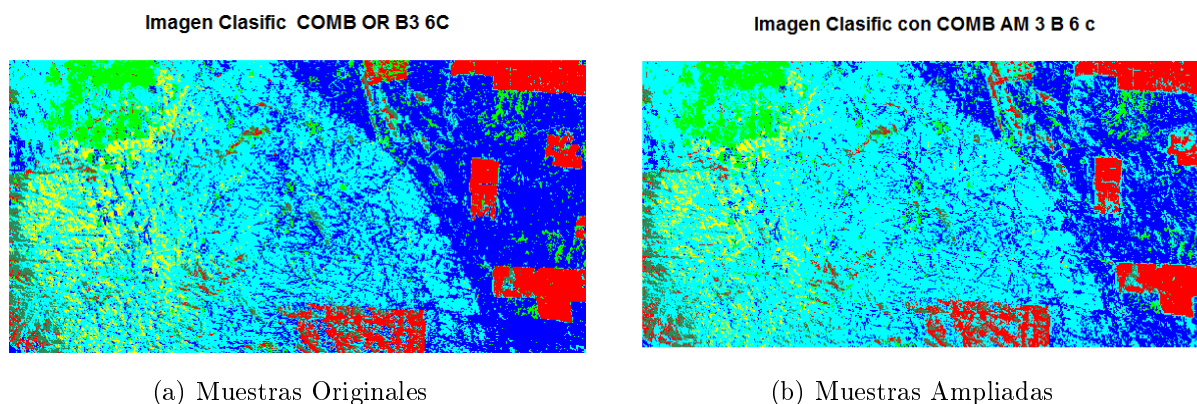


Figura 3.9: Mapas temáticos obtenidos con el clasificador COMB de la Subimagen 1 con 3 bandas

Los valores más altos en los coeficientes del Error y más bajos en los Kappa se refleja en los mapas temáticos de la Figura 3.9. El mapa temático (b) es más ruidoso mostrando confusión en las clases 2 y 4 (Azul y Cian). Aunque el mapa para la MO se ve algo mejor también muestra confusión entre esas clases. En la esquina superior derecha, en ambos mapas, aparecen píxeles asignados erróneamente a la clase roja, más en el (b) que en (a). Sin embargo, en ese mismo sector, el primero puede discriminar líneas correspondientes a cursos de agua, lo cual es un detalle importante para el mapa. Los mapas para los otros clasificadores se muestran en el Anexo A en la Figura A.3.

3.1.3. Resultados de las Clasificaciones de la Subimagen 1 con cuatro bandas espectrales

Otro método abordado para reducción de dimensión es Análisis Discriminante Lineal (ADL), presentado en el Capítulo 2, sección 2.2.2. Utilizando las 6 clases y las MO, se generan 5 *funciones discriminantes estandarizadas*, seleccionando las que más aportan a la variación total, según el criterio 2.17. La propuesta en este trabajo es seleccionar las variables (bandas) que tienen los coeficientes estandarizados más altos, en valor absoluto, en cada función. Las bandas que resultaron seleccionadas, con este método, son: 2,3,4 y 14 y se utilizan como *variables de entrada* para cada clasificador, en lugar de utilizar las funciones discriminantes.

El análisis fue realizado implementando los códigos en R (ver Anexo B). Por todo lo expuesto, en esta sección se trabaja con la Subimagen 1 pero de dimensión $360 \times 739 \times 4$. En el Anexo A, la Figura A.4 es una composición RGB, de la imagen con 4 bandas. Los tamaños de las MA generadas están en la Tabla A.3 del Anexo A.

Los algoritmos de clasificación fueron implementados tanto para las MO como para las MA. En la Figura 3.10 se presentan los mapas temáticos obtenidos con el clasificador COMB con

MO y MA para la Subimagen 1 de 4 bandas. Se puede observar que la calidad de los mapas es similar a los obtenidos con las 14 bandas, con la ventaja de una menor dimensión. Con las MA el mapa es más ruidoso, presentando confusión entre las clases Azul y Cian.

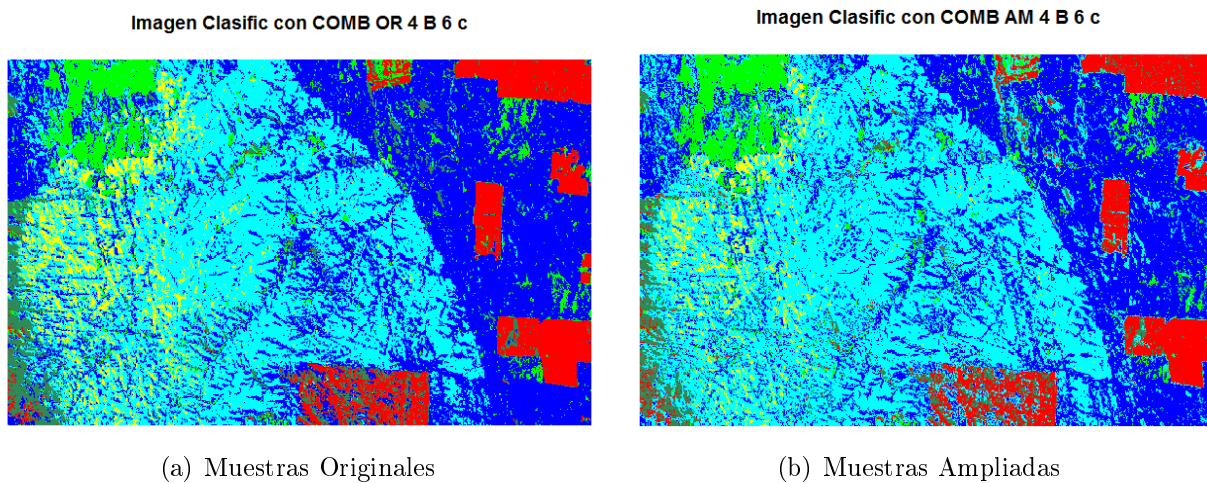


Figura 3.10: Mapas temáticos obtenidos con el clasificador COMB de Subimagen 1 con 4 bandas

Para la validación de la clasificación, la Tabla 3.11 muestra los coeficientes estimados del Error y Kappa para cada clasificador. Ambos tienen valores similares a los obtenidos con la imagen de las 14 bandas.

Los valores más bajos para el Error se obtuvieron con el Clasificador COMB tanto para MO como para MA y en la misma situación se obtuvieron los valores más altos para Kappa, alrededor del 94 %.

Tabla 3.11: Error y Kappa por clasificador con MO y MA para la Subimagen 1 con 4 bandas

Clasificador	Error (%)		Kappa(%)	
	MO	MA	MO	MA
ADL	6.0	8.3	92.5	89.6
ADC	5.2	5.8	93.4	92.6
SVM	4.4	6.5	94.5	91.8
COMB	4.4	4.4	94.4	94.4

La Figura 3.11 muestra que los valores de Kappa son mayores para las MO, salvo con el clasificador COMB donde coinciden con MA. El valor menor es en ADL con MA.

Los coeficientes de Fiabilidad por clase (*en %*), se presentan en las Tablas 3.12 y 3.13 para la Subimagen 1 con cuatro bandas.

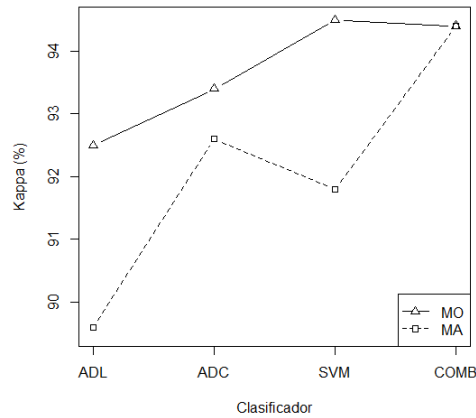


Figura 3.11: Kappa vs clasificador de la Subimagen 1 con 4 bandas

Tabla 3.12: Coeficientes de Fiabilidad por clase (en %) para la Subimagen 1 con 4 bandas y las MO

Clasificador	Clases					
	1	2	3	4	5	6
ADL	94.3	91.6	100	86.7	98.1	99.2
ADC	96.6	91.2	100	87.9	98.8	99.8
SVM	93.7	90.0	100	92.9	99.4	99.2
COMB	96.0	91.2	100	91.5	98.1	99.4

Tabla 3.13: Coeficientes de Fiabilidad por clase (en %) para la Subimagen 1 con 4 bandas y las MA

Clasificador	Clases					
	1	2	3	4	5	6
ADL	93.7	97.1	100	77.6	95.0	98.8
ADC	94.3	82.4	92.7	98.6	99.4	94.4
SVM	90.9	84.5	92.7	97.3	95.6	94.8
COMB	94.3	90.4	96.0	96.1	98.1	97.1

En los gráficos de líneas múltiples de la Figura 3.12 está la información presentada en las Tablas 3.12 y 3.13. Del gráfico se desprende que, con las MO la clase 3 fue asignada correctamente por todos los clasificadores (Fiabilidad 100%), las clases 2 y 4 son las que tienen valores más bajos. Todos los clasificadores presentan comportamiento similar en cada clase, salvo en la 4 donde ADL y ADC tienen valores más bajos. El coeficiente de Fiabilidad para el clasificador COMB se mantiene entre los valores más altos, para todas las clases.

Con las MA, los clasificadores tienen diferente comportamiento, por ejemplo ADL en las clases 2, 3 y 6 tiene Fiabilidad más alta, en cambio en la 4 es el valor más bajo. El clasificador COMB con MA tiene Fiabilidad alta para todas las clases, mayor al 90%.

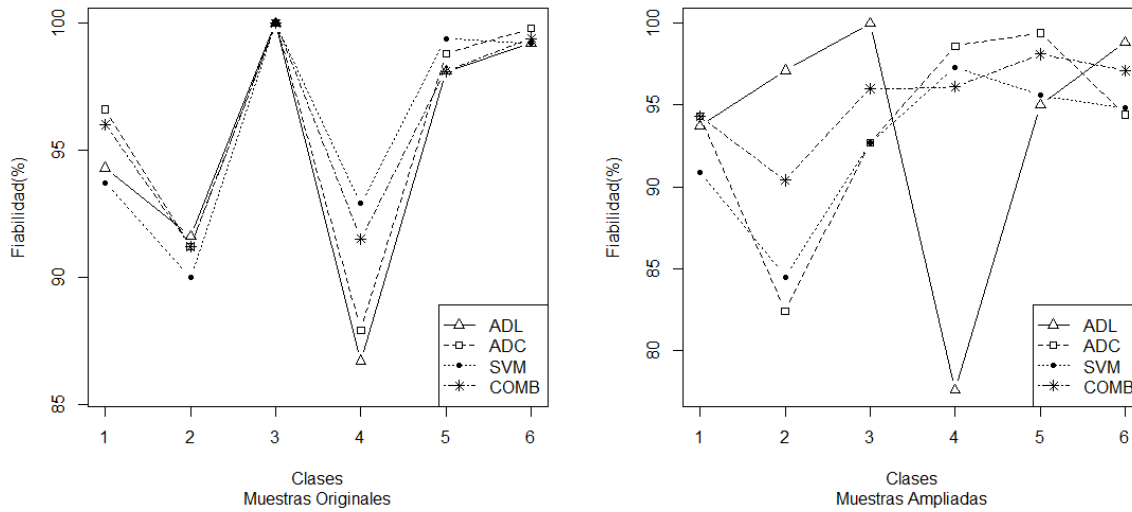


Figura 3.12: Coeficientes de Fiabilidad por cada clase, de la Subimagen 1 con 4 bandas

3.2. Análisis para la Subimagen 2

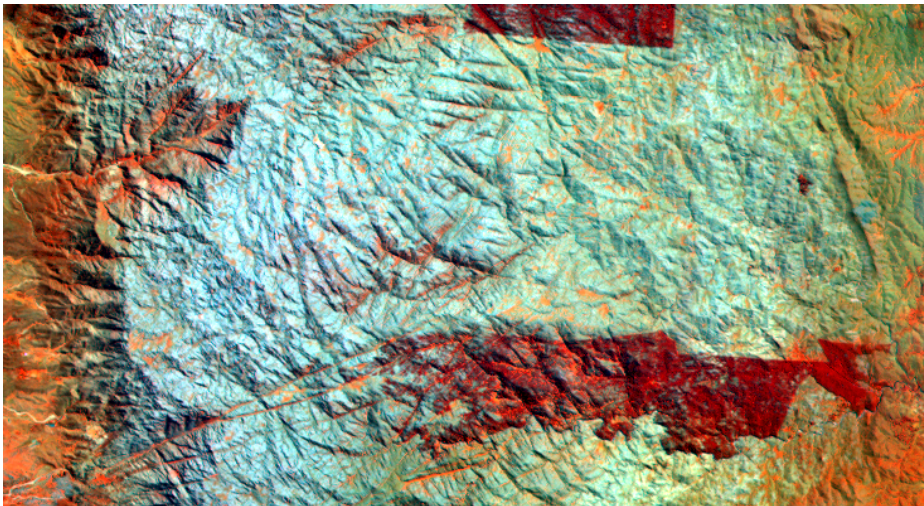


Figura 3.13: Composición RGB249 de la Subimagen 2

A continuación se presenta el análisis realizado para la Subimagen 2 (Figura 3.13) que corresponde a una sección de la imagen completa y es continuación de la Subimagen 1 en el extremo sur, su dimensión es $400 \times 739 \times 14$, o sea 295600 píxeles, para cada una de las 14 bandas espectrales.

Se determinaron 4 clases o regiones en la imagen, “Pinar”(verde), “Roca metamórfica” (azul), “Unidad externa” (amarillo) y “Zona de falla” (negro). Para cada una, se seleccionaron 1565 pí-

xeles para las “muestras de entrenamiento” y 1075 para los “datos de referencia”. La información de las muestras seleccionadas se puede leer en la Tabla A.4, en el Anexo A.

3.2.1. Resultados de las Clasificaciones de la Subimagen 2

Los resultados de las clasificaciones obtenidas para la Subimagen 2 se muestran en la Figura 3.14. El mapa temático obtenido con el clasificador COMB y 14 bandas es el que discrimina mejor las clases, se puede ver más ruidoso el que corresponde a las 3 bandas y el de las 4 bandas es similar al (a) con la ventaja de tener menor dimensión, pero no discrimina correctamente las clases 1 y 4.

