National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

NL 339 (r 88/04) c

Canada

Automated Detection of Multiple Sclerosis Lesions in Magnetic Resonance Images of the Human Brain

Micheline Kamber

A Thesis
in
The Department
of
Computer Science

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Computer Science at
Concordia University
Montréal, Québec, Canada

November, 1991

©Micheline Kamber, 1991

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation

ISBN 0 315-73635 6

Canada

ABSTRACT

Automated Detection of Multiple Sclerosis Lesions in Magnetic Resonance Images of the Brain

Micheline Kamber

Magnetic resonance (MR) imaging is a medical technique which permits the visualization of a variety of tumors, lesions, and abnormalities present within the soft biological tissues of the body. Segmentation of medical image data is the process of assigning anatomically-meaningful labels to each component of the image. This thesis describes the development of a tool for the segmentation of MR images of the head. In particular, the tool is designed for the detection of multiple sclerosis lesions of the brain. The design was based on two objectives: i) to evaluate the effectiveness of incorporating *a priori* knowledge of brain anatomy in the classification process, and ii) to compare the statistical and symbolic approaches to machine learning.

Knowledge of neuroanatomy is represented in the form of a tissue probability model. The model was constructed to provide *a priori* probabilities of brain tissue distribution per unit voxel in a standardized 3D 'brain space'. Use of the model to detect multiple sclerosis lesions reduces the number of false positive lesions by 50%.

The performance of the statistical minimum distance and Bayesian classifiers was compared to that of a symbolic decision tree learning algorithm. A version of this algorithm for the handling of noisy data was included in the comparative study. Each classifier performed at about the same level of accuracy. The statistical classifiers were the fastest in training, yet were the slowest in recall.

RESUME

La détection automatisée des lesions de la sclérose en plaques du cerveau en imagerie par résonance magnétique

Micheline Kamber

L'imagerie par résonance magnétique (IRM) est une technique médicale qui permet la visualisation d'une grande variété de tumeurs, lésions, ou autres anomalies anatomiques du corps humain. La segmentation des images médicales est le processus d'affectation d'une entité anatomique identifiée par un nom (matière grise, matière blanche, fluide cérébro spinal, ..) à chaque composant de l'image. Cette thèse décrit le développement d'un outil pour la segmentation des images IRM du cerveau. Cet outil a été construit en particulier pour la détection des lésions caracteristiques de la sclérose en plaques. Son développement a été réalisé avec deux objectifs: i) évaluer l'efficacité de l'utilisation de la connaissance *a priori* de l'anatomie cérébrale dans le processus de classification, et ii) comparer les approches statistiques et symboliques de l'apprentissage automatique.

La connaissance neuroanatomique est représentée par un modèle de probabilité. Ce modèle a été construit pour fournir les probabilités de distribution des tissues cérébraux par 'voxel' dans un système de coordonnées standard. L'utilisation du modèle a permis d'éliminer 50% des fausses positives lésions.

Les performances des algorithmes de classification des plus proches voisins (PPV) et Bayesien ont été comparées avec celles d'un algorithme symbolique d'arbre de décision. Une version de cet algorithme pour les données bruitées a aussi été évaluée.

Acknowledgements

I wish to express my sincere thanks to my supervisors, Dr. R. Shinghal of Concordia University, and Dr. A. Evans of the Montreal Neurological Institute (M.N.I.), for their invaluable direction, encouragement, and support. Their guidance is greatly appreciated.

My gratitude goes out to Dr. Partlow for patiently trying to instill in me the basics of neuroanatomy. I thank Dr. Francis and Dr. Del Carpio for their instruction on the detection of MS lesions. My gratitude is extended to Dr. Arnold for his evaluation of the results of the segmentation tool, and for his interest in the program.

I wish to thank Youpu Zhang and Wei Qian Dai of the NeuroImaging Laboratory at the M.N.I. for implementation of the program to transform MR image volumes into Talairach space, and Louis Collins, also of the NeuroImaging Lab, for the image homomorphic filtering procedures.

I am grateful for the many hours spent by Alain Gauvin and Atsushi Takahashi in operating the MRI scanner for acquisition of the image data for the model. My thanks goes out to all those who kindly agreed to volunteer for scanning.

I would like to take this opportunity to thank all my associates at the Montreal Neurological Institute, and my family, who have helped in one way or another. I thank Stanley Hum for the manual segmentation of MS lesions, and Dr. Peters, Chris Henri, Louis Collins, and Vasco Kollokian for their valuable suggestions and moral support.

I am grateful for the financial assistance provided by the Natural Sciences and Engineering Research Council of Canada and the Concordia University Fellowship program. I am also thankful for the hardware made available by the American Multiple Sclerosis Society.

Contents

List of Figures

Chapter 1

Introduction

1.1 Segmentation of Magnetic Resonance Images

Magnetic Resonance Imaging (MRI) is a noninvasive medical technique employed to produce cross-sectional (i.e. tomographic) images of the human body, permitting the detailed visualization of biological soft tissues. As a diagnostic tool, MRI allows the detection of a variety of tumors, lesions, and abnormalities present within internal anatomical structures. Although it can be applied to all parts of the body, MRI has been particularly useful for imaging the brain and spinal cord. This is due to its ability to discriminate between grey and white matter tissues, and fluid spaces, unlike computed tomography, [1] an earlier commonly used imaging system. MR data is generally acquired as a series of 2-dimensional (2D) images spanning the organ under study. The 2D images or 'slices' make up a volume of MR data.

Segmentation or *tissue classification* of magnetic resonance images is the division of the image data into regions corresponding to anatomical tissues, fluids, and structures. Regions occur as individual pixels are assigned class labels with each label representing a different tissue type. Figure 1.1 shows an example of a magnetic

[1] **Computed tomography**, or CT, is an ionizing imaging technique similar to X-ray imagery. A computer is used, instead of film, to hold images.

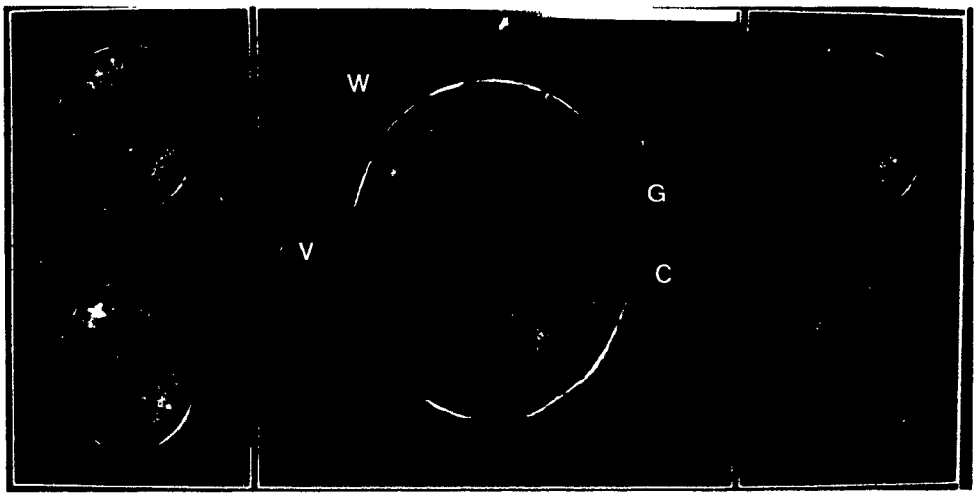


Figure 1.1: A typical magnetic resonance image of the head. Major tissue classes include: grey matter (G), white matter (W), ventricular cerebrospinal fluid (V), and external cerebrospinal fluid (C). Ventricular cerebrospinal fluid is located in structures referred to as the *ventricles*. External cerebrospinal fluid refers to the cerebrospinal fluid beneath the 'arachnoid' layer of the brain. 'Folds' or convolutions of the brain are referred to as *gyri*.

resonance image of the head. The classes of interest correspond to the 'gross' or major tissues of the brain. These include grey matter, white matter, ventricular cerebrospinal fluid (CSF), and external CSF.

Magnetic resonance imaging is typically used qualitatively with radiologists inspecting the 2D images for abnormal structures or deformations. There is a growing interest, however, in the development of computerized tools to permit quantitative analysis and visualization of anatomical structures in 3D, particularly in images of the brain [Kubler and Gerig, 1990; Spitzer and Stiehl, 1989]. These tools require accurate segmentation of MR volume data. The segmentation of head MR data in routine clinical applications gives important quantitative information about anatomy in normal and diseased brain. Salient examples include brain atrophy, tumor volume, and morphological changes of cerebral structures. Such quantitative analysis aids in the evaluation, diagnosis and study of neurological diseases and psychiatric disorders. Work in segmentation has been done to study changes in brain tissues in patients with Alzheimer's disease and schizophrenia [Press, Amaral, and Squire, 1989; Kolnne, 1989], and to study the process of aging [Jernigan, Press and Hesselink, 1990; Wahlund, Agartz, Almqvist, Basun, Forssell, Saaf, and Wetterberg, 1990; Pfefferbaum, Zatz, and Jernigan, 1986]. Quantitative measurements of cerebral structures such as the *amygdala* and *hippocampus* have been conducted in order to investigate their contribution to epilepsy and memory functions [Cedes, Andermann, Watson, Evans, Gloor, Melanson, Gotman, Leroux, Olivier, and Peters, 1991; Cascino, Jack, Parisi, Sharbrough, Hirschorn, Meyer, Marsh, and O'Brien, 1991]. The quantitative measurement of tissue volumes also assists in assessing the effectiveness of drug treatment and radiation therapy intended to reduce the size of abnormal tissues.

MR segmentation can be used to detect multiple sclerosis lesions of the brain. Multiple sclerosis (MS) is an autoimmune disease characterized by lesions or 'damage' of the myelin covering (fatty white substance) of neurons of cerebral white matter. In MR images, these lesions appear as regions of hyperintensities, initially

small and isolated, but becoming more extensive and connected as the disease progresses (Figure 1.2). Quantitative measurement of MS lesion volume is important for the study of the disease, the evaluation of drug treatments, and MS patient follow-up.

It is possible to view a set of 2D images and mentally reconstruct the 3D shape of the anatomical structures within them. This visualization task, however, is difficult and requires extensive training [Gerig, Martin, Kikinis, Kubler, Shenton, and Jolesz, 1991]. Automated segmentation of MR volume data allows the computerized display of individual brain structures in 3D. Procedures permitting the 3D visualization of anatomical structures can serve as educational tools, as well as facilitate surgical and radiotherapy planning. Surgeons, for example, would like to have 3D views of the internal structure of the brain so that they can assess the depth and shape of a lesion as well as its geometric relationship to other internal structures [Levin, Hu, Tan, Galhotra, Herrmann, Chen, Pelizzari, Balter, Beck, Chen, and Cooper, 1989]. This means of accurate tumor localization would enable radiation therapy to be planned more effectively to treat malignant tissues and to minimize the irradiation of adjacent tissues [Fan, Trivedi, Fellingham, and Gamboa-Aldeco, 1987].

Manual segmentation of individual structures is tedious, time-consuming, and costly. Errors occur due to poor hand-eye coordination, low tissue contrast, and unclear tissue boundaries made up of data elements appearing as partial volumes[2]. Thus, despite its potential clinical value, segmentation is seldom performed manually [Ozkan, Sprenkels, and Dawant, 1990]. Accurate automated segmentation of MR images is the first step towards promising quantitative and 3D visualization applications of magnetic resonance imaging.

Segmentation is a central problem in medical imaging. Human investigators implicitly segment an image into its anatomical components, drawing on a stored

[2]**Partial volume** refers to the case of a data element (pixel or voxel) containing more than one tissue type. (A voxel is a 3D pixel).

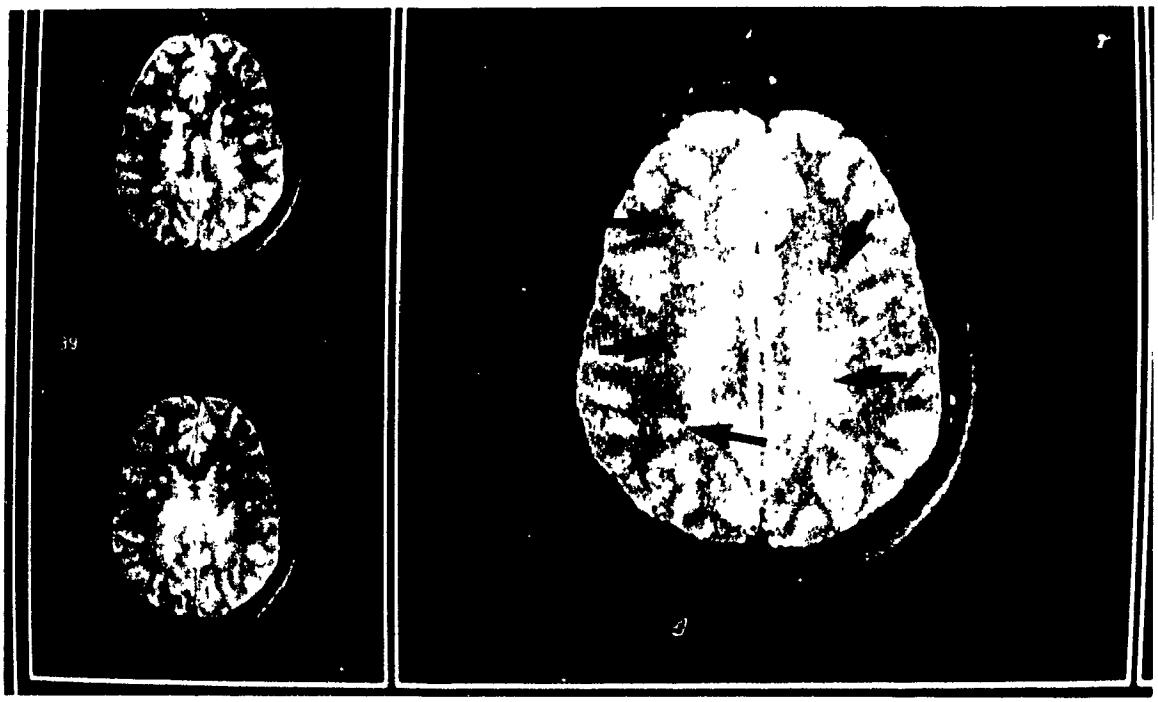


Figure 1.2: A typical magnetic resonance image of the brain of a patient afflicted with multiple sclerosis. Multiple sclerosis lesions (*arrows*) are demonstrated by the bright areas distributed across the area of the image.

knowledge of anatomy to overcome image artifacts, noise, and lack of tissue contrast. The pattern recognition qualities of the eye and brain are often overlooked until one attempts to imitate them with a computer algorithm. Thus, aside from being of interest to the medical community, the segmentation of magnetic resonance images presents an interesting challenge to computer scientists. It is a real-world problem for which pattern recognition and machine learning methods can be developed, applied, and tested.

1.2 Approaches to Image Segmentation

Magnetic resonance images have the following characteristics:

1. **MR images contain noise and artifacts.** While some artifacts are specific to the scanner used, the majority are inherent in the imaging method itself. Factors causing MR artifacts include patient motion as well as inhomogeneities and nonlinearities in the magnetic fields applied during image acquisition [Porter, Hastrup, Richardson, Wesbey, Olson, Cromwell, and Moss, 1987]. As a result, intensity values of pixels for the same tissue can vary within a slice, from slice to slice, or from one volume to another. Thus, approaches to segmentation such as thresholding and statistical classification based on pixel intensity alone are not reliable. The characteristic thresholds or intensities used to segment a particular tissue in one area of an image will not necessarily apply to other occurrences of the tissue type within the same slice or volume. For this reason, segmentation methods which are solely data-driven are limited by the quality of image data on which they are applied.

2. **Outlines of regions in MR images can be of varying sharpness or contrast** [Kapouleas, 1990b]. Depending on voxel size and on the angle at which a slice is acquired, a given voxel may contain a composite of tissue types. This phenomenon, known as the *partial volume effect*, makes the use

of edge detection methods difficult and unreliable. As the outline of an organ can contain edges of varying sharpness, its automatic detection may not be possible without the inclusion of edges from adjacent structures. Fan et al. [1987] note that when a physician manually segments imaged structures, he will base the hand-drawn outlines not only on statistical information available in the image, but also on *a priori* knowledge of anatomy and medical experience. This is the premise behind model-based approaches to segmentation which make use of knowledge such as the size and shape or relative location of anatomical structures in their classification. These methods generally use low-level operators to create regions which are then matched to anatomical models. However, the high-level decision component of these strategies is not always able to correct the mistakes make in the low-level segmentation. Difficulties in finding proper data structures to represent objects and models, and in controlling program flow must be overcome [Stansfield, 1986]. Knowledge-based analysis of image data has produced promising preliminary results; however, it is still at an early stage. Novel ideas for describing and representing natural structures in conjunction with artificial intelligence tools are needed for generally applicable methods of analysis [Kubler and Gerig, 1990].

1.3 Objectives and Scope of Thesis

Use of A Priori Knowledge. The segmentation method developed in this thesis uses a model-based approach to classification which incorporates *a priori* knowledge of average brain anatomy. The idea behind the development of the model was to use information about a voxel's location in 3D space as a heuristic in pred%%[PrinterError: (tissue type. Prior to segmentation, a head MR volume is affine transformed into a 3D proportional grid system which acts as a standardized 'brain space'. The space, referred to hereafter as *Talairach space*, was defined by Talairach, Szikla, Tournoux,

Prossalentis, Bordas-Ferrer, Covello, Jacob, Mempel, Buser, and Bancaud [1967] in their work on brain atlases. (A later version is found in Talairach and Tournoux [1988]). Talairach space is routinely used in neuroscience work involving the comparison of human brains as it adjusts for variation in individual cerebral size and proportion. The transformation into the Talairach coordinate system uses translation, rotation, and scaling so that brains within the space appear geometrically equivalent (ignoring non-linear morphological differences). If coordinate (x, y, z) of brain A is a voxel of grey matter, for example, then the same coordinate in brain B is likely to be of grey matter as well. A tissue probability model was constructed based on MR brain data obtained from a group of healthy volunteers, giving for each voxel in Talairach space, the probabilities of that voxel belonging to each of the possible tissue types. The model is used to guide brain MR image segmentation, serving as a 3D probabilistic template for gross tissue structures. Thus, the tissue classifier uses geometric information about the location of voxels, as well as statistical information based on grey scale values in the given image. This thesis explores the extent to which a prior voxel-based tissue probability model can overcome classification errors inherent in solely data-driven procedures due to the effects of noise, partial volume, field inhomogeneities, and other artifacts.

The tissue probability model or *canonical mask* should also be useful in segmenting lesions or tumors characterized by their location of occurrence. In this thesis, it is used in the detection of MS lesions of the brain. Ninety to ninety-five percent of MS lesions occur in white matter tissue [Maravilla, 1988]. Thus, knowledge of whether a voxel is in an area of the brain having a high probability of white matter can be used in detecting multiple sclerosis lesions.

Statistical vs. Symbolic Learning Approach. In selecting a classification algorithm, various learning techniques were studied. Empirical learning techniques can be divided roughly into three categories: statistical [Duda and Hart, 1973], neural networks, and symbolic learning techniques such as induction of decision

trees[3] or production rules.

Statistical techniques, neural networks and symbolic learning are approaches to supervised learning that can be used to classify a sample pattern into a specific class. Methods from these categories differ in the ways in which class descriptions are represented. A decision tree or rule-based approach has the advantage of offering a modularized, clearly explained format for a decision, and is compatible with a human's reasoning procedures and expert system knowledge bases [Weiss and Kapouleas, 1989]. Mooney, Shavlik, Towell, and Gove [1989] stress the issue of human interpretability by stating that symbolic learning can produce interpretable rules, while mathematical formulae or weights determined by neural networks are harder to grasp. Symbolic learning procedures, however, tend to be more complex, often requiring manipulation of a knowledge base.

Bayesian classifiers employ a statistical approach to classification which assumes that the classification problem can be posed in probabilistic terms and that all relevant probability values are known [Duda and Hart, 1973]. If the assumptions hold true, then Bayesian classifiers have the minimum error rate in comparison with all other classification algorithms. Relevant probability values include the *a priori* probabilities or distribution of the classes to be recognized. Such data is rarely complete in real-world problems. Thus, the assumption that classes are equally probable is common. Class conditional independence of features is assumed in order to reduce computation (the presence or absence of each feature is assumed to be independent of the presence or absence of others). *Proton density* is a magnetic resonance tissue-specific parameter commonly used as a feature in statistical pattern recognition for tissue classification. This parameter reflects the number of magnetized protons within a given voxel. Ozkan et al. [1990] note that the distribution of proton density values is not normal and thus violates the multivariate normal distribution assumption made for Bayesian classification. Hence, although

[3]A **decision tree** is a recursive structure for representing classification rules.

in theory, Bayesian classifiers have the minimum error rate, in practice this is often untrue due to the inaccuracy of assumptions made for its use.

In view of the advantage in representation of classification rules obtained by symbolic learning techniques, and of the theoretical superiority of Bayesian classification, it is interesting to compare the performance of a Bayesian classifier with that of a symbolic learning algorithm, applied to the tissue classification of MR images of the brain. ID3 [Quinlan, 1979, 1983, 1986b], a decision tree algorithm, was chosen as it is a popular and relatively simple symbolic learning algorithm and has been extensively tested on problems ranging from chess end games [Quinlan, 1983] to object recognition [Shepherd, 1983]. In addition, ID3 has been augmented with techniques for handling noisy data [Quinlan, 1986a; Quinlan, 1987c; Niblett and Bratko, 1986]. In separate studies performed by Weiss and Kapouleas [1989] and Mooney et al. [1989] on real-world data, ID3 was found to perform at least as well as statistical classifiers. Weiss and Kapouleas note that numerous experiments by Breiman, Friedman, Olshen, and Stone [1984] showed that, in most cases, a decision tree classifier is superior to alternative statistical classification techniques. If both the statistical and symbolic classifiers perform at the same level, then the latter may be preferred for its greater human interpretability.

The objectives of this thesis are the following:

1. to develop a tool for the segmentation of magnetic resonance images of the brain at the gross tissue-type level. In particular, the tool is to allow for the automated detection of multiple sclerosis lesions.

2. to evaluate the effectiveness of incorporating *a priori* knowledge of brain anatomy in tissue classification. A model of *a priori* tissue probabilities per voxel in a standard 3D 'brain space' is to be used along with statistical information based on image grey level values.

3. to compare the performance of statistical Bayesian and minimum distance

classifiers with that of the symbolic decision tree classifier, ID3, trained for MR image segmentation. The minimum distance classifier is included as it is often cited as a basis of comparison with other classification methods.

The desirable characteristics of the segmentation tool are:

1. The segmentation algorithm employed should be robust and accurate: Segmentation accuracy should not be dependent on optimal image quality.

2. Minimal or no user interaction is desired: This is especially important if computerized analysis tools are to become an accepted component of the medical environment where physicians and clinicians have limited amounts of time to spend in front of work stations.

3. Users should be able to modify the program's proposed segmentations or (*tissue maps*) in case of disagreement. In the ideal case, the system would adopt the user-provided corrections to improve its accuracy on subsequent trials. At the very least, users should be able to manually edit output segmentations if necessary.

4. The segmentation process should be efficient on standard work stations.

5. The segmentation process should make use of the 3D information inherent within volumes of MR data.

6. Knowledge of anatomy should be employed: It seems logical for the segmentation process to use *a priori* knowledge of anatomy during the classification process as this is what human experts do. Human experts require prior knowledge of anatomy and medical experience when performing manual segmentation.

The remainder of this thesis presents the steps taken in achieving the stated objectives. Chapter 2 describes the basic principles of magnetic resonance imaging.

Chapter 3 presents a review of segmentation techniques, with particular attention given to methods for classifying tissues in magnetic resonance images of the brain. Chapter 4 presents a review of machine learning techniques. The review was conducted to aid in selecting a symbolic learning algorithm for the segmentation tool. Readers familiar with the review material may wish to proceed directly to Chapter 5 which describes the development of the tissue probability model and the implementation of the Bayesian, minimum distance, and decision tree classifiers. Chapter 6 presents obtained results. The final chapter draws conclusions from the use of the model, and of the statistical and symbolic classifiers, followed by a discussion on future related work.

Chapter 2

Magnetic Resonance Imaging

This chapter presents a brief description of the principles behind magnetic resonance imaging.

2.1 Basic Principles

Magnetic resonance imaging is based on an inherent property of nuclear particles. Nuclei containing an odd number of protons, an odd number of neutrons, or both, have a spin. Every charged particle in motion has a magnetic field associated with it. Thus, as nuclei are charged particles, they exhibit small magnetic fields. Some nuclei also display a vibration effect or *precession.*

Every object that can be made to spin or 'precess' will do so more strongly under the influence of a force applied at the same frequency as the natural resonant frequency of the object. An example of this *resonance* is illustrated with a pair of identical tuning forks. If the first tuning fork is struck, it will start to vibrate. The vibration energy is transferred to the second tuning fork, causing it to vibrate as well. A similar resonance effect occurs when atomic nuclei are subjected to electromagnetic waves at their own vibration frequency [Philips, 1984].

Hydrogen is the body's most abundant nuclei which obeys this principle of mag-

netization. If hydrogen nuclei are placed in a magnetic field, a number of them will tend to line up in the field's direction. Since the protons of the nuclei are in motion, they will precess around the direction of the field like a spinning top. (This direction is referred to as the 'equilibrium direction').

In MRI, objects to be imaged are placed in a strong magnetic field of 0.5-1.5 Tesla (1 Tesla = 10 kgauss). A group of the object's hydrogen nuclei will align about the direction of the field as described above (Figure 2.1b). A radiofrequency (RF) pulse is then applied perpendicular to the direction of the magnetic field (Figure 2.1c). This stimulates the nuclei, causing them to tilt away from the aligned equilibrium direction and precess around the direction of the magnetic field. The frequency of this precession is equal to the frequency at which the RF field is applied and is dependent on the strength of the applied magnetic field and on the type of nuclei. Typically, the strength and duration of the applied radiofrequency signal is made to tilt the nuclei 90 or 180 degrees from the aligned equilibrium direction. During this displacement, the nuclei absorb energy from the radiofrequency pulse. Once the RF signal is removed, the nuclei gradually realign themselves in the direction of the original magnetic field while emitting their own radio or resonance signal (Figure 2.1d). These signals can be detected by an antenna and analyzed by a computer to create a grey scale image of the object.

2.2 MR Tissue-Specific Parameters

Data for MR images is generated by the electromagnetic signals originating from the imaged objects. The number of protons at a given image pixel is referred to as *proton density* and is dependent on the tissue type at the pixel. The greater the proton density, the larger the emitted signals with be, thus providing a means of distinguishing various tissues within an object.

Image contrast is also influenced by two other tissue-specific MR parameters

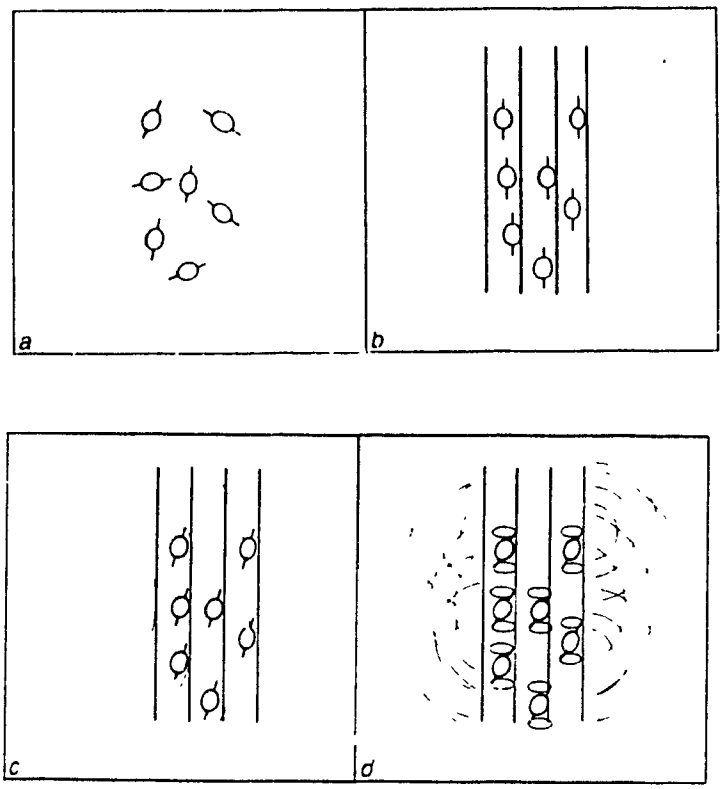


Figure 2.1: Sequence of events in MR observation.
a) Randomly oriented nuclei.
b) Nuclei align in direction of applied magnetic field (equilibrium direction).
c) Radiofrequency pulse is applied, causing the axis of rotation of the nuclei to tip.
d) The nuclei precess about the direction of the applied radiofrequency field and emit their own radiofrequency signal.
Reproduced from *Introduction to MR Imaging*, Philips Medical Systems, The Netherlands, 1984.

known as *T1 (longitudinal relaxation time)* and *T2 (transverse relaxation time).* T1 describes the rate at which a displaced axis of rotation returns to equilibrium. T2 is related to the time required for the emitted resonance signal to decay. White matter of the brain, for example, has a T1 of 690 ms and a T2 of 107 ms. Grey matter of the brain has a T1 of 825 ms and a T2 of 110 ms [Peters, 1988]. These differences are responsible for the contrast between white and grey matter in typical MR images.

The rate at which the radiofrequency pulse is applied is known as *TR (repetition time).* The resonance signals can be measured at various time points within a TR interval. This is analogous to taking different 'snapshots' of the emitted signals. Each snapshot or *echo* gives an image whose brightness at any point is related to the signal emitted by nuclei at that position at the the time of measurement. The time interval between the application of the radiofrequency pulse and the measurement of the resonance signals (or the time at which the snapshot is taken) is referred to as *TE (echo delay time).* If the resonance signals are measured more than once within a TR interval (i.e. more than one snapshot is taken), this is known as *multi-echo acquisition.* The repetition time and echo delay times are generally varied so as to best demonstrate a particular pathology in the image, or the contrast between grey and white matter tissues. The variance of the repetition and echo delay times is done to make use of the characteristic proton density, T1, and T2 values of the individual tissues.

Magnetic resonance images can be acquired with various imaging sequences, each of which generates a signal whose strength depends on the MR tissue-specific parameters mentioned above. Two imaging sequences used in this thesis are the Spin Echo (SE) and Fast Field Echo imaging (FFE) sequences. In the spin echo imaging sequence, tissues with a high proton density and a long T2 relaxation time appear bright. This sequence is considered to be the standard, although images can be acquired at a considerably faster rate with the fast field echo sequence.

Given an image, the proton density, T1, and T2 values at each pixel can be computed, hence providing tissue-specific information useful for classification. The calculation, however, requires large amounts of computation. Rather than calculating the pure proton density, T1, and T2 values, in clinical practice it is easier to acquire a set of *proton density-weighted* ('proton density-like'), *T1-weighted*, and *T2-weighted* images as multi-echoes. Each pixel can then be represented by a three-dimensional vector whose components are the intensity values in the proton density-weighted, T1-weighted, and T2-weighted images respectively. Classification schemes which make use of this vector or combinations of proton density, T1, and T2 values are examples of *multispectral* classification.

2.3 Concluding Remarks

The basic principles of magnetic resonance imaging have been described. The radio frequency energy applied during image acquisition is non-ionizing, and thus does not expose the patient to risks associated with X-ray imagery. Acquisition parameters TR and TE can be varied to enhance the contrast between individual tissues. Proton density, T1, and T2 are tissue-specific parameters available per pixel and can be used as features in classification.

Chapter 3

Review of Image Segmentation Techniques

This chapter presents a review of image segmentation techniques. Particular attention is given to methods for the segmentation of magnetic resonance images of the head and for the detection of multiple sclerosis brain lesions. The techniques studied include thresholding, edge detection, region growing, and statistical classification. Knowledge-based strategies, typically employing models of anatomy, are also discussed.

3.1 Segmentation Techniques

3.1.1 Thresholding

In segmentation by intensity thresholding, ranges of grey level values are assigned to characterize each tissue type. If a pixel's intensity value falls within the range for a specific class, then the pixel is assigned to that class. This procedure assumes that tissue classes are definable by non-overlapping ranges of intensities and that the intensity for each tissue class is the same at all points in the image.

Lim and Pfefferbaum [1989] describe a thresholding system for the segmentation of MR images of the head. They note that the major obstacle to any thresholding

approach to MRI segmentation is the artifact of radiofrequency (RF) inhomogeneity which causes some portions of an image to contain higher or lower intensities than others for equivalent tissues.

Kapouleas [1990b] describes a thresholding method employed to detect the outline of the brain and of multiple sclerosis lesions in MR images. A 3x3 gradient operator for edge detection is applied to a 256x256 image which is then divided into 16x16 subimages. A local threshold for each subimage is then calculated by the averaging of edge pixels detected by the gradient operator. The method works well and has the advantage of being automatic; no user interaction is required. The thresholding, however, is not successful on its own. Heuristics are necessary to correct errors. A geometric model or 'template' of the brain is used to eliminate *false positive* lesions (non-MS voxels which are mis-classified as MS) according to their location. Furthermore, postprocessing is required to join erroneously disconnected regions.

Thus, the thresholding approach to segmentation is unreliable due to image artifacts. The use of *a priori* knowledge of anatomy is suggested as a means for improving thresholding results.

3.1.2 Edge Detection and Region Growing

Edge detection and region growing are image analysis techniques that have been applied to the segmentation of MR images. In edge detection, gradient operators are employed to distinguish edges by the high rate of change in grey level value between neighboring pixels. Bomans, Hohne, Tiede, and Riemer [1990] presented a 3D extension of the Marr-Hildreth operator (a Laplacian of a Gaussian) which they used to segment structures of the brain. Kennedy, Filipek, and Caviness [1989] proposed a modification to the Sobel operator (estimation of partial derivatives) which was also applied to the classification of brain tissues. In region growing, a *seed* is planted (normally by a user) by marking a pixel or group of pixels within

an area to be segmented. An iterative algorithm checks for pixels that are adjacent to the seed and of an intensity similar to it. Pixels satisfying a similarity condition are added to the seed, allowing it to grow as a region. Jernigan et al. [1990] used region growing to locate regions of brain tissue within an image. Fan et al. [1987] presented an edge-limited region growing scheme where regions are grown according to a similarity condition and stopped at edge-labeled pixels. This was tested on the segmentation of images of the prostate and bladder. To the untrained observer, Fan et al.'s [1987] segmentation results appeared correct. However, the computer-generated segmentations differed from hand-drawn outlines made by a physician who accounted for the presence of a U-shaped sling of muscle lying between the two structures. The muscle had appeared blurred in the image because of partial volume effects, and so was improperly segmented by the algorithm.

