



INSTITUTO POLITÉCNICO NACIONAL

ESCUELA SUPERIOR DE INGENIERÍA
MÉCANICA Y ELÉCTRICA

“Sistema Experto para la Predicción de la Hipoacusia”

PROYECTO

Para Obtener el Título de:

Ingeniero en Comunicaciones y Electrónica

PRESENTAN:

CÉSAR ROJAS JUÁREZ

FABIOLA SARABIA LÓPEZ

ASESORES TÉCNICOS:

DRA. MARÍA ELENA ACEVEDO MOSQUEDA

M. en C. MARÍA JULIA CALDERÓN SAMBARINO

ASESOR METODOLÓGICO:

M. en C. GENARO ZAVALA MEJÍA



México D.F.

INSTITUTO POLITÉCNICO NACIONAL
ESCUELA SUPERIOR DE INGENIERÍA MECÁNICA Y ELÉCTRICA
UNIDAD PROFESIONAL "ADOLFO LÓPEZ MATEOS"

TEMA DE TESIS

QUE PARA OBTENER EL TÍTULO DE
POR LA OPCIÓN DE TITULACIÓN
DEBERA (N) DESARROLLAR

INGENIERO EN COMUNICACIONES Y ELECTRÓNICA
TESIS COLECTIVA Y EXAMEN ORAL INDIVIDUAL
C. CESAR ROJAS JUAREZ
C. FABIOLA SARABIA LOPEZ

"SISTEMA EXPERTO PARA LA PREDICCIÓN DE LA HIPOACUSIA"

DISEÑAR E IMPLEMENTAR UN SISTEMA EXPERTO QUE PERMITA LA PREDICCIÓN DE LA HIPOACUSIA.

- DISEÑO DE ALGORITMOS PARA UN SISTEMA EXPERTO IMPLEMENTANDO MEMORIAS ASOCIATIVAS MORFOLÓGICAS
- ANÁLISIS, CLASIFICACIÓN Y CODIFICACIÓN DE BASE DE DATOS PARA SER IMPLEMENTADO EN EL SISTEMA
- REALIZAR PRUEBAS DE EFECTIVIDAD
- COMPARACIÓN DE LOS RESULTADOS OBTENIDOS.

MÉXICO D.F. A 09 DE NOVIEMBRE DE 2015

ASESORES

M. EN C. MARÍA ELENA ACEVEDO MOSQUEDA

M. EN C. MARÍA JULIA CALDERON SAMBARINO

ING. GENARO ZAVALA MEJÍA

ING. PATRICIA LORENA RAMÍREZ RANGEL
JEFE DEL DEPARTAMENTO DE
INGENIERÍA EN COMUNICACIONES Y ELECTRÓNICA



RESUMEN

El presente trabajo se divide en dos partes; en la primera parte se presenta la implementación de un algoritmo basado en Memorias Asociativas Morfológicas Auto asociativas, con este clasificador se puede obtener el diagnóstico y poder saber si una persona padece hipoacusia (incapacidad parcial o total para escuchar sonidos) o no.

La base de datos que se ha utilizado contiene registros de personas con hipoacusia y personas sanas.

En la segunda parte se hace uso del primer algoritmo, ahora codificando la base de datos utilizando el método Johnson Mòbius modificado y así utilizar como datos de entrada los vectores generados en la etapa de codificación de la base de datos.

Después de la implementación del clasificador, se realizan varias pruebas de efectividad con ambos algoritmos: *Leave One Out*, *Hold Out* y *K-Fold Cross Validation*.

Con los resultados obtenidos en las pruebas, se hace un promedio de efectividad de ambos clasificadores para quedarse con los mejores resultados, ya que no hay trabajos de clasificación de pacientes hipoacusicos y no hipoacusicos con los cuales se puedan basar los resultados obtenidos.

La efectividad del primer clasificador se obtuvo de 93.07%, y la efectividad del segundo clasificador es del 98.08% , superando con ello la efectividad del primer clasificador.

ABSTRACT

This paper is divided into two parts; in the first part of the implementation of a Morphological Associative Memories Based on Auto associative with this classifier algorithm is presented can obtain the diagnosis and to find out if a person has hearing loss (partial or total inability to hear sound) [1] or not.

The database that was used contains records of people with hearing loss and healthy people.

In the second part using the first algorithm is now coding the database using the modified method Johnson Mobius and thus used as input vectors generated in step coding database.

After implementation of the classifier, several tests are performed effectiveness both algorithms: Leave One Out, Hold Out and Validation Cross-Fold K.

With the results obtained in the tests, an ERA of both classifiers to keep the best results is made, since there is no classification work hard of hearing patients rather hard of hearing with which to base the results.

First classifier effectiveness of 93.07% was obtained, and the effectiveness of the second classifier is 98.08%, thus exceeding the effectiveness of the first classifier.

Índice General

Resumen.....	III
Abstract.....	IV
Índice Tablas.....	VII
Capítulo 1: Introducción	
1.1 Contexto.....	1
1.2 Identificación del Problema.....	2
1.3 Justificación de la propuesta.....	3
1.4 Objetivo General.....	4
1.5 Objetivos Específicos.....	4
Capítulo 2: Antecedentes y Estado del Arte	
2.1 Hipoacusia.....	5
2.1.1 Clasificación.....	5
2.1.2 Oído.....	5
2.1.3 Anatomía del Oído.....	6
2.1.4 Epidemiología.....	7
2.1.5 Tratamiento.....	7
2.1.6 Prevención.....	7
2.2 Estado del Arte.....	8
2.2.1 Análisis de señales acústicas de la voz humana.....	8
2.2.2 Nuevas Tecnologías en la rehabilitación del niño sordo.....	10
2.2.2.1 Programas para trabajar la voz, habla y audición.....	10
2.3 Conceptos Básicos.....	12
2.4 Memorias Asociativas.....	15
2.4.1 Lernmatrix de Steinbuch.....	15
2.4.2 Correlograph de Willshaw, Buneman y LonguetHiggins.....	16
2.4.3 Linear Associator de AndersonKohonen.....	17
2.4.4 La memoria asociativa Hopfield.....	18
2.4.5 Memorias Asociativas Alfa-Beta.....	19
2.4.6 Memorias Asociativas Media.....	20

Capítulo 3: Marco Teórico

3.1 Memorias Asociativas Morfológicas	22
3.1.1 Memorias Heteroasociativas Morfológicas.....	22
3.1.2 Memorias Autoasociativas Morfológicas	23
3.2 Diseño de una Memoria Asociativa	23
3.3 Selección de Rasgos.....	27
3.3.1 Algunas definiciones	28
3.3.1.1 Relevancia fuerte y débil.....	29
3.3.1.2 Optimilidad del subconjunto de rasgos	30
3.4 Codificación de Jonson-Möbius Modificado.....	30
3.5 Base de datos de Hipoacusia.....	32

Capítulo 4: Desarrollo

4.1 Diseño del Clasificador.....	33
4.2 Selección de rasgos de la Base de datos.....	34
4.3 Generación de Memorias Asociativas Morfológicas	34
4.4 Transformación de la Base de Datos	38
4.5 Programación C#	39

Capítulo 5: Resultados

5.1 Pruebas de Efectividad.....	39
5.1.1 Leave One Out.....	39
5.1.2 Hold Out.....	44
5.1.3 K-Fold Cross Validation.....	69
5.2 Resultados Globales	82

Conclusiones	85
---------------------------	----

Trabajos Futuros	85
-------------------------------	----

Referencias	86
--------------------------	----

Índice de tablas

Algoritmo 1

Tabla 1.1 Resultados de prueba Leave One Out	
Base de Datos Completa	41
Tabla 1.2 Resultados de prueba Leave One Out	
Base de Datos sin Jitter Factor	41
Tabla 1.3 Resultados de prueba Leave One Out	
Base de Datos sin PromRelP	41
Tabla 1.4 Resultados de prueba Leave One Out	
Base de Datos sin Ambos Campos	42
Tabla 2.1 Resultados de prueba Hold Out	
Base de Datos Completa (A - I)	45
Tabla 2.2 Resultados de prueba Hold Out	
Base de Datos sin Jitter Factor (A - I)	48
Tabla 2.3 Resultados de prueba Hold Out	
Base de Datos sin PromRelP (A - I)	51
Tabla 2.4 Resultados de prueba Hold Out	
Base de Datos sin Ambos Campos (A - I)	54
Tabla 3.1 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos Completa	70
Tabla 3.2 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos sin Jitter Factor	72
Tabla 3.3 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos sin PromRelP	73
Tabla 3.4 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos sin Ambos Campos	75
Tabla 4.1 Resumen de las pruebas Leave One Out	82
Tabla 4.2 Resumen de las pruebas Hold Out.....	82
Tabla 4.3 Resumen de las pruebas K-FoldCross Validation	83

Algoritmo 2

Tabla 1.1 Resultados de prueba Leave One Out	
Base de Datos Completa	43
Tabla 1.2 Resultados de prueba Leave One Out	
Base de Datos sin Jitter Factor	43
Tabla 1.3 Resultados de prueba Leave One Out	
Base de Datos sin PromRelP	44
Tabla 1.4 Resultados de prueba Leave One Out	
Base de Datos sin Ambos Campos	44
Tabla 2.1 Resultados de prueba Hold Out	
Base de Datos Completa (A - I)	57
Tabla 2.2 Resultados de prueba Hold Out	
Base de Datos sin Jitter Factor (A - I)	60
Tabla 2.3 Resultados de prueba Hold Out	
Base de Datos sin PromRelP (A - I)	63
Tabla 2.4 Resultados de prueba Hold Out	
Base de Datos sin Ambos Campos (A - I)	66
Tabla 3.1 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos Completa	77
Tabla 3.2 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos sin Jitter Factor	78
Tabla 3.3 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos sin PromRelP	79
Tabla 3.4 Resultados de prueba <i>K-Fold Cross Validation</i>	
Base de Datos sin Ambos Campos	80
Tabla 4.1 Resumen de las pruebas Leave One Out	82
Tabla 4.2 Resumen de las pruebas Hold Out.....	83
Tabla 4.3 Resumen de las pruebas K-FoldCross Validation	84

CAPITULO 1.

INTRODUCCION

1.1 CONTEXTO

El área de las memorias asociativas ha adquirido gran importancia en investigaciones a nivel internacional, principalmente en los campos con temas relacionados a la aplicación del reconocimiento y clasificación de patrones. Siendo este último la parte central y fundamental del trabajo.

El propósito fundamental de una memoria asociativa es recuperar correctamente patrones a partir de patrones de entrada, los cuales pueden estar alterados con ruido aditivo, sustractivo o combinado [1]. Para el diseño de una memoria asociativa se requiere de una fase de aprendizaje, que es el proceso mediante el cual se crea la memoria asociativa, formándose parejas de patrones asociados, uno de entrada y uno de salida, seguida de la fase de recuperación en donde se recuperan los patrones. Si en cada asociación el patrón de entrada es igual al de salida, la memoria es autoasociativa, de lo contrario es heteroasociativa .

Es posible clasificar datos usando diversos métodos y algoritmos, uno de los más comunes en esta área son las redes neuronales, también se pueden implementar con teorías probabilísticas, pero existe una diferencia entre los resultados obtenidos, es decir, en la efectividad de clasificación pues algunos métodos lograran un mejor porcentaje respecto a otros.

Las memorias asociativas morfológicas tienen respuesta perfecta, capacidad infinita de aprendizaje y almacenamiento (Díaz-de-León & Yáñez, 1999), además de poder trabajar con conjuntos de números reales positivos o negativos, no hay requisitos previos ni limitaciones respecto a la elección de los conjuntos de entrada o salida, por lo que no necesariamente deben ser diferentes o poseer características especiales.

En base a este fundamento, se implementa un clasificador de pacientes con hipoacusia, mediante el uso de memorias asociativas morfológicas, para obtener un alto porcentaje de efectividad de la clasificación.

1.2 IDENTIFICACIÓN DEL PROBLEMA

La hipoacusia (también conocida como deficiencia auditiva) es la incapacidad parcial o total para escuchar sonidos en uno o ambos oídos. Dado que su grado de esta afección va desde una afección leve hasta la pérdida total de la capacidad auditiva.

El órgano de la audición y el sentido ligado a él, proporcionan al ser humano uno de los medios más importantes de comunicación personal, social y cultural. Por ello, cualquier afección que altere su funcionamiento normal y que produzca un grado severo de sordera, va a trastornar de manera intensa las relaciones afectivas y sociales, dificultando la participación e integración social de los afectados.

En muchos casos se trata de una forma de marginación muy poco valorada en la sociedad, lo que hace que el aislamiento de estos pacientes termine siendo muy importante. Para que sirva a modo de ejemplo, podemos manifestar que la mayoría de teléfonos, timbres, despertadores y otros aparatos que utilizan señales sonoras, emplean tonos agudos que, precisamente, son los que las personas con problemas de sordera no pueden escuchar bien.

Los problemas de audición no sólo pueden deberse a un problema congénito, actualmente en las grandes ciudades la contaminación de ruido, el mal uso de las nuevas tecnologías para escuchar música, afectan al oído, haciéndose este problema cada vez más frecuentes en personas más jóvenes [2].

La pérdida auditiva relacionada con la edad (presbiacusia) es la pérdida lenta de la audición que se presenta a medida que las personas envejecen [3].

Actualmente llegan a presentarse casos en los cuales el diagnóstico de esta afección (hipoacusia) resulta equivoco, lo que tiene por consecuencia el empleo de un mal tratamiento el cual puede ocasionar afecciones aún mayores de las que ya se tenían, afectando gravemente así las posibilidades de recuperación de pacientes cuyos casos disponían de un tratamiento especializado y efectivo para su recuperación.

1.3 JUSTIFICACIÓN DE LA PROPUESTA

Más de 360 millones de personas padecen pérdida de audición incapacitante, según las nuevas estimaciones mundiales de su prevalencia publicadas por la Organización Mundial de la Salud (OMS) [4].

Otros 32 millones de afectados por pérdidas de audición son niños menores de 15 años. La principal causa de dicha discapacidad son las infecciones del oído.

Casi una de cada tres personas mayores de 65 años padece pérdida de audición, si no se trata, la pérdida auditiva afecta la comunicación y puede agravar el aislamiento social, la ansiedad, la depresión y el declive cognitivo.

La situación de las personas que padecen pérdida de audición puede mejorar con la utilización de aparatos, como audífonos, dispositivos de ayuda auditiva e implantes cocleares, y con el empleo de subtítulos, la formación en el lenguaje de signos y el apoyo educativo y social.

La OMS está ayudando a los países a desarrollar programas de atención primaria del oído y la audición integrados en los sistemas nacionales de atención primaria de salud.

Aproximadamente la mitad de los casos se podría evitar a través de la prevención temprana o un diagnóstico médico adecuado, de lo contrario derivaría en daños mayores e incluso permanentes para el paciente, ya que esta afección no sólo puede limitar la capacidad auditiva a ciertas frecuencias lo que implicaría que las personas que tengan estas características aparentemente tendrían una percepción auditiva correcta siendo que no es así.

Es por eso que resulta de suma importancia poder realizar un diagnóstico efectivo y confiable, que permita la diferenciación entre una persona hipoacusica y una persona sana con afecciones leves y se le pueda dar al paciente tratamiento adecuado.

Por lo que el desarrollo de una nueva técnica o método de diagnóstico para esta afectación con un mejor porcentaje de efectividad y confiabilidad resultaría de gran ayuda para lograr mejores y correctos tratamientos para cada uno de los pacientes.

En el presente proyecto se propone realizar el diagnóstico de hipoacusia mediante el uso de memorias asociativas morfológicas teniendo como resultado un sistema de diagnóstico altamente eficiente y confiable.

1.4 OBJETIVO GENERAL

Diseñar e implementar un sistema experto basado en memorias asociativas que permita la predicción de la hipoacusia de una persona con un alto porcentaje de efectividad.