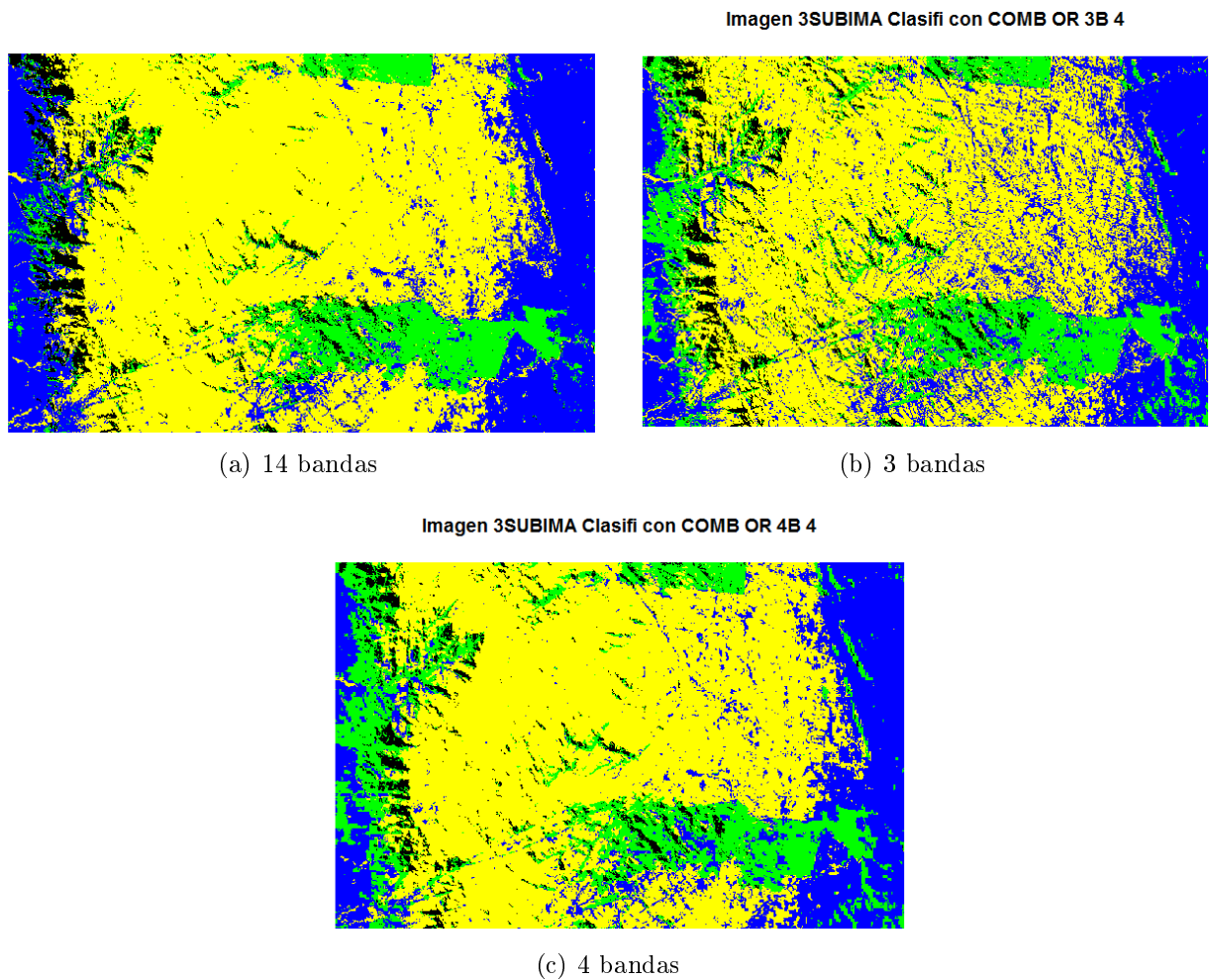


Figura 3.14: Mapas temáticos obtenidos con el clasificador COMB para Subimagen 2

3.2.2. Validación de la Clasificación para Subimagen 2

En la Tabla 3.14 se muestran los resultados para la Subimagen 2 en todas las situaciones analizadas. Los valores más bajos del Error (Kappa más alto) se obtienen para la imagen de 14 bandas, la imagen con 3 bandas tiene coeficientes de Error más altos (Kappa más bajos) y en las 4 bandas los errores son algo más bajos.

Tabla 3.14: Medidas de Fiabilidad para todas las situaciones de la Subimagen 2

Medidas		Error(%)						Kappa(%)					
Dimensión		14 Bandas		3 Bandas		4 Bandas		14 Bandas		3 Bandas		4 Bandas	
Clasificador	Muestras	MO	MA	MO	MA	MO	MA	MO	MA	MO	MA	MO	MA
		ADL	5.0	9.0	13.6	12.4	6.0	12.4	93.4	87.9	81.9	83.5	91.86
ADC	7.4	16.6	10.8	11.8	7.4	12.7	90.0	77.8	85.55	84.1	90.12	83.0	
SVM	9.9	16.5	11.9	19.1	9.3	23.5	86.8	77.8	84.1	74.2	87.53	68.2	
COMB	4.0	10.9	11.2	11.3	5.2	10.1	94.61	85.4	85.1	84.9	92.99	86.4	

En la Figura 3.15 se muestran los gráficos para Kappa en cada dimensión y con MO y MA. En todas las dimensiones, el coeficiente Kappa es más alto para COMB. Con 14 bandas, los coeficientes de Kappa, con MO, son más altos para todos los clasificadores. En 3 bandas, ADL toma el valor más alto de Kappa con las MA, en SVM el valor de Kappa es el más bajo y el clasificador COMB tiene Kappa con valor similar para MO y MA. En 4 bandas, los coeficientes Kappa tienen comportamiento similar en ADL;ADC y COMB con valores altos, salvo para SVM donde es bajo para las MA.

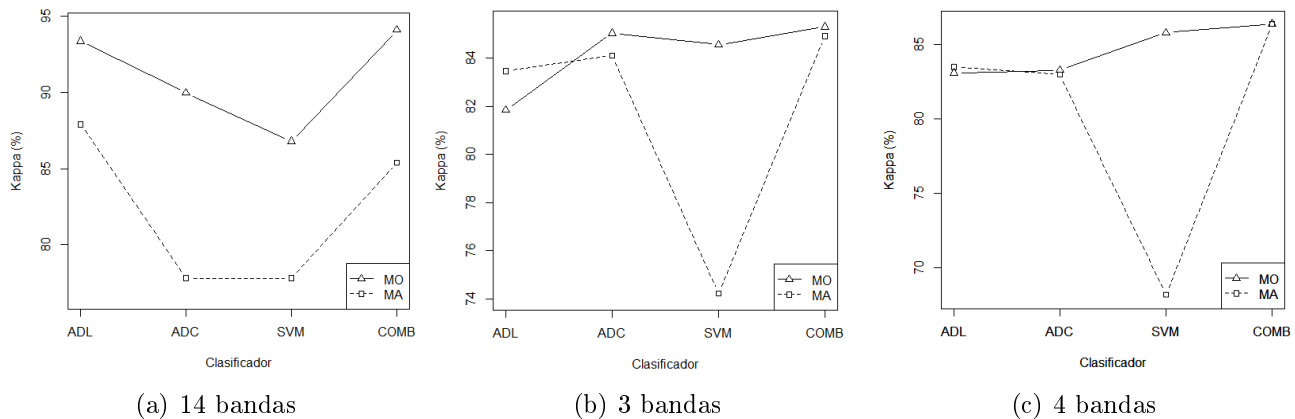


Figura 3.15: Coeficientes Kappa para tres dimensiones de la Subimagen 2

En la Figura 3.16 se muestra la Fiabilidad por clase, para la Subimagen 2 para 14 bandas. Tanto para MO como para MA, el coeficiente Kappa para COMB tiene los valores más altos, en todas las clases, con excepción de la clase 4 donde es más alto con ADL.

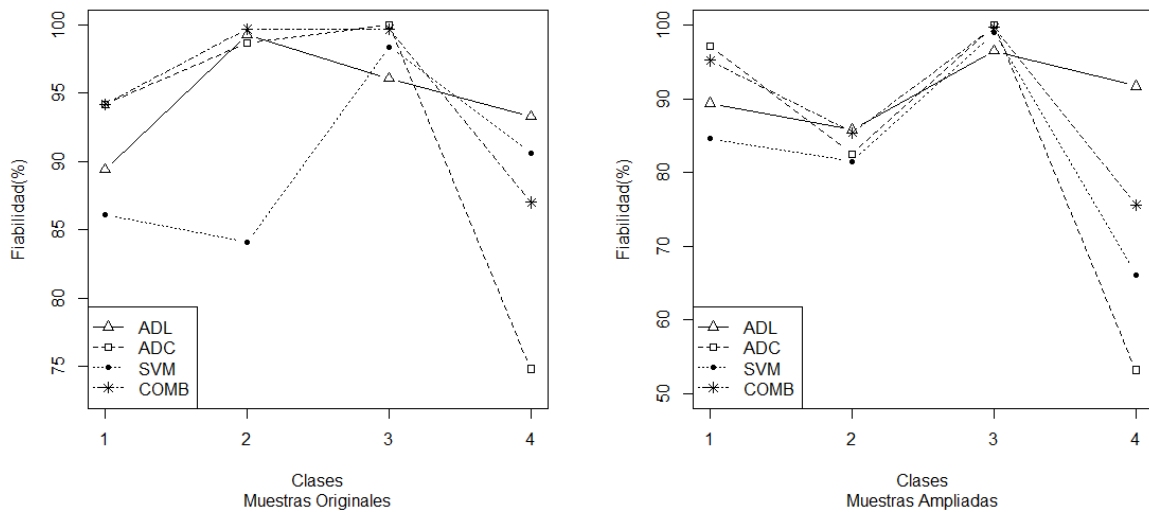


Figura 3.16: Coeficientes de Fiabilidad por clase, de la Subimagen 2 con 14 bandas

En la Figura 3.17 se muestra la Fiabilidad por clase para la imagen con 4 bandas, donde presenta un comportamiento algo diferente. Para MO, en la clase 1 es más alto Kappa con ADC, los coeficientes para la clase 2 son más bajos y en la clase 4 es más alto con ADL. Para las MA, se observa que Kappa para SVM es más bajo en todas las clases. En las Tablas A.5, A.6, A.9 y A.10 del Anexo A se muestran los valores de Fiabilidad por clase, correspondientes a los últimos gráficos analizados.

En la Figura A.9 del Anexo A se muestran los mapas temáticos para ADL y ADC con 14 y 4 bandas y con las MO.

Con respecto a la Subimagen 2 con tres bandas, se muestra en el Anexo A, una imagen RGB de la zona en la Figura A.7. Los resultados obtenidos para dicha imagen están en las Tablas A.7, A.8 y en la Figura A.8, del Anexo A.

3.3. Análisis para la Subimagen 3

La Figura 3.18 muestra la zona que abarca la Subimagen 3. Corresponde a la sección sur de la imagen completa, su dimensión es $300 \times 739 \times 14$, o sea 221700 píxeles, para cada una de las 14 bandas espectrales.

Se determinaron 5 clases, o regiones, en la imagen, seleccionando 1830 píxeles, en total, para las “muestras de entrenamiento” y 921 puntos para los “datos de referencia”. Los detalles de las muestras seleccionadas para cada clase, se presentan en la Tabla A.11 del Anexo A, mientras que en la Figura A.10 del citado anexo se describe la distribución espacial de las muestras

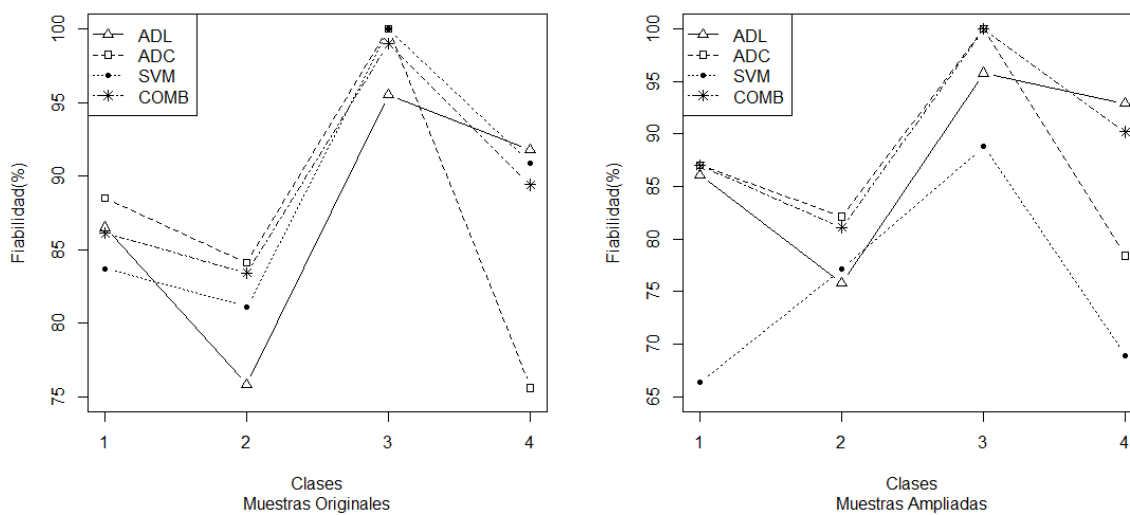


Figura 3.17: Coeficientes de Fiabilidad por clase, de la Subimagen 2 con 4 bandas

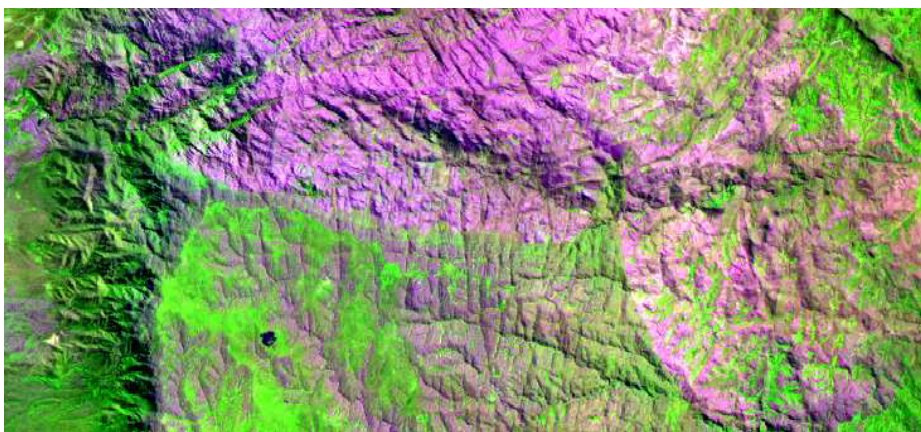


Figura 3.18: Composición RGB439 de la Subimagen 3

seleccionadas.

3.3.1. Resultados de las Clasificaciones de la Subimagen 3

El tamaño de las muestras ampliadas se puede ver en la Tabla A.12 del Anexo A. En la Tabla A.13, del Anexo A se presentan, para algunos píxeles, las clases obtenidas para el clasificador COMB, para Subimagen 3 con 14 bandas espectrales. La Figura 3.19, presenta los mapas temáticos obtenidos para las Subimagen 3, en todas las dimensiones y con el clasificador COMB. Comparando con el mapa de la imagen completa (14 bandas), se observa que el de 4 bandas es similar y que mapa el de 3 bandas presenta regiones mal clasificadas.

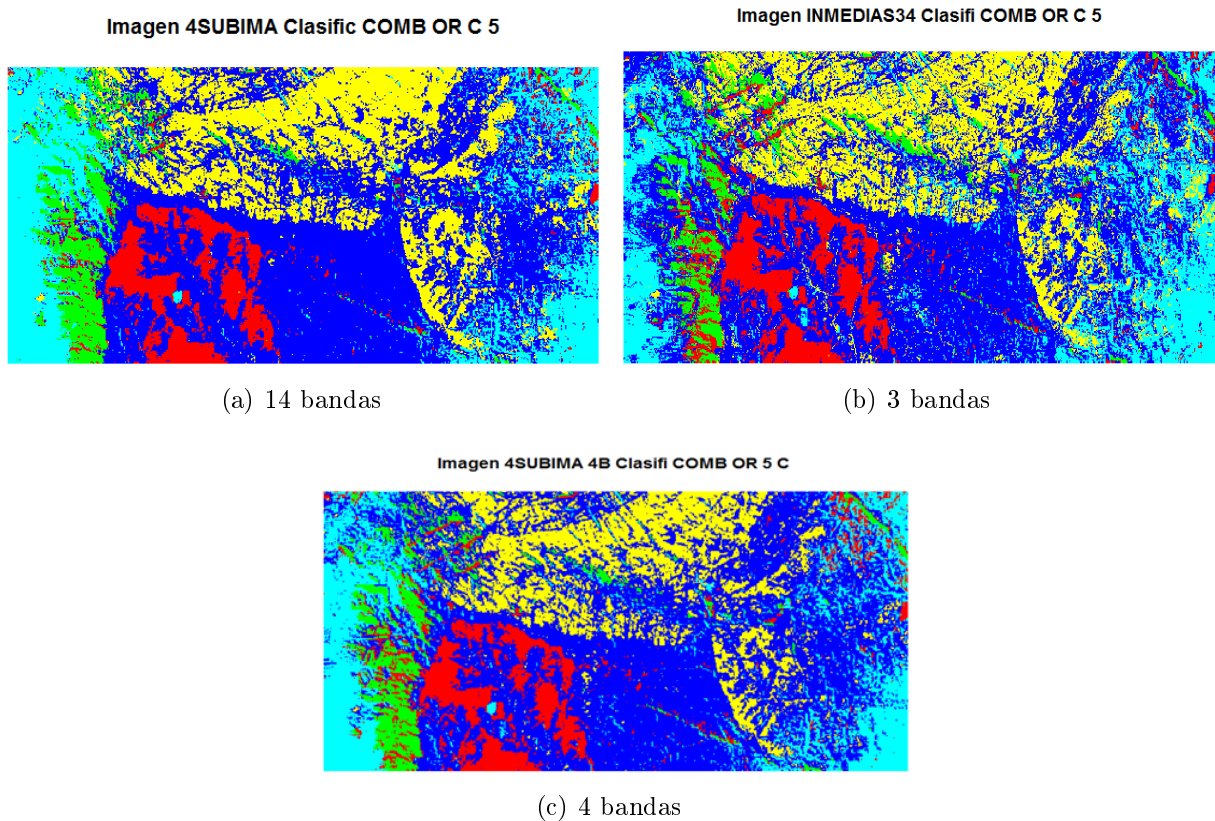


Figura 3.19: Mapas temáticos obtenidos con el clasificador COMB de la Subimagen 3

3.3.2. Validación de la Clasificación de la Subimagen 3

La Matriz de Confusión obtenida para el clasificador COMB con la Subimagen 3, se encuentra en la Tabla A.14 del Anexo A.

