

01170

A
29



Universidad Nacional Autónoma de México

División de Estudios de Posgrado
de la Facultad de Ingeniería

Redes Neuronales Artificiales: Aplicación
en el Diagnóstico Diferencial de la
Demencia Vasculiar y Tipo Alzheimer

Que para obtener el grado de
Maestría en Ingeniería Eléctrica

Presenta:

Ma. Elizabeth García Pérez

Director de Tesis:

Dr. Francisco Cervantes Pérez

Ciudad Universitaria, Agosto de 1996

**TESIS CON
FALLA DE ORIGEN**

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

A mi familia

Agradecimientos

Al Dr. Francisco Cervantes, por apoyar la idea de este proyecto y por su asesoría y amistad a lo largo del tiempo.

Muy especialmente al Dr. Arturo Violante, por su entusiasmo, interés y participación en la elaboración de este trabajo, sin la cual no hubiese sido posible realizarlo.

A Eduardo Cabrera, por su gran ayuda en la terminación de esta tesis.

Al Dr. Julio Sotelo por las facilidades brindadas para realizar este trabajo.

Al comité evaluador y jurado de esta tesis.

A los pacientes con demencia.

**REDES NEURONALES ARTIFICIALES: APLICACIÓN EN EL
DIAGNÓSTICO DIFERENCIAL DE LA DEMENCIA
VASCULAR Y TIPO ALZHEIMER**

Índice

I. INTRODUCCIÓN	2
I.1 MOTIVACIÓN Y PLANTEAMIENTO DEL PROBLEMA.....	2
I.2 OBJETIVOS DEL TRABAJO	5
I.3 ORGANIZACIÓN DE LA TESIS	6
II. ANTECEDENTES	9
II.1 DESCRIPCIÓN GENERAL DE LAS DEMENCIAS VASCULAR Y TIPO ALZHEIMER	9
II.1.1 DEFINICIÓN DE DEMENCIA.....	9
II.1.2 DEMENCIA VASCULAR.....	10
II.1.3 DEMENCIA TIPO ALZHEIMER	11
II.1.4 DIAGNÓSTICO DIFERENCIAL ENTRE DV Y DTA.....	12
II.2 REDES NEURONALES ARTIFICIALES COMO TECNOLOGÍA PARA CLASIFICACIÓN DE PATRONES.....	20
II.2.1 INTRODUCCIÓN	20
II.2.2 LA NEURONA COMO INTEGRADOR	23
II.2.3 ALGUNOS CONCEPTOS GENERALES RELACIONADOS CON EL PROCESO NEURONAL.....	26
II.2.3 MODELO DE NEURONA	28
II.2.5 ARQUITECTURAS Y ALGORITMOS DE ENTRENAMIENTO	30
II.2.6 ALGORITMO DE RETROPROPAGACIÓN	34
II.2.7 APLICACIÓN DE LAS RNA EN LA MEDICINA	41
II.2.8 EJEMPLOS DE APLICACIÓN.....	42

**III. DISEÑO CONSTRUCCIÓN Y ENTRENAMIENTO
DEL SISTEMA BASADO EN RNA 50**

III.1 METODOLOGÍA50
III.2 RECOLECCIÓN DE LA MUESTRA51
III.3 REPRESENTACIÓN DE LA INFORMACIÓN.....52
**III.4 SELECCIÓN DEL ALGORITMO DE ENTRENAMIENTO,
ARQUITECTURA Y PARÁMETROS DE LA RED.....54**
III.5 ENTRENAMIENTO DE LA RNA55
III.6 RESULTADOS DEL ENTRENAMIENTO57
III.7 PRUEBAS A LA RNA ENTRENADA60

**IV. ESTUDIO DE LAS PRUEBAS RELACIONADAS
CON EL DIAGNÓSTICO DIFERENCIAL
UTILIZANDO RNA 62**

IV.1 PLANTEAMIENTO.....62
IV.2 RESULTADOS64
IV.2.1 CONJUNTO DE COSTO MÍNIMO.....69

V. CONCLUSIONES 78

A. APENDICE 82

A.1 ANÁLISIS82
A.2 DISEÑO.....83
A.3 IMPLEMENTACIÓN.....84
A.4 PRUEBAS Y RESULTADOS90

VI. REFERENCIAS 93

I. Introducción

***1.1 Motivación y Planteamiento
del Problema***

1.2 Objetivos del Trabajo

1.3 Organización de la Tesis

I. INTRODUCCIÓN

Una de las principales características que diferencian a los individuos de los otros integrantes del reino animal y de los mismos individuos entre sí son los procesos de pensamiento, los cuales han sufrido un gran desarrollo en las etapas evolutivas del hombre, permitiéndole llegar a crear, desarrollar y descubrir una infinidad de conocimientos que nos colocan encima de los demás seres vivos, por lo que la pérdida de esas funciones de pensamiento asentadas en regiones cerebrales conocidas y llamadas globalmente funciones mentales es uno de los problemas más devastadores a los que nos enfrentamos y en los que día a día se adquieren más conocimientos para combatirlos.

Este trabajo pretende sumarse a esa serie de esfuerzos, no para combatir la demencia, pero sí para ayudar al paciente a llevar una mejor calidad de vida.

1.1. Motivación y Planteamiento del Problema

Durante los últimos años, al aumentar la expectativa de vida de los individuos, particularmente en los países desarrollados, la demencia ha adquirido mayor interés debido al incremento observado en su prevalencia. Se ha reportado demencia de 1.3 a 6.2% en personas de 65 años de edad y mayores (Ueda et al, 1992; Gorelick et al, 1993; Joachin et al, 1988), de la misma manera como ocurre con la Enfermedad Vasculiar Cerebral (EVC) que ha llegado a ocupar la tercer causa de muerte, y se ha identificado como una de las principales causas de Demencia. De esta forma, el impacto sobre la salud pública, dada la incapacidad a la que conduce, se considera vital.

Dentro del diagnóstico (dx) diferencial de la demencia se debe considerar a la Demencia Vasculiar (DV) y a la Demencia Tipo Alzheimer (DTA) como sus principales exponentes; entre ambas se ha llegado a reportar que representan hasta un 90% del total de los pacientes con demencia (O'Brien, 1992; Bolier et al, 1989). La Sociedad Mexicana de Alzheimer reporta que 6% de las personas mayores de 65 años cuentan con un dx de demencia lo que implica un número creciente de personas con dx de DTA en México.

Es sabido que este tipo de demencias pueden llegar a presentar dificultades en el dx diferencial por las similitudes en sus características clínicas y exámenes paraclínicos si no se llevan a cabo estudios invasivos, como la Biopsia Cerebral (McKhann et al, 1984), ya que es la única prueba para realizar en forma segura el dx diferencial de ambas entidades. Obviamente, estos estudios presentan problemas éticos y médicos por las posibles complicaciones y las mínimas posibilidades terapéuticas que se brindan una vez obtenido el dx de certeza (sobre todo en el caso de la DTA). No obstante, se requiere de parámetros confiables que permitan dar validez al dx de estas entidades por las diferencias terapéuticas y la posible modificación de su historia natural, principalmente si consideramos que la DV se puede asociar a factores de riesgo modificables como es la hipertensión arterial.

El hecho de que la prueba más confiable para obtener el dx certero sea invasiva, hace que se requiera una serie muy grande de estudios que se le deben realizar al paciente. Estos estudios y datos están formados con la información de los seis grupos siguientes:

1. Demográficos
2. Antecedentes
3. Síntomas
4. Escalas Neurológicas y Neuropsicológicas
5. Electrofisiología
6. Neuroimagen y otros estudios (laboratorio, rayos X, doppler de vasos de cuello).

Cada grupo está formado por diferentes datos y pruebas; en el grupo 1, 2, 3 y 4 se encuentran los datos que se pueden obtener en la primer consulta. Los del grupo 5 y 6 necesitan de tiempo para realizarse; el contenido de cada grupo se explicará en el siguiente capítulo.

Existen ciertos indicadores que nos muestran que el costo de este gabinete es muy alto, no sólo en forma económica. Para este trabajo se consideraron los siguientes factores:

1. Riesgo para el paciente
2. Infraestructura requerida
3. Tiempo de realización de la prueba
4. Factor económico (costo)

De los cuatro indicadores, el más importante es el que tiene que ver directamente con la persona enferma, esto es, el **riesgo**. Toda persona que se somete a estudios médicos sabe que existe la posibilidad de llegar a contraer algún padecimiento o quedar con alguna consecuencia debida a la aplicación de ciertos medios de contraste o anestesia. En otros casos, un mal procedimiento o algún error humano en alguna de estas pruebas tiene la probabilidad de llegar a dañar cualquier órgano o función que en un principio no presentaba problemas, se puede contagiar al paciente con algún virus e incluso, se podría llegar a provocar la muerte del paciente.

El segundo, tercero y cuarto factor están muy relacionados entre sí. Con respecto a la **infraestructura**, no todos los centros de salud del país cuentan con el equipo necesario para realizar estudios de Imagen por Resonancia Magnética (IRM) o de Topografía Axial Computarizada (TAC), por lo que se tiene que referir al paciente hacia alguna institución que cuente con servicio a público externo. Esto relaciona al segundo con el tercer factor, el **tiempo**, ya que el remitir a algún paciente a otro centro u hospital requiere de tiempo de transporte, el tiempo que tarda en esperar por su cita, el tiempo que tardan los resultados y finalmente, el tiempo que tarda en tener otra cita con el especialista que le ordenó el estudio (estos tiempos pueden ir desde semanas hasta meses). Lo anterior es sin considerar la demanda de pacientes que necesitan el estudio (por diferentes causas, no solo demencia).

Con todo esto, el dx final se retrasa, pues cada uno de esos estudios tiene que ser evaluado por el especialista, que a su vez manda los resultados al médico neurólogo que está atendiendo al paciente, quien evalúa el conjunto de resultados de todos los estudios para finalmente emitir un dx. Por lo anterior, el resultado puede ser prolongar el tiempo entre el debut del padecimiento y el dx adecuado.

La relación del tercero con el cuarto factor es directa, pues el requerir transporte para ir de un centro a otro (si es que se encuentran en la misma ciudad) genera un cierto costo (**factor económico**) y a esto hay que agregarle el costo mismo del estudio (como el de la IRM o la TAC que son pruebas de precio elevado), lo que hace que el gabinete completo tenga un costo monetario alto.

No obstante, se requiere de parámetros confiables que permitan dar validez al dx de estas entidades, principalmente por las diferencias terapéuticas y la posible modificación de su historia natural, recordando que no existe tratamiento curativo y que se desconoce la forma de evitar su progresión. De esta forma, el impacto sobre la salud pública, dada la incapacidad a la que conduce, se considera vital.

Una estrategia que se sigue en el dx neurológico es el análisis de los resultados obtenidos en las diferentes pruebas mencionadas a través de un sistema computacional de propósito específico. Para realizar el dx médico o la toma de decisiones asistida por computadora se han desarrollado nuevos y diferentes tipos de tecnologías con las que se han podido atacar problemas médicos cada vez más complejos (Schwartz, 1970, 1987; Shortliffe, 1987; Szolovitz et al, 1988). Entre ellas, destaca la creación de **Sistemas Expertos (SE)** y el desarrollo de las **Redes Neuronales Artificiales (RNA)**.

Los SE son programas computacionales, algunos de los cuales están basados en reglas de producción, que intentan ayudar en el dx. El ejemplo clásico en este campo es el realizado por Shortliffe en 1976, quien construye un SE llamado MYCIN que diagnostica diferentes tipos de enfermedades infecciosas (Shortliffe, 1976). Kulkowski en el '80 y Reggia en el '85 publican diferentes artículos en los que hablan acerca de las técnicas de **Inteligencia Artificial (IA)** disponibles y su relación con los sistemas de consulta médica (Kulkowski, 1980; Reggia et al, 1985). Otro SE llamado SETH,

creado por Darmoni y cols., se utiliza para controlar los niveles de envenenamiento por medicamentos en adultos (Darmoni et al, 1994). Un SE más que ayuda en la toma de decisiones en el dx médico es llamado DXplain, que fue desarrollado en colaboración con la Asociación Médica Americana (Barnett et al, 1987).

Otra tecnología que frecuentemente se está utilizando son las redes neuronales artificiales que, desde 1986 con el trabajo de Rumelhart y cols., han adquirido gran importancia en la solución de diferentes tipos de problemas como clasificación de patrones, dx médico, control de calidad, predicción en la bolsa de valores, finanzas, reconocimiento de firmas en bancos, control de robots, etc. (Widrow et al, 1994).

Las RNA son sistemas computacionales formados por elementos simples interconectados que procesan información y generan una respuesta (Lippmann, 1987). Han demostrado su eficiencia en el reconocimiento de imágenes radiológicas (Boone et al, 1990a, 1990b; Gross et al, 1990; Hallgren et al, 1992), señales biomédicas (Gevins et al, 1988; Mamelak et al, 1991; Aikon et al, 1990; Jandó et al, 1992, 1993), así como en la solución de problemas de predicción y diagnóstico (Casselman et al, 1990; Poil et al, 1991; Moallemi, 1991; Baxt, 1991). En el capítulo II.2 se ampliará la descripción de algunos de los sistemas a los que se hace referencia aquí.

1.2. Objetivos del trabajo

Conociendo las propiedades de las RNA y considerando que el dx diferencial de estas dos entidades es difícil, se propone el uso de esta tecnología como auxiliar del médico en la toma de decisiones para obtener un buen dx diferencial.

Por lo tanto, se plantean los siguientes objetivos generales:

- Diseñar y construir un modelo de Red Neuronal Artificial para el auxilio en el Diagnóstico Diferencial de la Demencia Vasculiar y Tipo Alzheimer.
- Determinar el conjunto mínimo de pruebas que permite hacer un diagnóstico diferencial certero y que no involucre estudios invasivos.

Para lograr estos objetivos y tomando en cuenta las características del problema, se desprenden los siguientes objetivos específicos:

1. Hacer una revisión bibliográfica.
2. Determinar una estructura global para el sistema de asistencia en el diagnóstico diferencial con base en un modelo de Redes Neuronales Artificiales.

3. Seleccionar una forma de representación computacional de los datos médicos y realizar la codificación de entrada a la RNA.
4. Definir el modelo a emplear, es decir, la arquitectura y el algoritmo de aprendizaje así como los diferentes parámetros que intervienen en el entrenamiento de la red.
5. Realizar diferentes combinaciones con los seis grupos de pruebas para encontrar el conjunto mínimo de parámetros con el que se pueda realizar un buen diagnóstico.
6. Finalmente, y como apéndice de esta tesis se propone diseñar e implementar una interfaz gráfica de usuario para que el sistema completo (red e interfaz) se pueda emplear como asistente en la toma de decisiones y de almacenamiento de nuevos datos. Estos enriquecerán a la Base de Datos (BD) original, con la que posteriormente se podrá reentrenar periódicamente a la RNA.

1.3. Organización de la tesis

En este primer capítulo se sitúan la motivación, el planteamiento del problema y los objetivos de la tesis, se describe en forma clara y resumida qué se va a hacer y por qué es importante realizarlo.

En el capítulo II se revisan los antecedentes del problema. Se describe qué es la demencia, cuáles son las características principales de la DV y la DTA, la forma en que se realiza actualmente el dx diferencial de estas dos entidades, se explica en qué consiste cada una de las pruebas que se realizan actualmente y el tiempo que tarda en realizarse el gabinete completo de pruebas, así como una introducción de lo que es la tecnología de redes neuronales artificiales en el reconocimiento de patrones, la descripción del algoritmo de retropropagación y algunos ejemplos de aplicaciones de las RNA en la medicina. También se justifica por qué se van a utilizar redes neuronales artificiales como una buena solución en la asistencia de la toma de decisiones del médico neurólogo.

En el capítulo III se describe la metodología que se siguió, esto es, la manera en que se realizó una primera aproximación al problema del dx diferencial utilizando RNA's; se explica también la forma en que se representaron los datos médicos (entradas y salida) para el entrenamiento de las redes, la división en grupos del conjunto total de pruebas médicas, la selección de la arquitectura de la red, el algoritmo de entrenamiento, la selección de parámetros para el buen entrenamiento de la red y la forma en que se eligieron los grupos de entrenamiento y prueba para la red

neuronal. También se describen los resultados obtenidos, así como la justificación de por qué hay que entrenar otras redes.

El capítulo IV describe los resultados obtenidos del estudio del conjunto de pruebas relacionadas con el dx diferencial y la manera en que se encontró el conjunto mínimo de parámetros para realizar un dx certero utilizando redes neuronales; además se muestran los resultados de confiabilidad obtenida en cada uno de los grupos de pruebas con diferentes redes neuronales y, a grosso modo, un análisis de costos y tiempos para encontrar el conjunto mínimo de estudios necesarios, con bajo costo y realizables en poco tiempo y que garantice un dx certero.

En el capítulo V se presentan las conclusiones del trabajo así como los pasos a seguir, a futuro, para el buen funcionamiento del sistema.

En el capítulo VI se mencionan las referencias bibliográficas de este trabajo.

Como un apéndice de esta tesis se describe el sistema de interfaz gráfica de usuario que se realizó y la metodología que se siguió para su implementación.

II. Antecedentes

II.1 Descripción General de las Demencias Vascular y Tipo Alzheimer

II.2 Redes Neuronales Artificiales como Tecnología para Clasificación de Patrones

II. ANTECEDENTES

En este capítulo se dan los antecedentes del problema, la definición de todos los términos que se utilizarán, definición del término demencia y también se definen los dos tipos de demencia con los que se trabajará a lo largo de este trabajo. También se describen todas las pruebas que se realizan en la actualidad para llevar a cabo el diagnóstico diferencial.

II.1. Descripción General de las Demencias Vasculares y Tipo Alzheimer.

En esta sección se explica el término general de "demencia" y cada uno de los dos tipos principales. También se mencionan algunos otros tipos de demencia y su clasificación actual.

II.1.1. Definición de Demencia

La población mundial envejece y este hecho, comúnmente aceptado, tiene como consecuencia un crecimiento en términos de frecuencia de las enfermedades asociadas con la vejez. Entre ellas, las que poseen los efectos más devastadores probablemente sean las que afectan el sistema nervioso.

La prevalencia de muchas enfermedades cerebrales aumenta con la edad; por ejemplo, las lesiones cerebrales isquémicas (Infartos) son más frecuentes en los ancianos. Sin embargo, el aspecto más importante del envejecimiento cerebral consiste en el incremento en la frecuencia de las denominadas enfermedades "degenerativas", término que indica simplemente que hasta hace poco tiempo la etiología y los mecanismos básicos de esas enfermedades eran desconocidos. Ahora, con el avance de la tecnología y aunque no se conocen por completo, se han podido realizar

más estudios con los que se ha llegado a saber un poco más sobre las enfermedades degenerativas en general.

Se puede considerar a la demencia como un síndrome cerebral orgánico, que ocasiona daño en las funciones intelectuales adquiridas y que se caracteriza por un déficit persistente en por lo menos tres -pueden ser más y en diferentes combinaciones- de las siguientes áreas de actividad mental: memoria, lenguaje, habilidades visoespaciales, personalidad o estado emocional y cognición (abstracción, cálculo, juicio) (Cummings et al, 1992). La naturaleza adquirida de este tipo de habilidades mentales distingue a la demencia del retardo mental mientras que su persistencia lo diferencia de los estados confusionales agudos. Adicionalmente, el requerimiento de encontrar varias áreas de las funciones mentales afectadas distingue a la demencia de la afasia (pérdida/disminución del habla o entendimiento de lenguaje), estados amnésicos y algunos otros déficits cognitivos monosintomáticos (DSM-III Working Group, 1987; Wisler, 1991; Bogousslavsky, 1992).

Como ya se mencionó, al aumentar la expectativa de vida en la población en general, la demencia ha adquirido mayor interés debido al incremento observado en su prevalencia.

Entre las demencias, la más frecuente es la DTA y la causada por lesiones vasculares llamada DV (O'Brien, 1992; Boller et al, 1989). Por lo menos 50% de los casos de demencias degenerativas se deben a DTA. Más aún, se observan lesiones cerebrales de "tipo DTA" junto con lesiones vasculares en otro 40% de los casos. En otras palabras, se encuentran lesiones cerebrales de tipo DTA y DV en alrededor del 90% del total de los casos de demencia. De esta forma, el impacto sobre la salud pública se considera vital, dada la incapacidad a la que conducen por sus fallas en las funciones mentales mencionadas.

II.1.2. Demencia Vascular

El término Demencia Multi-Infarto ha sido utilizado como un sinónimo de todas las demencias de etiología vascular implicando incorrectamente que los Infartos Cerebrales múltiples son la única causa de la DV (Drachman, 1993). Asimismo, un infarto único localizado en una área estratégica puede llegar a causar DV.

La Demencia Vascular puede tener un inicio súbito o agudo pero por lo general el deterioro del sujeto se da en forma lenta, como en la DTA. En estos casos, el paciente a menudo expresa su preocupación por fallas de memoria, es frecuente encontrar mayor lucidez por el día y confusión por la noche (Erkinjuntti et al, 1988; Hachinski et al, 1975). Los pacientes con DV suelen conservar mejores relaciones interpersonales, la depresión suele ser común y pueden existir intentos suicidas. En algunos casos pueden presentar una historia de síntomas neurológicos anormales debidos a ataques isqué-

micos transitorios. En muchos otros, la enfermedad puede progresar gradualmente, imitando el curso de una condición degenerativa tipo DTA. Los factores de riesgo para adquirir demencia vascular incluyen hipertensión, cardiopatías (enfermedades del corazón), tabaquismo, diabetes mellitus, alcoholismo y algunos otros (Meyer et al, 1988). Pueden presentar dificultades o trastornos en orientación, memoria reciente, abstracción y cálculo, al igual que en la DTA. La capacidad de expresar el lenguaje puede estar alterada en la escritura, al procesar instrucciones complejas y en gramática.

El perfil emergido de cuidadosas observaciones anatomo-clínicas de Tomlinson, Blessed y Roth basadas en pacientes con DV **confirmada por autopsia**, sugieren que esta demencia requiere una pérdida cerebral de 50 ml. o mayor, usualmente de 100 ml. También se encontró correlación entre el volumen del infarto y la severidad de la demencia en los casos vasculares; sin embargo, si existen múltiples infartos que involucren al tálamo y a la sustancia blanca subcortical pueden causar demencia, aún sin gran pérdida de volumen cerebral. Cabe señalar que aún con infartos cerebrales mayores de 50 ml. puede no encontrarse demencia a lo largo de la vida del sujeto y por tanto **el sitio del infarto también es crítico**. En estudios recientes se sugiere que la bilateralidad de los infartos, así como que se involucre el tálamo es de importancia crítica para el desarrollo de la DV.

No obstante lo anterior, en la mayor parte de los casos, el cuadro clínico de la DV suele ser indistinguible del que presenta la DTA o de algunos otros tipos de demencia.

II.1.3. Demencia Tipo Alzheimer

La Demencia Tipo Alzheimer es un trastorno degenerativo cerebral caracterizado por el deterioro progresivo de las funciones cognitivas. En sus etapas iniciales, la DTA es una alteración de la cognición y a medida que progresa también se presentan déficits emocionales y motores.

Las alteraciones cognoscitivas que predominan son fallas de memoria, lenguaje, alteraciones sensoriales y disminución en habilidades perceptuales (visual, auditiva), de manera similar a lo que sucede en pacientes con DV. Los déficits emocionales que se presentan incluyen trastornos afectivos, agitación, agresividad, intolerancia al estrés, modificaciones en el esquema alimenticio y alteraciones en la esfera de la sexualidad.

Los cambios motores incluyen rigidez y decremento de la actividad motriz (Gottfries, 1985).

La frecuencia de las enfermedades demenciales de la edad adulta es difícil de establecer con exactitud ya que los criterios diagnósticos y los métodos de recolección de información utilizados varían y pueden ser poco confiables (Katchaturian, 1989). Los estudios epidemiológicos han

Intentado delinear factores de riesgo para desarrollar DTA, de éstos, sólo la edad avanzada ha sido un factor etiológico comprobable. En algunos pacientes se ha observado que la predisposición genética juega un papel importante especialmente en sujetos con dx temprano (i.e. el dx se da en las primeras etapas de la enfermedad), otros factores que se han sugerido son vulnerabilidad a la exposición al aluminio, exposición a virus lentos, defectos inmunológicos, enfermedad tiroidea y traumatismo craneoencefálico. El padecimiento parece ser más frecuente en mujeres, esto puede deberse en parte a la mayor longevidad que presenta este sexo.

Es necesario recordar que existe un grupo de pacientes con enfermedad tipo Alzheimer quienes además de las características descritas, tienen factores de riesgo vascular e incluso pueden presentar datos de enfermedad vascular cerebral. Es en estos sujetos principalmente en quienes resulta difícil el dx etiológico del síndrome demencial.

La Sociedad Mexicana de Alzheimer reporta que 6% de las personas mayores de 65 años cuentan con un dx de demencia lo que implica un número creciente de personas con dx de DTA en México. Esto es comprensible si consideramos que si bien nuestro país esta conformado en su mayoría por una población joven, los avances que se han logrado en el desarrollo de cuidados prenatales y pediátricos en general, aunado a la posibilidad de prevenir y tratar exitosamente muchos padecimientos sobre todo infecciosos en más sectores de la población, han hecho posible que la edad promedio de vida de nuestros habitantes aumente y por lo tanto exista un mayor porcentaje de ancianos. El problema que enfrentan actualmente los países desarrollados con un alto porcentaje de población anciana presentan un panorama que podría ser similar al que va a enfrentar nuestro país una vez que aumente el grupo de la población más susceptible a sufrir enfermedades asociadas al envejecimiento.

11.1.4. Diagnóstico Diferencial entre DV y DTA

Para realizar el dx diferencial de la DV y la DTA es necesario tener en cuenta las siguientes consideraciones.

El primer elemento a considerar en el dx es saber si el paciente está o no demenciado. En ocasiones esto puede llegar a ser problemático ya que los síntomas tempranos de la DTA varían según las características del individuo en cuestión, su edad, status socioeconómico, nivel de funcionamiento previo, escolaridad, etc. Por ejemplo, un contador empezará a tener dificultades con los números, un ama de casa olvidará cómo cocinar, qué ingredientes llevan los alimentos, un obrero puede tener dificultad para llevar a cabo su rutina diaria, puede no recordar que ruta debe seguir para llegar a su trabajo, etc. Algunos sujetos pueden iniciar con alteraciones del lenguaje conservando todas las habilidades viso-espaciales o viceversa (Eslinger et al, 1985; Bayles, 1991).

Existen diversas causas que pueden alterar las funciones intelectuales. Es importante obtener una historia clínica detallada que incluya tipo de inicio (abrupto vs. lento y progresivo) para establecer la naturaleza de la disfunción inicial (memoria, lenguaje, orientación, apraxias -trastorno en la ejecución intencional de un gesto motor; por ejemplo, vestirse o servir agua-, agnosias -incapacidad para reconocer y discriminar patrones complejos como objetos, colores, fisonomías-, atención), e investigar acerca de su posible etiología (vascular, degenerativa, mixta). El interrogatorio acerca de cambios en la personalidad y síntomas depresivos (Bolla et al, 1991; Fisher et al, 1990; Krall, 1983; Rovner et al, 1989) son necesarios para el dx diferencial. También se debe considerar el tipo de fármacos que tome el paciente, ya que se ha visto que en los ancianos pueden causar o acelerar fallas en memoria (Spiegel et al, 1981). Otros exámenes que se deberán incluir son; análisis de sangre, deficiencias nutricionales y la valoración de enfermedad cerebrovascular. El examen motor en la DTA inicial no presenta anormalidades (Cummings, 1992).