1.5 OBJETIVOS ESPECÍFICOS

- Diseño de algoritmos para un sistema experto implementando memorias asociativas morfológicas.
- Análisis, clasificación y codificación de base de datos para ser implementado en el sistema.
- Realizar pruebas de efectividad.
- Comparación de los resultados obtenidos.

CAPITULO 2.

ANTECEDENTES Y ESTADO DEL ARTE

2.1 Hipoacusia

Hipoacusia, pérdida auditiva o sordera, es un déficit auditivo, que consiste en la incapacidad para escuchar sonidos en uno o ambos oídos, lo que puede tener un fuerte impacto en el estilo de vida, dificultando el desarrollo del habla, el lenguaje, la comunicación, el disfrute de las actividades y las interacciones sociales [5].

Los síntomas más típicos, es cuando el paciente habitualmente refiere oír menos sobre todo en ambientes ruidosos, en ocasiones acompañado de molestias, dolor o mareos [6].

2.1.1 Clasificación

La hipoacusia conductiva ocurre debido a un problema mecánico en el oído externo o el oído medio. Es posible que los tres minúsculos huesos del oído no conduzcan el sonido apropiadamente o que el tímpano no vibre en respuesta al sonido. La presencia de líquido en el oído medio puede causar este tipo de hipoacusia.

La hipoacusia neurosensorial se debe a un problema en el oído interno. Ocurre con mayor frecuencia cuando las terminales nerviosas que transmiten el sonido a través del oído están lesionadas o han muerto.

La hipoacusia conductiva es a menudo reversible, mientras que la neurosensorial no lo es. Las personas que tienen ambas formas de hipoacusia tienen hipoacusia mixta.

2.1.2 Oído

El oído, a pesar de su pequeño tamaño, es un órgano muy complejo. Actuando a modo de filtro, el oído transforma todos los sonidos que podemos oír en información precisa a la que el cerebro puede dar prioridad [7].

Cada oído está formado por unos mecanismos delicados. En el oído interno, una red de pequeñas células sensoriales y fibras nerviosas recogen las vibraciones del sonido y la transforma en impulsos eléctricos que el cerebro puede procesar.

Si el oído se expone a fuertes vibraciones de forma reiterada a lo largo del tiempo, las células sensoriales y las fibras nerviosas pueden resultar dañadas, y si no se pueden curar o remplazar, esto llevara a una pérdida auditiva permanente.

Un oído humano puede percibir diminutas variaciones de presión en el aire, si están en el rango de frecuencia audible, aproximadamente entre los 20Hz y los 20KHz, esto puede variar según cada persona y su edad [8].

2.1.3 Anatomía del oído

El oído está compuesto de tres partes [7,9]:

Oído externo: la forma de las orejas asegura que las ondas sonoras sean captadas y conducidas al tímpano a través del conducto auditivo.

Oído medio: está compuesto por el tímpano y tres huesecillos (martillo, yunque y el estribo), trabajan conjuntamente para amplificar las ondas sonoras.

Oído interno: el procesamiento empieza en el oído interno, donde las ondas sonoras se trasforman en impulsos eléctricos. La cóclea (también conocida como caracol) esta rellena de un líquido que las ondas sonoras mueven y cuyo moviendo es capturado por las células sensoriales que envían información sobre la posición en forma de impulsos nerviosos a través del nervio vestibular.

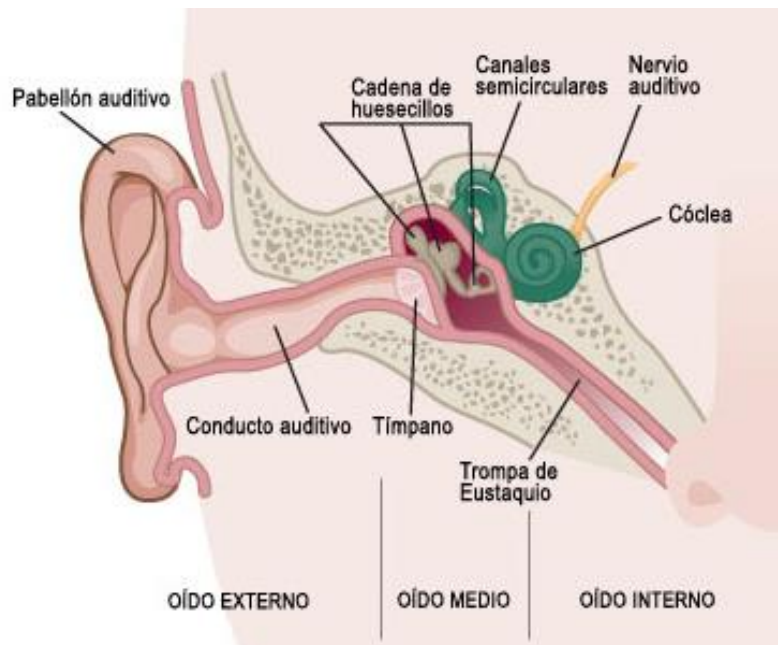


Figura1. Anatomía del Oído [9].

El cerebro: una vez que se envían los impulsos al cerebro, donde se interpreta como una sensación.

2.1.4 Epidemiología

Es común perder la capacidad auditiva a medida que se va envejeciendo, una de cada tres personas mayores de 65 años (165 millones de personas en todo el mundo) tienen pérdida de audición, no solo se relaciona con la edad, se asocia con más frecuencia al sexo masculino, la exposición al ruido, enfermedades infecciosas del oído, fármacos, daños del nervio auditivo, tabaquismo y diabetes [10].

La hipoacusia también se asocia a mayor número de personas con depresión, deterioro de la calidad de vida, deterioro cognitivo, lleva implícito un aumento de riesgo de accidentes, comprometiendo la seguridad física y la capacidad de vivir de forma independiente.

2.1.5 Tratamiento

El paciente se pueden realizar algunos de los exámenes, para diagnosticar el tipo de hipoacusia, como la Audiometría (una prueba auditiva empleada para verificar el tipo y magnitud de la hipoacusia), Tomografía computarizada de la cabeza (en caso de sospechar un tumor o una fractura) o una Resonancia magnética de oído y el cerebro [6,11].

Si la hipoacusia no desaparece con fármacos útiles para esta enfermedad en caso de alguna infección, entonces se puede practicar una cirugía para mejorar la audición. En otros casos un audífono le puede ayudar al paciente a oír. Los implantes cocleares solo se emplean en personas que son completamente sordas y que no se beneficiarían de un audífono.

En cada caso, el tratamiento farmacológico será individualizado, todo depende de las características y síntomas de cada paciente.

2.1.6 Prevención

Entre los medicamentos que pueden inducir a la disminución de la audición se encuentran los antibióticos administrados por vía endovenosa, medicamentos contra el cáncer también pueden causar pérdida auditiva y aquellos que disminuyen la irrigación en el oído interno como en los pacientes diabéticos e hipertensos [12].

Señala la OMS que las practicas idóneas como reducir la exposición a fuertes vibraciones de forma reiterada a lo largo del tiempo, pues puede provocar en las células sensoriales y las fibras nerviosas sean dañadas, y si no se pueden curar o reemplazar, esto llevara a una pérdida auditiva permanente, no introducir

objetos en los oídos, puede evitar que muchas personas desarrollen problemas auditivos .

Un gran porcentaje de personas con pérdida de audición pueden beneficiarse de un diagnóstico temprano y un tratamiento adecuado.

2.2 ESTADO DEL ARTE

El análisis acústico se considera como una herramienta objetiva, fiable y no invasiva en la medición de voz. Recientemente, el análisis acústico de voz se ha vuelto popular en situaciones clínicas debido a que la instrumentación es cada vez más fácil de usar y más accesible debido a los avances en la tecnología de procesamiento digital de señales, ya que hay una gran variedad de sistemas de análisis acústico comercial que ya están disponibles hoy en día. El documento (*Limitaciones de las medidas de perturbación en el análisis de voz acústica clínica*) presenta los resultados de un estudio que investiga el uso de análisis acústico para discriminar muestras de voz normales y patológicas [13].

La exactitud del análisis de la voz acústica está influenciada por la calidad de la grabación. El propósito del estudio (*Influencia, Adquisición de datos de Precisión para mediciones de calidad de voz acústica*) fue investigar la influencia del ambiente y el micrófono en la medición de la calidad de voz acústica. Más específicamente, se investigó el comportamiento relativo de los diferentes entornos de hardware, la relación entre sus características técnicas y la calidad de la voz acústica [14].

Se realizó un taller en Denver, Colorado para llegar a un mejor acuerdo sobre el propósito y los métodos de análisis acústico de las señales de voz. El patrocinio fue por el Centro Nacional para la voz y el habla, un centro de investigación y formación, financiado por el Instituto Nacional de la Sordera y Otros Trastornos de la Comunicación, y el Centro de Denver para las Artes Escénicas [15].

Los temas incluyeron las definiciones y la nomenclatura en el análisis de la voz, los algoritmos para la extracción de los parámetros, la grabación de alta fidelidad de las señales de micrófono, estructuras de archivos informáticos, el intercambio de bases de datos y desarrollo de las señales de prueba.

2.2.1 Análisis de señales acústicas de la voz humana

El análisis de señales acústicas de la voz humana tiene muchos propósitos. Desde un punto de vista tecnológico hay una necesidad cada vez mayor de

almacenar, cifrar, transmitir y sintetizar señales de voz. La industria de las telecomunicaciones ha dividido en dos la transmisión de la información ya sea en voz o datos, lo que sugiere que las señales de voz son una clase propia. Desde un punto de vista de la ciencia básica los investigadores han estudiado tradicionalmente las señales del micrófono para entender la producción y percepción del habla dado que la señal acústica es el vínculo común entre ellos. Por ultimo desde un punto de vista de la ciencia de la salud, la voz humana ha sido mostrada para llevar mucha información, acerca de la salud general y el bienestar de un individuo. La voz revela quienes somos y como nos sentimos, dando considerable información sobre la estructura y función sobre ciertas partes del cuerpo.

Este taller se limita al análisis de la voz en lugar del análisis del habla, se encuentra centrado en la extracción de información de la fuente de sonido de la señal de un micrófono.

Más específicamente, el taller fue una respuesta a la urgencia expresada por un grupo de científicos de voz, médicos de voz y fabricantes de instrumentos para llegar a un consenso sobre la utilidad, la viabilidad y la normalización de los métodos de perturbación de voz.

La producción de la voz se puede considerar como la activación de todo un sistema de osciladores acoplados.

Se puede asumir que el sistema de osciladores acoplados contiene y libera información sobre el cuerpo humano, en particular, acerca de su genética, el desarrollo, la edad, enfermedad, el idioma, la cultura, la comida, y la respuesta al medio ambiente (Figura 2). El análisis de la perturbación de voz tiene el objetivo de extraer algo de esta información de la señal de voz.

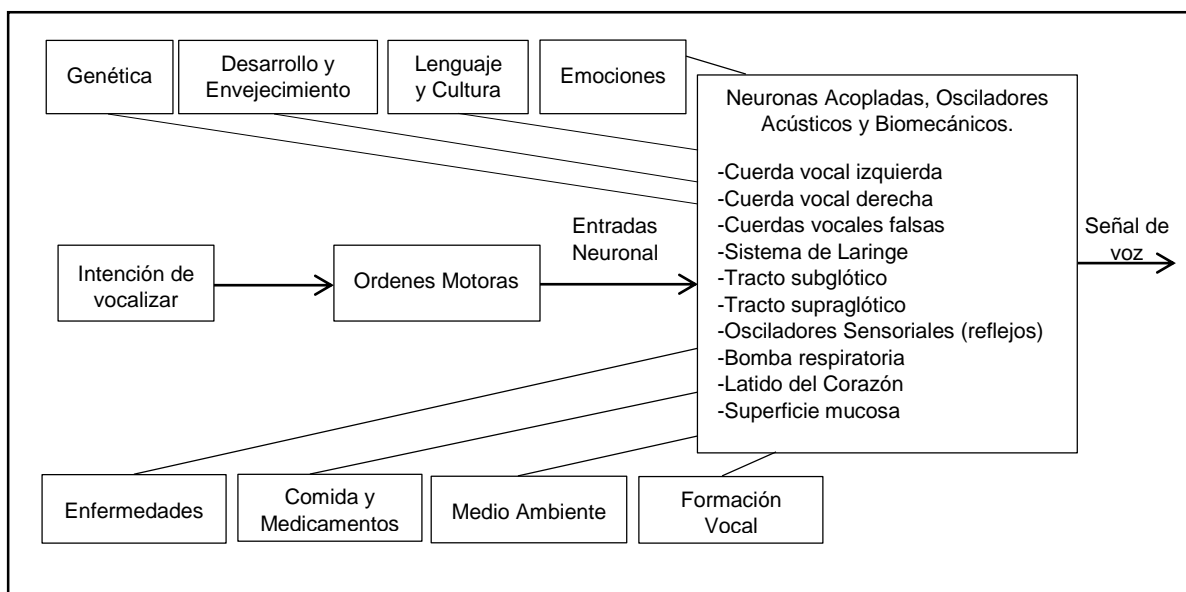


Figura 2. Osciladores biológicos implicados en la producción de la voz y los factores que pueden influir en ellos.

2.2.2 Nuevas tecnologías en la rehabilitación del niño sordo.

La interacción social es fundamental para el desarrollo del lenguaje y la comunicación el mecanismo de mediación que facilita este desarrollo. Esto implica que los aspectos funcionales de la lengua se aprendan mediante interacciones significativas.

Dentro de la población de deficiencia auditiva y teniendo en cuenta su diversidad (pérdida auditiva, momento del diagnóstico, tipo de prótesis, presencia de trastornos asociados.) se encuentran ciertas características comunes que dificultan el mantener una interacción comunicativa adecuada: retraso en la adquisición del habla y dificultades en el desarrollo del lenguaje. Este desfase hace que cometan errores en el empleo de las reglas gramaticales; un mal aprendizaje y un uso inadecuado del vocabulario; comprensión errónea de los mensajes orales. En el desarrollo de habilidades cognitivas: falta o dificultades para mantener la atención; en el almacenamiento correcto de las imágenes fonológicas de los sonidos del habla; en la estructuración espacio-temporal, pocas o malas estrategias para la resolución de problemas.

Dadas estas características es importante organizar y planificar actividades para crear situaciones comunicativas significativas en la intervención logopédica (la prevención, el diagnóstico, el pronóstico, el tratamiento y la evaluación integral de los trastornos de la comunicación humana, teniendo en cuenta los trastornos del habla o del lenguaje hablado o escrito, tanto en comprensión como en expresión). En la actualidad hay una gran variedad de materiales para trabajar durante la intervención con deficientes auditivos. Muchos de ellos se encontramos en soporte informático, no obstante, el logopeda deberá conocer qué programas informáticos pueden servirle en cada caso particular [32].

2.2.2.1 Programas para trabajar la voz, habla y audición

Algunos de los programas [32] que se pueden encontrar en el mercado para niños con deficiencia auditiva son los siguientes:

SPEECH VIEWER III

Distribuido por IBM (International Business Machines Corp), el cual se basa en producciones orales, mostrando las características del habla: tono, intensidad, sonoridad, duración, espectros, pronunciación. Ofrece interfaces muy atractivas para que los niños puedan comprender la relación entre los fenómenos visuales y el oral con gran motivación.

Sistemavisha

Distribuido por el Departamento de Ingeniería Electrónica ETSI de Telecomunicación de la Universidad Politécnica de Madrid. Desarrolla características de la voz y el habla (intensidad, tono y sonoridad).

Sistema AVEL

Elaborado por EGESA (Electrónica General ESPAÑOLA, S.A.) Es un programa para el reconocimiento de voz, basado en una tarjeta analógica incluida en un ordenador PC compatible. La entrada se realiza mediante un micrófono unido a una unidad de control que se conecta al ordenador. Permite el desarrollo de programas de articulación vocálica correcta, ejercicios de ritmo, control de tono y de intensidad.