Los resultados obtenidos de las medidas Error (%) y Kappa (%) para las tres dimensiones analizadas, con las MO y las MA para los cuatro clasificadores, se pueden leer en la Tabla A.15 del Anexo A.

En la Figura 3.20 se muestran los gráficos para Kappa con cada clasificador. Para todas las dimensiones, los valores de Kappa son mayores en las MO, particularmente, Kappa para COMB es el más alto. Con 3 bandas los valores de Kappa son más bajos para todos los clasificadores, tanto en las MO como en las MA, con respecto a los de 14 bandas. Para las MO, están alrededor de 88% y para MA, Kappa se encuentra entre 79 y 86%. Los valores de Kappa para la imagen de 4 bandas son más altos, alrededor del 97% y para las MA entre 86 y 90%.

En la Figura 3.21 se muestra la Fiabilidad por clase, para la Subimagen 3 con 14 bandas. Para las MO, el coeficiente Kappa para el clasificador COMB tiene los valores más altos en todas las clases, excepto para la clase 1 donde es más alto con ADL. Lo que se destaca, para

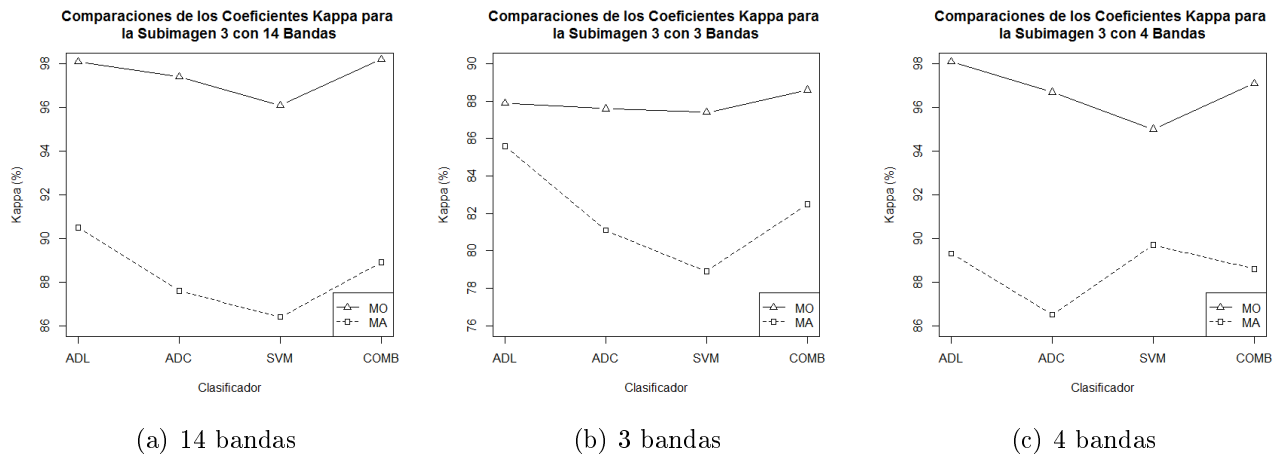


Figura 3.20: Coeficientes Kappa vs clasificador para las tres dimensiones de la Subimagen 3

MA es que en la clase 5 los Kappa, para todos los clasificadores, son más bajos, nuevamente ADL es el que toma mayor valor.

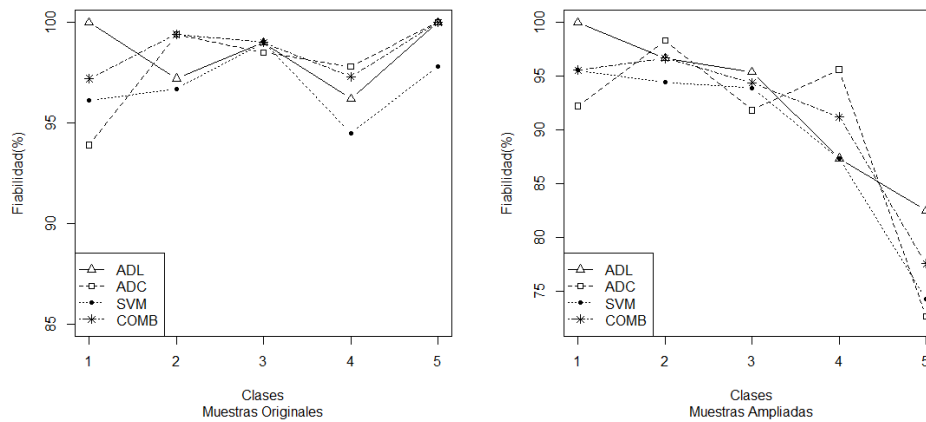


Figura 3.21: Coeficientes de Fiabilidad por clase de la Subimagen 3 con 14 bandas.

En la Figura 3.22 se muestra la Fiabilidad por clase para la imagen de 4 bandas, tiene un comportamiento similar al observado con 14, para las MO y las MA. Para las MO, Kappa es más alto, en la clase 1 con ADL, los coeficientes para la clase 4 son más bajos. Para las MA, se observa que Kappa con SVM es el más alto en las clases 4 y 5, superando al clasificador COMB. En las Tablas desde la A.16 hasta la A.21 del Anexo A se muestran los valores de fiabilidad para 14, 3 y 4 bandas, tanto para las MO como para las MA. En Ferrero y Bustos (2017) se presentaron los resultados de la clasificación de la Subimagen 3 con las dos técnicas propuestas de reducción de dimensión.

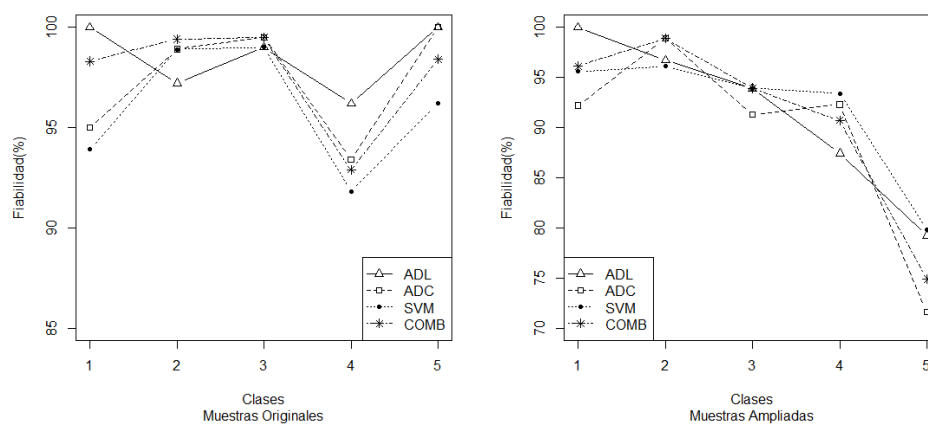


Figura 3.22: Coeficientes de Fiabilidad por clase de la Subimagen 3 con 4 bandas

Capítulo 4

Conclusiones y trabajos futuros

Las imágenes satelitales son una herramienta importante para la exploración de las rocas presentes sobre la corteza terrestre. Del análisis realizado a tres imágenes ópticas multiespectrales ASTER, en el presente Capítulo se presentan las conclusiones obtenidas a partir de las 24 situaciones generadas para cada imagen (cuatro clasificadores, dos tipos de muestras de entrenamiento y tres dimensiones espectrales). Las subimágenes consideradas son de zonas vecinas, con distinto número de clases.

4.1. Conclusiones

De los resultados obtenidos se puede concluir lo siguiente:

4.1.1. Muestras Originales vs Muestras Ampliadas

En este trabajo se propone realizar el análisis para las muestras seleccionadas manualmente, llamadas muestras originales. Se evalúa su desempeño tanto visualmente, mediante los mapas temáticos producidos, como numéricamente por medio del coeficiente de Concordancia Kappa. Además, se sugiere otra metodología que permite obtener muestras de manera objetiva, permitiendo a un usuario sin profundo conocimiento del área de estudio, obtener muestras de mayor tamaño y separabilidad entre clases. La metodología fue descrita en la Sección 2.4 del Capítulo 2. Éstas son llamadas muestras ampliadas. Se comparan los resultados obtenidos con los dos tipos de muestras. En las tres subimágenes los mayores valores de Kappa corresponden a las MO y 14 bandas. En casi todos los casos, las MA tienen Kappas más bajos y mapas temáticos más ruidosos, con algunas excepciones. De las 24 situaciones analizadas, con las MA Kappa es más alto: a) en la subimagen 1 con ADL y 14 bandas (Figura 3.4), b) en la subimagen 2 con 3 bandas y con 4 bandas, donde también coinciden en ADC (Figura 3.15). Una de las razones

puede ser que el tamaño de las muestras de entrenamiento utilizadas, eran de tamaño adecuado para la clasificación.

4.1.2. Comparando los Clasificadores

Otra propuesta que se presenta aquí es realizar el análisis con cada uno de cuatro clasificadores: ADL, ADC, SVM y COMB. Éstos se describen en el Capítulo 2, Sección 2.3. El clasificador COMB arroja mejores resultados ya que tiene valores de Kappa más altos en todas las subimágenes y para todas las situaciones. Aunque los valores de Kappa para MA son más bajos en la mayoría de las situaciones, para algunos casos, alcanzan los mismos valores que las MO con el clasificador COMB, subimagen 1 con 4 bandas, subimagen 2 con 3 y 4 bandas. Para SVM los valores son más bajos en casi todas las situaciones y está más afectado por las MA.

4.1.3. Clasificador por clase espectral

También, se comprobó que un único clasificador no es el mejor para todas las clases señaladas en la imagen. Es por ello que se analizaron todos los clasificadores en cada una de las clases espectrales. En las tres subimágenes, los coeficientes de Fiabilidad por clase (ver 2, Sección 2.5) para el clasificador COMB estuvieron entre los valores más altos. En la subimagen 1 con 14 bandas y MO, los valores de fiabilidad son más bajos en la clase 4 para todos los clasificadores considerados, seguida por la clase 2, las demás clases toman valores altos (Figura 3.5). El patrón que se presenta entre los clasificadores es un patrón regular, es decir, los coeficientes de Fiabilidad son altos en una clase y más bajos en otra para todos los clasificadores. Por el contrario, con MA tienen un comportamiento irregular, es decir, para una misma clase un coeficiente es alto con clasificador y bajo con otro. En la subimagen 2 los valores más bajos ocurren en la clase 4. El clasificador SVM tiene valores más bajos en todas las clases. ADC en algunas clases tiene valores altos y bajos en otras (Figuras 3.16 y 3.17). En la subimagen 3 los clasificadores tienen coeficientes de Fiabilidad alta para todas las clases, y con las MA los valores son más bajos en la clase 5, Figuras 3.21 y 3.22.

4.1.4. Las dimensiones espectrales

La dimensión de la imagen ASTER utilizada tiene 14 bandas espectrales. En este trabajo, también se proponen dos maneras de reducir la dimensión espectral, obteniendo una de 3 bandas y otra de 4. Los resultados demuestran que con la imagen de 3 bandas se logra una notable mejoría en la visualización, resaltando más detalles que con la imagen de mayor dimensión, pero no se logra una buena clasificación. Los valores de Kappa son menores con respecto a las otras

dimensiones analizadas y los mapas temáticos más ruidosos. Con la imagen de 4 bandas, que son un subconjunto de las variables originales, se alcanzan los valores de Kappa obtenidos con la de 14 bandas, lo cual es un resultado importante ya que se logra la misma información, con menor dimensión.

4.2. Contribuciones originales de la tesis

- Construcción de un mapa temático para la zona en estudio, no hay antecedentes al respecto.
- Metodología para la reducción de dimensión: Se plantea un método que da buenos resultados y además tiene la ventaja que es simple y utiliza un subconjunto de las variables originales.
- Se utiliza una técnica adecuada para encontrar grupos de variables no utilizada en trabajos referidos al tema.
- El clasificador propuesto (COMB) combina LDA, QDA y SVM diferentes a lo propuesto en la literatura.
- Se propone una metodología para obtener muestras ampliadas, reduciendo la intervención del investigador.
- Programas en R para desarrollar la propuesta.

4.3. Trabajos futuros

Cuando se aplica el clasificador, es necesario estimar parámetros. En este trabajo, se utilizan los que ofrece R en sus librerías, por defecto. Una alternativa es evaluar los resultados de SVM con distintas funciones *núcleos* (kernels), las cuáles requieren estimaciones para sus parámetros y en una de ellas, la utilizada aquí, RBF también se pueden proponer distintas estimaciones.

Analizar y comparar los resultados del coeficiente Kappa con otro, el coeficiente de Kappa ponderado, Cohen (1968).

En este trabajo se usó una metodología para construir el clasificador COMB, de acuerdo a lo propuesto en Kuncheva (2014), donde se le asigna a la clase donde se obtuvo la probabilidad más alta de asignación, pero hay otras opciones para asignar la clase y se puede analizar el comportamiento del clasificador con otras asignaciones.

Apéndice A

Anexo A

En este anexo se completa la información de los resultados obtenidos para todas las situaciones presentadas en el Capítulo 3.

A.1. Subimagen 1

A.1.1. Subimagen 1 con 14 bandas espectrales

En la Tabla A.1 se presenta, para algunos píxeles, las probabilidades y clases obtenidas con los clasificadores ADL, ADC y SVM. La última columna es la clase obtenida con el clasificador COMB. Por ejemplo, el dato 23 es asignado a la clase 4, por ser la que obtuvo probabilidad más alta.

Tabla A.1: Clase asignada por el clasificador COMB

pixel id	probabilidad asignada			clases			
	ADLp	ADCp	SVMp	ADL	ADC	SVM	COMB
20	0.99	1.00	0.80	2	2	2	2
21	0.95	0.85	0.48	2	2	1	2
22	0.93	0.97	0.70	2	4	2	4
23	0.86	0.98	0.44	2	4	1	4
24	0.84	0.99	0.56	2	2	1	2
25	0.89	0.99	0.85	1	2	1	2

En la Figura A.1 se presentan los mapas temáticos obtenidos con ADL; ADC y SVM para la imagen completa y con las muestras originales, para completar la información mostrada en el Capítulo 3, Sección 3.1.1.