Los síntomas clínicos de la DTA pueden estar causados por una variedad de condiciones médicas. Se han encontrado que aún en estudios de autopsia existen pacientes (10%) que no muestran las placas y las madejas características de esta entidad, situación que complica el dx diferencial. Muchas enfermedades pueden producir síntomas tipo DTA pero sólo unas cuantas ocurren con la suficiente frecuencia como para ser consideradas dentro del dx diferencial.

Heston (1983) propone una distinción útil para facilitar el estudio de las demencias, la primera división es: demencia primaria indiferenciada, demencia primaria diferenciada y demencia secundaria.

El primer grupo incluye aquellas enfermedades que afectan principalmente al cerebro y que producen demencia debido a los efectos directos en el tejido cerebral. Las enfermedades de este grupo son muy parecidas entre sí y generalmente no pueden ser distinguidas con procedimientos clínicos ordinarios. Se requiere de un examen directo del tejido cerebral que se obtiene por biopsia o por autopsia (DTA y Enfermedad de Pick). Este grupo incluye las demencias seniles entre las cuales la DTA constituye alrededor de un 50 al 70% del número total de demencias (Meyer et al, 1988) mientras que las demencias multi-infarto (DMI o DV) corresponden a un 15-25%. Otras demencias irreversibles incluyendo a las demencias mixtas, que resultan de la combinación de lesiones vasculares y daño neuronal primario, enfermedad de Parkinson, de Creutzfeld-Jakob, de Huntington, de Pick, esclerosis múltiple y otras más, incluyendo a las pseudo-demencias (un estado donde las condiciones somáticas y psiquiátricas, muchas de las cuales son reversibles, provocan una disfunción cognitiva) se presentan de un 15 a un 35%. En la mayoría de los casos, del 80 al 90%, el problema clínico más importante es realizar el dx diferencial entre DTA y DV.

Con respecto al dx de la DV, Hachinski (Hachinski et al, 1974) considera varias alternativas para caracterizarlo. La demencia multi-infarto se diagnostica en pacientes con lesiones cerebrales isquémicas (infartos) que presentan un cuadro clínico similar a la DTA. Estos infartos por lo general se producen cuando en forma súbita se suspende el flujo cerebral

intentado delinear factores de riesgo para desarrollar DTA, de éstos, sólo la edad avanzada ha sido un factor etiológico comprobable. En algunos pacientes se ha observado que la predisposición genética juega un papel importante especialmente en sujetos con dx temprano (i.e. el dx se da en las primeras etapas de la enfermedad), otros factores que se han sugerido son vulnerabilidad a la exposición al aluminio, exposición a virus lentos, defectos inmunológicos, enfermedad tiroidea y traumatismo craneoencefálico. El padecimiento parece ser más frecuente en mujeres, esto puede deberse en parte a la mayor longevidad que presenta este sexo.

Es necesario recordar que existe un grupo de pacientes con enfermedad tipo Alzheimer quienes además de las características descritas, tienen factores de riesgo vascular e incluso pueden presentar datos de enfermedad vascular cerebral. Es en estos sujetos principalmente en quienes resulta difícil el dx etiológico del síndrome demencial.

La Sociedad Mexicana de Alzheimer reporta que 6% de las personas mayores de 65 años cuentan con un dx de demencia la que implica un número creciente de personas con dx de DTA en México. Esto es comprensible si consideramos que si bien nuestro país está conformado en su mayoría por una población joven, las avances que se han logrado en el desarrollo de cuidados prenatales y pediátricos en general, aunada a la posibilidad de prevenir y tratar exitosamente muchos padecimientos sobre todo infecciosos en más sectores de la población, han hecho posible que la edad promedio de vida de nuestros habitantes aumente y por lo tanto exista un mayor porcentaje de ancianos. El problema que enfrentan actualmente los países desarrollados con un alto porcentaje de población anciana presentan un panorama que podría ser similar al que va a enfrentar nuestro país una vez que aumente el grupo de la población más susceptible a sufrir enfermedades asociadas al envejecimiento.

II.1.4. Diagnóstico Diferencial entre DV y DTA

Para realizar el dx diferencial de la DV y la DTA es necesaria tener en cuenta las siguientes consideraciones.

El primer elemento a considerar en el dx es saber si el paciente está o no demenciado. En ocasiones esto puede llegar a ser problemático ya que los síntomas tempranos de la DTA varían según las características del individuo en cuestión, su edad, status socioeconómico, nivel de funcionamiento previo, escolaridad, etc. Por ejemplo, un contador empezará a tener dificultades con los números, un ama de casa olvidará cómo cocinar, qué ingredientes llevan los alimentos, un obrero puede tener dificultad para llevar a cabo su rutina diaria, puede no recordar que ruta debe seguir para llegar a su trabajo, etc. Algunos sujetos pueden iniciar con alteraciones del lenguaje conservando todas las habilidades viso-espaciales o viceversa (Eslinger et al, 1985; Bayles, 1991).

Existen diversas causas que pueden alterar las funciones intelectuales. Es importante obtener una historia clínica detallada que incluya tipo de inicio (abrupto vs. lento y progresivo) para establecer la naturaleza de la disfunción inicial (memoria, lenguaje, orientación, apraxias -trastorno en la ejecución intencional de un gesto motor; por ejemplo, vestirse o servir agua-, agnosias -incapacidad para reconocer y discriminar patrones complejos como objetos, colores, fisonomías-, atención), e investigar acerca de su posible etiología (vascular, degenerativa, mixta). El interrogatorio acerca de cambios en la personalidad y síntomas depresivos (Bolla et al, 1991; Fisher et al, 1990; Krail, 1983; Rovner et al, 1989) son necesarios para el dx diferencial. También se debe considerar el tipo de fármacos que tome el paciente, ya que se ha visto que en los ancianos pueden causar o acelerar fallas en memoria (Spiegel et al, 1981). Otros exámenes que se deberán incluir son: análisis de sangre, deficiencias nutricionales y la valoración de enfermedad cerebrovascular. El examen motor en la DTA inicial no presenta anomalías (Cummings, 1992).

Los síntomas clínicos de la DTA pueden estar causados por una variedad de condiciones médicas. Se han encontrado que aún en estudios de autopsia existen pacientes (10%) que no muestran las placas y las madejas características de esta entidad, situación que complica el dx diferencial. Muchas enfermedades pueden producir síntomas tipo DTA pero sólo unas cuantas ocurren con la suficiente frecuencia como para ser consideradas dentro del dx diferencial.

Heston (1983) propone una distinción útil para facilitar el estudio de las demencias, la primera división es: demencia primaria indiferenciada, demencia primaria diferenciada y demencia secundaria.

El primer grupo incluye aquellas enfermedades que afectan principalmente al cerebro y que producen demencia debido a los efectos directos en el tejido cerebral. Las enfermedades de este grupo son muy parecidas entre sí y generalmente no pueden ser distinguidas con procedimientos clínicos ordinarios. Se requiere de un examen directo del tejido cerebral que se obtiene por biopsia o por autopsia (DTA y Enfermedad de Pick). Este grupo incluye las demencias seniles entre las cuales la DTA constituye alrededor de un 50 al 70% del número total de demencias (Meyer et al, 1988) mientras que las demencias multi-infarto (DMI o DV) corresponden a un 15-25%. Otras demencias irreversibles incluyendo a las demencias mixtas, que resultan de la combinación de lesiones vasculares y daño neuronal primario, enfermedad de Parkinson, de Creutzfeld-Jakob, de Huntington, de Pick, esclerosis múltiple y otras más, incluyendo a las pseudo-demencias (un estado donde las condiciones somáticas y psiquiátricas, muchas de las cuales son reversibles, provocan una disfunción cognitiva) se presentan de un 15 a un 35%. En la mayoría de los casos, del 80 al 90%, el problema clínico más importante es realizar el dx diferencial entre DTA y DV.

Con respecto al dx de la DV, Hachinski (Hachinski et al, 1974) considera varias alternativas para caracterizarlo. La demencia multi-infarto se diagnostica en pacientes con lesiones cerebrales isquémicas (infartos) que presentan un cuadro clínico similar a la DTA. Estos infartos por lo general se producen cuando en forma súbita se suspende el flujo cerebral

sanguíneo, pues para funcionar en forma adecuada, el cerebro requiere hasta el 20% del gasto cardíaco (cantidad de sangre por minuto que distribuye el corazón). Esta cantidad de sangre puede alterarse con la formación de aterosclerosis (formación de placas de colesterol, grasas, elementos sanguíneos y plaquetas, en las paredes interiores de las arterias) que provoca la disminución en el flujo sanguíneo cerebral y el consiguiente infarto en el territorio irrigado por la(s) arteria(s) dañada(s). También se puede producir cuando en forma súbita el flujo es interrumpido por un émbolo (coágulo que viaja de un sitio a otro hasta ocluir una arteria de menor calibre).

La investigación de las demencias de tipo vascular es un campo necesario de explorar. La presencia de infartos no siempre puede explicar la presencia de demencia, ya que en la edad adulta ambas pueden coexistir sin estar necesariamente correlacionadas, lo que impide poder determinar el origen de la demencia.

El dx clínico de la DTA por lo general se basa en una presentación clínica típica y la ausencia por **diagnóstico de exclusión** de otras causas de demencia. Los clínicos conjuntan los hallazgos de exámenes físicos, neurológicos, entrevista psiquiátrica y valoración psicológica y aplican los criterios diagnósticos más comúnmente utilizados. Primero, en la práctica diaria los clínicos deben ser capaces de llegar a un dx confiable debido al gran número de errores falsos-negativos y falsos-positivos, los cuales pueden tener consecuencias muy serias para los pacientes y los familiares ya que la persona puede pasar por un cuadro demencial parcialmente reversible (por ejemplo, hidrocefalia normotensa). Segundo, sin un método diagnóstico exacto resulta difícil comparar los resultados de investigaciones llevadas a cabo con diferentes tipos de pacientes y finalmente, se debe señalar que un dx acertado es necesario para poder valorar la eficacia de tratamientos farmacológicos propuestos para este tipo de alteraciones (Forette et al, 1989).

Existen tres tipos de criterios diagnósticos (Kukuli et al, 1983, 1990) que se utilizan con más frecuencia, estos son:

(a) El del **DSMIII-R**, "*Diagnostic and Statistical Manual of Mental Disorders*" (1987) que se resume en la tabla siguiente:

a. Se debe obtener el dx de demencia
b. Contar con inicio insidioso y un curso generalmente progresivo
c. Exclusión mediante historia clínica, examen físico y estudios de laboratorio de otras causas específicas de demencia.

Tabla 1 Criterios Diagnósticos del DSMIII-R.

(b) El **Diagnóstico Clínico de Enfermedad de Alzheimer** del Grupo de Trabajo del Instituto Nacional de Enfermedades Neurológicas y Contagiosas de los Institutos Nacionales de Salud de EU "**NINCDS/ARDA**" (McKhann et al, 1984). Este grupo se instituyó con el fin de formular y describir los criterios clínicos de la DTA, planear metodologías de investigación y describir el curso natural de

la enfermedad. Propusieron una distinción que se resume en la tabla siguiente (McKhann et al, 1984).

POSIBLE DTA , demencia diagnosticada mediante un examen clínico que incluye empeoramiento progresivo de una o más áreas de funcionamiento cognitivo en presencia de una alteración sistémica o cerebral secundaria pero no considerada como la causa principal de la demencia. En sujetos mayores (>75 años) es común que tengan más de un dx y por lo tanto los clínicos consideran la categoría de POSIBLE como más factible.
PROBABLE DTA , diagnosticada en ausencia de alteraciones sistémicas o cerebrales que pudieran explicar la demencia. En términos más prácticos el inicio temprano de la DTA (<65 años) se clasifica en muchas ocasiones como PROBABLE debido a la ausencia de enfermedades sistémicas
DTA DEFINITIVA , que comprende los criterios clínicos para DTA con estudios de biopsia o autopsia con base en una serie de signos y síntomas neurológicos y psiquiátricos.

Tabla 2 Distinción de tipos de DTA propuesta por el grupo de trabajo de los Institutos Nacionales de Salud de EU.

Y los criterios propuestos para el dx probable de DTA se resumen a continuación.

I. El dx incluye:

Decremento en el funcionamiento intelectual, establecido mediante un examen clínico.
Los déficits de memoria y otros procesos cognoscitivos importantes (lucidez, vigilancia, orientación a lugar y tiempo) deben ser documentados.
La demencia debe ser confirmada por pruebas neuropsicológicas.
Inicia insidioso después de los 40 años (más o menos después de los 65 años de edad).
Empeoramiento constante de pérdida de memoria.
Sin deterioro en el estado de conciencia.
Ausencia de enfermedad sistémica u otro trastorno cerebral que pudiera explicar la disfunción cognoscitiva.

II. El dx está apoyado por:

Deterioro progresivo de funciones específicas tales como lenguaje (afasia), habilidad motora (apraxia), percepción (agnosia).
Deterioro en actividades de la vida cotidiana y alteración en patrones de comportamiento.
Historia familiar de enfermedad similar.
Manifestación de atrofia cerebral en estudios de Imagen como TAC o IRM, cambios no específicos o sin cambios en el EEG, punción lumbar normal.

III. Otras características clínicas acordes con el dx son:

Mesetas en el curso de progresión de la enfermedad.
Six asociados como depresión, insomnio, incontinencia, delirios, ilusiones, alucinaciones, arranques emotivos, trastornos sexuales, pérdida de peso.
Especialmente en una fase avanzada de la enfermedad, otras anomalías neurológicas tales como crisis epilépticas y trastornos motores como tono muscular aumentado, mioclonus o trastornos de la marcha.

IV. El dx es incierto o poco probable si

La enfermedad tiene un inicio súbito o apopléjico
Hay hallazgos neurológicos focales tales como hemiparesia, pérdida sensorial, déficits del campo visual.
Las crisis epilépticas, la falta de coordinación y los trastornos de la marcha se presentan al inicio o muy temprano en el curso de la enfermedad.

(c) Criterio de Investigación Diagnóstica de Eldorfer y Cohen (1980). Estos criterios fueron establecidos en la Universidad de Washington para la investigación en la degeneración neuronal primaria tipo Alzheimer y están resumidos a continuación.

I. Características Clínicas para Inclusión:

Tipo de Inicio
Disminución o pérdida en al menos dos de las siguientes habilidades: aprendizaje, atención, memoria y orientación.
Disminución en al menos una de las siguientes habilidades cognitivas: cálculo, abstracción y juicio, comprensión.
Problemas en al menos una de las siguientes áreas: habilidad para trabajar, para relacionarse con la familia, con sus compañeros y su rol social.
Indicación de disfunción cerebral en al menos una de las siguientes: Atrfia cerebral en TAC, EEG anormal.
Escala de Hachinski menor a 4 puntos

II. Criterios médicos de exclusión:

Signos neurológicos focales
Historia médica de: infarto al miocardio o enfermedad cardiovascular crónica; accidente cardiovascular; alcoholismo o abuso de sustancias; enfermedad psiquiátrica crónica; sífilis; daño cerebral temprano provocado por hipoxia; enfermedad renal crónica, hepática, pulmonar o endocrina; enfermedad de Huntington, Parkinson, Pick, o desajustes neurológicos que afectan a una región específica del cerebro; demencia multi-infarto clara en IAM; enfermedad cardiovascular hipertensa.
Pseudodemencias: enfermedad maníaca primaria; enfermedad depresiva primaria; enfermedad física, metabólica; Intoxicaciones; Interacción con drogas.

Dada la gran variedad de posibles causas y la complejidad del análisis de los síntomas y signos presentados por los pacientes, para realizar el dx diferencial de DV y DTA se requiere llevar a cabo un conjunto grande de pruebas, que se han agrupado de la siguiente manera:

1. **Demográficos:** este grupo está constituido por el sexo, edad, estado civil, años de escolaridad y ocupación del paciente.
2. **Antecedentes:** se considera la historia clínica del paciente que incluye datos sobre tabaquismo, alcoholismo, antecedentes hereditarios, hipertensión arterial, obesidad, hiperlipidemia, trombosis e historia de depresión. Estos datos se obtienen haciendo preguntas al paciente o a algún familiar y se utilizan para encontrar factores de riesgo para aterosclerosis.
3. **Síntomas:** incluyen datos sobre el tiempo de evolución de la enfermedad, si tiene problemas de orientación, cambios en la personalidad, si existe pérdida en el control de esfínteres, apraxia, agnosia, problemas en cálculo numérico, rol social, trastornos del lenguaje y síntomas psicóticos.
4. **Escalas Neurológicas y Neuropsicológicas:** en un intento por confirmar el dx clínico de Demencia Vascular se han llevado a cabo diferentes métodos entre los que se tienen escalas isquémicas como la de Hachinski (Hachinski et al, 1974). Es un conjunto de observaciones obtenidas de la historia y del examen clínico que sugieren la presencia de enfermedad cerebrovascular o severidad suficiente como para ser causa de demencia. Sin embargo, en algunos estudios prospectivos (Villardita, 1993) se han demostrado las dificultades que se presentan en aquellos casos en que se puede encontrar una Demencia Tipo Alzheimer en un sujeto con historia de EVC previa, con lo que el índice sería sugestivo de etiología vascular sin brindar beneficio en el dx diferencial (Gorelick et al, 1993; Von Reutern, 1991). Lo mismo sucede ante sujetos que se traslapan en las alteraciones fisiopatológicas que ocurren en la demencia; es decir, sujetos con hallazgos histológicos de DTA en quienes se encuentran también cambios compatibles con un origen vascular de la demencia.

Otra escala neurológica es la de Loeb (Loeb, 1988; Cummings, 1985), otro intento por diferenciar a la DV de la DTA, lo que constituye un verdadero reto en la actualidad (López, 1990; Román et al, 1993; McKhann et al, 1984; Kukull et al, 1990; Mirsen et al, 1991).

En ambas escalas (Hachinski y Loeb) se valora el inicio de la demencia de manera brusca o abrupta, la evolución a pasos, la presencia de síntomas y/o signos de focalización neurológica y la historia de hipertensión arterial sistémica. También se valora la historia de EVC en el caso de la escala de Loeb. El resultado de la prueba se valora en un rango de 0-12 mientras que en la segunda es de 0-10, considerándose como sugestivo de DV en ambas escalas cuando se determina un valor mayor de cinco.

El perfil neuropsicológico se ha postulado como de utilidad para distinguir la demencia cortical, como es el caso de la DTA de aquella subcortical, como es la observada en diferentes padecimientos. En la siguiente tabla se muestra la clasificación de demencia subcortical y cortical.

Se ha encontrado que la demencia cortical se acompaña de afasia, apraxia y trastornos en la orientación visoespacial, a diferencia de la bradipsiquia y los trastornos motores observados en la demencia subcortical, estas características no permiten diferenciar la Demencia Tipo Alzheimer de la Demencia Vascular dado que esta última se considera de tipo mixto.

<p>DEMENCIAS CORTICALES <i>Enfermedad de Alzheimer</i> Degeneración Lóbulo Frontal</p>
<p>DEMENCIAS SUBCORTICALES Síndromes Extrapiromidiales Enfermedad de Parkinson Enfermedad de Huntington Parálisis Supranuclear Progresiva Enfermedad de Wilson Degeneraciones Espinocerebelosas Calcificación idiopática de Ganglios Basales Hidrocefalia Síndrome Demencial de Depresión Esclerosis Múltiple Encefalopatía por Virus de VIH <i>Demencia Hoosiers</i> Enfermedad de Binswanger</p>
<p>DEMENCIAS CON DISFUNCIÓN CORTICAL Y SUBCORTICAL COMBINADA <i>Demencia Multi-Infarto</i> Demencias Infecciosas Demencias por Virus Lentos Encefalopatías Tóxicas y Metabólicas Enfermedades Sistémicas Endocrinopatías Estados de Deficiencia Intoxicación por Fármacos Exposición a Metales Pesados Demencias Industriales Síndromes Demenciales Misceláneos Post-traumático, Post-anóxico, Neoplásico, etc.</p>
<p>(Cummings & Benson, 1992)</p>

Tabla 3 Clasificación de los cuadros demenciales.

Las pruebas neuropsicológicas que se utilizaron en este trabajo son:

Mini Mental State Examination (MMSE): es una prueba diseñada para valorar funciones cognitivas en forma rápida. Esta formada por 30 reactivos, divididos en dos secciones, la primera requiere respuestas

verbales en las que se evalúa orientación, memoria y atención, con un puntaje máximo de 21 puntos. La segunda, valora actividades básicas de lenguaje, cálculo y constructivas; su puntaje máximo es de 9 puntos (Folstein et al, 1975).

Escala de Depresión Geriátrica: fue diseñada específicamente para valorar depresión en sujetos ancianos. Contiene 30 reactivos, los puntajes de 0-10 se consideran normales y los mayores de 11 son sugestivos de depresión (Mattis, 1976; Díaz et al, 1993).

Escala de Actividades Cotidianas: se utiliza para valorar habilidades adaptativas sociales y personales en las actividades de la vida diaria. Esta formada por 280 reactivos que no se administran directamente al sujeto, sino al familiar o cuidador primario y comprende las siguientes áreas: cuidado personal, actividades del hogar y laboral, relaciones interpersonales, afecto, memoria episódica y sistema motor (Kachaturian, 1985; Díaz et al, 1993)

5. **Electrofisiología**: en este grupo se encuentran los estudios de electroencefalograma (EEG) y de P300. Los estudios de EEG son normales en un inicio y a medida que la enfermedad progresa puede presentarse una lentificación en la actividad de fondo. El EEG (Helkala et al, 1991) tradicionalmente se ha considerado inespecífico en el estudio de las demencias degenerativas. Estudios de electrofisiología como los potenciales evocados, específicamente el P300 han servido para corroborar los hallazgos de la valoración neuropsicológica ya que se han relacionado al recortamiento de la latencia del componente P300 en pruebas de atención y de memoria (Patterson et al, 1983).
6. **Neuroimagen y otros estudios**: Los estudios de imagen ya sea TAC o IRM valoran los aspectos patológicos de la DTA tales como:
 - signos de atrofia cerebral,
 - agrandamiento de las cavidades ventriculares, especialmente del tercer ventrículo,
 - ampliación temporalmente acentuada de los surcos encefálicos, y las uniones corticomedulares indistintas en el parénquima.

Se ha demostrado que el curso de deterioro clínico en pacientes demenciados está correlacionado de manera importante con los cambios antes mencionados (DeLeon et al, 1980 y 1983; Fox et al, 1975) y con el número de infartos cerebrales presentes en el caso de la DMI (Mirsenth et al, 1988).

Otras pruebas que pertenecen a este grupo son la radiografía simple de tórax, exámenes de laboratorio y doppler de vasos de cuello. La radiografía y el doppler son para identificar enfermedad vascular carotídea, mientras que el examen de laboratorio está encaminado a detectar factores de riesgo para aterosclerosis como es la presencia de hiperlipidemia e hiperglucemia.

En ausencia de un marcador específico para el dx de DTA, que no implique estudios invasivos como es la **Biopsia Cerebral**, es difícil encontrar una clara diferenciación entre ambas entidades. Sobre todo, cuando ambas coexisten la contribución relativa de cada una de ellas al deterioro cognitivo del paciente es casi imposible de lograr y se requiere de marcadores diagnósticos objetivos y reproducibles para intentar establecer esta diferencia y así poder decidir el tratamiento, **que no es el mismo empleado para ambas entidades**. E incluso, se pretende obtener marcadores con valor pronóstico ante la existencia de factores de riesgo conocidos para enfermedad vascular, con lo que se lograría incluir en la curva de prevalencia de la DV, ya que esta se puede asociar a factores de riesgo modificables como es la hipertensión arterial.

II.2. Redes Neuronales Artificiales como Tecnología para Clasificación de Patrones.

En este capítulo se definirá, en forma resumida qué es una RNA, como funciona y cuales son las aplicaciones donde se han utilizado. Se hace una pequeña revisión de la historia de cómo surgieron las RNA's y una revisión del algoritmo de retropropagación.

II.2.1. Introducción

El comportamiento de todo organismo está influenciado por aspectos del ambiente en el que se desarrolla y el ambiente mismo responde a aspectos de la actividad actual del organismo (Romanes, 1981). Cuando el ambiente interactúa con el organismo es como si se le aplicara una entrada a un sistema y el organismo genera una respuesta a esa interacción que se puede considerar como la salida del sistema, que a su vez interactúa con el ambiente. Esta interacción se puede representar como se muestra en la figura 1.

Todo organismo, entonces, está constituido por muchos sistemas que ayudan a mantener en equilibrio al cuerpo humano. Uno de los más importantes en nuestro cuerpo es el sistema nervioso, tiene como función principal controlar e integrar la actividad de las diversas partes del cuerpo. Constituye un mecanismo mediante el cual el individuo es capaz de reaccionar a los cambios en el ambiente externo, al tiempo que mantiene su ambiente interno en equilibrio dinámico dentro de los estrechos límites

necesarios para la supervivencia. Está dividido en sistema nervioso central y sistema nervioso periférico. En este trabajo solo haremos referencia al sistema nervioso central.

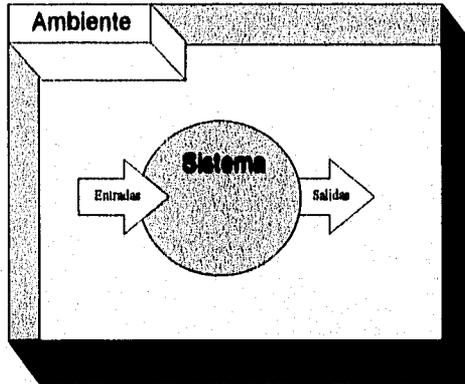


Fig. 1 Interacción ambiente-sistema (Adaptado de Arbib, 1969).

El sistema nervioso central está constituido, en forma simple, de dos partes: el cerebro y la médula espinal, siendo el cerebro la parte más importante. La división anatómica del sistema nervioso se observa en la siguiente figura (Fig. 2).

La capacidad de un organismo de reaccionar a un cambio presupone la existencia de un mecanismo capaz de detectar dicho cambio y que ponga en marcha la respuesta apropiada. Esto requiere un sistema sensorial capaz de registrar toda la gama de cambios que se imponen al organismo desde el exterior, sean éstas de naturaleza química o física. Así, el control de los gases y los electrolitos en la sangre depende de la existencia de un medio para medirlos, del mismo modo que el control del movimiento depende de un mecanismo capaz de medir como se desarrolla continuamente este movimiento de un momento a otro. Este sistema sensorial o receptivo debe ser capaz de distinguir un enorme número de circunstancias diferentes así como de activar un mecanismo capaz de integrar toda la información que le llega del sistema sensorial. Sólo así puede el sistema nervioso producir una respuesta apropiada a todas las circunstancias presentes en un momento dado.

En toda actuación del sistema nervioso, por simple o compleja que sea, se hallan implicados, por tanto, tres elementos: el sensorial o receptivo, el integrador y el efector o motor. De los tres, es el elemento integrador el que sufre un mayor grado de desarrollo en los animales superiores, para llegar a constituir la mayor parte del sistema nervioso.

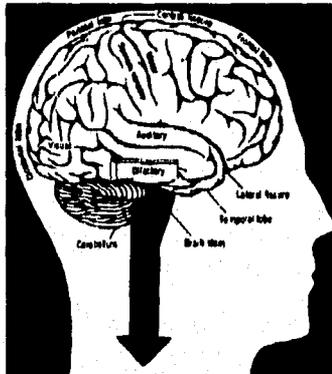


Fig. 2 División anatómica del sistema nervioso central (Tomada de Arbib, 1967).