Isoton

Elaborado por el Departamento de Ingeniería Electrónica E.T.S.I. y Telecomunicaciones. Permite entrenar parámetros del habla intensidad, sonoridad y tono fundamental. Es un sistema para el entrenamiento del habla, basado en la imitación de patrones establecidos o en el control de videojuegos por voz.

PIPA

Distribuido por TEA Ediciones. Es un programa interactivo de percepción auditiva, dirigido a psicólogos y pedagogos para trabajar con niños la percepción y discriminación auditiva.

Programa Sedea (programa secuenciado de desarrollo auditivo)

Desarrollado entre el Equipo de Rehabilitación del Colegio La Purísima para niños sordos y los profesionales de la empresa onda Educa (España). Programa que desarrolla aspectos como la atención, localización, discriminación, asociación entre imagen y sonido, identificación, reconocimiento, memoria y comprensión auditiva.

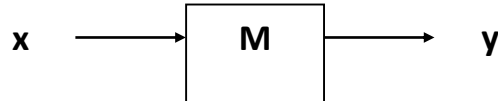
Memoria auditiva secuencial

Distribuido por el Hospital Santa Creu y Sant Pau (España). Programa con ejercicios para trabajar la memoria auditiva secuencial y para los problemas de reconocimiento auditivo.

2.3 Conceptos Básicos

El propósito fundamental de una memoria asociativa es recuperar correctamente patrones completos a partir de patrones de entrada, los cuales pueden estar alterados con ruido aditivo, sustractivo o combinado. Los conceptos utilizados en esta sección se encuentran en las referencias [17, 18, 19].

Una **Memoria Asociativa** puede formularse como un sistema de entrada y salida, idea que se esquematiza a continuación:



En este esquema, los patrones de entrada y salida están representados por vectores columna denotados por \mathbf{x} y \mathbf{y} , respectivamente. Cada uno de los patrones de entrada forma una asociación con el correspondiente patrón de salida, la cual es similar a la una pareja ordenada; por ejemplo, los patrones \mathbf{x} y \mathbf{y} del esquema anterior forman la asociación (\mathbf{x}, \mathbf{y}) .

No obstante que a lo largo de las dos secciones restantes del presente capítulo se respetarán las notaciones originales de los autores de los modelos presentados aquí, a continuación se propone una notación que se usará en la descripción de los conceptos básicos sobre memorias asociativas, y en el resto de los capítulos de esta tesis.

Los patrones de entrada y salida se denotarán con las letras negrillas, \mathbf{x} y \mathbf{y} , agregándoles números naturales como superíndices para efectos de discriminación simbólica. Por ejemplo, a un patrón de entrada \mathbf{x}^1 le corresponderá el patrón de salida \mathbf{y}^1 , y ambos formarán la asociación $(\mathbf{x}^1, \mathbf{y}^1)$; del mismo modo, para un número entero positivo k específico, la asociación correspondiente será $(\mathbf{x}^k, \mathbf{y}^k)$.

La memoria asociativa **M** se representa mediante una matriz, la cual se genera a partir de un conjunto finito de asociaciones conocidas de antemano: este es el **conjunto fundamental de aprendizaje**, o simplemente **conjunto fundamental**.

El conjunto fundamental se representa de la siguiente manera:

$$\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \dots, p\}$$

Donde p es un número entero positivo que representa la cardinalidad del conjunto fundamental.

A los patrones que conforman las asociaciones del conjunto fundamental se les llama **patrones fundamentales**. La naturaleza del conjunto fundamental proporciona un importante criterio para clasificar las memorias asociativas:

Una memoria es **Autoasociativa** si se cumple que $\mathbf{x}^\mu = \mathbf{y}^\mu \quad \forall \mu \in \{1, 2, \dots, p\}$, por lo que uno de los requisitos que se debe de cumplir es que $n = m$.

Una memoria **Heteroasociativa** es aquella en donde $\exists \mu \in \{1, 2, \dots, p\}$ para el que se cumple que $\mathbf{x}^\mu \neq \mathbf{y}^\mu$. Nótese que puede haber memorias heteroasociativas con $n = m$.

En los problemas donde intervienen las memorias asociativas, se consideran dos fases importantes: La fase de aprendizaje, que es donde se genera la memoria asociativa a partir de las p asociaciones del conjunto fundamental, y la fase de recuperación que es donde la memoria asociativa opera sobre un patrón de entrada, a la manera del esquema que aparece al inicio de esta sección.

A fin de especificar las componentes de los patrones, se requiere la notación para dos conjuntos a los que llamaremos arbitrariamente A y B . Las componentes de los vectores columna que representan a los patrones, tanto de entrada como de salida, serán elementos del conjunto A , y las entradas de la matriz **M** serán elementos del conjunto B .

No hay requisitos previos ni limitaciones respecto de la elección de estos dos conjuntos, por lo que no necesariamente deben ser diferentes o poseer características especiales. Esto significa que el número de posibilidades para escoger A y B es infinito.

Por convención, cada vector columna que representa a un patrón de entrada tendrá n componentes cuyos valores pertenecen al conjunto A , y cada vector columna que representa a un patrón de salida tendrá m componentes cuyos valores pertenecen también al conjunto A . Es decir:

$$\mathbf{x}^\mu \in A^n \text{ y } \mathbf{y}^\mu \in A^m \quad \forall \mu \in \{1, 2, \dots, p\}$$

La j -ésima componente de un vector columna se indicará con la misma letra del vector, pero sin negrilla, colocando a j como subíndice ($j \in \{1, 2, \dots, n\}$ o $j \in \{1, 2, \dots, m\}$ según corresponda). La j -ésima componente del vector columna \mathbf{x}^μ se representa por: x_j^μ

Con los conceptos básicos ya descritos y con la notación anterior, es posible expresar las dos fases de una memoria asociativa:

1. **Fase de Aprendizaje** (Generación de la memoria asociativa). Encontrar los operadores adecuados y una manera de generar una matriz \mathbf{M} que almacene las p asociaciones del conjunto fundamental $\{(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^2, \mathbf{y}^2), \dots, (\mathbf{x}^p, \mathbf{y}^p)\}$, donde $\mathbf{x}^\mu \in A^n$ y $\mathbf{y}^\mu \in A^m \quad \forall \mu \in \{1, 2, \dots, p\}$. Si $\exists \mu \in \{1, 2, \dots, p\}$ tal que $\mathbf{x}^\mu \neq \mathbf{y}^\mu$, la memoria será heteroasociativa; si $m = n$ y $\mathbf{x}^\mu = \mathbf{y}^\mu \quad \forall \mu \in \{1, 2, \dots, p\}$, la memoria será autoasociativa.
2. **Fase de Recuperación** (Operación de la memoria asociativa). Hallar los operadores adecuados y las condiciones suficientes para obtener el patrón fundamental de salida \mathbf{y}^μ , cuando se opera la memoria \mathbf{M} con el patrón fundamental de entrada \mathbf{x}^μ ; lo anterior para todos los elementos del conjunto fundamental y para ambos modos: autoasociativo y heteroasociativo.

Se dice que una memoria asociativa \mathbf{M} exhibe **recuperación correcta** si al presentarle como entrada, en la fase de recuperación, un patrón \mathbf{x}^ω con $\omega \in \{1, 2, \dots, p\}$, ésta responde con el correspondiente patrón fundamental de salida \mathbf{y}^ω .

Una memoria asociativa bidireccional también es un sistema de entrada y salida, solamente que el proceso es bidireccional. La dirección hacia adelante se describe de la misma forma que una memoria asociativa común: al presentarle una entrada \mathbf{x} , el sistema entrega una salida \mathbf{y} . La dirección hacia atrás se lleva a cabo presentándole al sistema una entrada \mathbf{y} para recibir una salida \mathbf{x} .

2.4 Memorias Asociativas

A continuación, en esta sección haremos un breve recorrido por los modelos de memorias asociativas, con objeto de establecer el marco de referencia en el que surgieron las memorias asociativas bidireccionales.

Las memorias asociativas que se presentarán en esta sección, son los modelos más representativos que sirvieron de base para la creación de modelos matemáticos que sustentan el diseño y operación de memorias asociativas más complejas. Para cada modelo se describe su fase de aprendizaje y su fase de recuperación.

Se incluyen cuatro modelos clásicos basados en el anillo de los números racionales con las operaciones de multiplicación y adición: *Lernmatrix*, *Correlograph*, *Linear Associator* y Memoria Hopfield, además de tres modelos basados en paradigmas diferentes a la suma de productos, a saber: memorias asociativas Morfológicas, memorias asociativas Alfa-Beta y memorias asociativas Media.

2.4.1 *Lernmatrix* de Steinbuch

Karl Steinbuch fue uno de los primeros investigadores en desarrollar un método para codificar información en arreglos cuadrículados conocidos como *crossbar* [29]. La importancia de la *Lernmatrix* [20, 21] se evidencia en una afirmación que hace Kohonen [22] en su artículo de 1972, donde apunta que las matrices de correlación, base fundamental de su innovador trabajo, vinieron a sustituir a la *Lernmatrix* de Steinbuch.

La *Lernmatrix* es una memoria heteroasociativa que puede funcionar como un clasificador de patrones binarios si se escogen adecuadamente los patrones de salida; es un sistema de entrada y salida que al operar acepta como entrada un patrón binario $\mathbf{x}^{\mu} \in A^n$, $A = \{0,1\}$ y produce como salida la clase $\mathbf{y}^{\mu} \in A^p$ que le corresponde (de entre p clases diferentes), codificada ésta con un método que en la literatura se le ha llamado *one-hot* [23]. El método funciona así: para representar la clase $k \in \{1, 2, \dots, p\}$, se asignan a las componentes del vector de salida \mathbf{y}^{μ} los siguientes valores: $y_k^{\mu} = 1$, y $y_j^{\mu} = 0$ para $j = 1, 2, \dots, k-1, k+1, \dots, p$.

Algoritmo de la *Lernmatrix*

Fase de Aprendizaje

Se genera el esquema (*crossbar*) al incorporar la pareja de patrones de entrenamiento $(\mathbf{x}^\mu, \mathbf{y}^\mu) \in A^n \times A^p$. Cada uno de los componentes m_{ij} de \mathbf{M} , la *Lernmatrix* de Steinbuch, tiene valor cero al inicio, y se actualiza de acuerdo con la regla $m_{ij} + \Delta m_{ij}$, donde:

$$\Delta m_{ij} = \begin{cases} +\varepsilon & \text{si } x_j^\mu = 1 = y_i^\mu \\ -\varepsilon & \text{si } x_j^\mu = 0 \text{ y } y_i^\mu = 1 \\ 0 & \text{en otro caso} \end{cases}$$

donde ε una constante positiva escogida previamente: es usual que ε es igual a 1.

Fase de Recuperación

La i -ésima coordenada y_i^ω del vector de clase $\mathbf{y}^\omega \in A^p$ se obtiene como lo indica la siguiente expresión, donde \vee es el operador *máximo*:

$$y_i^\omega = \begin{cases} 1 & \text{si } \sum_{j=1}^n m_{ij} \cdot x_j^\omega = \vee_{h=1}^p \left[\sum_{j=1}^n m_{hj} \cdot x_j^\omega \right] \\ 0 & \text{en otro caso} \end{cases}$$

2.4.2 *Correlograph* de Willshaw, Buneman y Longuet-Higgins

El *correlograph* es un dispositivo óptico elemental capaz de funcionar como una memoria asociativa [24]. En palabras de los autores “el sistema es tan simple, que podría ser construido en cualquier laboratorio escolar de física elemental”.

Algoritmo del *Correlograph*

Fase de Aprendizaje

La *red asociativa* se genera al incorporar la pareja de patrones de entrenamiento $(\mathbf{x}^\mu, \mathbf{y}^\mu) \in A^n \times A^m$. Cada uno de los componentes m_{ij} de la *red asociativa* \mathbf{M} tiene valor cero al inicio, y se actualiza de acuerdo con la regla:

$$m_{ij} = \begin{cases} 1 & \text{si } y_i^\mu = 1 = x_j^\mu \\ \text{valor anterior} & \text{en otro caso} \end{cases}$$

Fase de Recuperación

Se le presenta a la red asociativa \mathbf{M} un vector de entrada $\mathbf{x}^\omega \in A^n$. Se realiza el producto de la matriz \mathbf{M} por el vector \mathbf{x}^ω y se ejecuta una operación de umbralizado, de acuerdo con la siguiente expresión:

$$y_i^\omega = \begin{cases} 1 & \text{si } \sum_{j=1}^n m_{ij} \cdot x_j^\omega \geq u \\ 0 & \text{en otro caso} \end{cases}$$

donde u es el valor de umbral. Una estimación aproximada del valor de umbral u se puede lograr con la ayuda de un número indicador mencionado en el artículo [24] de Willshaw *et al.* de 1969: $\log_2 n$

2.4.3 Linear Associator de Anderson-Kohonen

El *Linear Associator* tiene su origen en los trabajos pioneros de 1972 publicados por Anderson y Kohonen [25,22].

Para presentar el *Linear Associator* consideremos de nuevo el conjunto fundamental:

$$\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \dots, p\} \text{ con } A = \{0, 1\}, \mathbf{x}^\mu \in A^n \text{ y } \mathbf{y}^\mu \in A^m$$

Algoritmo del Linear Associator

Fase de Aprendizaje

1) Para cada una de las p asociaciones $(\mathbf{x}^\mu, \mathbf{y}^\mu)$ se encuentra la matriz $\mathbf{y}^\mu \cdot (\mathbf{x}^\mu)^T$ de dimensiones $m \times n$

2) Se suman la p matrices para obtener la memoria

$$\mathbf{M} = \sum_{\mu=1}^p \mathbf{y}^\mu \cdot (\mathbf{x}^\mu)^T = [m_{ij}]_{m \times n}$$

de manera que la ij -ésima componente de la memoria \mathbf{M} se expresa así:

$$m_{ij} = \sum_{\mu=1}^p y_i^\mu x_j^\mu$$

Fase de Recuperación

Esta fase consiste en presentarle a la memoria un patrón de entrada \mathbf{x}^ω , donde $\omega \in \{1, 2, \dots, p\}$ y realizar la operación

$$\mathbf{M} \cdot \mathbf{x}^\omega = \left[\sum_{\mu=1}^p \mathbf{y}^\mu \cdot (\mathbf{x}^\mu)^T \right] \cdot \mathbf{x}^\omega$$

2.4.4 La memoria asociativa Hopfield

El artículo de John J. Hopfield de 1982, publicado por la prestigiosa y respetada *National Academy of Sciences* (en sus *Proceedings*), impactó positivamente y trajo a la palestra internacional su famosa memoria asociativa [26].

En el modelo que originalmente propuso Hopfield, cada neurona x_i tiene dos posibles estados, a la manera de las neuronas de McCulloch-Pitts: $x_i = 0$ y $x_i = 1$; sin embargo, Hopfield observa que, para un nivel dado de exactitud en la recuperación de patrones, la capacidad de almacenamiento de información de la memoria se puede incrementar por un factor de 2, si se escogen como posibles estados de las neuronas los valores $x_i = -1$ y $x_i = 1$ en lugar de los valores originales $x_i = 0$ y $x_i = 1$.

Al utilizar el conjunto $\{-1, 1\}$ y el valor de umbral cero, la fase de aprendizaje para la memoria Hopfield será similar, en cierta forma, a la fase de aprendizaje del *Linear Associator*. La intensidad de la fuerza de conexión de la neurona x_i a la neurona x_j se representa por el valor de m_{ij} , y se considera que hay simetría, es decir, $m_{ij} = m_{ji}$. Si x_i no está conectada con x_j entonces $m_{ij} = 0$; en particular, no hay conexiones recurrentes de una neurona a sí misma, lo cual significa que $m_{ij} = 0$. El estado instantáneo del sistema está completamente especificado por el vector columna de dimensión n cuyas coordenadas son los valores de las n neuronas.