En la Figura A.2 se muestran los histogramas por bandas y las correlaciones entre pares de bandas, donde se observan diagramas de dispersión que indican alta asociación lineal y además

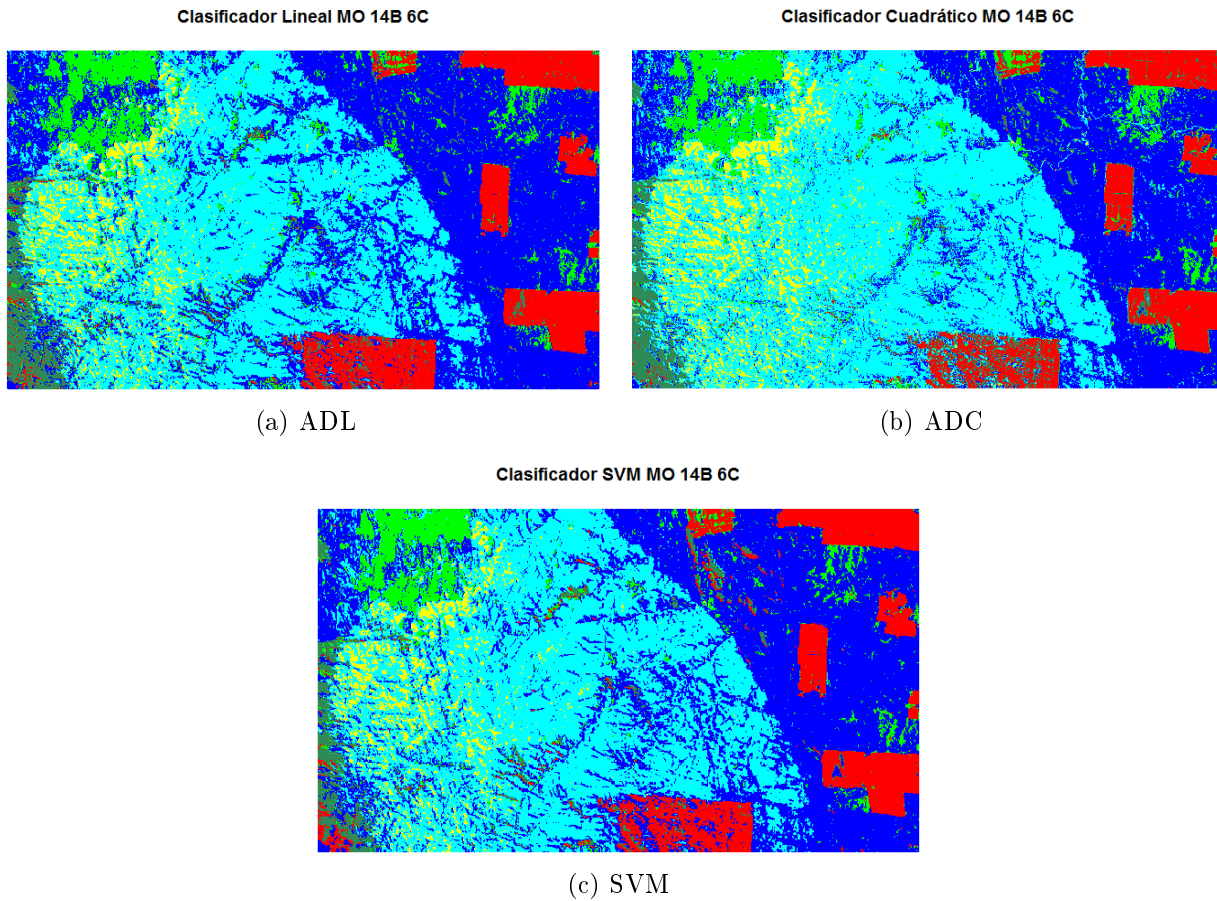


Figura A.1: Mapas temáticos de la Subimagen 1 con 14 bandas y las muestra originales

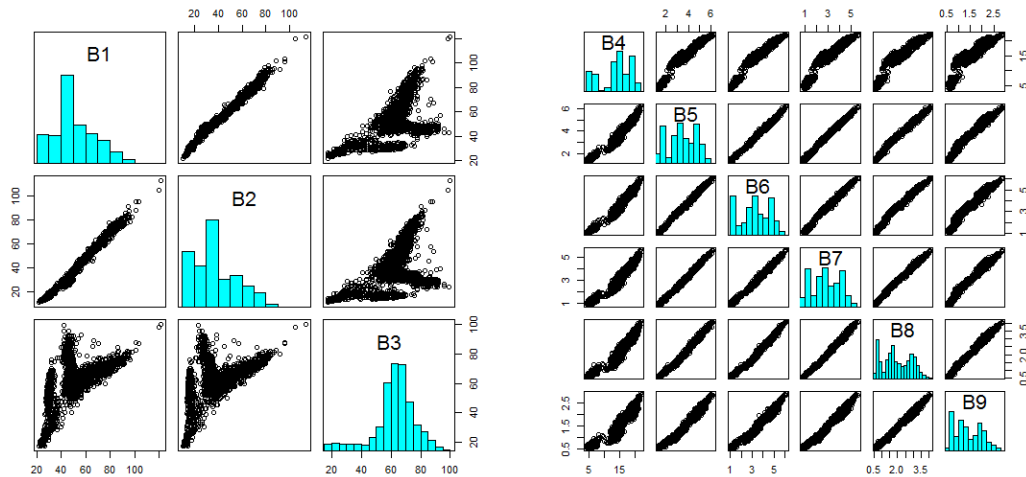
valores similares por grupo. El primer gráfico muestra las bandas 1, 2 y 3 , correspondiente al espectro Visible, el segundo corresponde al Infrarrojo Cercano, bandas 4 a la 9, y el tercer gráfico muestra los diagramas correspondientes al Infrarrojo Térmico, bandas 10 a 14.

A.1.2. Subimagen 1 con tres bandas espectrales

El tamaño de las muestras ampliadas para cada clase con la imagen de tres bandas se presenta en la Tabla A.2. Se puede observar que las muestras para todas las clases aumentaron considerablemente su tamaño, con respecto a las obtenidas en la imagen con catorce bandas. En la Figura A.3 se presentan los mapas temáticos obtenidos con los clasificadores ADL;ADC

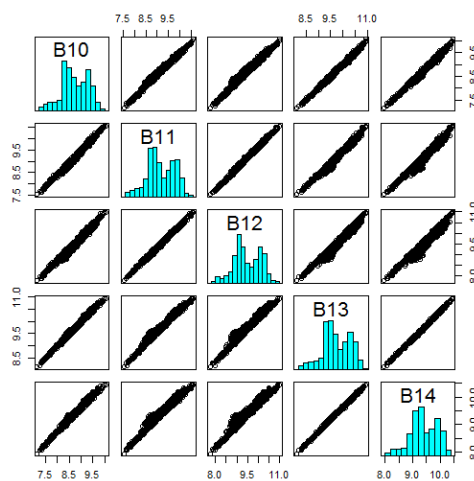
Tabla A.2: Tamaño de las muestras ampliadas para la Subimagen 1 con 3 bandas

1	2	3	4	5	6
903	8079	1278	25580	343	1532



(a) B1,B2 y B3

(b) B4-B9



(c) B10-B14

Figura A.2: Histogramas y Diagramas de dispersión de la Subimagen 1 y las muestras originales

y SVM con las MO para la imagen de tres bandas. Se observan píxeles asignados erróneamente a la clase Pinar (roja) con ADL. El clasificador ADC es el que presenta más ruido y con SVN, en la esquina superior derecha, aparece más claramente un curso de agua.

A.1.3. Subimagen 1 con cuatro bandas espectrales

En la Figura A.4 se muestra la composición RGB para la Subimagen 1, con 4 bandas.

Para la Subimagen 1 con cuatro bandas espectrales se generaron las muestras ampliadas. El tamaño para cada clase se muestra en la Tabla A.3. Al igual que en el caso de tres bandas, las

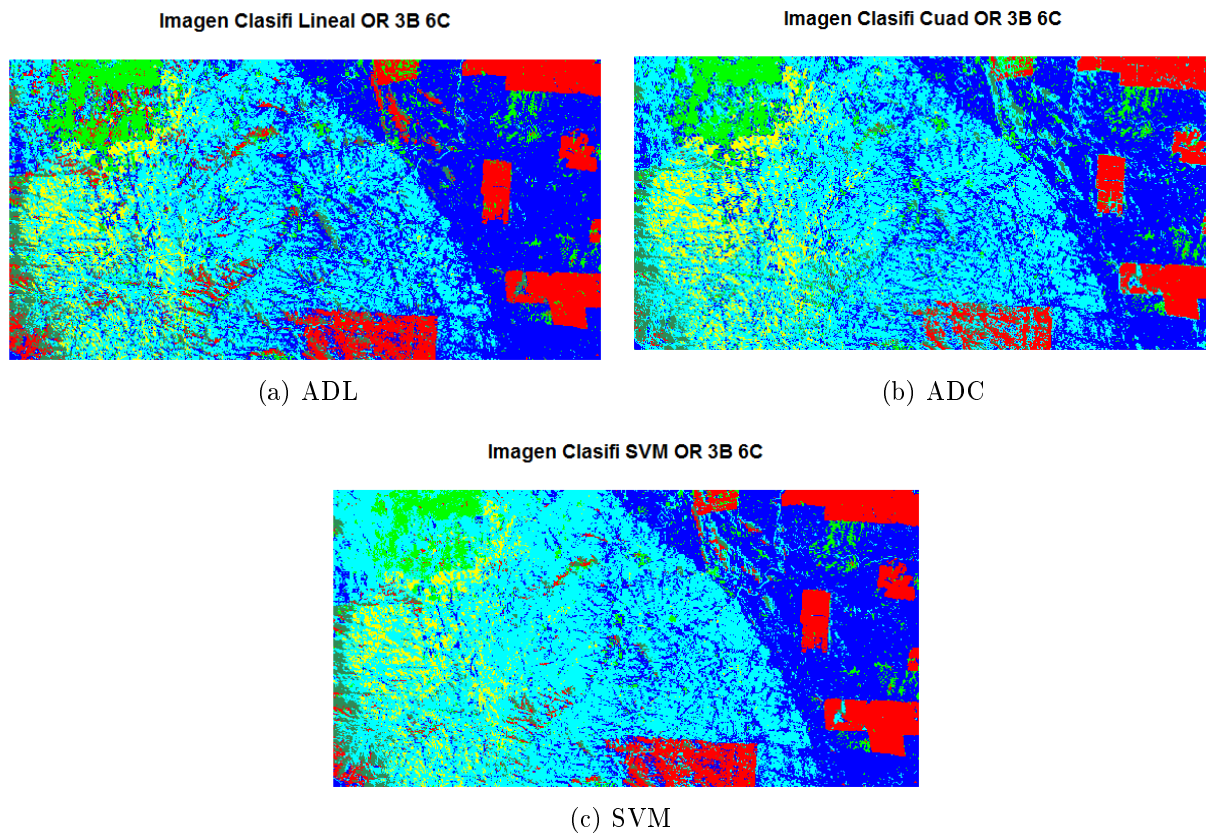


Figura A.3: Mapas temáticos de la Subimagen 1 con 3 bandas y las muestras originales.

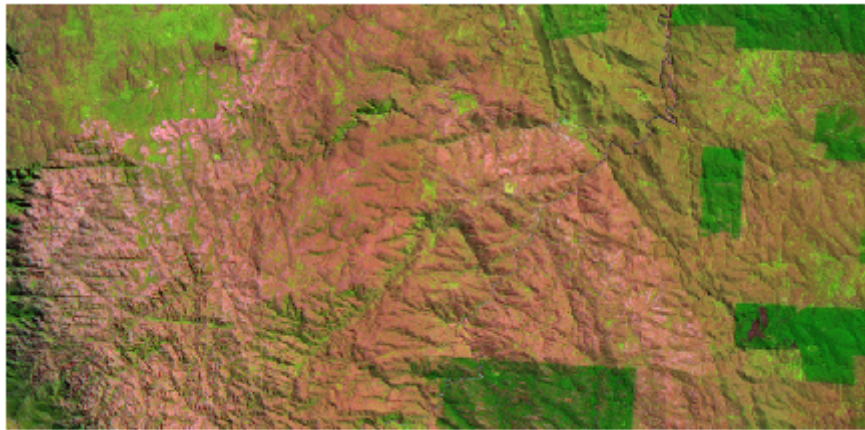


Figura A.4: Composición RGB123 de la Subimagen 1, con 4 bandas

Tabla A.3: Tamaño de las muestras ampliadas para la Subimagen 1 con 4 bandas

1	2	3	4	5	6
771	6492	632	21070	192	987

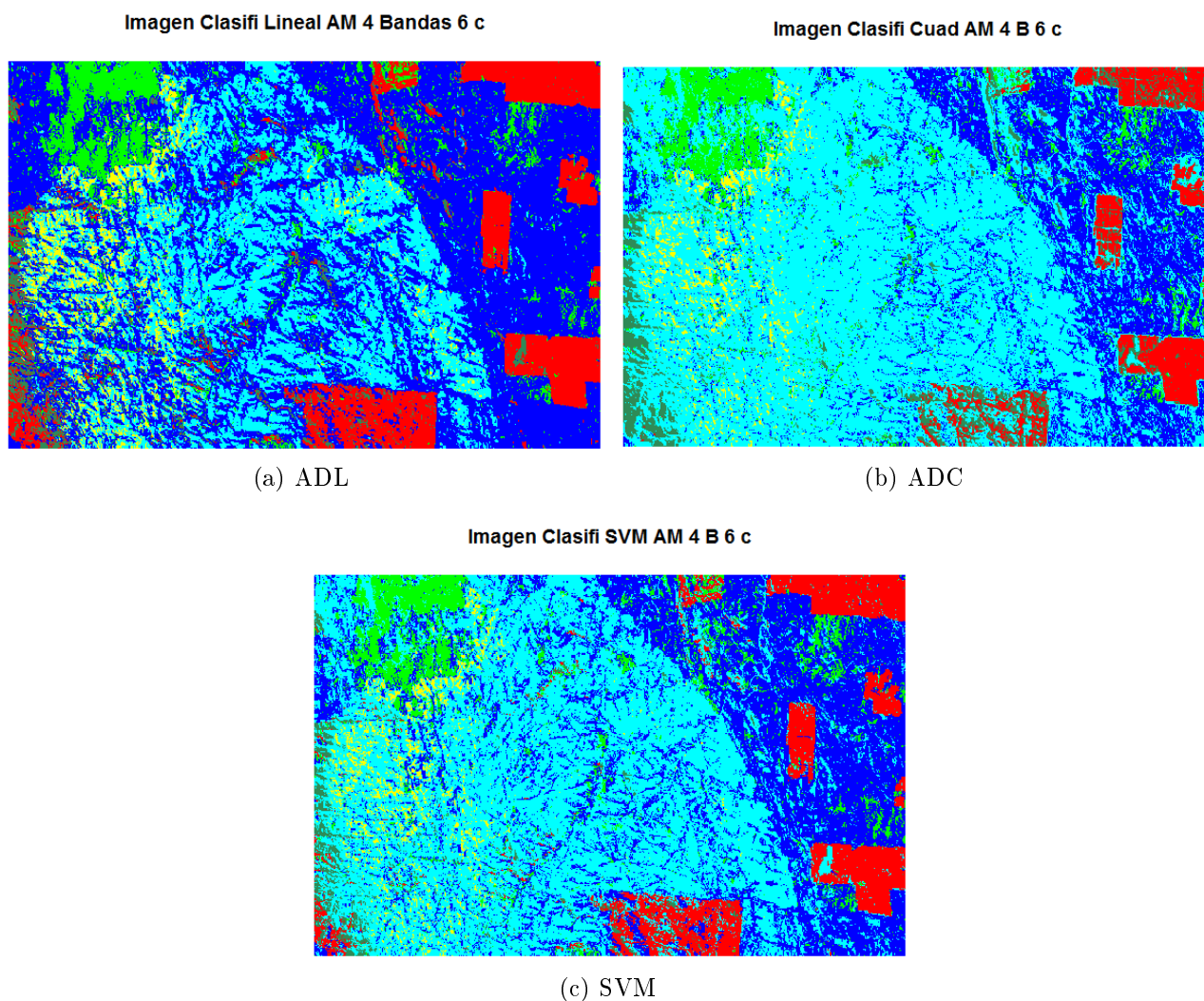


Figura A.6: Clasificación de la Subimagen 1 con 4 bandas y las muestras ampliadas

Tabla A.4: Detalles de las muestras seleccionadas para cada región de la Subimagen 2

Muestras de entrenamiento				Datos de referencia	
Nro.	Clases	Color	píxeles	Color	píxeles
1	pinar	Verde	218	Rojo	208
2	roca metam	Azul	883	Cyan	302
3	unidadexter	Amarillo	353	Magenta	311
4	zona de falla	Negro	111	Maroon	254
Total			1565		1075

A.2.1. Clasificaciones de la Subimagen 2 con 3 bandas

A partir de la imagen completa de 14 bandas se utilizó el método de “Análisis por Factores”, de acuerdo a lo descrito en el Capítulo 2, Sección 2.2.1. El método resume la información en

Tabla A.5: Coeficientes de Fiabilidad (en %) por clase para la Subimagen 2 con 14 bandas y las MO

Clasificador	Clases			
	1	2	3	4
ADL	89.4	99.3	96.1	93.3
ADC	94.2	98.7	100.0	74.8
SVM	86.1	84.1	98.4	90.6
COMB	94.2	99.7	100.0	88.1

Tabla A.6: Coeficientes de Fiabilidad por clase (en %) para la Subimagen 2 con 14 bandas y las MA

Clasificador	Clases			
	1	2	3	4
ADL	89.4	85.8	96.5	91.7
ADC	97.1	82.5	100.0	53.2
SVM	84.6	81.5	99.0	66.1
COMB	95.2	85.4	99.7	75.6

una imagen de 3 bandas. El resultado obtenido muestra una mejora en la visualización, tal como se observa en la Figura A.7.