Edge detection and region growing schemes operate on the assumption that gradient-based edges separate structures of interest. This may not always hold true. These methods of segmentation are highly constrained by image noise, artifacts, and partial volumes. The edges of an object can vary in contrast, making the automatic detection of its outline difficult without the inclusion of edges or areas of adjacent structures. Furthermore, as noted by Fan et al. [1987], physicians do not base their manually drawn outlines solely on visible sharp edges. Knowledge of anatomy is also employed. They suggest that systems for medical image segmentation utilize *a priori* knowledge of anatomy.

3.1.3 Statistical Classification

Statistical classification techniques, such as Bayesian classifiers, linear regression, neural networks, and cluster analysis, have been applied to image segmentation. Each pixel of an image is represented by a set of features or *feature vector*. Features may be based on local mean intensity, variance from the local mean, as well as on spatial, morphological, and other textural measures [Meinzer, Engelmann,

Schleppelmann, and Schafer, 1990]. Statistical classifiers, when applied to tissue classification, operate on the assumption that each tissue type has a characteristic *signature* or partition in feature space. A statistical classifier receives, as input, a training set of user-selected samples from each class. These are examined and used to define classification rules describing each tissue type.

As described in section 2.2, multi-echos or 'snapshots' of an object can be acquired so as to reflect the tissue-specific parameters of proton density, T1, and T2. The feature vector of a pixel can consist of its intensity value in each of the echo images. Statistical approaches to segmentation which make use of this vector are examples of *multispectral* classification. The multispectral image classification technology was originally developed for and by NASA with the intention of processing multispectral satellite images [Vannier, Butterfield, Rickman, Jordan, Murphy, and Biondetti, 1987]. Algorithms employed in previous work in multispectral classification for the segmentation of MR images include a supervised Bayesian classifier [Amamoto, Kasturi, and Mamourian, 1990; Gerig et al., 1991; Hyman, Kurland, Levy, and Shoop, 1989] and multiple linear regression [Jernigan et al., 1989]. Gerig et al. [1991] also developed a clustering technique. The method of classification described in Cline, Lorensen, Kikinis, and Jolesz [1990] works on the assumption that each tissue has a bivariate normal probability distribution, estimated from training samples of dual-echo images.

Experiments in MR segmentation with neural networks include Ozkan et al.'s [1990] and Dawant, Margolin, Ozkan, Aramata, and Kawamura's [1990] use of the backpropagation algorithm to classify tissues of the brain. Katz and Merickel [1989] applied the same algorithm to the segmentation of the aorta from MR images of the heart. Raff and Newman [1990] explored the use of autoassociative memory and a novelty filter for the detection of multiple sclerosis lesions of the brain. In autoassociative memories, a pattern is retrieved by an incomplete version of itself. The novelty filter stores the feature vectors of training samples and is used to extract characteristics of the memorized patterns.

The success of statistical classification schemes based on features derived from image grey scale values depends largely on image quality and on the selection of adequate features. Gerig et al. [1990] note that the statistical classification approach to MR segmentation requires optimal image data, acquired so that each tissue forms a distinct partition in feature space. A drawback of statistical classification is its requirement of a training set. This can involve considerable user interaction. In supervised learning, where a teacher assigns a class label to each training sample, the subjective bias of the teacher is introduced. Fully automatic methods are necessary for accurate reproducible results.

3.1.4 Knowledge-Based Segmentation

Difficulties in data-driven approaches to segmentation due to noise, blurred edges, partial volume, and artifactual variations in intensities within tissue types have led to the increased development of knowledge-based systems. When applied to medical image segmentation, these systems are typically rule-based, encoding knowledge of anatomy, of image acquisition parameters, and of the nature of the imaging technique employed (MRI, X-ray, computed tomography, etc.). Anatomical information can be modeled symbolically (describing the properties and relationships of individual structures) or geometrically (serving as masks or templates of anatomy).

Symbolic Models. Stansfield [1986] developed a rule-based expert system for the segmentation of coronary vessels in images obtained by digital subtraction angiography (DSA). This imaging technique requires the injection of a contrast material into the blood flow. The subtraction of X-ray images acquired before and during the flow of the contrast media yields an image of the vasculature. The expert system employs edge detection and region growing to create segments which are grouped and classified with the use of production (IF-THEN) rules. These rules encode low-level knowledge of segmentation and shape analysis, as well as high-level knowledge of cardiac anatomy and physiology. Stansfield comments that

without the use of *a priori* knowledge, the system's search procedures succumb to a combinatorial explosion. A noted problem was the incorrect labeling of noise structures as vessels.

Use of a symbolic anatomical model of the brain is described by Sokolowska and Newell [1986] in their system for the segmentation of computed tomography scans. The system employs a bottom-up approach, grouping pixels in blocks to form segments, which are in turn grouped and classified as anatomical regions. The classification is guided by a 'jigsaw puzzle strategy' that attempts to match regions to the anatomical model. The model lists the structures that may be present in a transverse [1] brain image, describing their characteristic intensity-based properties, as well as probable spatial relationships between structures.

An anatomical model of the brain is also used in an expert system devised by Vernazza, Serpico, and Dellepiane [1987] for the classification of organs in three types of slices of head MR data (a transverse slice through the eyes, a transverse slice through the ventricles, and a sagittal [2] slice). Elementary regions, created by region growing, are classified with the use of production rules that define search heuristics, and a semantic network [3] describing the organs (such as brain, eyes, nose, skin, and bone) possible in each slice type. As in Sokolowska and Newell's [1986] model, organs are described in terms of shape, densiometry and spatial interrelationships.

Menhardt's [1988a, 1988b] approach to segmentation of MR images uses fuzzy logic where uncertainty values ranging from 0 to 1 are assigned to pixels to represent their degree of membership in each class. The characteristics of tissues in T1-weighted and T2-weighted images are used to define fuzzy logic operators for the identification of brain, skull, skin, and cerebrospinal fluid. Menhardt [1988a]

---

[1] A **transverse** plane is any plane which divides the brain into top and bottom.

[2] A **sagittal** plane is any plane which divides the brain into left and right pieces.

[3] A **semantic network** is a graph that uses nodes to represent objects, and arcs to represent the relationships between them.

uses a hierarchical description of anatomy and of the signs and symptoms of neurological diseases, as well as grey scale intensity values, to evaluate hypotheses for the presence of various pathologies of the brain.

Geometric Models. A geometric model can be used to provide prior information regarding the positions of anatomical regions. As mentioned earlier, Kapouleas [1990a, 1990b] employs a geometric model of the brain in the segmentation of multiple sclerosis lesions. The model is deformable to comply with the variations in brain shape amongst individuals. The top-down system employs a thresholding technique to suggest possible lesions. The modeling method, based on bicubic B-spline surfaces, uses surfaces that are easy to identify (such as the brain's outer surface, the ventricles, and the interhemispheric fissure [4]) as landmarks when fitting the geometric model to a given slice. The model is used to approximately locate the area of white matter adjacent to the ventricles ('periventricular white matter'), where the majority of MS lesions can occur. Proposed lesions which are not within this area are rejected. Another example of the use of geometric models for image segmentation can be found in Dann, Hoford, Kovacic, Reivich, and Bajesy [1989] in their work on the elastic matching between an idealized anatomic atlas and a given data slice. Ayache, Boissonnat, Brunet, Cohen, Chieze, Geiger, Monga, Rocchisani, and Sander [1989] also developed a method to fit deformable models of structures to detected edges.

The intention of this review of knowledge-based techniques was to describe ways in which *a priori* knowledge has been used for the segmentation of medical images. Symbolic or geometric models of anatomy are typically employed. Difficulties to overcome in the use of knowledge-based segmentation include the following:

- Appropriate knowledge representation schemes are needed [Kubler and Gerig, 1990; Stansfield, 1986]. Knowledge representations are often incomplete. In regards to the segmentation of brain MR data, models are necessary for de-

[4]The **interhemispheric fissure** is the space separating the left and right hemispheres.

scribing the entire brain, and not just particular types of slices.

- Incorrect labeling can occur due to noise and errors in low-level segmentation. Model-based systems generally use segmentation techniques such as edge detection and region growning to create segments for anatomical labeling. Errors which occur in the segments due to noise, partial volume, and image artifacts are difficult to overcome, even with the use of a model [Kapouleas, 1990b].

- Complex forms of control (involving heuristics) are required when searching to select plausible hypotheses and avoid exhausting available memory resources [Vernazza et al., 1987; Stansfield, 1986]

- Knowledge-based systems depend heavily on heuristics, making their application to other domains difficult.

Nonetheless, the use of *a priori* knowledge to guide the segmentation process is promising as it resembles the human processes involved in manual segmentation.

3.2 Concluding Remarks

This chapter has described data-driven approaches to medical image segmentation as well as knowledge-based techniques. Data-driven methods are highly susceptible to noise, artifacts, and variations in tissue intensities across slices and image volumes. Knowledge-based systems tend to exhibit greater robustness due to their assimilation of symbolic or geometric models of anatomy. Problems exist, however, in finding appropriate schemes for representing knowledge, and in controlling search procedures.

The review was conducted to serve as a background in the design of an image segmentation tool for the detection of MS lesions. The following decisions were made regarding the design of the system:

- Methods which involve low-level segmentation such as edge detection or region growing were not be employed as they depend heavily on the quality of image data.

- Multispectral data was to be used, as it provides more tissue-specific information than a single image.

- The Bayesian and minimum distance statistical classifiers were to be implemented. These are used, however, with *a priori* knowledge of gross neuroanatomy. A geometric model of brain tissue probability was to be constructed based on a group of healthy volunteers. The model must provide *a priori* probabilities of grey matter, white matter, and cerebrospinal fluid per voxel in a standardized 3D 'brain space'. The model should aid in the detection of MS lesions, which occur primarily in highly probable white matter areas. The tissue model is to cover the entire brain and thus should be useful in the segmentation of volumes of brain MR image data.

Statistical classification has been applied extensively to the segmentation of MR images. This approach has been criticized for its poor human interpretability, and because of the assumption of conditional class independence generally made for the use of Bayesian classification. A goal of this thesis is to evaluate the extent to which a symbolic learning approach may overcome these limitations. The next chapter presents a review of machine learning techniques, with the intent of aiding in the selection of a learning algorithm to be used in combination with the model of brain anatomy for the detection of MS lesions. The performance of the symbolic algorithm chosen is to be compared with that of the statistical classifiers.

Chapter 4

Review of Machine Learning

This chapter presents a review of research in machine learning. The review was conducted to provide a background for the selection of a symbolic learning algorithm for the development of an MR image segmentation tool. The chapter starts with a discussion of the rationale for the study of machine learning, followed by a description of its basic terminology. A historical review of activity in the field is presented. The remainder of the chapter details past work in machine learning, arranged in the form of a taxonomy. The proposed taxonomy separates the development of practical automated learning algorithms from their theoretical analysis.

4.1 Machine Learning

The field of artificial intelligence has grown out of the desire to automate intelligence in machines. *Intelligence* can be defined to include the ability to reason, to acquire and apply knowledge, and to perceive and manipulate objects in the physical world [Winston, 1984]. Artificial intelligence, or AI, is a domain containing a broad range of sciences. It interests computer scientists and engineers, working to find ways to make computers more useful, and psychologists, biologists and philosophers, seeking to understand the principles of intelligence. Since AI first gained recognition as a

discipline in the mid 1950s, machine learning has been a key area of research. Such prominent attention is highly conceivable, as any attempt to understand and automate intelligence must embody an understanding of learning itself [Quinlan, 1986b].

4.1.1 Rationale for the Study of Machine Learning

Two reasons for the study of machine learning are:

- to reach beyond the limits of traditional computer science by providing computers with the ability of learn. To enable a computer to perform not only tasks which it has been programmed to do, but new tasks which it has learned to do.
- to gain understanding of the phenomenon of learning by simulating the learning process. Simon [1983] notes that once learning is understood, this knowledge may be used to help make humans more efficient at their work.

Machine learning plays an important role in knowledge acquisition for the design of expert systems. Automated learning techniques would enable a system to develop decision rules from examples of experts' decisions and through the automated analysis of facts in a database [Michalski, 1986]. Learning as discovery or 'the finding of new things' [Simon, 1983] may uncover concepts not yet known to exist. The design of successful learning techniques should involve the input of computer scientists in artificial intelligence and software engineering, researchers in neural networks as well as in the field of cognitive neuroscience.

4.1.2 Machine Learning Terminology

Definitions of Learning. Simon [1983] proposes the following definition of learning:

Learning denotes changes in a system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the same population more efficiently and more effectively the next time.

This definition includes *speed-up learning* (the ability to perform a task faster than one had done previously). It comprises learning as an acquisition of new methods and knowledge, or as an improvement in existing methods and knowledge to make them faster, more accurate, and more robust [Cohen and Feigenbaum, 1982].

Dietterich [1990] presents a definition of learning which involves *induction*, the reaching of a conclusion by reasoning from individual facts. An agent (being a person or a program) learns when it is told a fact, F, which it did not know, or when it can logically infer F from its existing knowledge. To illustrate, Dietterich provides an example of an agent learning the game of poker. The agent is told that a hand containing three Jacks is superior to a hand containing only two Queens. The agent also knows that a hand containing three Tens is superior to a hand containing two Eights. Learning occurs when the agent is able to infer that any hand containing three cards of rank R_1 is superior to any hand containing at most two cards of rank R_2. According to this definition, learning is not said to result when a system finds a faster way to infer a fact that it already knew. Thus, unlike Simon's definition of learning, Dietterich's definition does not include speed-up learning.

A definition commonly adopted by researchers in the field of expert systems is that learning is the acquisition of explicit knowledge, often represented in the form of rules. An explicit representation is important so that the knowledge can be easily verified, modified, and explained [Cohen and Feigenbaum, 1982]. Learning as the acquisition of explicit rules from examples can include rules for efficiency improvement.

An alternative definition of learning is that it is skill acquisition. This reflects the accepted view that performance of a task improves through practice. Cognitive

scientists and psychologists are particularly interested in understanding the ways in which knowledge is acquired in order to perform skillfully.

Learning can also be described as discovery. This definition views learning as theory and hypothesis formulation and *inductive inference* (the process of inferring general rules from specific examples). This learning perspective centers on understanding how scientists form hypotheses to construct theories to explain complex phenomena.

Laird [1990] avoids defining learning by instead describing the nature of a learning program. Most learning systems can be thought of as search programs. As searching proceeds, changes in state occur which represent learning if they exhibit an improvement in computation or prediction. He describes learning programs ideally as search procedures whose space requirements increase slowly in comparison to the size of input data, and which exhibit gradual improvements in hypotheses. His view is similar to that of Simon.

We shall accept the view that learning is any process by which a system acquires explicit knowledge and/or improves its performance. A learning system can receive input from its environment (usually a teacher) and a knowledge base. A system learning how to recognize tumors visible in medical images, for example, may access a knowledge base containing explicit information about the types of different tumors and their characteristic size, shape and location. The environment supplies information to the system in the form of examples of each tumor type. As the system learns, it may update or modify the contents of its knowledge base. The learning process typically involves searching through a large space of possible hypotheses.

Basic Terminology. Learning as an acquisition of knowledge can be divided into two phases: training and recall. During the former phase, the learning program is presented with a training set consisting of examples of each class or *concept* to be learned. The learning program must derive a function or classification rule which will allow it to accurately classify or label input examples. An example is repre-

sented by a set of features or attributes which describe it (eg. size, shape, color). The set of features for an example make up its feature vector. The classification rules can be expressed in several ways (e.g. as a set of production rules, a decision tree, an artificial neural network, or a logical definition). Since classification rules represent concepts, this form of learning is also referred to as *concept learning* or *concept acquisition.* During the recall phase, the system uses the learned rules to classify new examples.

Learning can be either supervised or unsupervised. In supervised learning, a teacher labels each training example with the class to which it belongs. In unsupervised learning, the examples presented during the training phase have not been class labeled. Discovery is a form of unsupervised learning designed to investigate domains in an unaided, exploratory way to discover 'unknown' concepts and relationships between them. Clustering is another form of unsupervised learning in which the learning program examines the examples and separates them into groups or 'clusters', where members of a cluster appear similar. The task is then to find a classification rule to define each cluster. Clustering is also called *learning from observation* or *concept formation.*

The following sections present an account of the history of machine learning, followed by a review of individual works in the field.

4.2 Historical Background

In 1956, John McCarthy, coiner of the term 'artificial intelligence', helped organize a conference to bring together scientists interested in this new field. Since then, machine learning has been a central area of research in artificial intelligence. The history of machine learning in AI can be divided into three phases of activity. These are i) exploration (during the 1950s and 1960s), ii) development of practical algorithms (1970s), and iii) increase of research directions (1980s and onwards)

[Shavlik and Dietterich, 1990].

The most recognized landmark in early machine learning history is Rosenblatt's [1958] perceptron algorithm. The perceptron, a system of randomly connected linear threshold units, can learn to associate specific responses to specific stimuli. It characterizes machine learning work of the exploration phase, kindled by psychologists and neurobiologists who sought to develop and test computational analogues of neurons.

Similar early work in machine learning includes that of Feigenbaum [1961], whose Elementary Perceiver and Memorizer Program (EPAM) was developed to simulate the behavior of subjects in experiments involving the rote learning of nonsense syllables.

Amongst the earliest contributions to speed-up learning is Samuel's [1959] work in evaluating static functions for game-tree strategies in checkers by memorizing previous board positions.

Minsky and Papert [1969] unveiled the drawbacks of the perceptron learning algorithm. Single-layer perceptrons are incapable of learning concepts which are not linearly separable. This finding, along with limitations in hardware, dampened enthusiasm for research in computational neuronal modeling for nearly 20 years.

The development of practical learning algorithms characterizes the second phase of research in machine learning, occurring during the 1970s. Many researchers, in recognizing that learning is a difficult and complex process, shifted their efforts towards the easier problem of learning single concepts. Workers adopted the view that the learning of complex, high-level concepts can not be accomplished without providing the learning system with background or domain knowledge of the application to be learned. An influential publication which prompted this change in view was Winston's [1970] thesis on blocks-world learning. Other work during this phase includes Buchanan and Mitchell's [1978] METADENDRAL system for the learning of mass-spectroscopy prediction rules, and Michalski and Chilausky's

[1980] AQ11 program. AQ11 has successfully been applied to the diagnosis of soybean diseases, producing results superior to that of human experts. An example of early work in discovery is Lenat's [1977] AM program. AM performs a heuristic search on a knowledge base of mathematical concepts to discover concepts in elementary mathematics and set theory.

The latest stage of activity in machine learning, from the 1980s onwards, has seen an increase in research directions. Efforts have centered on the use of machine learning as a tool for knowledge acquisition. Research directions include the further development of symbolic, clustering, discovery, and explanation-based learning strategies, a resurgence of neural networks, and advances in the theoretical analysis of learning. Work from each of these areas is described in the following section.

4.3 Taxonomic Review of Machine Learning

This section presents a taxonomy of machine learning as summarized in Figure 4.1. The proposed classification separates the development of practical machine learning techniques from their theoretical analysis.

Machine Learning Techniques. Machine learning techniques can be classified according to the degree of inference [1] they involve [Cohen and Feigenbaum, 1982; Michalski, Carbonell, and Mitchell, 1983, 1986]. A learning system uses inference to transform information provided by the environment (knowledge from an external source) into some usable representation which in turn may be added to a knowledge base. The degree of inference performed reflects the *level* of information provided by the environment. 'High-level' information is very abstract, requiring a great deal of transformation, whereas 'low-level' information requires little or no transformation. For example [adapted from Roch, Pun, Hochstrasser, and Pelle-

[1] **Inference** is the deriving of a conclusion from induction or deduction - a conclusion arrived at in logic.

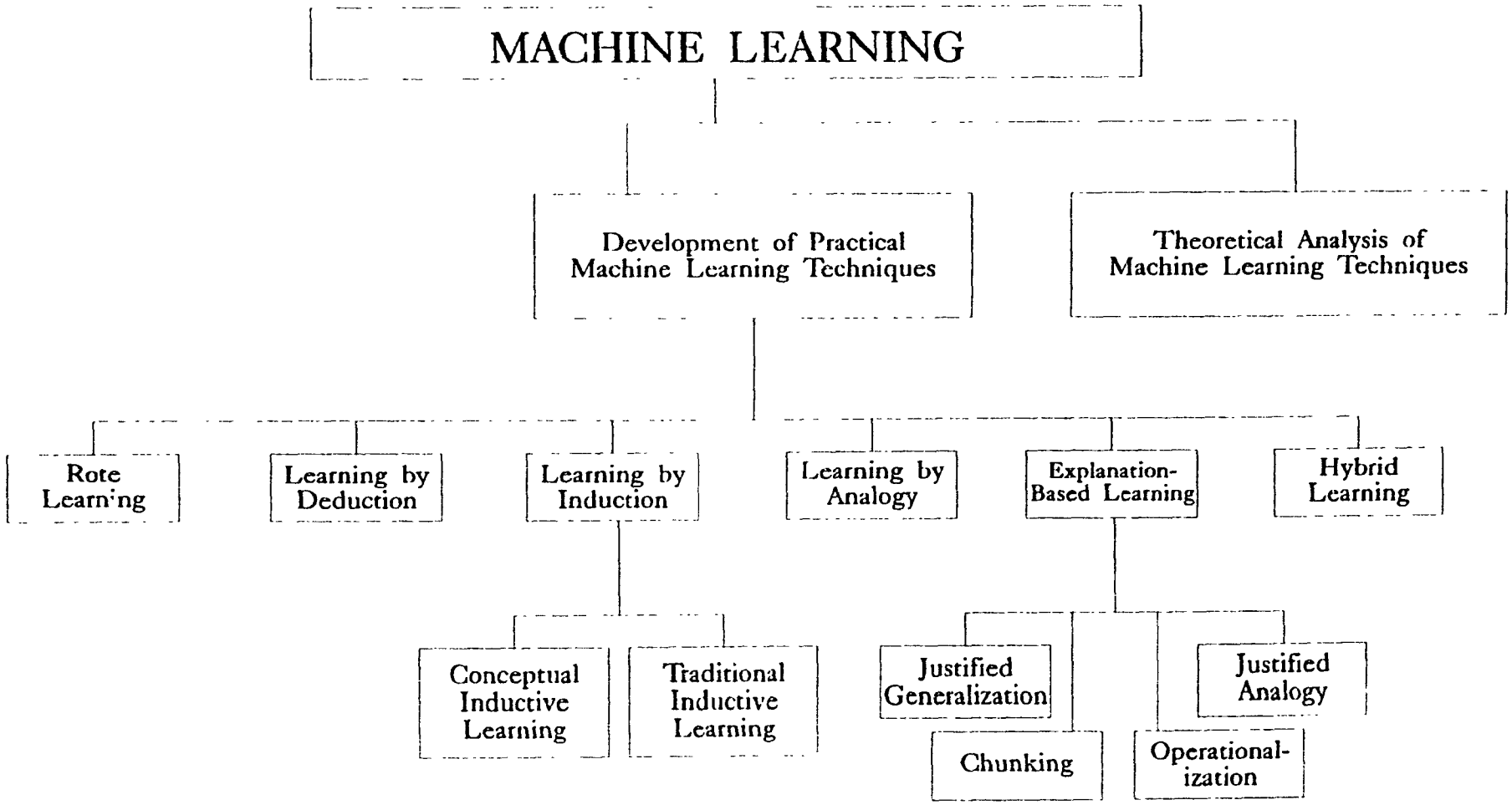


Figure 4.1 Taxonomy of Machine Learning

grini, 1989], a medical image is stored digitally in pixels (low-level representation). A more abstract representation of the image could be in pixel regions, made up of pixels with similar characteristics. An even higher level of information may use attributes such as the number, size, and shape of each region, etc. The taxonomy of machine learning strategies in Figure 4.1 is based on the amount of inference the learning system must perform to bridge the gap between environmental and system knowledge levels. Four categories are defined. These are [Cohen and Feigenbaum, 1982]:

1. *rote learning*, in which no inference is performed. All information is in a form directly usable by the learning system.

2. *learning by deduction*, in which the information given by the environment is too general or abstract to be understood by the system. The learning system must *specialize* the given data by hypothesizing to fill in the missing details.

3. *learning by induction*, in which the information provided by the environment is too specific and detailed. The learning element must hypothesize more general rules (*generalize*).

4. *learning by analogy*, a combination of inductive and deductive learning in which the information provided by the environment is analogous to the concept the learning system is trying to learn.

) these four categories, one can add:

5. *explanation-based learning*, an analytical learning technique which generalizes rules after the observance of just one example. This is in contrast to the above four categories which are empirically-based.

6. *hybrid leaning*, which combines artificial intelligence and numerically-oriented machine learning techniques.

The explanation-based learning and hybrid learning categories are not distinct in that they may exist as combinations of the first four groups.

Other classification schemes for machine learning strategies have been proposed by Dietterich [1990] and Michalski et al. [1983]. Dietterich suggests a three-part taxonomy divided into speed-up learning, learning by being told, and inductive learning. Michalski et al. [1983] note that machine learning techniques can also be classified according to the knowledge representation used (eg. decision trees, semantic networks, production rules), or according to the domain of application (eg. medicine, chemistry).

Each of the six categories of machine learning techniques, as listed above, is further described within the remainder of this section.

4.3.1 Rote Learning

Rote learning, also referred to as *learning by memorization* or *learning by being programmed*, is the most elementary form of learning. A rote learning system simply stores the information it obtains for future retrieval and thus involves no inference or reasoning at all. The information it receives (from a teacher or program) must already be in a directly usable form.

In general, a learning system can be considered as performing a function, f, that takes an input pattern $(x_1, x_2, \ldots, x_n)$ and computes an output pattern $f(x_1, x_2, \ldots, x_n) = (y_1, y_2, \ldots, y_m)$. A rote learning system memorizes the associated pair $[(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_m)]$. During recall, whenever the input pattern is again $(x_1, x_2, \ldots, x_n)$, the rote learning system need simply retrieve $(y_1, y_2, \ldots, y_m)$ from memory and thus avoid re-computation of $f(x_1, x_2, \ldots, x_n)$.

A good example of experiments in rote learning is Samuel's [1959] work on programming computers to play the game of checkers. Samuel's goal was to teach a computer to play well by having it memorize and recall the worth of previously

played moves. The program uses the minimax search algorithm which employs a 'look-ahead' strategy to search for optimal moves. The minimax search procedure uses a static evaluation function to estimate the 'goodness' of a move based on the resulting board positions should the move be played. For each move that the computer can play, in building the game-tree the minimax search will consider the opponent's possible countermoves, as well as the countermoves of those moves, and so on. The time required to generate moves and evaluate board positions increases as the search descends deeper into the tree. The minimax procedure is limited by the time and memory required to find a good move. Ideally, the deeper the search can continue in the tree, the better the move found is expected to be.

Samuel used rote learning to store the board positions encountered and their estimated worth. By accessing this memory record, computing time could be saved. The idea was that if the computer used the saved time to compute further in depth, then it would improve with time.

Samuel's seemingly simple approach resulted in a fairly powerful learning scheme. His program was able to play a better-than-average game. The method was, however, constrained by limitations of storage space. Thus, an issue relevant to rote-learning systems is the storage-vs-computation trade-off. Once the cost required to find and retrieve an associate pair (X, Y) becomes greater than the cost required to re-compute $f(Y)$, rote learning becomes ineffective. For such learning, careful consideration must also be given to the memory organization technique employed.

4.3.2 Learning by Deduction

Deduction is the process of inferring specific facts from general data. This process is also referred to as *specialization*. For example, given the background knowledge that 'All demyelinating diseases attack the central nervous system' and the teacher provided sentence 'Multiple sclerosis is a demyelinating disease', then the fact 'Multiple sclerosis attacks the central nervous system' can be deduced.

In deductive learning systems, a teacher provides advice or information that is too vague, general, or high-level for the system to understand. In order to use this information, the system must relate it to procedures or a format which it already knows - to something more specific. This transformation is known as *operationalization.* Deductive learning systems are often implemented as tools for acquiring or editing knowledge in expert systems. This form of learning is also referred to as *advice-taking.*

Mostow's [1979] FOO (First Operational Operationalizer) is an example of a deductive system which learns to take advice on how to play the card game of hearts. The human advice-giver can input playing hints such as 'Avoid taking points'. This advice is not operational since FOO has no procedures to avoid taking points. It does, however, have methods for selecting and playing cards. Thus, the advice can be converted into a form usable by FOO, such as 'Play a low card'. FOO has a few noted shortcomings. Errors in operationalization are prone to occur when there is a large gap between the advice in the form it is input and in a form useable by the program. FOO does not have any way to integrate its acquired operational advice into a knowledge base that could drive a Hearts-playing program. It must be instructed by a user as to which operationalization rules to apply. In spite of its weaknesses, FOO presents an important study in techniques of operationalization.

4.3.3 Learning by Induction

Learning by induction has been the most active area of research in machine learning [Michalski et al., 1983, 1986; Cohen and Feigenbaum, 1982; Michalski, 1983]. *Induction* is the process of inferring general hypotheses from specific facts. This process, also known as *generalization,* is based on the inductive paradigm which can be stated as follows: given facts and background knowledge, find a hypothesis or general rule which together with the background knowledge implies or explains the given facts i.e.

hypothesis + background knowledge → facts

For example, from the facts 'Joe, a gardener, plants tomatoes' and 'Sue, a gardener, plants tomatoes', and the background knowledge 'All gardeners plant the same foods', the rule 'All gardeners plant tomatoes' can be induced. In contrast to deductive learning, which results in the adding of deduced facts to a knowledge base, inductive learning systems can be used to add or refine general rules to the knowledge base. The design of an expert system typically involves a series of interviews between an expert and a knowledge engineer, whose task is to represent the expert's knowledge explicitly. This can be an extremely lengthy process. Feigenbaum [1981], in noting that the interview approach to knowledge acquisition cannot keep pace with the increasing demand for expert systems, has termed this the 'bottleneck' problem. Inductive learning algorithms are generally acknowledged to be most valuable in overcoming the bottleneck problem of constructing a knowledge base in the development of any AI system [Elomaa and Holsti, 1989].

Michalski [1983] presents a formalism of inductive learning theory. He describes *conceptual inductive learning*, a form of inductive learning whose final products (hypothesized rules) are symbolic concept or class descriptions expressed in high-level, human-oriented terms. This is in contrast, he notes, to classification rules generated by traditional mathematical and statistical data analysis techniques, such as regression analysis, numerical taxonomy, and factor analysis, whose results are 'merely mathematical formulas'. Conceptual inductive learning has a strong cognitive science flavor as it emphasizes inducing human-oriented rather than machine-oriented class descriptions. As illustrated in Figure 4.1, the taxonomy of learning by induction separates conceptual inductive learning techniques from the traditional mathematical and statistical inductive learning techniques. These two subdivisions can in turn be organized into supervised and unsupervised strategies.

Conceptual Inductive Learning. Amongst the supervised conceptual inductive learning strategies are Michalski's [1983] STAR algorithm and Quinlan's [1979,

1983, 1986b] ID3 decision-tree algorithm. Examples of unsupervised conceptual inductive learning include Michalski and Stepp's [1983; Stepp and Michalski, 1986] conceptual clustering, Fisher's [1987] COBWEB, and Lenat's [1977] AM. These algorithms will be discussed briefly below.

Michalski's STAR algorithm presents a general methodology for learning structural descriptions from examples. It requires a teacher to input a set of positive and negative examples of the concept to be learned. Michalski's algorithm defines the concept of a star as a set of expressions capable of describing all of the given positive examples and none of the negative ones. The expressions must be maximally general, that is, there can be no other relation expressing them which is more general [Genesereth and Nilsson, 1987]. The algorithm uses a number of generalization rules to transform the expression from specific to general. For example, the expression 'MS lesions occur near the ventricles of the brain and in deep white matter tissue' can be generalized with the *dropping generalization rule* to 'MS lesions occur near the ventricles of the brain' and 'MS lesions occur in deep white matter tissue'.

An appealing aspect of STAR is that it is a fairly simple algorithm for generating symbolic concept descriptions of objects. However, its success relies greatly on its background knowledge or knowledge base. Formalizing domain background knowledge is a difficult task. Furthermore, the algorithm is nonincremental, requiring all training examples to be input at once at the beginning of the training phase.

Quinlan's [1979, 1983, 1986b] ID3 (Iterative Dichotomiser 3) is a well-known example of conceptual inductive learning. ID3 uses a *decision tree* to represent its acquired knowledge, the classification rules. A decision tree is a form of a flow chart in which each tree node represents a test on an attribute, and each outgoing branch corresponds to a possible outcome of this test. For instance, given is a set of training examples with the feature 'color' which can take one of two values, either 'red' or 'blue'. A test on this feature would be represented by an internal node, 'color', having two branches: one for 'red' and one for 'blue'. Each leaf node represents a

classification to be assigned to an example.

In order to classify an example, the algorithm tests the example's feature values against the decision tree. A path is traced from the root to a leaf node. The leaf at the end of the path holds the class prediction for that example.

Building the tree is a recursive process. At each new node, ID3 uses an entropy function to select the feature that will best partition the node's examples into classes. The feature vector with the lowest entropy is chosen as the test feature. Branches are grown from the node for each of the test feature's possible outcomes. The examples are then sorted amongst the branches (according to their test feature values) and stored as nodes of the branches. A node containing examples which all belong to the same class becomes a leaf labeled with the class. The process is repeated until no more leaves can be created.

The advantages of ID3 are that it is a simple learning scheme which can perform generalization, organization, and compression of data. It represents classification rules symbolically as a decision tree, although this form of knowledge representation has been criticized for being difficult to interpret by humans [Cendrowska, 1987]. It also does not allow for the learning of new concepts without reconstruction of the tree from scratch. Difficulties arise in the classifier's predictive ability when it is either trained or tested on noisy data. Several works have appeared in the literature suggesting ways in which to improve ID3's noise-handling ability [Quinlan, 1986a; Quinlan, 1987c; Niblett and Bratko, 1986; Mingers, 1989a]. Like STAR, ID3 is nonincremental, although an incremental version was proposed by Utgoff [1988a]. ID3 also assumes that the given set of features are adequate in discriminating the examples of one class from another.

Conceptual clustering [Stepp and Michalski, 1986; Michalski and Stepp, 1983]. is a form of unsupervised conceptual inductive learning in which a group of objects form a class only if it is describable by a concept from a predefined concept class. Conceptual clustering was proposed as an alternative to cluster analysis and numeri-

cal taxonomy, which are often inadequate for they arrange objects solely on the basis of numerical measures of object similarity. Stepp and Michalski's CLUSTER/S algorithm is able to learn structural relationships between objects. Its classification is hierarchical. Learned classification rules are represented by a single conjunctive statement (list of logical ANDs). This representation form is described as one of the drawbacks of the technique, providing a limited means of knowledge representation. Furthermore, the teacher must specify the number of desired clusters in advance, which can be a hindrance in situations where the number of classes in unknown. CLUSTER/S is a nonincremental learning technique.