Con base en lo anterior y en el diagrama de la Fig. 1, se puede modelar al sistema nervioso como un sistema de tres estados, como se muestra en la Fig. 3.

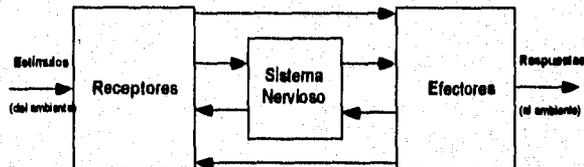


Fig. 3 El sistema nervioso considerado como un sistema de tres estados (Adaptada de Arbib, 1967).

Desde Ramón y Cajal se conoce que el sistema nervioso en general y el cerebro en particular no es una masa uniforme de tejido neural, sino que está constituido por un conjunto de zonas heterogéneas. Cada una de éstas tiene una arquitectura distinta de las otras, comprendiendo un variado número de neuronas de forma diferente, tamaño, respuestas fisiológicas, etc.

La unidad anatómica básica del sistema nervioso es una célula llamada neurona y el número de ellas que tiene un cerebro humano es, sin contar las células del cerebelo, de aproximadamente 10^{11} , es decir, cien mil millones de neuronas.

Los estudios de Ramón y Cajal demostraron que, desde el punto de vista morfológico, la neurona es una célula muy diferente a las otras del organismo, ya que del cuerpo neuronal surgen finas prolongaciones que se desarrollan y ramifican a diferentes distancias; estas prolongaciones pueden ser de dos tipos de acuerdo al papel fisiológico que desempeñan en el paso de señales de una neurona a otra:

- a) la recepción a través de las *dendritas* que junto con el *soma* reciben aferencias de otras neuronas y.
- b) la transmisión por el *axón*, que termina en las *dendritas* o *somas* de otras células para establecer un contacto funcional con ellas.

Una vez que se estableció el hecho de que el sistema nervioso está constituido por neuronas se hizo evidente la necesidad de estudiar cómo envían y reciben señales que se transmiten entre sí. Un importante número de investigadores entre los que se puede citar a Sherrington (1906), Hodgkin y Huxley (1952), Katz (1966, 1969), Eccles (1964), Kuffler y Nicholls (1976) y Miledi (1973) entre muchos otros, han estudiado cómo una neurona se comunica con otra. Precisamente, al punto de contacto entre dos neuronas Sherrington lo denominó *sinapsis*.

A medida que el axón se aleja del cuerpo neuronal, su diámetro disminuye y se ramifica ensanchándose en los extremos digitales para formar el botón sináptico, también llamado terminal nerviosa. Este botón puede hacer contacto con la superficie de una dendrita o del cuerpo celular de la neurona (contacto sináptico axo-somático); con la terminal axónica de otra célula (contacto axo-axónico); mientras que una dendrita podría interactuar con otra (contacto dendro-dendrítico).

De acuerdo con la dirección que sigue el flujo de información de una neurona a otra, se denomina *presinapsis* a la terminal axónica y *postsinapsis* a la membrana de la dendrita receptora (Garduño et al, 1986).

La transmisión de señales de una célula a otra en una *sinapsis* es un proceso muy complejo en el cual sustancias neurotransmisoras específicas son liberadas hacia el receptor que se encuentra en otra célula. El efecto que se provoca es un incremento (**excitación**) de potencial eléctrico en el interior del cuerpo de la célula receptora con respecto al exterior. Si el efecto es excitador y el cambio que se produce en el potencial alcanza un valor de **umbral**, se genera un impulso (**potencial de acción**) de intensidad y duración fija que se envía por el axón hacia las células de la red que reciben proyecciones axónicas de la neurona que ha disparado. En la Fig. 4 se muestra esta comunicación.

II.2.2. La Neurona como Integrador

Cada neurona es un integrador de estímulos (neurotransmisores) que van del campo dendrítico al cuerpo celular. Algunas de las áreas receptoras (membrana subsináptica) en las dendritas y el cuerpo celular son excitadoras, otras inhibitorias. Además, la actividad inhibitoria presináptica puede afectar indirectamente algunos receptores excitadores. En cualquier momento, una neurona puede recibir cientos o aún miles de estímulos en sus áreas excitadora e inhibitoria de la membrana.

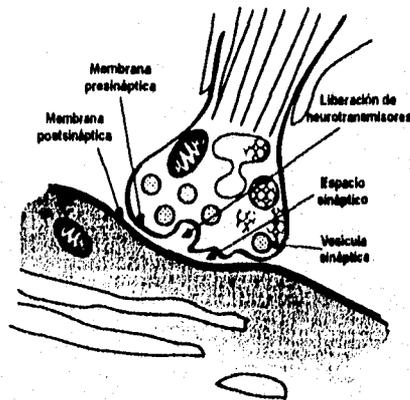


Fig. 4 Sinapsis biológica (Adaptada de Kandel, 1968).

La mayor parte de las neuronas reciben gran cantidad de sinapsis. En esta gran actividad, la neurona reacciona y puede responder. Si la suma del Potencial Post-sináptico Excitador (PPSE) excede de la de los Potenciales Post-sinápticos Inhibitorios (PPSIs), el segmento inicial del axón puede excitarse para producir un potencial de acción en el axón (Fig. 5). Si la suma algebraica de estos potenciales (PPSEs y PPSIs) no basta para estimular el segmento inicial, no se genera el potencial de acción en el axón.

La despolarización del segmento inicial al voltaje crítico es un prerrequisito para la generación de un potencial de acción. Así cada complejo dendrita-cuerpo celular de una neurona es un centro de integración en miniatura, responderá con un potencial de acción según el efecto neto de la actividad sináptica Inhibitoria en la membrana receptora de la neurona. El axón es el vehículo para señalar la información codificada, a través de los potenciales de acción, del complejo dendrita-cuerpo celular a otras neuronas o efectores (músculos o células glandulares).

Cada membrana postsináptica de las neuronas y los efectores señalados contiene cientos o miles de receptores. Cada receptor, compuesto de proteínas macromoleculares que actúan como descifradores especializados, responde a un estímulo dado en forma probablemente predeterminada. Por ejemplo, la acetilcolina es un agente excitador de la placa motora terminal (contrae el músculo voluntario) e inhibitorio en la sinapsis del nervio vago en el tejido cardíaco (disminución de la frecuencia cardíaca).

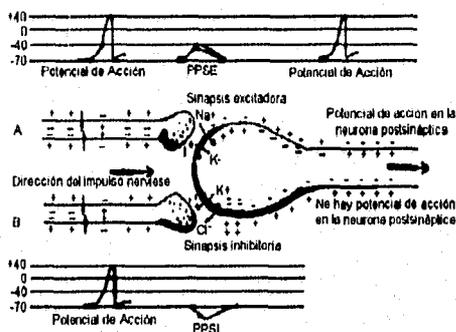


Fig. 5 Secuencias en las transmisiones (A) excitadora y (B) inhibitoria de neuronas presinápticas (izquierda) a través de sinapsis a una neurona postsináptica (derecha). (Noback et al, 1984).

A su vez, una neurona puede estar influida por su propia actividad a través de un circuito de retroacción negativa que comprende a la interneurona. Ésta, llamada célula de Renshaw, está entre el axón de una neurona motora inferior de la médula espinal y la región dendrita-cuerpo celular de la misma neurona motora y otras. El axón colateral termina en una sinapsis excitadora en la célula de Renshaw; a su vez, los axones de esta célula tienen sinapsis inhibitorias con la neurona motora matriz inferior.

Las neuronas del SNC pueden clasificarse en dos grandes grupos:

1. de proyección: tienen axones largos que van de una región del sistema nervioso central a otra;
2. locales de circuito: tienen axones que pasan e interactúan con neuronas cercanas.

Excitación e inhibición presinápticas. Una neurona puede ejercer una influencia presináptica a través de sinapsis axo-axónica con otra neurona. Así, el axón de la neurona A hace sinapsis con el de la neurona B. Los neurotransmisores liberados por la neurona A ejercen sus efectos en los receptores localizados en la membrana celular de la terminal presináptica de la neurona B. Se supone que estos receptores actúan controlando la cantidad de la entrada de Ca^{++} en las terminales nerviosas de la neurona B.

En la excitación presináptica (facilitación), la entrada de Ca^{++} del espacio sináptico es incrementada y seguida por un aumento en la cantidad de transmisor liberado por la neurona B. En la inhibición presináptica, hay una disminución en la entrada de Ca^{++} del espacio sináptico y esto es seguido por una disminución en la cantidad de transmisor liberado por la neurona B. Las interacciones de la excitación e inhibición pre y postsinápticas contribuyen al proceso neural refinado en el sistema nervioso central.

Importancia de la inhibición. La inhibición es un proceso muy importante. En una explicación sencilla imagine las consecuencias de las influencias excitadoras sin actividad de freno (como los frenos en un coche) para conservar el control de estas influencias. Las influencias inhibitorias pueden ayudar a conducir y regular los efectos de excitación y la actividad directa para lograr el fin deseado. Por ejemplo, al aprender los movimientos complicados de la escritura, al principio un niño tiene dificultades de control porque, en parte, se ejecutan demasiados movimientos innecesarios. Durante el proceso de aprendizaje estos movimientos se eliminan gradualmente a través de la inhibición. Se conservan los conductos excitadores deseados para producir los movimientos focales. En este caso es importante la inhibición de los movimientos no esenciales e innecesarios en el proceso de aprendizaje.

11.2.3. Algunos conceptos generales relacionados con el proceso neuronal.

El proceso consiste en una serie de acciones u operaciones dirigidas hacia algún extremo terminal. En el sistema nervioso, el proceso neuronal se efectúa por interacciones entre complejos de neuronas en un centro (centro procesador). Siguiendo el proceso, la información neural es transmitida a otros centros. Las secuencias de los centros procesadores se pueden organizar como vías sensitivas o motoras o circuitos de retroacción. Por ejemplo, al tocar un objeto caliente puede provocar varias reacciones, entre ellas las percepciones de dolor y calor. El objeto caliente estimula los receptores de dolor y calor en la piel y, en respuesta, se envían impulsos nerviosos a los centros en la médula espinal. Al seguir el proceso neuronal en la médula espinal, la información neural puede ser distribuida a otros centros.

1. Se transmite alguna información a centros motores en la médula espinal, y el resultado es la estimulación de los músculos para retirar la mano.
2. Se puede transmitir otra información a través de una secuencia de centros (vía del dolor) para un proceso posterior.

Los centros superiores de la vía del dolor participan en el conocimiento del mismo y en la apreciación de características como la intensidad y localización del dolor.

Proceso en serie. En el proceso en serie, las neuronas están dispuestas en orden secuencial; por ejemplo, una neurona (o grupo de neuronas) hace sinapsis con otra neurona (o grupo de neuronas) que a su vez hace sinapsis con otra neurona, músculo o células glandulares. Esa secuencia en serie aparece en el reflejo rotuliano en el que la neurona sensitiva de un receptor

del huso neuromuscular hace sinapsis en la médula espinal con neuronas motoras que inervan y estimulan las fibras musculares para contraerse.

Proceso en paralelo. En el proceso en paralelo, en una constelación de receptores, cada uno de estos puede hacer sinapsis con secuencias diferentes de neuronas, con frecuencia llamados conductos. Estos actúan para procesar y llevar información neural en secuencias paralelas. En algún centro o centros estos conductos se integran entre sí y con otros. Así, los receptores en la cabeza y cuerpo mandan impulsos simultáneamente por varios conductos paralelos a través del sistema nervioso central. Estas señales se pueden procesar paralelamente y después compararse (mecanismos de comparación); su "significado" se transmite a otros centros. En un contexto más amplio, el proceso paralelo es una especialización filogenéticamente desarrollada del organismo para facilitar más interpretaciones sutiles de los estímulos del medio exterior e integrar respuestas apropiadas a las demandas complejas del ambiente.

Organización jerárquica. La jerarquía es una organización sistemática, cuyos miembros o niveles pueden estar asignados a rangos específicos, relacionándose entre sí. Muchas vías y sistemas funcionales del sistema nervioso se organizan jerárquicamente con otro nivel sucesivo en el proceso de secuencia en un nivel más complejo. Por ejemplo, la retina está estructurada con los fotorreceptores (bastones y conos) para responder a la luz, seguido de un nivel más avanzado de proceso por las células bipolares y, a su vez, seguido de un nivel o proceso más avanzado por las células ganglionares de la retina.

Modulación. La modulación es una expresión del proceso neuronal por el cual el sistema nervioso regula y adapta el nivel en reposo o basal de excitabilidad y sensibilidad de una población o poblaciones de neuronas. Las influencias de modulación pueden elevar el nivel basal de excitabilidad y hacer más eficaces ciertas aferencias sensitivas o disminuir el nivel basal y hacer menos eficaz la misma aferencia sensitiva. La modulación conlleva la influencia del nivel de actividad de una neurona o población de neuronas de un nivel a otro. Un ejemplo generalizado de modulación se expresa en las variaciones en nuestra atención a la estimulación durante los diferentes estados y fases del sueño, somnolencia, aburrimiento y vigilia. La modulación altera el grado, cualidad y velocidad de sensibilidad a la estimulación tanto en el sistema sensitivo como en el motor.

Algunas de estas propiedades de las neuronas biológicas también se han tomado para modelar neuronas. En la siguiente sección se describe un modelo de neurona.

11.2.4. Modelo de Neurona

McCulloch y Pirts (1943) propusieron un modelo formal muy simple de neurona que incluye sólo algunas propiedades básicas de las neuronas biológicas mencionadas anteriormente. Estas son:

(a) **Suma Espacial:** el cambio global en el potencial de membrana de la neurona es igual a la suma algebraica de los efectos ponderados de todas las entradas que hayan estado activas en el tiempo anterior.

(b) **Propiedad de Umbral:** toda célula tiene un valor de umbral que debe ser alcanzado por los cambios en sus entradas para poder generar potenciales de acción que son transmitidos a otros elementos de la red neuronal.

Estas características son similares a las de las neuronas biológicas, en la Fig. 6 se muestra esta similitud.

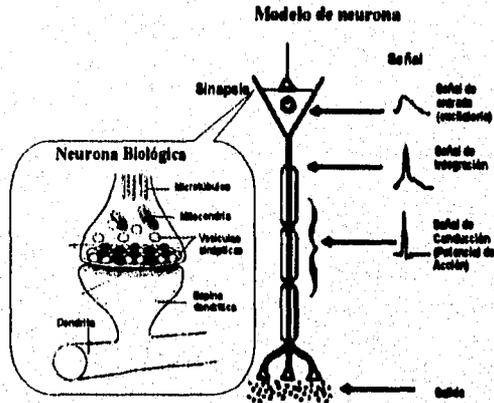


Fig. 6 Modelo de neurona y sinapsis biológica.

Específicamente, cada elemento de procesamiento (neurona artificial básica) se puede modelar como un dispositivo no lineal con interconexiones ponderadas w_{ij} también llamadas pesos sinápticos. El valor de salida de cada neurona está basado en todas sus conexiones de entrada. El valor de salida de la j -ésima unidad se calcula como:

$$net_j = \sum_{i=1}^n x_i w_{ij} \quad (1)$$

donde: net_j representa el estado en la neurona j ;
 x_i son todas las conexiones que llegan al elemento j , y

w_{ij} es el peso asociado a cada conexión que va de la neurona i a la j .

El peso w_{ij} puede ser positivo o negativo correspondiendo a una sinapsis excitatoria o inhibitoria respectivamente; si es cero entonces no existe conexión entre i y j (Hertz et. al., 1991). Una vez que el estado de la neurona se ha calculado, este valor se convierte a un valor de activación para el elemento. Este valor se puede escribir como:

$$a_j(t) = F_j(a_j(t-1), net_j(t)) \quad (2)$$

donde se puede notar que el valor de activación es una función del estado de la red. Ya calculado este valor de activación se puede determinar el valor de salida de la neurona aplicando una función de salida:

$$y_j = f_j(a_j) \quad (3)$$

Como normalmente $a_j = net_j$, entonces esta función se puede escribir como:

$$y_j = f_j(net_j) \quad (4)$$

donde f_j es una función no lineal. La función que comúnmente se utiliza es la sigmoide que se define como:

$$f_j(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

y como el operando es la suma ponderada de la unidad j , se tiene:

$$f(net_j) = \frac{1}{1 + e^{-net_j}} = \frac{1}{1 + e^{-\sum_i w_{ij} x_i}} \quad (6)$$

Después de calcular el valor de la función sigmoide, el resultado nos da la salida de la unidad j , que se envía a lo largo de todas las interconexiones. En la Fig. 7 se muestra esta unidad de procesamiento básica.

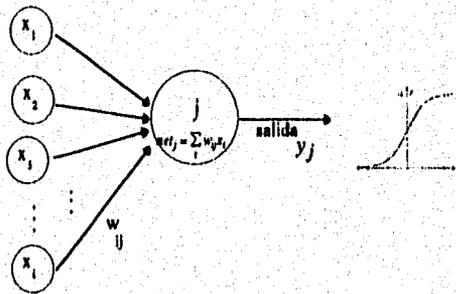


Fig. 7 Esquema de una unidad de procesamiento básica.

Las neuronas biológicas involucran otro tipo de características que no se toman en cuenta en esta descripción; las más significativas son:

1. No siempre pueden aproximarse con dispositivos de umbral como en este modelo; responden a su entrada en forma continua, pero la relación no lineal entre la entrada y la salida de la célula es una característica universal. La hipótesis es que la no linealidad es esencial, no así su forma específica; en cualquier caso, las unidades continuas también pueden ser modeladas y algunas veces son más útiles que las unidades de umbral.
2. Producen una secuencia de impulsos, no un sólo nivel de salida.
3. No todas las neuronas tienen el mismo retraso fijo ($t \rightarrow t+1$); no se adaptan todas al mismo tiempo (en sincronía) con respecto a un reloj central.
4. La cantidad de transmisores sinápticos varía impredeciblemente. Este efecto se puede modelar con una generalización estocástica de la dinámica de McCulloch y Pitts.

En resumen, en esta sección se presentó un modelo de neurona, el de McCulloch y Pitts y algunas de sus propiedades. En la siguiente sección se describe la historia de como nacieron las RNA's y los modelos que fueron surgiendo, los problemas que se presentaron y la forma en que se resolvieron.

11.2.5. Arquitecturas y Algoritmos de Entrenamiento

El campo de las Redes Neuronales Artificiales ha crecido debido a un esfuerzo colectivo entre neurofisiólogos y científicos en computación -el primero utiliza a las computadoras para modelar diferentes teorías sobre funciones neuronales biológicas, mientras que el segundo busca la manera de construir computadoras más inteligentes que puedan imitar (sin pretender igualar) la forma en la cual el cerebro trabaja. El procesamiento neuronal es llevado a cabo no por un solo procesador muy poderoso sino por un gran número de procesadores muy simples. La potencia y velocidad del cerebro humano se postula que es consecuencia del trabajo colectivo de millones de procesadores (neuronas) sobre la misma tarea. Cuando muchos procesadores son utilizados en una tarea, la consecuente redundancia hace que el punto final del procesamiento (la respuesta) no se vea significativamente afectada por errores ocurridos en algún o algunos procesadores individuales.

Una red neuronal artificial relativamente simple consiste de una capa de entrada, una capa intermedia (o escondida) y una de salida. Cada capa contiene un conjunto de nodos; la analogía biológica con los nodos son las neuronas. Los nodos en las capas adyacentes están interconectados, estas interconexiones realizan la misma función que las dendritas en el caso biológico. Cada interconexión tiene un cierto peso caracterizado por un valor numérico; si el valor es positivo, la conexión es excitatoria, si es negativo, se considera como inhibitoria. Las observaciones o datos representan la entrada (conjunto de entrenamiento) a la red mientras que el resultado (clasificación, predicción, dx o respuesta) representa la salida.

La señal de entrada presentada a la red se propaga a través de las conexiones dendríticas y resulta en un patrón diferente en la capa intermedia, evocando un patrón en la capa de salida; este patrón es la respuesta (Dayhoff, 1990).

La historia de esta clase de ideas se origina con Aristóteles, aún cuando las bases del modelado neuronal fueron dadas en el trabajo hecho por McCulloch y Pitts en 1943 en el cual se introduce el modelo explicado en la sección anterior.

En 1949, Donald Hebb construye una descripción matemática del concepto de aprendizaje. Con base en los estudios de la neurona y sus observaciones de aprendizaje y condicionamiento clásico descritos en su libro "*Organization of Behavior*" (Hebb, 1949), desarrolló un paradigma de aprendizaje que ahora es conocido con su nombre, *Aprendizaje Hebbiano*. Esta regla establece la eficiencia de un incremento de sinapsis si se tiene una actividad presináptica seguida muy de cerca en el tiempo por una actividad postsináptica. Otras versiones de esta regla la describen como un incremento en el peso entre neuronas en proporción con la correlación de potenciales pre y postsinápticos. En forma clara, si una sinapsis tiene un potencial presináptico positivo y un potencial postsináptico positivo, se debe incrementar el peso entre ellas y si el potencial presináptico es negativo (positivo) y el potencial postsináptico es positivo (negativo), entonces la conductancia sináptica se decrementa. Estas modificaciones al paradigma de Hebb fueron muy significativas en las RNA y actualmente se utilizan.

En 1957 el grupo de Frank Rosenblatt atrajo la atención de todos los investigadores en el área, pues se había centrado en el problema de como encontrar los pesos apropiados para realizar tareas computacionales particulares. Se concentraron en redes de elementos procesadores que llamaron perceptrones, en las cuales las unidades de procesamiento fueron organizadas en capas con conexiones entre una capa y la siguiente. Rosenblatt estudió perceptrones de dos y tres capas pero solamente pudo probar que el perceptrón de dos capas podía separar entradas en dos clases si las dos clases eran linealmente separables (i.e. que un hiperplano pudiera separar el espacio en entradas en dos).

El tipo más simple de perceptrones sin capas intermedias (Rosenblatt, 1962) sirvió para probar la convergencia de un algoritmo de aprendizaje, es decir, un modo de cambiar los pesos iterativamente hasta encontrar el conjunto deseado. Mucha gente expresó gran entusiasmo con

el modelo y esperaba que esas máquinas pudieran ser la base de la Inteligencia Artificial.

Redes muy similares, llamadas "ADALINES" fueron inventadas en la misma época por Widrow y Hoff (Widrow et al, 1960; Widrow, 1962). El adalino de dos capas, como el perceptrón, ajusta sus pesos entre la capa de entrada y la de salida en respuesta al error calculado por la diferencia del patrón obtenido y el deseado. Matemáticamente prueban que el error entre la respuesta deseada y la obtenida debería encontrar un mínimo local bajo ciertas condiciones, lo más importante, que los entornos debían ser linealmente separables.

Morvin Minsky es responsable por algunos de los primeros resultados utilizando RNA. Inspirado en el trabajo de McCulloch y Pitts diseñó una máquina de 40 neuronas con sinapsis que ajustaban sus pesos de acuerdo a una tarea específica (aprendizaje Hebbiano). A mediados de los 60's, estudia el modelo realizado por Rosenblatt y cols. y escribe un documento sobre las limitaciones de éste (Minsky et al, 1969). Uno de esas limitaciones era que no podía resolver el problema de la OR exclusiva, pues no se tenía un conjunto linealmente separable.

Rosenblatt también estudió estructuras con más capas de unidades y creyó que éstos podrían solucionar los problemas presentados por Minsky y Papert en su libro "Perceptrons" (Minsky et al, 1969); sin embargo, no encontró un algoritmo de aprendizaje que pudiera determinar los pesos necesarios para implementar una tarea dada. Minsky y Papert dudaron que se pudieran encontrar alguno y sugirieron explorar otras aproximaciones para la Inteligencia Artificial. Con esto, mucha gente de la comunidad científica dejó de estudiar este paradigma durante 20 años.

Sin embargo, algunos investigadores continuaron desarrollando la teoría de redes neuronales. Un tema de impacto fue el de las memorias direccionables por contenido en las cuales, diferentes patrones de entrada eran asociados con otros suficientemente similares. Sin embargo, esto había sido propuesto por muchos otros antes (Taylor, 1956; Steinbuch, 1961) y después fueron redescubiertos por Anderson (Anderson, 1968, 1970; Anderson et al, 1981), Willshaw (Willshaw et al, 1969), Marr (1969, 1971) y Kohonen (1974-1988). Grossberg (1967-1987) hace una reformulación del problema general de aprendizaje en las redes. Marr (Marr, 1969, 1970, 1971) desarrolla teorías de redes del cerebelo, neocórtex cerebral e hipocampo, asignando funciones específicas a cada tipo de neuronas. Mucho gente incluyendo a Marr (Marr, 1982), Von der Malsburg (1973) y Cooper (1974; Noss et al, 1975) estudiaron el desarrollo y funcionamiento del sistema visual.

Otro parte de este desarrollo fue hecho por Cragg y Temperley (Cragg et al, 1954, 1955); ellos reformulan la red de McCulloch y Pitts como un sistema magnético típico entre los físicos ("spin glass"). Caltonello (1961) construye una teoría estática utilizando ideas de la Mecánica Estadística y le incorporó aprendizaje basado en los ideas de Hebb (1949) acerca del aprendizaje en el cerebro. El mismo tema fue tomado en los '70 por Little (Little et al, 1975, 1978; Little, 1974) y de nuevo tomado por Hopfield

(Hopfield, 1982) quien introduce una función de energía y enfatiza la noción de memoria como atractores estables dinómicamente. Hinton y Sejnowski (1983, 1986) y Peretto (Peretto, 1984) construyeron formulaciones utilizando unidades estocásticas.

Pero, a pesar de todos estos intentos, el trabajo que más influencia tuvo en esta década fue el de Rosenblatt; mucha gente se dedicó a desarrollar algoritmos para realizar el ajuste de pesos. Entre ellos, Bryson (Bryson et al, 1969) publica por primera vez un algoritmo llamado "Back-propagation" (retropropagación), posteriormente aparece publicado a mediados de los 70 por Werbos (Werbos, 1974) e independientemente redescubierto aproximadamente en 1985 por Rumelhart, Hinton y Williams (Rumelhart et al, 1986a, b) y por Parker (1985). Le Cun (1985) también propone un algoritmo muy relacionado con los anteriores. Con estos algoritmos se podían resolver muchos de los problemas relacionados a la no separabilidad de los datos. En la Fig. 8 se muestran diferentes arquitecturas y diferentes regiones de decisión. En la primer columna se muestran las diferentes arquitecturas, la segunda indica las diferentes regiones que pueden ser formadas con diferentes redes. Las siguientes columnas están formadas por el problema de la OR Exclusiva y problemas con diferentes tipos de regiones.

Barto (Barto et al, 1981) propone un algoritmo en el cual se tiene información del ambiente relacionado con la red neuronal, obteniendo un tipo diferente de arquitectura. Aunque aún no se encuentra el algoritmo ideal para realizar cualquier tarea computacional, ya se pueden resolver muchos problemas que el perceptrón de una sola capa no podía realizar.

En general, para crear una red neuronal se debe tener una arquitectura y un algoritmo que realice el entrenamiento de la red. También, para garantizar que la red esté trabajando de manera adecuada y que va a llegar a una solución final, con el resultado correcto, se debe tener un teorema de convergencia que lo garantice y esto se debe tener para cada uno de los modelos anteriores.