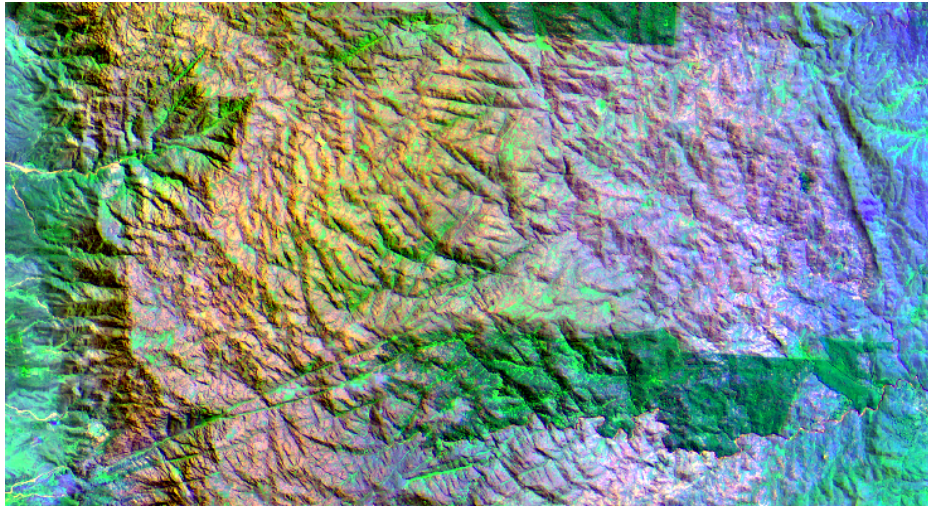


Figura A.7: Composición RGB de la Subimagen 2, con 3 bandas

Para la Subimagen 2 con 3 bandas, en las Tablas A.7 y A.8 se muestran los coeficientes de Fiabilidad para las MO y las MA, respectivamente. Los mismos resultados, están graficados en la Figura A.8.

Tabla A.7: Coeficientes de Fiabilidad por clase para la Subimagen 2 con 3 bandas y las MO

Clasificador	Clases			
	1	2	3	4
ADL	87.98	80.13	87.14	91.73
ADC	94.71	87.41	91.96	83.46
SVM	84.13	82.45	91.64	93.70
COMB	90.39	83.44	91.32	90.94

Tabla A.8: Coeficientes de Fiabilidad por clase para la Subimagen 2 con 3 bandas y las MA

Clasificador	Clases			
	1	2	3	4
ADL	0.8702	0.7914	0.9164	0.9331
ADC	0.8702	0.7947	0.9807	0.8740
SVM	0.6538	0.7914	0.9904	0.7362
COMB	0.8654	0.7947	0.9711	0.9134

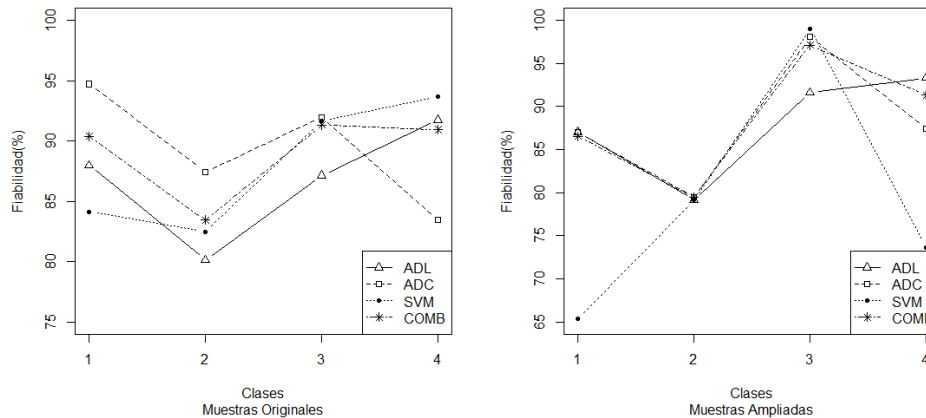


Figura A.8: Coeficientes de Fiabilidad por clase de la Subimagen 2, con 3 bandas

A.2.2. Clasificaciones de la Subimagen 2 con 4 bandas

A partir de la imagen completa de 14 bandas se utilizó el Análisis Discriminante Lineal (estandarizado), de acuerdo a lo descrito en el Capítulo 2 Sección 2.2.2. El método resume la información en una imagen de 4 bandas, ellas son: 3, 9, 12, 13. Los mapas temáticos obtenidos para la Subimagen 2 con 14 y 4 bandas y con las MO se muestran en la Figura A.9.

Tabla A.9: Coeficientes de Fiabilidad por clase para la Subimagen 2 con 4 bandas y las MO.

Clasificador	Clases			
	1	2	3	4
ADL	0.8798	0.9868	0.9421	0.9291
ADC	0.9471	0.9834	0.9968	0.7559
SVM	0.8413	0.8576	0.9936	0.9134
COMB	0.9375	0.9834	0.9871	0.8661

Tabla A.10: Coeficientes de Fiabilidad por clase para la Subimagen 2 con 4 bandas y las MA

Clasificador	Clases			
	1	2	3	4
ADL	0.8606	0.7583	0.9582	0.9291
ADC	0.8702	0.8212	1.0000	0.7835
SVM	0.6635	0.7715	0.8875	0.6890
COMB	0.8702	0.8113	1.0000	0.9016

Imagen 3SUBIMA Clasifi Lineal OR 4B 4

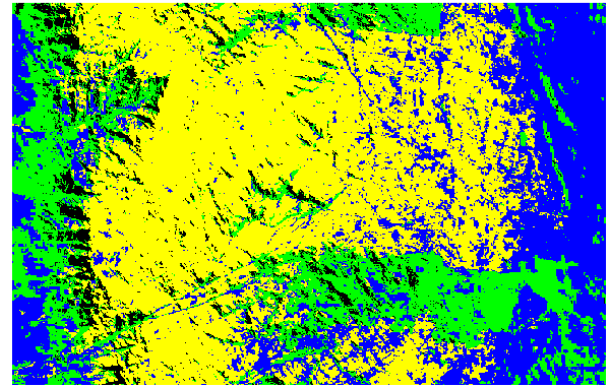
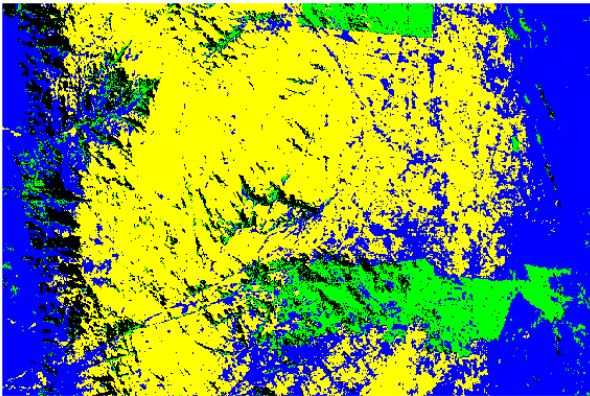


Imagen 3SUBIMA Clasifi Cuad OR 4B 4

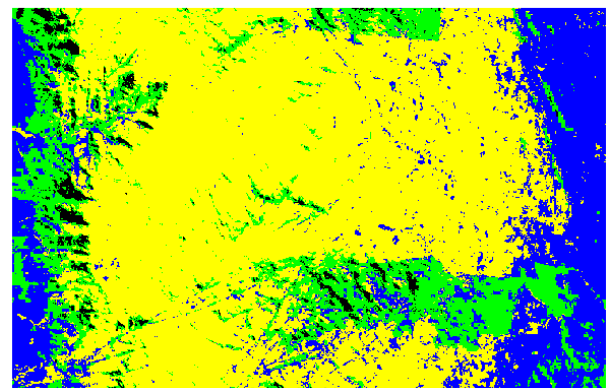
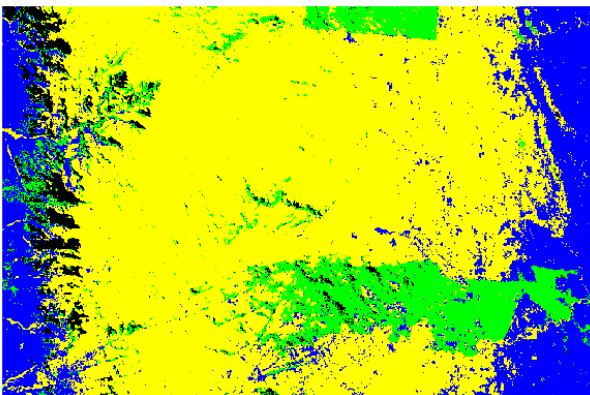


Figura A.9: Mapas temáticos para ADL y ADC con las MO de la Subimagen 2

A.3. Subimagen 3

Los detalles de las muestras de entrenamiento y datos de referencia seleccionadas en la Subimagen 3 se describen en la Tabla A.11. En la Figura A.10 se puede observar la ubicación

Tabla A.11: Detalles de las muestras seleccionadas para cada región de la Subimagen 3

Muestras de entrenamiento				Datos de referencia	
Nro.	Clases	Color	píxeles	Color	píxeles
1	pasto de altura	Rojo	492	Magenta	180
2	zona de falla	Verde	174	Maroon	180
3	roca tipo I	Azul	497	Sea Green	196
4	roca tipo II	Amarillo	414	Purple	182
5	roca tipo III	Cyan	253	Coral	183
Total			1830		921

espacial de los píxeles seleccionados para las muestras de entrenamiento y para los puntos test. La Figura A.11 es una Composición RGB para la Subimagen 3, con 3 y 4 bandas, obtenidas

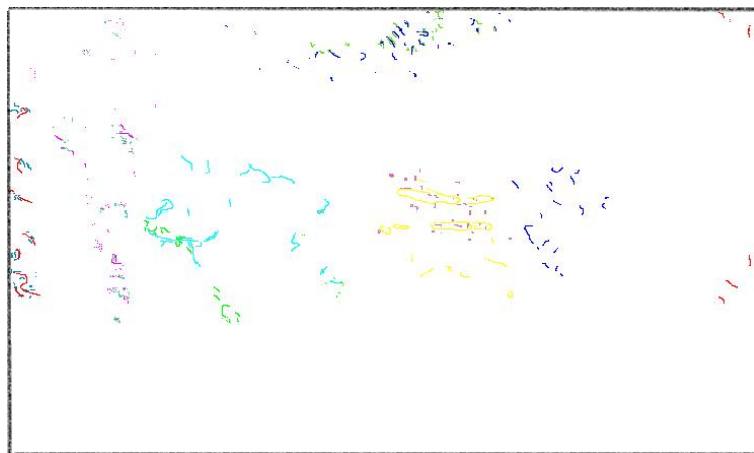


Figura A.10: Muestras de entrenamiento y datos de referencia en la Subimagen 3

según los métodos propuestos en este trabajo. Se obtiene una imagen con una notable mejoría en la visualización, tal como ocurre en las otras imágenes analizadas. Para la imagen de 4 bandas, las que resultaron seleccionadas, fueron: 3,4, 9 y 11.

A.3.1. Resultados de las Clasificaciones de la Subimagen 3

Las tamaños para muestras ampliadas obtenidas, para cada clase con la Subimagen 3, se detallan en la Tabla A.12. En la Tabla A.13 se exhiben, para algunos píxeles, las clases obtenidas para el clasificador COMB.

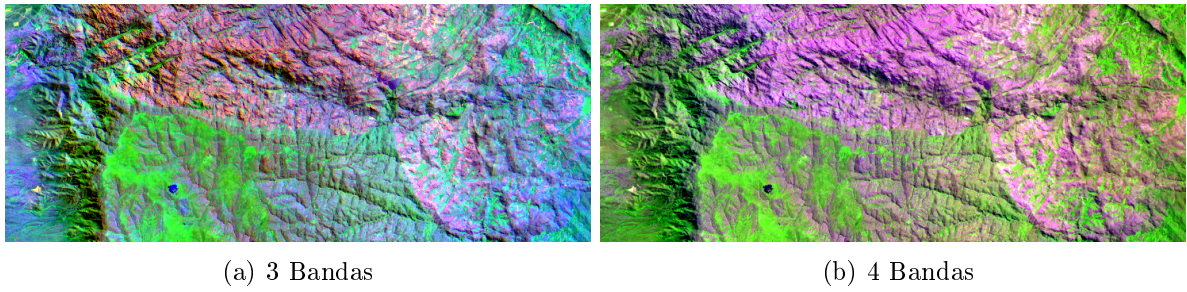


Figura A.11: Composición RGB123 de la Subimagen 3 con 3 y 4 bandas

Tabla A.12: Tamaño de las muestras ampliadas para la Subimagen 3 con 14 bandas

1	2	3	4	5
248	140	2866	3042	334

A.3.2. Validación de la Clasificación para la Subimagen 3

Considerando la Subimagen 3 con 14 bandas, la Tabla A.14 es la Matriz de Confusión para COMB con las MO. Se obtienen muy buenos resultados ya que, son muy pocos los píxeles mal asignados (fuera de la diagonal).

En la tabla A.15 se presentan los resultados obtenidos de las medidas Error (%) y Kappa (%) para todas las situaciones consideradas.

En las Tablas A.16 hasta la A.21 se describe el comportamiento de los clasificadores, en cada clase, para las situaciones consideradas con la Subimagen 3.