La memoria Hopfield es autoasociativa, simétrica, con ceros en la diagonal principal. En virtud de que la memoria es autoasociativa, el conjunto fundamental para la memoria Hopfield es $\{(\mathbf{x}^\mu, \mathbf{x}^\mu) \mid \mu = 1, 2, \dots, p\}$ con $\mathbf{x}^\mu \in A^n$ y $A = \{-1, 1\}$

Algoritmo Hopfield

Fase de Aprendizaje

La fase de aprendizaje para la memoria Hopfield es similar a la fase de aprendizaje del *Linear Associator*, con una ligera diferencia relacionada con la diagonal principal en ceros, como se muestra en la siguiente regla para obtener la ij -ésima componente de la memoria Hopfield \mathbf{M} :

$$m_{ij} = \begin{cases} \sum_{\mu=1}^p x_i^{\mu} x_j^{\mu} & \text{si } i \neq j \\ 0 & \text{si } i = j \end{cases}$$

Fase de Recuperación

Si se le presenta un patrón de entrada \mathbf{x} a la memoria Hopfield, ésta cambiará su estado con el tiempo, de modo que cada neurona x_i ajuste su valor de acuerdo con el resultado que arroje la comparación de la cantidad $\sum_{j=1}^n m_{ij} x_j$ con un valor de umbral, el cual normalmente se coloca en cero.

Se representa el estado de la memoria Hopfield en el tiempo t por $\mathbf{x}(t)$; entonces $x_i(t)$ representa el valor de la neurona x_i en el tiempo t y $x_i(t+1)$ el valor de x_i en el tiempo siguiente ($t+1$).

Dado un vector columna de entrada \mathbf{x} , la fase de recuperación consta de tres pasos:

- 1) Para $t = 0$, se hace $\mathbf{x}(t) = \mathbf{x}$; es decir, $x_i(0) = x_i$, $\forall i \in \{1, 2, 3, \dots, n\}$
- 2) $\forall i \in \{1, 2, 3, \dots, n\}$ se calcula $x_i(t+1)$ de acuerdo con la condición siguiente:

$$x_i(t+1) = \begin{cases} 1 & \text{si } \sum_{j=1}^n m_{ij} x_j(t) > 0 \\ x_i(t) & \text{si } \sum_{j=1}^n m_{ij} x_j(t) = 0 \\ -1 & \text{si } \sum_{j=1}^n m_{ij} x_j(t) < 0 \end{cases}$$

- 3) Se compara $x_i(t+1)$ con $x_i(t) \forall i \in \{1, 2, 3, \dots, n\}$. Si $\mathbf{x}(t+1) = \mathbf{x}(t)$ el proceso termina y el vector recuperado es $\mathbf{x}(0) = \mathbf{x}$. De otro modo, el proceso continúa de la siguiente manera: los pasos 2 y 3 se iteran tantas veces como sea necesario hasta llegar a un valor $t = \tau$ para el cual $x_i(\tau+1) = x_i(\tau) \forall i \in \{1, 2, 3, \dots, n\}$; el proceso termina y el patrón recuperado es $\mathbf{x}(\tau)$.

En el artículo original de 1982, Hopfield había estimado empíricamente que su memoria tenía una capacidad de recuperar $0.15n$ patrones, y en el trabajo de Abu-Mostafa & St. Jacques [30] se estableció formalmente que una cota superior para el número de vectores de estado arbitrarios estables en una memoria Hopfield es n .

2.4.5 Memorias Asociativas Alfa-Beta

Las memorias Alfa-Beta [19] utilizan máximos y mínimos, y dos operaciones binarias originales α y β de las cuales heredan el nombre.

Para la definición de las operaciones binarias α y β se deben especificar los conjuntos A y B , los cuales son:

$$A = \{0, 1\} \quad \text{y} \quad B = \{0, 1, 2\}$$

La operación binaria $\alpha: A \times A \rightarrow B$ se define como:

x	y	$\alpha(x, y)$
0	0	1
0	1	0
1	0	2
1	1	1

La operación binaria $\beta: B \times A \rightarrow A$ se define como:

x	y	$\beta(x, y)$
0	0	0
0	1	0
1	0	0
1	1	1
2	0	1
2	1	1

Los conjuntos A y B , las operaciones binarias α y β junto con los operadores \wedge (mínimo) y \vee (máximo) usuales conforman el sistema algebraico $(A, B, \alpha, \beta, \wedge, \vee)$ en el que están inmersas las memorias asociativas Alfa-Beta.

2.4.6 Memorias Asociativas Media

Las Memorias Asociativas Media [28] utilizan los operadores A y B , definidos de la siguiente forma:

$$A(x, y) = x - y$$

$$B(x, y) = x + y$$

Las operaciones utilizadas se describen a continuación.

Sean $P = [p_{ij}]_{m \times r}$ y $Q = [q_{ij}]_{r \times n}$ dos matrices.

Operación \diamond_A : $P_{m \times r} \diamond_A Q_{r \times n} = [f_{ij}^A]_{m \times n}$ donde $f_{ij}^A = \mathbf{med}_{k=1}^r A(p_{ik}, q_{k,j})$

Operación \diamond_B : $P_{m \times r} \diamond_B Q_{r \times n} = [f_{ij}^B]_{m \times n}$ donde $f_{ij}^B = \mathbf{med}_{k=1}^r B(p_{ik}, q_{k,j})$

Algoritmo Memorias Media

Fase de Aprendizaje

Paso 1. Para cada $\xi = 1, 2, \dots, p$, de cada pareja $(\mathbf{x}^\xi, \mathbf{y}^\xi)$ se construye la matriz:

$$[\mathbf{y}^\xi \diamond_A (\mathbf{x}^\xi)^t]_{m \times n}$$

Paso 2. Se aplica el operador media a las matrices obtenidas en el paso 1 para obtener la matriz \mathbf{M} , como sigue:

$$\mathbf{M} = \mathbf{med}_{\xi=1}^p [\mathbf{y}^\xi \diamond_A (\mathbf{x}^\xi)^t]$$

El ij -ésimo componente \mathbf{M} está dado como sigue:

$$m_{ij} = \mathbf{med}_{\xi=1}^p A(y_i^\xi, x_j^\xi)$$

Fase de Recuperación

Se tienen dos casos:

Caso 1. Recuperación de un patrón fundamental. Un patrón \mathbf{x}^w , con $w \in \{1, 2, \dots, p\}$ se le presenta a la memoria \mathbf{M} y se realiza la siguiente operación:

$$\mathbf{M} \diamond_B \mathbf{x}^w$$

El resultado es un vector columna de dimensión n , con la i -ésima componente dada como:

$$(\mathbf{M} \diamond_B \mathbf{x}^w)_i = \mathbf{med}_{j=1}^n B(m_{ij}, x_j^w)$$

Caso 2. Recuperación de un patrón alterado. Un patrón \mathbf{x} , que es una versión alterada de un patrón \mathbf{x}^w , se le presenta a la memoria \mathbf{M} y se realiza la siguiente operación:

$$\mathbf{M} \diamond_B \mathbf{x}$$

De nuevo, el resultado es un vector de dimensión n , con la i -ésima componente dada como:

$$(\mathbf{M} \diamond_B \mathbf{x})_i = \mathbf{med}_{j=1}^n B(m_{ij}, x_j)$$

CAPÍTULO 3.

MARCO TEORICO

En este capítulo se presentan aquellos conceptos que dan soporte a la metodología propuesta en esta investigación, además se defienden los conceptos y términos que se utilizarán. En especial, se describirán los aspectos relacionados con memorias asociativas.

3.1 Memorias Asociativas Morfológicas

Hay dos tipos de memorias asociativas morfológicas: las memorias *max*, simbolizadas con \mathbf{M} , y las memorias *min*, cuyo símbolo es \mathbf{W} ; en cada uno de los dos tipos, las memorias pueden funcionar en ambos modos: heteroasociativo y autoasociativo [27].

Se definen dos nuevos productos matriciales:

El *producto máximo* entre \mathbf{D} y \mathbf{H} , denotado por $\mathbf{C} = \mathbf{D} \nabla \mathbf{H}$, es una matriz $[c_{ij}]_{m \times n}$ cuya ij -ésima componente c_{ij} es

$$c_{ij} = \bigvee_{k=1}^r (d_{ik} + h_{kj})$$

El *producto mínimo* de \mathbf{D} y \mathbf{H} denotado por $\mathbf{C} = \mathbf{D} \Delta \mathbf{H}$, es una matriz $[c_{ij}]_{m \times n}$ cuya ij -ésima componente c_{ij} es

$$c_{ij} = \bigwedge_{k=1}^r (d_{ik} + h_{kj})$$

Los productos máximo y mínimo contienen a los operadores máximo y mínimo, los cuales están íntimamente ligados con los conceptos de las dos operaciones básicas de la morfología matemática: *dilatación* y *erosión*, respectivamente.

3.1.1 Memorias Heteroasociativas Morfológicas

Algoritmo de las memorias morfológicas max

Fase de Aprendizaje:

1. Para cada una de las p asociaciones $(\mathbf{x}^\mu, \mathbf{y}^\mu)$ se usa el producto mínimo para crear la matriz $\mathbf{y}^\mu \Delta (-\mathbf{x}^\mu)^t$ de dimensiones $m \times n$, donde el negado transpuesto del patrón de entrada \mathbf{x}^μ se define como $(-\mathbf{x}^\mu)^t = (-x_1^\mu, -x_2^\mu, \dots, x_n^\mu)$.
2. Se aplica el operador máximo \vee a las p matrices para obtener la memoria \mathbf{M} .

$$\mathbf{M} = \bigvee_{\mu=1}^p [\mathbf{y}^\mu \Delta (-\mathbf{x}^\mu)^t]$$

Fase de Recuperación:

Esta fase consiste en realizar el producto mínimo Δ de la memoria \mathbf{M} con el patrón de entrada \mathbf{x}^ω , donde $\omega \in \{1, 2, \dots, p\}$, para obtener un vector columna \mathbf{y} de dimensión m :

$$\mathbf{y} = \mathbf{M} \Delta \mathbf{x}^\omega$$

Las fases de aprendizaje y de recuperación de las **memorias morfológicas min** se obtienen por dualidad.

3.1.2 Memorias Autoasociativas Morfológicas

Para este tipo de memorias se utilizan los mismos algoritmos descritos anteriormente y que son aplicados a las memorias heteroasociativas; lo único que cambia es el conjunto fundamental. Para este caso, se considera el siguiente conjunto fundamental:

$$\{(\mathbf{x}^\mu, \mathbf{x}^\mu) \mid \mathbf{x}^\mu \in A^n, \text{ donde } \mu = 1, 2, \dots, p\}$$

3.2 Diseño de Memoria Asociativa

La diferencia fundamental entre las memorias asociativas clásicas (*Lernmatrix*, *Correlograph*, *Linear Associator* y Memoria Asociativa Hopfield) y las memorias asociativas morfológicas radica en los fundamentos operacionales de éstas últimas, que son las operaciones morfológicas de *dilatación* y *erosión*; el nombre de las memorias asociativas morfológicas está inspirado precisamente en estas dos operaciones básicas. Estas memorias rompieron con el esquema utilizado a través de los años en los modelos de memorias asociativas clásicas, que utilizan operaciones convencionales entre vectores y matrices para la fase de aprendizaje y suma de productos para la recuperación de patrones. Las memorias asociativas morfológicas cambian los productos por sumas y las sumas por máximos o mínimos en ambas fases, tanto de aprendizaje como de recuperación [27].

Fase de Aprendizaje:

Sea el conjunto fundamental: $\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \dots, p\}$ con $\mathbf{x} \in \mathbf{R}^n$ y $\mathbf{y} \in \mathbf{R}^m$. La asociación de cada pareja se realiza de la manera siguiente:

$$\mathbf{y}^\mu \otimes (-\mathbf{x}^\mu)^t = \begin{bmatrix} y_1^\mu - x_1^\mu & \dots & y_1^\mu - x_n^\mu \\ \vdots & \ddots & \vdots \\ y_m^\mu - x_1^\mu & \dots & y_m^\mu - x_n^\mu \end{bmatrix}$$

Las memorias Máximas y mínimas se obtienen aplicando las operaciones morfológicas de erosión.

$$M = \bigvee_{K=1}^P (y^k \otimes (-x^k)^t)$$

$$W = \bigwedge_{K=1}^P (y^k \otimes (-x^k)^t)$$

Fase de recuperación:

Se aplica la operación morfológica de dilatación.

$$M\Delta x = \begin{bmatrix} \bigvee_{i=1}^n (y_1 - x_i + x_i) \\ \vdots \\ \bigvee_{i=1}^n (y_m - x_i + x_i) \end{bmatrix} = y$$

$$W\nabla x = \begin{bmatrix} \bigvee_{i=1}^n (y_1 - x_i + x_i) \\ \vdots \\ \bigvee_{i=1}^n (y_m - x_i + x_i) \end{bmatrix} = y$$

Ejemplo

Sean las siguientes 3 (p=3) parejas de vectores (x,y) con dimensiones n=3 y m=4, respectivamente

$$x^1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow y^1 = \begin{pmatrix} 3 \\ 2 \\ -2 \\ 1 \end{pmatrix}$$

$$x^2 = \begin{pmatrix} 0 \\ -2 \\ -4 \end{pmatrix} \rightarrow y^2 = \begin{pmatrix} -1 \\ -1 \\ 0 \\ 2 \end{pmatrix}$$

$$x^3 = \begin{pmatrix} 0 \\ -3 \\ 0 \end{pmatrix} \rightarrow y^2 = \begin{pmatrix} 0 \\ -2 \\ 0 \\ 5 \end{pmatrix}$$

Fase de Aprendizaje: 1ª Asociación

$$y^1 \otimes (-x^1)^t = \begin{pmatrix} 3 \\ 2 \\ -2 \\ 1 \end{pmatrix} \otimes (0 \ 0 \ 0) = \begin{pmatrix} 3-0 & 3-0 & 3-0 \\ 2-0 & 2-0 & 2-0 \\ -2-0 & -2-0 & -2-0 \\ 1-0 & 1-0 & 1-0 \end{pmatrix}$$

$$y^1 \otimes (-x^1)^t = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 2 & 2 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix}$$

2ª Asociación

$$y^2 \otimes (-x^2)^t = \begin{pmatrix} -1 \\ -1 \\ 0 \\ 2 \end{pmatrix} \otimes -(0 \ -2 \ -4) = \begin{pmatrix} -1-0 & -1-(2) & -1-(-4) \\ -1-0 & -1-(2) & -1-(-4) \\ 0-0 & 0-(-2) & 0-(-4) \\ 2-0 & 2-(-2) & 2-(-4) \end{pmatrix}$$

$$y^2 \otimes (-x^2)^t = \begin{pmatrix} -1 & 1 & 3 \\ -1 & 1 & 3 \\ 0 & 2 & 4 \\ 2 & 4 & 6 \end{pmatrix}$$

3ª Asociación

$$y^3 \otimes (-x^3)^t = \begin{pmatrix} 0 \\ -2 \\ 0 \\ 5 \end{pmatrix} \otimes -\begin{pmatrix} 0 & -3 & 0 \end{pmatrix} = \begin{pmatrix} 0-0 & 0-(-3) & 0-0 \\ -2-0 & -2-(-3) & -2-0 \\ 0-0 & 0-(-3) & 0-0 \\ 5-0 & 5-(-3) & 5-0 \end{pmatrix}$$

$$y^3 \otimes (-x^3)^t = \begin{pmatrix} 0 & 3 & 0 \\ -2 & 1 & -2 \\ 0 & 3 & 0 \\ 5 & 8 & 5 \end{pmatrix}$$