COBWEB [Fisher, 1987] is an incremental conceptual clustering algorithm. It uses a hill-climbing strategy to construct hierarchical classification trees.

AM [Lenat, 1977] is a discovery inductive learning program that uncovers concepts in elementary mathematics and set theory. Unlike the machine learning techniques previously described in this chapter, AM does not learn concepts for use in any performance task. It simply seeks to define and evaluate concepts based on its knowledge of mathematics. AM's knowledge base consists of 115 concept definitions selected from finite set theory, as well as 242 heuristic rules to guide it in its search for new theories. Each concept in AM's knowledge base is represented by a *frame*. Frames are a form of knowledge representation useful for describing stereotypes of objects or situations. A frame consists of a number of 'slots' describing features of the concept it represents (eg. name, definition, examples of, ..). Attached to each frame is a set of heuristic rules, used to guide the search for concepts not yet known to AM. AM was able to generate over 200 concepts in mathematics, including the definition of prime numbers.

Lenat worked to design AM as application independent as possible. Heuristic rules which would provide guidance in only a single situation were avoided. AM demonstrated the power of a knowledge base. Its knowledge base and small set of heuristics were able to guide a nontrivial discovery process. Weaknesses of AM,

however, lie in its inability to improve its set of heuristic rules. The more concepts AM discovers, the less efficient its initial set of heuristics is. It is not able to generate heuristic rules to use the newly discovered concepts in later searches. Solutions to this were proposed by Lenat [1983] with the EURISKO project, an extension of AM, which uses heuristics to develop new heuristic rules.

Traditional Inductive Learning Strategies. This category of inductive learning strategies refers to mathematical and statistical classification techniques whose classification rules are represented as mathematical formulae. They differ from conceptual inductive learning in that they do not provide symbolic descriptions of learned concepts. Artificial neural networks, for example, fall into this category. Applications implemented by these networks are training problems rather than concept-learning problems [Laird, 1990]. Given a set of positive and negative examples, the training problem task is to construct a network of linear threshold units that agrees with the training set by generating a 1 for each positive example and a 0 for each negative example. A multi-layer feed-forward network is shown in Figure 4.2. The inputs $x_1, x_2, \ldots, x_n$, representing the feature values of an example, are fed simultaneously into a layer of neuron-like units. These units make up the input layer. The outputs of these units are, in turn, fed simultaneously to a second layer of units, known as a 'hidden layer'. The hidden layer's output can be input to another hidden layer, and so on. The number of these layers is arbitrary, although in practice, usually only a few are used. The last layer of units is the output layer which emits the network's prediction for training examples (1 or 0). Multi-layer feed-forward networks of linear threshold functions, given enough hidden units, can closely approximate any function [K. Hornik et al., unpublished manuscript referenced in Laird 1990]. The problem lies in finding a learning algorithm that can process a set of training examples, setting the weights and thresholds of each unit correctly.

One such algorithm is the backpropagation algorithm of Rumelhart, Hinton, and Williams [1986]. The backpropagation algorithm works by minimizing the squared-

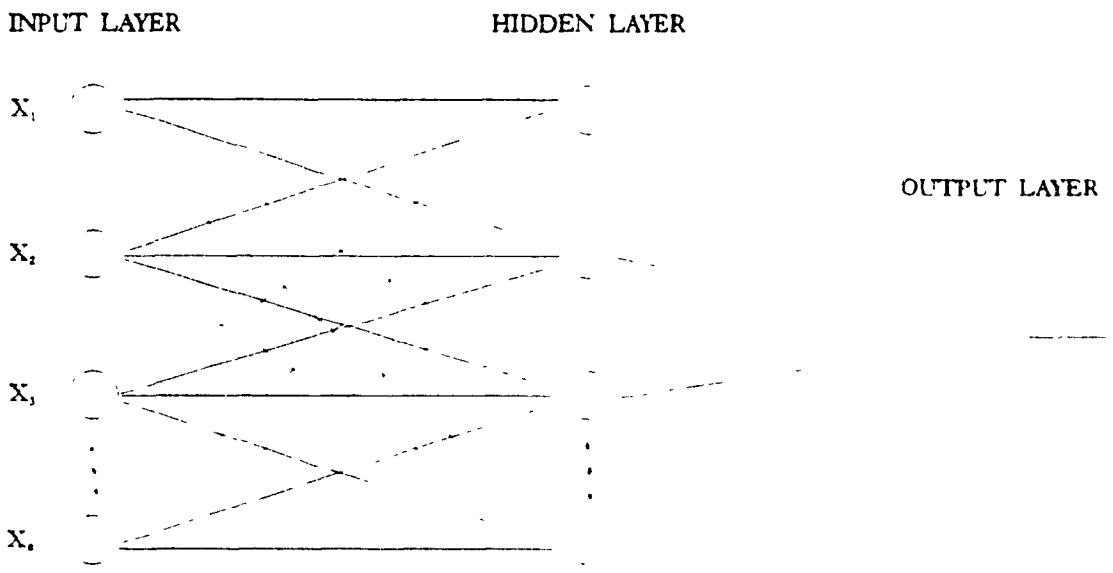


Figure 4.2: A 3-layer feed-forward network.

error between the network's output and the correct outputs provided by the labeled training examples. In order to reduce this error, the algorithm iteratively computes a slight change in all of the network's weights and thresholds. After initializing each weight to a small, randomly chosen value, for each training example the weights are updated as follows. Each layer in the network is evaluated and the output value of each unit is saved. A generalized error is then calculated. The weights of the output unit are adjusted to minimize the error. The weights of each hidden layer are then updated layer by layer, towards the input layer. In this manner, the error is backpropagated, hence giving the algorithm its name. In general, the training set must be processed several hundreds or thousands of times before the weight values converge.

Advantages of neural networks include their high tolerance to noisy data as well as their ability to classify patterns on which they have not been trained [Le Cun, Matan, Boser, Denker, Henderdon, Howard, Hubbard, Jackel, and Baird, 1990]. Neural networks are easily adaptable to parallel machines.

Problems with neural networks include the difficulty in choosing a good network topology. Choosing the right number of hidden units is a trial-and-error task. If there are too few hidden units, then there may be no setting of the weights consistent with the training set examples. Initial weights can greatly affect how well a concept is learned [Shavlik and Towell, 1989]. Neural networks have been shown to classify as well or slightly better than symbolic learning algorithms like ID3, yet they require 10 to 1000 times as much training time [Mooney et al., 1989; Weiss and Kapouleas, 1989]. Another disadvantage of the connectionist approach lies in the difficulty of interpreting the learned network. The weights of hidden units can remain unknown.

To summarize, a series of inductive learning algorithms have been described in the literature. STAR and ID3 are characterized by their ability to generate conceptual descriptions of the classes on which they are trained to recognize. AM is an unsupervised learning program capable of discovering concepts which are not part

of its initial knowledge base. These three approaches are examples of conceptual inductive learning. They differ from neural network models, including that of back-propagation, which represent learned concepts with mathematical formulae rather than symbolic descriptions.

4.3.4 Learning by Analogy

Learning by analogy is a combination of inductive and deductive learning. It is also referred to as *concept learning by analogy* or *analogical problem solving*. An example of this learning strategy is Winston's [1980] ANALOGY program which, for instance, can learn the structural definition of a cup given its functional description. The program draws analogies between the cup's functional description ('has a FLAT-BOTTOM' and 'has a HANDLE') and a library of definitions of other objects (bricks, bowls, suitcases, etc.) derived from previous cases. ANALOGY infers that because a cup has a FLAT-BOTTOM it must be STABLE by drawing an analogy to a stored description of a brick whose FLAT-BOTTOM causes it to be STABLE. Similarly, it uses a stored description of a suitcase to learn, by analogy, why a HANDLE causes the cup to be LIFTABLE. Thus, ANALOGY is inductive learning in that it generalizes to find the analogy between specific examples and precedent cases. It is deductive in that it deduces new rules describing specific training examples, based on rules defining already-seen cases. ANALOGY's rules, however, consist of a collection of instances of causal relations, rather than general rules of causality. This 'weak' domain theory cannot lead deductively to assertions about new causal links [Mitchell, Keller, and Kedar-Cabelli, 1986].

4.3.5 Explanation-Based Learning

A recent research topic to evolve in the history of machine learning is *explanation-based learning*, or EBL. As opposed to empirically-based learning which formulates

generalizations after observing several examples, EBL generalizes from just one example. Examples of EBL systems may perform varying degrees of inference. Mostow's [1979] FOO (deductive learning), Lenat's [1977] AM (inductive learning), and Winston's [1980] ANALOGY (analogical learning) are all EBL systems. A formalism of explanation-based learning is presented in [Mitchell et al., 1986].

There are two basic steps employed in EBL. These are:

- *Explain*: Given *one* input example, build an explanation using background knowledge of the concept to be learned.

- *Generalize*: Analyze the explanation and construct a generalized concept, using other training examples.

DeJong and Mooney [1986] describe an alternate view to EBL called explanation-based generalization (EBG) which combines the explanation and generalization steps into one.

EBL programs can be classified in the following four categories [Ellman, 1989]:

1. *Justified Generalization.* Given initial background knowledge and a set of training examples, find a generalization of a concept that is a logical consequence of the background knowledge and the training set. We say the generalization is 'justified' because it can be expressed in terms of the background knowledge.

2. *Chunking.* Convert a given linear or tree-structure sequence of operators into a single operator.

3. *Operationalization.* Convert a given nonoperational expression into an operational one (expressed in terms of data and actions available to an agent). This is a form of learning by deduction.

4. *Justified Analogy.* Given background knowledge, an analogous example and a training example, find a feature that is true of the analogous example and infer that it is also true for the training example.

These categories are not disjoint. Depending on the language and interpretation used in its description, an EBL program may fit into several categories.

EBL requires a great deal of background knowledge. Results of EBL depend critically on the representation of background knowledge and explanations. Explanations or rules generated by EBL systems can waste storage space and time. Heuristics are needed in deciding when to create and keep rules. EBL is most useful in improving the efficiency of an inference program (i.e. for speed-up learning). An advantage is its ability to explain its predictions.

4.3.6 Hybrid Learning

Hybrid learning comprises the last category of practical machine learning techniques in the presented taxonomy. Hybrid systems merge symbolically-oriented AI approaches to automated learning with numerically-oriented approaches.

Examples of hybrid systems can by found in Shavlik and Towell [1989] and Handelman, Lane, and Gelfand [1989]. Both systems use a combination of rule-based and neural network learning. The rule-based component of the system provides examples to the network which then learns and generalizes from these examples. Although based on preliminary results, Shavlik and Towell found that their hybrid system out performed stand-alone EBL and backpropagation versions of the system.

In summary, practical machine learning techniques have been described. Learning systems can be divided into four categories based on the degree of inference involved. These categories are rote learning (which performs no inference), learning by deduction, learning by induction, and learning by analogy. Two other categories, explanation-based learning and hybrid learning, represent recent approaches

to machine learning which combine methods used in the first four categories.

4.3.7 Theoretical Analysis of Machine Learning Techniques

The final branch of the proposed taxonomy of machine learning comprises work towards the theoretical analysis of machine learning techniques.

The theoretical analysis of machine learning techniques deals with the following kinds of questions [Laird, 1990]:

- How complex is a learning problem in a particular domain? Can learning occur in polynomial time?
- Does the learning time change with the method of knowledge representation employed?
- How can a learning algorithm be designed with provable performance guarantees?

An important step towards answering these questions is found in Valient's [1984] *probably approximately correct* (PAC) theory of the learnable. Valient's theory uses statistics to judge if a learned hypothesis is probably approximately correct, as follows:

Let F be a relation for a class within a universe. $\hat{F}$ is approximately correct if the symmetric difference between F and $\hat{F}$ is small. A learning system is probably approximately correct (*PAC learnable*) if:

$$Probability(error(F, \hat{F}) > \varepsilon) < \delta$$

where δ is the *confidence parameter.*

A bound on the number of training examples (m) required to guarantee that $\hat{F}$ is PAC can be estimated with:

$$m \geq \frac{1}{\varepsilon}(ln\frac{1}{\delta} + ln|H|)$$

where H is a set of hypotheses over a universe and $|H|$ is the number of hypotheses in H.

Valient's theory also allows one to measure the computational complexity of a learning algorithm based on its representation.

Other advances in the theoretical analysis of learning techniques are seen in Mitchell's [1980] work introducing the notion of the *bias* of an inductive learning algorithm, used to restrict the number of inductive conclusions. Mitchell [1982] also defined the notion of a *version state* of hypotheses in his development of the candidate-elimination algorithm. A version state consists of the set of all hypotheses satisfying each of the positive examples and none of the negative examples of a given concept.

4.4 Concluding Remarks

This chapter has described research in machine learning. A historical background was presented, followed by a taxonomic review of work in the field. Advantages and disadvantages of various learning techniques were discussed, as well as a means for approximating the correctness of a learning system. This review was conducted to serve as a background in selecting a machine learning algorithm for the tissue classification of magnetic resonance images. The symbolic learning algorithm chosen is to be compared to the statistical minimum distance and Bayesian classifiers. Examples of each tissue class represent specific facts from which classification rules can be generalized. Therefore, inductive learning is best suited for the classification task. Of the two types of inductive learning, conceptual inductive learning is selected for its symbolic representation of class descriptions. ID3 appears to be a good choice of algorithm amongst the conceptual inductive learning strategies discussed here. ID3 and its successors have been tested on several large medical data sets, including appendicitis, cancer, and thyroid data [Weiss and Kapouleas, 1989; Quinlan, 1987c;

Elomaa and Holsti, 1989], audiological disorders [Mooney et al., 1989], and lymphography [Niblett and Bratko, 1986]. Shepherd [1983] applied ACLS, a modified version of ID3, to image classification in a system for the recognition of chocolates from a production line. ID3 has been the basis of several commercial rule-induction systems [Quinlan, 1986b]. ID3 and the use of decision trees do, however, pose a number of disadvantages:

- Decision trees may contain replicated branches. Pagallo [1989] has proposed a hybrid system, however, which integrates decision trees and Boolean feature combination. The system is able to overcome the replication in decision trees and can express complex DNF [2] expressions which ID3 cannot.

- No additional domain specific knowledge is used to control search other than the training examples themselves.

- The decision tree algorithm is totally dependent on an adequate set of features, allowing the discrimination of examples into classes [Quinlan, 1990].

- When constructing a decision tree, the sets of training examples become smaller and smaller, reducing their statistical significance.

- Mingers [1989a] notes that the decision tree algorithm cannot backtrack if the choice of 'best' feature later seems incorrect.

- Decision trees have been criticized because they do not convey potential uncertainties in classification decisions. Quinlan [1987a], however, proposed a method for assigning uncertainties to decision tree predictions.

- Missing feature values is a problem in the construction of decision trees. Quinlan [1989] proposed methods to deal with this issue. (The segmentation tool does not face this problem).

[2] Disjunctive Normal Form (DNF): an explanation written as a disjunction (OR) of conjunctions (ANDs) eg. $f_1 f_2 f_3 \; OR \; f_1 \neg f_2 f_3 \; OR \; f_1 \neg f_2 \neg f_3 \; OR \; \ldots$

- It is difficult to incorporate new knowledge into an already formed decision tree. Utgoff [1988a] proposed ID5R, an incremental version of ID3 which allows for the incorporation of new training data without requiring the tree to be rebuilt from scratch.

In spite of the disadvantages, there are several arguments towards to the use of a decision tree algorithm:

- By expressing knowledge explicitly, decision trees offer greater understandability over statistical techniques [Mingers, 1989b]. Quinlan [1990] emphasizes the need for classifiers to justify the way in which decisions are arrived, in a form that human decision-makers can understand and scrutinize. Decision trees represent knowledge symbolically, which is in contrast to statistical classifiers and neural networks where knowledge is represented as a collection of numbers. [Quinlan, 1986b]. However, in comparison to other representations used in symbolic learning (such as production rules or frames), decision trees can be difficult to understand [Cendrowska, 1987], particularly deep trees [Mingers, 1989b].

- When making a classification, only features occurring on a decision path need be computed. This may reduce the time spent on feature extraction in comparison to classification methods which require the extraction of all features.

- The computation time required in building a tree increases only linearly as modeled by the product of the size of the training set, the number of features available, and the number of nodes in the tree [Quinlan, 1983]. No exponential growth in time or space has been observed as the dimensions of the induction task increase, making ID3 applicable to large training sets [Quinlan, 1986b].

- Decision trees can represent structural as well as numerical data [Shepherd, 1983].

- The decision tree algorithm is context sensitive. Features may not be uniformly helpful in classification, yet the nature of the algorithm allows it to select the more discriminating features, based on the training samples.

- Decision trees can be simplified to production rules suitable for use in expert systems [Quinlan, 1987b].

- Empirical comparisons of ID3 to the backpropagation algorithm [Mooney et al., 1989] found the two comparable in terms of accuracy. However, ID3 was significantly faster in training and in recall. The backpropagation algorithm was more accurate than ID3 on noisy data, where the noise-handling version used was the chi-square statistic [Quinlan, 1986a]. (More effective noise-handling strategies appear in Mingers [1989a] and Niblett and Bratko [1986]). The comparison of ID3 to linear discriminant, minimum distance, and Bayesian classifiers by Weiss and Kapouleas [1989] on four types of real-world data found that complex problems are better off represented with decision trees. The minimum distance classifier was accurate only when the features were good discriminators. The backpropagation's performance was not the best overall, consuming enormous amount of CPU, and whose accuracy was comparable to that of ID3.

Hence, ID3 was chosen as the symbolic learning algorithm for implementation within the segmentation tool.

The following chapter describes the segmentation tool. The implementation of ID3 and a noise-handling version of the decision tree algorithm is discussed as well as that of the minimum distance and Bayesian classifiers. Each classification algorithm is to be used individually in conjunction with a brain tissue probability model for the tissue classification of MR images.

Chapter 5

Method of Segmentation

This chapter describes the implementation of a tool for the segmentation of MR images of the head at the gross tissue-type level. In particular, the tool is designed to detect MS lesions of the brain. A tissue probability model was developed to provide the segmentation process with *a priori* knowledge of brain anatomy. The development of the model is discussed, followed by a description of the classifiers implemented. These include a minimum distance classifier, a Bayesian classifier, ID3, and a noise-handling version of ID3.

5.1 Brain Tissue Probability Model

5.1.1 Background and Rationale

A tissue probability model was constructed to provide *a priori* probabilities of brain tissue distribution per unit voxel in a standardized 3D anatomy-based 'brain space'. The space, referred to here as Talairach space, is a 3D coordinate system advanced by Talairach et al. [1967, 1988] as a method of reference for the location of cerebral structures. The development of the coordinate system stemmed from studies conducted by Talairach during the late 1940's. Upon examination of a series of human

cadaver brain hemispheres [1], Talairach found that the deep grey matter nuclei of the *thalamus* and *basal ganglia* had the same coordinates in each of the hemispheres surveyed. Using this notion, Talairach then developed a proportional grid system as a means of studying and comparing individual brains of varying size and proportion. The axes of the system are derived from the *AC-PC* or *intercommissural line* (joining structures known as the anterior and the posterior commissures), and from the midline orthogonal to it which reaches the top of the brain. The AC-PC line was chosen as an axis due to its proximity to the 'reference' structures of thalamus and basal ganglia. Upon definition of the axes, all brains within an MR image volume can be affine transformed to fit a standardized rectangular block or 'brain space'. With the exception of non-linear morphological differences, each brain mapped to Talairach space is of the same size, shape, and orientation. Talairach et al. [1967, 1988] used the proportional grid system to define a brain atlas - a series of 2D maps or templates of the more significant cerebral structures. Talairach space has become an integral part of gross neuroanatomy where it is used to determine the coordinates of cerebral blood vessels for their avoidance during the insertion of surgical probes. It is also used in studies of morphometric variability in the normal brain [Evans, Dai, Collins, Peelin, and Marrett, 1991], and in functional studies of cognition or sensory physiology where small signals can be detected only by adding 3D images from different subjects to increase the signal-to-noise ratio.

Talairach's proportional system provided the basis for constructing the tissue probability model of this thesis. MR brain image data was obtained from a group of healthy volunteers. The image volumes were transformed into Talairach space and segmented. From the segmented volumes, the respective probabilities of grey matter, white matter, ventricular CSF, and external CSF in each voxel in Talairach space were derived to create the model. Thus, the tissue probability model is composed of a probabilistic mask for each of the gross-tissue types. The segmentation of a given brain image volume in Talairach space makes use of the model to provide

[1] The brain consists of two **hemispheres**, left and right

a priori or *a posteriori* knowledge of tissue class distribution. As 90-95% of MS lesions occur in white matter tissue [Maravilla, 1988], a large number of which are periventricular (i.e. adjacent to the ventricles), the white matter and ventricular masks in particular are used to guide lesion detection. (This use of ventricular probability is the reason the distinction is made between ventricular and external CSF). The following sections detail the construction of the tissue probability model and its use in the segmentation of MS lesions.

5.1.2 Construction of the Model

Data Acquisition. The tissue probability model was based on MR brain image data obtained from a group of 12 healthy volunteers (three women and nine men) with a mean age of 33.7 years. Images were acquired from a Philips Gyroscan 1.5 Tesla superconducting magnet system. Using a 3D Fast Field Echo (FFE) sequence, 56 non-overlapped transverse slices were collected (TR=75 ms, TE=7 ms) at 3 mm intervals over the entire brain. The data were stored as 256x256 images with 1 mm square pixels. Instead of a head coil, a mirror coil was used for its greater signal-to-noise ratio. The imaging time was 37 minutes. Figure 5.1 shows MR images of a subset of the group of volunteers.

Transformation into Talairach Space. Each MR volume destined for the model was transformed into Talairach space. As the anterior commissure is difficult to locate in MRI, the AC-PC line was approximated from the user-identified location of six cerebral structures within or near the midsagittal [2] plane. These landmarks are: the anterior commissure (AC), the posterior commissure (PC), the inferior aspects of the anterior and posterior corpus callosum, the inferior aspect of the thalamus, and the posterior aspect of the cerebellum. The approximate location of each landmark is illustrated in Figure 5.2a. From an averaged midsagittal view, the top of the brain and the extremes of the AC-PC line in the frontal and

[2]The **midsagittal** plane separates the left and right hemispheres.

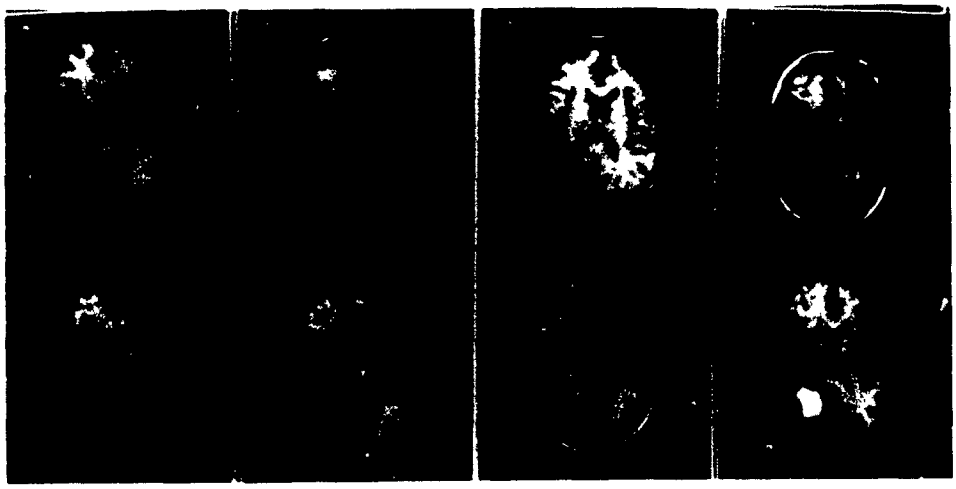
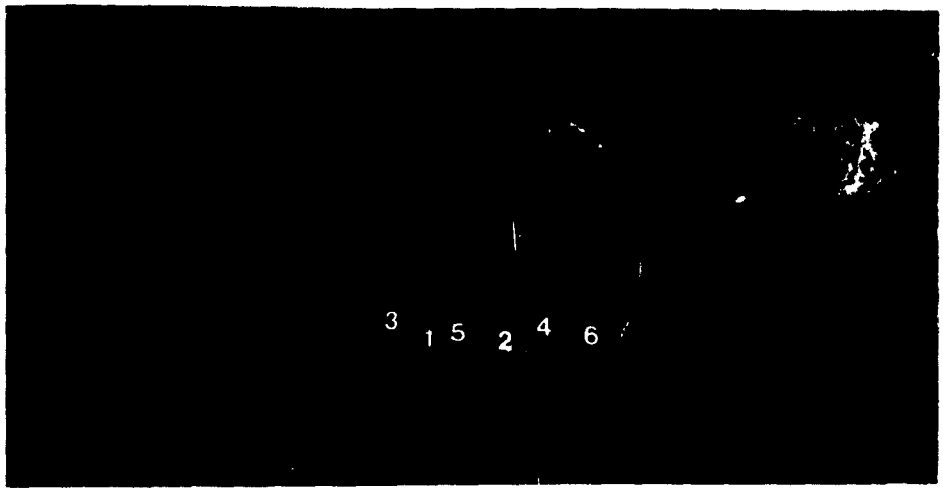


Figure 5.1: MR images of eight of the twelve healthy volunteers used to construct the brain tissue probability model.

a)



b)

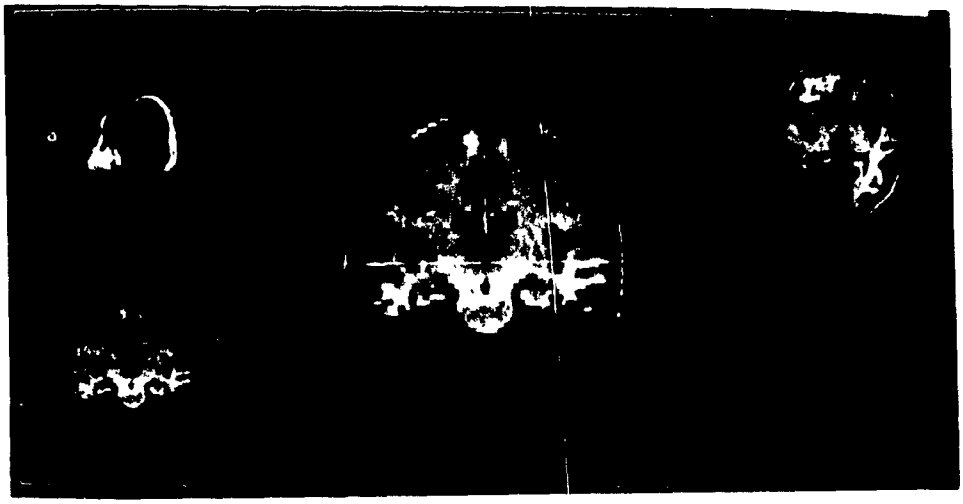


Figure 5.2: Procedure for the transformation of a brain MR volume into Talairach space:
a) Approximate location of the landmarks used to estimate the AC-PC line. These landmarks are: the anterior commissure (1), the posterior commissure (2), the inferior aspect of the anterior (3) and posterior (4) corpus callosum, the inferior aspect of the thalamus (5), and the top of the cerebellum (6). Once the AC-PC line is defined, the extremes of the frontal and occipital lobes and the top of the brain are identified from an averaged midsagittal view.
b) The extremes of the left and right cerebral hemispheres and the top of the brain are identified on a coronal view.

occipital lobes are identified by the user for scaling in the coronal [3] plane (Figure 5.2a). From a coronal view, scaling in the transverse direction is derived from the user-identification of the top of the brain and the extremes of the left and right hemispheres (Figure 5.2b) Thus, each brain volume is transformed into a standardized rectangular block. Transformed volumes contain eighty slices, with a slice-thickness of approximately 1.5 mm. Figure 5.3 shows the image volumes of Figure 5.1 after having been transformed into Talairach space. Aside from minor variations due to uncertainties in choosing the AC-PC line and to non-linear anatomical variability, each brain appears the same at all levels of Talairach space.

Preprocessing of Volume Data. Before tissue segmentation can proceed, the MR image volumes are modified to reduce artifactual intensity variations created by inhomogeneities in the radiofrequency field applied during image acquisition. This 'RF inhomogeneity artifact' causes the intensity of given tissue types to be a function of position in the field and will compromise any attempt to define tissue-specific intensity. The artifact was reduced by applying homomorphic filtering [Axel, Costantini, and Listerud, 1987; Levine, 1985]. Homomorphic filtering is a non-linear filtering process whereby a representation or model of systematic artifact can be used for its removal from an image. The volume artifact was approximately modeled by blurring each slice with an empirically-determined kernel size of 60x60 pixels. Although the procedure was not able to completely remove the inhomogeneity artifact, it did improve the images significantly. Figure 5.4 shows an MR image slice before and after filtering. The non-uniform intensity of white matter caused by the artifact is highly visible in the upper left and bottom right corners of the 'before' image. Individual slice registration or 'alignment' was not necessary due to the 3D image acquisition sequence used.

Segmentation of Volumes for the Model. The CSF ventricles in each volume were manually outlined with an edge tracing routine so as to distinguish

[3] A coronal plane is any plane which separates the brain into front and back.

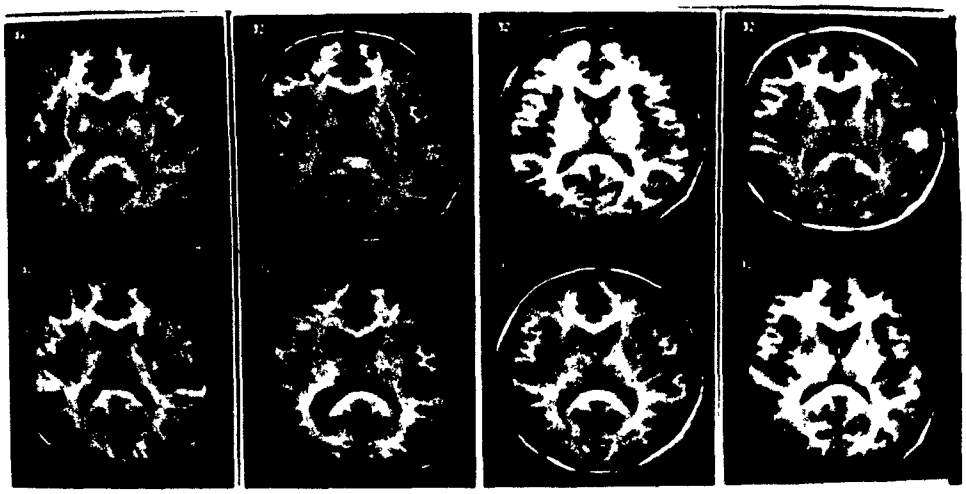


Figure 5.3: The volumes of Figure 5.1 after transformation into Talairach space. Each brain volume now fits into a standardized rectangular parallelepiped or 'block'. All are of the same orientation within the block.

Figure 5.4: A transverse MR slice before (left) and after (right) homomorphic filtering. The RF inhomogeneity artifact can be seen as a diagonal band from the upper left to the bottom right of the image.

ventricular from external CSF. External CSF refers to the CSF beneath the *arachnoid* layer of the brain. A supervised Bayesian classifier was then used to segment each volume into grey matter, white matter, and external CSF. Voxels which did not belong to these categories were classified as 'background'. The features used for classification were the mean intensity and standard deviation of the mean from a 3x3x3 window about each voxel. The training sets contained tissue samples selected from throughout the volume. Due to differences in intensity values for like tissues across different datasets, the classifier was trained on each volume separately. The area of the brain ranging from the level of the eyeballs to the top of the head was segmented. Obvious errors in segmentation were manually corrected with a tissue map editing function. As the volumes had been acquired with high tissue contrast, the resulting segmentations looked generally correct and did not require a great deal of manual correction. (Figure 5.5 shows a *scatter plot* of a typical volume used in the construction of the model. The graph plots the intensity values of samples of each tissue type, illustrating the separability of each tissue cluster). Voxels containing partial volumes of grey matter and CSF tended to be labeled as grey matter. Voxels containing partial volumes of grey and white matter tended to be classified as white matter. These were left unchanged.

Averaging of Segmented Volumes to Create Model. The twelve segmented brain volumes were averaged to create the tissue probability model. The model itself is considered as a volume, with each voxel corresponding to a coordinate in Talairach space. Five probability values are stored at each voxel: the probability of grey matter, of white matter, of ventricular CSF, of external CSF, and of background. These values are derived by calculating the average of each tissue class per voxel from the set of segmented volumes. The model is illustrated in Figures 5.6 to 5.10. Figure 5.6 shows three slices of the model. Each voxel reflects the most probable tissue class at that location in Talairach space, based on the population of volunteers. Tissue classes are color-coded for display. In Figures 5.7 and 5.8, the probability masks for external CSF, ventricular CSF, external and ventricular CSF

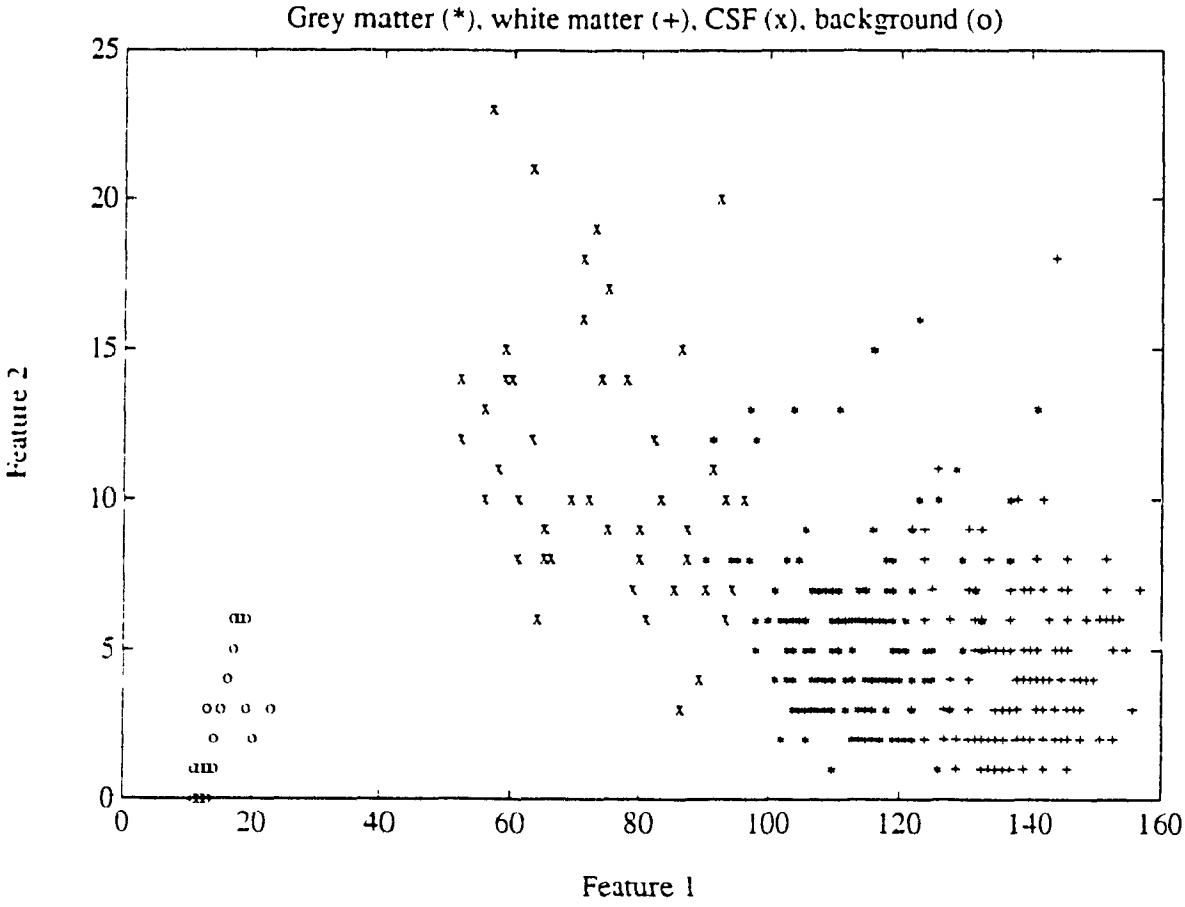


Figure 5.5: Scatter plot showing tissue clusters in an MR image volume of a typical healthy volunteer. The plotted features are the mean intensity (feature 1) and standard deviation about the mean (feature 2) from a 3x3 pixel neighborhood about each voxel.