STRUCTURE	TYPE OF DECISION REGION	EXCLUSIVE OR PROBLEM	CLASSIC XOR PROBLEM REGION	ANY GENERAL DECISION REGION
 SINGLE-LAYER	HALF PLANE BOUNDED BY HYPERPLANE			
 TWO-LAYER	CONVEX SETS OF CLASSIC REGION			
 THREE LAYER	ARBITRARY CONVEXLY BOUNDED BY NUMBER OF NODES			

Fig. 8 Tipos de regiones de decisión que pueden ser formadas por perceptrones de una y dos capas de unidades secundarias y dos entradas (Tomado de Lippmann, 1987).

Como ya se mencionó, existen muchos algoritmos de entrenamiento y diferentes tipos de arquitecturas, pero el algoritmo que comúnmente se utiliza es el de retropropagación ("back-propagation") debido a que puede resolver problemas de clasificación de patrones complejos y desarrollando funciones de mapeo altamente no lineales. Se debe aclarar que no es propósito ni objetivo de esta tesis realizar una revisión exhaustiva de todos los algoritmos de entrenamiento, por eso sólo se revisará el algoritmo de retropropagación. En la siguiente sección se da una pequeña explicación del algoritmo y su teorema de convergencia.

II.2.6. Algoritmo de Retropropagación

El algoritmo de retropropagación emplea una arquitectura de tres o más capas de unidades de procesamiento. La primera capa es la de entrada (las únicas unidades en la red que reciben entrada externa); la siguiente es la intermedia, en la que las unidades de procesamiento están interconectadas tanto a la capa anterior como a la siguiente. La última capa es la capa de salida. Cada unidad de procesamiento recibe señales de todos los nodos de la capa anterior y envía conexiones a todos los de la capa siguiente, sin embargo, se debe notar que las unidades no están conectadas a otras unidades de la misma capa.

La red de retropropagación es entrenada con una técnica llamada de aprendizaje supervisado, en la que a la red se le presenta una serie de pares de patrones y cada par consta de un patrón de entrada y uno de salida; cada patrón es un vector de números reales. El patrón de salida es la respuesta deseada al patrón de entrada y es usado para determinar el error en la red cuando se realiza el ajuste de pesos.

El patrón de salida algunas veces se diseña para representar una clasificación del patrón de entrada. De esta forma, la red puede ser representada con una serie de patrones de entrada junto con la clasificación para cada patrón de entrada. En otras aplicaciones, la salida deseada es simplemente el patrón deseado por el patrón de entrada y la red es entrenada para ser un sistema de mapeo de patrones. Un entrenamiento típico de retropropagación puede realizar cientos o miles de iteraciones de entrenamiento.

La unidad de procesamiento básica de retropropagación también está basada en el modelo de McCulloch y Pitts, por lo tanto, el elemento j realiza una suma pesada de sus entradas x_i y tiene un valor de salida y_j , que se utiliza durante el ajuste de pesos en la etapa de aprendizaje. Después de que se ha terminado el entrenamiento, los pesos se dejan fijos y estos valores son los que se utilizan durante la sesión de pruebas.

Como el entrenamiento se realiza con un algoritmo de aprendizaje supervisado, a la red se le presentan pares de patrones (un patrón de entrada con un patrón de salida deseado). En cada presentación, los pesos

son ajustadas para decrementar la diferencia entre la salida de la red y la salida deseada.

El algoritmo de aprendizaje de retropropagación lleva a cabo un paso de propagación hacia adelante y uno de propagación hacia atrás. Ambas pasadas se realizan para cada presentación del patrón durante la etapa de aprendizaje.

La **propagación hacia adelante** comienza con la presentación de un patrón de entrada a la capa de entrada de la red y continúa con el cálculo del nivel de activación que se propaga hacia la capa escondida. En cada capa, todas las unidades de procesamiento suman sus entradas y aplican una función no lineal (sigmoide) para calcular la salida. Las unidades de la capa de salida producen la salida de la red.

El paso de **propagación hacia atrás** se inicia con la comparación del patrón de salida de la red con el patrón deseado, lo que permite calcular una diferencia o "error". La propagación hacia atrás calcula el valor de error para las unidades escondidas y cambia las pesas comenzando con la capa de salida y moviéndose hacia atrás a través de las capas de unidades escondidas sucesivas. En este paso, la red corrige sus pesas de tal forma que el error observado se decremente en la siguiente etapa del entrenamiento.

El valor de error δ refleja el error asociado con la unidad de procesamiento; este parámetro se utiliza durante el procedimiento de corrección de pesas mientras se lleva a cabo el entrenamiento de la red. Un valor muy grande muestra que se debe realizar una corrección grande a las pesas y el signo dice la dirección en la cual las pesas deben ser cambiadas. La siguiente figura (Fig. 9) servirá como referencia para toda la explicación del algoritmo.

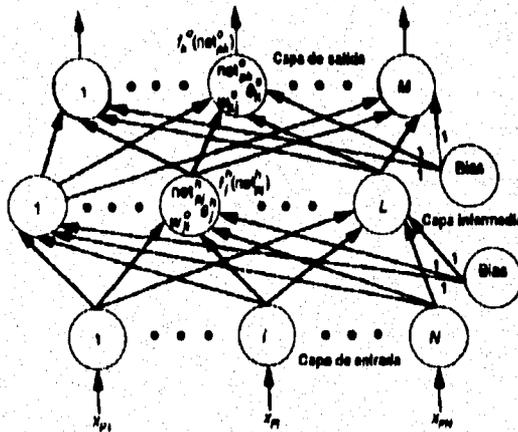


Fig. 9 Arquitectura de tres capas para el algoritmo de retropropagación (tomada de Freeman et al, 1991).

Propagación hacia adelante.

Como ya se mencionó, el paso de propagación hacia adelante se inicia cuando se presenta un patrón de entrada a la red (i.e. un vector de entrada, por ejemplo, $x_p = (x_{p1}, x_{p2}, \dots, x_{pn})$). Cada unidad de entrada corresponde a un valor en el vector de patrones de entrada. Después de que se determinan los niveles de activación para la primera capa de unidades, las siguientes capas llevan a cabo la propagación hacia adelante, la cual determina los niveles de las otras capas de unidades. Los valores de salida de las unidades que llegan a la unidad j se suman como en el modelo de McCulloch y Pitts.

La capa de entrada es un caso especial, sus unidades no llevan a cabo la suma pesada de sus entradas puesto que cada unidad simplemente asume el valor correspondiente del vector de entrada.

Algunas redes de retropropagación emplean una unidad llamada "bías" como parte de todas las capas; esta unidad tiene un valor de activación constante de 1; en la Fig. 9 se denota como θ_j^h para la capa intermedia, los índices cambian dependiendo de la neurona y la capa a la que pertenezca. Cada unidad "bías" está conectada a todas las unidades en la siguiente capa y sus pesos se ajustan durante la propagación del error hacia atrás. Esta unidad contribuye como un término constante en la suma net_j . Esto es equivalente a trasladar la función sigmoide a la izquierda o a la derecha del eje y .

Entonces, las unidades de entrada distribuyen los valores a las unidades de la capa escondida. La entrada a la j -ésima unidad está dada por:

$$net_n^h = \sum_{p=1}^n w_{pn}^h x_p + \theta_n^h \quad (7)$$

donde: w_{pn}^h es el peso de la conexión de la j -ésima unidad de la capa de entrada;
 θ_n^h es el término bías.

Se asume que la activación de este nodo es igual al estado de la neurona, entonces la salida del nodo es:

$$i_n = f_i^h(net_n^h) \quad (8)$$

Las ecuaciones para los nodos de salida son:

$$net_\mu^o = \sum_{n=1}^N w_n^o i_n + \theta_\mu^o \quad (9)$$

$$o_\mu = f_o^o(net_\mu^o) \quad (10)$$

El conjunto inicial de pesos representa una primera aproximación a el conjunto de pesos finales referentes al problema. El cambio de pesos para diferentes tipos de redes se puede hacer con la regla delta:

$$w(t+1)_i = w(t)_i + 2\mu e_k x_{ki} \quad (11)$$

donde: μ es una constante positiva;

x_{ki} es la i -ésima componente del k -ésimo vector de entrenamiento;

E_k es la diferencia entre la salida actual y el valor correcto:

$$e_k = (d_k - y_k)$$

Se tienen ecuaciones similares cuando la red tiene más de dos capas o cuando la función de salida es no lineal. Más adelante se mostrarán estas ecuaciones.

Propagación hacia atrás.

a) Cambio de pesos en la capa de salida.

Siguiendo la regla delta, el error para el k -ésimo vector de entrada es $e_k = (d_k - y_k)$, donde la salida deseada es d_k y la salida obtenida es y_k . Pero cuando se tienen múltiples capas es necesario hacer un cambio de notación. Entonces, se define el error en una sola unidad de salida como $\delta_\mu = (y_\mu - o_\mu)$, donde el subíndice " μ " se refiere al μ -ésimo vector de entrenamiento y " k " se refiere a la k -ésima unidad de salida. En este caso, y_μ es la salida deseada y o_μ es la salida de la k -ésima unidad. El error que por la regla delta generalizada (Rumelhart et al, 1986a, b) se minimiza es la suma de los cuadrados de los errores de todas las unidades de salida:

$$E_p = \frac{1}{2} \sum_{k=1}^n \delta_k^2 \quad (12)$$

El factor de $1/2$ en la ec. (12) es para el cálculo de derivadas que más adelante se necesitará.

Para determinar la dirección en la cual se deben cambiar los pesos se calcula el negativo del gradiente de E_p , ∇E_p , con respecto a los pesos w_p . Entonces, se pueden ajustar los valores de los pesos tal que el error total se reduzca; esto nos hace pensar en una superficie en el espacio de pesos. Un ejemplo de este tipo de superficies se encuentra en la Fig. 10.

Para facilitar el proceso, se considerará cada componente de ∇E_p separadamente. De la ec. (12) y de la definición de δ_μ se tiene:

$$E_p = \frac{1}{2} \sum_i (y_{\mu i} - o_{\mu i})^2 \quad (13)$$

y

$$\frac{\partial E_p}{\partial w_v^i} = -(y_n - o_n) \frac{\partial f_n'}{\partial (net_n^i)} \frac{\partial (net_n^i)}{\partial w_v^i} \quad (14)$$

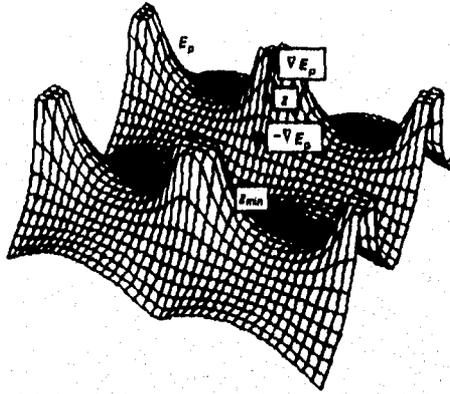


Fig. 10 Superficie imaginaria en el espacio de pesos. Se muestra la dirección del gradiente descendente (Tomada de Freeman et al, 1991).

Por el momento no se evaluará la derivada de $f_n'(net_n^i)$. El último factor en la ec. (14) es:

$$\frac{\partial (net_n^i)}{\partial w_v^i} = \left(\frac{\partial}{\partial w_v^i} \sum_{j=1}^i w_{vj}^i + o_i \right) = 1_n \quad (15)$$

Combinando las ecs. (14) y (15) se tiene para el gradiente descendente:

$$\frac{\partial E_p}{\partial w_v^i} = (y_n - o_n) f_n''(net_n^i) 1_n \quad (16)$$

Los pesos en la capa de salida son cambiados de acuerdo a:

$$w_v^i(t+1) = w_v^i(t) + \Delta_p w_v^i(t) \quad (17)$$

donde:

$$\Delta_p w_v^i(t) = \eta (y_n - o_n) f_n''(net_n^i) 1_n \quad (18)$$

El factor η es el parámetro de velocidad de aprendizaje (que comúnmente se encuentra entre 0.25 y 0.75) y es seleccionado por el usuario. Valores muy grandes en este parámetro pueden llevar a la red a la inestabilidad y a un mal aprendizaje. Por otro lado, los valores muy pequeños pueden hacer que el aprendizaje sea muy lento. Algunas veces,

la velocidad de aprendizaje es variable para producir un aprendizaje más eficiente de la red; por ejemplo, permitiendo que el valor de η comience en un valor grande e ir decreméntándolo durante la sesión de aprendizaje.

Como ya se mencionó, la curva de umbral que se utiliza comúnmente es la sigmoide, aunque puede utilizarse también la función lineal. Sus ecuaciones son, respectivamente:

- $f_{j_n}(net_{j_n}^o) = \frac{1}{1 + e^{-net_{j_n}^o}}$
- $f_{j_n}(net_{j_n}^o) = net_{j_n}^o$

Para la salida con sigmoide, que es sobre la que se trabajará, se tiene:

$$w_{j_n}^o(t+1) = w_{j_n}^o(t) + \eta(y_{j_n} - o_{j_n})o_{j_n}(1 - o_{j_n})j_n \quad (19)$$

Entonces, se puede definir ya la ecuación de cambio de pesos como:

$$w_{j_n}^o(t+1) = w_{j_n}^o(t) + \eta \delta_{j_n}^o j_n \quad (20)$$

Si se está tratando de hacer un análogo entre la regla delta generalizada y el método de mínimos cuadrados, no se puede cambiar ninguno de los pesos hasta que todos los patrones de entrenamiento hayan sido presentados una vez a la red. Entonces, simplemente hay que acumular los cambios por cada patrón y después hacer el cambio. Este proceso se debe repetir hasta que el error sea aceptablemente bajo. El error se calcula como:

$$E = \sum_{p=1}^P E_p \quad (21)$$

donde P es el número de patrones en el conjunto de entrenamiento.

b) Cambio de pesos en la capa intermedia.

Se debe repetir para la capa escondida el mismo tipo de cálculo que se hizo para la capa de salida. Pero se presenta un problema, se sabe cuál es la salida de cada elemento de la capa intermedia, pero no se sabe cuál debería ser. Si se retoma la ec. (21) y se sustituyen en ella todas las variables se obtiene:

$$E_p = \frac{1}{2} \sum_i \left(y_{j_n} - f_{j_n} \left(\sum_i w_{ij_n}^o j_i + \theta_{j_n}^o \right) \right)^2$$

Se sabe que j_n depende de las ecs. (7) y (8). Explotando este hecho se calcula el gradiente de E_p con respecto a los pesos de la capa escondida:

$$\frac{\partial E_p}{\partial w_{\mu}^n} = -\sum_i (y_{\mu} - o_{\mu}) \frac{\partial o_{\mu}}{\partial (net_{\mu}^n)} \frac{\partial (net_{\mu}^n)}{\partial i_{\mu}} \frac{\partial i_{\mu}}{\partial (net_{\mu}^n)} \frac{\partial (net_{\mu}^n)}{\partial w_{\mu}^n} \quad (22)$$

Cada uno de los factores de la ec. (22) pueden ser calculados explícitamente de las ecuaciones anteriores. El resultado es:

$$\frac{\partial E_p}{\partial w_{\mu}^n} = -\sum_i (y_{\mu} - o_{\mu}) f_i'(net_{\mu}^o) w_{\mu}^o f_i'(net_{\mu}^n) x_{\mu} \quad (23)$$

El cambio de pesos de la capa escondida se hace en proporción al negativo de la ec. (24):

$$\Delta_{\mu} w_{\mu}^n = \eta f_i'(net_{\mu}^n) x_{\mu} \sum_i (y_{\mu} - o_{\mu}) f_i'(net_{\mu}^o) w_{\mu}^o \quad (24)$$

Utilizando la definición de δ_{μ}^o se puede escribir:

$$\Delta_{\mu} w_{\mu}^n = \eta f_i'(net_{\mu}^n) x_{\mu} \sum_i \delta_{\mu}^o w_{\mu}^o \quad (25)$$

Nótese que cada cambio de pesos en la capa Intermedia depende de todos los términos de error δ_{μ}^o de la capa de salida. Este resultado es justamente la justificación del nombre de retropropagación. Los errores ya conocidos de la capa de salida son propagados a la capa Intermedia para determinar el cambio apropiado de pesos en cada capa. Definiendo el término de error de la capa escondida se tiene:

$$\delta_{\mu}^n = f_i'(net_{\mu}^n) \sum_i \delta_{\mu}^o w_{\mu}^o \quad (26)$$

y se nota claramente que es análogo al de la capa de salida:

$$w_{\mu}^n(t+1) = w_{\mu}^n(t) + \eta \delta_{\mu}^n x_{\mu} \quad (27)$$

Finalmente, se puede observar que las ecs. (20) y (27) tienen la misma forma que la ec. (11), la regla delta.

Convergencia.

Cuando una red es entrenada continuamente, poco a poco se obtienen respuestas correctas; es importante, por tanto, tener una medida cuantitativa del aprendizaje. El error cuadrático medio (RMS) es usualmente calculado para mostrar el grado en el cual se ha llevado a cabo el aprendizaje en la red; esta medida nos dice hasta cuando la red da la respuesta correcta. Como la red va aprendiendo, el RMS disminuye, generalmente, un valor RMS menor a 0.1 indica que la red ha aprendido su conjunto de entrenamiento.

Note que cuando una respuesta es correcta o no, se tiene una decisión binaria (sí/no). El valor deseado de la red es un número real, y es el valor de salida. Entonces, la red no da una respuesta de sí o no que en realidad quiere decir correcto o incorrecto. La red se va acercando al valor

deseado incrementándose en cada paso. Es posible definir un punto de corte cuando la salida de la red coincide con los valores deseados y permite definir una respuesta correcta. La convergencia es un proceso que se lleva a cabo mientras el valor RMS para la red se acerca a 0.

La red de retropropagación generalmente converge a un valor pequeño de RMS cuando el conjunto de entrenamiento es claramente distinguible. Cuando se encuentra un mínimo local, la red debe ser hábil, mediante algunas técnicas, para evitar entrar a ese mínimo; por ejemplo, cambiando la velocidad de aprendizaje o el número de elementos en la capa escondida.

Agregando pequeños valores aleatorios a los pesos se le permite a la red escapar de un mínimo, una vez que lo ha encontrado, al mover su posición hacia un punto aleatorio a distancia del mínimo no deseado. Si la nueva posición está suficientemente lejos del valle del mínimo local, entonces puede proceder con la convergencia en una nueva dirección sin tener que volver a caer en el mismo mínimo local. La cantidad de ruido requerido depende de la superficie, que generalmente es desconocida para el investigador.

El proceso de convergencia de la propagación de error hacia atrás es prácticamente el mismo que en el método del gradiente descendente, que se deriva de la estadística tradicional.

Cuando la red ya ha sido entrenada, encuentra el modo de mapear un conjunto arbitrario de patrones de entrada a un conjunto arbitrario de patrones de salida. Este mapeo se encuentra sin conocer la función matemática que relaciona a los patrones de salida con los patrones de entrada. Las técnicas de análisis de ajuste de curvas tradicionales necesitan conocer a priori la forma de esa función matemática.

En la siguiente sección se muestran algunos ejemplos de aplicación de las RNA's en diferentes áreas de la medicina.

II.2.7. Aplicación de las RNA en la Medicina.

Las Redes Neuronales Artificiales han demostrado ser de gran utilidad en diferentes disciplinas científicas. El hecho de que la medicina se esté convirtiendo en un área en la que las RNA se pueden aplicar en forma directa se debe a las características que presenta. Por ejemplo, el tipo de datos que se obtienen son "blandos" y los criterios de evaluación son subjetivos; es decir, la información dada al especialista es evaluada con respecto a lo que el paciente *siente* o *creo sentir* y a *que tanto lo siente*.

Esto hace que las escalas de evaluación cotidiana tengan un rango amplio en la clasificación de enfermedades al hacer algún dx; por ejemplo, dos diferentes pacientes pueden tener un dolor de cabeza muy fuerte y con la misma intensidad, pero alguno de ellos puede quejarse más

que el otro. Esto provocaría que se les clasificara en diferentes niveles de dolor y por tanto, les recetarían diferentes tipos de medicamentos.

Otro factor que hay que tomar en cuenta y que tal vez sea el más importante es la multicausalidad. Todas las enfermedades involucran diferentes factores de riesgo y siempre hay que ir descartándolos uno a uno y quedarse con los más probables para ir también reduciendo el número de posibles etiologías. Esto involucra información especializada de múltiples áreas de la medicina, y con ello se obtiene una gran cantidad de datos que tiene que ser evaluada en conjunto por una sola persona para llevar a cabo la toma de decisiones y obtener un dx diferencial certero o lo más exacto posible.

Para aplicar las RNA a cierto problema se necesita tanto el conocimiento de un experto en el tema que se vaya a trabajar para la selección del conjunto de datos de entrenamiento y de prueba, como de un Ingeniero especialista en Neurocomputación para resolver la parte que corresponde al diseño y construcción de la RNA. Esto es, el Ingeniero debe seleccionar la técnica que se va a aplicar, la arquitectura que debe tener la red de acuerdo a la representación que se seleccione para los datos con los que se cuente, decidiendo si se debe o no aplicar un preproceso a la información. Una vez que se tienen codificados los datos de entrada a la red, esta se entrena y prueba con la participación de ambos profesionistas para que la red resultante pueda ser utilizada por los usuarios finales para quien se ha diseñado. Esto se logra diseñando una interfaz gráfica, quizás en ambiente Windows, por ejemplo, con la que se facilite el manejo. Todo el proceso anterior se describe en el diagrama de la Fig. 11. En general, se define la metodología que se debe seguir al aplicar las RNA a cualquier problema médico de dx y/o clasificación.

Esta herramienta, como se puede ver en los siguientes ejemplos, puede ser muy útil en el campo de la medicina (Kulkowski, 1980), ya que la herramienta pueden proveer asistencia en el dx, ya sea en la etapa de reconocimiento de patrones y/o sugiriendo un dx basado en las características del padecimiento.

II.2.8. Ejemplos de Aplicación

La clasificación de patrones es una tarea general que puede incluir desde un ejemplo muy sencillo hasta uno muy complicado. En esta sección se describirán diferentes ejemplos donde se aplican las RNA's en la medicina. Al final de la tesis se comparará el resultado que se obtenga de este trabajo con las características y conclusiones obtenidas en cada uno de los trabajos siguientes.

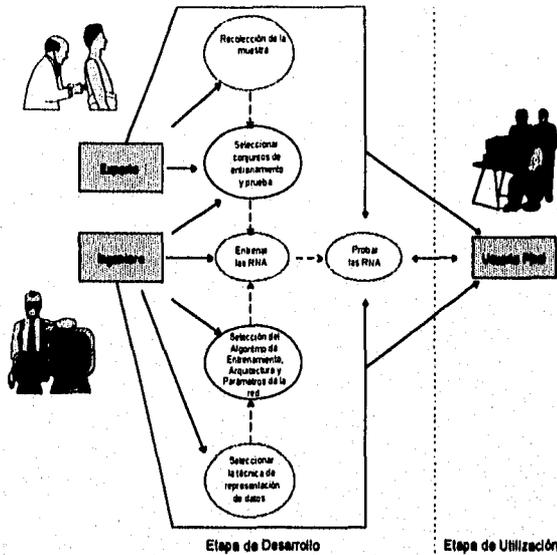


Fig. 11 Diagrama general de la aplicación de RNA's en el diagnóstico médico.

a) Radiología

En este trabajo, que consta de tres partes (Boone et al, 1990a, 1990b; Gross, 1990) se realiza el reconocimiento de patrones en placas radiográficas de tórax de niños recién nacidos. Este trabajo se realizó para ayudar al radiólogo en la dura y compleja tarea de reconocimiento que rutinariamente llevan a cabo y para esto se requiere de mucha trabajo y un enorme poder computacional.

El radiólogo utiliza el conjunto de Información de un paciente dado (hallazgos en las radiografías e Información clínica pertinente) para realizar un dx. Esta entrada/salida (hallazgos/dx) constituye una lección didáctica para las redes neuronales durante la fase de entrenamiento.

La red fue entrenada en una arquitectura de tres capas, formada por 21 nodos de entrada, 15 neuronas en la capa intermedia y 12 salidas para 12 diagnósticos diferentes. Cada salida daba un número entre 0 y 1. Las entradas estaban codificadas como ausente o presente (0/1). El algoritmo utilizado para el entrenamiento fue el de retropropagación.

El sistema se entrenó para seleccionar uno o varios diagnósticos de una lista de doce posibilidades. Se basa en 21 observaciones realizadas en cada una de las series de radiografías. Inicialmente, un radiólogo pediatra experimentado proporciona las observaciones radiográficas y lista diferentes diagnósticos para cada una de las 77 radiografías

(conjunto de entrenamiento) en la fase preliminar utilizadas para el entrenamiento.

Subsecuentemente, dos radiólogos pediatras (uno de los cuales proporcionó los datos para la fase de entrenamiento), en forma independiente, leen una serie de 103 radiografías de tórax de recién nacidos (conjunto de prueba) y recopilan una lista de anomalías y diferentes diagnósticos para cada placa.

Posteriormente, se le da a la red neuronal (previamente entrenada) la lista de hallazgos de cada uno de los radiólogos y así genera un reporte con los diagnósticos sugeridos para cada caso.

El acuerdo independiente entre la red y cada radiólogo fue mayor que entre los dos radiólogos. De sus resultados, los autores concluyen que un sistema que utilice una red neuronal tiene el potencial suficiente para diagnosticar enfermedades cardiopulmonares en recién nacidos (con una consistencia equivalente, a la de un radiólogo pediatra) siempre y cuando se le proporcionen los datos apropiados. Sin embargo, los autores piensan seguir trabajando en el proyecto para mejorar estos resultados.

b) Medicina Interna

Otro trabajo que involucra RNA en la medicina es el realizado por Riccardo Poli y colaboradores (1991). Ellos desarrollan un sistema llamado HYPERNET (Hypertension Neural Expert Therapist) para el dx y tratamiento de la hipertensión. El sistema está formado por una red neuronal entrenada con los datos clínicos de sujetos y con las series de 24 horas de presión sistólica y diastólica. La salida del sistema incluye cuatro arreglos de 24 nodos cuyos valores especifican la dosis que se le administrará al paciente de cada una de las drogas antihipertensas más comunes.

La presencia y grado de hipertensión (dx) puede ser inferido de la dosis de droga: no se requiere tratamiento para sujetos normales. Los datos utilizados para entrenar a la red son las series de presión sistólica-diastólica de 300 sujetos sanos y 85 sujetos hipertensos. Los datos incluyen un reporte clínico con toda la información requerida por Hypernet. En cuanto a los medicamentos, sólo se tomaron en cuenta a los más comunes. Los datos clínicos del expediente son tomados en cuenta para seleccionar el medicamento adecuado. En el artículo no se menciona el tipo de codificación utilizada para el conjunto de datos.

La arquitectura de la red está dividida en tres módulos:

1. El módulo de generación de referencias (RGM). Está formado por dos redes idénticas con dos entradas, cuatro neuronas intermedias y 24 salidas en cada una. Estas redes comparan una serie de tiempo de 24 horas de presión arterial con la serie típica para sujetos normales

del mismo sexo y edad. El propósito del módulo es encontrar casos similares.