Memoria MÁX:

$$M = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 2 & 2 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix} \vee \begin{pmatrix} -1 & 1 & 3 \\ -1 & 1 & 3 \\ 0 & 2 & 4 \\ 2 & 4 & 6 \end{pmatrix} \vee \begin{pmatrix} 0 & 3 & 0 \\ -2 & 1 & -2 \\ 0 & 3 & 0 \\ 5 & 8 & 5 \end{pmatrix}$$

$$M = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 2 & 3 \\ 0 & 3 & 4 \\ 5 & 8 & 6 \end{pmatrix}$$

Memoria MÍN:

$$W = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 2 & 2 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix} \wedge \begin{pmatrix} -1 & 1 & 3 \\ -1 & 1 & 3 \\ 0 & 2 & 4 \\ 2 & 4 & 6 \end{pmatrix} \wedge \begin{pmatrix} 0 & 3 & 0 \\ -2 & 1 & -2 \\ 0 & 3 & 0 \\ 5 & 8 & 5 \end{pmatrix}$$

$$W = \begin{pmatrix} -1 & 1 & 0 \\ -2 & 1 & 0 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix}$$

Fase de Recuperación

Presentamos el patrón x^1 a la memoria MAX

$$M\Delta x^1 = \begin{pmatrix} 3 & 3 & 3 \\ 2 & 2 & 3 \\ 0 & 3 & 4 \\ 5 & 8 & 6 \end{pmatrix} \Delta \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} (3+0) & \wedge(3+0) & \wedge(3+0) \\ (2+0) & \wedge(2+0) & \wedge(3+0) \\ (0+0) & \wedge(3+0) & \wedge(4+0) \\ (5+0) & \wedge(8+0) & \wedge(6+0) \end{pmatrix}$$

$$M\Delta x^1 = \begin{pmatrix} 3\wedge & 3\wedge & 3 \\ 2\wedge & 2\wedge & 3 \\ 0\wedge & 3\wedge & 4 \\ 5\wedge & 8\wedge & 6 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 0 \\ 5 \end{pmatrix} \neq y^1$$

Presentamos el patrón x^1 a la memoria MÍN

$$W\nabla x^1 = \begin{pmatrix} -1 & 1 & 0 \\ -2 & 1 & -2 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix} \nabla \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} (-1+0) & \vee(1+0) & \vee(0+0) \\ (-2+0) & \vee(1+0) & \vee(-2+0) \\ (-2+0) & \vee(-2+0) & \vee(-2+0) \\ (1+0) & \vee(1+0) & \vee(1+0) \end{pmatrix}$$

$$W\nabla x^1 = \begin{pmatrix} -1\vee & 1\vee & 0 \\ -2\vee & 1\vee & -2 \\ -2\vee & -2\vee & -2 \\ 1\vee & 1\vee & 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -2 \\ 1 \end{pmatrix} \neq y^1$$

Presentamos el patrón x^2 y x^3 a la memoria MÍN

$$W\nabla x^2 = \begin{pmatrix} -1 & 1 & 0 \\ -2 & 1 & -2 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix} \nabla \begin{pmatrix} 0 \\ -2 \\ -4 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -2 \\ 1 \end{pmatrix} \neq y^2$$

$$W\nabla x^3 = \begin{pmatrix} -1 & 1 & 0 \\ -2 & 1 & -2 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{pmatrix} \nabla \begin{pmatrix} 0 \\ -3 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -2 \\ -2 \\ 1 \end{pmatrix} \neq y^3$$

3.3 Selección de Rasgos

La selección de rasgos se ha vuelto el foco de atención de muchos trabajos de investigación, principalmente en áreas del conocimiento humano donde los conjuntos de datos (entrenamiento/prueba) consisten en cientos o miles de rasgos. Algunos ejemplos clásicos de su aplicación son: la selección de

características en microarreglos genéticos, la predicción de ocurrencia de enfermedades en el ser humano.

Diversos son los aspectos que deben ser observados para llevar a cabo procesos de aprendizaje automático (Machine Learning); sin embargo, la discriminación de información redundante y la preservación de información relevante son piezas clave para lograr la reducción dimensional de los datos. La mayoría de los autores coinciden en que la selección de rasgos (Feature Selection) puede ser dividida en dos grandes tareas: decidir que rasgos son los que mejor describen el contexto y seleccionar cual es la mejor combinación de éstos que mejore la precisión predictiva [33]; sin embargo, las discrepancias ocurren por los criterios usados para definir tanto la relevancia como la redundancia en los datos.

3.3.1 Algunas definiciones

En 1991 varios autores coincidieron en que la relevancia de la información debe ser considerada booleana, libre de ruido y definida en términos de distribuciones de probabilidad, que permitan obtener una estimación confiable que sugiera la eliminación de rasgos claramente identificables [49].

Definición 1

Un rasgo X_i se dice que es relevante en un concepto C , si X_i aparece en toda la representación del concepto C y varía sistemáticamente para cada categoría o clase.

Definición 2

Un rasgo X_i se dice que es relevante si existe un x_i y y para el cual $p(X_i = x_i) > 0$, tal que

$$p(Y = y | X_i = x_i) \neq p(Y = y).$$

Bajo esta definición, X_i es relevante, si al conocer su valor, produce un cambio en la asignación de la etiqueta de clase. En otras palabras, si Y es condicionalmente dependiente de X_i ; sin embargo, cuando se tienen datos sin etiquetar, esta definición es incapaz de identificar información relevante, ya que, la asignación de la etiqueta de clase bajo estas condiciones es equiprobable para toda instancia considerada en la fase de aprendizaje. Cabe mencionar que esta situación es meramente hipotética, ya que, en procesos relacionados con el aprendizaje automático no se utilizan instancias sin etiquetar para la fase de entrenamiento.

Definición 3

Sea, $S_i = \{X_1, \dots, X_{i-1}, \dots, X_{i+1}, \dots, X_m\}$ el conjunto de todos los rasgos sin contar a X_i . Sea, s_i el valor asignado para todos los rasgos en S_i .

Un rasgo X_i se dice que es relevante, si existe un x_i , y y para el cual $p(X_i = x_i | Y = y, S_i = s_i) > 0$, tal que

$$p(Y = y, S_i = s_i | X_i = x_i) \neq p(Y = y, S_i = s_i).$$

Bajo esta definición, X_i es relevante, cuando la información derivada del conocimiento de X_i puede cambiar la asignación de la etiqueta de clase.

Definición 4

Un rasgo X_i se dice que es relevante, si existe un x_i , y y para el cual $p(X_i = x_i | Y = y, S_i = s_i) > 0$, tal que

$$p(Y = y | X_i = x_i, S_i = s_i) \neq p(Y = y | S_i = s_i).$$

Bajo esta definición, X_i es relevante, si la probabilidad de que sea asignada una determinada etiqueta de clase puede cambiar cuando es eliminada la información que proporciona el valor del rasgo X_i .

Sería deseable que, con las definiciones antes mencionadas, se pudiera identificar información relevante en un conjunto de instancias determinadas; sin embargo, en situaciones donde existe alta correlación en los datos que describen un problema, las definiciones anteriores no son suficientes, consecuentemente, surge la necesidad de definir dos grados de relevancia: fuerte y débil.

3.3.1.1 Relevancia fuerte y débil

Existen dos grados de relevancia en la información: fuerte y débil. Para poder diferenciar ambos grados, es necesario considerar que, la relevancia debe ser definida en términos de un clasificador Bayesiano, basado en el principio matemático de que la mayoría de los sucesos son dependientes y que la probabilidad de que un suceso ocurra en el futuro puede ser deducida de las ocurrencias anteriores de dicho evento [50].

Un rasgo X_i es fuertemente relevante si al eliminar la información que aporta éste, se observa un decremento en la precisión predictiva del modelo en cuestión, es decir, la simple remoción de la información que proporciona este rasgo, deteriora el desempeño del proceso de clasificación. En términos de funciones de probabilidad condicional para la asignación de la etiqueta de clase, se puede

enunciar que un rasgo X_i es fuertemente relevante si cumple con la siguiente expresión.

$$p(Y = y | X_i = x_i, S_i = s_i) \neq p(Y = y | S_i = s_i).$$

Un rasgo X_i es débilmente relevante, si no es fuertemente relevante y existe un subconjunto de rasgos, S , para el cual el desempeño del modelo en cuestión es inferior que el alcanzado por el subconjunto de rasgos dado por $S \cup \{X_i\}$; en otras palabras, la precisión predictiva alcanzada por el subconjunto de rasgos S , es menor que la precisión predictiva alcanzada cuando se considera la información contenida en el rasgo X_i . En términos de funciones de probabilidad condicional para la asignación de la etiqueta de clase, se puede enunciar que un rasgo X_i es débilmente relevante si existe un subconjunto de características $S' \subset S$ para el cual existe alguna x_i , y y s'_i con $p(X_i = x_i | S' = s'_i) > 0$ y cumple con la siguiente expresión.

$$p(Y = y | X_i = x_i, S'_i = s'_i) \neq p(Y = y | S'_i = s'_i).$$

Dadas ambas definiciones para los dos grados de relevancia posibles, se dice que un rasgo X_i es relevante, ya sea fuerte o débilmente relevante; de lo contrario se dice que X_i es irrelevante.

3.3.1.2 Optimalidad del subconjunto de rasgos

En la búsqueda de subconjuntos de rasgos que permitan representar la información de manera reducida, el algoritmo de inducción, típicamente está dado por un clasificador que usa todos los rasgos fuertemente relevantes y posiblemente algunos de relevancia débil, no obstante que, la selección de rasgos propuesta, generalmente es un subconjunto de carácter subóptimo, ocasionado por el desconocimiento de la distribución de probabilidad sobre los datos nunca antes vistos (conjunto de prueba).

Otro aspecto delicado es que, con las definiciones anteriormente propuestas, no se puede asegurar que un rasgo fuertemente relevante aparezca en el subconjunto de rasgos seleccionado, así como tampoco se puede asegurar que algunos rasgos débilmente relevantes no aparezcan en esta selección.

Claramente estas limitaciones apuntan a la existencia de algún tipo de procedimiento encargado de esta tarea nada trivial. Desde una perspectiva subyacente a los requerimientos de cómputo necesarios para conducir esta búsqueda, sería deseable que, el algoritmo de inducción se comportara

adecuadamente aún en presencia de información irrelevante, no obstante que, la complejidad computacional, generalmente aumenta de manera exponencial en presencia de instancias con estas características [35].

Algunos beneficios derivados de la selección de rasgos son:

- Facilitar la visualización y comprensión de los datos
- Reducción en los requerimientos de almacenamiento
- Reducción en los tiempos de entrenamiento y recuperación
- Incrementos en la precisión predictiva

3.4 Codificación Jonson-Möbius Modificado

Para generar el código Jonhson-Möbius modificado [31], se elige el número mayor del grupo de números a codificar, y se tomará como el tamaño para cada vector resultante, a manera de que si el número mayor del grupo a codificar el 10, cada vector resultante tendrá 10 elementos. Posteriormente, cada que se toma un número a codificar, debe restarse el número mayor. Por ejemplo, en este caso, el número mayor es 10, si se toma el 6 para codificar, se restaría $10-6=4$. Y el resultado indica la cantidad de ceros que se deben de poner en el vector resultante, el resto deberá ser llenado con unos. Para este caso, el vector resultante sería: [0 0 0 0 1 1 1 1 1]. Nótese que si fuese necesario codificar el número máximo, el vector resultante no tendría ceros, por ejemplo, si se codificara el 10, el resultado sería: $10-10=0$; [1 1 1 1 1 1 1 1 1 1].

Si se observa, la cantidad de unos que contendrá el vector resultante, será siempre la representada por el numero a codificar, por ejemplo, si se codificara el 3, el vector resultante contendría 3 unos.

Algoritmo

Sea R un conjunto de números reales, $R=\{r_1, r_2, r_3, \dots, r_i, \dots, r_n\}$.

1.- si uno de los números del conjunto (digamos r_i) es negativo, se crea un nuevo conjunto transformado a través de la operación “restar r_i a cada uno de los números”. Si hubiese más de un número negativo, se elige el menor. Nótese que r_i será igual a cero después de la transformación.

2.-Escoger un número fijo de decimales y truncar cada uno de los números del conjunto transformado.

3.-Realizar un escalamiento de 10^d en el conjunto del paso 2, para que finalmente nos quede un conjunto de n enteros no negativos. Que será representado de la siguiente manera: $\{e_1, e_2, e_3, \dots, e_m, \dots, e_n\}$, donde e_m es el número mayor.

4.-El código Jonhson-Möbius modificado para cada $j=1,2,2,\dots,n$ se obtiene al generar (e_m-e_j) ceros concatenados por la derecha con e_j unos.

Ejemplo

Encontrar el código Jonhson-Möbius modificado de $\{0.2, 1.25, -0.3, 2.147\}$.

1.-Existe un número negativo entre los números a convertir (-0.3), es por eso que debe de restarse a todos los números, de tal manera que el conjunto con números negativos generados es: $\{0.5, 1.5, 0, 2.447\}$.

2.-Se escoge el número fijo $d=1$, con el cual el número de decimales se reducirá a unos para obtener: $\{0.5, 1, 5, 0, 2.4\}$.

3.- Se realiza un escalamiento 10^d para obtener un nuevo conjunto: $\{5, 15, 0, 24\}$ donde $e_m=24$ es el número mayor.

4.-Para cada número e_i del conjunto e , se generan e_m-e_i ceros concatenados con e_i unos, los resultados se muestran a continuación.

```

5-[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1]
15-[0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]
0-[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
24-[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]

```

3.5 Base de datos de Hipoacusia

El nombre de la base de datos, *base de datos hipoacusia*, la cual fue diseñada en una Escuela de Educación Especial en Villahermosa, Tabasco.

La base de datos se compone de una serie de mediciones Biomedicas de la voz realizada a 255 personas, 191 de ellas con problemas de Hipoacusia. Cada columna en la tabla es una medición de voz particular, y cada fila corresponde a una de las 255 voces grabadas. El objetivo principal de la obtención de los datos, es discriminar registros de personas sanas de aquellas con hipoacusia. En la columna 9 se incluye FALSO si la persona no padece de hipoacusia y VERDADERO si tiene hipoacusia.

Los rasgos con los que cuenta la base de datos hipoacusia se muestran a continuación:

Id_arch_wav – número de la grabación

F0 media – Frecuencia fundamental media

F0 max- Frecuencia fundamental máxima

F0 min- Frecuencia fundamental mínima

Desv. Std.- Desviación estándar

Coef var- Coeficiente de variabilidad

Jitter medio –Jitter medio

Jitter factor – Jitter factor

Jitter ratio – Jitter ratio

Prom.Rel.Pert.- Promedio

Hipoacusico – Indica la clase a la que pertenece cada registro (VERDADERO o FALO)

CAPITULO 4.

DESARROLLO

4.1 DISEÑO DEL CLASIFICADOR

Una vez aprendida la teoría necesaria sobre memorias asociativas morfológicas, es posible implementar el modelo que nos permitirá hacer la clasificación. Para ello, es necesario plantear la estructura del clasificador en base al modelo asociativo a usar.

4.2 SELECCIÓN DE RASGOS DE LA BASE DE DATOS

El primer paso es hacer una selección de rasgos (Feature Selection) a la base de datos hipoacusia, para seleccionar la mejor combinación de elementos de modo que mejore la precisión predictiva.

Se ha utilizado Microsoft Excel, que ha servido para ordenar la base de datos. Se utilizó una columna para cada campo de la base de datos y una fila para cada registro.

Para las pruebas de efectividad se ha modificado la base de datos primero quitando el campo *Jitter factor*, después el campo *Prom.Rel.Pert* y por último los dos campos (*Jitter ratio* y *Prom.Rel.Pert*).

La base de datos originalmente cuenta con 255 registros quedando con 252, debido a que los datos de estos campos se ven desproporcionados numéricamente en comparación con los otros campos.