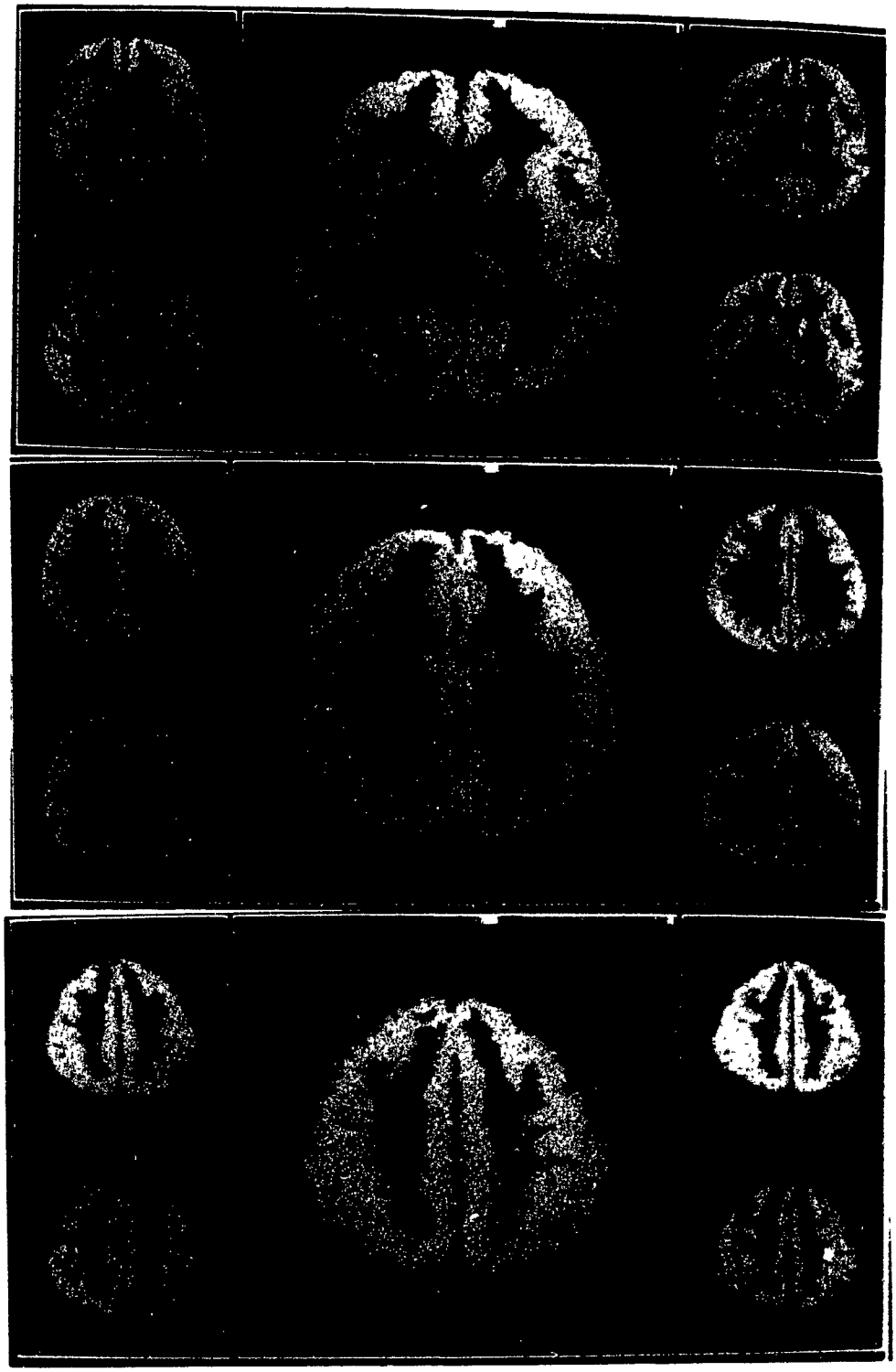


Figure 5.6: The tissue probability model is shown for three slices of the brain in Talairach space. For illustration, each voxel displays the most probable tissue type at that location in Talairach space (yellow for grey matter, green for white matter, brown for CSF).

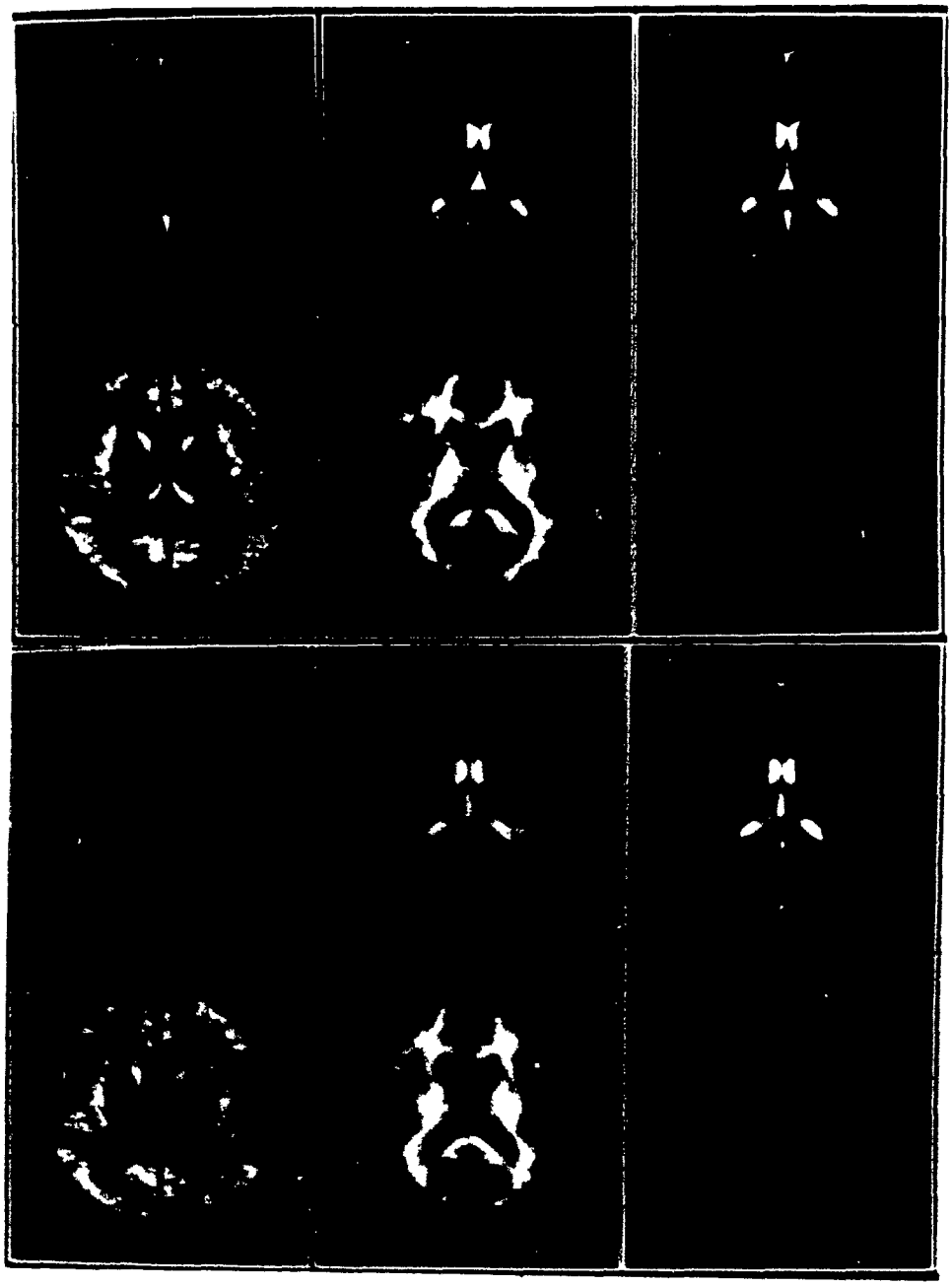


Figure 5.7: The tissue probability masks for external CSF, ventricular CSF, external and ventricular CSF combined (top row) and grey and white matter (bottom row) are mapped to grey scale values. Low intensity voxels indicate areas of low tissue probability for the respective mask. High intensity voxels indicate voxels with high tissue type probability.

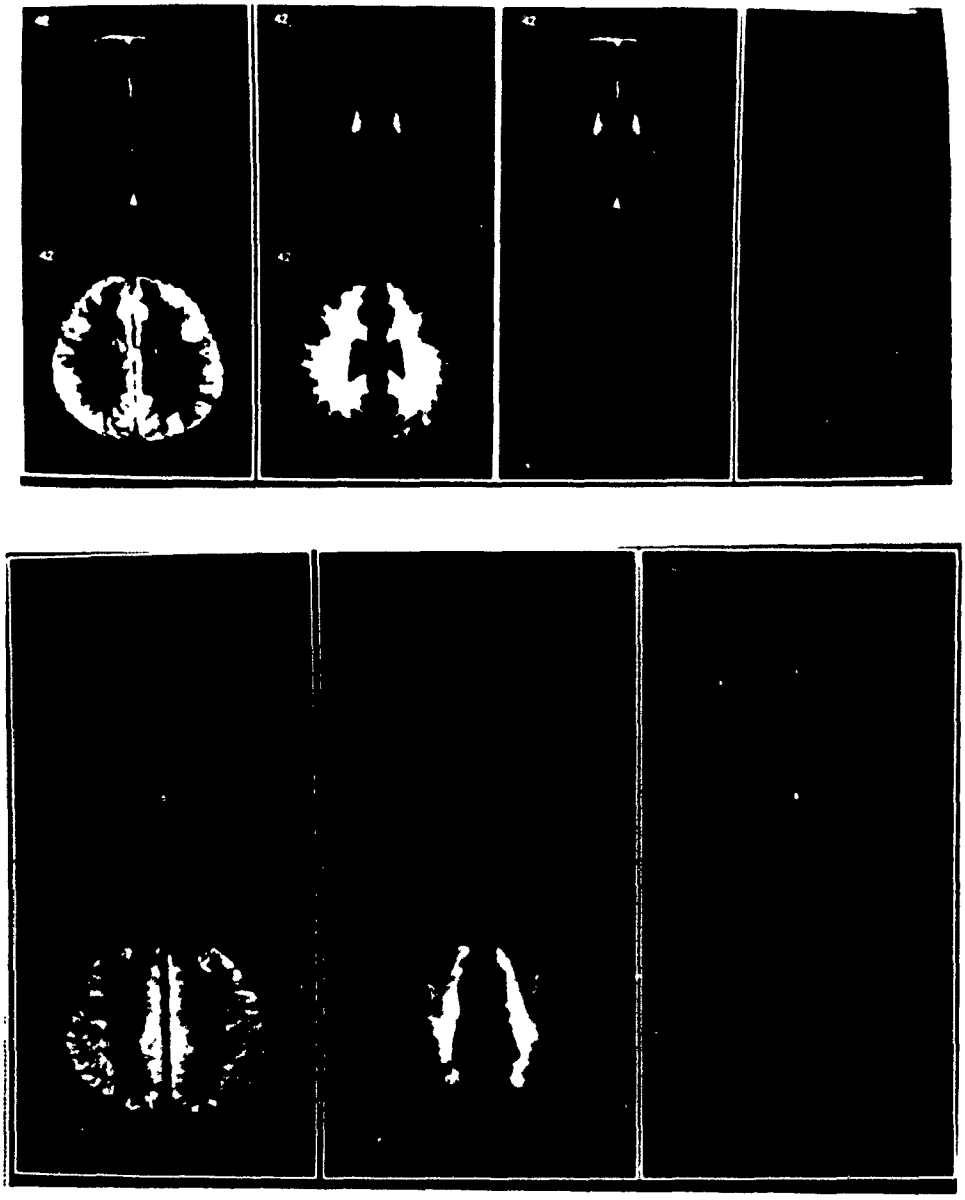


Figure 5.8: The tissue probability masks for external CSF, ventricular CSF, external and ventricular CSF combined (top row) and grey and white matter (bottom row) are mapped to grey scale values. Low intensity voxels indicate areas of low tissue probability for the respective mask. High intensity voxels indicate voxels with high tissue type probability.

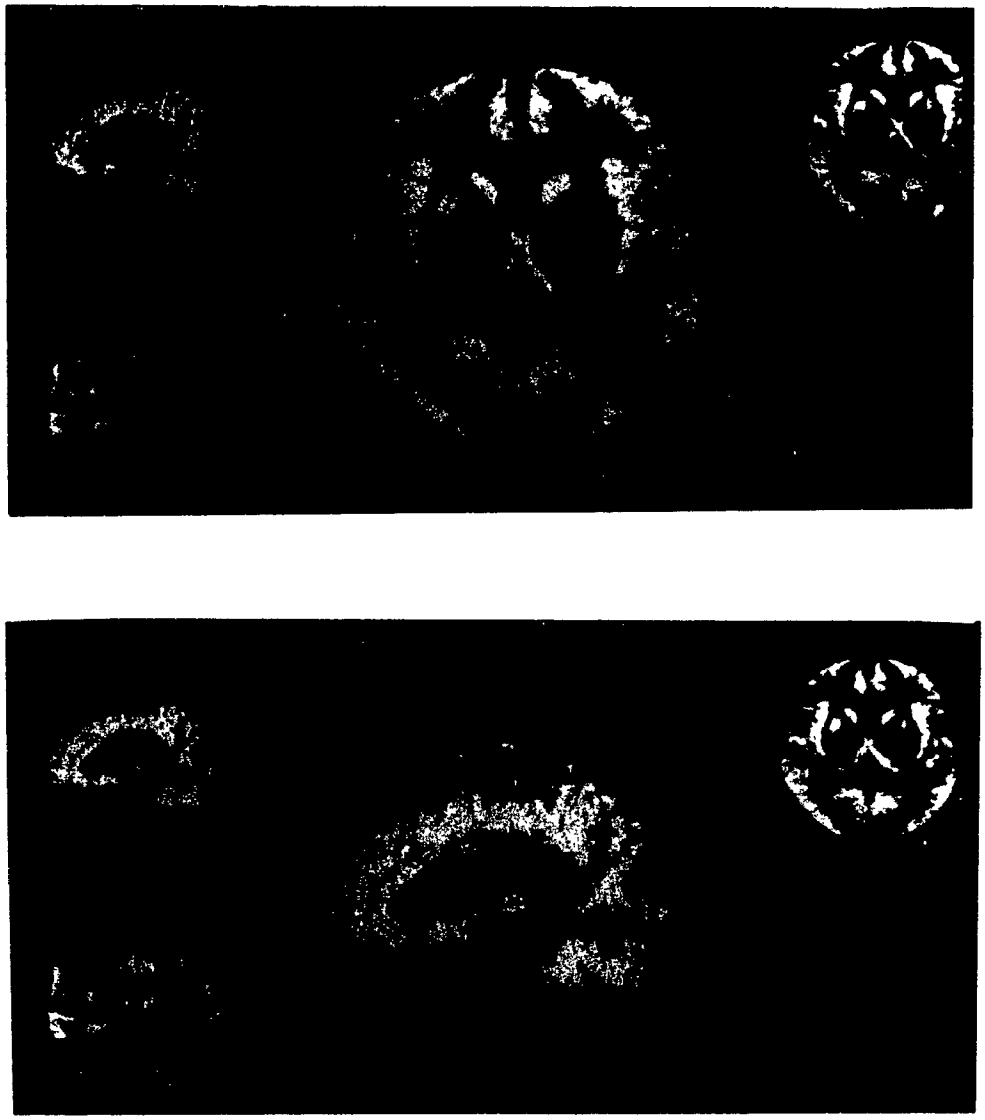


Figure 5.9: Transverse, sagittal, and coronal views of the grey matter tissue probability mask.

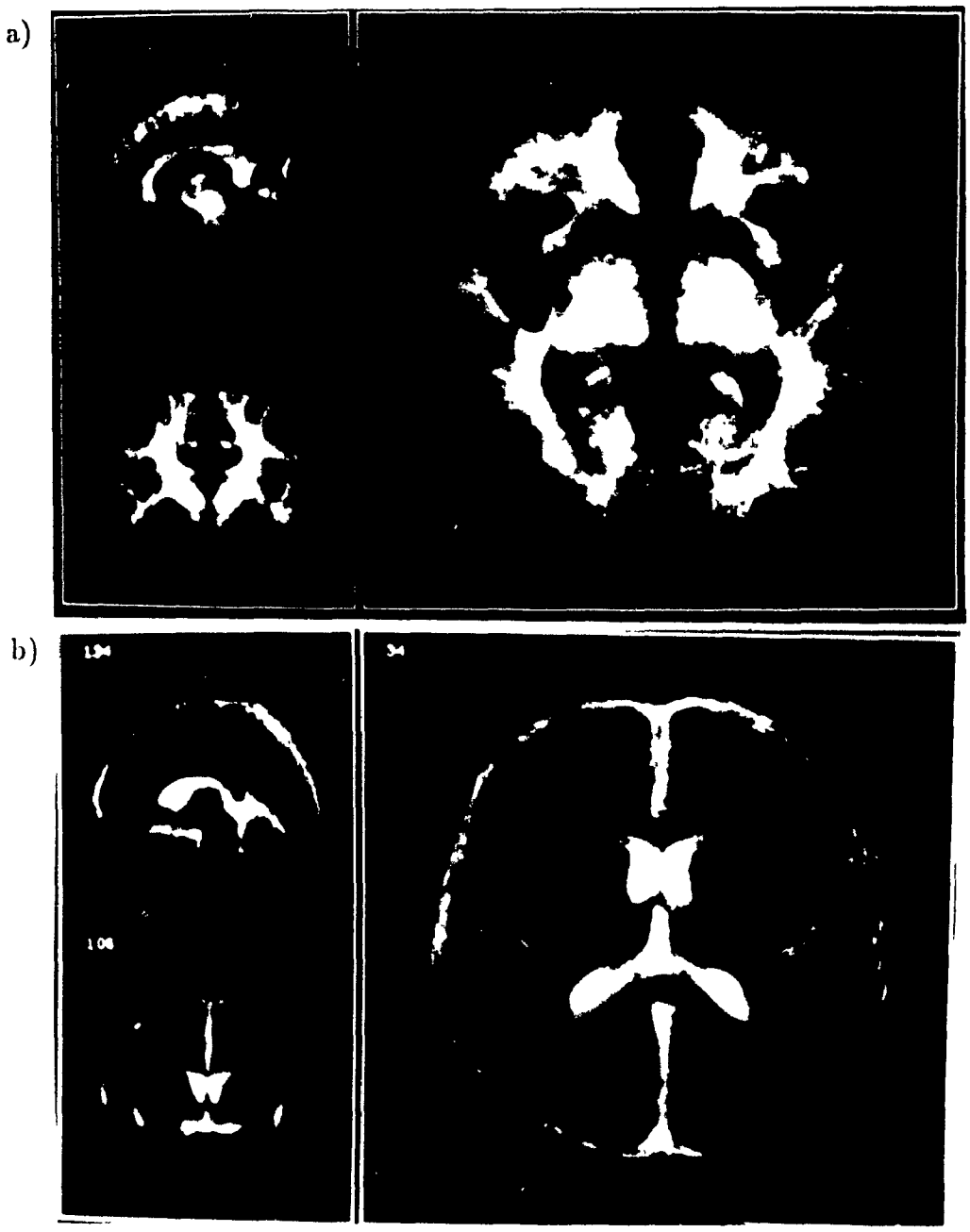


Figure 5.10: Transverse, sagittal, and coronal views of the tissue probability masks for a) white matter, and b) CSF.

combined, grey matter, and white matter have been mapped to grey scale values to illustrate the distribution of these respective tissue groups. Low intensity voxels in the probability mask for tissue T represent voxels with a low probability of being of type T. Similarly, high intensity voxels represent coordinates with a high probability of being of tissue type T. For example, in the second photograph of Figure 5.8 we note that the probability of ventricular CSF occurring towards the top of the head is zero, indicated by a lack of intensities in the ventricular mask. Figure 5.9 shows the tissue probability mask for grey matter in transverse, coronal, and sagittal views. Likewise, Figure 5.10 shows the tissue probability mask for white matter and CSF in each of the three views.

The model contains some anatomical errors due to the partial volume effect. Errors include an under-representation of the thalamus, and the presence of periventricular grey matter in the grey matter mask (Figures 5.7 to 5.9). Errors in the white matter mask include an overestimation of white matter of the gyri and towards the top of the head (Figure 5.10). The white matter mask is to be used during classification to indicate the likely location of MS lesions. Due to the excess of white matter within the model, one may expect a number of voxels to be incorrectly classified as MS lesion (false positive lesions). As the model's periventricular grey matter should in fact be white matter, one may expect the mis-classification of MS lesions in this area. These anatomical errors, however, are small and should not have a great effect on the segmentation results.

5.1.3 Use of the Model within the Segmentation Tool

This section describes the use of the tissue probability model within the image segmentation tool. An examination of other ways in which the model can be used is given in the discussion on future related work in section 7.2.

The probability values of the model per voxel can be used in three ways within the segmentation tool - as *a priori* probabilities of class distribution in Bayesian

classification, as geometric features, and as *a posteriori* information to disallow proposed MS lesions which do not occur in probable areas of white matter. As the model contains tissue probabilities per voxel in Talairach space, a given volume for segmentation must first be transformed into Talairach space.

A Priori. The tissue probability model is used to provide *a priori* probabilities per voxel of grey matter, white matter, cerebrospinal fluid (external and ventricular), and background for the Bayesian classification of these tissue classes in volumes of *healthy* brain data.

The tissue probability model cannot be used to provide Bayesian prior probabilities of tissue class distribution when MS lesion is one of the classes to be recognized. The *a priori* probability of MS lesion is not known as it varies throughout the brain and spinal cord during the stages of the disease.

When using the Bayesian classifier, the user specifies whether or not *a priori* probabilities are to be used. If the model is not used, tissue classes are assumed to be equally likely.

As Geometric Features. The tissue probability values can be used as features, providing heuristic information of probable tissue types based on voxel location. Since the majority of MS lesions, for example, occur in white matter, the white matter mask may be selected as a feature in classification. Thus, the feature vector for a given voxel can contain:

- geometric 'knowledge-based' information, consisting of the grey matter, white matter, ventricular CSF, external CSF, and background probabilities at the corresponding voxel in the model (the user selects which masks to include as features), and

- statistical information, namely the mean intensity value and standard deviation about the mean, based on a neighborhood window about the voxel. The window or 'kernel' size and dimension (2D or 3D) can be specified by the user.

These features are extracted per echo volume.

The user has the option of turning the model 'on' or 'off'. Thus, the segmentation process can be purely data-driven (model off) or data- and model-driven (model on).

A Posteriori. The model can be used *a posteriori* to disallow proposed MS lesions which occur in implausible locations. MS lesions appear as hyperintensities in T2-weighted images. The majority of false positive lesions occur in hyperintense areas caused by noise and the RF inhomogeneity artifact. The choroid plexus (a structure within the ventricles responsible for the production of CSF) may also appear as a hyperintensity. It therefore tends to be mis-classified as MS lesion. Voxels that have been classified as MS are accepted as such if their corresponding value in the white matter probability mask is equal to or above a user-defined threshold. Proposed lesion voxels with white matter probabilities below the threshold are accepted if their ventricular CSF probability is greater than zero. This avoids eliminating true MS lesions which occur near the ventricles, where the probability of white matter may not be high. When the model indicates that a proposed MS voxel is in an unlikely location, the voxel is relabeled as 'other'.

5.2 Classification Methods

The segmentation tool allows the user to select from four classification algorithms: minimum distance, Bayesian, ID3, and a noise-handling version of ID3. Each of these classifiers is described below.

5.2.1 Minimum Distance

The minimum distance classification method [Vannier et al., 1987] bases predictions on distance measurements between each sample to be classified and the class centers,

estimated by their means, for each class in the training set.

A training set of samples is given where each sample is represented by an n-dimensional feature vector $X = (x_1, x_2, \ldots, x_n)$. For each class i, the mean feature vector M_i is computed as:

$$M_i = \frac{\sum_{k=1}^{S} X_k}{S},$$

where S is the number of training samples for class i, and X_k is a sample of class i. For example, if a training set contains a total of 3 samples of grey matter, represented by the feature vectors (13,20), (17,21), and (12,19), then the mean feature vector for the grey matter class is:

$$M_{grey\_matter} = \frac{(13,20)+(17,21)+(12,19)}{3} = (14, 20).$$

To classify an unknown sample, a distance, D, is calculated from the sample's feature vector to each of the class means. The minimum distance classifier of the segmentation tool employs the Euclidean distance:

$$D_i(X, M_i) = \sqrt{(x_1 - M_{i1})^2 + (x_2 - M_{i2})^2 + \ldots + (x_n - M_{in})^2}.$$

X is assigned the class i for which D_i is the minimum:

$$D_i(X, M_i) < D_j(X, M_j) \text{ for } 1 \leq j \leq m,$$

where m is the number of classes. Thus if the unknown sample X is (15,20), and the mean feature vectors for white matter and CSF are (60,80) and (2,5) respectively (assuming three classes), then X is classified as grey matter.

5.2.2 Bayesian

Supervised Bayesian classification for continuous-valued features was implemented within the segmentation tool. This classification algorithm requires the calculation of a covariance matrix for each class, representing the covariance of samples for each class in the given training set [Duda and Hart, 1973]. Difficulties were encountered, however, due to the frequent occurrence of non-invertible matrices, particularly

when tissue probabilities from the model were used as features. The algorithm was then replaced by a supervised Bayesian classifier for discrete-valued features.

Bayesian classification is based on Bayes' probability rule. Given m classes, $C_1, C_2, \ldots, C_m$, and an unknown sample represented by an n-dimensional feature vector $X = (x_1, x_2, \ldots, x_n)$, Bayes' rule states that the *a posteriori* probability of sample X belonging to class i is:

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

where

$$P(X) = \sum_{i=1}^{m} P(X|C_i)P(C_i).$$

As $P(X)$ is constant for $P(C_i|X)$, only $P(X|C_i)P(C_i)$ need be maximized.

When classifying MS lesions, the *a priori* probability of lesion is not known. In this case, classes are assumed to be equally likely i.e. $P(C_1) = P(C_2) = \ldots = P(C_m)$. When classifying healthy brain tissues into grey matter, white matter, CSF, and background, the tissue probability model is used to provide the *a priori* probabilities of $P(C_i)$.

In order to reduce computation in evaluating $P(X|C_i)$, feature class conditional independence (the presence or absence of each feature in a given class is independent of the presence or absence of the others) is assumed. Thus,

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i),$$

where the probabilities $P(x_1|C_i), P(x_2|C_i), \ldots, P(x_n|C_i)$ can be estimated from the training samples by:

$$P(x_k|C_i) = \frac{S_{i,x_k}}{S_i}.$$

Here S_i is the number of training samples of class i, and S_{i,x_k} is the number of training samples of class i for x_k.

In order to classify an unknown sample X, the product $P(X|C_i)P(C_i)$ is evaluated for each class i. Sample X is assigned the class i iff

$P(X|C_i)P(C_i) > P(X|C_j)P(C_j)$ for $1 \leq j \leq m,\ j \neq i$.

5.2.3 ID3

The Algorithm. Classification rules generated by ID3 are represented in the form of a decision tree (see section 4.3.3). Decision-trees are constructed with a top-down recursive divide-and-conquer approach. In building a decision tree, ID3 uses an entropy function (described below) to examine the feature values of each training example and determine the feature that will best partition the given examples into classes. Each node of a decision tree represents a test on a feature, and each leaf denotes a class, as shown in Figure 5.11. The basic algorithm [Quinlan, 1979, 1983, 1986b] is presented in Figure 5.12. The tree starts as a single node containing the training set of labeled examples. If the examples are all of the same class, then a leaf is created and labeled with the class. If they are not, then ID3 needs to grow branches from the node which will test a feature of the examples. The examples are then sorted into groups reflecting the different possible values of the test feature. If the feature's values are discrete, then one branch is created for each of the possible values. If the feature's values are continuous, then two branches are grown corresponding to the conditions $feature \leq value$ and $feature > value$ where the feature and value pair are determined with the entropy function. The feature with the lowest entropy is chosen as the test feature. Each of the branches of the newly generated node is then examined. If a branch contains examples all belonging to the same class, then a leaf node is attached to it and labeled with the common class. Otherwise, the procedure of i) considering all features as tests, ii) choosing the best one, and iii) growing branches for the possible outcomes of the test, is repeated until there are no more nodes to expand.

Entropy Function. Quinlan [1979, 1983] proposed an entropy function as a measure of selecting the best discriminating feature ('test' feature) from a given set. The original function, which handled just two classes, has since been modified for

Training Examples Input to ID3

Feature		*Class*
Age	Origin	
5	France	expensive
5	Quebec	inexpensive
10	Quebec	expensive
8	France	expensive
6	Quebec	inexpensive
6	France	expensive
10	France	expensive

Decision Tree Output by ID3

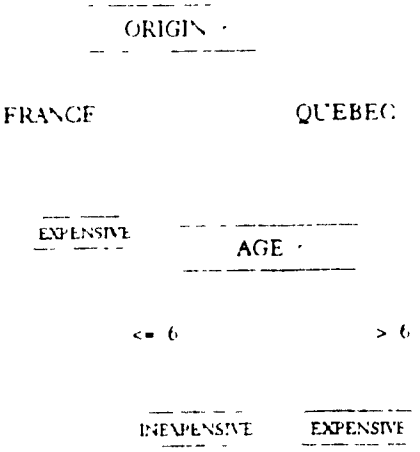


Figure 5.11: Example of a Decision Tree to Classify Expensive and Inexpensive Wines.

```
procedure ID3(examples)

    if examples are all of the same class, c
    then    create a leaf node labeled with class c
            return

    else
            select best-feature, the features that minimizes
            the expected entropy

            for each value v_i of best-feature
                    select the examples, e_i from examples
                    for which best-feature = v_i

                    construct subtree_i using ID3(e_i)
            endfor

            Return a node which tests best-feature and
            has subtree_i attached.
    endif

endprocedure
```

Figure 5.12: The ID3 Algorithm.

multiple classes.

The entropy of feature F is [Clark, 1990]:

$Entropy(F) = \sum_{i=1}^{V} w_i Entropy(F_i)$

where V is the number of values that feature F can take. ($V = 2$ for continuous-valued features, corresponding to the number of branches). w_i is the weight of the i^{th} branch and is defined as the number of examples in branch i divided by the total number of examples at the given node. $Entropy(F_i)$ is the entropy of the i_{th} branch, given by

$Entropy(F_i) = -\sum_{j=1}^{M} p_j log_2 p_j$

where M is the number of classes. p_j is the probability of the j_{th} class in branch i, estimated from the number of training examples of class j having the i_{th} value of feature F. The feature chosen as the best discriminator is the one with the lowest entropy. For a given continuous feature, the value on which to branch is the one with the lowest entropy for that feature.

Using the data in Figure 5.11, if the test feature at the root of the tree were the *origin*, the examples would be partitioned into two groups as follows:

origin = France: 4 expensive, 0 inexpensive

origin = Quebec: 1 expensive, 2 inexpensive

To calculate the entropy for the feature *origin*:

$Entropy(origin = France) = -(1log_2 1 + 0log_2 0) = 0$

$Entropy(origin = Quebec) = -(\frac{1}{3}log_2\frac{1}{3} + \frac{2}{3}log_2\frac{2}{3}) = 0.918$

$Entropy(origin) = \frac{4}{7}Entropy(origin{=}France) + \frac{3}{7}Entropy(origin{=}Quebec) =$ 0.394.

Similarly, the entropy for the feature *age* is 0.571 for value 6 (i.e. among the given *age* values of 5, 6, 8, and 10, testing on $age \leq 6$ and $age > 6$ has the lowest entropy). The *origin* feature has the lowest entropy and is chosen for the test at the root.

In order to keep the trees at a manageable size during implementation, the features were considered as continuous (implying a single feature value to split the samples at a node into two branches), thereby avoiding the growth of a branch for every possible test feature value. If all of the examples at a node have the same feature vector, but the examples are not of the same class, then Quinlan's [1986a] majority voting method for the handling of inadequate features is followed. This involves converting the node into a leaf and labeling it with the class in majority among the given examples.