2. El módulo de compatibilidad de medicamentos (DCM); es una red de dos capas con 17 entradas y cuatro unidades de salida. Su función es analizar las características más importantes del reporte clínico del paciente (entradas) y determinar el grado de compatibilidad con cada medicamento considerado (salidas).
3. El módulo para seleccionar la terapia adecuada (TSM). Es una red de seis capas cuyas entradas son: el grado de compatibilidad; la diferencia entre el cambio en la presión sistólica y el cambio en la presión diastólica obtenidas en el monitoreo al paciente; los datos de cinco anestésicos que pueden influenciar el dx. Se tienen 96 unidades de salida (24 valores x 4 medicamentos) y representan la salida final del sistema. La primera capa contiene 12 unidades, se utiliza para identificar el aumento de presión, en el tiempo y con respecto a la presión. La segunda capa está formada por 12 nodos conectados con la capa anterior y 5 unidades que representan el anestésico que pudiera influenciar el dx. La tercer capa consta de 12 unidades cuyas entradas son las salidas activadas en la capa anterior y los cuatro medicamentos compatibles estimados por el DCM. Esta capa, junto con otras 12 unidades y la salida del TSM determinan la terapia del paciente, si es necesaria. Del artículo no se puede obtener la información de cual era el número de nodos en las otras tres capas y para que se utilizaban.

El algoritmo utilizado para el entrenamiento de todas las redes fue el de retropropagación.

El sistema fue evaluado con 35 sujetos, 10 no tenían signos de presión alta y 25 tenían sospecha de hipertensión. El dx del sistema lo consideraron correcto si coincide con el dx del médico. Los resultados que obtuvieron son: Hypernet diagnosticó correctamente 33 de 35 pacientes, que representan el 94%. De estos, solo a 11 les prescribió tratamiento. El especialista evaluó el dx y lo juzgó correcto en siete casos, aceptable en dos y erróneo en dos, por tanto, el 82% fue correcto o aceptable.

c) Electrofisiología

Otro trabajo es el realizado por Jandó y Seyal (1992), quienes entrenan una RNA para reconocer espigas de alto voltaje en registros de pacientes epilépticos.

El sistema está formado por dos RNA's de tres capas, entrenadas con el algoritmo de retropropagación. El número de neuronas de entrada en las dos redes está determinado por el número de canales de

EEG que se utilicen. Cada canal corresponde a una entrada (8 en este caso). La capa intermedia está formada por 8 nodos y la salida es solo un nodo.

Los datos utilizados para el entrenamiento de la red se obtienen de la siguiente manera: se digitaliza y graba la señal de EEG, se limpia con un filtro paso altas de 2 Hz. Después se toma el promedio de la derivada del voltaje con respecto al tiempo, y esto se hace para cada canal. El promedio es la suma de los cambios de voltaje entre cada muestra (pre y postdisparo) en el intervalo dividido por el número de muestras en el intervalo representando el intervalo total. La selección de los patrones de entrenamiento fue hecha por el usuario viendo una pantalla gráfica del EEG. El conjunto de entrenamiento está formado por los datos de 5 pacientes.

La red B es entrenada con el conjunto de datos que se obtienen del promedio de voltajes pre-disparo, es decir, el promedio de voltaje que se obtiene antes de que se presente un disparo anormal en el registro de EEG. La red A está entrenada con los datos que se obtiene después del disparo (promedio).

Para obtener la capacidad de reconocimiento de la red, realizan un análisis de selectividad y sensibilidad. La selectividad fue calculada en términos del número de reconocimientos falsos-positivos definidos por inspección visual. Después se obtiene el número total de reconocimientos de falsos-positivos (porcentaje) en función del tiempo total de grabación (promedio de falsos-positivos/hora).

La sensibilidad fue calculada con base en el número de veces que un disparo anormal seguro permanece sin reconocerse (error de falsos-negativos). Se calcula entonces el porcentaje de errores falsos-negativos como función de disparos anormales definidos por inspección visual y el promedio de errores falsos-negativos por hora. Obviamente el resultado correcto depende por completo de la certidumbre del Interpretador.

Sus resultados muestran lo siguiente: dos de los cinco casos tiene focos epilépticos independientes, dos pacientes tienen focos unilaterales sencillos y un paciente tiene una anomalía caracterizada por un disparo y onda compleja de 3 seg.

Se probaron con el conjunto de entrenamiento 7 conjuntos de redes ya entrenadas (el artículo no menciona si las siete se entrenaron con el mismo conjunto de datos y cual es la diferencia entre ellas, aparentemente es el umbral de reconocimiento) y los dos pacientes con focos independientes fueron tratados como casos especiales para propósitos de análisis.

El análisis de sensibilidad arroja un resultado de 94.2%. La selectividad de las redes depende de la selección del intérprete. Los resultados fueron: cuando el umbral de reconocimiento fue alto (salida ≥ 0.9), el promedio de falsos-positivos fue bajo ($3.7\% \pm 6.8$). El reconocimiento de falsos-positivos se incrementó ($20.9\% \pm 22.6$) cuando

se fijó un umbral bajo (salida ≥ 0.825). En el peor de los casos representa alrededor del 50% del total de reconocimientos de la red comparado con el 3.6% en el mejor de los casos.

Como sus resultados no los dejaron muy convencidos desarrollaron una nueva red para tratar de disminuir estos errores. Un año más tarde, el mismo grupo de investigadores (Jandó et al, 1993) publican otro artículo donde resuelven el problema utilizando el EEG de ratas epilépticas, pero utilizan como entrada el registro de un solo canal, que es la derivada de los picos de todos los canales, con lo que logran un mejor porcentaje de reconocimiento.

d) Cardiología

Las RNA también se han utilizado para el dx de infarto al miocardio en pacientes que ingresan al área de urgencias con historia de dolor en tórax (Baxt, 1991).

El sistema se forma por una red neuronal con cuatro capas. La primera contiene 20 unidades, las dos capas intermedias tienen 10 y la salida consta de un solo nodo. El algoritmo de entrenamiento fue el de retropropagación.

Se seleccionaron en forma retrospectiva los datos clínicos de 351 casos de pacientes con alta probabilidad de tener infarto al miocardio, con esta información se formó el conjunto de entrenamiento. Los datos están codificados con 1/0 (presencia/ausencia) con excepción de la edad.

Se probó en forma prospectiva con 331 pacientes adultos (conjunto de prueba). El resultado de la red fue comparado con el del especialista.

Para el análisis estadístico de los resultados de la red se realizaron intervalos de confianza (CI) y una prueba de J-cuadrada.

Los resultados son: el médico diagnosticó correctamente 28 pacientes con infarto al miocardio para una sensibilidad de 77% (95% CI, 77% a 82.9%) y diagnosticó correctamente 250 pacientes sin infarto para una especificidad de 84.7% (CI, 84% a 86.4%). La red diagnosticó incorrectamente un paciente con infarto al miocardio para una sensibilidad del 92.7% (CI, 92.2% a 97.5%) y también clasificó mal 11 pacientes sin infarto para una especificidad de 96.2% (CI, 96.2% a 96.4%). La diferencia en la sensibilidad es estadísticamente significativa ($p=0.003$). La especificidad también fue estadísticamente significativa ($p<0.001$)

El trabajo se concluye planeando probar con más pacientes a la red para entonces considerarla lista para su uso cotidiano. Sin embargo, con los resultados actuales postulan que su sistema puede ser de gran

ayuda para identificar infarto al miocardio, en pacientes adultos que se presentan a una sala de urgencias.

Así como estos ejemplos existen muchos otros (Dhawan et al., 1993; Edenbrandt et al., 1992; Moallemi, 1991; Mamelak et al., 1991; Jandó et al., 1992, 1993; Hailgren et al., 1992; Weinstein et al., 1992; etc.) que utilizan redes neuronales en el dx médico. Con todas estas aplicaciones como base podemos decir que el área de las Redes Neuronales Artificiales se está convirtiendo en una herramienta muy poderosa en todas las áreas de la medicina.

En la siguiente sección se explica el trabajo realizado en esta tesis siguiendo el mismo formato de las aplicaciones anteriores, incluso detallando más algunas partes, como los parámetros de entrenamiento, el tiempo que tardó en entrenarse la red y otros. Posteriormente, como ya se mencionó se hará referencia a estos trabajos para hacer una comparación entre todos.

III. Diseño, Construcción y Entrenamiento del Sistema basado en RNA

III.1 Metodología

III.2 Recolección de la Muestra

*III.3 Representación de la
Información*

*III.4 Selección del Algoritmo de
Entrenamiento, Arquitectura y
Parámetros de la Red*

III.5 Entrenamiento

*III.6 Resultados del
Entrenamiento*

*III.7 Pruebas a la RNA
Entrenada*

III. DISEÑO, CONSTRUCCION Y ENTRENAMIENTO DEL SISTEMA BASADO EN RNA

En este capítulo se describe la metodología que se siguió a lo largo del proyecto, se desglosa cada parte de la metodología y se explica en forma detallada cada una de las partes, también se muestra el resultado de la red entrenado con esta metodología y algunos comentarios sobre por qué se realizará otro entrenamiento.

III.1. Metodología

Con base en el diagrama de la fig. 11 del capítulo anterior se puede generar un nuevo esquema para mostrar la metodología seguida en este trabajo. En éste, se sustituyen los datos generales por los utilizados en esta tesis (ver Fig. 1).

La metodología aplicada a este problema quedó de la siguiente manera:

1. Recolección de la muestra de pacientes con DV y DTA
2. Seleccionar los datos del grupo de pacientes del INNN a incluir en el proyecto, ya sea para formar el conjunto de entrenamiento o el de prueba.
3. Definir la representación de la información de cada paciente para formar una base de datos que contenga todos los datos necesarios ordenados por grupos.
4. Seleccionar el algoritmo de entrenamiento, la arquitectura y los parámetros de la red.
5. Llevar a cabo el proceso de entrenamiento de la RNA hasta obtener un error en el conjunto de prueba menor o igual a 2×10^{-1} .

6. Probar la RNA ya entrenada para analizar casos nuevos no contemplados en el conjunto de entrenamiento.

III.1.1. Recolección de la Muestra

Con ayuda del médico neurólogo se reunieron 58 casos de sujetos con dx clínico y por Imagen de DV y DTA que fueron atendidos en las clínicas de Enfermedad Vascul ar Cerebral y Cognición del Instituto Nacional de Neurología y Neurocirugía "Manuel Velasco Suárez".

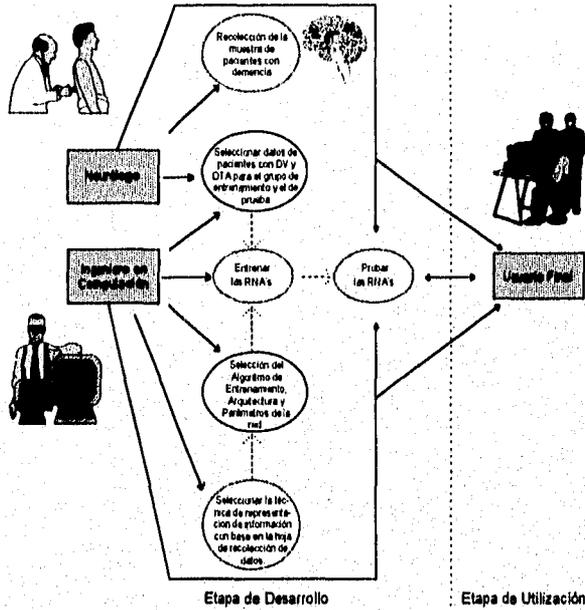


Fig. 1 Metodología empleada.

Del total de la muestra se formaron tres grupos que se integraron de la siguiente manera:

- Grupo I:** 19 sujetos con diagnóstico de DV.
- Grupo II:** 16 sujetos con diagnóstico de DTA.
- Grupo III:** 23 sujetos con diagnóstico de demencia, sin tener especificado el tipo de demencia.

Las características clínicas de la muestra se obtuvieron revisando el expediente de cada sujeto, se seleccionó la información necesaria y se integró una BD que incluye 46 datos diferentes por sujeto.

A todos estos pacientes se les realizó una evaluación neuropsicológica utilizando las pruebas explicadas en el capítulo anterior: Mini Mental State Examination, Escala de Actividades Cotidianas y la Escala de Depresión Geriátrica; también se sometió a evaluación a cada uno de los sujetos de la muestra utilizando las escalas de Hachinski y Loeb. Todos los sujetos de la muestra reúnen criterios diagnósticos para demencia de acuerdo a los parámetros de la Asociación Psiquiátrica Americana (APA) y del Manual Estadístico y de Diagnóstico para Trastornos Mentales (DSMIII). Los sujetos con DV y aquéllos con DTA reúnen los criterios diagnósticos de los Institutos Nacionales de Enfermedades Neurológicas de E.U. (NINDS).

También, a todos los sujetos de la muestra se les realizó estudios de neuroimagen: TAC y/o IRM, así como EEG, P300, Radiografía de tórax (Rx Tx) y laboratorio básico.

Para la recolección de la muestra fue necesario diseñar el formato de una hoja de datos con la que el médico pudiera recabar la misma información de todos los pacientes. Ésta se realizó de común acuerdo entre médico-Ingeniero; esta hoja se muestra en la Fig. 2. En la hoja se muestran sólo 33 variables, pero hay que agregarle información del sexo del paciente y como en algunos de los incisos marcados se pueden presentar varias características al mismo tiempo, es necesario agregar algunas columnas más a la BD, pero toda la información está recabada en esa hoja.

Lo anterior forma la primera parte del diagrama de la Fig. 1 es decir, la interacción del neurólogo con el Ingeniero para seleccionar los conjuntos de entrenamiento y prueba de la RNA.

III.1.2. Representación de la información

Para este punto fue necesario estudiar la información que se obtuvo con las hojas de recolección de datos y codificarla en la BD. Lo primero, en colaboración con el neurólogo, fue la división en los 6 grupos ya mencionados:

1. Demográficos
2. Antecedentes
3. Síntomas
4. Escalas Neurológicas y Neuropsicológicas
5. Electrofisiología
6. Neuroimagen y otros estudios (laboratorio, rayos X, doppler de vasos de cuello).

Posteriormente y con base en la hoja de recolección de datos se realizó una codificación en una hoja de cálculo del paquete comercial Excel versión 5.0., para facilitar así el acceso de datos a la RNA.

La división de grupos se realizó para observar el comportamiento de cada una de las variables de los grupos en la RNA.

Esta BD incluye a los grupos I y II, es decir, se guardó junta la información del grupo de pacientes con DV (19) y con DTA (16), para obtener una base de 35 sujetas con dx de demencia y formar así el conjunto de entrenamiento de la RNA. Es importante resaltar que en esta BD se incluyó el dx canacida de cada uno de las pacientes para que la red se conformara con todas las características de la muestra incluyendo el dx.

En otra base de datos se almacenó la información referente al grupo III, es decir, la información de los pacientes del grupo de prueba (23), con la misma estructura y cantidad de información que la primera, pero sin incluir el dx correspondiente para que la red, ya entrenada, genere su clasificación.

III.1.3. Selección del Algoritmo de Entrenamiento, Arquitectura y Parámetros de la Red.

Actualmente existen muchos tipos de arquitecturas y algoritmos de entrenamiento. Para saber que tipo de red utilizar se debe poner atención en las datos con las que se cuenta. En primer lugar, las datos no se pueden mapear con una función lineal, al contrario, se tiene un conjunto de datos que probablemente no es linealmente separable¹ y que, para realizar un mapeo se necesitaría una función extremadamente no lineal entre variables continuas. Segunda, el conjunto de entrenamiento está formada por parejas de entrada-salida (características-dx) y tercera, las datos pueden representarse de manera espacial, es decir, no cambian con el tiempo.

Si bien, con base en estas características se pueden seleccionar varias modelos de redes neurales artificiales, algunas datos del problema nos dan una idea más clara de cómo debe ser:

1. el que las datos sean estáticos requieren de una red "feedforward";
2. el número de nodos de entrada (46 características de cada paciente) define una capa de entrada;
3. el dx diferencial entre DV y DTA se puede obtener con un sola nodo (esta es, en la capa de salida sola se tendrá una neurona);
4. el tipo de mapeo necesaria (no lineal, entre variables continuas).

¹La definición de conjunto linealmente separable se encuentra en el capítulo I. Introducción.

Tomando como base estas características y los ejemplos mencionadas en el capítulo anterior, se definió que se utilizará una arquitectura de tres capas y el algoritmo empleado para el entrenamiento de la red será el de retropropagación, debido a su gran capacidad para realizar clasificación de patrones complejos.

III.1.4. Entrenamiento de la RNA

Para el entrenamiento de la RNA se buscó un simulador comercial, que mínimamente debería:

1. Importar información de las bases de datos y hojas de cálculo más comerciales;
2. tener la capacidad de extraer un pequeño conjunto de datos del conjunto de entrenamiento para probarlo durante el aprendizaje;
3. permitir el suficiente control sobre los parámetros de la red, como por ejemplo, fijar el error mínimo para el conjunto de entrenamiento o el de prueba, el número de iteraciones máximo cuando la red permanece en un mínimo local;
4. contener diferentes funciones de activación por capa para seleccionar la deseada así como sugerir un número de nodos en la capa intermedia, pero permitiendo cambiarlo en el caso deseado;
5. obtener gráficas de error del conjunto de entrenamiento y del de prueba durante el entrenamiento;

Se buscó alguno que las tuviera y el indicado fue el NeuroShell 2 versión 1.5, para windows. Este permite, además de las características mencionadas, probar cualquier conjunto de entrenamiento, generar un archivo con los resultados de la red que puede ser analizado tanto en el mismo simulador como en Excel, o en cualquier otra BD, permite generar archivos DDL de redes ya entrenadas para correr dentro de otras aplicaciones tipo windows. Cuenta con dos módulos:

1. uno es el módulo para principiantes, donde el mismo simulador hace el ajuste de todos los parámetros, dependiendo del tipo y la cantidad de datos con los que se cuente y el usuario solo tiene que decir que emplee el entrenamiento;
2. el otro es con el módulo de avanzados, donde el usuario tiene más control sobre los parámetros y los ajusta a como crea conveniente, dependiendo de la experiencia con la que se cuente.

Para este trabajo se utilizó el segundo módulo, pues era importante tener cierto control en cuanto a parámetros para detener el entrenamiento y en el manejo del porcentaje de error permitido a cada una de las redes, la

forma en que se deben presentar las parejas de entrenamiento y si se quiere presentar un conjunto de prueba al mismo tiempo que se está llevando a cabo el entrenamiento.

Como se mencionó, la BD formada se utilizó como conjunto de entrenamiento de la RNA. La red fue entrenada con el algoritmo de retropropagación en una arquitectura de tres capas y con ellas, se realizó el dx asistido por computadora. Es importante mencionar que en todos los casos del conjunto de entrenamiento de las RNA se incluyó el dx conocido de DTA y DV.

Lo primero que se hizo fue alimentar la BD a la red, especificando cuáles eran las entradas y cuál la salida. Se define la arquitectura y los parámetros de la red. Incluyendo la función de activación en cada una de las capas. Para esta red se definieron como funciones de activación, en la entrada, una función lineal ("linear"), en la capa intermedia una función logística ("logistic") y en la salida una tangente hiperbólica ("tanh"). La decisión para seleccionar estas funciones se tomó con base en los resultados de un estudio previo². Estas funciones se muestran en la Fig. 3.

En esta parte también se deben establecer tres parámetros más como son la velocidad de aprendizaje, que se ajustó en $\eta = 0.1$, el término de momentum en $m = 0.1$ y los pesos en todas las capas en $w = 0.3$.

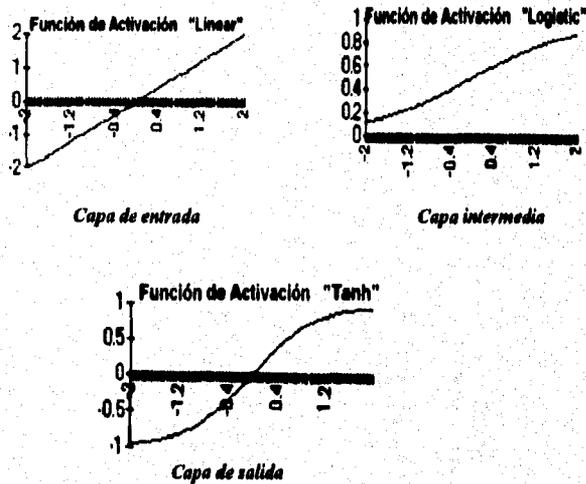


Fig. 3 Funciones de activación utilizadas en la RNA entrenada

² En esta tesis no se muestran los resultados ni la forma en que se realizaron las pruebas para decidir el tipo de función a utilizar por capa. Sin embargo, a grosso modo, la prueba consistió en entrenar diferentes redes con diferentes funciones de activación, utilizando parte de los datos de este trabajo.

La presentación de los pares de entradas a la hora del entrenamiento se realizó en forma aleatoria. El entrenamiento se debía detener cuando el promedio de error fuera menor o igual a 2×10^{-7} o el número de presentaciones del conjunto de entrenamiento fuera igual a 100000. Finalmente, se inicia el entrenamiento de la red. Una pantalla del módulo de entrenamiento se muestra en la Fig. 4.

Como ya se mencionó, el vector de entrada corresponde a las características clínicas de los pacientes y la salida es el dx para cada caso. Estos son los patrones que se presentan repetidamente al modelo neuronal con el fin de modificar los pesos sinápticos para reducir los errores en las clasificaciones ofrecidas por la red.

Con los parámetros ya seleccionados se procedió al entrenamiento de la red y los resultados se encuentran en la siguiente sección.

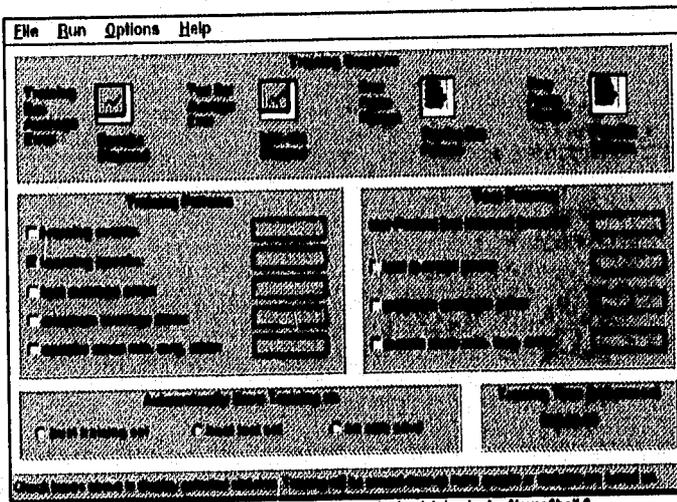


Fig. 4 Pantalla del Módulo de Entrenamiento del simulador NeuroShell 2

III.1.5. Resultados del Entrenamiento

El modelo neuronal que se entrenó (para llevar a cabo la asociación de patrones) fue una red estándar de retropropagación con tres capas. En la Fig. 5 se muestra, a modo de ejemplo, el tipo de arquitectura utilizada para el entrenamiento, el número de nodos en la capa de entrada fue de 46, en la capa intermedia 29 y uno en la salida.

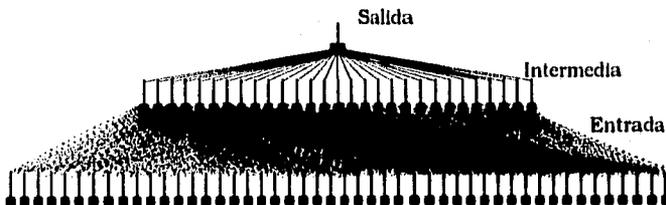


Fig. 5 Arquitectura de la RNA empleada

La red se entrenó durante casi 65 hrs para llegar a un promedio de error de 0.0000002. En la Fig. 6 se muestra la pantalla final del entrenamiento de esta red. En ella se pueden observar como quedaron los demás parámetros. La Gráfica 1 y la Gráfica 2 muestran en forma de barras el error de la red tanto en el conjunto de entrenamiento como en el conjunto de prueba seleccionados para comparar al mismo tiempo que se lleva a cabo el aprendizaje. Se debe aclarar **que no es el mismo empleado en la etapa de pruebas de la red**, es decir, el simulador permite seleccionar un pequeño conjunto del conjunto de entrenamiento para utilizarlo durante la fase de aprendizaje e ir probando este grupo conforme se va encontrando un conjunto de pesos con el error mínimo.

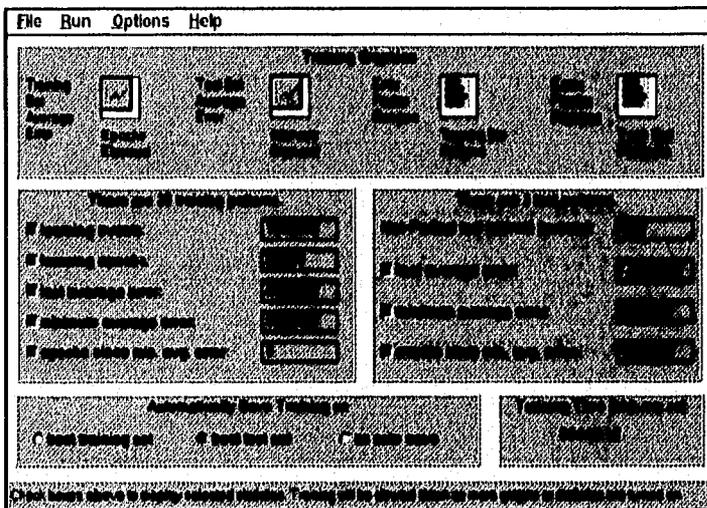
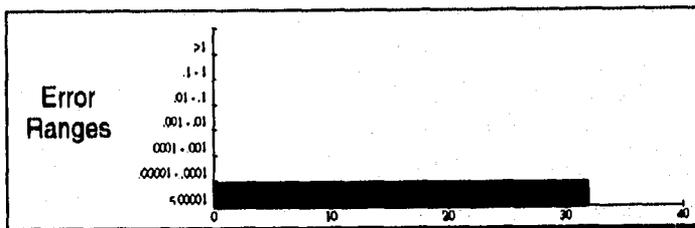


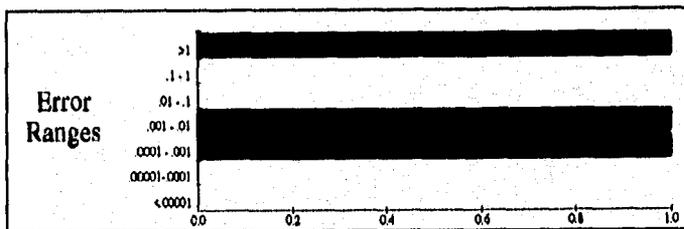
Fig. 6 Pantalla final del entrenamiento de la RNA

La Gráfica 3 muestra cuales son las 14 variables más significativas para la red, de todo el conjunto de entrenamiento.

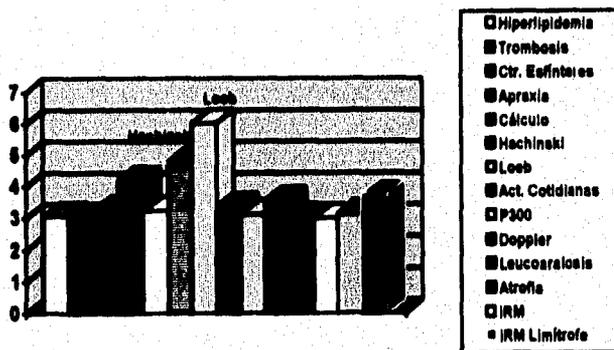
En esta gráfica se puede observar que las variables que más están contribuyendo al dx diferencial de DV y DTA son la escala de Hachinski y la de Loeb. Ambas pertenecen al grupo de escalas. Las demás variables, aunque aportan, no es mucho lo que ayudan a dar un dx correcto.