Resultando 4 bases de datos para pruebas de efectividad con 252 registros:

1. Base de Datos Completa.
2. Base de Datos sin el campo Jitter factor.
3. Base de Datos sin el campo Prom.Rel.Pert.
4. Base de Datos sin ambos campos (Jitter factor y Prom.Rel.Pert).

4.3 Generación de Memorias Asociativas Morfológicas

Algoritmo 1

Una memoria asociativa morfológica, requiere de patrones de entrada x asociados a patrones y , los cuáles serán nuestro conjunto de patrones de entrenamiento. En este caso, cada registro de la base de datos es un patrón x el cual forman parte del conjunto de patrones de entrenamiento.

Ejemplo para el registro 1 de la Base de Datos Completa:

$$x = [9,1] = \begin{bmatrix} 222.757 \\ 227.319 \\ 112.5 \\ 10.7776 \\ 4.8383 \\ 1.53702 \\ 0.689999993 \\ 144.92755 \\ -0.211905 \end{bmatrix}$$

Para el caso de los patrones y , es necesario considerar lo siguiente, puesto que la base de datos cuenta con 252 registros, se crean vectores de 252 filas por una

columna; los cuales estarán formados por un elemento con valor igual a 1000 y el resto de ellos serán cero.

$$y = [252,1] = \begin{bmatrix} 1000 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Estas son condiciones suficientes para implementar la fase de aprendizaje de la memoria asociativa morfológica.

Fase de aprendizaje

Una vez generados los conjuntos de patrones x y y procedemos a realizar las asociaciones.

Generación de Asociaciones

Se toma en cuenta un primer conjunto fundamental $(\mathbf{x}^\mu, \mathbf{y}^\mu)$; donde $\mu = 1, 2, \dots, 252$; para obtener una primera asociación, se consideró el conjunto fundamental $(\mathbf{x}^1, \mathbf{y}^1)$.

$$x^1 = \begin{pmatrix} 222.757 \\ 227.319 \\ 112.5 \\ 10.7776 \\ 4.8383 \\ 1.53702 \\ 0.689999993 \\ 144.92755 \\ -0.211905 \end{pmatrix} \rightarrow y^1 = \begin{pmatrix} 1000 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Ahora se toma el patrón x^1 , se multiplica por -1 y posteriormente se transpone.

$$\begin{aligned} & (-x^1)^t = \\ & = (-222.757, -227.319, -112.5, -10.7776, -4.8383, -1.53702, -0.689999993, -144.92755, +0.211905) \end{aligned}$$

Para terminar de realizar la primera asociación se realiza la operación:

$$y^\mu \otimes (-x^\mu)^t = \begin{bmatrix} 1000 - 222.757 & \dots & 1000 + 0.211905 \\ \vdots & \ddots & \vdots \\ 0 - 222.757 & \dots & 0 + 0.211905 \end{bmatrix}$$

Este procedimiento se le realiza al resto de los conjuntos fundamentales, obteniendo al final 252 (o n número de conjuntos fundamentales) asociaciones.

Generación de Memoria Max Min

Memoria MÁX:

Se genera una matriz ($m \times n$); m es el número de patrones x , n es el número de elementos del patrón x .

$$M = \begin{pmatrix} 777.757 & \dots & 100.211 \\ \vdots & \ddots & \vdots \\ -222.757 & \dots & 0.211 \end{pmatrix} \vee \begin{pmatrix} 877.757 & \dots & 100.111 \\ \vdots & \ddots & \vdots \\ -122.757 & \dots & 0.111 \end{pmatrix} \vee \dots \vee \begin{pmatrix} 677.757 & \dots & 100.511 \\ \vdots & \ddots & \vdots \\ -122.957 & \dots & 0.311 \end{pmatrix}$$

$$M = \begin{pmatrix} 877.757 & \dots & 100.511 \\ \vdots & \ddots & \vdots \\ -122.957 & \dots & 0.311 \end{pmatrix}$$

Esta matriz está formada por los valores máximos obtenidos de la comparación elemento a elemento de las matrices resultantes en las asociaciones.

Memoria MÍN:

Se genera una matriz Min ($m \times n$); m es el número de patrones x , n es el número de elementos del patrón x .

$$W = \begin{pmatrix} 777.757 & \dots & 100.211 \\ \vdots & \ddots & \vdots \\ -222.757 & \dots & 0.211 \end{pmatrix} \wedge \begin{pmatrix} 877.757 & \dots & 100.111 \\ \vdots & \ddots & \vdots \\ -122.757 & \dots & 0.111 \end{pmatrix} \wedge \dots \wedge \begin{pmatrix} 677.757 & \dots & 100.511 \\ \vdots & \ddots & \vdots \\ -122.957 & \dots & 0.311 \end{pmatrix}$$

$$W = \begin{pmatrix} 677.757 & \dots & 100.111 \\ \vdots & \ddots & \vdots \\ -222.757 & \dots & 0.111 \end{pmatrix}$$

Esta matriz está formada por los valores mínimos obtenidos de la comparación elemento a elemento de las matrices resultantes en las asociaciones.

Y de esta forma se obtienen las dos matrices morfológicas Max y Min.

Fase de recuperación.

A la memoria máxima y mínima se le presentan uno a uno los patrones x del conjunto de pruebas y se les aplica la operación morfológica de dilatación. Obteniendo así un patrón y^* recuperado.

$$M\Delta x^1 = \begin{pmatrix} 877.757 & \dots & 100.511 \\ \vdots & \ddots & \vdots \\ -122.957 & \dots & 0.311 \end{pmatrix} \Delta \begin{pmatrix} 222.757 \\ 227.319 \\ 112.5 \\ 10.7776 \\ 4.8383 \\ 1.53702 \\ 0.689999993 \\ 144.92755 \\ -0.211905 \end{pmatrix}$$

$$= \begin{pmatrix} (877.757 + 222.757) \wedge \dots \wedge (877.757 + (-0.211905)) \\ \vdots \wedge \ddots \wedge \vdots \\ (-122.957 + 222.757) \wedge \dots \wedge (0.311 + (-0.211905)) \end{pmatrix}$$

$$M\Delta x^1 = \begin{pmatrix} 1100.514 \wedge \dots \wedge 877.545095 \\ \vdots \wedge \ddots \wedge \vdots \\ 99.8 \wedge \dots \wedge 0.099095 \end{pmatrix} = \begin{pmatrix} 877.545095 \\ \vdots \\ 0.099095 \end{pmatrix} \neq y^1$$

Presentamos el patrón x^1 a la memoria MÍN

$$W\nabla x^1 = \begin{pmatrix} 877.757 & \dots & 100.511 \\ \vdots & \ddots & \vdots \\ -122.957 & \dots & 0.311 \end{pmatrix} \nabla \begin{pmatrix} 222.757 \\ 227.319 \\ 112.5 \\ 10.7776 \\ 4.8383 \\ 1.53702 \\ 0.689999993 \\ 144.92755 \\ -0.211905 \end{pmatrix}$$

$$= \begin{pmatrix} (877.757 + 222.757)v & \dots v & (877.757 + (-0.211905)) \\ \vdots v & \ddots v & \vdots v \\ (-122.957 + 222.757)v & \dots v & (0.311 + (-0.211905)) \end{pmatrix}$$

$$W\nabla x^1 = \begin{pmatrix} 1100.514 v & \dots v & 877.545095 \\ \vdots v & \ddots v & \vdots v \\ 99.8 v & \dots v & 0.099095 \end{pmatrix} = \begin{pmatrix} 1100.514 \\ \vdots \\ 99.8 \end{pmatrix} \neq y^1$$

Para el caso de los patrones y^* recuperados si son iguales a los patrones y asociados, entonces y^* es un patrón recuperado y en este caso es también considerado un registro clasificado correctamente.

4.4 TRANSFORMACION DE LA BASE DE DATOS

Algoritmo 2

En el segundo algoritmo se ha transformado la base de datos utilizando el método de Johnson-Môbuis modificado. Para esto se realizó lo siguiente:

- 1- A los registros del campo *Prom.Rel.Pert* de la base de datos hipoacusia se obtuvo su valor absoluto, de tal manera que no existen números negativos dentro de la base de datos.
- 2- Se realiza una multiplicación por 100 o por 1000, dependiendo el caso a todos los campos de la base de datos.
- 3- Se redondean y se truncan todos los valores de la base de datos para que sean enteros.
- 4- Se busca el número mayor de toda la base de datos hipoacusia.
- 5- Se convertirá cada campo de cada registro en un vector que tendrá como tamaño el valor obtenido en el punto 4. Por ejemplo, si el valor mayor es = 20 se tendrá como resultado:

[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]

- 6- El vector creado en el punto anterior estará formado únicamente por unos y ceros, siendo el número de unos igual al valor del campo y el resto de los elementos serán ceros. Siguiendo con el ejemplo del punto 5, si alguno de

los campos de un registro de la base de datos tuviese como valor un numero 5, el vector resultante sería:

$$5 = [1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0].$$

- 7- Se realizan agrupaciones de 10 elementos (cada una) en los vectores resultantes del punto 6, reduciendo así el tamaño de cada registro para que el tiempo de procesamiento sea menor; para finalmente realizar una conversión de binario a decimal a cada una de las agrupaciones. Siguiendo con el ejemplo planteado se muestra a continuación el proceso mencionado.

$$5 = [(1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0) (0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0)].$$

Paquete 1 → (1 1 1 1 1 0 0 0 0 0) = 992

&

Paquete 2 → (0 0 0 0 0 0 0 0 0 0) = 0

$$5 = [992 , 0].$$

- 8- Se realizan los pasos 6 y 7 para cada registro, de esta forma se tiene la base de datos codificada y lista para continuar con el proceso de entrenamiento.
- 9- Ahora se realiza el procedimiento del algoritmo 1 descrito anteriormente a la base de datos codificada.

4.5 Programación C#

Existen muchas maneras de implementar el método del modelo asociativo, incluso es posible hacerlo manualmente. Sin embargo hacer uso de una computadora trae ventajas, una de ellas es la precisión en el manejo de grandes cantidades de información. Por esta razón, se ha decidido diseñar los clasificadores en una computadora.

Se ha elegido C# debido a que tiene un entorno amigable para el usuario, es gratuito en su versión Express, contiene muchas herramientas que facilitan la programación.

CAPITULO 5.

RESULTADOS

5.1 Pruebas de efectividad

Para poner a prueba la efectividad del clasificador de Hipoacusia, se han realizado tres diferentes pruebas: Leave One Out, Hold Out y K-Fold Cross Validation.

5.1.1 Leave One Out

Esta prueba consiste en tomar un registro de la base de datos como conjunto de prueba y los registros restantes serán el conjunto de entrenamiento. Este procedimiento se realizara un número r de veces, siendo $r = \text{número de registros de la base de datos}$, esto debido a que todos los registros serán tomados individualmente como conjunto de prueba y así asegurar que el registro del conjunto de prueba no esté incluido en los registros del conjunto de entrenamiento.

Posteriormente, en la etapa de recuperación, se podrá determinar a qué clase pertenece y de esta manera ser clasificarlo como verdadero o falso (hipoacusico, o no hipoacusico). Este proceso se lleva acabo con cada uno de los registros de la base de datos y sus respectivos conjuntos de entrenamiento. Después se compara cada registro clasificado con el registro original, se determina el número de registros perdidos y recuperados en la clasificación. Con estos datos, se obtiene un porcentaje de efectividad.

Algoritmo

- 1- Tomar un registro n de la base de datos
- 2-Elegir n como conjunto de prueba y el resto como aprendizaje
- 3- Clasificar n
- 4- Comparar registró n clasificada, con el registro n de la base de datos.
- 5- Conteo de patrones Recuperados y perdidos.

Algoritmo 1

Con el primer algoritmo, se realizaron 10 pruebas de efectividad para cada una de las bases.

Tabla 1.1 Porcentaje de efectividad de Leave One Out Base de Datos Completa

Base de Datos Completa			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	228	24	90.47%
2	228	24	90.47%
3	228	24	90.47%
4	228	24	90.47%
5	228	24	90.47%
6	228	24	90.47%
7	228	24	90.47%
8	228	24	90.47%
9	228	24	90.47%
10	228	24	90.47%
Promedio			90.47%

Tabla 1.2 Porcentaje de efectividad de Leave One Out Base de Datos sin Jitter Factor

Base de Datos sin Jitter Factor			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	227	25	90.07%
2	227	25	90.07%
3	227	25	90.07%
4	227	25	90.07%
5	227	25	90.07%
6	227	25	90.07%
7	227	25	90.07%
8	227	25	90.07%
9	227	25	90.07%
10	227	25	90.07%
Promedio			90.07%

Tabla 1.3 Porcentaje de efectividad de Leave One Out Base de Datos sin PromRelP

Base de Datos Sin PromRelP			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	230	22	91.26%
2	230	22	91.26%
3	230	22	91.26%
4	230	22	91.26%
5	230	22	91.26%
6	230	22	91.26%
7	230	22	91.26%
8	230	22	91.26%
9	230	22	91.26%
10	230	22	91.26%
Promedio			91.26%

Tabla 1.4 Porcentaje de efectividad de Leave One Out Base de Datos sin Ambos Campos

Base de Datos Sin Ambos Campos			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	229	23	90.87%
2	229	23	90.87%
3	229	23	90.87%
4	229	23	90.87%
5	229	23	90.87%
6	229	23	90.87%
7	229	23	90.87%
8	229	23	90.87%
9	229	23	90.87%
10	229	23	90.87%
Promedio			90.87%

Algoritmo 2

Con el segundo algoritmo, se realizaron 10 pruebas de efectividad para cada una de las bases.

Tabla 1.5 Porcentaje de efectividad de Leave One Out Base de Datos Completa

Base de Datos Completa			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	241	11	95.63%
2	241	11	95.63%
3	241	11	95.63%
4	241	11	95.63%
5	241	11	95.63%
6	241	11	95.63%
7	241	11	95.63%
8	241	11	95.63%
9	241	11	95.63%
10	241	11	95.63%
Promedio			95.63%

Tabla 1.6 Porcentaje de efectividad de Leave One Out Base de Datos sin Jitter Factor

Base de Datos sin Jitter Factor			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	240	12	93.65%
2	240	12	93.65%
3	240	12	93.65%
4	240	12	93.65%
5	240	12	93.65%
6	240	12	93.65%
7	240	12	93.65%
8	240	12	93.65%
9	240	12	93.65%
10	240	12	93.65%
Promedio			93.65%

Tabla 1.7 Porcentaje de efectividad de Leave One Out Base de Datos sin PromRelP

Base de Datos Sin PromRelP			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	244	8	96.82%
2	244	8	96.82%
3	244	8	96.82%
4	244	8	96.82%
5	244	8	96.82%
6	244	8	96.82%
7	244	8	96.82%
8	244	8	96.82%
9	244	8	96.82%
10	244	8	96.82%
Promedio			96.82%

Tabla 1.8 Porcentaje de efectividad de Leave One Out Base de Datos sin Ambos Campos

Base de Datos Sin Ambos Campos			
No. de Prueba	Recuperados	Perdidos	% Efectividad
1	242	10	96.03%
2	242	10	96.03%
3	242	10	96.03%
4	242	10	96.03%
5	242	10	96.03%
6	242	10	96.03%
7	242	10	96.03%
8	242	10	96.03%
9	242	10	96.03%
10	242	10	96.03%
Promedio			96.03%

5.1.2 Hold Out

La segunda prueba de efectividad que se aplicó al clasificador de Hipoacusia fue Hold Out, que consiste en tomar un porcentaje de la base de datos como conjunto de prueba, y dejar el resto como conjunto de aprendizaje. Este porcentaje ira del 10% al 90% en incrementos de 10% para el caso de los conjuntos de entrenamiento e inversamente proporcional para el conjunto de pruebas.