5.2.4 ID3 with Noise-Handling

Extensions to ID3 have been proposed for the handling of noisy data [Quinlan, 1986a, 1987c; Niblett and Bratko, 1986]. Noise can result from inaccurately measured or missing data. When a decision tree is built, many of the branches will reflect noise present in the training data. Pruning methods identify and remove the least reliable branches. Pruning a tree will decrease its accuracy on the training set, yet should increase the accuracy of the classification of independent test data. Mingers [1989a] compared the accuracy of pruned and unpruned trees and found that, for three out of four domains tested, pruning improved recognition rates by 19% to 25%. As part of the work presented here, a pruning version of ID3 was implemented. The error-cost complexity pruning algorithm [Breiman et al., 1984] was selected as it produces small and accurate trees [Mingers, 1989a]. A drawback of the pruning algorithm is, however, its requirement of an additional test set.

Breiman et al.'s [1984] algorithm for the error-cost complexity pruning of decision trees is summarized in Figure 5.13. In essence, the method generates a group of trees pruned at different degrees. Selection of the 'best' pruned tree is made with the use of an independent test set, employed to measure the accuracy of each tree.

A decision tree is built by recursively partitioning a set of training examples. Each node records the class distribution of the training examples it represents.

procedure **error-cost_complexity_prune(***tree*, *testset*)

while *tree* has internal nodes

Search for internal node n_{minerr} with minimum error-cost complexity. This cost for a given node reflects the resulting mis-classification rate (based on the training set used to create *tree*) should the node be pruned.

Prune *tree* by cutting off subtrees at node n_{minerr}.

Classify *testset* on pruned *tree* and calculate its mis-classification rate.

endwhile

Return the pruned tree whose mis-classification rate (R_{miss}) on *testset* is within 1 standard error (SE) of the tree with the minimum mis-classification error, where $SE = \frac{R_{min} \times (100 - R_{min})}{(number\ samples\ in\ testset)}$.

endprocedure

Figure 5.13: The Error Cost Complexity Pruning Algorithm.

Assuming the features are adequate, each leaf will contain examples of a unique class. A tree can be pruned as long as it contains internal (non-leaf) nodes. A node is pruned by removing its branches. The pruned node becomes a leaf and is labeled by the most frequent class of its former branches.

The pruning algorithm computes the error-cost complexity of each internal node, n. This cost measures the percentage of mis-classified training examples that would occur should node n be pruned ($err_{prune}(n)$) versus the resulting mis-classification rate if n were to be kept ($err_{keep}(n)$):

$$err\_cost\_complexity(n) = \frac{err_{prune}(n) - err_{keep}(n)}{L-1},$$

$$err_{prune}(n) = \frac{M_n}{S},$$

and

$$err_{keep}(n) = \sum_i err_{prune}(n_i),$$

where L is the number of leaves in the subtree at node n; S is the number of training samples; M_n is the number of training samples that would be mis-classified if node n were pruned, and n_i are the leaves of n.

The node selected for pruning is the one with the minimum error-cost complexity. The procedure is repeated with each pruned version of the original tree until no more nodes can be pruned.

An independent test set is used to measure the accuracy of each pruned tree. Only the 'best' pruned tree is retained. Breiman's method selects the smallest tree with a mis-classification rate (R_{miss}) within one standard error (SE) of the minimum. The standard error of the mis-classification rate, assuming a binomial distribution, is:

$$SE = \sqrt{\frac{R_{miss} \times (100 - R_{miss})}{T}}$$

where R_{miss} is the percentage of incorrect classifications on the test set, and T is the number of samples in the test set. The use of a pruned tree generally results in

improved accuracy on test data, and faster classification.

5.3 Implementation Details

The segmentation tool was developed on a SUN Sparc station with 24 Mbytes of memory and consists of over ten thousand lines of 'C' language code. This work was conducted at the NeuroImaging Laboratory located within the McConnell Brain Imaging Center of the Montreal Neurological Institute. The tool was built on top of a software package called MSP (Multiple Sclerosis Package). The package was originally designed to allow clinicians to segment MS lesions in MRI manually (by tracing or painting in lesions with the use of a mouse-controlled cursor), or by using edge tracing or region growing routines on each individual lesion.

The steps for the segmentation of a volume are the following:

Input: single or dual-echo MR volume.

1. Transformation of volume into Talairach space.

2. Preprocessing of data with homomorphic filtering to diminish RF inhomogeneity artifact.

3. Definition of a training set (and test set if desired). Sets can contain samples from throughout the volume.

4. Selection of a classification algorithm: minimum distance classifier, Bayesian classifier, ID3, or an error-cost complexity pruning version of ID3. If the Bayesian classifier is selected, the user can also indicate if the tissue model is to be used to provide *a priori* probabilities.

5. Selection of features: These can include the grey matter, white matter, ventricular CSF, external CSF, and background probabilities per voxel, as well

as the local mean and standard deviation about the mean in each echo volume for a neighborhood centered at the voxel. The neighborhood size and dimension (2D or 3D) can be specified.

6. Indication if the model is to be used *a posteriori* to disallow proposed lesions below a threshold probability for white matter tissue. If so, the threshold can be set.

7. Training of the classifier on a specified training file or by indicating training samples interactively. Training samples can be input as individual pixels or as hand-drawn and labeled regions.

8. Selection of the slices to be segmented. Individual slices or the entire volume can be selected. If a test set is used, confusion matrices are computed and can be displayed on screen or written to a file.

Output: Segmented image data.

Segmented images can be saved as 'tissue maps'. A 'paint-tissue' function was implemented to allow users to edit maps interactively in case of errors produced by the automated segmentation. Segmented volumes are left in Talairach space to facilitate comparison with volumes of other patients and of the same patient at different time points.

When using ID3, the user can specify the 'window size' on the training set (the number of training samples to be used in building the tree). By default, the entire set is selected. If a subset is used, the constructed tree is tested on the remaining samples. If the user finds that the accuracy is less than satisfactory, more training samples are added to the window. The user can select the initial window size as well as the number of examples to be added on each trial. The process of building a tree and testing on the remaining samples is repeated until all samples have been added to the window, or the user wishes the process to stop.

All of the segmentation options have default values. The minimum input required from the user is to provide a training set.

5.4 Concluding Remarks

The method of segmentation employed in the development of an MR image segmentation tool has been described. A tissue probability model was constructed from the segmented MR image volumes of twelve healthy volunteers to provide *a priori* probabilities, per voxel in Talairach space, of the gross tissues of the brain. These tissue groups are: grey matter, white matter, ventricular CSF, external CSF, and background. The model is used in three ways:

- as *a priori* probabilities of class distribution for the segmentation of images of healthy brains,
- as geometric features in addition to image intensity-based features, and
- to disallow proposed MS lesions which do not occur in probable areas of white matter.

In addition to allowing the user to specify the model's use, the segmentation tool also permits the user to select the features to be extracted and the classification algorithm to be employed. The user can choose from four classifiers for the task of tissue classification. The classifiers have been described in this chapter, and are, namely, a minimum distance classifier, a Bayesian classifier, ID3, and an error-cost complexity pruning version of ID3. All classifiers are supervised.

Experiments designed to evaluate the usefulness of the model and the performance of each of the four classifiers are described in the following chapter.

Chapter 6

Experimental Results

This chapter describes the experiments conducted to evaluate the usefulness of a 3D voxel-based tissue probability model employed for the detection of MS lesions in magnetic resonance images of the brain. Four classifiers trained for the segmentation task, namely a minimum distance classifier, a Bayesian classifier, ID3, and a pruning version of ID3, are compared.

6.1 The Problem of Validation

Validation of segmented images is a difficult task. Various methods of validating classifier accuracy were considered in designing the experiments of this thesis.

The standard approach to the validation of pattern recognition results is to test the given classifier on a set of labeled examples (a test set). A confusion matrix, indicating the number of correct and incorrect predictions per class, is computed as well as the overall recognition rate (percentage of correct classifications) for the test data. When a user selects samples for a test set, he will tend to include only those samples whose class membership he is sure of. The test set can then be biased to contain the 'easier' samples and it thus not always a good measurement of a classifier's accuracy.

A common approach to the validation of MR segmentation results is to compare the output tissue maps with a manually segmented version. Classifier predictions which do not agree with an expert's manual segmentation are regarded as incorrect. Errors can occur with manual segmentation, however, due to poor hand-eye coordination and partial volumes. As Gerig et al. [1991] note, it is doubtful whether a manual tracing of external CSF along the brain surface can give a clear segmentation to be used as ground truth. Individual voxels may contain several tissue types, and tissue boundaries can be of varying contrast, making the accurate outlining of structures difficult.

A third method of appraising the validity of segmentation results is to ask an expert's opinion. The expert can judge the results as 'acceptable', 'acceptable with modifications to be made by hand', or 'unacceptable', for example.

A fourth method quantifies the segmented tissues and compares the amounts with known ratios obtained from postmortem studies. Lim and Pfefferbaum [1989] employed this technique in their segmentation of grey and white matter tissues.

The first three methods each require input from an expert. Ideally the expert should repeat the task (either of manual segmentation, the provision of test sets, or overall assessment of results) on the same data on different occasions so that his consistency in decision-making can be estimated. Individual experts can disagree, for example, as to whether hyperintense voxels within an image represent MS lesion. They may also disagree as to the extent of the lesion. Thus, it is preferable to have a team of experts validate the accuracy of the segmentation tool. Measurements of inter-observer variance (differences in opinion between experts) should be considered.

An alternative method for validating the accuracy of a tissue classifier is to test the classifier on an artificial image volume whereby the class label of each voxel is known. The artificial volume can be created in two ways. One method is to build an imitation of the real-world object for imaging. A plexiglass model of the brain,

for example, can be constructed. Components within the model can then be filled with various chemical solutions, each representing a different tissue type [Rousset, Jacquemet, Lavenne, Chaze, Le Bars, and Cinotti, 1990]. The plexiglass model can be placed in an MRI system and scanned, creating an image volume for which the class labels of voxels are known. Alternatively, the artificial image volume can be generated by a computer program. The testing on such data for which class labels are known eliminates the problem of inter-observer and intra-observer variability described above when real-world image data are used. However, artificial data rarely model their real-world counterpart exactly. It may thus not be an accurate indication of a classifier's ability to segment tissues of actual MR image data. The experiments of this thesis employed both real and artificial data in order to exploit the advantages of each approach.

6.2 Experimental Methods

This section describes the experiments conducted to evaluate the usefulness of the brain tissue probability model in the detection of MS lesions. The performance of four classifiers applied to the tissue classification task is also compared.

6.2.1 Image Data Sets

The experiments were based on both artificial and actual brain MR images of patients with multiple sclerosis.

Artificial Data. An artificial brain volume depicting MS lesions was generated as followed. MR image data of the brain of a healthy individual was acquired and segmented in the manner described for the construction of the tissue probability model (section 5.1.2). The resulting segmented volume was saved so that each voxel contained a class label corresponding to either grey matter, white matter, CSF, or background. With the use of a 'painting' function, MS lesions were manually added

to the volume according to the size, shape, location, and relative intensity in which they can be found in real MR brain images of multiple sclerosis patients. The segmented volume was then used to create an artificial dual-echo image volume by replacing class labels with the mean grey scale values of each tissue type as determined from actual MR T1-weighted and T2-weighted echos (see section 2.2 for discussion of T1 and T2). The mean tissue values (and their variance) are listed in Figure 6.1. Additional versions of the dual-echo artificial brain volume were generated at various levels of noise. Noise was added to each voxel according to a gaussian distribution with variances [1] of 20, 40, 60, and 80 percent. Figure 6.2 shows the T1-weighted volume at the various noise levels. The corresponding T2-weighted images are shown in Figure 6.3. The noise level per tissue type is uniform within an artificial volume. In typical real-world image data, the degree of noise per tissue may vary from 4-17%, as seen in Figure 6.1. The artificial volumes were created with the intent of studying the accuracy of each classifier on data at varying levels of noise.

Real Multiple Sclerosis Data. MR brain image data was obtained from two patients diagnosed as having multiple sclerosis. Images were acquired from a Philips Gyroscan 1.5 Tesla superconducting magnet system. Using a 2D Spin Echo sequence, 64 non-overlapped transverse slices were collected (TR=1700, TE=30/80ms) at 2 mm intervals over the entire brain. MRI data were stored as 256x256 images with 1 mm pixels. The imaging time was 29 minutes. Image data was transferred via Ethernet to the NeuroImaging Laboratory. The image volumes were then transformed into Talairach space consisting of 80 slices, with a slice thickness of approximately 1.5 mm. Homomorphic filtering was applied to reduce the RF inhomogeneity artifact (section 5.1.2). A scatter plot of an MS image volume is given in Figure 6.4. Using the original version of the MSP software package (section 5.3), manual segmentation of each volume into 'MS' and 'other' was performed

[1] gaussian_noise(level) = $\sqrt{ln(rand1) \times (-2level)} \times cos(2\pi \times rand2)$ where $rand1$ and $rand2$ are random real numbers between 0 and 1.0, and $level$ represents gaussian distribution variance.

Tissue Type	*Echo 1* *(T1-weighted)* mean (variance)	*Echo 2* *(T2-weighted)* mean (variance)
grey matter	167.24 (5.76)	151.62 (5.80)
white matter	146.06 (6.17)	135.20 (9.36)
CSF	134.46 (7.59)	154.74 (8.55)
MS lesion	173.72 (6.33)	190.03 (17.17)
background	11.14 (3.89)	17.30 (5.22)

Figure 6.1: Table of typical mean and standard deviation grey scale values for each tissue type. Note that MS lesions are similar to grey matter in echo 1 but are brighter in echo 2.

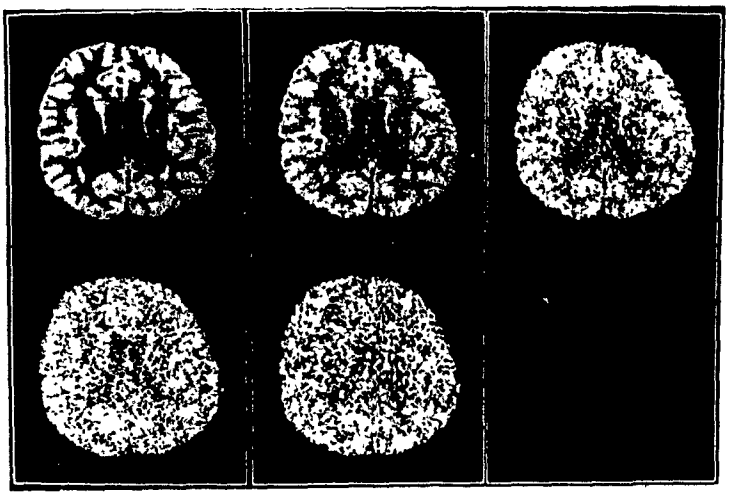


Figure 6.2: Artificial T1-weighted MR brain volumes at levels of 0%, 20%, 40%, 60%, and 80% noise.

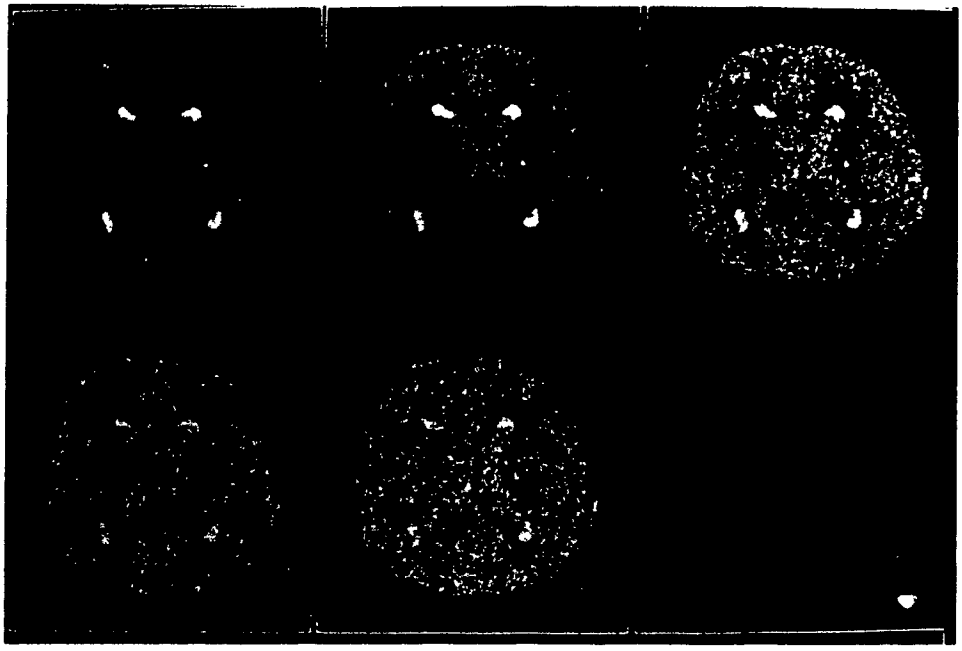


Figure 6.3: Artificial T2-weighted MR brain volumes at levels of 0%, 20%, 40%, 60%, and 80% noise.

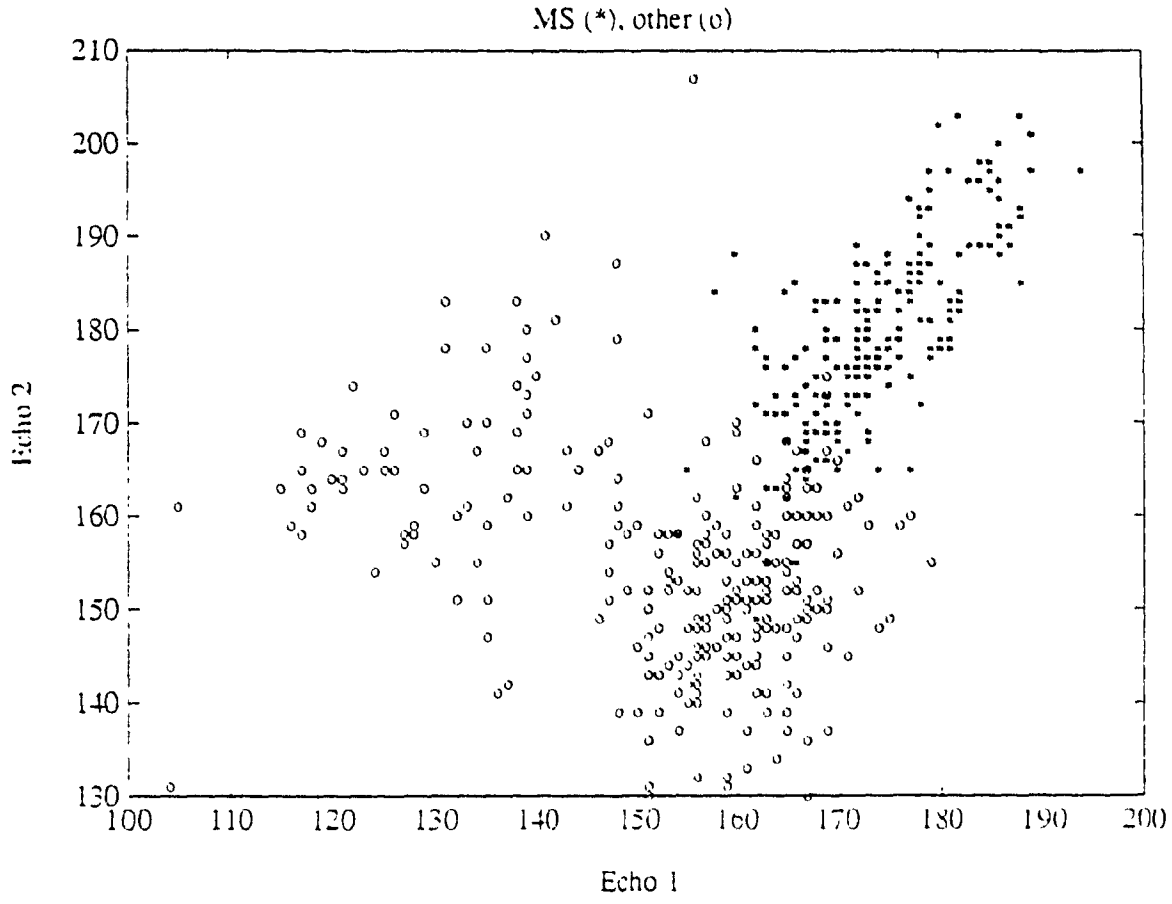


Figure 6.4: Scatter plot of an MR volume of multiple sclerosis data showing the separation in feature space of MS and non-MS ('other') tissues. The features used are the dual-echo mean intensity based on a 3x3 neighborhood about each voxel.

by a research assistant trained by a neurologist in the detection of MS lesions. In all further experiments described later, classifier predictions which did not agree with the manual segmentation were regarded as incorrect. For an experienced user of the MSP package, the procedure of outlining MS lesions in an MR volume requires 4-6 hours. For an inexperienced user, the procedure can take 12-16 hours. For this reason, it was not possible to find a more qualified expert (such as a radiologist, neurologist, or neurosurgeon) willing to perform the tedious and time-consuming task of manual segmentation. (This is also the reason that only two volumes were manually segmented). The research assistant has, however, been outlining MS lesions on MR images for over a year as part of a study conducted by the Multiple Sclerosis Clinic of the Montreal Neurological Hospital to quantify the volume of lesion in patients at various time intervals [Francis, Evans, Baer, Kamber, Collins, and Antel, 1991]. (The manual segmentations obtained for the MS Clinic study were done for volumes which were not in Talairach space. Therefore, they could not be used in this experiment). The research assistant, hereafter referred to as 'the expert', has an intra-observer variability of 3-5% in measurement of total MS lesion volume.

6.2.2 Experiments

Methods. Two sets of experiments were performed, the first based on the artificial volumes and the second based on the real MS data. The segmentation of each volume with each classifier was repeated varying the kernel size (1, 3, 5 pixels) and dimension (2D or 3D) used in the extraction of features. For example, the window from which the feature vector for each voxel is extracted was varied from 1x1, 1x1x1, 3x3, 3x3x3, ..., 5x5x5. These parameters were varied in order to study their effect on the classification of data at varying degrees of noise.

The area of the volumes segmented extended from above the eyeballs (where the tissue probability model starts) to just below the top of the brain. Slices towards

the top of the brain were omitted due to strong partial volume effects as the brain surface curves over to the horizontal. In order to reduce computation time, every third slice was segmented. As the percentage of MS lesions is small with respect to the entire volume, only the tissues within the brain were segmented. (The brain outline of each slice was detected semi-automatically with an edge-tracing function of the MSP package. Voxels outside the region are non-brain and were therefore excluded).

Artificial Data. The artificial image volumes, representing increasing levels of noise, were first segmented in a purely data-driven manner using only the local mean and standard deviation features described in section 5.1.3. In addition, the volumes were segmented using the probability masks of grey matter, white matter, ventricular and external CSF, and background as geometric knowledge-based features per voxel. Each voxel within the volumes was classified into one of five classes: grey matter, white matter, CSF, MS lesion, or background. The training set contained 500 samples. The test set employed for decision tree pruning contained 460 samples. The experiments on the artificial data were designed to test the usefulness of the model in providing geometric features for the classification of normal tissues (grey and white matter, CSF), as well as for the detection of MS lesions.

Real MS Data. The segmentation of the volumes of real MS data was conducted in four different ways:

- data-driven, using just the dual-echo local mean and standard deviation features,
- using the model to provide geometric features as above (employing all five probability masks),
- using the model *a posteriori* to disallow proposed MS lesions in implausible locations, and

- using the model to provide features and *a posteriori* information (a combination of the two preceding cases).

When used *a posteriori*, proposed lesions were accepted if their probability of white matter (from the model) was greater than the empirically determined threshold of 50%. To avoid eliminating periventricular MS lesions, whose probability of white matter may not be high, proposed lesions with white matter probabilities below the threshold were accepted if then ventricular CSF probability was greater than zero, except on slices for which it is possible for choroid plexus to occur [2]. (This range of slices was user-determined upon observing a number of MR volumes in Talairach space and represents another way for the model to incorporate knowledge about MS lesion location). As the contrast between grey and white matter was very poor, each voxel was classified as either 'MS' or 'other'. The training set for each volume contained an average of 450 samples. The test set used for pruning contained an average of 400 samples.

6.3 Results

This section describes the results of the experiments conducted in the segmentation of the artificial and real MR image volumes containing MS lesions. The pruning version of ID3 is referred to as PruneID3.

6.3.1 Results on Artificial Data

Classifier Accuracy. The overall classification accuracy of the minimum distance, Bayesian, ID3, and PruneID3 classifiers is shown without the use of the model in Figure 6.5. Figure 6.6 shows classifier accuracy when the model was employed to

[2] **Choroid plexus**, a structure within the ventricles responsible for the production of CSF, can appear as hyperintensities, similar to MS lesion.

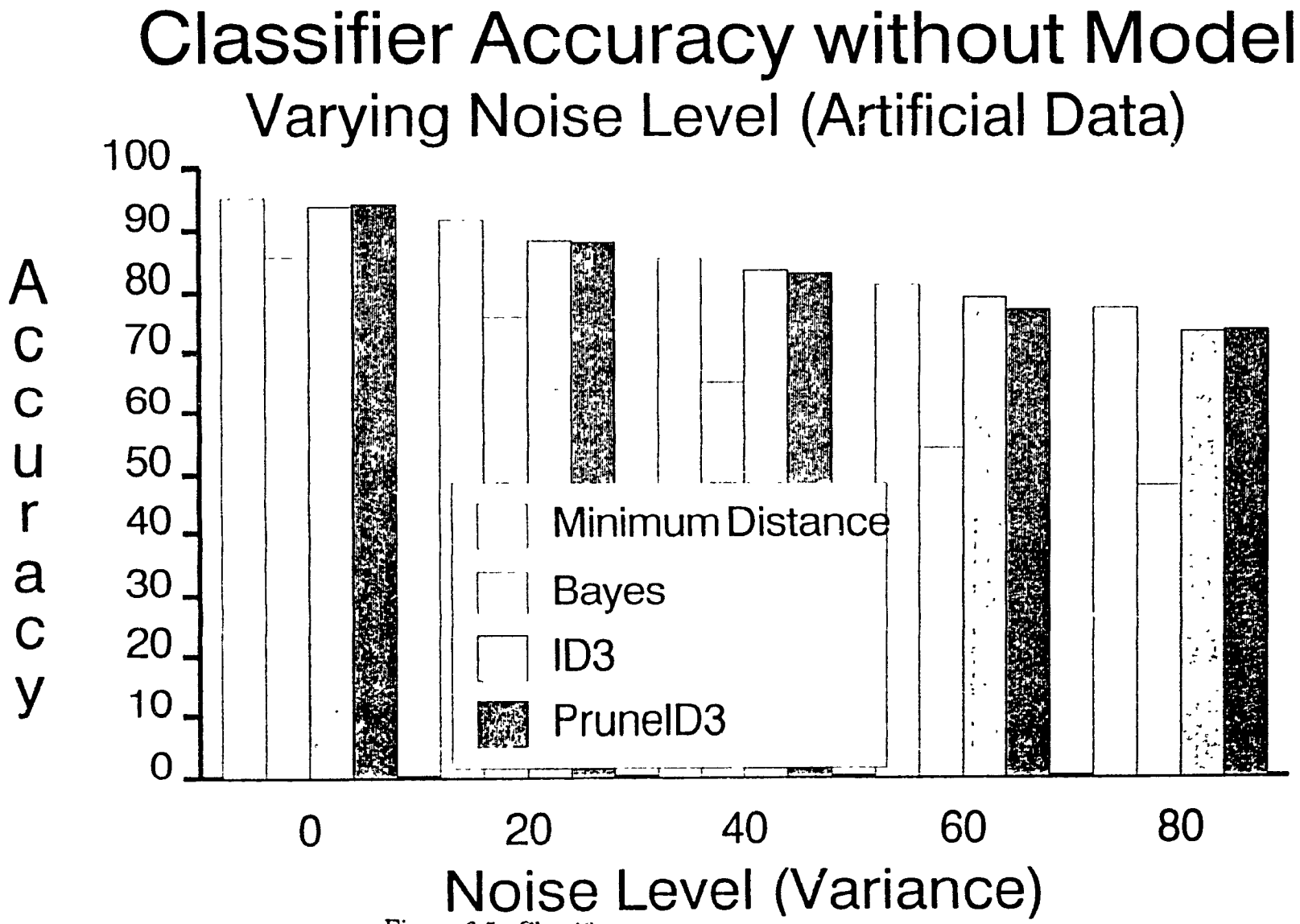


Figure 6.5: Classifier accuracy without use of model on artificial data at increasing levels of noise.

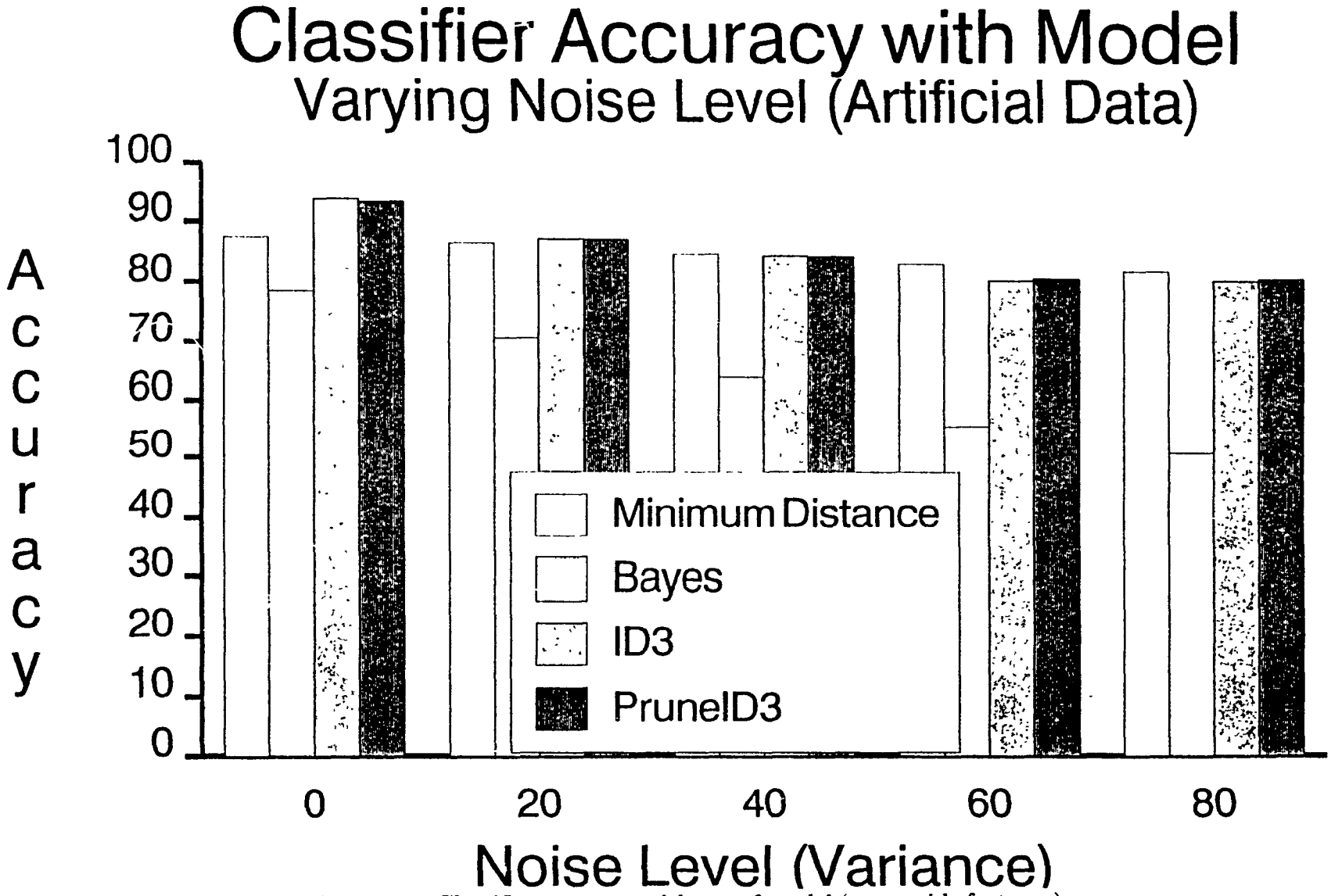


Figure 6.6: Classifier accuracy with use of model (to provide features) on artificial data at increasing levels of noise.

provide geometric features. Classifier accuracy or 'recognition rate' is computed as the number of correct classifier predictions over the total number of samples for the given test set. The minimum distance classifier had the highest overall recognition rates (ranging from 95.67% on the noise-free data to 77.79% on the noisiest data when the model was not used). Overall, ID3 and its pruned version performed as well as the minimum distance algorithm (within at most 5%). ID3 and PruneID3, however, out performed the minimum distance and Bayesian classifiers when the geometric features were added for the classification of the noise-free data. This addition of features, unimportant for the classification of clean data, created confusion for the statistical classifiers whose recognition rates dropped by around 8%. This illustrates ID3's ability to base its predictions on the more discriminating features. The Bayesian classifier was the least accurate, with recognition rates ranging from 85.73% without the model (78.40% with the model) for the noise-free data to 48.26% (50.89% with the model) on the noisiest data.

Pruning did not improve ID3's overall accuracy. (The test set used in pruning should perhaps have contained fewer samples). Pruning did reduce the size of decision trees by 50-60% (Figure 6.7).

Each classifier's accuracy with and without the model is shown individually in Figures 6.8 to 6.11. The model is useful in the classification of noisy data, particularly at the noise levels of 60 and 80%. As expected, the overall recognition rates decrease as the degree of noise within a volume increases. The recognition rates are surprisingly high for the noisy data. This is most likely due to the dual-echo nature of the data, contributing more tissue specific information than a single-echo volume. A future experiment of interest could be to measure the effect of the model on the segmentation of a single-echo volume where one would expect the model to play a greater role in classifying noisy data. (Figure 6.12 shows a segmented slice from the artificial image volume with a noise level of 40%, obtained by the minimum distance classifier without the model).

Noise	Number of Leaves	
Level (%)	ID3	PruneID3
0	16	8
20	28	11
40	51	18
60	70	37
80	86	28

Figure 6.7: Average number of leaves in decision trees for classification of artificial data at varying levels of noise.