Tabla A.13: Clase asignada con el clasificador COMB para Subimagen 3

pixel id	probabilidad asignada			clases			
	ADLp	ADCp	SVMp	ADL	ADC	SVM	COMB
1	0.96	0.95	0.92	1	5	3	1
2	1.00	0.78	0.61	1	3	1	1
3	0.99	0.76	0.56	1	3	1	1
4	0.99	0.79	0.57	1	5	1	1
5	0.92	1.00	0.57	1	5	3	5
6	1.00	0.96	0.50	1	5	3	1
7	1.00	0.61	0.48	1	3	3	1
8	0.99	1.00	0.97	1	5	3	5

Tabla A.14: Matriz de confusión para el clasificador COMB, con MO y Subimagen 3, 14 bandas

clasif	verdad					Sum
	1	2	3	4	5	
1	175.00	0.00	0.00	0.00	0.00	175.00
2	0.00	179.00	0.00	0.00	0.00	179.00
3	5.00	0.00	194.00	5.00	0.00	204.00
4	0.00	0.00	2.00	177.00	0.00	179.00
5	0.00	1.00	0.00	0.00	183.00	184.00
Sum	180.00	180.00	196.00	182.00	183.00	921.00

Tabla A.15: Medidas de confiabilidad para todas las situaciones de la Subimagen 3

Medidas	Error(%)						Kappa(%)					
	14 Bandas		3 Bandas		4 Bandas		14 Bandas		3 Bandas		4 Bandas	
Dimensión	MO	MA	MO	MA	MO	MA	MO	MA	MO	MA	MO	MA
Muestras												
Clasificador												
ADL	1.5	7.6	9.7	11.5	1.5	8.6	98.1	90.5	87.9	85.6	98.1	89.3
ADC	2.1	9.9	9.9	15.1	2.6	11.0	97.4	87.6	87.6	81.1	96.7	86.5
SVM	3.2	11.0	10.1	16.8	4.0	8.3	96.1	86.4	87.4	78.9	95.0	89.7
COMB	1.4	8.9	9.1	14.0	2.3	9.1	98.2	88.9	88.6	82.5	97.1	88.6

Tabla A.16: Coeficientes de Fiabilidad por clase para la Subimagen 3 con 14 bandas y las muestras originales

Clasificador	Clases				
	1	2	3	4	5
ADL	100.0	97.2	99.0	96.2	100.0
ADC	93.9	99.4	98.5	97.8	100.0
SVM	96.1	96.7	99.0	94.5	97.8
COMB	97.2	99.4	99.0	97.3	100.0

Tabla A.17: Coeficientes de Fiabilidad por clase para la Subimagen 3 con 14 bandas y MA

Clasificador	Clases				
	1	2	3	4	5
ADL	100.0	96.7	95.4	87.4	82.5
ADC	92.2	98.3	91.8	95.6	72.7
SVM	95.6	94.4	93.9	87.4	74.3
COMB	95.6	96.7	94.4	91.2	77.6

Tabla A.18: Coeficientes de Fiabilidad por clase para la Subimagen 3 con 3 bandas y las muestras originales

Clasificador	Clases				
	1	2	3	4	5
ADL	99.4	99.4	95.9	78.0	78.7
ADC	94.4	98.9	91.8	80.8	84.7
SVM	93.9	97.8	96.9	78.6	82.0
COMB	98.3	98.9	94.4	79.1	83.6

Tabla A.19: Coeficientes de Fiabilidad por clase para la Subimagen 3 con 3 bandas y MA

Clasificador	Clases				
	1	2	3	4	5
ADL	99.4	99.4	95.9	73.1	74.3
ADC	93.3	99.4	84.7	85.7	61.7
SVM	88.9	93.9	89.8	79.7	63.4
COMB	97.8	99.4	89.3	79.7	63.9

Tabla A.20: Coeficientes de Fiabilidad por clase para la Subimagen 3 con 4 bandas y MO

Clasificador	Clases				
	1	2	3	4	5
ADL	100.0	97.2	99.0	96.2	100.0
ADC	95.0	98.9	99.5	93.4	100.0
SVM	93.9	98.9	99.0	91.8	96.2
COMB	98.3	99.4	99.5	92.9	98.4

Tabla A.21: Coeficientes de Fiabilidad por clase para la Subimagen 3 con 4 bandas y MA

Clasificador	Clases				
	1	2	3	4	5
ADL	100.0	96.7	93.9	87.4	79.2
ADC	92.2	98.9	91.3	92.3	71.6
SVM	95.6	96.1	93.9	93.4	79.8
COMB	96.1	98.9	93.9	90.7	74.9

Apéndice B

CODIGOS en R

En este anexo se detallan los programas implementados en R, para realizar el análisis completo que se presenta en esta tesis.

Los archivos necesarios, que utilizan los programas, están disponibles para el lector que así lo requiera.

El programa que se muestra a continuación es el Principal para realizar la CLASIFICACIÓN:

PROGRAMA PRINCIPAL PARA REALIZAR LAS CLASIFICACIONES

```
#####  
rm(list=ls(all=TRUE));  
library(e1071);library(lattice)  
library(ripa);library(MASS)  
library(xtable)  
#####LEER ARCHIVOS NECESARIOS. DAR EL CAMINO DEL DIRECTORIO#####  
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\imagematrix.R')  
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\panelhist.R')  
##  
#####  
##LEER MUESTRAS DE ENTRENAMIENTO: MO  
#####  
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\leerme.R')  
#####  
##### SE REALIZA LA CLASIFICACION  
# Clasificación con los clasificadores: ADL,ADC,SVM y COMB  
#####
```

```

#Clasificador lineal ADL
#####
options(digits=5)
z <- lda(cl ~ .,cla, prior = c(1,1,1,1,1,1)/6, method='mle')
z
####
### GRAFICO PARA LAS MO EN LAS DOS PRIMERAS FUNC.DISC. LD1 Y LD2#####
n1 = sum(cl == 1)
n2 = sum(cl == 2)
n3 = sum(cl == 3)
n4 = sum(cl == 4)
n5 = sum(cl == 5)
n6 = sum(cl == 6)
colores=c(rep('Green', n1), rep('Blue', n2),
          rep('Yellow', n3), rep('Cyan', n4),
          rep('Sea Green', n5), rep('Red',n6))
plot(z, dimen=2,main='ADL para las muestras OR comp 6', col=colores)
####
#####
#Clasificador cuadrático: ADC
#####
zq <- qda(cl ~ ., cla, prior = c(1,1,1,1,1,1)/6, method='mle')
zq
####
#####
#Clasificador Maquinas de Soporte Vectorial: SVN estandarizado
#####
zs <- svm(cl ~ .,cla, prior =c(1,1,1,1,1,1)/6,kernel='radial',scale=T,
          probability = TRUE)
summary(zs)

#####
#####Clasificando toda la imagen #####
#####LEER ARCHIVO CON DATOS DE LA IMAGEN#####
#####
#####

```

```
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\leerimagenC.R')
#####
#####SE PREDICEN LAS CLASES PARA TODOS LOS PIXELES DE LA IMAGEN
#####
dat=cbind(datos[3:16])
#####
#### Con análisis discriminante lineal
#####
pred1p=predict(z, dat)$posterior
pred1c=predict(z,dat)$class
pred1=cbind(pred1p,pred1c)
max1=apply(pred1p,1,max);head(max1)
pread1=cbind(max1,pred1c);head(pread1)
pread1=cbind(max1,pred1c);tail(pread1)
####
#####SE VISUALIZA Y GRABA EL MAPA TEMATICO CON ADL
######3
vi=as.numeric(pred1c)
c3=matrix(vi,ncol=360, byrow=T);dim(c3)=c(739,360)
dim(c3)
x11()
image(c3[,360:1], col=c('Green','Blue','Yellow','Cyan',
    'Sea Green','Red'), main='Clasificador Lineal MO 14B 6C',
    axes=FALSE)
write.table(c3,file='clasifdlmOR6C.txt',sep=' ',quote=FALSE, row.names=F)
###
#####
##### Con análisis discriminante cuadrático ADC #####
#####
pred2c=predict(zq, dat)\$class
pred2p=predict(zq, dat)\$posterior
max2=apply(pred2p,1,max);head(max2)
preadq=cbind(max2,pred2c);head(preadq)
## SE VISUALIZA Y GRABA EL MAPA TEMATICO CON ADC
viq=as.numeric(pred2c)
c5=matrix(viq,ncol=360, byrow=T);dim(c5)=c(739,360)
```

```

dim(c5)
x11()
image(c5[,360:1], col=c('Green','Blue','Yellow','Cyan',
  'Sea Green','Red'), main='Clasificador Cuadrático MO 14B 6C',
  axes=FALSE)
write.table(c5,file='clasifdqOR6C.txt',sep=' ',quote=FALSE, row.names=F)
#####
#####
###Con análisis clasificador SVM #####
#####
pred3c=predict(zs, dat)
pred3=predict(zs, dat, decision.values = TRUE, probability = TRUE)
pred3p=attr(pred3, 'probabilities')
max3=apply(pred3p,1,max);head(max3)
presvm=cbind(max3,pred3c);head(presvm)
#####
###SE VISUALIZA Y GRABA EL MAPA TEMATICO CON SVM
vis=as.numeric(pred3c)
c7=matrix(vis,ncol=360, byrow=T);dim(c7)=c(739,360)
dim(c7)
x11()
image(c7[,360:1], col=c('Green','Blue','Yellow',
  'Cyan','Sea Green','Red'), main='Clasificador SVM MO 14B 6C',
  axes=FALSE)
write.table(c7,file='clasifsvmOR6C.txt',sep=' ',quote=FALSE, row.names=F)
###
#Gráfico
###
plot(cmdscale(dist(c7)),
  col = colores,
  pch = c('o','+')[1:2127 %in% zs$index + 1])
#####
###
#####PARA CONSTRUIR CLASIFICADOR COMBINADO
#####
com=cbind(max1,max2,max3);head(com)

```

```

dim=dim(com)
dimc=dim[1];dimc
comp=cbind(pred1c,pred2c,pred3c);head(comp)
comple=cbind(max1,max2,max3,pred1c,pred2c,pred3c)
head(comple)
tail(comple)
##
#
rango=apply(comp,1,max)- apply(comp,1,min)
head(rango)
###
max=max.col(com)
max4=cbind(1:dimc,max)
clnu=comp[max4]
head(clnu)
su=data.frame(ADLp = max1, ADCp = max2, SVMp = max3, ADL=pred1c,
  ADC=pred2c, SVM=pred3c, cln = ifelse(rango == 0,comple[,4:6],clnu))
su[1:8,]
tail(su)
print(xtable(su[20:25,]))
####
### SE VISUALIZA Y GRABA EL MAPA TEMATICO CON COMB
#####
vin=as.numeric(su\$cln)
c9=matrix(vin,ncol=360, byrow=T);dim(c9)=c(739,360)
dim(c9)
x11()
image(c9[,360:1], col=c('Green','Blue','Yellow',
  'Cyan','Sea Green','Red'), main='Clasificador Combinado MO 14B 6C',
  axes=FALSE)
write.table(c9,file='clasifcommOR6C.txt',sep=' ',quote=FALSE, row.names=F)
###
#####
#

```

El programa anterior llama dos programas que serán nombrados “leerme.R” y “leerima-

genC.R” y leen las muestras de entrenamiento y la imagen completa, respectivamente.

Programa “*leerme.R*”

```
#####
#LEER MUESTRAS ENTRENAMIENTO ORIGINALES: MO
#####
cla=read.table("6clac.txt", header=TRUE)
attach(cla)
cl=factor(rep(1:6,c(200,480,256,721,213,257)))
t2=data.frame(cla)
summary(t2)
head(t2)
#Grafico
xyplot(B3~B2,group=cl,data=cla,cex=1,pch=13)
xyplot(B3~B4,group=cl,data=cla,cex=1,pch=13)
xyplot(B14~B4,group=cl,data=cla,cex=1,pch=13)
xyplot(B4~B2,group=cl,data=cla,cex=1,pch=13)
by(t2,cl,summary)
#Se realiza un Histograma y diagrama de dispersión, utilizando panel.hist.R
###
#luego se escribe la sentencia donde se usa la función
pairs(t2[1:3], diag.panel=panel.hist)
x11()
pairs(t2[4:9], diag.panel=panel.hist)
x11()
pairs(t2[10:14], diag.panel=panel.hist)
#####
```

Las sentencias que se muestran a continuación son para leer la imagen completa, que está guardada en un archivo *.txt. Una vez acomodados los datos quedan guardados en la variable “datos”.

Programa “*leerimagenC.R*”

```
#####
#####LEER Y VISUALIZAR IMAGEN SUBIMAGEN 1 PARA TRABAJAR CON R##
#####
#####
```

```
comp=read.table('1SUBIMACOMPL30.txt', header=F)
attach(comp)
icompleta=data.frame(comp)
dim=dim(icompleta)
dim
#####
####SE NECESITA SEPARAR LOS DATOS EN CADA BANDA (14 EN TOTAL)
#####
m1=icompleta[1:360, ]; b1=as.matrix(m1,nrow=360,ncol=739)
dim(b1); B1=c(b1)
m2=icompleta[361:720,]; b2=as.matrix(m2,nrow=360,ncol=739)
dim(b2); B2=c(b2)
m3=icompleta[721:1080,]; b3=as.matrix(m3,nrow=360,ncol=739)
dim(b3); B3=c(b3)
m4=icompleta[1081:1440,]; b4=as.matrix(m4,nrow=360,ncol=739)
dim(b4); B4=c(b4)
m5=icompleta[1441:1800,]; b5=as.matrix(m5,nrow=360,ncol=739)
dim(b5); B5=c(b5)
m6=icompleta[1801:2160,]; b6=as.matrix(m6,nrow=360,ncol=739)
dim(b6); B6=c(b6)
m7=icompleta[2161:2520,]; b7=as.matrix(m7,nrow=360,ncol=739)
dim(b7); B7=c(b7)
m8=icompleta[2521:2880,]; b8=as.matrix(m8,nrow=360,ncol=739)
dim(b8); B8=c(b8)
m9=icompleta[2881:3240,]; b9=as.matrix(m9,nrow=360,ncol=739)
dim(b9); B9=c(b9)
m10=icompleta[3241:3600,]; b10=as.matrix(m10,nrow=360,ncol=739)
dim(b10); B10=c(b10)
m11=icompleta[3601:3960,]; b11=as.matrix(m11,nrow=360,ncol=739)
dim(b11); B11=c(b11)
m12=icompleta[3961:4320,]; b12=as.matrix(m12,nrow=360,ncol=739)
dim(b12); B12=c(b12)
m13=icompleta[4321:4680,]; b13=as.matrix(m13,nrow=360,ncol=739)
dim(b13); B13=c(b13)
m14=icompleta[4681:5040,]; b14=as.matrix(m14,nrow=360,ncol=739)
dim(b14); B14=c(b14)
```

```

#
#Como necesitamos una matriz formada por Filas, Columnas y las 14 bandas,
#entonces las sentencias siguientes generan cada fila 739 veces, donde 739
#es la cantidad de columnas y 360 es la cantidad de filas.
# Entonces el 1 aparece 739 veces, el 2 aparece 739 veces,....., el 360
# aparece 739 veces, entonces 360x739=266040.
#
C=rep(1:739, each = 360, len = 266040)
d=cbind(C,F=1:360,B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,B11,B12,B13,B14)
datos=data.frame(d)
dim=dim(datos);dim
datos[1:2,]
#####
#para visualizar por banda
#####
#se abre una ventana nueva
####
x11()
vb1=plot(normalize(imagematrix(equalize(b1), noclip=T)))
vb2=plot(normalize(imagematrix(equalize(b2), noclip=T)))
vb3=plot(normalize(imagematrix(equalize(b3), noclip=T)))
vb4=plot(normalize(imagematrix(equalize(b4), noclip=T)))
vb5=plot(normalize(imagematrix(equalize(b5), noclip=T)))
vb6=plot(normalize(imagematrix(equalize(b6), noclip=T)))
vb7=plot(normalize(imagematrix(equalize(b7), noclip=T)))
vb8=plot(normalize(imagematrix(equalize(b8), noclip=T)))
vb9=plot(normalize(imagematrix(equalize(b9), noclip=T)))
vb10=plot(normalize(imagematrix(equalize(b10), noclip=T)))
vb11=plot(normalize(imagematrix(equalize(b11), noclip=T)))
vb12=plot(normalize(imagematrix(equalize(b12), noclip=T)))
vb13=plot(normalize(imagematrix(equalize(b13), noclip=T)))
vb14=plot(normalize(imagematrix(equalize(b14), noclip=T)))
###
#####
#####
#para visualizar una composición color

```



```
#####
####guardamos una matriz#####
#####SE GUARDA UNA MATRIZ DE TRES DIMENSIONES
matriz3M=array(dim=c(360,739,3))
matriz3M[, ,1] = matrix(normalize(B4), byrow=T)
matriz3M[, ,2] = matrix(normalize(B3), byrow=T)
matriz3M[, ,3] = matrix(normalize(B9), byrow=T)
plot(imagematrix(matriz3M))
#####FIN#####
```