Ya que la forma de elegir los registros que formaran los conjuntos de entrenamiento será de forma aleatoria, para esta prueba fue necesario separar la base de datos en dos clases (hipoacusico y no hipoacusico) para asegurarnos que los conjuntos de entrenamiento y

prueba tendrán el porcentaje de registros real de cada clase. Una vez terminados los entrenamientos y obtenidas las memorias, procedemos con la etapa de recuperación y de esta forma poder clasificar los registros del conjunto de prueba y obtener la efectividad de los registros clasificados.

Algoritmo

- 1- Tomar un porcentaje de la base de datos como conjunto de aprendizaje.
- 2- El resto de la base de datos se toma como conjunto de prueba.
- 3- Se clasifica los registros de prueba.
- 4- Comparar cada registro clasificado con su correspondiente en la base de datos.
- 5- Se hace un conteo de patrones recuperados y perdidos.
- 6- El algoritmo se repite del 10% al 90% como conjunto de prueba.

Se hicieron pruebas con 9 diferentes porcentajes (10%, 20%.....90%) como conjunto de aprendizaje y cada una se repitió 10 veces.

Los registros de cada conjunto (Aprendizaje y Prueba) se toman aleatoriamente, es por eso que se hacen 10 veces cada prueba, para obtener un promedio y el resultado sea más preciso.

Los resultados se muestran a continuación de los dos clasificadores.

Algoritmo 1

Tabla 2.1 Porcentaje de efectividad de Hold Out Base de Datos Completa A-I

Base de Datos Completa A										
Aprendizaje: 10%						Prueba: 90%				
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	188	193	184	199	184	200	193	188	189	196
Perdidos	40	35	44	29	44	28	35	40	39	32
Efectividad (%)	82.45	84.65	80.70	87.28	80.70	87.72	84.65	82.46	82.89	85.96
Efectividad: 83.94%										

Base de Datos Completa B										
Aprendizaje: 20%					Prueba: 80%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	192	197	183	187	175	183	187	185	202	175
Perdidos	20	15	19	15	27	19	15	17	10	27
Efectividad (%)	90.57	92.92	90.59	92.57	86.63	90.59	92.57	91.58	95.28	86.63
Efectividad: 90.99%										

Base de Datos Completa C										
Aprendizaje: 30%					Prueba: 70%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	156	157	153	152	158	152	157	148	159	151
Perdidos	22	21	25	26	20	26	21	30	19	27
Efectividad (%)	87.64	88.20	85.96	85.39	88.76	85.39	88.20	83.15	89.33	84.83
Efectividad: 86.68%										

Base de Datos Completa D										
Aprendizaje: 40%					Prueba: 60%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	135	132	136	138	132	133	132	131	135	134
Perdidos	17	20	16	14	20	19	20	21	17	18
Efectividad (%)	88.81	86.84	89.47	90.79	86.84	87.50	86.84	86.18	88.82	88.16
Efectividad: 88.02%										

Base de Datos Completa E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	111	107	110	111	109	110	108	114	114	110
Perdidos	15	19	16	15	17	16	18	12	12	16
Efectividad (%)	88.09	84.92	87.30	88.10	86.51	87.30	85.71	90.48	90.48	87.30
Efectividad: 87.61%										

Base de Datos Completa F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	86	90	90	95	93	95	89	89	87	91
Perdidos	16	10	12	7	9	7	13	13	15	11
Efectividad (%)	84.31	90.00	88.24	93.14	91.18	93.14	87.25	87.25	85.29	89.22
Efectividad: 88.90%										

Base de Datos Completa G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	68	69	67	70	69	68	71	67	68	70
Perdidos	8	7	9	6	7	8	5	9	8	6
Efectividad (%)	89.47	90.79	88.16	92.11	90.79	89.47	93.42	88.16	89.47	92.11
Efectividad: 90.39%										

Base de Datos Completa H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	49	47	47	49	47	49	47	49	47	49
Perdidos	3	5	5	3	5	3	5	3	5	3
Efectividad (%)	94.23	90.38	90.38	94.23	90.38	94.23	90.38	94.23	90.38	94.23
Efectividad: 92.30%										

Base de Datos Completa I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	24	23	22	24	25	22	24	23	22	25
Perdidos	2	3	4	2	1	4	2	3	4	1
Efectividad (%)	92.31	88.46	84.62	92.31	96.15	84.62	92.31	88.46	84.62	96.15
Efectividad: 90.00%										

Tabla 2.2 Porcentaje de efectividad de Hold Out Base de Datos sin Jitter Factor A-I

Base de Datos sin Jitter Factor A										
Aprendizaje: 10%					Prueba: 90%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	198	201	181	193	198	189	193	192	198	181
Perdidos	30	27	21	15	30	39	15	36	30	21
Efectividad (%)	86.84	88.16	89.60	92.79	86.84	82.89	92.79	84.21	86.84	89.60
Efectividad: 88.05%										

Base de Datos sin Jitter Factor B										
Aprendizaje: 20%					Prueba: 80%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	182	181	177	180	184	180	182	177	181	180
Perdidos	20	21	25	22	18	22	20	25	21	22
Efectividad (%)	90.1	89.60	87.62	89.11	91.09	89.11	90.10	87.62	89.60	89.11
Efectividad: 89.30%										

Base de Datos sin Jitter Factor C										
Aprendizaje: 30%					Prueba: 70%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	156	152	157	158	160	157	156	157	156	150
Perdidos	22	26	21	20	18	21	22	21	22	28
Efectividad (%)	87.64	85.39	88.20	88.76	89.89	88.20	87.64	88.20	87.64	84.27
Efectividad: 87.58%										

Base de Datos sin Jitter Factor D										
Prueba: 40%					Prueba: 60%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	130	130	136	136	133	134	132	134	136	132
Perdidos	22	22	16	16	19	18	20	18	16	20
Efectividad (%)	85.52	85.53	89.47	89.47	87.50	88.16	86.84	88.16	89.47	86.84
Efectividad: 87.69%										

Base de Datos sin Jitter Factor E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	115	114	110	115	111	108	112	113	110	115
Perdidos	11	12	16	11	15	18	14	13	16	11
Efectividad (%)	91.27	90.48	87.30	91.27	88.10	85.71	88.89	89.68	87.30	91.27
Efectividad: 89.12%										

Base de Datos sin Jitter Factor F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	87	94	87	93	91	93	87	88	88	93
Perdidos	15	8	15	9	11	9	15	14	14	9
Efectividad (%)	85.29	92.16	85.29	91.18	89.22	91.18	85.29	86.27	86.27	91.18
Efectividad: 88.33%										

Base de Datos sin Jitter Factor G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	70	69	68	66	69	68	68	70	69	69
Perdidos	6	7	8	10	7	8	8	6	7	7
Efectividad (%)	92.10	90.79	89.47	86.84	90.79	89.47	89.47	92.11	90.79	90.79
Efectividad: 90.26%										

Base de Datos sin Jitter Factor H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	49	48	47	48	44	46	48	46	49	48
Perdidos	3	4	5	4	8	6	4	6	3	4
Efectividad (%)	94.23	92.31	90.38	92.31	84.62	88.46	92.31	88.46	94.23	92.31
Efectividad: 90.96%										

Base de Datos sin Jitter Factor I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	22	25	24	24	23	24	22	23	22	21
Perdidos	4	1	2	2	3	2	4	3	4	5
Efectividad (%)	84.61	96.15	92.31	92.31	88.46	92.31	84.62	88.46	84.62	80.77
Efectividad: 88.46%										

Tabla 2.3 Porcentaje de efectividad de Hold Out Base de Datos sin PromRelP A-I

Base de Datos sin PromRelP A										
Aprendizaje: 10%					Prueba: 90%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	203	192	189	203	202	175	194	203	197	188
Perdidos	25	36	39	25	26	27	34	25	31	40
Efectividad (%)	89.04	84.21	82.89	89.04	88.60	86.63	85.09	89.04	86.40	82.46
Efectividad: 86.34%										

Base de Datos sin PromRelP B										
Aprendizaje: 20%						Prueba: 80%				
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	177	181	180	182	180	181	177	181	179	181
Perdidos	25	21	22	20	22	21	25	21	23	21
Efectividad (%)	87.62	89.60	89.11	90.10	89.11	89.60	87.62	89.60	88.61	89.60
Efectividad: 89.06%										

Base de Datos sin PromRelP C										
Aprendizaje: 30%						Prueba: 70%				
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	154	157	152	157	155	154	155	154	159	155
Perdidos	24	21	26	21	23	24	23	24	19	23
Efectividad (%)	86.51	88.20	85.39	88.20	87.08	86.52	87.08	86.52	89.33	87.08
Efectividad: 87.19%										

Base de Datos sin PromRelP D										
Aprendizaje: 40%						Prueba: 60%				
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	134	137	134	136	134	137	136	134	132	136
Perdidos	18	15	18	16	18	15	16	18	20	16
Efectividad (%)	88.15	90.13	88.16	89.47	88.16	90.13	89.47	88.16	86.84	89.47
Efectividad: 88.82%										

Base de Datos sin PromReIP E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	115	111	116	112	113	112	113	112	110	116
Perdidos	11	15	10	14	13	14	13	14	16	10
Efectividad (%)	91.26	88.10	92.06	88.89	89.68	88.89	89.68	88.89	87.30	92.06
Efectividad: 89.68%										

Base de Datos sin PromReIP F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	96	91	87	92	90	93	89	87	84	91
Perdidos	6	11	15	10	12	9	13	15	18	11
Efectividad (%)	94.11	89.22	85.29	90.20	88.24	91.18	87.25	85.29	82.35	89.22
Efectividad: 88.23%										

Base de Datos sin PromReIP G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	67	69	68	69	68	67	69	68	67	69
Perdidos	9	7	8	7	8	9	7	8	9	7
Efectividad (%)	88.16	90.79	89.47	90.79	89.47	88.16	90.79	89.47	88.16	90.79
Efectividad: 89.61%										

Base de Datos sin PromRelP H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	49	45	48	49	46	49	46	45	49	46
Perdidos	3	7	4	3	6	3	6	7	3	6
Efectividad (%)	94.23	86.54	92.31	94.23	88.46	94.23	88.46	86.54	94.23	88.46
Efectividad: 90.77%										

Base de Datos sin PromRelP I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	24	22	26	23	24	23	24	23	22	22
Perdidos	2	4	0	3	2	3	2	3	4	4
Efectividad (%)	92.31	84.62	100.00	88.46	92.31	88.46	92.31	88.46	84.62	84.62
Efectividad: 89.62%										

Tabla 2.4 Porcentaje de efectividad de Hold Out Base de Datos sin Ambos Registros A-I

Base de Datos sin ambos Registros A										
Aprendizaje: 10%					Prueba: 90%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	205	201	198	205	201	199	201	205	201	197
Perdidos	23	27	30	23	27	29	27	23	27	31
Efectividad (%)	89.91	88.16	86.84	89.91	88.16	87.28	88.16	89.91	88.16	86.40
Efectividad: 88.28%										

Base de Datos sin ambos Registros B										
Aprendizaje: 20%					Prueba: 80%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	182	183	182	182	183	181	182	181	183	182
Perdidos	20	19	20	20	19	21	20	21	19	20
Efectividad (%)	90.1	90.59	90.10	90.10	90.59	89.60	90.10	89.60	90.59	90.10
Efectividad: 90.14%										

Base de Datos sin ambos Registros C										
Aprendizaje: 30%					Prueba: 70%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	163	160	161	157	163	161	160	157	160	157
Perdidos	15	18	17	21	15	17	18	21	18	21
Efectividad (%)	91.57	89.89	90.45	88.20	91.57	90.45	89.89	88.20	89.89	88.20
Efectividad: 89.83%										

Base de Datos sin ambos Registros D										
Aprendizaje: 40%					Prueba: 60%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	139	132	137	139	135	131	132	137	135	139
Perdidos	13	20	15	13	17	21	20	15	17	13
Efectividad (%)	91.45	86.84	90.13	91.45	88.82	86.18	86.84	90.13	88.82	91.45
Efectividad: 89.21%										

Base de Datos sin ambos Registros E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	116	113	112	113	116	113	113	112	111	116
Perdidos	10	13	14	13	10	13	13	14	15	10
Efectividad (%)	92.06	89.68	88.89	89.68	92.06	89.68	89.68	88.89	88.10	92.06
Efectividad: 90.07%										

Base de Datos sin ambos Registros F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	93	92	91	93	88	92	93	90	93	91
Perdidos	9	10	11	9	14	10	9	12	9	11
Efectividad (%)	91.18	90.20	89.22	91.18	86.27	90.20	91.18	88.24	91.18	89.22
Efectividad: 89.80%										

Base de Datos sin ambos Registros G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	69	67	68	69	70	71	70	67	71	70
Perdidos	7	9	8	7	6	5	6	9	5	6
Efectividad (%)	90.78	88.16	89.47	90.79	92.11	93.42	92.11	88.16	93.42	92.11
Efectividad: 91.05%										

Base de Datos sin ambos Registros H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	50	48	47	49	49	47	50	48	49	49
Perdidos	2	4	5	3	3	5	2	4	3	3
Efectividad (%)	96.15	92.31	90.38	94.23	94.23	90.38	96.15	92.31	94.23	94.23
Efectividad: 93.46%										

Base de Datos sin ambos Registros I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	24	23	24	22	23	24	25	22	23	25
Perdidos	2	3	2	4	3	2	1	4	3	1
Efectividad (%)	91.31	88.46	92.31	84.62	88.46	92.31	96.15	84.62	88.46	96.15
Efectividad: 90.28%										

Algoritmo 2

Tabla 2.5 Porcentaje de efectividad de Hold Out Base de Datos Completa A - I

Base de Datos Completa A										
Aprendizaje: 10%					Prueba: 90%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	215	203	210	212	213	208	203	208	210	215
Perdidos	13	25	18	16	15	20	25	20	18	13
Efectividad (%)	94.3	89.04	92.11	92.98	93.42	91.23	89.04	91.23	92.11	94.30
Efectividad: 91.97%										

Base de Datos Completa B										
Aprendizaje: 20%					Prueba: 80%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	202	206	204	197	200	204	203	205	202	204
Perdidos	10	6	8	15	12	8	9	17	10	8
Efectividad (%)	95.28	97.17	96.23	92.92	94.34	96.23	95.75	92.34	95.28	96.23
Efectividad: 95.17%										

Base de Datos Completa C										
Aprendizaje: 30%					Prueba: 70%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	165	167	159	160	158	161	160	165	159	167
Perdidos	13	11	19	18	20	17	18	13	19	11
Efectividad (%)	92.7	93.82	89.33	89.89	88.76	90.45	89.89	92.70	89.33	93.82
Efectividad: 91.06%										

Base de Datos Completa D										
Aprendizaje: 40%					Prueba: 60%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	145	141	140	142	139	140	141	144	142	145
Perdidos	7	11	12	10	13	12	11	8	10	7
Efectividad (%)	95.39	92.76	92.11	93.42	91.45	92.11	92.76	94.74	93.42	95.39
Efectividad: 93.35%										

Base de Datos Completa E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	117	120	113	115	117	116	112	122	114	118
Perdidos	9	6	13	11	9	10	14	4	12	8
Efectividad (%)	92.86	95.24	89.68	91.27	92.86	92.06	88.89	96.83	90.48	93.65
Efectividad: 92.38%										

Base de Datos Completa F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	95	94	93	95	93	95	98	95	97	94
Perdidos	7	8	9	7	9	7	4	7	5	8
Efectividad (%)	93.14	92.16	91.18	93.14	91.18	93.14	96.08	93.14	95.10	92.16
Efectividad: 93.03%										

Base de Datos Completa G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	73	71	73	75	72	70	71	74	73	71
Perdidos	6	5	3	1	4	6	5	2	3	5
Efectividad (%)	96.05	93.42	96.05	98.68	94.74	92.11	93.42	97.37	96.05	93.42
Efectividad: 95.13%										