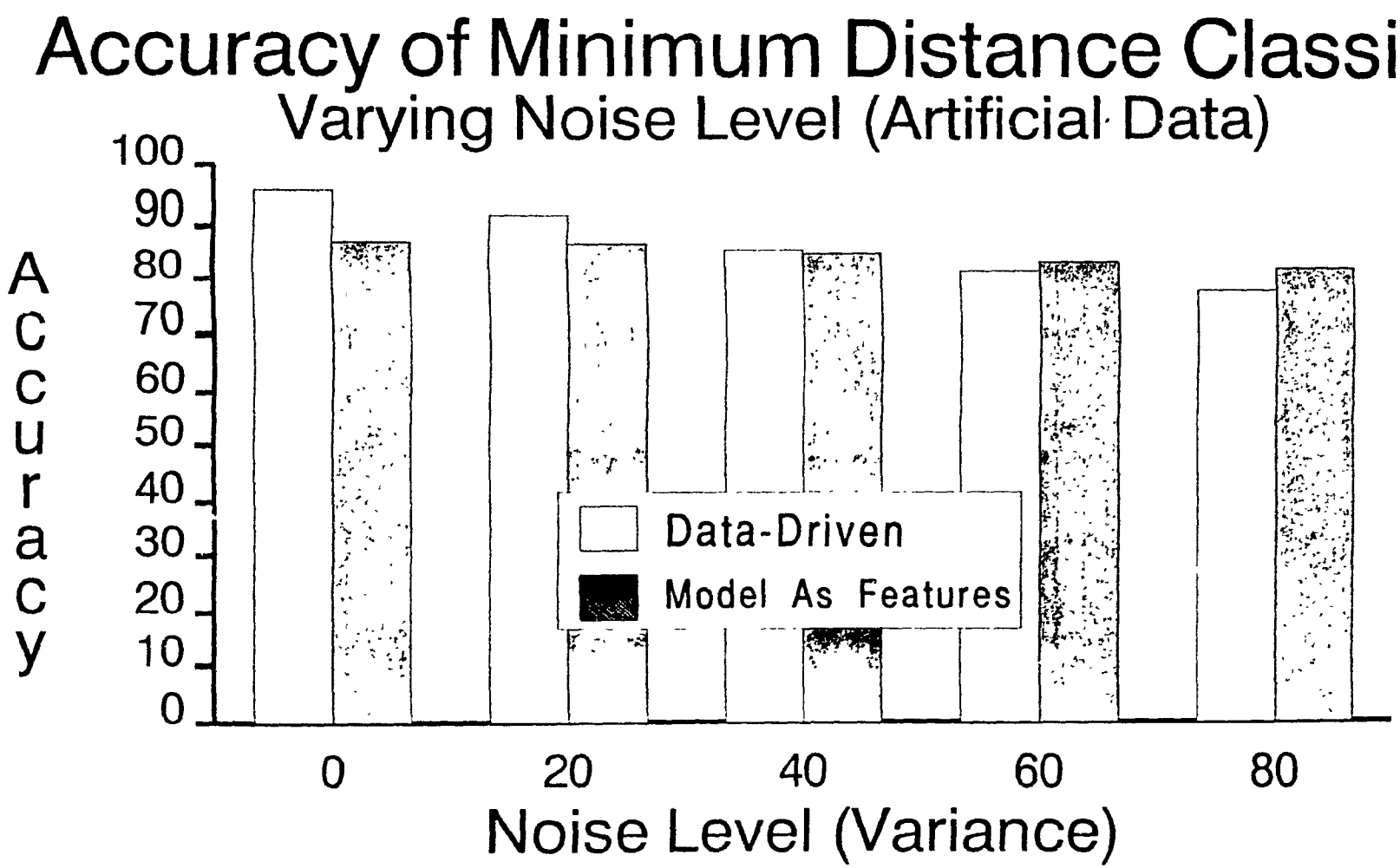


Figure 6.8: Minimum distance classifier accuracy with and without model at varying levels of noise.

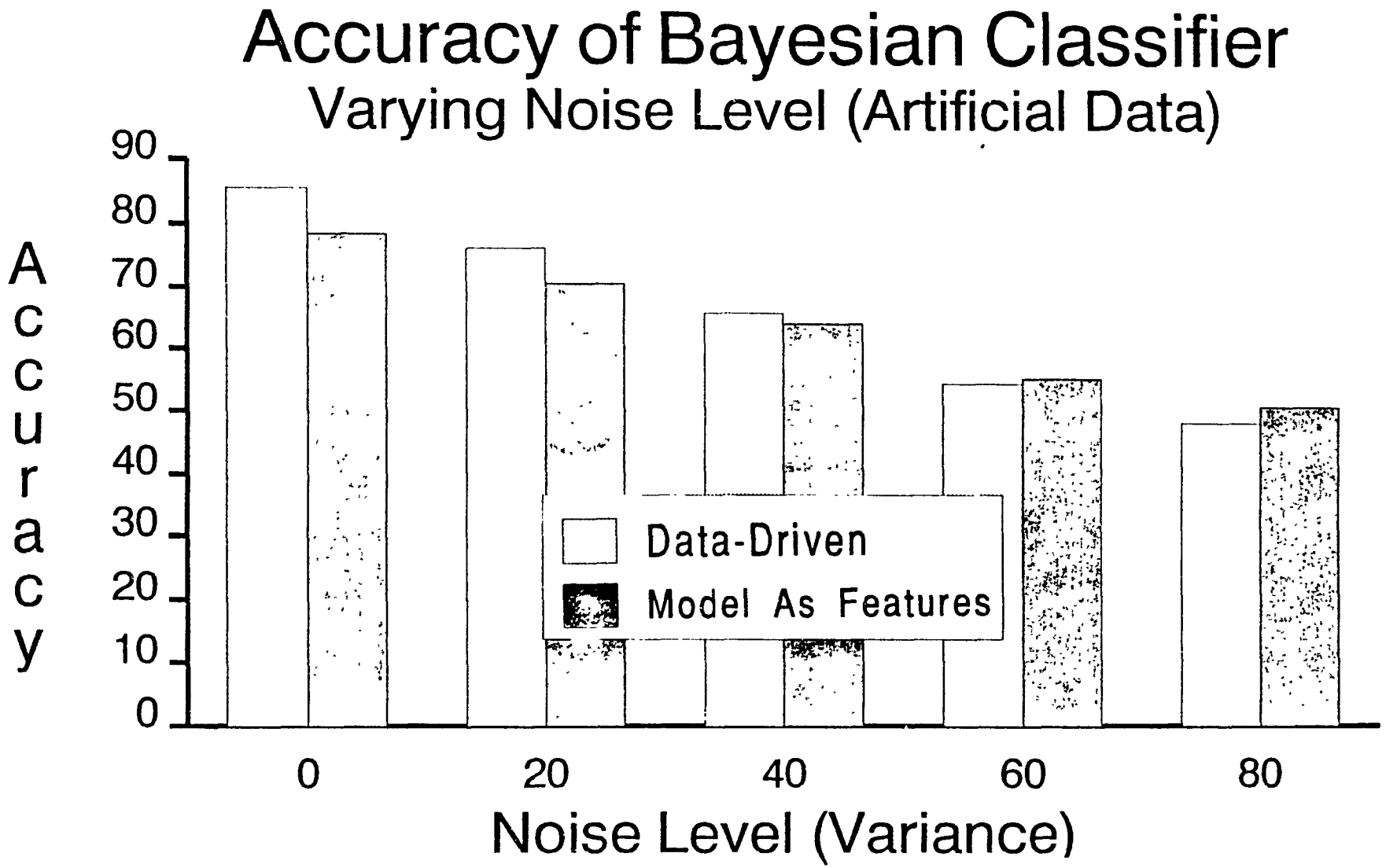


Figure 6.9: Bayesian classifier accuracy with and without model at varying levels of noise.

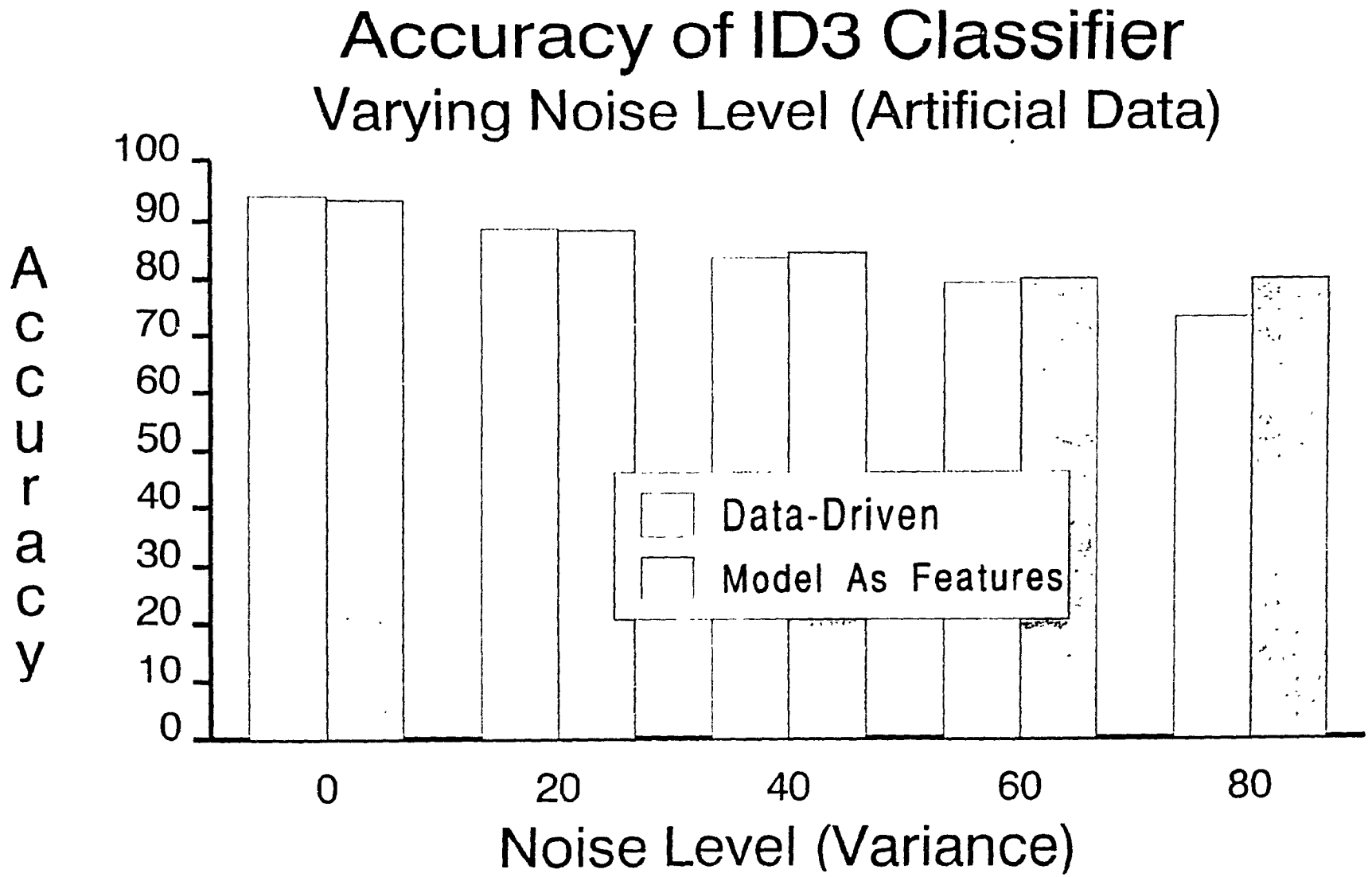


Figure 6.10: ID3 classifier accuracy with and without model at varying levels of noise.

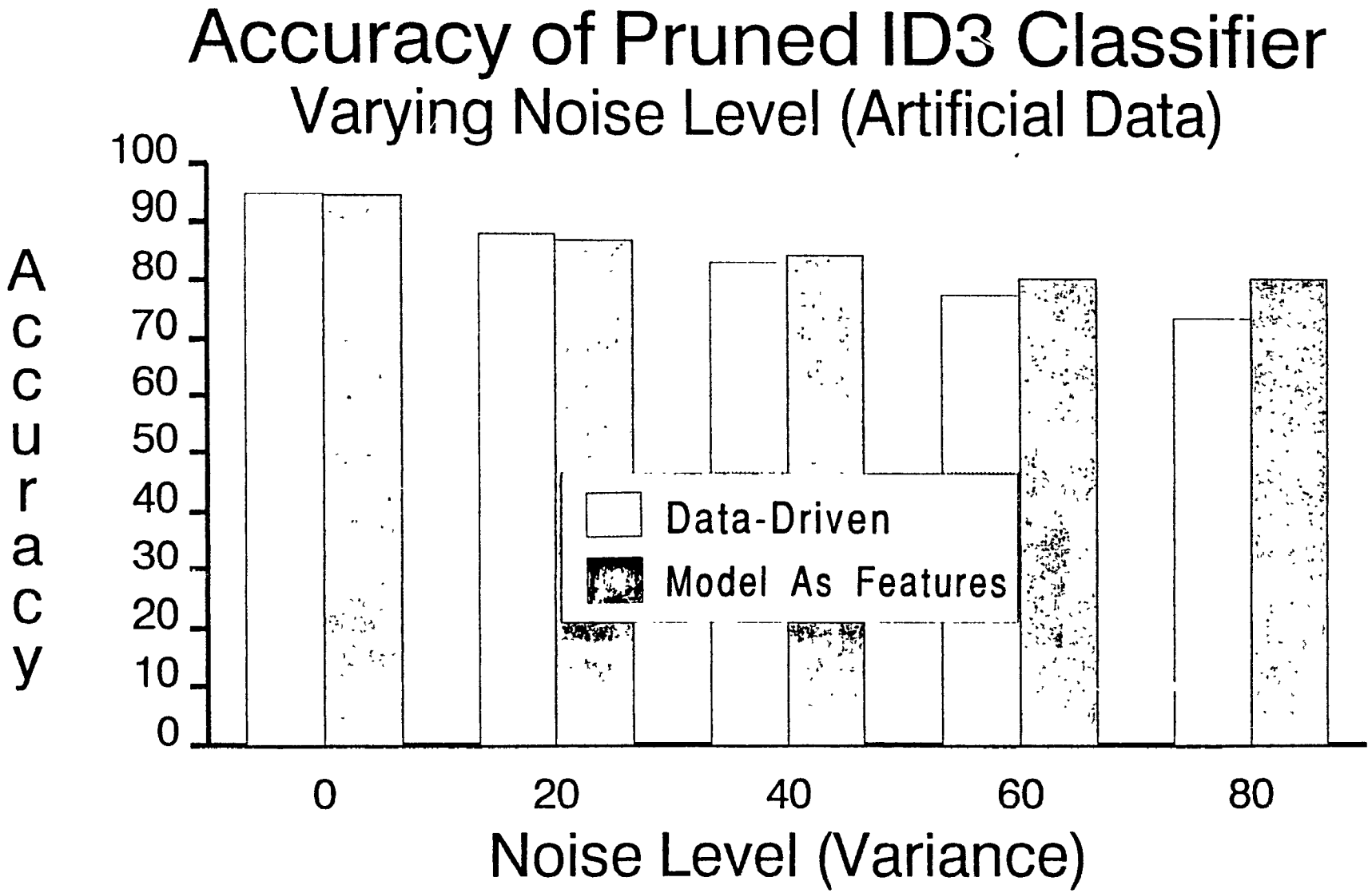


Figure 6.11: PruneID3 classifier accuracy with and without model at varying levels of noise.

Figure 6.12: Segmented slice of the artificial image volume at the 40% noise level. Segmentation results was obtained with the minimum distance classifier without use of the model.

Feature Extraction Time. The average time in CPU seconds required to extract four features per slice (the dual-echo local mean and standard deviation about a voxel) is shown in Figure 6.13 at varying kernel size and dimensions.

Training Time. The overall training time in CPU seconds is given in Figure 6.14 for 500 samples. The minimum distance and Bayesian classifiers are the fastest classifiers in training. The training times of the statistical classifiers are unaffected by increases in the level of noise. The time required for the construction and pruning of decision trees, however, increases as noise levels rise. The noisier the data, the larger ID3's decision tree is (Figure 6.7). Training times of the decision tree classifiers also increase as features are added.

Classification Time. The average amount of CPU seconds required to classify one slice is given for each classifier in Figure 6.15. The classification time for decision trees increases with the level of noise, due to the larger size of trees on noisy data. Classification with pruned trees is up to 2 CPU seconds faster per slice than classification with unpruned trees.

Effect of Kernel Size on Noisy Data. Figures 6.16 to 6.19 show the effect of varying kernel size and dimension (from 1x1 to 5x5x5) for each classifier at each level of noise. The use of large windows within the same dimension (3x3, 5x5, or 3x3x3, 5x5x5) improves classification accuracy of noisy volumes (noise levels at 60-80%), particularly in the Bayesian classifier. Classification of clean data prefers small windows (1x1(x1)). The extraction of features in 3D (considering voxels in slices above and below the given slice) improves classification of the noisy data, particularly for the Bayesian classifier. For example, the use of 3D over 2D for the Bayesian classification of the 80% noise volume improved results by up to 20%.

Kernel Size	*CPU seconds per Slice*
1x1	1.158
1x1x1	1.700
3x3	12.133
3x3x3	25.617
5x5	22.753
5x5x5	92.767

Figure 6.13: Average feature extraction time in CPU seconds per slice at various kernel sizes and dimensions.

Noise	*CPU seconds for Classifier Training*			
Level(%)	Minimum Distance	Bayesian	ID3	PruneID3
0	0.019	0.050	0.867	5.153
20	0.017	0.039	2.256	10.951
40	0.019	0.044	3.300	17.461
60	0.019	0.050	3.767	31.784
80	0.019	0.050	3.750	41.683

Figure 6.14: Classifier average training time in CPU seconds for artificial data (500 training samples).

Noise	*CPU seconds per slice for Classification*			
Level(%)	Minimum Distance	Bayesian	ID3	PruneID3
0	24.054	11.942	4.911	4.032
20	24.075	12.118	8.439	7.322
40	24.159	12.839	10.283	8.925
60	24.244	10.826	12.019	10.029
80	24.422	12.305	10.284	9.999

Figure 6.15: Classifier average classification time in CPU seconds per slice for artificial data.

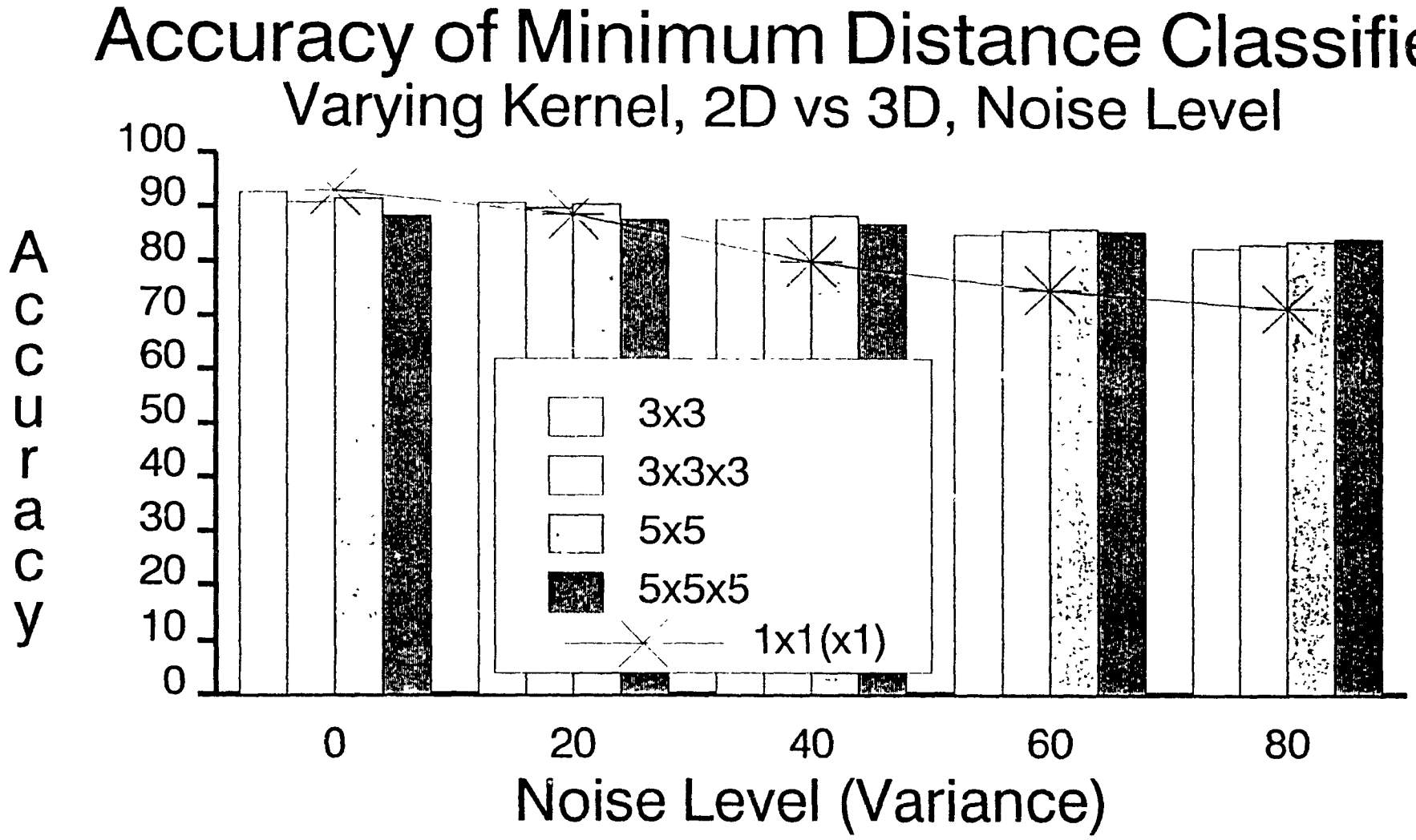


Figure 6.16: Minimum distance classifier: Effect of kernel size on classification of noisy data.

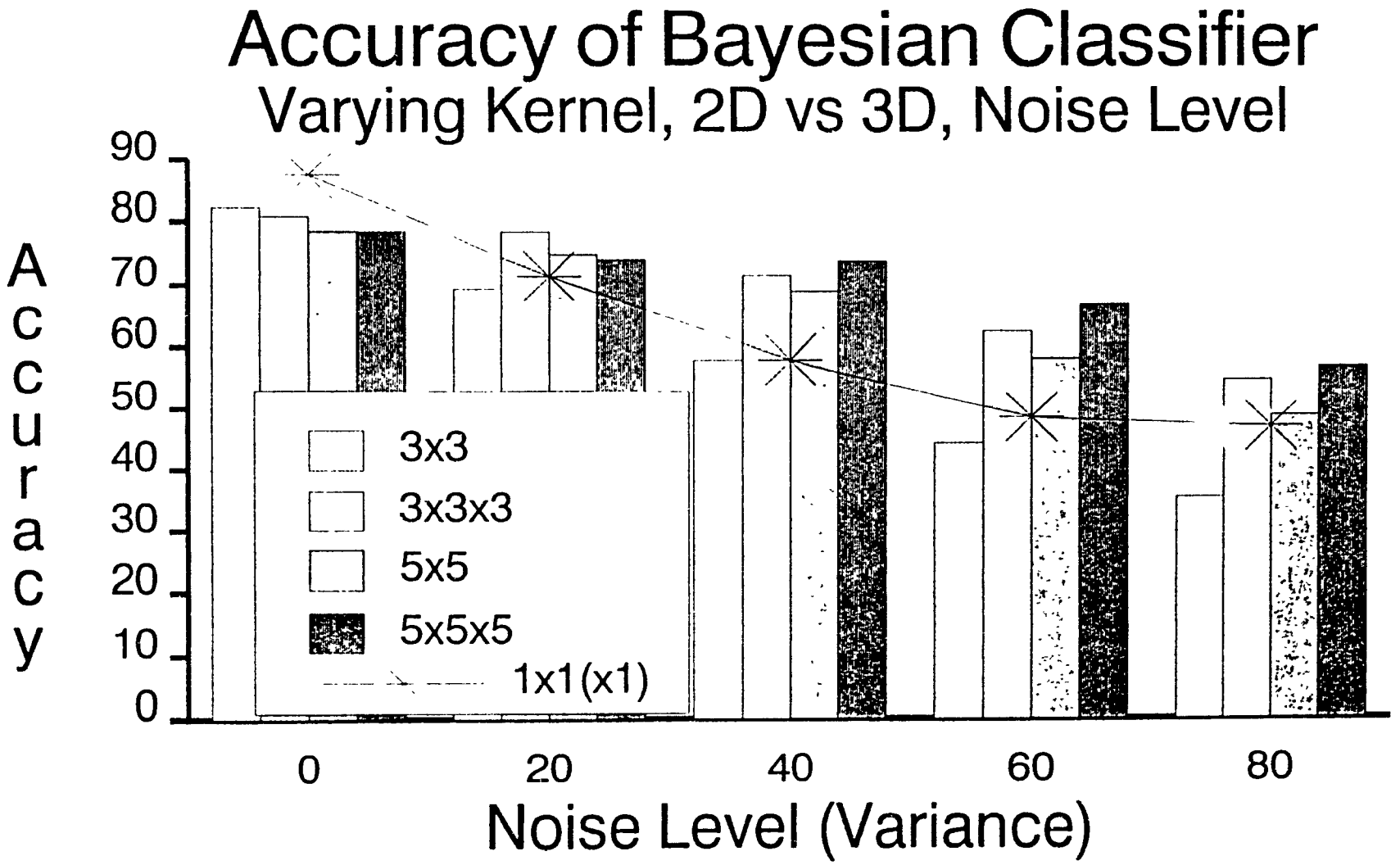


Figure 6.17: Bayesian classifer: Effect of kernel size on classification of noisy data.

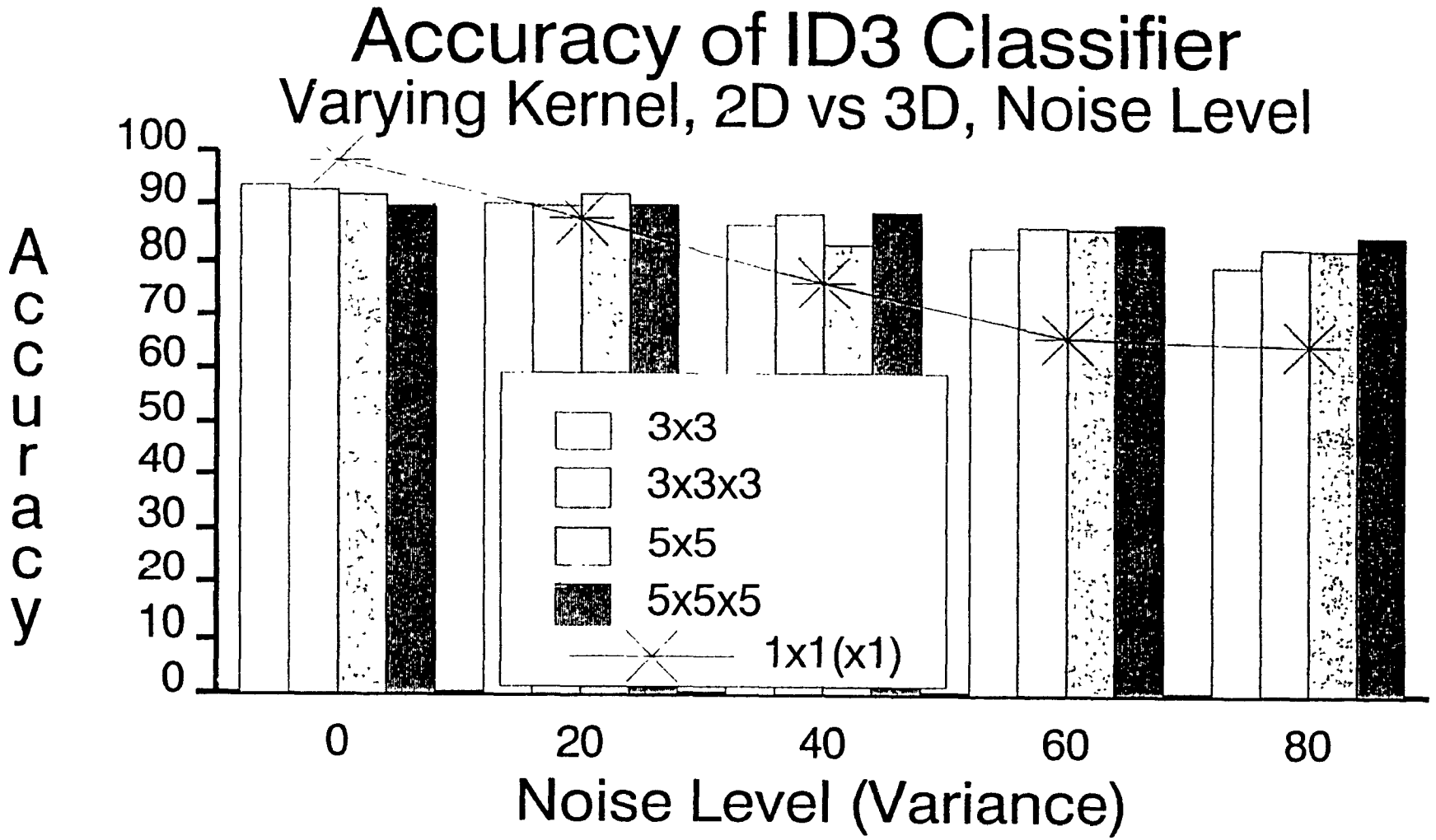


Figure 6.18: ID3: Effect of kernel size on classification of noisy data.

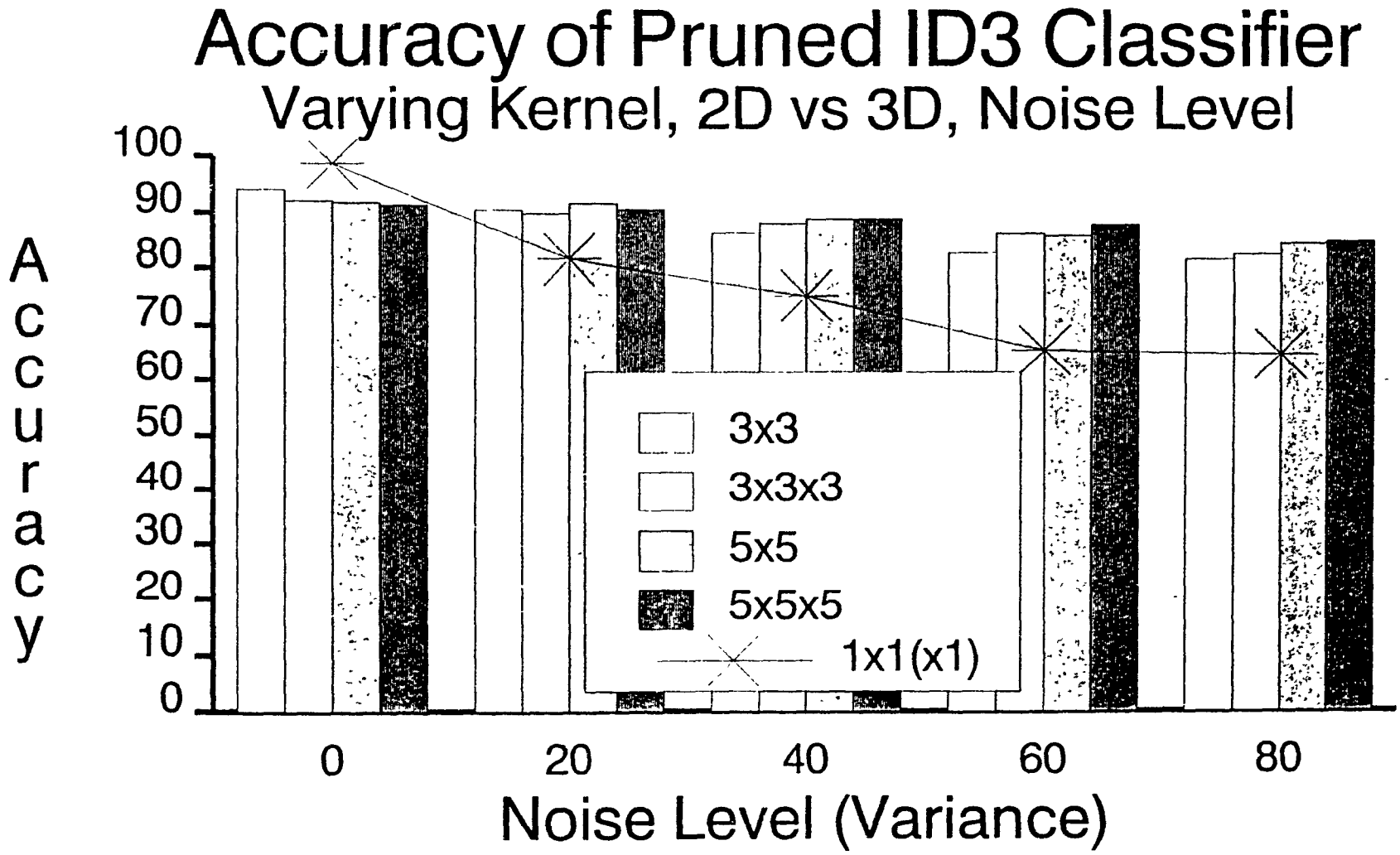


Figure 6.19: PruneID3: Effect of kernel size on classification of noisy data.

6.3.2 Results on Real MS Data

Figures 6.20 to 6.22 present examples of the results obtained in the segmentation of MR volumes of patients with multiple sclerosis. (Figures 6.20 and 6.21 are from the same volume). Each figure contains a series of four photographs. The first and second photographs correspond respectively to the T1-weighted (first echo) and T2-weighted (second echo) images of the given slice. The third photograph in each set shows classification results obtained in a purely data-driven manner; in these photographs, the tissue probability model was not used in any way. Voxels classified as MS lesion are displayed in dark brown. The fourth photograph shows the segmentation obtained when the model was used *a posteriori* to eliminate false positive lesions. In the third and fourth photographs of each set, the expert's manual segmentation of MS lesion is displayed on top of the obtained segmentation results.

The model was effective in eliminating false positive lesions (caused in majority by the RF inhomogeneity artifact). The elimination of false positive lesions is indicated with the 'F' pointers in the third photograph of Figures 6.20-6.22 and in their corresponding locations in the fourth photograph of each example. Although the use of the model has not eliminated all false positives (such as those at the 'E' pointers of Figures 6.20-6.22), it has reduced the number of mis-classifications considerably (quantitative measures are discussed below). Use of the model was successful in refusing proposed MS lesions in voxels corresponding to choroid plexus (not shown). The use of the model, however, caused the elimination of some true positive MS lesions occurring in the white matter tracts below the posterior horns of the lateral ventricles (see mis-classified portion of lesion in area indicated by pointer 'M' of Figure 6.20).