Para realizar la validación de la Clasificación se corre el siguiente programa: *“verifi6Clcsvm.c.R”*

```
#Este programa evalúa la validez de la clasificación, arma la matriz de confusión
#y los coeficientes de Fiabilidad Global, Error, de Concordancia Kappa y
#Fiabilidad por clase.
#Primero: Leer la imagen clasificada, es una matriz (739,360) (Porque lee
#primero las columnas y luego las filas)
#
rm(list=ls(all=TRUE))
library(MASS);library(psych)
library(xtable);library(mda)
#####Con Discriminante Lineal, IMAGEN CON TODAS BANDAS MO#####
a=read.table(“clasifdlmOR6C.txt”, header=T)
attach(a)
dim(a)
#####Con Discriminante Cuadrático, IMAGEN CON TODAS BANDAS MO###
b=read.table(“clasifdqOR6C.txt”, header=T)
attach(b)
dim(b)
#####Con SVM, IMAGEN CON TODAS BANDAS MO#####
c=read.table(“clasifsvmOR6C.txt”, header=T)
attach(c)
dim(c)
#####Con COMB, IMAGEN CON TODAS BANDAS MO#####
d=read.table(“clasifcommOR6C.txt”, header=T)
attach(d)
dim(d)
```

```
#luego leemos los puntos test, conjunto del cual sabemos la clase a que
#pertenece cada dato, es la verdad. Un único archivo donde estan las columnas
#y filas (en ese orden), de todas las clases, en este caso 6. Tanto la imagen
#clasificada como los puntos test, están ordenados de manera que aparecen
#primero columnas y despues filas.
#
cf=read.table('puntest6C.txt', header=T)
attach(cf);dim(cf)
cf1=as.matrix(cf)
#para controlar la dimensión
dim=dim(cf1);dim
#la sentencia siguiente extrae el dato del archivo original 'a', en la posición
#indicada en el archivo (de los puntos test) 'cf1'.
clas1=a[cf1]
clas2=b[cf1]
clas3=c[cf1]
clas4=d[cf1]
verd=rep(1:6,c(175,239,150,481,160,484))
mat1=cbind(clas1,verd)
mat2=cbind(clas2,verd)
mat3=cbind(clas3,verd)
mat4=cbind(clas4,verd)
tab1=table(clas1,verd)
tab2=table(clas2,verd)
tab3=table(clas3,verd)
tab4=table(clas4,verd)
###
##Otra opción para matriz de confusión con el error
##
confusion(clas1, verd)
confusion(clas2, verd)
confusion(clas3, verd)
confusion(clas4, verd)
#####
#para contruir la matriz de confusión agregando los márgenes y controlar
```

```

#la dimensión
matconf=addmargins(tab1)
print(xtable(matconf))
matconf1=addmargins(tab2)
print(xtable(matconf1))
matconf2=addmargins(tab3)
print(xtable(matconf2))
matconf3=addmargins(tab4)
print(xtable(matconf3))

cat('Matriz de Confusión con DL, 6 clases, Imagen Completa MO', '\n')
matconf
cat('Matriz de Confusión con DC, 6 clases, Imagen Completa MO', '\n')
matconf1
cat('Matriz de Confusión con SVM, 6 clases, Imagen Completa MO', '\n')
matconf2
cat('Matriz de Confusión con COMB, 6 clases, Imagen Completa MO', '\n')
matconf3

#####
##Cálculo del coeficiente de Fiabilidad Global DL
F1=(sum(diag(tab1)/sum(tab1)))*100
cat('Fiabilidad Global= ', F1, '\n')
n3=sum(tab1)*sum(diag(tab1))- sum(rowSums(tab1)*colSums(tab1))
d3=sum(tab1)^2 - sum(rowSums(tab1)*colSums(tab1))
K1=(n3/d3)*100
cat('Kappa = ', K1, '\n')###
###calcula Fiabilidad por clase
diag(prop.table(tab1, 2))
# total percent correct
sum(diag(prop.table(tab1)))

###
##Cálculo del coeficiente de F Global, Kappa y F por clase con DC
F2=(sum(diag(tab2))/sum(tab2))*100
cat('Fiabilidad Global= ', F2, '\n')

```

```

n3c=sum(tab2)*sum(diag(tab2))- sum(rowSums(tab2)*colSums(tab2))
d3c=sum(tab2)^2 - sum(rowSums(tab2)*colSums(tab2))
Kc=(n3c/d3c)*100
cat('Kappa = ', Kc, '\n')###
diag(prop.table(tab2, 2))
# total percent correct
sum(diag(prop.table(tab2)))
#####
#
##Cálculo del coeficiente de Fiabilidad Global SVM
F3=(sum(diag(tab3))/sum(tab3))*100
cat('Fiabilidad Global= ', F3, '\n')
n3c=sum(tab3)*sum(diag(tab3))- sum(rowSums(tab3)*colSums(tab3))
d3c=sum(tab3)^2 - sum(rowSums(tab3)*colSums(tab3))
Ks=(n3c/d3c)*100
cat('Kappa = ', Ks, '\n')###
diag(prop.table(tab3, 2))
# total percent correct
sum(diag(prop.table(tab3)))
#####
###
##Cálculo del coeficiente de Fiabilidad Global COMB
F4=(sum(diag(tab4))/sum(tab4))*100
cat('Fiabilidad Global= ', F4, '\n')
n3c=sum(tab4)*sum(diag(tab4))- sum(rowSums(tab4)*colSums(tab4))
d3c=sum(tab4)^2 - sum(rowSums(tab4)*colSums(tab4))
Kco=(n3c/d3c)*100
cat('Kappa = ', Kco, '\n')###
#####
diag(prop.table(tab4, 2))
# total percent correct
sum(diag(prop.table(tab4)))
#####
#####

```

Calcular “MUESTRAS AMPLIADAS”.

En el programa que se presenta a continuación “preproc06C.R” es para obtener la metodología propuesta para obtener las *muestras ampliadas* .

```
rm(list=ls(all=TRUE));
source('http://www.phaget4.org/R/myImagePlot.R')
library(e1071);library(lattice)
library(ripa);library(MASS)
library(psych)
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\imagematrix.R')
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\panelhist.R')

#####
###LEE MUESTRAS DE ENTRENAMIENTO
#####
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\leerme.R')

#ARMAR EL ARCHIVO PARA CADA CLASE
cla1=cla[1:200,];head(cla1)
cla2=cla[201:680,];head(cla2)
cla3=cla[681:936,];head(cla3)
cla4=cla[937:1657,];head(cla4)
cla5=cla[1658:1870,];head(cla5)
cla6=cla[1871:2127,];head(cla6)
###
###
###
#Calcular los cuartiles para cada clase y cada banda
b1=boxplot(B1~cl, plot=F)
b2=boxplot(B2~cl, plot=F)
b3=boxplot(B3~cl, plot=F)
b4=boxplot(B4~cl, plot=F)
b5=boxplot(B5~cl, plot=F)
b6=boxplot(B6~cl, plot=F)
b7=boxplot(B7~cl, plot=F)
b8=boxplot(B8~cl, plot=F)
b9=boxplot(B9~cl, plot=F)
```

```
b10=boxplot(B10~c1, plot=F)
b11=boxplot(B11~c1, plot=F)
b12=boxplot(B12~c1, plot=F)
b13=boxplot(B13~c1, plot=F)
b14=boxplot(B14~c1, plot=F)
#####
#####
b18=b1$stats[c(2,4),];b18
b28=b2$stats[c(2,4),];b28
b38=b3$stats[c(2,4),];b38
b48=b4$stats[c(2,4),];b48
b58=b5$stats[c(2,4),];b58
b68=b6$stats[c(2,4),];b68
b78=b7$stats[c(2,4),];b78
b88=b8$stats[c(2,4),];b88
b98=b9$stats[c(2,4),];b98
b108=b10$stats[c(2,4),];b108
b118=b11$stats[c(2,4),];b118
b128=b12$stats[c(2,4),];b128
b138=b13$stats[c(2,4),];b138
b148=b14$stats[c(2,4),];b148
#####
qu_c1a1=cbind(b18[,1],b28[,1],b38[,1],b48[,1],b58[,1],b68[,1],b78[,1],b88[,1],
b98[,1],b108[,1],b118[,1],b128[,1],b138[,1],b148[,1]);qu_c1a1
qu_c1a2=cbind(b18[,2],b28[,2],b38[,2],b48[,2],b58[,2],b68[,2],b78[,2],b88[,2],
b98[,2],b108[,2],b118[,2],b128[,2],b138[,2],b148[,2]);qu_c1a2
qu_c1a3=cbind(b18[,3],b28[,3],b38[,3],b48[,3],b58[,3],b68[,3],b78[,3],b88[,3],
b98[,3],b108[,3],b118[,3],b128[,3],b138[,3],b148[,3]);qu_c1a3
qu_c1a4=cbind(b18[,4],b28[,4],b38[,4],b48[,4],b58[,4],b68[,4],b78[,4],b88[,4],
b98[,4],b108[,4],b118[,4],b128[,4],b138[,4],b148[,4]);qu_c1a4
qu_c1a5=cbind(b18[,5],b28[,5],b38[,5],b48[,5],b58[,5],b68[,5],b78[,5],b88[,5],
b98[,5],b108[,5],b118[,5],b128[,5],b138[,5],b148[,5]);qu_c1a5
qu_c1a6=cbind(b18[,6],b28[,6],b38[,6],b48[,6],b58[,6],b68[,6],b78[,6],b88[,6],
b98[,6],b108[,6],b118[,6],b128[,6],b138[,6],b148[,6]);qu_c1a6

#####
```

```
#####  
# recorriendo la matriz  
#####  
#####  
#####leer imagen Completa ‘leerimagenC.R’  
#  
source(‘E:\\SUSANA\\tesisdoc\\PROGDIC16\\leerimagenC.R’)  
#  
dat=datos  
l = dim(dat);l  
nuevosDatos = matrix(data = NA, nrow = l[1], ncol = 17);  
nuevosDatos_cla1 = matrix(data = 0, nrow = 739, ncol = 360);  
nuevosDatos_cla2 = matrix(data = 0, nrow = 739, ncol = 360);  
nuevosDatos_cla3 = matrix(data = 0, nrow = 739, ncol = 360);  
nuevosDatos_cla4 = matrix(data = 0, nrow = 739, ncol = 360);  
nuevosDatos_cla5 = matrix(data = 0, nrow = 739, ncol = 360);  
nuevosDatos_cla6 = matrix(data = 0, nrow = 739, ncol = 360);  
j=1;  
  
ptm <- proc.time()  
for(i in 1:l[1]){  
  
if(qu_cla1[1,1] <= dat[i,3] & qu_cla1[2,1] >= dat[i,3] &  
    qu_cla1[1,2] <= dat[i,4] & qu_cla1[2,2] >= dat[i,4]&  
    qu_cla1[1,3] <= dat[i,5] & qu_cla1[2,3] >= dat[i,5]&  
    qu_cla1[1,4] <= dat[i,6] & qu_cla1[2,4] >= dat[i,6] &  
    qu_cla1[1,5] <= dat[i,7] & qu_cla1[2,5] >= dat[i,7]&  
    qu_cla1[1,6] <= dat[i,8] & qu_cla1[2,6] >= dat[i,8]&  
    qu_cla1[1,7] <= dat[i,9] & qu_cla1[2,7] >= dat[i,9]&  
    qu_cla1[1,8] <= dat[i,10] & qu_cla1[2,8] >= dat[i,10]&  
    qu_cla1[1,9] <= dat[i,11] & qu_cla1[2,9] >= dat[i,11]&  
    qu_cla1[1,10] <= dat[i,12] & qu_cla1[2,10] >= dat[i,12]&  
    qu_cla1[1,11] <= dat[i,13] & qu_cla1[2,11] >= dat[i,13]&  
    qu_cla1[1,12] <= dat[i,14] & qu_cla1[2,12] >= dat[i,14]&  
    qu_cla1[1,13] <= dat[i,15] & qu_cla1[2,13] >= dat[i,15] &  
    qu_cla1[1,14] <= dat[i,16] & qu_cla1[2,14] >= dat[i,16]){
```

```

flag_cla1 = 1
}
else{
flag_cla1 = 0
}

    if(qu_cla2[1,1] <= dat[i,3] & qu_cla2[2,1] >= dat[i,3] &
       qu_cla2[1,2] <= dat[i,4] & qu_cla2[2,2] >= dat[i,4]&
       qu_cla2[1,3] <= dat[i,5] & qu_cla2[2,3] >= dat[i,5]&
       qu_cla2[1,4] <= dat[i,6] & qu_cla2[2,4] >= dat[i,6] &
       qu_cla2[1,5] <= dat[i,7] & qu_cla2[2,5] >= dat[i,7]&
       qu_cla2[1,6] <= dat[i,8] & qu_cla2[2,6] >= dat[i,8]&
       qu_cla2[1,7] <= dat[i,9] & qu_cla2[2,7] >= dat[i,9]&
       qu_cla2[1,8] <= dat[i,10] & qu_cla2[2,8] >= dat[i,10]&
       qu_cla2[1,9] <= dat[i,11] & qu_cla2[2,9] >= dat[i,11]&
       qu_cla2[1,10] <= dat[i,12] & qu_cla2[2,10] >= dat[i,12]&
       qu_cla2[1,11] <= dat[i,13] & qu_cla2[2,11] >= dat[i,13]&
       qu_cla2[1,12] <= dat[i,14] & qu_cla2[2,12] >= dat[i,14]&
       qu_cla2[1,13] <= dat[i,15] & qu_cla2[2,13] >= dat[i,15] &
       qu_cla2[1,14] <= dat[i,16] & qu_cla2[2,14] >= dat[i,16]){
flag_cla2 = 1
}
else{
flag_cla2 = 0
}

    if(qu_cla3[1,1] <= dat[i,3] & qu_cla3[2,1] >= dat[i,3] &
       qu_cla3[1,2] <= dat[i,4] & qu_cla3[2,2] >= dat[i,4]&
       qu_cla3[1,3] <= dat[i,5] & qu_cla3[2,3] >= dat[i,5]&
       qu_cla3[1,4] <= dat[i,6] & qu_cla3[2,4] >= dat[i,6] &
       qu_cla3[1,5] <= dat[i,7] & qu_cla3[2,5] >= dat[i,7]&
       qu_cla3[1,6] <= dat[i,8] & qu_cla3[2,6] >= dat[i,8]&
       qu_cla3[1,7] <= dat[i,9] & qu_cla3[2,7] >= dat[i,9]&
       qu_cla3[1,8] <= dat[i,10] & qu_cla3[2,8] >= dat[i,10]&
       qu_cla3[1,9] <= dat[i,11] & qu_cla3[2,9] >= dat[i,11]&
       qu_cla3[1,10] <= dat[i,12] & qu_cla3[2,10] >= dat[i,12]&
       qu_cla3[1,11] <= dat[i,13] & qu_cla3[2,11] >= dat[i,13]&

```



```

        qu_cla3[1,12] <= dat[i,14] & qu_cla3[2,12] >= dat[i,14]&
        qu_cla3[1,13] <= dat[i,15] & qu_cla3[2,13] >= dat[i,15] &
        qu_cla3[1,14] <= dat[i,16] & qu_cla3[2,14] >= dat[i,16]){
flag_cla3 = 1
}
else{
flag_cla3 = 0
}
    if(qu_cla4[1,1] <= dat[i,3] & qu_cla4[2,1] >= dat[i,3] &
        qu_cla4[1,2] <= dat[i,4] & qu_cla4[2,2] >= dat[i,4]&
        qu_cla4[1,3] <= dat[i,5] & qu_cla4[2,3] >= dat[i,5]&
        qu_cla4[1,4] <= dat[i,6] & qu_cla4[2,4] >= dat[i,6] &
        qu_cla4[1,5] <= dat[i,7] & qu_cla4[2,5] >= dat[i,7]&
        qu_cla4[1,6] <= dat[i,8] & qu_cla4[2,6] >= dat[i,8]&
        qu_cla4[1,7] <= dat[i,9] & qu_cla4[2,7] >= dat[i,9]&
        qu_cla4[1,8] <= dat[i,10] & qu_cla4[2,8] >= dat[i,10]&
        qu_cla4[1,9] <= dat[i,11] & qu_cla4[2,9] >= dat[i,11]&
        qu_cla4[1,10] <= dat[i,12] & qu_cla4[2,10] >= dat[i,12]&
        qu_cla4[1,11] <= dat[i,13] & qu_cla4[2,11] >= dat[i,13]&
        qu_cla4[1,12] <= dat[i,14] & qu_cla4[2,12] >= dat[i,14]&
        qu_cla4[1,13] <= dat[i,15] & qu_cla4[2,13] >= dat[i,15] &
        qu_cla4[1,14] <= dat[i,16] & qu_cla4[2,14] >= dat[i,16]){
flag_cla4 = 1
}
else{
flag_cla4 = 0
}
    if(qu_cla5[1,1] <= dat[i,3] & qu_cla5[2,1] >= dat[i,3] &
        qu_cla5[1,2] <= dat[i,4] & qu_cla5[2,2] >= dat[i,4]&
        qu_cla5[1,3] <= dat[i,5] & qu_cla5[2,3] >= dat[i,5]&
        qu_cla5[1,4] <= dat[i,6] & qu_cla5[2,4] >= dat[i,6] &
        qu_cla5[1,5] <= dat[i,7] & qu_cla5[2,5] >= dat[i,7]&
        qu_cla5[1,6] <= dat[i,8] & qu_cla5[2,6] >= dat[i,8]&
        qu_cla5[1,7] <= dat[i,9] & qu_cla5[2,7] >= dat[i,9]&
        qu_cla5[1,8] <= dat[i,10] & qu_cla5[2,8] >= dat[i,10]&
        qu_cla5[1,9] <= dat[i,11] & qu_cla5[2,9] >= dat[i,11]&