Gráfica 1 Error en el conjunto de entrenamiento.



Gráfica 2 Errores en el conjunto de prueba durante el entrenamiento



Gráfica 3 Factores de contribución dados por la red.

III.1.6. Pruebas a la RNA entrenada

Una vez que el modelo neuronal se entrenó, se obtuvo la clasificación correcta de todos los casos utilizados en el entrenamiento. Se realizó una prueba con el conjunto seleccionado para ello (Grupo III), que incluía datos de 23 sujetos con dx de demencia, pero que no estaban incluidos en el conjunto de entrenamiento. Esto se realizó con el fin de verificar la forma en que la red clasificaba.

El resultado fue el siguiente: la red entrenada con todos los datos que formaban la BD clasificó correctamente el diagnóstico en un 82.6% de los casos; es decir, acertó en 19 de los 23 casos presentados. Este resultado fue analizado y confirmado por el médico neurólogo.

Comparando este resultado con los ejemplos descritos en el capítulo II se puede decir que el resultado obtenido en este trabajo es mejor o muy similar a los que se presentan, dependiendo del trabajo al que se haga referencia. Es muy superior comparado con el trabajo de Jandó y Seyal, quienes entrenaron una red para el reconocimiento de espigas de alto voltaje en señales de EEG. Es muy similar al realizado por Ricardo Poll y cols. para el dx y tratamiento de hipertensión, ya que registran un 82% de eficiencia. Sin embargo, el trabajo de Baxt obtiene mejores resultados que éste. El trabajo realizado con placas radiográficas no se puede comparar, ya que los autores no especifican cual es el resultado de la red.

Con base en lo anterior, se puede decir que se obtuvo un excelente resultado, aunque realmente no cumple con las expectativas que teníamos al inicio de este trabajo, es decir, realizar un dx certero al 100% utilizando este tipo de tecnología. De esta forma, el médico aún tendría que utilizar en gran parte su experiencia para determinar si el resultado sugerido por la RNA es el correcto.

Se puede observar que el número de variables con el que actualmente se realiza el dx es muy grande; aparentemente, mientras la tecnología avanza y van surgiendo nuevos métodos diagnósticos, el médico los ha ido integrando para tratar de obtener un dx más certero. Sin embargo, conforme avanza la tecnología, el costo de los nuevos estudios va en aumento, no permitiendo que la mayor parte de la gente tenga acceso a ellos.

Tomando como referencia la Gráfica 3, donde se muestran las variables más significativas obtenidas por la red, se determinó que existen algunas variables que son las que más contribuyen al dx. Probablemente, algunas de ellas no ayudan a obtener el dx correcto, sino que al contrario, hacen más difuso el caso.

Por lo anterior, se decidió hacer un estudio de todas las variables de este problema para saber si entrenando diferentes RNA's con distintos conjuntos de entrenamiento (i.e. tomando combinaciones de los seis grupos señalados) se alcanza un porcentaje más alto de reconocimiento, con lo que el médico estaría más seguro del resultado del sistema y se cumplirían nuestros objetivos. Por tanto, el siguiente paso en el proyecto es analizar los resultados de las diferentes RNA's para ver si se obtiene un mejor porcentaje.

IV. Estudio de las Pruebas Relacionadas con el Diagnóstico Diferencial utilizando RNA

IV.1 Planteamiento

IV.2 Resultados

IV. ESTUDIO DE LAS PRUEBAS RELACIONADAS CON EL DIAGNÓSTICO DIFERENCIAL UTILIZANDO RNA

En este capítulo se describe, con base en el resultado de la sección anterior, el nuevo planteamiento al problema para lograr un porcentaje mejor de reconocimiento. Se describen los resultados de este estudio de pruebas, así como un pequeño estudio costo-beneficio para saber cual es el grupo que tiene menor costo con respecto a los cuatro puntos señalados anteriormente, esto es, riesgo para el paciente, costo en tiempo y dinero de los estudios e infraestructura necesaria para llevar a cabo todo el gabinete de pruebas.

IV.1 Planteamiento

El objetivo de la serie de simulaciones descritas en esta parte fue explorar la confiabilidad de la información de los diferentes grupos de datos para determinar que tanto contribuye cada una de ellos en el diagnóstico diferencial de la DV y la DTA. Este estudio se realizó en forma exhaustiva para determinar con exactitud cuáles son los parámetros que más contribuyen al diagnóstico y ver si con ellos se obtiene una eficiencia mejor que la encontrada en el capítulo anterior.

Como ya se tiene la información dividida en grupos, la primer aproximación fue crear una BD nueva para cada conjunto; es decir, se formaron 6 bases de datos con la información separada de cada paciente:

1. Demográficos
2. Antecedentes
3. Síntomas

4. Escalas Neurológicas y Neuropsicológicas
5. Electrofisiología
6. Neuroimagen y otros estudios (laboratorio, rayos X, doppler de vasos de cuello).

Estas BD fueron utilizadas para entrenar diferentes RNA's. Como son seis grupos y no se sabe exactamente cuáles son las variables que más contribuyen al diagnóstico diferencial, excepto por los resultados obtenidos en el capítulo anterior, se organizaron tantas bases de datos como combinaciones hay con los 6 grupos.

Para encontrar el número de combinaciones se tiene la siguiente fórmula:

$$C_r^n = \frac{n!}{r!(n-r)!} \quad (9)$$

y como necesitamos las combinaciones de 6 en 6, de 5 en 5, ..., hasta de uno en uno, se tiene que el número total de combinaciones es:

$$C_r^n = C_6^6 + C_5^6 + C_4^6 + C_3^6 + C_2^6 + C_1^6 \quad (10)$$

donde el número total de combinaciones es la suma de los resultados en (10), esto es:

$$C_r^n = 63 \quad (11)$$

Por tanto, se tienen 63 BD diferentes con las que se entrenaron 63 RNA's diferentes; con esto se completa el análisis exhaustivo a los parámetros utilizados en el dx. diferencial; se probaron todas las redes con el mismo conjunto de prueba con que se probó la red obtenida en el capítulo anterior (combinación de 6 en 6).

En cuanto a los parámetros de entrenamiento de estas nuevas redes, se utilizaron los mismos de las redes anteriores:

$$\eta=0.1 \quad m=0.1 \quad w=0.3.$$

Las funciones de transferencia también fueron las mismas, es decir, a la entrada una función lineal ("linear"), en la capa intermedia una función logística ("logistic") y en la salida una tangente hiperbólica ("tanh"). El entrenamiento también se detiene cuando el promedio de error es menor o igual a 2×10^{-7} o el número de presentaciones del conjunto de entrenamiento fuera igual a 100000.

El propósito de entrenar 63 redes con diferentes conjuntos de entrenamiento es obtener también información sobre cuál es el conjunto de

costo mínimo, tomando en cuenta los factores mencionados en el capítulo 1, que son:

1. Riesgo para el paciente
2. Infraestructura requerida
3. Tiempo de realización de pruebas
4. Factor económico (costo)

Al final de la siguiente sección se analizarán los posibles conjuntos de datos que mejor cumplen con las características requeridas en los objetivos.

IV.2 Resultados

Todas las RNA's fueron entrenadas con los mismos parámetros y utilizando las mismas funciones de umbral, así como con el mismo conjunto de entrenamiento; esto es, se utilizaron a los mismos 35 casos de pacientes con demencia solo que dependiendo de la BD seleccionada fueron los datos que se incluyeron en el conjunto de entrenamiento.

Se entrenaron 63 RNA's (Incluyendo a la red entrenada en la etapa anterior), el tiempo de entrenamiento estuvo entre los 11 min. y las 65 hrs. donde la red que menos tiempo utilizó para entrenarse es la que contiene la información de los datos demográficos y escalas (la red que mejor funciona) y la que más tardó fue la que contiene a los seis grupos. A todas las redes se les probó con el mismo conjunto de entrenamiento, esto es, con el grupo III. Los resultados emitidos por la red fueron evaluados por el neurólogo, quien determinó si el diagnóstico era correcto o no.

En las siguientes tablas (Tabla 1, Tabla 2) se muestran los porcentajes y el número de casos clasificados correctamente por las 63 redes. La información de las tablas está organizada con respecto al número de datos que se incluyen en su conjunto de entrenamiento.

De las tablas 1 y 2 se pueden observar varias cosas:

1. Existen 11 redes que obtienen mejores resultados que la red entrenada con los 6 grupos.
2. La red que contiene datos demográficos y datos de escalas es la que obtiene un mayor número de casos bien clasificados (22 de los 23 casos).
3. Resultaron 5 redes que solo clasifican mal 2 casos y
4. 5 redes que fallan en 3 casos.

Grupo con que se entrenó	% casos bien clasificados	# de casos reconocidos
Uno en uno		
Demográficos	13.0	3
Antecedentes	52.2	12
Síntomas	39.1	9
Escalas	81.3	21
Neuroimagen	56.5	13
EEG	52.2	12
Dos en dos		
Demográficos, antecedentes	65.2	15
Demográficos, síntomas	52.2	12
Demográficos, neuroimagen	60.8	14
Demográficos, eeg	52.2	12
Antecedentes, síntomas	56.5	13
Antecedentes, escalas	81.3	21
Antecedentes, neuroimagen	82.6	19
Antecedentes, eeg	56.5	13
Síntomas, escalas	82.6	19
Síntomas, neuroimagen	60.8	14
Síntomas, eeg	47.8	11
Escalas, neuroimagen	78.3	18
Escalas, eeg	81.3	21
Neuroimagen, eeg	78.3	18
Tres en tres		
Demográficos, antecedentes, síntomas	52.2	12
Demográficos, antecedentes, escalas	78.3	18
Demográficos, antecedentes, neuroimagen	78.3	18
Demográficos, antecedentes, eeg	60.8	14
Demográficos, síntomas, escalas	87.0	20
Demográficos, síntomas, neuroimagen	60.8	14
Demográficos, síntomas, eeg	47.8	11
Demográficos, escalas, neuroimagen	82.6	19
Demográficos, escalas, eeg	81.3	21
Demográficos, neuroimagen, eeg	60.8	14
Antecedentes, síntomas, escalas	78.3	18
Antecedentes, síntomas, neuroimagen	60.8	14
Antecedentes, síntomas, eeg	43.5	10
Antecedentes, escalas, neuroimagen	82.6	19
Antecedentes, escalas, eeg	87.0	20
Antecedentes, neuroimagen, eeg	85.2	15
Síntomas, escalas, neuroimagen	82.6	19
Síntomas, escalas, eeg	78.3	18
Síntomas, neuroimagen, eeg	56.5	13
Escalas, neuroimagen, eeg	73.9	17

Tabla 1 Resultados obtenidos con los primeros tres grupos. Porcentaje de casos correctos y número de casos bien clasificados.

Grupo con que se entrenó	% casos bien clasificados	# de casos bien clasificados
Cuatro en cuatro		
Demográficos, antecedentes, síntomas, escalas	47.8	11
Demográficos, antecedentes, síntomas, neuroimagen	69.9	16
Demográficos, antecedentes, síntomas, eeg	73.9	17
Demográficos, antecedentes, escalas, neuroimagen	87.0	20
Demográficos, antecedentes, escalas, eeg	78.3	18
Demográficos, antecedentes, neuroimagen, eeg	73.9	17
Demográficos, síntomas, escalas, neuroimagen	78.3	18
Demográficos, síntomas, escalas, eeg	87.3	21
Demográficos, síntomas, neuroimagen, eeg	60.9	14
Demográficos, escalas, neuroimagen, eeg	39.1	9
Antecedentes, síntomas, escalas, neuroimagen	87.0	20
Antecedentes, síntomas, escalas, eeg	82.6	11
Antecedentes, síntomas, neuroimagen, eeg	69.6	16
Antecedentes, escalas, neuroimagen, eeg	82.6	19
Síntomas, escalas, neuroimagen, eeg	82.6	19
Cinco en cinco		
Demográficos, antecedentes, síntomas, escalas, neuroimagen	69.6	16
Demográficos, antecedentes, síntomas, escalas, eeg	82.6	19
Demográficos, antecedentes, síntomas, irm, eeg	69.6	16
Demográficos, antecedentes, escalas, neuroimagen, eeg	87.0	20
Demográficos, síntomas, escalas, neuroimagen, eeg	82.6	19
Antecedentes, síntomas, escalas, neuroimagen, eeg	82.6	19
Séis		
Demográficos, antecedentes, síntomas, escalas, neuroimagen, eeg	82.6	19

Tabla 2 Resultados de los últimos tres grupos.

Sólo una red responde con un error mínimo de 1 caso mal clasificado, lo que nos da idea del grado de complejidad del problema, y como ya se había mencionado, pensamos que esto se debe a que probablemente, el conjunto de entrenamiento no es linealmente separable.

Las 11 redes que se desempeñaron mejor están señaladas con sombras dentro de la tabla. La que mejor funciona es la que contiene en su conjunto de entrenamiento los datos demográficos y las escalas (solo clasificó mal uno de los 23 casos de prueba). La red que solamente contiene a las escalas obtuvo el diagnóstico correcto en 21 de los 23 casos de prueba, que es también un resultado mejor que el obtenido con la red en que se toman en cuenta los 6 grupos de pruebas.

Así como la red de las escalas, existen otras 4 redes que sólo clasifican mal dos de los casos y otras 5 que fallan en tres, obteniendo de esta manera mejores porcentajes de reconocimiento que la red donde se utilizan los seis grupos de pruebas.

Si se numeraran progresivamente las 11 mejores redes de la tabla, se tendría que la número uno es la que solo contiene al grupo de escalas, la número dos contiene los datos demográficos y las escalas, y así sucesivamente, con lo que se puede crear una tabla (Tabla) con los datos de las redes para obtener cuales son los grupos de datos que más participan en el diagnóstico.

La red marcada con color más oscuro corresponde a la red que da mejores resultados (22 aciertos), las marcadas en color más claro corresponden a las que aciertan en 21 de los casos y las de color blanco son las que solo clasifican correctamente 20 de los 23 casos de prueba.

De la Tabla 3 se puede observar que en el conjunto de entrenamiento de las 11 redes se encuentran los datos del grupo de escalas, solo en la primera red señalada en la tabla y combinado en las otras 10.

Red	Grupo Demográficos	Antecedentes	Síntomas	Escalas	EEG	Neuroimagen
1				✓		
2	✓					
3		✓		✓		
4				✓	✓	
5	✓		✓	✓		
6	✓			✓	✓	
7		✓		✓	✓	
8	✓	✓		✓		✓
9	✓		✓	✓	✓	
10		✓	✓	✓		✓
11	✓	✓		✓	✓	✓

Tabla 3 Grupos de variables que más contribuyen al diagnóstico.

El grupo que se repite en 6 ocasiones es el de los datos demográficos, pero la red que se entrena sólo con los datos demográficos obtiene un resultado muy pobre (3/23). Los antecedentes y el grupo de electrofisiología (EEG) se repiten en 5 redes, pero los resultados individuales son de 12/23 casos correctos en ambas redes, y cuando se utiliza la combinación de los dos grupos, el resultado es de 13/23 casos correctos; ninguno de estos resultados es aceptable. Pero, al combinarse con el grupo de las escalas, obtienen un resultado de 20/23, que es mejor que el obtenido con los seis grupos.

El grupo de síntomas y el de neuroimagen se repiten en 3 ocasiones cada uno. Por separado, los grupos obtienen resultados de 9/23 y 13/23 respectivamente. Al combinar estos dos se obtiene un resultado de 14/23. Ninguno de estos resultados nos indica que realmente contribuyan al diagnóstico, pero al combinarse con otros de los grupos se obtienen mejores resultados, como es el caso de la red 10. En esta red se encuentran combi-

nados los grupos de antecedentes, síntomas, escalas y neuroimagen, lo que no ayuda a saber cual es la verdadera contribución de estos dos grupos.

Otro punto que es importante analizar en esta parte es saber cuáles son los casos en los que se equivocan las redes. La Tabla 4 muestra la red y el número de caso en el que se equivocó al dar el diagnóstico.

Red \ # caso	4	5	12	14	16	17	18	20	22	24
Escalas						✓		✓		
Demográficos, Escalas						✓				
Antecedentes, Escalas			✓					✓		
Escalas, EEG			✓						✓	
Demográficos, Síntomas, Escalas	✓	✓						✓		
Demográficos, Escalas, EEG				✓			✓			
Antecedentes, Escalas, EEG			✓	✓			✓			
Demográficos, Antecedentes, Escalas, Neuroimagen					✓	✓				✓
Demográficos, Síntomas, Escalas, EEG		✓		✓						
Antecedentes, Síntomas, Escalas, Neuroimagen		✓			✓					✓
Demográficos, Antecedentes, Escalas, Neuroimagen, EEG						✓		✓		✓

Tabla 4 Número de caso en el conjunto de prueba en los que las 11 mejores redes se equivocaron.

El grupo de casos en los que se equivoca la red es muy pequeño, lo que implica que las redes se equivocan casi siempre en los mismos. Los casos en los que más se equivocan son el 17 y el 20, incluyendo a la red que mejor funciona, pues se equivoca en el 17. Los casos 5, 12, 14 y 24 fueron mal clasificados por tres redes diferentes cada uno.

Entonces, existe un subconjunto en el grupo de prueba formada por 10 casos (para incluir a todos en los que hubo error) en los que se pueden presentar problemas a la hora de hacer el diagnóstico utilizando una RNA. La solución sería reentrenar a la red incluyendo en su conjunto de entrenamiento a estos 10 nuevos casos y posteriormente volver a probarlos. Esto se discutirá más a fondo en el capítulo de discusión de resultados.

El que existan 11 RNA's con menos características en su grupo de entrenamiento, que obtengan mejores resultados que la que se entrenó con todos los datos de los pacientes nos lleva directamente a una pregunta, ¿por qué funciona mejor una red con menos parámetros de entrada que la que contiene a todos? Es una pregunta con difícil respuesta. Una respuesta formada por varios puntos podría ser:

1. al ir incluyendo grupos de pruebas al conjunto de entrenamiento se aumenta la complejidad del problema;

2. en algunos casos, al agregar datos no se modifica la eficacia de la clasificación; en otros la disminuye y en otros la supera.

IV.2.1 Conjunto de costo mínimo

Existe otro criterio que hay que tomar en cuenta a la hora de decidir cuál es la mejor red; éste se refiere al costo que implica el realizar los estudios necesarios para obtener un buen diagnóstico. Evidentemente y con base en los resultados obtenidos, cualquiera de las 11 redes que se tienen como mejores, involucra un costo menor en cualquiera de los cuatro puntos que ya se han señalado y que son:

1. Riesgo para el paciente.
2. Infraestructura requerida.
3. Tiempo de realización de la prueba.
4. Factor económico (costo).

El análisis se hará tomando cada uno de los 6 grupos de datos y evaluándolo en función de los cuatro puntos anteriores, para al final obtener la red de menor costo. Los grupos están ordenados de la siguiente manera:

1. Demográficos
2. Antecedentes
3. Síntomas
4. Escalas Neurológicas y Neuropsicológicas
5. Electrofisiología
6. Neuroimagen y otros estudios (laboratorio, rayos X, doppler de vasos de cuello).

1. Factor "Riesgo"

Este, como ya se ha mencionado, es el más importante de los cuatro porque involucra directamente al paciente.

Con respecto al grupo 1, 2, 3 y 4 es notorio que no implican ningún riesgo de salud, pues todos esos datos se obtienen con preguntas y respuestas, es decir, el especialista se dedica a hacerle una larga lista de preguntas a las que el paciente o el familiar que lo acompaña tiene que responder y los datos son registrados en las hojas de recolección de información. Por tanto, no existe ninguna clase de riesgo asociado con la obtención de datos en estos cuatro grupos.

El grupo 5 son los estudios de EEG y P300. Para estos estudios, lo único que se hace es pegar electrodos en la cabeza por medio de un gel conductor. Solo se le dice al paciente que se acueste y haga lo que el técnico le pida, ya sea abrir o cerrar los ojos o estar atento para que pueda

diferenciar un sonido especial entre otros. Por tanto, estos estudios tampoco representan riesgo alguno para el paciente.

En el grupo 6 se realizan un mayor número de pruebas. Como en este caso están agrupados estudios con diferentes características se analizarán de uno por uno. Primero, estudios de laboratorio, principalmente el de sangre. Este estudio se hace cotidianamente e implica un riesgo cuando se llegan a dañar las venas del paciente o se puede contagiar con algún virus si no se realiza un manejo adecuado de las jeringas. Además, algunas veces este estudio se repite en el caso de que se esté aplicando algún tratamiento, lo que implica que tengan que volver a practicarlo con los consiguientes riesgos.

Los estudios de rayos X se piden cuando se quiere descartar algún otro problema, estos implican una radiación hacia el paciente que, aunque mínima, no es aconsejable estar sometiendo continuamente al paciente a este tipo de estudios.

La TAC y la IRM implican diferentes tipos de riesgo y problemas, a saber:

- a) la duración del estudio;
- b) medios de contraste.

Normalmente estos estudios requieren entre 30 y 120 minutos dependiendo del padecimiento, de si la enfermedad es casi evidente y con cortes grandes se tienen hallazgos, y de si el paciente permanece quieto, ya que no debe moverse. Esto es un problema porque no todos los pacientes pueden estar todo el tiempo quietos; si se llega a mover, hay que volver a sacar la placa en la que se movió. Si se necesitan cortes muy pequeños, esto retrasará más el estudio y el paciente se puede poner muy nervioso por estar dentro de un tubo grande en el que hace mucho frío. En ocasiones, si el paciente es claustrofóbico o ya está muy inquieto, se necesita llamar a un anestesiólogo para que evalúe si el paciente se encuentra en condiciones para recibir algún tipo de anestesia, que implica un riesgo mayor. En estos casos, se debe detener el estudio y en algunas ocasiones hay que posponerlo.

El otro factor de riesgo es el medio de contraste que se utiliza. Comúnmente se obtienen placas del mismo paciente con y sin contraste. En principio, el paciente pudiera no sentir ninguna molestia, con excepción del piquete en la vena por el que se pasa el medio de contraste y el riesgo que implica la jeringa. Después, el paciente puede presentar algunas molestias como calor interno o dolor, y debe empezar a tomar muchos líquidos para desechar el químico. Como se puede observar, estos estudios sí implican riesgo directo para el paciente y se debe tener cuidado con su aplicación.

Con respecto al "doppler", como es una aplicación de la ultrasonografía, permite la evaluación no invasiva del flujo sanguíneo, con la posibilidad de efectuar registros repetidos cuando así se requiere sin representar un peligro para el paciente.

2. Factor "Infraestructura"

Con respecto a la infraestructura de cada Institución ya se había hablado un poco en el capítulo 1. Éste, como país en desarrollo, no cuenta con el equipo suficiente en todas las Instituciones. Incluso muchas veces no se cuenta ni con el mínimo para tener una atención mediana. Para este punto no es necesario analizar a todos los grupos porque sólo en algunos se requiere de equipo especial como el grupo 4 y 5. Para los estudios de electrofisiología se necesita un polígrafo, equipo que es relativamente costoso, dependiendo de las características del mismo. No todos los hospitales cuentan con un polígrafo para realizar el EEG, y mucho menos con el equipo necesario para realizar una prueba de potenciales evocados, como el P300, ya que se requiere, además del aparato para graficar, un equipo de audífonos y un aparato generador de señales. Además, no toda la gente está capacitada para interpretar este tipo de estudios.

En el grupo 5 se encuentran las pruebas que requieren de grandes equipos que son muy costosos. El doppler es una técnica reciente, la interpretación del mismo requiere de personal capacitado y del equipo necesario para efectuarlo, así como un graficador. Son muy pocos los hospitales que cuentan con este equipo.

En la actualidad casi todos los hospitales cuentan con aparato para sacar placas de rayos X y este no sería un equipo difícil de conseguir.

Uno de los equipos más costosos es el de Resonancia Magnética. En todo el país existen alrededor de 10 equipos para realizar este estudio. Sólo algunos hospitales especializados y otros privados pueden tener un equipo como éste, lo que dificulta el acceso para realizarlos. El equipo de Tomografía Axial Computarizada también es muy costoso pero no tanto como el de resonancia magnética. Y no sólo el costo del equipo en sí es importante, también se necesitan instalaciones especiales, aire acondicionado, que se encuentren apartados de los demás equipos, como en el caso del aparato de resonancia, y además de todo esto, el costo por su mantenimiento es muy elevado. Por lo anterior, no todos los hospitales cuentan con este tipo de tecnología.

2. Factor "Tiempo"

Éste es otro factor muy importante, porque el diagnóstico y la posible modificación de la enfermedad depende en gran medida del tiempo que tarde en obtenerse los resultados de todos los estudios solicitados.

La información de los primeros tres grupos se puede obtener en la primer consulta con el médico, e incluso los datos de las cinco escalas del grupo 4 se pueden obtener en esa primera consulta. Por tanto, no se necesita de mucho tiempo para tener la información. Las otras tres pruebas del grupo 5 son estudios que tiene que realizar un psicólogo, y para ello, normalmente se tiene que pedir una interconsulta para que un especialista pueda evaluar al paciente. Sin embargo, no son citas que estén muy

alejadas y los resultados se pueden obtener en 2 o 3 días, dependiendo de la cantidad de personas que atienda la Institución.

Para el estudio de "doppler" y rayos X tampoco se requiere de mucho tiempo para obtener una cita. Se pueden tener los resultados en poco menos de una semana.

El problema se presenta con los estudios de tomografía y resonancia, que debido a que son muy pocas las instituciones que cuentan con los equipos, se debe hacer una orden para otro hospital en donde le puedan sacar las placas; considerando que es mucha la gente que por diversas enfermedades necesita un estudio de este tipo, que los hospitales deben tener espacio tanto para la gente que se atiende en el mismo hospital como para la gente que es canalizada de otra institución, las citas llegan a tardar de semanas a meses, con lo que se retrasa el que lleguen todos los resultados de los estudios realizados al paciente hasta el médico que lo está atendiendo y con ello el retraso para dar el diagnóstico definitivo.

4. Factor "Económico"

El cuarto y último factor es el económico. El precio de las pruebas del grupo 5 no es muy alto, incluso son de las pruebas más baratas. Debido precisamente a que se necesita de equipo muy costoso y a que no todas las instituciones cuentan con lo necesario, el grupo 6 contiene a las pruebas con costo muy elevado. En una institución como el Instituto Nacional de Neurología, donde la gran mayoría de la gente que se atiende es de bajos recursos, se utiliza un sistema de diferentes tarifas dependiendo del estatus económico del paciente. Existen 7 claves diferentes que son asignadas dependiendo de los resultados del estudio económico al paciente, estas van desde la A hasta la G, donde la A es la más barata y por consiguiente los estudios son pagados en precios mucho más bajos de lo que realmente valen. Con todo y esto, existe mucha gente que no cuenta con los recursos suficientes como para pagar tantos estudios como le sean necesarios. En hospitales privados, el costo es muy elevado y prácticamente inaccesible para la mayor parte de la población.

Todo esto es sin tomar en cuenta que mucha gente es canalizada de algunos de los estados de la república, con lo que el costo de su atención se eleva muchísimo al tener que pagar hospedaje y comidas de él y de los familiares que lo acompañan. Es por esta razón que se trata de encontrar el menor conjunto de pruebas y al menor costo posible evaluando los cuatro factores analizados.

En la Tabla 5 se encuentra un resumen de los costos valuados para los seis grupos de pruebas con los cuatro factores anteriores.