The automated segmentation was able to detect MS lesions which had not been included in the manual segmentation. This is indicated by the 'T' pointers in each of the Figures 6.20 to 6.22. Dr. Douglas Arnold, a neurologist at the Montreal Neurological Institute and Hospital, verified that the computer identification of

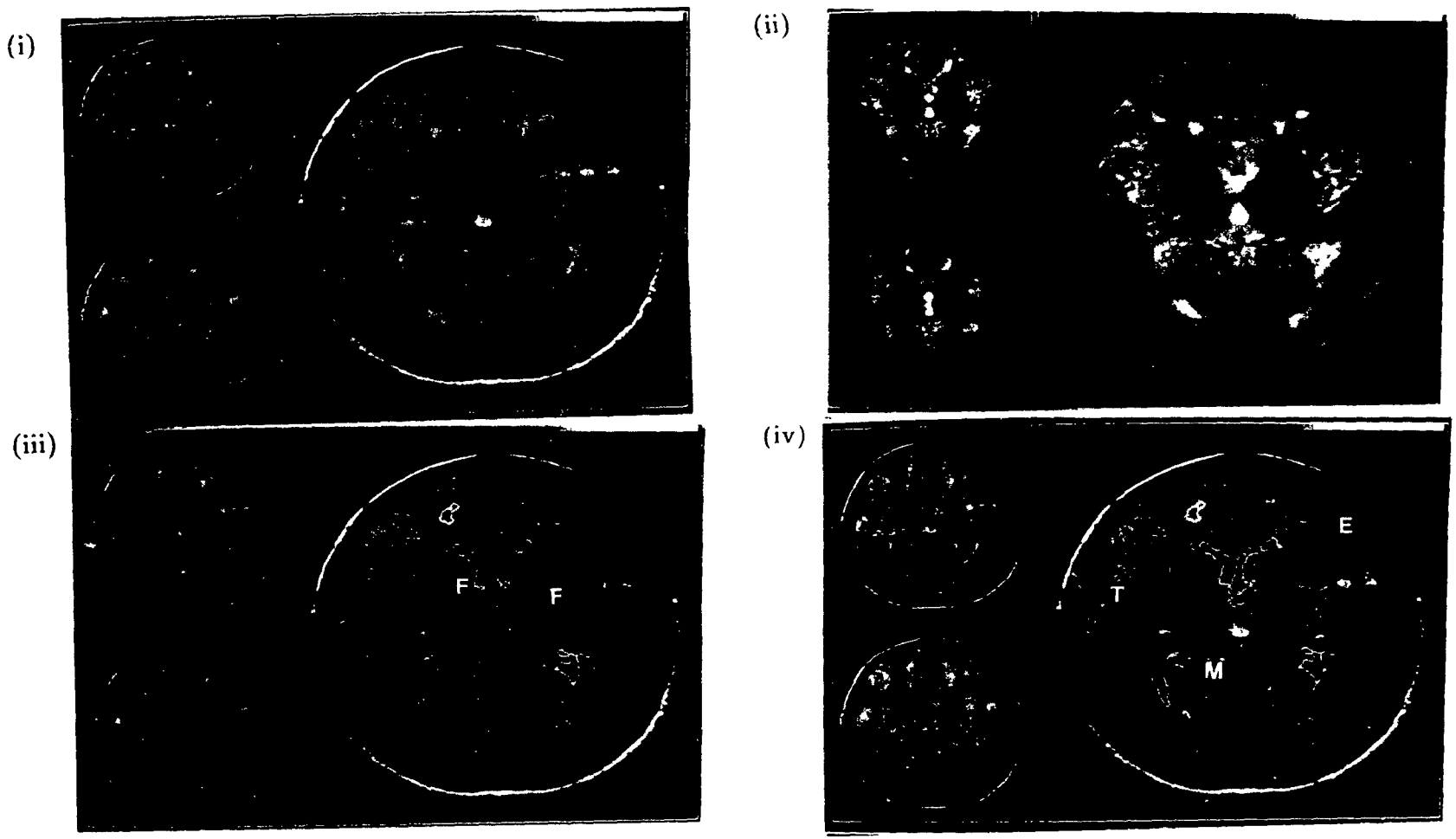


Figure 6.20: Example 1: Results of segmentation. (i) T1-weighted image. (ii) T2-weighted image. (iii) Data-driven segmentation result (without use of model). (iv) Segmentation result with model used a posteriori.
(The ventricles have also been manually outlined and appear pinkish). T = true positive MS. F = false positive MS. E = errors not corrected with model. M = error caused by model.

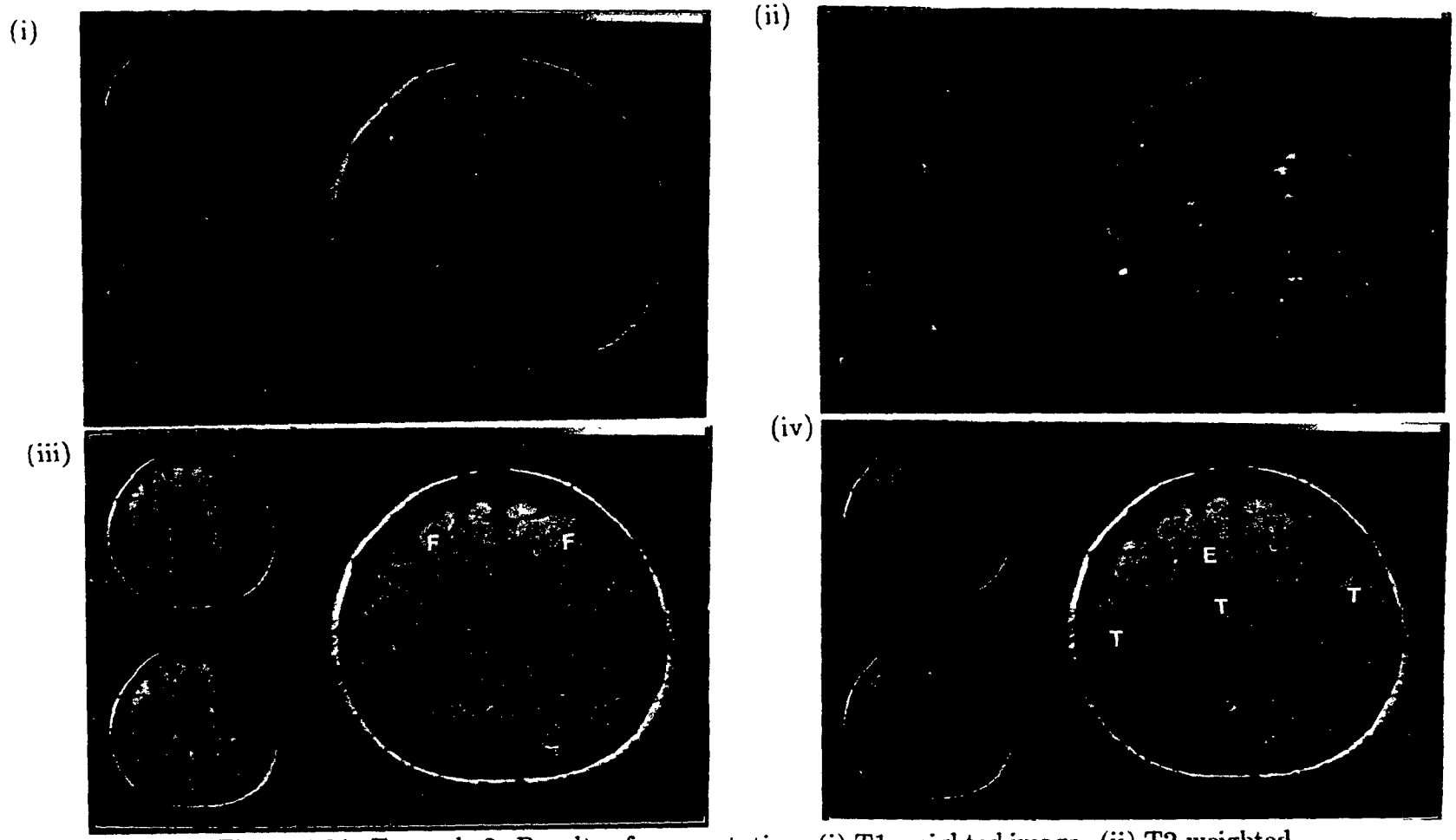


Figure 6.21: Example 2: Results of segmentation. (i) T1-weighted image. (ii) T2-weighted image. (iii) Data-driven segmentation result (without use of model). (iv) Segmentation result with model used a posteriori.
T = true positive MS, F = false positive MS, E = errors not corrected with model.

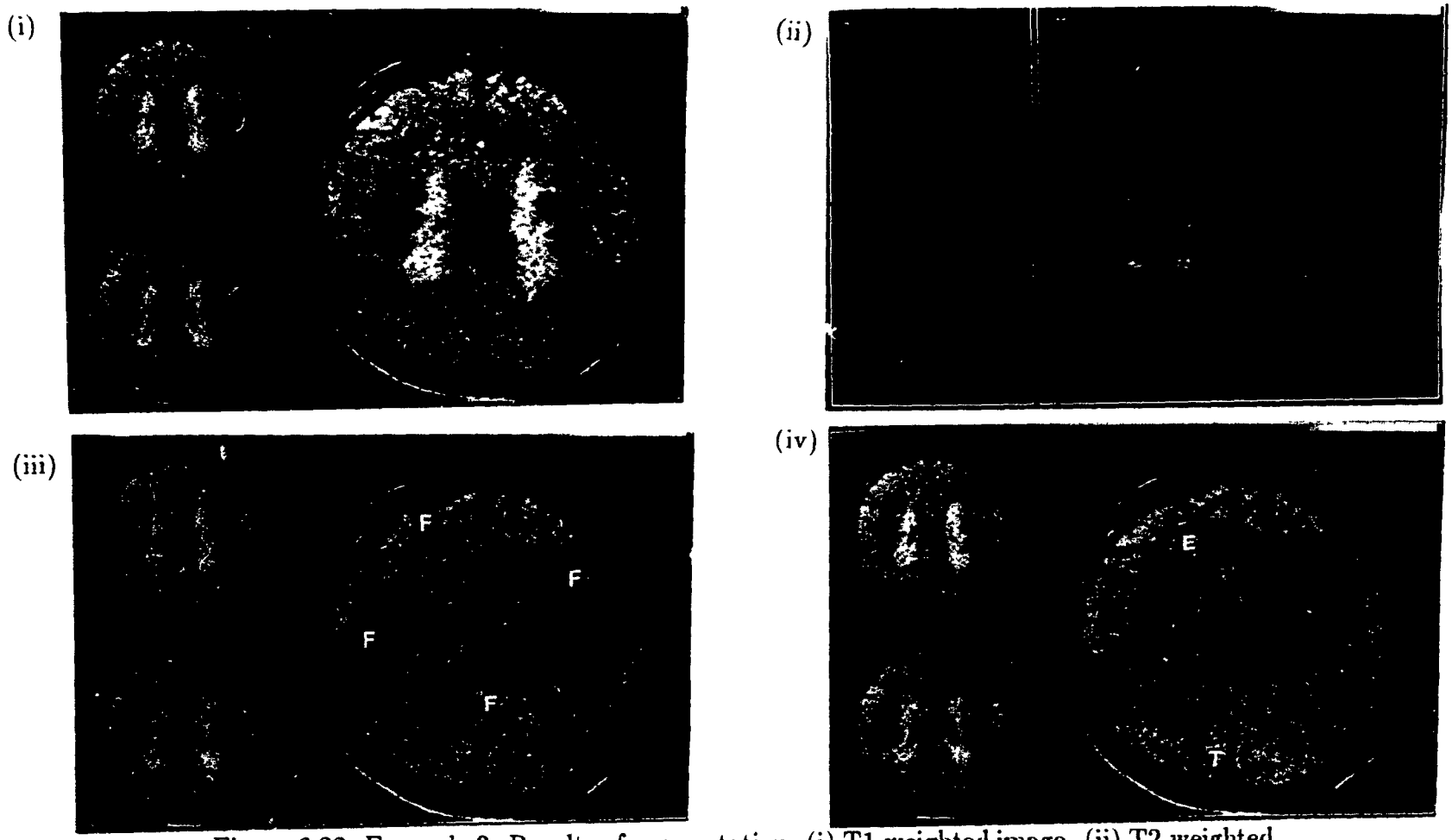


Figure 6.22: Example 3: Results of segmentation. (i) T1-weighted image. (ii) T2-weighted image. (iii) Data-driven segmentation result (without use of model). (iv) Segmentation result with model used a posteriori.
T = true positive MS, F = false positive MS, E = errors not corrected with model.

lesions at these pointers is correct. Dr. Arnold reviewed the segmentation results obtained with the use of the model and found the program to be 'more accurate and more consistent than manual segmentation has been or could be' (see Appendix for Dr. Arnold's evaluation of the segmentation tool). He notes that it is extremely tedious and difficult for humans to manually segment lesions, particularly in volumes where lesions are tiny and numerous.

Classifier Accuracy. Figure 6.23 shows the overall classification accuracy of the minimum distance, Bayesian, ID3, and PruneID3 classifiers. Purely data-driven classification (without any use of the model) resulted in recognition rates ranging from 88.96% to 92.51%. Use of the model *a posteriori* improved classification accuracy in each classifier by 3-5%. Recognition rates are also shown for classification where the model is used to provide features (*a posteriori* and non *a posteriori*). Using the model's probability values as features did not aid classification. As noted in the experiments on noise-free artificial data, the additional features cause the minimum distance classifier's accuracy to drop considerably. The accuracy of the Bayesian and decision tree classifiers remained the same. This shows the minimum distance classifier's dependence on good features. The pruning version of ID3 performed as well or slightly better (1-2%) than ID3. Pruning reduced the average number of leaves in decision trees from 38 to 8.

It is important to draw the increase in accuracy brought about by the *a posteriori* use of the model into perspective. Purely data-drive segmentation, with recognition rates of around 90%, would seem to be quite accurate. The quantity of actual MS lesions, however, is very small in comparison to brain size. The percentage of MS lesions in the brains of the patients of this experiment ranges from 3.00-4.23%. Therefore, recognition rates of 90% are not acceptable when detecting MS lesions. The challenge lies in overcoming mis-classifications in the remaining 10%. Thus, an increase of about 4% (when the model is used *a posteriori*) is considered to be a good improvement.

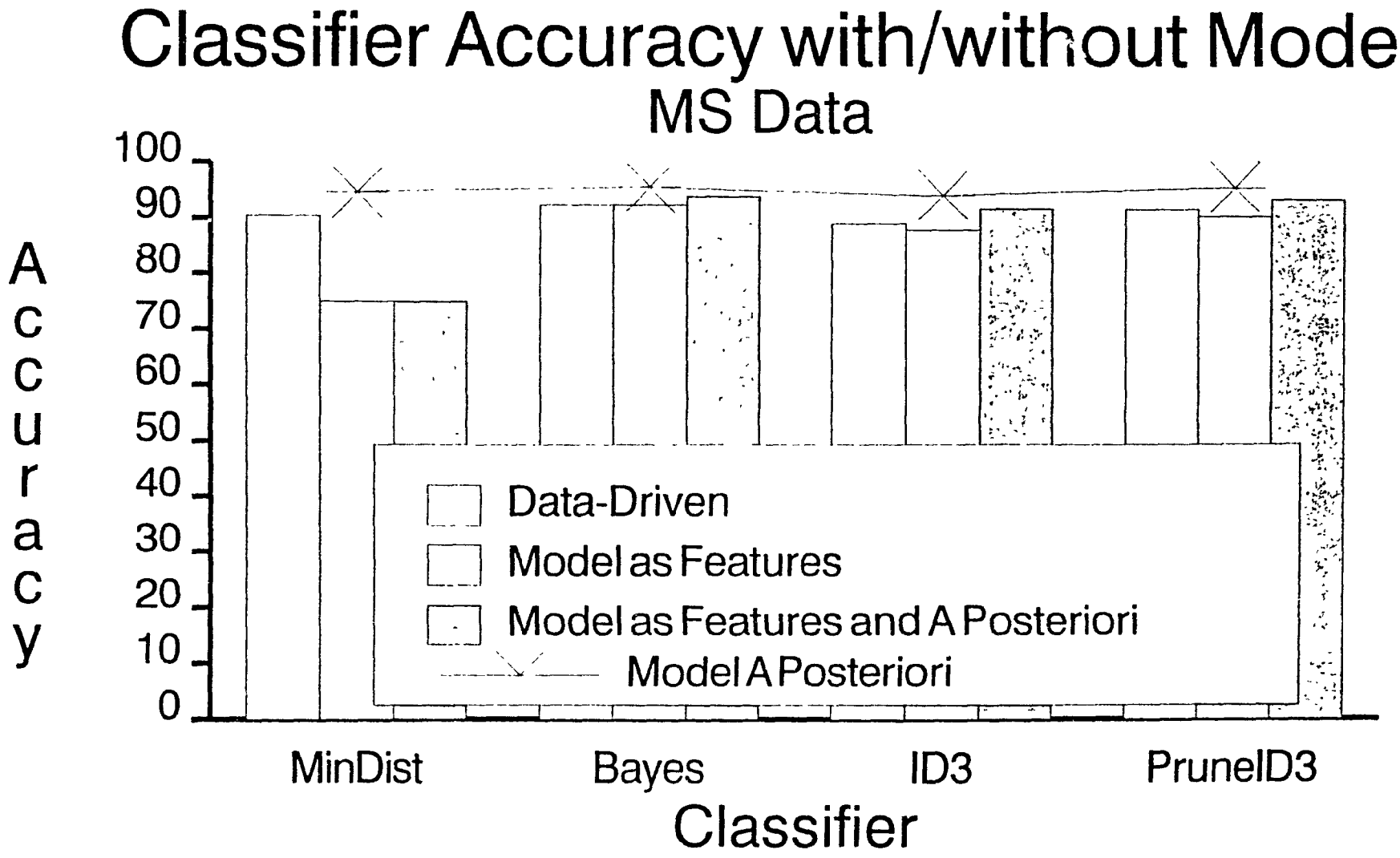


Figure 6.23: Classifier accuracy on MS data with and without use of model.

Sensitivity and Specificity. The overall sensitivity, specificity, and accuracy of each classifier is shown in Figure 6.24. Sensitivity refers to the recognition rate of the MS lesion samples and is defined as [Williams, 1987]:

$$Sensitivity = \frac{TP}{T_{MS}}$$

where TP is the total of true positives (samples of MS which are classified as such) and T_{MS} is the total number of MS samples.

Specificity refers to the recognition rate of non-MS samples and is defined as:

$$Specificity = \frac{TN}{T_{OTH}}$$

where TN is the total of true negatives (samples of 'other' classified as such) and T_{OTH} is the total of 'other' (non-MS) samples. It can be shown that accuracy is a function of sensitivity and specificity:

$$Accuracy = Sensitivity\frac{T_{MS}}{T_{MS}+T_{OTH}} + Specificity\frac{T_{OTH}}{T_{MS}+T_{OTH}}.$$

Use of the model *a posteriori* decreased the sensitivity of each classifier in detecting MS, but increased their specificity and accuracy. Specificity was improved by 4-6%.

Percentage of False Positive MS. Use of the model *a posteriori* reduced the number of false positive MS lesions (FPMS) by around 50% (Figure 6.25). The number of FPMS still largely exceeds the number of actual MS (the lowest percentage of false positive MS obtained was 124% for the pruned ID3 classifier using the model *a posteriori*). The recognition rates, however, were computed considering the manual segmentation to be the ground truth. It has been shown and verified that the manual segmentation incorrectly omits several examples of MS lesion. (The amount of MS lesions is underestimated). Due to inaccuracies in the manual segmentation, a number of voxels which the program had correctly identified as MS were incorrectly recorded as mis-classifications in the confusion matrix. Confusion matrices are shown in Figure 6.26 for the minimum distance classification of an MS volume with and without the use of the model *a posteriori* to reduce the number of false positive lesions.

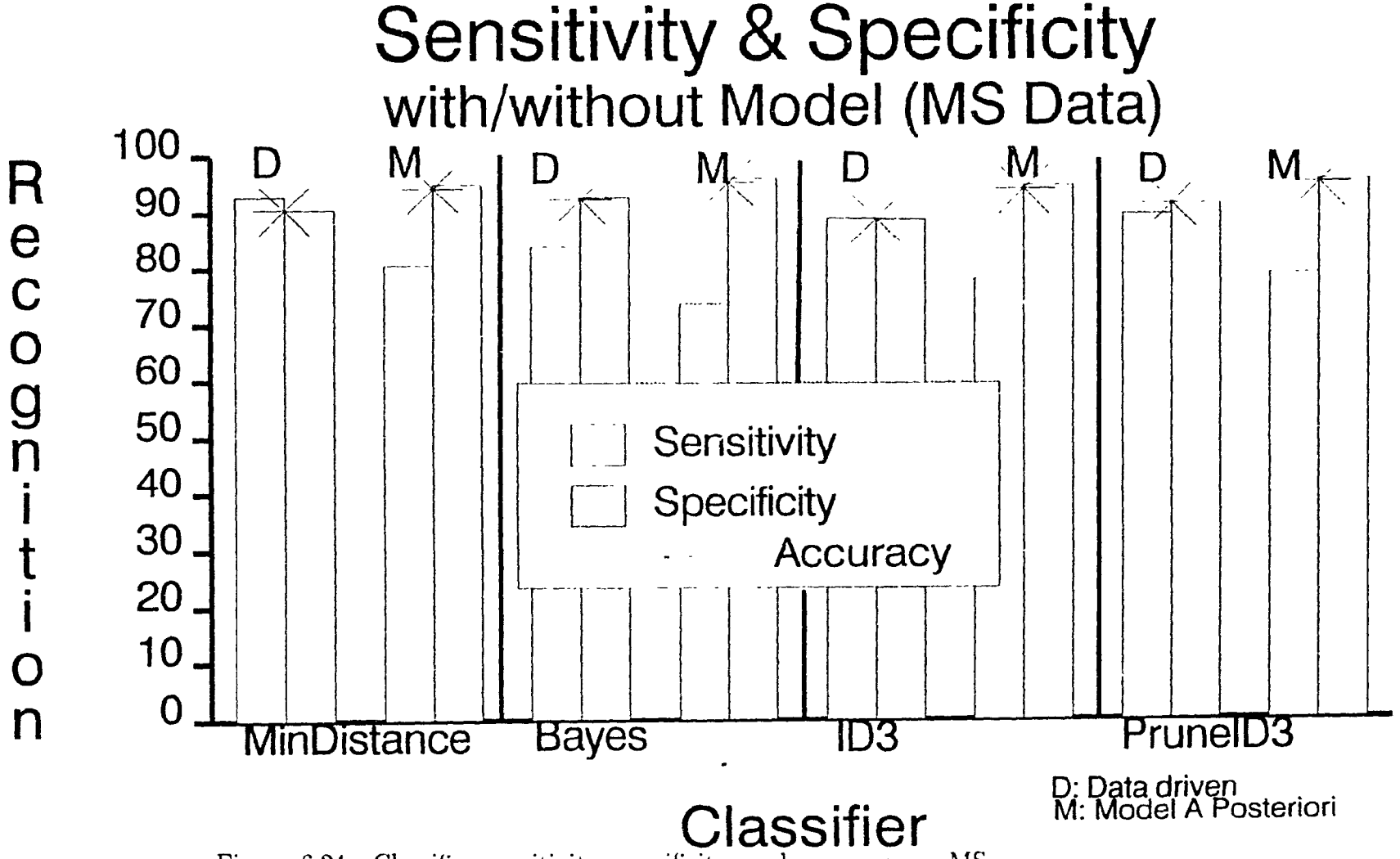


Figure 6.24: Classifier sensitivity, specificity, and accuracy on MS data with and without use of Model.

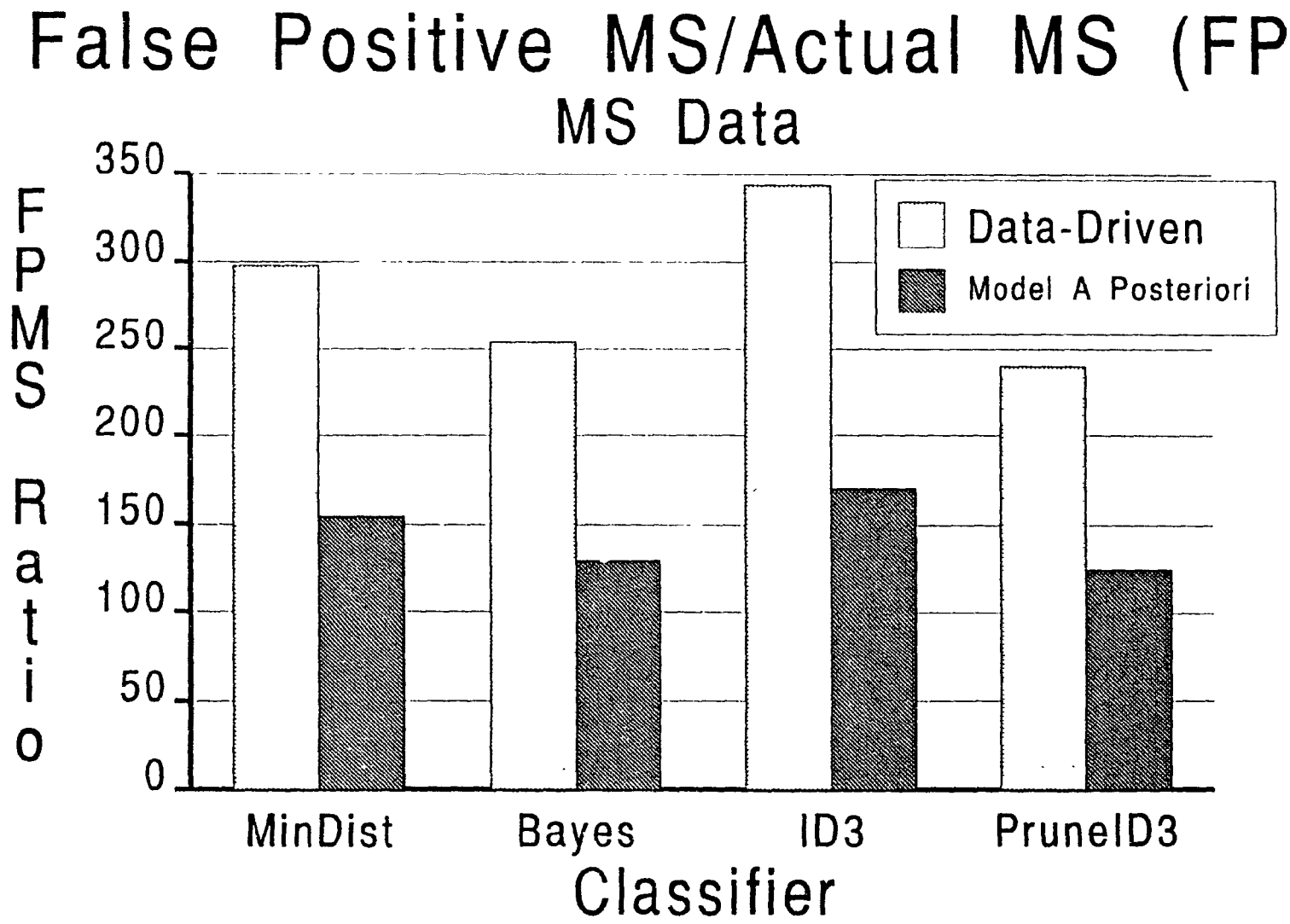


Figure 6.25: Percentage of false positive MS lesions to actual MS lesions for each classifier with and without use of the model.

Minimum Distance Classification of MS lesions without Use of Model

Classes	MS	Other	Error (%)	Recognition (%)
MS	3481	252	6.75	93.25
Other	14990	129211	10.40	89.60
Total	18471	129463	10.30	89.70

Minimum Distance Classification of MS lesions with Use of Model *A Posteriori* to Disallow Proposed Lesions in Voxels Where the Probability of White Matter is $< 50\%$

Classes	MS	Other	Error (%)	Recognition (%)
MS	3063	670	17.95	82.05
Other	7892	136309	5.47	94.53
Total	10955	136561	5.79	94.21

Figure 6.26: Confusion matrices for the minimum distance classification of an MS volume with and without the use of the model *a posteriori* to reduce the number of false positive lesions. (An entry in row i and column j indicates the number of samples of tissue type i which were classified as tissue type j). Use of the model improved specificity and accuracy (recognition rate) by around 5%. The number of false positive lesions decreased by over 50%.

Training Time. The overall training time in CPU seconds of each classifier, based on four features (dual-echo local mean and standard variance about a voxel neighborhood) for a training set with an average of 450 samples was compared. As with the experiments on the artificial data, the minimum distance and Bayesian classifiers had the lowest training times (0.019 and 0.050 CPU seconds respectively). On the average ID3 required 1.622 CPU seconds and PruneID3 required 7.822 CPU seconds. PruneID3's longer training time is attributed to the additional CPU seconds required to prune the decision trees.

Classification Time. The classifier with the fastest classification time was PruneID3 with an average classification time of 2.283 CPU seconds per slice (brain region only). The next fastest classifier was ID3 at 4.443 CPU seconds per slice. The slowest classifiers were the Bayesian and minimum distance classifiers, requiring on the average of 5.523 and 8.537 CPU seconds per slice respectively.

2D vs. 3D. Classification in 3D did not show a significant improvement over classification in 2D.

6.4 Discussion and Concluding Remarks

Experiments were conducted to study the usefulness of a tissue probability model in the segmentation of magnetic resonance images of the brain, particularly in the detection of multiple sclerosis lesions. The performance of a minimum distance and Bayesian classifier was compared to that of ID3 and an error-cost complexity pruning version of ID3.

Artificial Data. The first group of experiments was conducted on a set of artificial, computer-generated brain-MR-like volumes, each with a different level of uniform noise. The class label of each voxel of the artificial volumes was known, providing a means of quantitatively comparing the classifiers' accuracy, training and recall time, and robustness under varying noise conditions. Purely data-driven clas-

sification was compared to model-and- data-driven classification, where the tissue probabilities within the model were used per voxel as geometric or knowledge-based features in addition to statistical measures based on image grey scale intensity. Voxels were classified as either grey matter, white matter, CSF, MS lesion, or background. The findings were as follows:

1. The minimum distance classifier was overall the most accurate of the four algorithms tested.

2. The statistical classifiers' (minimum distance and Bayesian) use of the model as features is detrimental in the classification of data with *low* levels of noise. In contrast, the addition of the geometric features for the classification of such data has no effect on the decision tree algorithms. When the model was used to provide features, ID3 out performed the minimum distance classifier in the segmentation of clean data. This illustrates the decision tree algorithm's ability to select the more discriminating features. When a minimum distance classifier is presented with non-discriminating features (as is the case of the geometric features when used on clean data), it becomes confused. In this set of experiments, the classifier's recognition rate dropped 10%. This is similar to observations by Weiss and Kapouleas [1989] who note that the minimum distance classifier performs well when the features are good. Shepherd's [1983] comparison of the minimum distance, Bayesian, and ACLS (ID3 successor) classifiers also found the minimum distance algorithm to have the higher recognition rate.

3. In all other cases, the ID3 classifier and its pruning version were as accurate or as almost as accurate as the minimum distance classifier (within less than 5%).

4. The statistical classifiers had the faster training times, however, the decision tree classifiers were faster in recall by up to 2-14 CPU seconds per slice.

5. Use of the model did improve classification on the noisier data by 2-7%.

6. The extraction of features in 3D as opposed to 2D overall did improve recognition rates on noisy data.

7. The classification of clean data prefers a kernel size of 1x1 while classification of noisier data prefers larger kernel sizes of 3x3(x3) or 5x5(x5).

Real MS Data. A second set of experiments was conducted on actual MR volumes of two patients with multiple sclerosis. The MS lesions on each slice were manually outlined and used to evaluate the classifiers' accuracy. Voxels were classified as either MS or 'other'. The volumes were segmented under the following conditions:

- without the use of the tissue probability model (data-driven),
- with the model used to provide geometric features of tissue probability per voxel,
- with the model used *a posteriori* to disallow proposed MS lesions in anatomically implausible locations,
- with the model used to provide geometric features of tissue probability and *a posteriori* information.

The results were as follows:

1. The use of the model *a posteriori*, in comparison to the data-driven approach, decreased classifier sensitivity to MS lesions, yet increased specificity (4-6%) and accuracy (3-5%). The percentage of false positive MS lesions decreased by 50%. As the proportion of MS lesions within the brain is typically very small, this is considered to be a good improvement in classification.

2. All classifiers had similar recognition rates. Pruning improved accuracy of the ID3 classifier slightly, by 1-2%.

3. The minimum distance classifier, as observed in the experiment on artificial data, was confused by the addition of all five tissue masks as features. This shows the minimum distance classifier's dependence on good features. Use of the model to provide features did not aid classification. When this use was combined with the model *a posteriori*, the results were almost as good as those obtained when the model was used *a posteriori* with just the image grey scale features.

4. The segmentation tool, when employed with the model *a posteriori*, was found to be more accurate and more consistent than manual segmentation (see Appendix for the comments of Dr. Arnold, a neurologist at the M.N.I. who reviewed the segmentation results). Dr. Arnold noted [3] that the false positive lesions which are not eliminated with the use of the model are likely due to an overestimate of white matter within the mask. These mis-classifications should be reduced once inaccuracies in the white matter mask are corrected.

A method for the detection of MS lesions has been developed in this thesis. While the use of the tissue probability model appears promising, results should be validated with further experiments. This would depend, however, on the contribution of a team of experts involved in the manual segmentation or provision of training sets for several volumes of image data. Ideally the experts would segment or sample the same data more than once so that measurements of intra- as well as inter-observer variability can be considered in accessing the accuracy of the segmentation tool.

Interpretability of Classification Rules. On actual MR data, each classifier performed at about the same level of accuracy. A difference is seen in the interpretability of each classifier's learned rules. The minimum distance and Bayesian classifiers represent class descriptions with mathematical formulae (sections 5.2.1, 5.2.2). ID3 represents its acquired knowledge in the form of a decision tree. The

[3] Personal communication.

average decision tree ouput by the ID3 algorithm was approximately ten levels deep. The classification rules represented by such large decision trees are barely intelligible. However, the pruned decision trees obtained were much smaller. An example of a pruned tree for the classification of MS lesion and 'other' (non-MS) is shown in Figure 6.27. The features used were the dual-echo mean and standard deviation about the mean for a 3x3 window about each voxel. The pruned decision tree contains five nodes, three of which were leaves, and is two levels deep. Although a limited form of knowledge representation, the pruned decision tree is small enough to be easily understood.

The next chapter, which is the final chapter, summarizes the work of this thesis and discusses areas of future related work.