```

```

    qu_cla5[1,10] <= dat[i,12] & qu_cla5[2,10] >= dat[i,12]&
    qu_cla5[1,11] <= dat[i,13] & qu_cla5[2,11] >= dat[i,13]&
    qu_cla5[1,12] <= dat[i,14] & qu_cla5[2,12] >= dat[i,14]&
    qu_cla5[1,13] <= dat[i,15] & qu_cla5[2,13] >= dat[i,15] &
    qu_cla5[1,14] <= dat[i,16] & qu_cla5[2,14] >= dat[i,16]){
flag_cla5 = 1
}
else{
flag_cla5 = 0
}
    if(qu_cla6[1,1] <= dat[i,3] & qu_cla6[2,1] >= dat[i,3] &
    qu_cla6[1,2] <= dat[i,4] & qu_cla6[2,2] >= dat[i,4]&
    qu_cla6[1,3] <= dat[i,5] & qu_cla6[2,3] >= dat[i,5]&
    qu_cla6[1,4] <= dat[i,6] & qu_cla6[2,4] >= dat[i,6] &
    qu_cla6[1,5] <= dat[i,7] & qu_cla6[2,5] >= dat[i,7]&
    qu_cla6[1,6] <= dat[i,8] & qu_cla6[2,6] >= dat[i,8]&
    qu_cla6[1,7] <= dat[i,9] & qu_cla6[2,7] >= dat[i,9]&
    qu_cla6[1,8] <= dat[i,10] & qu_cla6[2,8] >= dat[i,10]&
    qu_cla6[1,9] <= dat[i,11] & qu_cla6[2,9] >= dat[i,11]&
    qu_cla6[1,10] <= dat[i,12] & qu_cla6[2,10] >= dat[i,12]&
    qu_cla6[1,11] <= dat[i,13] & qu_cla6[2,11] >= dat[i,13]&
    qu_cla6[1,12] <= dat[i,14] & qu_cla6[2,12] >= dat[i,14]&
    qu_cla6[1,13] <= dat[i,15] & qu_cla6[2,13] >= dat[i,15] &
    qu_cla6[1,14] <= dat[i,16] & qu_cla6[2,14] >= dat[i,16]){
flag_cla6 = 1
}
else{
flag_cla6 = 0
}
if((flag_cla1 + flag_cla2 + flag_cla3 + flag_cla4 + flag_cla5 + flag_cla6 ) == 1){
if(flag_cla1 == 1){
nuevosDatos[j,] = c(d[i,],1);
nuevosDatos_cla1[(d[i,1]),(d[i,2])]=1;
j=j+1;
}
else{}

```

```
if(flag_cla2 == 1){
nuevosDatos[j,] = c(d[i,],2);
nuevosDatos_cla2[(d[i,1]),(d[i,2])]=1;
j=j+1;
}
else{}
if(flag_cla3 == 1){
nuevosDatos[j,] = c(d[i,],3);
nuevosDatos_cla3[(d[i,1]),(d[i,2])]=1;
j=j+1;
}
else{}
if(flag_cla4 == 1){
nuevosDatos[j,] = c(d[i,],4);
nuevosDatos_cla4[(d[i,1]),(d[i,2])]=1;
j=j+1;
}
else{}
        if(flag_cla5 == 1){
nuevosDatos[j,] = c(d[i,],5);
nuevosDatos_cla5[(d[i,1]),(d[i,2])]=1;
j=j+1;
}
        else{}
                if(flag_cla6 == 1){
nuevosDatos[j,] = c(d[i,],6);
nuevosDatos_cla6[(d[i,1]),(d[i,2])]=1;
j=j+1;
}
                else{}
}
}
nuevas_muestras = nuevosDatos[1:(j-1),];
time=proc.time() - ptm; print(time);
write.table(nuevas_muestras,file='nuevasmuestras6c.txt',sep=' ',
            quote=FALSE, row.names=F)
```

```

x11()
myImagePlot(nuevosDatos_cla1)
x11()
myImagePlot(nuevosDatos_cla2)
x11()
myImagePlot(nuevosDatos_cla3)
x11()
myImagePlot(nuevosDatos_cla4)
x11()
myImagePlot(nuevosDatos_cla5)
x11()
myImagePlot(nuevosDatos_cla6)

```

REDUCCION DE DIMENSION UTILIZANDO **FA** y la metodología propuesta en este trabajo. El Programa que se presenta a continuación “*fa.R*” realiza el Análisis por Factores y calcula las nuevas variables.

```

rm(list=ls(all=TRUE))
library(psych); library(GPARotation)
library(ripa); library(MASS)
library(xtable)
#
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\imagematrix.R')
#

#LEER IMAGEN PARA TRABAJAR CON R
#
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\leerimagenC.R')
#
dat1=cbind(B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,B11,B12,B13,B14)
dim1=dim(dat1); dim1
##
#análisis por factores
##
f3t= fa(r=dat1,nfactors=2,n.obs=266040)
f3t

```

```
f3lo=f3t\$load
x11()
fa.diagram(f3t)
x11()
plot(f3t)
###
##
# De acuerdo a los grupos se combinan las variables del mismo grupo, sacando
# las medias y formando nuevas variables en este caso, son tres las var nuevas
BN1=cbind(B1,B2,B4,B5,B6,B7,B8,B9)
MBN1=apply(BN1,1,mean)
length(MBN1)
BN3=cbind(B10,B11,B12,B13,14)
MBN3=apply(BN3,1,mean)
length(MBN3)
MBN2=B3
length(MBN2)
datnuevo=cbind(MBN1,MBN2,MBN3)
write.table(datnuevo,file='datnuevo.txt',sep=' ',quote=FALSE, row.names=F)
###
#compruebo que las variables no están correlacionadas
###
cor(datnuevo)

####
##
##LEER Y VISUALIZAR LA IMAGEN NUEVA
##
ncomp=read.table('datnuevo.txt', header=T)
attach(ncomp)
incompleta=data.frame(ncomp)
dim=dim(incompleta)
dim
###
###
##
```

```
#convierto cada banda en una matriz para visualizar
##
x=array(MBN1,dim=c(360,739))
y=array(MBN2,dim=c(360,739))
z=array(MBN3,dim=c(360,739))
###grabo para importar y visualizar en ENVI
#acomodo cada banda por fila, una debajo de otra
#
nuev=rbind(x,y,z)
write.table(nuev,file='nuev.txt',sep=' ',quote=FALSE, row.names=F)
####
###

#
#se abre una ventana nueva
####
x11()
vbn1=plot(normalize(imagematrix(equalize(x), noclip=T)))
vbn2=plot(normalize(imagematrix(equalize(y), noclip=T)))
vbn3=plot(normalize(imagematrix(equalize(z), noclip=T)))
#
#Visualizar una composición color
#
matriz3M=array(dim=c(360,739,3))#guardo una matriz
matriz3M[, ,1] = matrix(normalize(x), byrow=T)
matriz3M[, ,2] = matrix(normalize(y), byrow=T)
matriz3M[, ,3] = matrix(normalize(z), byrow=T)
plot(imagematrix(matriz3M))
###
##
C=rep(1:739, each = 360, len = 266040)
nd=cbind(F=1:360,C,MBN1,MBN2,MBN3)
ndatos=data.frame(nd)
dim=dim(ndatos);dim
ndatos[1:5,1:4]
datn1=cbind(MBN1,MBN2,MBN3)
```

```
datosn1=data.frame(datn1)
dim1=dim(datn1);dim1
```

Otra propuesta de este trabajo es reducir la dimensión, utilizando Análisis Discriminante Canónico (estandarizado). El siguiente programa calcula las Funciones Discriminantes Estandarizadas y se analizan los coeficientes y también los gráficos que resultan.

```
#####
#ANALISIS DISCRIMINANTE ESTANDARIZADO
#####
rm(list=ls(all=TRUE))
library(ripa);library(MASS)
library(psych)
library(e1071)
library(memisc)
library(candisc)
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\imagematrix.R')
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\panelhist.R')
##Leer las Muestras de Entrenamientos
source('E:\\SUSANA\\tesisdoc\\PROGDIC16\\leerme.R')
#
y=cbind(B1,B2,B3,B4,B5,B6,B7,B8,B9,B10,B11,B12,B13,B14)
modelodis <- lm(y ~cl, data=cla)
canl=candiscList(modelodis)
canl
names(canl)
names(canl\\$cl)
can <- candisc(modelodis, data= cla)
summary(can, coeff='std')
uno=can\\$coeffs.std
uno
x11()
heplot(can)
x11()
# 1-dim plot
can1m <- candisc(modelodis, data= cla, ndim=1)
```

```
can1m2 <- candisc(modelodis, data= cla, ndim=2)
x11()
plot(can1m, type='n', ask=FALSE)
x11()
plot(can1m2,ylim=c(-8,8))
###
```

Una vez que se reduce la dimensión, con las variables resultantes, 3 y 4 en este caso, los pasos serían: realizar la Clasificación con la nueva dimensión, para las MO; calcular nuevamente las MA y repetir el análisis y por último llevar a cabo la validación de los resultados.

Apéndice C

Anexo Digital

El archivo incorporado al CD, que acompaña al documento impreso de tesis, es “tesisSusanaFerrero.pdf”

Bibliografía

- Agresti, A. (1990). *Categorical Data Analysis*. John Wiley & Sons, Inc.
- Alpaydin, E. (2010). *Introduction to Machine Learning*. The MIT Press, Cambridge, Massachusetts, second edition.
- Ben-Hur, A. y Weston, J. (2010). A users guide to support vector machines. *Data mining techniques for the life sciences*, pages 223–239.
- Benkhoff, J. and Helbert, J. y the MERTIS team (2006). Thermal infrared spectroscopy to investigate the composition of mercury the MERTIS instrument on BepiColombo. *Advances in Space Research*, 38(4):647–658.
- Chuvieco, E. (2002). *Teledetección Ambiental*. Editorial Ariel S.A. España.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4):213–220.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37:35–46.
- Congalton, R. G. y Green, K. (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. Lewis Publishers, CRC Press, Inc.
- Crosta, A., Filho, C. D. S., Azevedo, F., y Brodie, C. (2003). Targeting key alteration minerals in epithermal deposits in patagonia, argentina, using aster imagery and principal component analysis. *International Journal of Remote Sensing*, 24(21):4233–4240.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer Berlin Heidelberg.
- Ding, C., Li, X., Liu, X., y Zhao, L. (2015). Quartzose mafic spectral feature space model: A methodology for extracting felsic rocks with ASTER thermal infrared radiance data. *Ore Geology Reviews*, 66(0):283 – 292.

- Ding, C., Liua, X., Liu, W., Liu, M., y Li, Y. (2014). Mafic– ultramafic and quartz– rich rock indices deduced from ASTER thermal infrared data using a linear approximation to the planck function. *Ore Geology Reviews*, 60:161–173.
- Duda, R. O., Hart, P. E., y Stork, D. G. (2001). *Pattern classification*. John Wiley & Sons, New York, second edition.
- ENVI (2009). *The Environment for Visualizing Images*. Research Systems Inc. USA,, 4.7 edition.
- Ferrero, S. (2012). Criterios para aumentar la separabilidad en muestras de entrenamiento utilizadas para clasificación supervisada. Primera reunión del proyecto Universal financiado por el CNPq (número 471805/2011-4 Teoria da Informacon na Análise de Imagens SAR Polarimétricas. INPE (Instituto Nacional de Pesquisas Espaciais). Brasil. Sao Jose dos Campos, SP.
- Ferrero, S., Bustos, O., y Frery, A. (2005). Assessing classification accuracy with simulated multispectral imagery. *XI Reunión de Trabajo en Procesamiento de la Información y Control (RPIC)*, 0:0.
- Ferrero, S. B. y Bustos, O. H. (2017). Dimension reduction: two proposals using multivariate techniques before classifying multispectral images. IV JIAAIS Jornadas Interdisciplinarias de Análise Avançada de Imagens e Sinais. Universidade Federal de Alagoas, Maceió, Brasil. 10-15/NOV.
- Fisher, R. A. (1936). The use of múltiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188.
- Foody, G. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80:185–201.
- Frery, A., Ferrero, S., y Bustos, O. (2006). Accuracy of satistical classification strategies in remote sensing imagery. *19th Brazilian Simposium on Computer Graphics and Image Processing. SIBGRAPI '06*, 1 (ISBN: 0-7695-2686-1):255–262.
- Frery, A., Ferrero, S., y Bustos, O. (2009). The influence of training errors, context and number of bands in the accuracy of image classification. *International Journal Remote Sensing*, 30:1425–1440.
- Gordon, A. D. (1999). *Classification*. Chapman & Hall / CRC, second edition.

- Imani, M. y Ghassemian, H. (2014). Band clustering-based feature extraction for classification of hyperspectral images using limited training samples. *IEEE Geoscience and Remote Sensing Letters*, 11(8):1325–1329.
- Kalinowski, A. y Oliver, S. (2004). *ASTER Mineral Index Processing Manual*. Remote Sensing Applications Geoscience Australia, www.ga.gov.au/image-cache/GA7833.pdf.
- Krzanowski, W. J. y Marriott, F. H. (1995). *Multivariate analysis: classification*. London:Arnold.
- Kuncheva, L. I. (2014). *Combining Pattern Classifiers Methods and Algorithms*. John Wiley & Sons, Inc., Hoboken, New Jersey., 2nd edition.
- Pérez, D. J., Ducart, D. F., y Crósta, A. P. (2007). Mapeo geológico con datos ASTER y ETM en Sierra de Umango (29°00´S) Sierras Pampeanas Occidentales, provincia de La Rioja, Argentina. *Anais XIII Simpósio Brasileiro de Sensoriamento Remoto. Florianópolis Brasil*, pages 2103–2110.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rencher, A. C. (1998). *Multivariate Statistical Inference and Applications*. John Wiley & Sons.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. John Wiley & Sons, Inc., second edition.
- Richards, J. A. y Jia, X. (2006a). *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag Berlin, 4th edition.
- Richards, J. A. y Jia, X. (2006b). *Remote Sensing Digital Image Analysis. An Introduction*. 4th edition.
- Tulyakov, S., Jaeger, S., Govindaraju, V., y Doermann, D. (2008). *Review of Classifier Combination Methods*, pages 361–386. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA.
- Vargas, J., Conde, M. B., Paccapelo, M. V., y Zingaretti, M. L. (2012). Máquinas de soporte vectorial: Metodología y aplicación en R. *Congreso Latinoamericano de Sociedades de Estadística; Décimo Congreso Latinoamericano de Sociedades de Estadística (X CLATSE)*.

- Vaughan, R. y Calvin, W.M. and Taranik, J. (2003). SEBASS hyperspectral thermal infrared data: surface emissivity measurement and mineral mapping. *Remote Sensing of Environment*, 85(1):48–63(16).
- Wang, S., Lu, J., Gu, X., Du, H., y Yang, J. (2016). Semi-supervised linear discriminant analysis for dimension reduction and classification. *Pattern Recognition*, 57(Supplement C):179 – 189.
- Watanabe, H. y Matsuo, K. (2003). Rock type classification by multi-band TIR of ASTER. *Geosciences Journal*, 7(4):347–358.
- Wolpert, D. H. y Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82.
- Zhang, X., Pazner, M., y Duke, N. (2007). Lithologic and mineral information extraction for gold exploration using aster data in the south chocolate mountains (california). *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(4):271–282.