El signo "—" significa que no lleva ningún costo implicado o el costo es mínimo, el signo "½" implica un costo mediano en algunos de los

cuatro factores y el signo "+" implica un alto costo. En la última columna se tiene el costo total de los cuatro factores por grupo de pruebas.

Factores	Riesgo	Infraestructura	Tiempo	Económico	Costo total
Grupos					
Demográficos	—	—	—	—	—
Antecedentes	—	—	—	—	—
Síntomas	—	—	—	—	—
Escalas	—	—	—	—	—
Electrofisiología	—	½	½	½	½
Neuroimagen	+	+	+	+	+

Tabla 5 Factores de costo por grupos de pruebas.

De la tabla anterior se puede observar que el grupo de pruebas más costoso es el de neuroimagen, seguida por el de electrofisiología. Casualmente, el diagnóstico como ahora se hace, está reforzado por los estudios de TAC e IRM, es decir, de todo el conjunto de pruebas el "patrón oro" para los neurólogos es la neuroimagen.

Con base en estos resultados, ahora sí se puede saber cuál o cuáles son las redes menos costosas que obtienen un diagnóstico certero. Es obvio que la red que contenga en su conjunto de entrenamiento el grupo de neuroimagen será la red más costosa, con buenos resultados pero costosa en todos los sentidos. Es por esto que se debe observar con cuidado la Tabla 3, pues con ello podemos darnos cuenta de cuál grupo de variables interviene en los conjuntos de entrenamiento de los mejores redes.

Se puede observar que:

- La red número 1 sólo incluye en su conjunto de entrenamiento al grupo de las escalas, que tiene un costo muy bajo.
- La red 2 incluye a los datos demográficos y las escalas, por tanto su costo también es muy pequeño.
- El costo de la red 3 también es mínimo, porque utiliza los datos de los antecedentes y nuevamente a las escalas.
- La red 4 utilizó datos de escalas y EEG, por lo que su costo se eleva un poco más, pero aún es muy pequeño comparado con los estudios de neuroimagen.

Estas redes utilizan uno o dos grupos de pruebas cuando mucho, por lo que su costo es bajo.

- La quinta red utiliza tres grupos de pruebas, los datos demográficos, los síntomas y las escalas; a pesar de que son tres grupos, su costo es muy pequeño porque son datos que se obtienen por entrevista en la consulta.
- El costo de las redes 6 y 7 también es bajo porque utiliza dos grupos de pruebas de entrevista y solo uno que

requiere de equipo especializado, la electrofisiología. Estos, como ya se dijo, tienen un costo medio.

- La red 8 es de costo alto, porque aunque tiene datos obtenidos por entrevista contiene al grupo más caro, el de neuroimagen.
- El grupo 9 tiene cuatro grupos de datos, pero tres son de muy bajo costo y el cuarto (EEG) es de costo medio, por tanto su valor es aceptable.
- Las redes 10 y 11 incluyen 4 y cinco grupos de variables respectivamente, pero ambas incluyen al grupo de neuroimagen, por lo que su costo se eleva. Incluso la red 10 también tiene la información del grupo de EEG.

Haciendo una comparación de redes vs costo total se podría obtener algo como la Tabla 6, donde además se agrega una columna con el número de casos equivocados por red, para un posterior análisis:

Red	Costo total	Casos equivocados
1	—	17,80
2	—	17
3	—	18,80
4	½	18,82
5	—	4,5,80
6	½	16,18
7	½	18,14,18
8	+	16,17,24
9	½	5,14
10	+	5,16,24
11	+	17,20,24

Tabla 6 Costo total y casos equivocados por cada una de las redes seleccionadas

Con base en la tabla anterior, se puede afirmar que las redes de menor costo en todos los sentidos son la 1, 2, 3 y 5. Las redes 4, 6, 7 y 9 son de costo intermedio y las redes 8, 10 y 11 son las más costosas. En teoría, cualquiera de las 11 redes son potencialmente utilizables, pero si se toma en cuenta al costo entonces tenemos que eliminar a las redes 8, 10 y 11, por ser las más caras.

Hay otro factor que se debe de considerar antes de decidir cuál es la red que mejor funciona con un conjunto mínimo de datos y con un bajo costo: considerar en cuáles y en cuántos casos se equivocan. Para esto, hay que analizar las tres columnas de la Tabla 6 y con base en ella ahora sí decidir cuál es la mejor red.

La tabla por sí sola nos muestra el resultado, la red que menos se equivoca, tiene un conjunto de entrenamiento muy pequeño y un costo más bajo es la número 2. Existen otras redes como la número 1 y 3 que únicamente se equivocan en dos casos y su costo es muy bajo, además su conjunto de entrenamiento es sólo de dos grupos. La red 5 es de bajo costo,

su conjunto de entrenamiento es de 3 grupos y solo da un diagnóstico incorrecto en 3/23 casos. Exceptuando a las redes 8, 10 y 11, las demás serían buenas opciones para utilizar.

Ahora, es claro que con ninguna de las redes se puede obtener un 100% de certeza, es decir, ninguna de las redes reconoce todos los casos que se le presentan. Una solución para esto podría ser una combinación de las 7 redes con costos menores y grupos pequeños en su conjunto de entrenamiento para encontrar, por mayoría, el diagnóstico final. Esta solución se obtiene al observar que, como ya se mencionó, no todos las redes se equivocan en el mismo caso.

Lo primero es partir de la segunda red, ya que es la que se equivoca en el caso 17. Un criterio de inclusión podría ser no incluir, además de las que ya fueron excluidas, a las redes que se equivocaron en el mismo caso, es decir, el 17. Esta condición elimina a la red número 1, que contiene sólo a las escalas y que se equivoca en el caso 17 y además en el 20. Así quedan ya solo 7 redes con costo bajo y medio (redes 2, 3, 4, 5, 6, 7 y 9 según la tabla 7).

El tener 7 redes cuyos errores se encuentren entre 1 y 3 de 23 casos totales nos garantiza que se obtendrá un buen resultado al reunir el diagnóstico de las 7 y obtener, por mayoría, el diagnóstico final.

La nueva arquitectura propuesta se muestra a continuación:

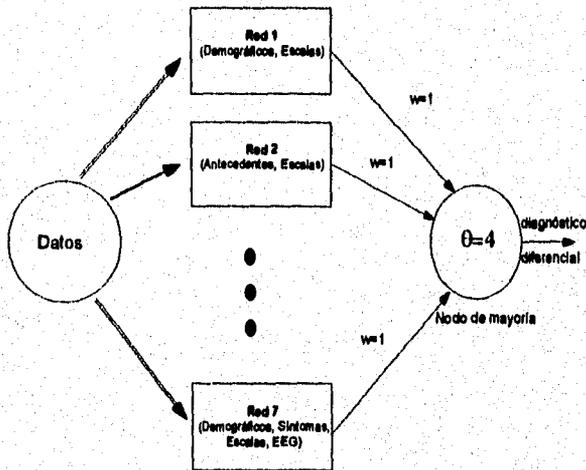


Fig. 1 Arquitectura propuesta.

Ahora obtendremos los resultados que se obtienen de esta nueva arquitectura. Este proceso se mostrará, como ejemplo, sólo en algunos de los casos mal clasificados.

a) Caso 17

Pongamos primero el caso 17, porque es el caso en el que se equivoca la mejor red. Para ello, lo primero que se hace es dar los datos necesarios a cada una de las redes y al final el nodo de mayoría decidirá el diagnóstico. El caso se probó en las 7 redes e individualmente se probaron obteniéndose los siguientes resultados:

La red 2 dice que los datos corresponden a DV
La red 3 dice que los datos corresponden a DTA
La red 4 dice que los datos corresponden a DTA
La red 5 dice que los datos corresponden a DTA
La red 6 dice que los datos corresponden a DTA
La red 7 dice que los datos corresponden a DTA
La red 9 dice que los datos corresponden a DTA

Dx. final obtenido por el nodo de mayoría: DTA

Dx. correcto del paciente: DTA

b) Caso 12

La red 2 dice que los datos corresponden a DTA
La red 3 dice que los datos corresponden a DV
La red 4 dice que los datos corresponden a DV
La red 5 dice que los datos corresponden a DTA
La red 6 dice que los datos corresponden a DTA
La red 7 dice que los datos corresponden a DV
La red 9 dice que los datos corresponden a DTA

Dx. final obtenido por la red: DTA

Dx. correcto del paciente: DTA

b) Caso 20

La red 2 dice que los datos corresponden a DV
La red 3 dice que los datos corresponden a DTA
La red 4 dice que los datos corresponden a DV
La red 5 dice que los datos corresponden a DTA
La red 6 dice que los datos corresponden a DV
La red 7 dice que los datos corresponden a DV
La red 9 dice que los datos corresponden a DV

Dx. final obtenido por la red: DV

Dx. correcto del paciente: DV

Los tres casos anteriores son en los que más errores tuvieron las 63 redes, y podemos observar que utilizando el nodo por mayoría estos errores se corrigen hasta alcanzar el 100% de confiabilidad.

V. Conclusiones

V. Conclusiones

El diseño de un sistema como el descrito en este trabajo junto con todas las herramientas y conocimientos necesarios para su preparación y funcionamiento requiere de la interacción de varias disciplinas, tanto científicas como tecnológicas, que definen lo que ahora se llama la Sexta Generación de computadoras, en la que se combinan máquinas seriales (con gran capacidad de proceso) con modelos conexionistas (que pueden ser entrenados para efectuar mapeos no lineales que difícilmente podrían caracterizarse). En el caso de este sistema, el modelo conexionista se encarga del mapeo entre las características que presentan los pacientes con demencia y el diagnóstico correspondiente.

En la actualidad se están desarrollando sistemas de este tipo en el área de la medicina (cada vez con más frecuencia), que ayudan a realizar el diagnóstico o la clasificación de diferentes tipos de enfermedades, como ya se ha mostrado. Es en esta área donde se hace más evidente la necesidad de contar con un sistema que auxilie al médico a tomar cierto tipo de decisiones (aunque ahora se utilizan en casi todas las áreas del conocimiento, por lo que no se debe despreciar su importancia en cada una de ellas), donde cada decisión que se toma tiene que ver con la vida de un ser humano, así que la precisión que debe tener el sistema, aunado a la experiencia del médico para determinar si es correcta la respuesta del sistema se vuelve de vital importancia.

Del sistema que se ha desarrollado a lo largo de esta tesis se pueden concluir varios puntos:

1. El primer objetivo general se cumplió, se construyó un modelo de RNA que realiza el diagnóstico diferencial entre DV y DTA, con un 82.6% de eficiencia. Aunque no es la precisión requerida para un problema de este tipo sí es la eficiencia necesaria para ser considerado un buen modelo que puede ser comparado con otros de su tipo (no sobre lo mismo, pero sí para diagnóstico) que en la actualidad se están utilizando.
2. El segundo objetivo se cumplió sobrepasando todas las expectativas planteadas. Se encontró un conjunto mínimo de pruebas con el que se puede realizar el diagnóstico diferencial con un 95.7% de eficiencia (solo un caso mal clasificado) y que implica un mínimo costo en los puntos ya mencionados. Otro punto muy importante referente a este resultado es que los estudios implicados en el grupo se pueden realizar en la **primer consulta con el médico**.
3. Además del conjunto mínimo anterior se encontraron otros seis grupos que tienen un porcentaje mayor de reconocimiento

que el de la primera red (entre 87 y 91.3%) y cuyo costo es mínimo o medio (ver tabla 6 del capítulo anterior).

4. Existen otras tres redes que lograron un 87% de eficiencia, pero cuyo conjunto de pruebas incluye a los estudios más caros con respecto a los cuatro puntos mencionados, además de que emiten una clasificación errónea en el mismo caso que la red que mejor funciona.
5. Se obtuvo una arquitectura que logró el 100% de reconocimiento al combinar las siete redes anteriores en un nodo de mayoría, con lo que se logra un sistema confiable.

Ahora bien, si se realiza una comparación con los sistemas ya implementados que se describen en el capítulo II, y con los que se mencionan en la bibliografía, se puede afirmar que no existe, entre estos, un sistema (con cualquier tipo de arquitectura de la RNA) con el que se pueda verificar un 100% de eficiencia, como el que se diseñó aquí.

Otro punto que es importante señalar es el desempeño del médico con sistema y sin él. Aunque no se realizaron pruebas como tales, sí se puede decir algo al respecto. La fase de prueba de las redes se realizó en varias partes, pues al principio solo se tenían muy pocos casos para probar. En un principio, se tenían redes que obtenían un 100% de reconocimiento. Sin embargo, a medida que el conjunto de prueba iba creciendo, el porcentaje de reconocimiento iba disminuyendo, con lo que la red que mejor funcionaba fallaba en seis casos. Casualmente casi todas las redes se equivocan en los mismos casos, aunque algunas los clasificaban bien, pero era la minoría. Entonces se realizó un análisis de estos casos, en conjunto con el médico, para tratar de descubrir que era lo que pasaba. Para realizar este análisis se revisaron nuevamente los expedientes con el fin de descubrir algún error a la hora de llenar las hojas de recolección de datos.

Para nuestra sorpresa, descubrimos que todas las entradas a la red eran correctas, sin embargo hubo dos errores. El primero fue que a la hora de hacer la BD por error se transcribieron mal los diagnósticos de dos de los casos, con lo que el porcentaje de reconocimiento de algunas redes aumentó y en otras disminuyó. Pero aún quedaban cuatro casos en los que fallaba el diagnóstico. Se analizaron nuevamente los casos, pero estos eran precisamente los que alcanzaban puntajes intermedios en algunas pruebas, con lo que el diagnóstico se vuelve incierto. El segundo error se encontró cuando se revisaron más detenidamente las placas de IRM. Entonces se pudo definir el diagnóstico correcto; la red había clasificado correctamente a tres de los casos que se habían marcado como erróneos, con lo que el sistema solo falló en uno que tenía características de demencia muy particulares.

De lo anterior se puede resumir que el desempeño del médico, cuando utilice este desarrollo en forma sistemática, va a mejorar, con lo que se lograría proporcionar mejor calidad de vida para el paciente con DTA al realizar un diagnóstico temprano (en las primeras fases de la

enfermedad) pues los pacientes en etapas severas tienen aún menos posibilidades de responder a cualquier intervención que se intente. En el caso de la DV a veces es probable detener la enfermedad, siempre y cuando ésta se encuentre relacionada con hipertensión arterial.

En el caso de las demencias, como ya se ha mencionado, cada caso presenta características muy particulares, con las que se puede complicar el diagnóstico. Para tratar de resolver este problema se debe contemplar dentro de este sistema un mantenimiento periódico, es decir, cuando se tengan más casos se debe re-definir el conjunto de entrenamiento, incluyendo en él los casos de más difícil diagnóstico y realizar así un re-entrenamiento de la red periódicamente. Esto con el fin de que cuando lleguen casos similares se puedan clasificar adecuadamente.

Este proceso de mantenimiento no llevaría inmerso el que el sistema se tenga que dejar de utilizar por mucho tiempo; el re-entrenamiento se llevaría a cabo en forma muy rápida, pues ya se cuenta con un conjunto de pesos definidos y a partir de ellos se empezaría de nuevo el proceso. Además, como se puede observar en el capítulo IV, aunque se tuviera que empezar desde cero, el entrenamiento de las siete redes que mejor funcionan no se lleva mucho tiempo, y comparado con el que se lleva entrenar a la red que contiene a los seis grupos de datos se puede decir que es un tiempo mínimo.

La DTA pronto se convertirá, en nuestro país, en la enfermedad más frecuente de las personas de la tercera edad de la vida. Las consecuencias de esto, desde el punto de vista médico, social, familiar y laboral son de enormes dimensiones. Cualquier esfuerzo por conocer la enfermedad y mitigarla, seguramente aportará beneficios importantes.

Por tanto, es importante hacer notar el posible uso de este sistema en forma sistemática, para obtener un diagnóstico certero en forma rápida, para el "beneficio" del paciente con demencia.

A. Apéndice

A. Interface de Usuario

Apéndice A

A. INTERFACE DE USUARIO

Los resultados del sistema nos llevan a afirmar que se obtendrá un diagnóstico más certero y en forma más rápida de lo que actualmente se hace. Por esta razón, es importante contar con una interface de usuario amigable, ya que el sistema será empleado por médicos que no están muy familiarizados con la codificación de datos y el manejo de las RNA. La idea es, entonces, obtener un sistema que contenga un lenguaje médico cotidiano y de fácil manejo, además de que obtenga el diagnóstico en forma transparente al usuario.

La interface de usuario es el mecanismo a través del cual se establece un diálogo entre el programa y el humano. Si se tienen en cuenta los factores humanos, el diálogo será fluido y se establecerá un ritmo entre el usuario y el programa.

Para este propósito se realizará una interface siguiendo los cuatro pasos mencionados a continuación (Pressman, 1993):

1. Análisis
2. Diseño
3. Implementación
4. Pruebas y resultados

A.1. Análisis

En esta parte se establecerán los requisitos de todo el sistema para que pueda interrelacionarse con otros elementos como el simulador de redes neuronales artificiales y el usuario (médicos). Es muy importante saber a quién va dirigido el sistema para tomar en cuenta los conocimientos que el usuario posee sobre sistemas de cómputo y el lenguaje que se manejará (en este caso, debe ser un lenguaje médico); con base en ellos se diseñará la interface. Un requisito importante es que el usuario no deberá realizar la codificación de datos a mano, es decir, solo manejará datos que él conozca e internamente el sistema, en forma automática y transparente para el usuario, realizará el proceso de codificación para posteriormente llevar esos datos a la RNA. Para este fin se utilizará la misma codificación

que se empleó a la hora de realizar el entrenamiento de la red y que fue diseñada de común acuerdo con el médico (Cap. 3).

Otra característica importante es que el usuario prefiere manejar un sistema que sea amigable, sobre todo si se toma en cuenta que los médicos no tienen contacto continuo con las computadoras y por tanto no están muy familiarizados con ellas. Por esta razón, el sistema deberá contar con una interface de usuario tipo "Windows", en la que el manejo que tenga que hacer el médico sólo sea para seleccionar opciones de una lista de términos conocidos para él y con el uso del "mouse", desplegar diferentes pantallas y menús.

Para este propósito se utilizará el programa Microsoft Access y Microsoft Visual Basic, por las facilidades que ofrecen para programar un ambiente tipo "windows" y porque la interface entre ellos y con el NeuroShell es directa. El uso de Access facilitará también el tener una base de datos con la información necesaria de cada paciente para su posterior consulta o actualización.

A.2. Diseño

El diseño, considerado como parte de la Ingeniería de Software está en gran parte oculto al usuario final, aunque es muy importante para la calidad global del software. El diseño de la interface de usuario es muy diferente a diseñar un sistema completo. Si es bueno, el usuario se adaptará a un ritmo normal de comunicación con la máquina. Pero si no lo es, el usuario lo sabrá inmediatamente y se sentirá a disgusto con un modo de interacción poco amigable.

El diseño de la interface de usuario tiene tanto que ver con el estudio de la gente como con aspectos de la tecnología. ¿Quién es el usuario? ¿Cómo aprende el usuario a interactuar con un sistema nuevo basado en computadora? ¿Qué espera el usuario del sistema? Estas son algunas de las preguntas que deben ser planteadas y respondidas como parte del diseño de la interface del usuario.

En la mayoría de los casos, más importante que la cultura general o la inteligencia del usuario es el conocimiento específico del terreno o del contexto para que el usuario comprenda el ámbito del problema y pueda interactuar efectivamente mediante una interface diseñada específicamente para ese propósito, cualquiera que éste sea.

Para el diseño de la interface se definirán tareas o acciones que puede realizar el usuario. Las acciones principales son:

1. Dar de alta casos nuevos.
2. Dar de baja.
3. Modificar o actualizar.
4. Consultar algún caso en especial.

El objetivo de cada una de las cuatro acciones anteriores está implícito en los mismos enunciados y éstas constituirán el menú de la base de datos. Las acciones se harán con base en el número de expediente del paciente. El propósito de la interface es que el usuario no tenga que aprenderse instrucciones complicadas, por lo que se propone sólo el uso de mouse y elegir entre botones de opciones múltiples, con excepción del número de expediente que sí tendría que teclearse.

En cualquiera de los cuatro casos de tareas, la secuencia a seguir será, después de seleccionar la opción a realizar:

1. seleccionar si se quiere obtener el dx de algún caso o se quiere ir a alguno de los grupos de datos;
2. Indicar con el número de expediente a cuál se le van a aplicar las acciones
3. realizar la acción requerida;
4. seleccionar si continúa con el siguiente grupo o se va al menú de grupos;
5. regresar al menú de la base de datos;
6. salir.

A grosso modo, estas son todas las secuencias que se realizan en el sistema, además de la interacción con las RNA, para lo que sólo se especificará que se quiere el diagnóstico con la ayuda de un botón. Todas estas acciones y secuencias se ejecutarán en ventanas y con botones; la información estará codificada (de la forma que se mencionó en el capítulo 3) y el usuario reconocerá el lenguaje médico, internamente se tendrán claves que irán a la base de datos para su posterior análisis.

Este diseño fue aprobado por el usuario, así como los primeros prototipos. Con la aceptación de éste, se procedió a implementar la interface.

A.3. Implementación

Como ya se mencionó, para realizar la interface de usuario se utilizarán los programas de Microsoft: Access v. 2.0 y Visual Basic v. 3.0; el simulador de RNA es el NeuroShell 2 v. 1.5.

Teniendo ya el diseño del sistema se implementaron las ventanas y botones, la distribución y los colores se eligieron para simular un ambiente tipo windows.

A partir de aquí, se explicará el manejo del sistema y se intercalarán imágenes de algunas de las ventanas más importantes.

El sistema inicia con una pantalla de presentación, que se puede observar en la Fig. 1. Después de la pantalla de presentación se presiona el botón "Empezar..." y se da el acceso a la pantalla del menú principal del programa. En esta pantalla se muestran las opciones disponibles como son: Ir a la base de datos ó analizar con RNA's. Esta ventana se muestra en la Fig. 2.

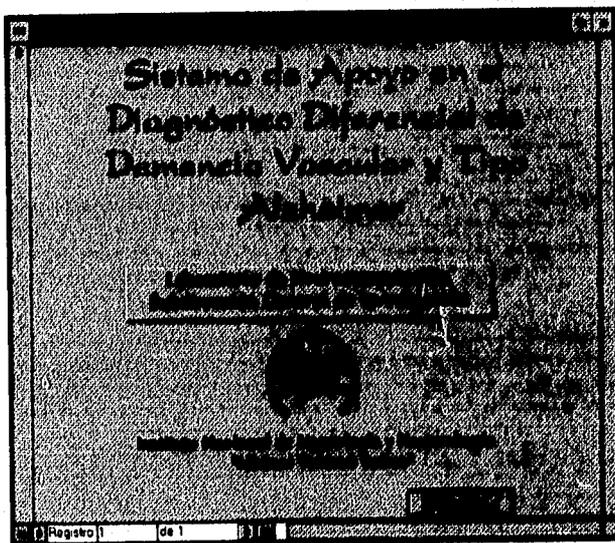


Fig. 1 Pantalla de presentación del sistema.

Si se elijera la opción de "Ir a la Base de Datos..." el programa abre la pantalla que contiene las opciones de las diferentes actividades que se pueden realizar. Esta ventana se muestra en la Fig. 4.

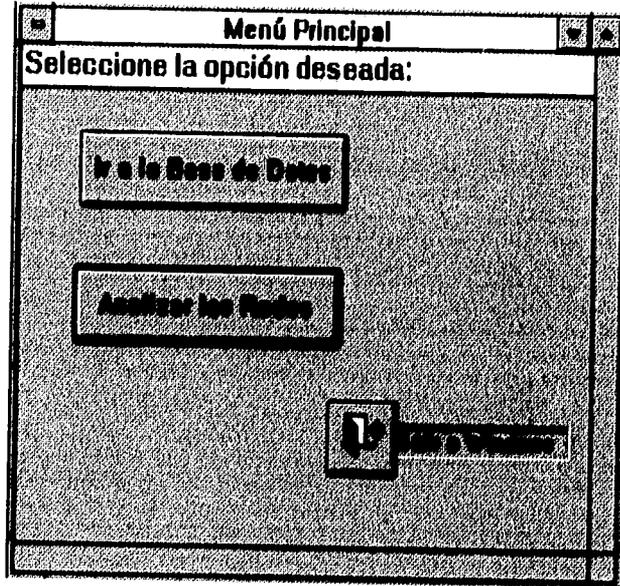


Fig. 2 Menú Principal del sistema.

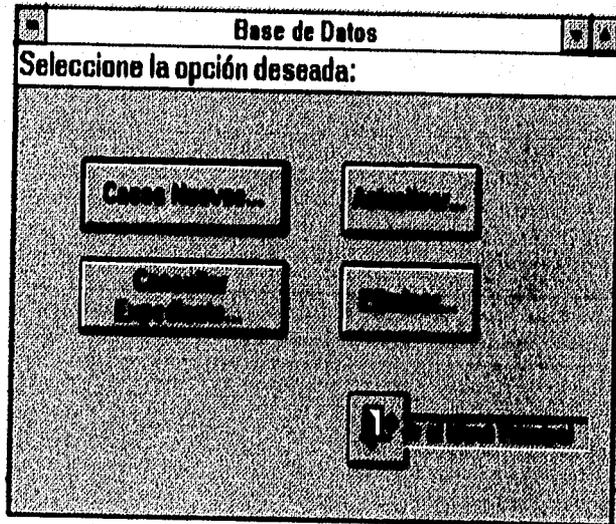


Fig. 3 Menú Base de Datos.

Cuando se selecciona la opción "Casos Nuevos..." el programa se va a otra pantalla llamada "Grupos de Datos" que se muestra en la Fig. 4.

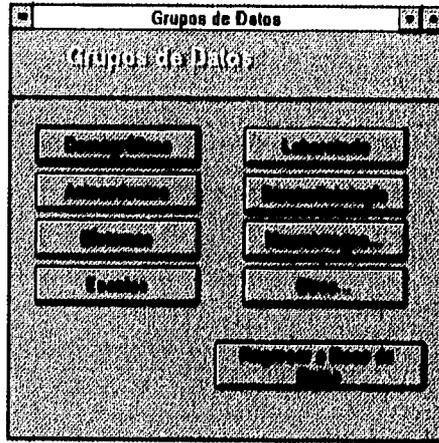


Fig. 4 Grupos de datos a los que se pueden acceder.

Al seleccionar cualquiera de las opciones de la pantalla se puede acceder a una nueva ventana en donde se pueden llenar los datos del paciente. Todas las pantallas son del mismo tipo, a continuación se muestra sólo una de ellas a manera de ejemplo.

Fig. 5 Opción de datos demográficos.

Como se puede observar en la Fig. 5, el lenguaje de los datos no está basado en números de codificación, sino en un lenguaje médico característico, aunque en esta pantalla se observan palabras comunes, existe en el campo de la medicina una clasificación estándar para cada una de las opciones que se presentan en la pantalla y que es la que se toma como base para desarrollar esta interfaz. Por ejemplo, el caso de la ocupación, es un menú desplegable en el que se encuentran las opciones de:

1. Desempleado/Jubilado
2. Ama de casa
3. Campesino/Obrero
4. Comercio/Técnico
5. Profesionalista

que es una clasificación que se utiliza comúnmente en estudios donde se implica la participación del paciente.