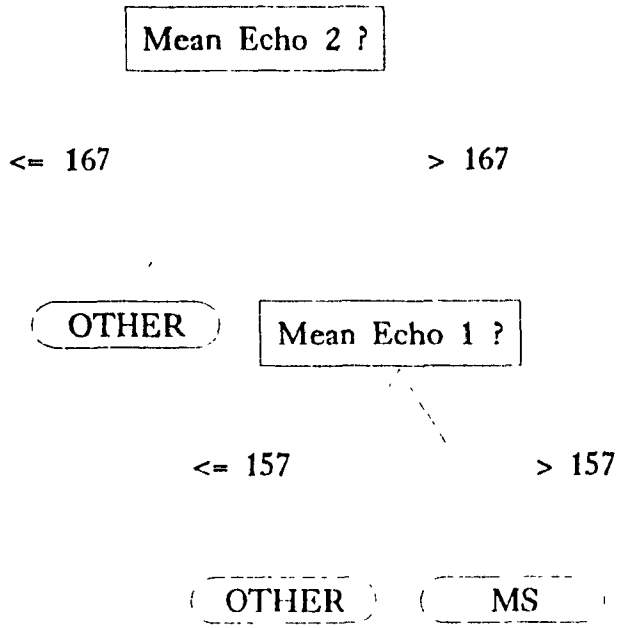


Figure 6.27: Pruned decision tree for the classification of MS lesion. Non-MS voxels are labeled as 'Other'.

Chapter 7

Conclusions and Future Related Work

7.1 Conclusions

A tool for the gross tissue segmentation of magnetic resonance images of the brain has been developed. In particular, the tool was designed for the automated detection of multiple sclerosis lesions. The goals of the thesis were twofold: to evaluate the effectiveness of incorporating knowledge of gross neuroanatomy in the segmentation task, and to compare the performance of statistical and machine learning classifiers applied to MR image segmentation. The classifiers under comparison were a minimum distance classifier, a Bayesian classifier, ID3 (a decision tree classifier), and a pruning version of ID3 for the handling of noisy data.

Knowledge of average brain anatomy was represented in the form of a tissue probability model giving for each voxel in a standardized 'brain space' (Talairach space), the respective probabilities of the voxel being of grey matter, white matter, ventricular CSF, external CSF, or background. The model was based on MR image data obtained from a group of twelve healthy volunteers. Brain image volumes for segmentation are first transformed into Talairach space in order to make use of the anatomical knowledge stored within the model.

The segmentation tool was designed to allow three possible uses of the tissue

probability model:

- to provide *a priori* tissue probabilities per voxel for the Bayesian classification of healthy brain tissues into grey matter, white matter, and CSF,
- to act as geometric features, providing heuristic information of probable tissue types per voxel location,
- to provide classifiers with *a posteriori* knowledge of probable multiple sclerosis lesion locations. Proposed MS lesions which do not occur in voxels with a probability of white matter (stored in the model) above a user-defined threshold can be disallowed.

The model was tested on the detection of MS lesions.

Effectiveness of Model. Experiments indicate that the use of the model improves segmentation results when compared to a purely data-driven approach. Use of the model *a posteriori* improved classification by 3-5% for each of the classifiers while increasing classifier specificity by 4-6% and reducing the number of false positive lesions by 50%. Use of the model to provide features improved accuracy only on extremely noisy data ($> 40\%$ noise).

Comparison of Statistical and Symbolic Learning Classifiers. When tested on real MR image data, each classifier performed at about the same level of accuracy. The statistical classifiers were the fastest in training, yet they were also the slowest in recall.

The decision tree classifier has an advantage over the statistical classifiers in providing a concise, 'human-like' explanation of each prediction. This may be easier to grasp then the mathematical formulae of minimum distance and Bayesian classification rules. The intelligibility of the decision trees, however, is difficult to measure, particularly in trees extending several levels deep.

The automated segmentation, regardless of the classifier used, was able to cor-

rectly identify MS lesions which were not included in the manual segmentation. Dr. Arnold, a neurologist at the M.N.I., reviewed the results of the segmentation tool obtained with the use of the model *a posteriori* and found the automated classification to be more accurate and consistent than manual segmentation (Appendix).

7.2 Future Related Work

This section discusses future research related to the further development of the segmentation tool of this thesis. Research directives can be divided into two categories:

- work concerning improvements and applications of the model, and
- work concerning machine learning and knowledge representation in general.

Future Developments and Applications of the Model. The model can be refined. First and foremost, its representation of gross neuroanatomy should be verified and corrected by an individual qualified in neuroanatomy. Present areas of error within the model include an overestimation of white matter of the gyri, particularly towards the top of the head. Likewise, the model incorrectly indicates a presence of periventricular grey matter. These errors are due to the partial volume effect.

The tissues of the model can be subdivided to provide more specific anatomical information. Ninety to ninety-five percent of MS lesions occur in white matter tissue. The majority of the remaining MS lesions can occur within the basal ganglia, a type of grey matter tissue located near the ventricles. Hence, further subdivision of the grey matter mask to distinguish basal ganglia should increase the recognition rate for lesions within this area. Similarly, the white matter mask can be partitioned to indicate the more probable locations for the occurrence of MS lesions, observed from the segmentation of a group of brain image volumes displaying the disease.

In subdividing the tissue model to represent the presence of pathologies, insight into the involvement of neurological diseases can be studied. For instance, Lim and Pfefferbaum [1989] note that although the characteristic lesions of Alzheimer disease are typically found in neocortical and subcortical limbic grey matter, pathological data has indicated white matter involvement in the disease. The further refinement of the model may aid in the segmentation of pathological image data, contributing to the study of disease.

The model could be extended to represent a brain atlas of structures defined by their functionality, as well as anatomy. Functional activity, such as glucose metabolism, is not necessarily mapped to the borders of anatomical structures. While MR images are well suited for imaging anatomy, a technique known as positron emission tomography is used to measure functionality in terms of energy metabolism and regional hemodynamics. Using data from positron emission scans, functional activity can be measured and incorporated into the model. Thus, in addition to the tissue classification of brain MR image data, future work can involve the model's use in the segmentation of functional areas. Furthermore, the model can be used to study and measure the variability in cerebral structures amongst individuals [von Keyserlingk, 1988].

The experiments of this thesis used the model *a posteriori*. A threshold was selected regarding the white matter tissue mask. Proposed MS voxels whose corresponding model white matter probability was below the threshold were disallowed and relabeled as 'other'. An alternative method would be to vary the threshold for different regions within the model. For example, a hyperintense voxel A within the white matter of a gyrus may have a white matter probability (from the model) of p. A hyperintense voxel B near the ventricles may also have a white matter probability of p. Neurologists are more reluctant, however, in accepting voxel A as lesion, as lesion is less likely to occur in that area. Setting a higher white matter threshold for proposed MS lesions occurring near the cortical rim would be similar to the reasoning process employed by neurologist and radiologists in this situation.

The model should also be tested on the segmentation of healthy brain tissues. The probability values within the model of tissue distribution can be used as prior probabilities in Bayesian classification.

Errors may occur in segmentation due to an improper fitting of individual brains to the tissue probability model in Talairach space. At present, the transformation into Talairach space is implemented as a linear function. A more accurate transformation would be non-linear, involving the input of several more landmarks. Work is currently being done in identifying possible landmarks [Evans et al., 1991].

Future Development Regarding Learning and Knowledge Representation. Research is proposed regarding machine learning and knowledge representation. Such work can be applied to the further development of the learning component of the segmentation tool.

It would be interesting to compare the performance of other classifiers to those studied here. Most researchers seem to believe that the ultimate learning algorithm will involve a combination of explanation-based, empirically-based, and analogical algorithms [DeJong et al., 1986]. Dietterich [1990] expects that a great deal of future research will be aimed towards the development of hybrid learning methods. Utgoff [1988b], for example, proposed a perceptron tree hybrid where leaf nodes are perceptrons and internal nodes are standard decision trees. The connectionist backpropagation algorithm has been observed to perform more accurately than ID3 when classifying noisy data although it requires longer training and recall times [Mooney et al., 1989]. Symbolic learning approaches, such as ID3, have an advantage over connectionist learning in that they tend to require fewer training examples, can be less computationally expensive, and can explain their reasoning. Therefore it is of interest to develop a hybrid learning algorithm which exploits the respective strengths of the connectionist and symbolic learning approaches. When applied to magnetic resonance image segmentation, it would be interesting to study how such a system could use a voxel-based model of *a priori* tissue probability to advantage.

Breiman et al. [1984] proposed a decision tree algorithm which incorporates prior probabilities. Such an algorithm could perhaps be combined with a connectionist form of learning.

Unsupervised algorithms should also be explored as they remove subjective bias and decrease the amount of interaction required by the user in training. Unsupervised learning may reveal characteristics in the data which were unobserved by humans [Gerig et al., 1991].

Future work in needed in finding representation structures for the encoding of more complex forms of knowledge. Decision trees are a limited form of knowledge and do not provide a very compact representation for Boolean concepts in disjunctive normal form [Dietterich, 1990]. With respect to the segmentation tool, a knowledge base facility could be set up to allow experts performing manual segmentation of MS lesions to explain their reasoning behind the inclusion and omission of hyperintense regions. The representation and use of such knowledge, in addition to anatomical models and image grey scale statistics, should result in improved classification.

Further work can be done regarding the association of uncertainties with class predictions. This topic is particularly important in the segmentation of MR images due to the inherent 'partial volume' nature of the images, where individual voxels can be composed of more than one tissue type. Uncertainties can be used to approximate the distribution of tissues within a voxel.

The segmentation tool should be extended to allow the generalization from volume to volume without the need for retraining. This is considered as not currently possible [Katz and Merickel, 1989] as intensity values for volumes of the same patient with identical acquisition parameters can vary for like tissues. When approached as a classification problem in itself, perhaps this problem of 'volume generalization' can be solved. Features for an image volume may include acquisition parameters as well as the age, gender, and known pathologies of the patient.

7.3 Concluding Remarks

The goals of this thesis were the following:

1. To develop a tool for the segmentation of MR images, particularly for the detection of multiple sclerosis lesions. This goal has been realized. The segmentation tool can be used in a data or model-driven manner.

2. To evaluate the effectiveness of a brain tissue probability model in the detection of MS lesions. The model was developed and stores the individual probabilities of grey matter, white matter, ventricular CSF, and external CSF for each voxel in a standardized 3D brain space, based on a group of heathly volunteers. When used *a posteriori* in segmentation to disallow proposed lesions in implausible white matter areas, the model was found to improve accuracy by 3-5%. The number of false positive MS lesions was cut in half. The segmentation tool's overall performance was found to be more accurate and more consistent than manual segmentation.

3. To compare the performance of the statistical minimum distance and Bayesian classifiers with that of the symbolic ID3 learning algorithm. On actual MR image data, each classifier performed at about the same level of accuracy. The statistical classifiers were faster in training, while the decision tree classifier and its pruning version were faster at recall. The ID3 classifier expresses its classification rules explicitly in the form of a decision tree. Although decision trees provide a limited form of knowledge representation (particularly when the trees are deep), pruned trees are fairly small and understandable.

Future related research may focus on the refinement of the model and the ways in which it can be used in classification. The development of more sophisticated learning systems, such as hybrid methods combining connectionist and symbolic strategies, may be investigated. The development of more encaptive forms of knowledge

representation schemes capable of encoding complex forms of reasoning is another area of work. For example, manual segmentations typically performed by experts to serve as a means for measuring a tissue classifier's accuracy may be accompanied by explanations of the reasoning behind the outlining of each lesion and omission of other lesion-like regions. This knowledge, in addition to models of anatomy and image grey scale information, could be used to improve classifier accuracy.

Artificial intelligence in medical applications, such as the segmentation of magnetic resonance images, is valuable to research in computer science. It encompasses a large and interesting domain of real-world problems for which AI theories can be tested, providing insight into future AI research as well as the development of useful clinical tools.

References

[1] *Introduction to MR Imaging.* Philips Medical Systems, The Netherlands, 1984.

[2] David Y. Amamoto, Rangachar Kasturi, and Alexander Mamourian. Tissue type discrimination in magnetic resonance images. In *Proceedings of the 10th International Conference on Pattern Recognition*, pages 603–607, IAPR, Atlantic City, NJ, June 1990.

[3] Leon Axel, Jay Costantini, and John Listerud. Intensity correction in surface-coil MR imaging. *American Journal of Roentgenology*, 148:418–420, Feb 1987.

[4] N. Ayache, J. D. Boissonnat, E. Brunet, L. Cohen, J. P Chieze, B. Geiger, O. Monga, J. M. Rocchisani, and P. Sander. Computer assisted neurology. In H. U. Lemke, M. L. Rhodes, C. C. Jaffe, and R. Felix, editors, *Proceedings of the 3rd International Symposiom on Computer Assisted Radiology*, pages 765–772, CAR, 1989.

[5] Michael Bomans, Karl-Heinz Hohne, Ulf Tiede, and Martin Reimer. 3D segmentation of MR images of the head for 3D display. *IEEE Transactions on Medical Imaging*, 9(2):177–183, 1990.

[6] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classification and Regression Trees.* Wadsworth, Monterrey, CA, 1984.

[7] B. G. Buchanan and T. M. Mitchell. *Model-Direction Learning of Production Rules.* Academic Press, New York, NY, 1978.

[8] Gregory D. Cascino, Clifford R. Jack Jr., Joseph E. Parisi, Frank W. Sharbrough, Kathryn A. Hirschorn, Frederic B. Meyer, W. Richard Marsh, and Peter C. O'Brien. Magnetic resonance imaging-based volume studies in temporal lobe epilepsy: Pathological correlations. *Annals of Neurology*, 30:31-36, 1991.

[9] F. Cedes, F. Andermann, C. Watson, A. Evans, P. Gloor, D. Melanson, M. J. Gotman, G. Leroux, A. Olivier, and T. Peters. Volumetric measurements of amygdaloid body and hippocampal formation in temporal lobe epilepsy. *Neurology*, submitted Sept 1991.

[10] J. Cendrowska. PRISM: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, 27:349-380, 1987.

[11] P. Clark. Machine learning: Techniques and recent developments. In A. R. Mirzai, editor, *Artificial Intelligence: Concepts and Applications in Engineering*, pages 65–93, MIT Press, Cambridge, MA, 1990.

[12] Harvey E. Cline, William E. Lorensen, Ron Kikinis, and Ferenc Jolesz. Three-dimensional segmentation of MR images of the head using probability and connectivity. *Journal of Computer Assisted Tomography*, 14(6):1037 1045, Nov/Dec 1990.

[13] Paul R. Cohen and Edward A. Feigenbaum, editors. *The Handbook of Artificial Intelligence.* Volume 3, Heuris Tech Press, Los Altos, California, 1982.

[14] Robert Dann, John Hoford, Steve Kovacic, Martin Reivich, and Ruzena Bajcsy. Evaluation of an elastic matching system for anatomic (CT/MR) and functional (PET) cerebral images. *Journal of Computer Assisted Tomography*, 13(4):603–611, July/Aug 1989.

[15] Benoit M. Dawant, Richard A. Margolin, Mehmed Ozkan, Hiroshi Aramata, and K. Kawamura. A neural network approach to the detection of white matter lesions in magnetic resonance images. In *Proceedings of the 9th Annual Scientific Meeting and Exhibition of the Society of Magnetic Resonance in Medicine, Book of Abstracts*, page 99, SMRM, New York, NY, Aug 1990.

[16] G. DeJong and R. Mooney. Explanation-based learning: An alternative view. *Machine Learning*, 1:145–176, 1986.

[17] T. G. Dietterich. Machine learning. *Annual Reviews in Computer Science*, 4:255-306, 1990.

[18] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, NY, 1973.

[19] T. Ellman. Explanation-based learning: A survey of programs and perspectives. *ACM Computing Surveys*, 21(2):163 - 221, June 1989.

[20] Tapio Elomaa and Niklas Holsti. An experimental comparison of inducing decision trees and decision lists in noisy domains. In Katharina Morik, editor, *Proceedings of the 4th European Working Session on Learning*, pages 59–69, Montpellier, France, Dec 1989.

[21] A. C. Evans, W. Dai, L. Collins, P. Neelin, and S. Marrett. Warping of a computerized 3D atlas to match brain image volumes for quantitative neuroanatomical and functional analysis. In *S.P.I.E: Medical Imaging V*, 1991.

[22] R. T. Fan, S. S. Trivedi, L. L. Fellingham, and A. Gamboa-Aldeco. Soft tissue segmentation and 3D display from computerized tomography and magnetic resonance-imaging. In *Medical Imaging*, pages 494–504, SPIE, 1987.

[23] E. A. Feigenbaum. Expert systems in the 1980s. In A. Bond, editor, *State of the Art Report on Machine Intelligence*, Pergamin-Infotech, Maidenform, 1981.

[24] E. A. Feigenbaum. The simulation of verbal learning behavior. In *Proceedings of the Western Joint Computer Conference*, pages 121–132, 1961.

[25] Douglas H. Fisher. Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2:139–172, 1987.

[26] Gordon S. Francis, Alan C. Evans, Larry Baer, Micheline Kamber, Louis Collins, and Jack P. Antel. Effect of MRI slice thickness on lesion volume determination in mult ple sclerosis. *Neurology*, submitted Sept 1991.

[27] Michael R. Genesereth and Nils J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, Palo Alto, CA, 1987.

[28] Guido Gerig, John Martin, Ron Kikinis, Olaf Kubler, Martha Shenton, and Ferenc A. Jolesz. Automating segmentation of dual-echo MR head data. In A. C. F. Colchester and David J. Hawkes, editors, *Information Processing in Medical Imaging, 12th International Conference*, pages 175 187, IPMI, Springer-Verlag, London, UK, July 1991.

[29] David A. Handelman, Stephen H. Lane, and Jack J. Gelfand. Integration of knowledge-based system and neural network techniques for autonomous learning machines. In *Proceedings of the INNS International Joint Conference on Neural Networks*, pages 683–688, IJCNN, 1989.

[30] Timothy J. Hyman, Robert J. Kurland, George C. Levy, and Jon D. Shoop. Characterization of normal brain tissue using seven calculated MRI parameters and a statistical analysis system. *Magnetic Resonance in Medicine*, 11:22–34, 1989.

[31] Terry L. Jernigan, Gary A. Press, and John R. Hesselink. Methods for measuring brain morphologic features on magnetic resonance images. *Archives of Neurology*, 47:27–32, Jan 1990.

[32] Ioannis Kapouleas. Automatic detection of white matter lesions in magnetic resonance imaging. *Computer methods and Programs in Biomedicine*, 32:17–35, 1990.

[33] Ioannis Kapouleas. Segmentation and feature extraction for magnetic resonance brain image analysis. In *Proceedings of the 10th International Conference on Pattern Recognition*, pages 583–590, Atlantic City, NJ, 1990.

[34] William T. Katz and Michael B. Merickel. Translation-invariant aorta segmentation from magnetic resonance images. In *Proceedings of the International Conference on Neural Networks*, pages 327–333, IEEE INNS, Washington, DC, 1989.

[35] David N. Kennedy, Pauline A. Filipek, and Verne S. Caviness Jr. Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Transactions of Medical Imaging*, 8(1):1–7, March 1989.

[36] Mark I. Kohn. Segmentation of magnetic resonance brain images. In *Proceedings of the 10th Annual Conference and Exposition Dedicated to Computer Graphics*, pages 168–171, NCGA, Philadelphia, PA, 1989.

[37] Olaf Kubler and Guido Gerig. Segmentation and analysis of multidimensional datasets in medicine. In K. H. Holme, H. Fuchs, and S. M. Pizer, editors, *3D Imaging in Medicine*, pages 63–79, Springer-Verlag, London, 1990.

[38] P. D. Laird. A survey of computational learning theory. In R. B. Banerji, editor, *Formal Techniques in Artificial Intelligence*, pages 173–215, Elsevier Science B.V., North Holland, 1990.

[39] Y. Le Cun, O. Matan, B. Boser, J. S. Denker, D. Henderdon, R. E. Howard, W. Hubbard, L. D. Jackel, and H. S. Baird. Handwritten zip code recognition with multilayer networks. In *Proceedings of the 10th International Conference on Pattern Recognition*, pages 35–39, IAPR, Atlantic City, NJ, June 1990.

[40] D. B. Lenat. The role of heuristics in learning by discovery: Three case studies. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 243 306, Tioga, Palo Alto, CA, 1983.

[41] D. B. Lenat. The ubiquity of discovery. *Artificial Intelligence*, 9:257 285, 1977.

[42] David N. Levin, Xiaoping Hu, Kim K. Tan, Simranjit Galhotra, Andreas Herrmann, George T.Y. Chen, Charles A. Pelizzari, James Balter, Robert N. Beck, Chin-Tu Chen, and Malcolm D. Cooper. Integrated 3D display of MR, CT, and PET images of the brain. In *Proceedings of the National Computer Graphics Association*, pages 179–186, Philadelphia, PA, April 1989.

[43] Martin D. Levine. *Vision in Man and Machine*. McGraw-Hill, New York, NY, 1985.

[44] Kelvin O. Lim and Adolf Pfefferbaum. Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. *Journal of Computer Assisted Tomography*, 13(4):588–593, 1989.

[45] Kenneth R. Maravilla. Multiple sclerosis. In David D. Stark and William G. Bradley, editors, *Magnetic Resonance Imaging*, pages 344 358, C. V. Mosby Company, St. Louis, 1988.

[46] H. P. Meinzer, U. Engelmann, D. Scheppelmann, and R. Schafer. Volume visualization of 3D tomographies. In K. H. Hohne, H. Fuchs, and S. M. Pizer, editors, *3D Imaging in Medicine - Algorithms, Systems, Applications*, pages 253–259, Springer-Verlag, 1990.

[47] Wido Menhardt. *Iconic Fuzzy Sets for MR Image Segmentation*. Technical Report, Philips Research Labs, Hamburg, Germany, 1988.

[48] Wido Menhardt and Karl-Heinrich Schmidt. Computer vision on magnetic resonance images. *Pattern Recognition Letters*, 8:73–85, Sept 1988.

[49] R. S. Michalski. Understanding the nature of learning: Issues and research directions. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 3–25, Morgan Kaufmann, Los Altos, CA, 1986.

[50] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors. *Machine Learning: An Artificial Intelligence Approach.* Volume 1, Tioga, Palo Alto, CA, 1983.

[51] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors. *Machine Learning: An Artificial Intelligence Approach.* Volume 2, Morgan Kaufmann, Los Altos, CA, 1986.

[52] R. S. Michalski and R. L. Chilausky. Learning by being told and learning from examples: An experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis. *Policy Analysis and Information Systems*, 4:125–160, 1980.

[53] R. S. Michalski and R. E. Stepp. Learning from observation: Conceptual clustering. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 331–363, Tioga, Palo Alto, CA, 1983.

[54] Ryszard. S. Michalski. A theory and methodology of inductive learning. *Artificial Intelligence*, 20:111–161, 1983.

[55] John Mingers. An empirical comparison of pruning methods for decision-tree induction. *Machine Learning*, 4(3):227–243, 1989.

[56] John Mingers. An empirical comparison of selection measures for decision-tree induction. *Machine Learning*, 3(4):319–342, 1989.

[57] M. L. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry.* MIT Press, Cambridge, MA, 1969.

[58] T. M. Mitchell. Generalization as search. *Artificial Intelligence*, 18:203 226, 1982.

[59] T. M. Mitchell. *The Need for Biases in Learning Generalizations.* Technical Report CBM-TR-117, Rutgers University, Department of Computer Science, New Brunswick, NJ, 1980.

[60] T. M. Mitchell, R. M. Keller, and S. T. Kedar-Cabelli. Explanation-based learning: A unifying view. *Machine Learning*, 1(1):47 80, 1986.

[61] Raymond Mooney, Jude Shavlik, Geoffrey Towell, and Alan Grove. An experimental comparison of symbolic and connectionist learning algorithms. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 775-787, Morgan Kaufmann, Detroit, MI, Aug 1989.

[62] D. J. Mostow. *Mechanical Transformation of Task Heuristics into Operational Procedures.* Technical Report CMU-CS-81-113, Carnegie-Mellon University, Computer Science, Pittsburgh, PA, 1979.

[63] Tim Niblett and Ivan Bratko. Learning decision rules in noisy domains. In M. A. Bramer, editor, *Expert Systems '86: Research and Development in Expert Systems III*, pages 25-34, British Computer Society Specialist Group on Expert Systems, Dec 1986.

[64] Mehmed Ozkan, Hendrick G. Sprenkels, and Benoit M. Dawant. Multispectral magnetic resonance image segmentation using neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, pages 429–434, San Diego, CA, June 1990.

[65] G. Pagallo. Learning DNF by decision trees. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 639 644, IJCA,

Morgan-Kaufmann, San Mateo, CA, 1989.

[66] T. M. Peters. Principles and applications of magnetic resonance imaging (MRI) in neurology and neurosurgery. *Journal of Mind and Behavior*, 9(3):241-262, summer 1988.

[67] Adolf Pfefferbaum, Leslie M. Zatz, and Terry L. Jernigan. Computer-interactive method for quantifying cerebrospinal fluid and tissue in brain CT scans: Effects of aging. *Journal of Computer Assisted Tomography*, 10(4):571-578, July/Aug 1986.

[68] Bruce A. Porter, William Hastrup, Michael L. Richardson, George E. Wesbey, Dana O. Olson, Laurence D. Cromwell, and Albert A. Moss. Classification and investigation of artifacts in magnetic resonance imaging. *RadioGraphics*, 7(2):271-287, 1987.

[69] G. A. Press, D. G. Amaral, and L. R. Squire. Hippocampal abnormalities in amnesic patients revealed by high-resolution magnetic resonance imaging. *Nature*, 341(7):54-57, Sept 1989.

[70] J. R. Quinlan. Decision trees and decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2):339-346, March/April 1990.

[71] J. R. Quinlan. Decision trees as probabilistic classifiers. In *Proceedings of the 4th International Workshop on Machine Learning*, pages 31-37, Morgan Kaufmann, Los Altos, CA, June 1987.

[72] J. R. Quinlan. Discovering rules by induction from large collections of examples. In D. Michie, editor, *Expert systems in the Micro Electronic Age*, pages 168 201, Edinburgh University Press, UK, 1979.

[73] J. R. Quinlan. The effect of noise on concept learning. In Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell, editors, *Machine Learning:*

An Artificial Intelligence Approach, pages 149–166, Morgan Kaufmann, Los Altos, CA, 1986.

[74] J. R. Quinlan. Generating production rules from decision trees. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 304–307, Aug 1987.

[75] J. R. Quinlan. Induction of decision trees. *Machine Learning*, 1:81 106, 1986.

[76] J. R. Quinlan. Learning efficient classification procedures and their application to chess end games. In Ryszard S. Michalski, Jaime G. Carbonell, and Tom M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 463–482, Tioga, Palo Alto, CA, 1983.

[77] J. R. Quinlan. Simplifying decision trees. *International Journal of Man-Machine Studies*, 27:221–234, 1987.

[78] J. R. Quinlan. Unknown attribute values in induction. In A. M. Segre, editor, *Proceedings of the 6th International Workshop on Machine Learning*, pages 164–168, Ithaca, NY, June 1989.

[79] Ulrich Raff and Francis D. Newman. Lesion detection in radiologic images using an autoassociative paradigm: Preliminary results. *Medical Physics*, 17(5):926–928, Sept/Oct 1990.

[80] Christian Roch, Thierry Pun, Denis F. Hochstrasser, and Christian Pellegrini. Automatic learning strategies and their application to electrophoresis analysis. *Computerized Medical Imaging and Graphics*, 13(5):383 391, Sept/Oct 1989.

[81] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386 408, 1958.

[82] O. Rousset, M. Jacquemet, F. Lavenne, M. Chaze, D. Le Bars, and L. Cinotti. Conception of a brain phantom for the calibration of quantitative studies in

positron emission tomography. In *Proceedings of the 2nd European Workshop on PET Instrumentation: Advances in Quantitation and Imaging Methods*, Nov. 1990.

[83] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Learning Internal Representations by Error Propagation.* MIT Press, Cambridge, MA, 1986.

[84] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3:211–229, 1959.

[85] Jude W. Shavlik and Thomas G. Dietterich, editors. *Readings In Machine Learning.* Morgan Kaufmann, San Mateo, CA, 1990.

[86] Jude W. Shavlik and Geoffrey G. Towell. An approach to combining explanation-based and neural learning algorithms. *Connection Science*, 1(3):233-255, 1989.

[87] B. A. Shepherd. An appraisal of a decision tree approach to image classification. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 473-475. IJCAI, Karlsruhe, Germany, 1983.

[88] Herbert A. Simon. Why should machines learn? In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 25-37, Tioga, Palo Alto, CA, 1983.

[89] E. Sokolowska and J. A. Newell. Multi-layered image representation: Structure and application in recognition of parts of brain anatomy. *Pattern Recognition Letters*, 4:223 230, Sept 1986.

[90] K. Spitzer and H. S. Stiehl. The paradigm of computer assisted radiology and computer assisted neurology. In H. U. Lemke, M. L. Rhodes, C. C. Jaffe, and R. Felix, editors, *Proceedings of the 3rd International Symposiom on Computer Assisted Radiology*, pages 292–295, CAR, 1989.

[91] Sharon A. Stansfield. ANGY: A rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):188 199, 1986.

[92] Robert E. Stepp and Ryszard S. Michalski. Conceptual clustering: Inventing goal-oriented classifications of structured objects. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, *Machine Learning: An Artificial Intelligence Approach*, pages 471–498, Morgan Kaufmann, Los Altos, CA, 1986.

[93] J. Talairach, G. Szikla, P. Tournoux, A. Prossalentis, M. Bordas-Ferrer, L. Covello, M. Jacob, A. Mempel, P. Buser, and J. Bancaud. *Atlas d'anatomie stereotaxique du telencephale.* Masson, Paris, France, 1967.

[94] Jean Talairach and Pierre Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - An Approach to Cerebral Imaging.* Thieme Medical Publishers, New York, NY, 1988.

[95] Paul E. Utgoff. ID5: An incremental ID3. In *Proceedings of the 5th International Conference on Machine Learning*, pages 107 120. Morgan-Kaufmann, San Matea, CA, 1988.

[96] Paul E. Utgoff. Perceptron trees: A case study in hybrid concept representation. In *Proceedings of the 7th National Conference on Artificial Intelligence*, pages 601–606, Morgan-Kaufmann, San Matea, CA, 1988.

[97] L. G. Valient. A theory of the learnable. *Communications of the ACM*, 27:1134–1142, 1984.

[98] Michael W. Vannier, Robert L. Butterfield, Douglas L. Rickman, Douglas M. Jordan, William A. Murphy, and Pietro R. Biondetti. Multispectral magnetic resonance image analysis. *CRC Critical Reviews in Biomedical Engineering*, 15(2):117–144, 1987.

[99] Gianni L. Vernazza, Sebastiano B. Serpico, and Silvana G. Dellepiane. A knowledge-based system for biomedical image processing and recognition. *IEEE Transactions on Circuits and Systems*, 34(11):1399–1416, 1987.

[100] D. Graf von Keyserlingk, K. Niemann, and J. Wasel. A quantitative approach to spatial variation of human cerebral sulci. *Acta Anatomica*, 131:127–131, 1988.

[101] Lars-Olof Wahlund, Ingrid Agartz, Ove Almqvist, Hans Basun, Lars Forssell, Jan Saaf, and Lennart Wetterberg. The brain in healthy aged individuals: MR imaging. *Neuroradiology*, 174:675–679, 1990.

[102] Craig Watson. *Basic Neuroanatomy: An Introductory Atlas*. Little, Brown and Company, Boston, MA, 1991.

[103] Sholom M. Weiss and Ioannis Kapouleas. An empirical comparison of pattern recognition, neural nets, and machine learning classification methods. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 781- 787, Morgan Kaufmann. Detroit, MI, Aug 1989.

[104] Lawrence E. Williams. The diagnostic process. In Lawrence E. Williams, editor, *Nuclear Medical Physics*, pages 157–174, CRC Press, Boca Raton, FL, 1987.

[105] Patrick Henry Winston. *Artificial Intelligence*. Addison-Wesley, Massachusetts, 1984.

[106] Patrick Henry Winston. Learning and reasoning by analogy. *Communications of the ACM*, 23:689-703, 1980.

[107] Patrick Henry Winston. *Learning Structure Descriptions from Examples*. Technical Report TR-231, Massachusetts Institute of Technology, Cambridge, MA, 1970.

Appendix

Glossary

Acquisition Process of measuring and storing image data.

Choroid plexus A structure within the ventricles responsible for the production of CSF. In magnetic resonance images, it can appear as a hyperintensity similar to multiple sclerosis lesions.

Coronal Plane Any plane which separates the brain into front and back.

Computed tomography, CT An ionizing imaging technique similar to X-ray imagery. A computer is used, instead of film, to hold images.

Decision Tree A recursive structure for representing classification rules. Each tree node represents a test on a feature, and each leaf represents a class.

Deduction The process of inferring specific facts from general data.

Echo time, TE Time during a TR interval at which the signals are recorded when acquiring MR images.

External CSF Refers to the cerebrospinal fluid under the arachnoidal layer of the brain.

Gyrus, gyri (plural) A 'fold' or convolution of the brain.

Heuristic A guideline or pointer towards the general area in which a solution to a problem may be found.

Induction The process of inferring general hypotheses or rules from specific facts.

Inference The deriving of a conclusion from induction or deduction - a conclusion arrived at in logic.

Magnetic resonance, MR Absorption or emission of electromagnetic energy by nuclei in a (static) magnetic field after excitation by suitable (RF) radiation.

MR See magnetic resonance.

MR imaging, MRI Non-invasive medical technique which permits the detailed visualization of internal anatomical structures in living subjects; production of tomographic (cross-sectional) views of a body, by use of the phenomenon of magnetic resonance.

Multiple Sclerosis, MS A neurological disease characterized by lesions to the myelin covering (fatty white substance) of neurons of cerebral white matter.

Partial volume effect Refers to case of a data element (pixel or voxel) containing more than one tissue type.

Periventricular Adjacent to the ventricles.

Production rule Rules in the form: IF condition THEN action.

Proton Density In the context of MR, the number of magnetized protons per voxel.

Radio frequence (RF) pulse Electromagnetic radiation pulse used in magnetic resonance imaging, commonly in the 1-100 megahertz range. Their principal effect on the body is energy deposition in the form of tissue heating, mainly at the surface.

Repetition time, TR Period between the beginning of a pulse sequence and the start of the succeeding sequence.

RF inhomogeneity artifact An image artifact caused by nonuniformities in the radiofrequency field applied during image acquisition.

Segmentation Process of identifying regions of an image which are uniform and homogeneous with respect to some given characteristics.

Sagittal plane Any plane which divides the brain into left and right pieces.

T1, longitudinal relaxation time An MR tissue-specific time constant characterizing the rate at which excited nuclei re-align with the external magnetic field.

T2, transverse relaxation time MR tissue-specific time constant characterising the rate at which nuclei reach equilibrium.

Transverse Plane Any plane which separates the brain into top and bottom.

Ventricle, ventricles A structure within the brain in which cerebrospinal fluid is produced.

Voxel 3-dimensional version of a pixel; a unit cube of image data.