En esta pantalla, el médico tiene que teclear los datos del número de expediente, las iniciales del nombre y la edad del paciente pero es la única en la que se requieren datos que no se pueden conocer a priori. En las demás pantallas no se necesita teclear nada, solo seleccionar la opción adecuada. Los datos que se presentan en la pantalla son de un caso real. Cuando se quiere introducir datos de un caso nuevo todas las opciones aparecen en blanco.

Si se elige consultar o actualizar del menú de la base de datos mostrado en la Fig. 2, se accesa al menú Grupos de Datos, al seleccionar una opción aparece una pequeña pantalla donde se pregunta qué número de expediente se quiere consultar o actualizar. Esta ventana se muestra a continuación:

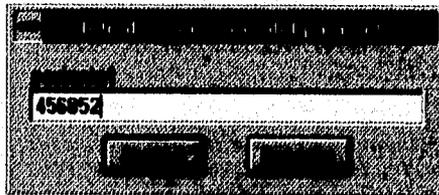


Fig. 6 Ventana para ir a un expediente específico.

En el caso de la opción "Eliminar" aparecerá una ventana en la cual se pregunta el número de expediente que será eliminado, se ejecuta la acción al seleccionar "aceptar" o se cancela. Después se regresa al menú de la Base de Datos.

Con estos procedimientos se realiza el almacenamiento y las diferentes opciones que se pueden realizar en una BD.

Cuando ya se han realizado todas las acciones anteriores y se quiere saber el diagnóstico sugerido por la red se debe ir al Menú principal y se selecciona la opción de "Analizar Redes", donde se debe escoger cada una de las siete mejores redes para que procesen los datos. En la siguiente figura se muestra la pantalla de presentación de este módulo.

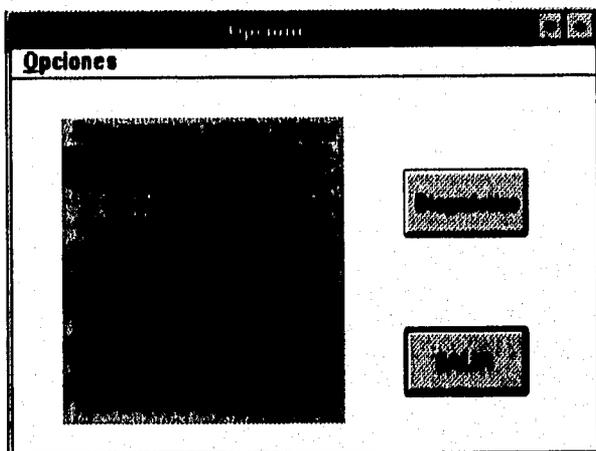


Fig. 7 Módulo para analizar redes.

Esta ventana contiene dos botones, al seleccionar el llamado "Diagnóstica" aparece una nueva ventana donde se muestra el conjunto de datos correspondientes a la red seleccionada (Fig. 8).

Al presionar el botón denominada "Cargar datos" aparecen automáticamente en pantalla los datos del caso que se está analizando.

Existe otra botón en la misma ventana llamada "Ejecutar red neuronal" que al presionarla, obtiene el diagnóstico (en forma transparente al usuario) y regresa un mensaje donde dice que ya obtuvo el diagnóstico y que continúe con la siguiente red.

Cuando se ha obtenido el diagnóstico de todas las redes, regresamos a la pantalla anterior, donde se selecciona el botón: "Resultada" y una nueva ventana como la de la Fig. 9 aparece dando el diagnóstico final.

d:/nshel2/demencia/trs/d_es_esp.dbl

RECORDS	ECOSIAS	FFD
Sexo <input type="text" value="2"/>	Hechinski <input type="text" value="2"/>	EEG <input type="text" value="24"/>
Edad <input type="text" value="24"/>	Leeb <input type="text" value="24"/>	P300 <input type="text" value="2"/>
Escolaridad <input type="text" value="19"/>	Minimal <input type="text" value="2"/>	
	Depresión <input type="text" value="19"/>	
	Act. cotidianas <input type="text" value="4"/>	

Buttons: [] []

Fig. 8 Ventana que ejecuta una de las siete mejores redes.

DE MENCIA

**El Diagnóstico es:
Demencia tipo Alzheimer**

[]

Fig. 9 Ventana de opciones con el diagnóstico final.

Finalmente, ya que se obtuvo el dx se puede regresar a la ventana del menú principal o salir del sistema.

En el caso de que se quiera ejecutar solo alguna de las 7 redes para obtener el dx, aparecerá un mensaje de error como el que se muestra a continuación.

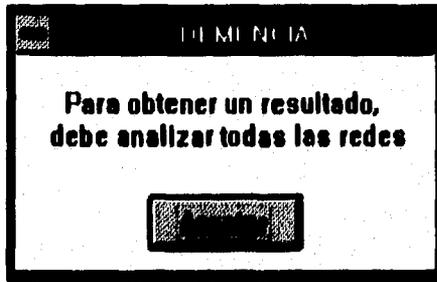


Fig. 10 Mensaje de error.

A.4. Pruebas y Resultados

El sistema fue probado por el médico neurólogo, quien realizó algunos cambios y sugerencias para que el sistema resultara más fácil de utilizar. Los cambios se realizaron y nuevamente se le presentó al médico.

El neurólogo dio su aprobación y procedió a probarlo con algunos pacientes del grupo de entrenamiento, quedando conforme con el resultado, que es el presentado en la sección anterior.

Se procedió a realizar el manual de usuario que también fue evaluado por el médico y concluyó que era lo suficientemente claro para su utilización.

VI. Referencias

VI. REFERENCIAS

- Alkon, D.; Blackwell, S.; Barbour, S.; Rigler, A.; Vogl, P. Pattern-Recognition by an Artificial Network Derived from Biologic Neuronal Systems. *Biological Cybernetics*. 1990;62:363-376
- Anderson, J. A Memory Model Using Spatial Correlation Functions. *Kybernetik* 5 1968;113-119.
- Anderson, J. Two Models of Memory Organization. *Math. Biosciences* 1991;8:137-160.
- Anderson, J. A.; Mozer, M.C. Categorization and Selective Neurons. *Parallel Models of Associative Memory*. eds G. E. Hinton and J. A. Anderson. Hillsdale: Erlbaum. 1981;213-236.
- Arbib, Michael. *Brains, Machines and Mathematics*. 2a. ed. Springer-Verlag. New York. 1987.
- Arbib, Michael. *The Metaphorical Brain 2. Neural Networks and Beyond*. John Wiley & Sons. New York. 1989.
- Barnett, Octo; Cimino, James; Hupp, Jon; Hoffer, Edward. Explain: An Evolving Diagnostic Decision-Support System. *JAMA* 1987;258:87-74
- Barto, A. G.; Sutton, R. S.; Brouwer, P. S. Associative Search Network: A Reinforcement Learning Associative Memory. *Biological Cybernetics*. 1981;40:201-211
- Baxt, William. Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction. *Annals of Internal Medicine*. 1991;115:843-848
- Bayles, K. Age at Onset of Alzheimer's Disease. Relation to Language Dysfunction. *Arch Neurol* 1991;48:155-159.
- Bogousslavsky J. The plurality of subcortical Infarction Stroke 1992;23:629-631.
- Bolla, K. Memory Complaints In Older Adults Fact or Fiction. *Arch Neurol* 1991;48:81-84.
- Boller F., López O., Mooney J. Diagnosis of Dementia Clinicopathologic Correlations. *Neurology* 1989;38:76-79.
- Boone, J.; Sigillito, V.; Sheber, G. Neural Networks in Radiology: An Introduction and Evaluation in a Signal Detection Task. *Medical Physics*. 1990a;17:2:234-241
- Boone, John M.; Gross, George W.; Greco-Hunt, Valerie. Neural Networks in Radiologic Diagnosis. I. Introduction and Illustration. *Investigative Radiology*. 1990b; 25:9:1012-1018.
- Bryson, A. E.; Ho, Y. C. *Applied Optimal Control*. New York: Blaisdell. 1969
- Calanillo, E.R. Outline of a Theory of Thought and Thinking Machines. *Journal of Theoretical Biology* 1961;1:204-235
- Casseiman, F.; Maj J. D. DASA/LARS: A Large Diagnostic System Using Neural Networks. *Proceedings of IJCNN 90-Wash DC Washington, DC (1990)* II-539-542
- Cooper, L. A Possible Organization of Animal Memory and Learning. *Proceedings of the Nobel Symposium on Collective Properties of Physical Systems*. Lundquist, B.; Lundquist, S., eds. New York: Academic Press. 1974
- Cragg, B. G.; Temperley, H. N. V. Memory: The Analogy with Ferromagnetic Hysteresis. *Brain* 1955;78,II:304-316.
- Cragg, B. G.; Temperley, H.N.V. The Organizations of Neurones: A Cooperative Analogy. *EEG and Clinical Neurophysiology* 1954;8:85-92.
- Cummings J.L.: Dementia. In *Clinical Neuropsychiatry* Grune and Stratton, Inc. (ed.) 1985;75-95.

- Cummings, Jeffrey; Benson, Frank. *Dementia. A Clinical Approach*. 2a ed. Butterworth-Heinemann, Stonham, MA. 1992.
- Dayhoff, J. *Neural Networks Architectures: An Introduction*. Van Nostrand Reinhold. 1990.
- DeLeon, M.J. et al. Computed Tomography in Evaluations of Brain-Behavior Relationships in Senile Dementia of the Alzheimer's Type. *Neurobiol Aging*. 1980;1:69-79.
- DeLeon, M.J. et al. Positron Emission Tomographic Studies of Aging and Alzheimer's Disease. *Am J Neuroradiol*. 1983;4:568-571.
- Dhawan, A; Arata, L. Segmentation of Medical Images through Competitive Learning. *Computer Methods and Programs in Biomedicine*. 1993;(40):203-215.
- Diaz, C.; García de la Cadena, C. Compendio de Pruebas Neuropsicológicas para la investigación Clínica en Pacientes Neurológicos y Psiquiátricos. *Lab. Psicología Experimental INNN MVS (Ed.)*1993.
- Drachman D. New Criteria for the Diagnosis of Vascular Dementia: Do we Know enough yet? *Neurology* 1993;43:243-45
- DSMIII Working Group: Diagnostic criteria for dementia in American Psychiatric Association (eds.): *Diagnostic and Statistical Manual of Mental disorders DSMIII R*, ed. 3, revised. Washington DC, Am Psychiatric Association, 1987, pp. 103-107.
- Eccles, J.C. *The Physiology of Synapsis*. Springer Verlag. Nueva York, 1964
- Edenbrandt, L.; Devine, B.; Macfarlane, P. Neural Networks for Classification of ECG ST-T Segments. *Journal of Electrocardiography* 1992;25;(3):167-173.
- Eisdorfer, C; Cohen, D. Diagnostic Criteria for Primary Neuronal Degeneration of the Alzheimer's Type. *The Journal of Family Practice*. 1980;11:4:553-557.
- Eisinger, P.; Damasio, A. Neuropsychologic Detection of Abnormal Mental Decline in Older Persons. *JAMA* 1985;253:5:670-674.
- Fisher, P. et al. Depression in Dementia of the Alzheimer Type and Multi-infarct Dementia. *Am J Psychiatry*. 1990;147;1484-1487.
- Folstein, M.; Folstein, S. et al. Mini-Mental State Examination. *Journal of Psychiatric Research* 1975;12:189-198.
- Forette, F. y cols. Reliability of Clinical Criteria for the Diagnosis of Dementia. *Arch Neurol* 1989;46:846-848.
- Fox, J.H. et al. The Use of Computerized Tomography in the Diagnosis of Senile Dementia. *J Neurol Neurosurg Psychiatr* 1975;38:948-953.
- Freeman, J.; Skapura, D. *Neural Networks: Applications and Programming Techniques*. Addison Wesley. 1991
- Garduño, R.; Lara, R.; Sandoval, E. *Hacia un enfoque de sistemas biológicos. Cerebro Metáforas y Modelos*. Conacyt. México. 1986
- Gevins, Alan; Morgan, Nelson. Applications of Neural-Network (NN) Signal Processing in Brain Research. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 1988;36:7:1152-1161
- Gorelick Ph., Roman G. Vascular Dementia: A time to Seize the moment. *Neuroepidemiology* 1993;12:139-40.
- Gross, George W.; Boone, John M.; Greco-Hunt, Valerie; Greenberg, Bruce. Neural Networks in Radiologic Diagnosis. II. Interpretation of neonatal chest radiographs. *Inv Radiology*. 1990;25:9:1017-1023.
- Grossberg, S. Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors. *Biol. Cybern* 1976a;23:121-134.
- Grossberg, S. Adaptive Pattern Classification and Universal Recoding: II. Feedback, Expectation, Offaction, Illusions. *Biol. Cybernetics* 1976b;23:187-202.
- Grossberg, S. Competitive Learning: From Interactive Activation to Adaptive Resonance. *Cognitive Science* 1987b;11:23-63.

- Grossberg, S. Embedding Fields: A Theory of Learning with Physiological Implications. *Journal of Mathematical Psychology* 1969;6:209-239.
- Grossberg, S. How Does the Brain Build a Cognitive Code? *Psychol Review*. 1980;87:1-51
- Grossberg, S. Neural Expectation: Cerebellar and Retinal Analogs of Cells Fired by Learnable or Unlearned Pattern Classes. *Kybernetik* 1972;10:49-57
- Grossberg, S. Nonlinear Difference-Differential Equations in Prediction and Learning Theory. *Proceedings of the National Academy of Sciences USA*. 1967;58:1329-1334.
- Grossberg, S. Some Nonlinear Networks Capable of Learning a Spatial Pattern of Arbitrary Complexity. *Proceedings of the National Academy of Sciences USA* 1968a;59:368-372.
- Grossberg, S. Some Physiological and Biochemical Consequences of Psychological Postulates. *Proceedings of the National Academy of Sciences USA* 1968b;60:758-765.
- Grossberg, S. *The Adaptive Brain*, 2 vols. Amsterdam: Elsevier. 1987a.
- Hachinski V. C.; Lassen N.A. Multi-infarct Dementia: A Cause of Mental Deterioration in the Elderly. *Lancet* 1974;2:207-210.
- Hallgren, R.; Reynolds, H. Computer Display of Multidimensional Biomedical Data. *Journal of Clinical Engineering*. 1992;17:3:235-243
- Hebb, D. O. *The Organization of Behavior*. New York: Wiley. 1949
- Heikala, E. L. et al. Different Patterns of Cognitive Decline Related to Normal or Deteriorating EEG a 3 year follow-up study of Patients with Alzheimer's Disease. *Neurology*. 1991;41:528-532.
- Hertz, J.; Krogh, A.; Palmer, R. G. *Introduction to the Theory of Neural Computation*. Addison Wesley Publishing Company. 1991
- Heston, L. *Dementia*. W.H. Freeman & co. 1983
- Hinton, G.; Sejnowski, T. Learning and Relearning in Boltzmann Machines. *Parallel Distributed Processing*. 1986:1.
- Hinton, G. E.; Sejnowski, T. J. Optimal Perceptual Inference. *Proceedings of the IEEE Conference in Computer Vision and Pattern Recognition (Washington, 1983)*. New York: IEEE. 1983;448-453.
- Hodgkin, A.L.; Huxley, A.F. A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve. *Journal of Physiology*. 1952;117:500-544.
- Hopfield, J. Neural Networks and Physical Systems with Emergent Computational Abilities. *Proc. Natl. Acad. Sci. USA*. vol 1982;79:2554-2558.
- Jandó, Gábor; Sayal, Masud. Automated Interictal EEG Spike Detection Using Artificial Neural Networks. *Electroenceph. Clin. Neurophysiol.*, 1992;83:271-280.
- Jandó, Gábor; Siegel, Ralph; Horváth, Zolt; Buzeáki, György. Pattern Recognition of the Electroencephalogram by Artificial Neural Networks. *Electroenceph Clin Neurophysiol*, 1993;86:100-109.
- Joachim C., Morris J., Selkoe A. Clinically diagnosed Alzheimer disease: Autopsy results in 150 cases. *Ann Neurol*. 1988;24:50-6.
- Kandel, R.; Shwartz, J. *Principles of Neural Science*. 2a. Ed. Elsevier. New York. 1985.
- Katz, B. *Nerve, Muscle and Synapse*. McGraw Hill. Nueva York . 1966.
- Katz, B. *The Release of Neural Transmitter Substances*. University Press Liverpool 1969.
- Khachaturain Z. Diagnosis of Alzheimer's Disease *Arch Neurol* 1985;42:1097-1104.
- Khachaturian, Z.S. y cols. Classification and Diagnosis of Alzheimer's Disease. An International Perspective WHO, Hogffre & Huber, 1989
- Kohonen, T. An Adaptive Associative Memory Principle. *IEEE Transactions on Computers C-23*. pp 1974;444-445.

- Kohonen, T. Analysis of a Simple Self-Organizing Process. *Biological Cybernetics* 44. 1982a;135-140.
- Kohonen, T. *Associative Memory: A System theoretic Approach*. Berlin: Springer. 1977
- Kohonen, T. *Self-Organization and Associative Memory*. Springer Verlag, Berlin. 1984.
- Kohonen, T. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*, 1982b;43:59-69.
- Kohonen, T.; Barna, G.; Chrisley, R. *Statistical Pattern Recognition with Neural Networks: Benchmarking Studies*. IEEE International Conference on Neural Networks (San Diego, 1988). New York: IEEE. 1988;1:61-68.
- Kohonen, T.; Mäkelä, K; Saramäki. Phonotopic Maps-Insightful Representation of Phonological Features for Speech Recognition. *Proceedings of the Seventh International Conference on Pattern Recognition (Montreal, 1984)*. New York: IEEE. 1984;182-185.
- Krall, V.A. The Relationship between senile dementia of the Alzheimer Type and Depression. *Canad J Psychiatr*. 1983;28:304-306.
- Kuffler, S.W.; Nichols, J.G. *From Neuron to Brain*. Sinayer Associates, Inc. MA 1976.
- Kukull W.A. y cols. Inter-lexer Reliability for Alzheimer's Disease Diagnosis. *Neurology* 1983;40:257-260.
- Kukull W.A., Larson E. B: The Validity of 3 Clinical Diagnostic Criteria for Alzheimer's Disease. *Neurology* 1990;40:1364-1369.
- Kulkowski, C.A. *Artificial Intelligence. Methods and Systems for Medical Consultation*. IEEE Trans Patt Anal Mech Intell. Sept. 1980.
- Le Cun, Y. Une Procédure d'Apprentissage pour Réseau à Seuil Asymétrique. *Cognitiva 85: A la Frontière de l'Intelligence Artificielle des Sciences de la Connaissance des Neurosciences Paris*. CESTA. 1985;599-604.
- Lippman R. An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*. 1987;4:4-22.
- Little, W. A. The Existence of Persistent States in the Brain. *Mathematical Biosciences* 1974;19:101-120.
- Little, W. A.; Shaw, G. L. A Statistical Theory of Short and Long Term Memory. *Behavioral Biology* 1975;14:115-133.
- Little, W. A.; Shaw, G. L. Analytic Study of the Memory Storage Capacity of a Neural Network. *Mathematical Biosciences* 1978;39:281-290.
- Loeb C. Clinical Criteria for Diagnosis of Vascular Dementia. *Eur. Neurol*. 1988;28:87-92.
- López O.L., Swihart A.A. Reliability of NINCDS-ADRDA Clinical Criteria for the Diagnosis of Alzheimer's Disease. *Neurology* 1990;40:1517-22.
- Mamelak, A.; Quattrochi, J.; Hobson, A. Automated Staging of Sleep in Cats Using Neural Networks. *Electroenceph. Clin. Neurophysiol*. 1991;79:52-61
- Marr, D. Simple Memory: A Theory for Archicortex. *Philosophical Transactions of the Royal Society of London B* 1971;262:23-81.
- Marr, D. A Theory of Cerebellar Neocortex. *Proceedings of the Royal Society of London B* 1970;178:161-234.
- Marr, D. A Theory of Cerebellar Cortex. *Journal of Physiology* 1969;202:437-470.
- Marr, D. *Vision*. San Francisco: Freeman. 1982
- Mattia, S. Mental Status Examination for Organic Mental Syndrome in the Elderly Patient. Bellack R. Krausu (Ed.) *Geriatric Psychiatry Grunne and Stratton*. 1978;77-121.
- Mc Culloch, W. S; Pitts, W. A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*. 1943;5,115-133.
- McKhann G., Drachman D., Folstein M. Clinical Diagnosis of Alzheimer Disease: Report of the NINCDS-ADRDA Work Group under the auspices of

- Department of the Health and Human Services Task Force on Alzheimer Disease. *Neurology*, 1984;34:939-944
- Meyer, JS; McClintic, KL; Rogers, RL; Sims, P.; Mortel, KF. Aetiological Considerations and risk factors for Multi-Infarct Dementia. *J Neurol Neurosurg Psychiatry*. 1988;51:1489-1497
- Miledi, R. Transmitter Release Induced by injection of Ca ions into Nerve Terminals. *Proceedings of the Royal Society of London, Biology* 1973;183:421-425.
- Minsky, M.; Papert, S. *Perceptrons: An Essay in Computational Geometry*. MIT Press. 1969
- Mirsen T., Lee D., Wong C. Clinical correlates of White-Matter Changes on Magnetic Resonance Imaging Scans of the Brain. *Arch Neurol* 1991;48:1015-1021.
- Mirsenth, Hachinski V. Epidemiology and Classification of Vascular and Multi-Infarct Dementia. In Meyer J., Lecher H. Marshall J. and Toole J. (eds.): *Vascular and Multi-Infarct Dementia*. Futura Publishing Co.Inc. 1988:47-61 pp.
- Moallem, Ciamac. Classifying Cells for Cancer Diagnosis Using Neural Networks. *IEEE Expert* 1991:8-12
- Nass, M. M.; Cooper, L. N. A Theory for Development of Feature Detecting Cells in Visual Cortex. *Biological Cybernetics* 1975;19:1-18.
- Noback, C.; Deamrest, R. *El Sistema Nervioso: Introducción y Repaso*. Interamericana McGraw-Hill. México, 1989.
- O'Brien M.D. Vascular Dementia: Problems with definitions nomenclature and classification. *J. Stroke Cerebrovasc. Dis.* 1992;2:217.
- Parker, D.B. *Learning Logic*. Technical Report TR-47, Center for Computational Research in Economics and Management Science. MIT, Cambridge, MA. 1985
- Patterson, J.V. et al. Average Evoked Potentials. In *Dementia en Reisberg, B. (eds.) Alzheimer's Disease*. The Free Press. 1983.
- Peretto, P. Collective Properties of Neural Networks: A Statistical Physics Approach. *Biological Cybernetics* 1984;50:51-62.
- Poli, R.; Cagnoni, S.; Livi, R.; Coppini, G.; Valli, G. A Neural Network Expert System for Diagnosis and Treating Hypertension. *Computer*. 1991:64-71
- Pressman, R. *Ingeniería del Software: Un enfoque práctico*. 3a ed. Mc Graw Hill México, 1993
- Reggia J.A., Tuhim S.; eds. *Computer-Assisted Medical Decision Making. Computers and Medicine*. v2 Springer Verlag, New York 1985
- Román, G.; Tatemichi, T.; Erkinjuntti, T. Vascular Dementia: Diagnostic Criteria for Research Studies. Report of the NINDS-AIREN International Workshop. *Neurology* 1993;43:250-260.
- Romanes, G.J.; Cunningham, *Tratado de Anatomía*. 12 ed. Interamericana, McGraw Hill. Madrid. 1981
- Rosenblatt, F. *Principles of Neurodynamics*, New York, Spartan Books. 1962
- Rovner, B.W.; Broadhead, J.; Spencer, M.; Carson, K.; Foistein, M. Depression and Alzheimer's Disease. *Am. J Psychiatry*. 1989;146:353-355.
- Rumelhart, D.; Hinton, G. & Williams, R. (1986b). Learning Internal Representations by Error Propagation, Chapter 8 in *Parallel Distributed Processing, Explorations in the Microstructure of Cognition*. vol. 1:Fundations. Cambridge, MA. MIT Press.
- Rumelhart, D.; Hinton, G. & Williams, R. Learning Representations by Back-Propagating Errors. *Nature* 1986a;323:533-536.
- Rumelhart, D.; Hinton, G.; Williams, R. Learning Representations by Retropropagación Errors. *Nature*. 1986;323:533-536
- Schwartz, W. B. *Medicine and the Computer. The Promise and Problems of Change*. *The New England Journal of Medicine* 1970;283:23;1257-1264

- Schwartz, W.; Patil, R.; Szolovitz, P. Artificial Intelligence in Medicine. Where do we stand? *The New England Journal of Medicine* 1987;316:11:685-688
- Sherrington, C.S. *The Integrative Action of the Nervous System*. Yale University Press, New Haven. 1906.
- Shortliffe, E.H. *Computer-based medical consultations: MYCIN*. Elsevier, New York. 1976
- Shortliffe, Edward. *Computer Programs to Support Clinical Decision Making*. *JAMA* 1987;258:81-66.
- Sparrow O., Strauss E.A. *A compendium of Neuropsychological tests administration, norms and commentary*. Oxford University, 1991.
- Spiegel, R. et al. *Psychopharmacology An Introduction*. John Wiley & Sons. 1981
- Steinbuch, K. (1961). *Die Lernmatrix*. *Kybernetik* 1. 1961;36-45.
- Szolovitz, P.; Patil, R.; Schwartz, W. Artificial Intelligence in Medical Diagnosis. *Annals of Internal Medicine*. 1988;108:80-87
- Taylor, W. K. *Electrical Simulation of Some Nervous System Functional Activities*. *Information Theory* ed. C. Cherry. London: Butterworths 1956;314-328.
- Ueda K., Kawano H. *Prevalence and Etiology of Dementia in a Japanese Community*. *Stroke* 1992;23:798-803.
- Villardita C. *Alzheimer's Disease compared with Cerebrovascular Dementia. Neuropsychological similarities and differences*. *Acta Neurol Scand* 1993;87:299-308.
- Von der Malsburg, Ch. *Self Organization of Orientation Sensitive Cells on the Striate Cortex*. *Kybernetik* 1973;14:85-100.
- Von Reutern A. *Transcranial Doppler Sonography Examination technique and normal reference values*. *Ultrasound Med. Biol.* 1991;8(8):746-761.
- Weinstein, J.; Kohn, K.; Grever, M.; Viswanadhan, V.; Rubinstein, L.; Monks, A.; Scudiero, D.; Welch, L.; Koutsoukos, A.; Chlause, A.; Paull, K. *Neural Computing in Cancer Drug Development: Predicting Mechanism of Action*. *Science*. 1992;258:447-451.
- Werbos, P. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph. D. Thesis, Harvard University. 1974
- Widrow, B. *Generalization and Information Storage in Networks of Adaline "Neurons"*. In *Self-Organizing Systems 1962*. eds. M.C. Yovits, G.T. Jacobs, & G.D. Goldstein, Washington: Spartan. 1962;435-461.
- Widrow, B.; Hoff, M. E. *Adaptive Switching Circuits*. In 1960 IRE WESCON Convention Record, part 4. New York: IRE. 1960;96-104.
- Widrow, B.; Rumelhart, D.; Lehr, Michael. *Neural Networks: Applications in Industry, Business and Science*. *Communications of the ACM*. 1994;37:3
- Willshaw, D.J.; Buneman, O.P.; Longuet-Higgins, H.S. *Non-holographic Associative Memory*. *Nature* 222. 1969;960-962.
- Wisler R. *Update on the Pathogenesis of Atherosclerosis*. *The Am J Med*. 1991; 91 (suppl. 1B):1B35-1B95.