Base de Datos Completa H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	49	52	50	51	52	49	50	51	52	50
Perdidos	3	0	2	1	0	3	2	1	0	2
Efectividad (%)	94.23	100.00	96.15	98.08	100.00	94.23	96.15	98.08	100.00	96.15
Efectividad: 97.30%										

Base de Datos Completa I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	26	24	24	24	25	24	24	25	26	25
Perdidos	0	2	2	2	1	2	2	1	0	1
Efectividad (%)	100	92.31	92.31	92.31	96.15	92.31	92.31	96.15	100.00	96.15
Efectividad: 95%										

Tabla 2.6 Porcentaje de efectividad de Hold Out Base de Datos sin Jitter Factor A - I

Base de Datos sin Jitter Factor A										
Aprendizaje: 10%					Prueba: 90%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	215	213	211	215	216	213	193	210	213	211
Perdidos	13	15	17	13	12	15	15	18	15	17
Efectividad (%)	94.3	93.42	92.54	94.30	94.74	93.42	92.79	92.11	93.42	92.54
Efectividad: 93.35%										

Base de Datos sin Jitter Factor B										
Aprendizaje: 20%					Prueba: 80%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	194	191	189	190	187	190	194	187	189	190
Perdidos	8	11	13	12	15	12	8	15	13	12
Efectividad (%)	96.04	94.55	93.56	94.06	92.57	94.06	96.04	92.57	93.56	94.06
Efectividad: 94.10%										

Base de Datos sin Jitter Factor C										
Aprendizaje: 30%					Prueba: 70%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	168	161	165	164	166	162	168	163	161	166
Perdidos	10	17	13	14	12	16	10	15	17	12
Efectividad (%)	94.38	90.45	92.70	92.13	93.26	91.01	94.38	91.57	90.45	93.26
Efectividad: 92.35%										

Base de Datos sin Jitter Factor D										
Prueba: 40%					Prueba: 60%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	144	140	139	141	140	145	144	140	141	142
Perdidos	8	12	13	11	12	7	8	12	11	10
Efectividad (%)	94.74	92.11	91.45	92.76	92.11	95.39	94.74	92.11	92.76	93.42
Efectividad: 93.15%										

Base de Datos sin Jitter Factor E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	119	118	122	115	120	115	123	119	122	115
Perdidos	7	8	4	11	6	11	3	7	4	11
Efectividad (%)	94.44	93.65	96.83	91.27	95.24	91.27	97.62	94.44	96.83	91.27
Efectividad: 94.28%										

Base de Datos sin Jitter Factor F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	95	94	93	95	98	93	96	95	98	95
Perdidos	7	8	9	7	4	9	6	7	4	7
Efectividad (%)	93.14	92.16	91.18	93.14	96.08	91.18	94.12	93.14	96.08	93.14
Efectividad: 93.33%										

Base de Datos sin Jitter Factor G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	73	75	72	73	71	73	72	70	73	71
Perdidos	3	1	4	3	5	3	4	6	3	5
Efectividad (%)	96.05	98.68	94.74	96.05	93.42	96.05	94.74	92.11	96.05	93.42
Efectividad: 95.13%										

Base de Datos sin Jitter Factor H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	50	49	51	48	50	49	51	48	50	51
Perdidos	2	3	1	4	2	3	1	4	2	1
Efectividad (%)	96.15	94.23	98.08	92.31	96.15	94.23	98.08	92.31	96.15	98.08
Efectividad: 95.57%										

Base de Datos sin Jitter Factor I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	24	25	24	24	25	24	25	24	23	25
Perdidos	2	1	2	2	1	2	1	2	3	1
Efectividad (%)	92.31	96.15	92.31	92.31	96.15	92.31	96.15	92.31	88.46	96.15
Efectividad: 93.46%										

Tabla 2.7 Porcentaje de efectividad de Hold Out Base de Datos sin PromRelP A - I

Base de Datos sin PromRelP A										
Aprendizaje: 10%					Prueba: 90%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	210	210	209	208	215	209	210	215	209	210
Perdidos	18	18	19	20	13	19	18	13	19	18
Efectividad (%)	92.11	92.11	91.67	91.23	94.30	91.67	92.11	94.30	91.67	92.11
Efectividad: 92.33%										

Base de Datos sin PromRelP B										
Aprendizaje: 20%						Prueba: 80%				
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	195	191	192	189	195	192	191	193	191	195
Perdidos	7	11	10	13	7	10	11	9	11	7
Efectividad (%)	96.53	94.55	95.05	93.56	96.53	95.05	94.55	95.54	94.55	96.53
Efectividad: 95.25%										

Base de Datos sin PromRelP C										
Aprendizaje: 30%						Prueba: 70%				
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	168	166	162	165	168	165	164	166	168	166
Perdidos	10	12	16	13	10	13	14	12	10	12
Efectividad (%)	94.38	93.26	91.01	92.70	94.38	92.70	92.13	93.26	94.38	93.26
Efectividad: 93.15%										

Base de Datos sin PromRelP D										
Aprendizaje: 40%						Prueba: 60%				
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	144	140	143	142	143	145	144	143	144	146
Perdidos	8	12	9	10	9	7	8	9	8	6
Efectividad (%)	94.74	92.11	94.08	93.42	94.08	95.39	94.74	94.08	94.74	96.05
Efectividad: 94.34%										

Base de Datos sin PromRelP E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	121	120	116	118	122	124	120	119	121	124
Perdidos	5	6	10	8	4	2	6	7	5	2
Efectividad (%)	96.03	95.24	92.06	93.65	96.83	98.41	95.24	94.44	96.03	98.41
Efectividad: 95.63%										

Base de Datos sin PromRelP F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	96	94	95	96	98	93	99	97	98	96
Perdidos	6	8	7	6	4	9	3	5	4	6
Efectividad (%)	94.11	92.16	93.14	94.12	96.08	91.18	97.06	95.10	96.08	94.12
Efectividad: 94.31%										

Base de Datos sin PromRelP G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	73	72	70	71	73	75	72	71	75	73
Perdidos	3	4	6	5	3	1	4	5	1	3
Efectividad (%)	96.05	94.74	92.11	93.42	96.05	98.68	94.74	93.42	98.68	96.05
Efectividad: 95.39%										

Base de Datos sin PromRelP H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	49	50	50	49	51	52	51	50	49	51
Perdidos	3	2	2	3	1	1	1	2	3	1
Efectividad (%)	94.23	96.15	96.15	94.23	98.08	98.11	98.08	96.15	94.23	98.08
Efectividad: 96.35%										

Base de Datos sin PromRelP I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	25	26	25	25	24	26	24	26	26	24
Perdidos	1	0	1	1	2	0	2	3	0	2
Efectividad (%)	96.15	100.00	96.15	96.15	92.31	100.00	92.31	89.66	100.00	92.31
Efectividad: 95.50%										

Tabla 2.8 Porcentaje de efectividad de Hold Out Base de Datos sin Ambos Registros A - I

Base de Datos sin ambos Registros A										
Aprendizaje: 10%					Prueba: 90%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	220	215	210	222	215	215	215	220	215	222
Perdidos	8	13	18	6	13	13	13	8	13	6
Efectividad (%)	96.49	94.30	92.11	97.37	94.30	94.30	94.30	96.49	94.30	97.37
Efectividad: 95.13%										

Base de Datos sin ambos Registros B										
Aprendizaje: 20%					Prueba: 80%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	195	194	193	195	193	194	193	195	195	193
Perdidos	7	8	9	7	9	8	9	7	7	9
Efectividad (%)	96.53	96.04	95.54	96.53	95.54	96.04	95.54	96.53	96.53	95.54
Efectividad: 96.03%										

Base de Datos sin ambos Registros C										
Aprendizaje: 30%					Prueba: 70%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	173	169	170	171	169	173	170	170	173	171
Perdidos	5	9	8	7	9	5	8	8	5	6
Efectividad (%)	97.51	94.94	95.51	96.07	94.94	97.19	95.51	95.51	97.19	96.61
Efectividad: 96.09%										

Base de Datos sin ambos Registros D										
Aprendizaje: 40%					Prueba: 60%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	147	143	145	143	147	148	143	147	145	148
Perdidos	5	9	7	9	5	4	9	5	7	4
Efectividad	96.71	94.08	95.39	94.08	96.71	97.37	94.08	96.71	95.39	97.37
Efectividad: 95.78%										

Base de Datos sin ambos Registros E										
Aprendizaje: 50%					Prueba: 50%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	122	120	119	120	122	123	120	123	119	122
Perdidos	4	6	7	6	4	3	6	3	7	4
Efectividad (%)	96.83	95.24	94.44	95.24	96.83	97.62	95.24	97.62	94.44	96.83
Efectividad: 96.03%										

Base de Datos sin ambos Registros F										
Aprendizaje: 60%					Prueba: 40%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	97	94	96	98	95	99	97	97	99	98
Perdidos	5	8	6	4	7	3	5	5	3	4
Efectividad	95.1	92.16	94.12	96.08	93.14	97.06	95.10	95.10	97.06	96.08
Efectividad: 95.09%										

Base de Datos sin ambos Registros G										
Aprendizaje: 70%					Prueba: 30%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	75	74	75	71	74	71	74	72	71	74
Perdidos	1	2	1	5	2	5	2	4	5	2
Efectividad	98.68	97.37	98.68	93.42	97.37	93.42	97.37	94.74	93.42	97.37
Efectividad: 96.18%										

Base de Datos sin ambos Registros H										
Aprendizaje: 80%					Prueba: 20%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	51	51	50	51	52	51	51	51	50	52
Perdidos	1	1	2	1	0	1	1	1	2	0
Efectividad	98.04	98.08	96.15	98.08	100.00	98.08	98.08	98.08	96.15	100.00
Efectividad: 98.07%										

Base de Datos sin ambos Registros I										
Aprendizaje: 90%					Prueba: 10%					
No. de Pruebas	1	2	3	4	5	6	7	8	9	10
Recuperados	25	25	26	24	24	26	25	24	25	26
Perdidos	1	1	0	2	2	0	1	2	1	0
Efectividad	96.15	96.15	100.00	92.31	92.31	100.00	96.15	92.31	96.15	100.00
Efectividad: 96.15%										

5.1.3 K-FoldCross Validation

Para poder implementar esta prueba, es necesario dividir la base de datos en 10 partes, cada parte será una K_n | $n=1,2,3,\dots,10$.. Posteriormente se toma una K como conjunto de prueba, y el resto se elige como conjunto de aprendizaje. Se clasifica cada uno de los registros de prueba, y se compara con su registro correspondiente de la base de datos. Se determina el número de errores, y con ello, se obtiene el porcentaje de efectividad.

Algoritmo

- 1- Dividir la base de datos en 10 partes.
- 2- Tomar una parte como conjunto de prueba.
- 3- Elegir el resto de la base de datos como conjunto de aprendizaje.
- 4- Clasificar los registros de prueba.
- 5- Comparar cada registro clasificado con su correspondiente de la base de datos.
- 6- Hacer un conteo de patrones recuperados y perdidos.
- 7- Se repite los pasos 2 al 6 para cada una de las K.
Mencionar lo de la clasificación de la base de datos Clase 1 clase 2

Algoritmo 1

Tabla 3.1 Porcentaje de efectividad de KFoldCross Validation Base de Datos Completa

Base de Datos Completa

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K1-1	206	34	85.83	K2-1	181	59	75.41
K1-2	200	40	83.33	K2-2	179	61	74.58
K1-3	181	59	75.42	K2-3	202	38	84.17
K1-4	189	51	78.75	K2-4	184	56	76.67
K1-5	201	39	83.75	K2-5	192	48	80.00
Promedio			81.42%	Promedio			78.17%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K3-1	180	60	75	K4-1	201	39	83.75
K3-2	196	44	81.67	K4-2	174	66	72.50
K3-3	209	31	87.08	K4-3	197	43	82.08
K3-4	163	77	67.92	K4-4	203	37	84.58
K3-5	140	100	58.33	K4-5	173	67	72.08
Promedio			74.00%	Promedio			79.00%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K5-1	184	56	76.67
K5-2	202	38	84.17
K5-3	184	56	76.67
K5-4	200	40	83.33
K5-5	186	54	77.50
Promedio			79.67%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K6-1	182	58	75.83
K6-2	211	29	87.92
K6-3	197	43	82.08
K6-4	193	47	80.42
K6-5	189	51	78.75
Promedio			81.00%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K7-1	191	49	79.58
K7-2	187	53	77.92
K7-3	178	62	74.17
K7-4	180	60	75.00
K7-5	204	36	85.00
Promedio			78.33%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K8-1	212	28	88.33
K8-2	212	28	88.33
K8-3	171	69	71.25
K8-4	192	48	80.00
K8-5	189	51	78.75
Promedio			81.33%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K9-1	190	50	79.16
K9-2	153	87	63.75
K9-3	165	75	68.75
K9-4	201	39	83.75
K9-5	197	43	82.08
Promedio			75.50%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K10-1	135	105	56.25
K10-2	198	42	82.50
K10-3	187	53	77.92
K10-4	207	33	86.25
K10-5	177	63	73.75
Promedio			75.33%

Tabla 3.2 Porcentaje de efectividad de KFoldCross Validation Base de Datos sin Jitter Factor

Base de Datos sin Jitter Factor

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K1-1	168	72	70.00	K2-1	208	32	86.66
K1-2	208	32	86.67	K2-2	185	55	77.08
K1-3	204	36	85.00	K2-3	154	86	64.17
K1-4	167	73	69.58	K2-4	196	44	81.67
K1-5	198	42	82.50	K2-5	168	72	70.00
Promedio			78.75%	Promedio			75.92%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K3-1	168	72	70	K4-1	191	49	79.58
K3-2	193	27	87.73	K4-2	213	27	88.75
K3-3	191	49	79.58	K4-3	208	32	86.67
K3-4	196	44	81.67	K4-4	180	60	75.00
K3-5	194	46	80.83	K4-5	204	36	85.00
Promedio			79.96%	Promedio			83.00%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K5-1	207	33	86.25	K6-1	176	64	73.33
K5-2	163	77	67.92	K6-2	179	61	74.58
K5-3	166	74	69.17	K6-3	185	55	77.08
K5-4	190	50	79.17	K6-4	203	37	84.58
K5-5	191	49	79.58	K6-5	159	81	66.25
Promedio			76.42%	Promedio			75.16%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K7-1	198	42	82.5	K8-1	201	39	83.73
K7-2	158	82	65.83	K8-2	203	37	84.58
K7-3	164	76	68.33	K8-3	193	47	80.42
K7-4	193	47	80.42	K8-4	206	34	85.83
K7-5	201	39	83.75	K8-5	178	62	74.17
Promedio			76.17%	Promedio			81.75%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K9-1	169	71	70.41	K10-1	207	33	86.25
K9-2	200	40	83.33	K10-2	168	72	70.00
K9-3	172	68	71.67	K10-3	202	38	84.17
K9-4	201	39	83.75	K10-4	196	44	81.67
K9-5	168	72	70	K10-5	201	39	83.75
Promedio			75.83%	Promedio			81.17%

Tabla 3.3 Porcentaje de efectividad de KFoldCross Validation Base de Datos sin PromRelP

Base de Datos Sin PromRelP

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K1-1	203	37	84.58	K2-1	204	36	85.00
K1-2	176	64	73.33	K2-2	210	30	87.50
K1-3	198	42	82.50	K2-3	197	43	82.08
K1-4	193	47	80.42	K2-4	183	57	76.25
K1-5	187	53	77.92	K2-5	189	51	78.75
Promedio			79.75%	Promedio			81.92%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K3-1	189	51	78.75
K3-2	193	47	80.42
K3-3	213	27	88.75
K3-4	195	45	81.25
K3-5	196	44	81.67
Promedio			82.17%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K4-1	213	27	88.75
K4-2	199	41	82.92
K4-3	183	57	76.25
K4-4	186	54	77.50
K4-5	209	31	87.08
Promedio			82.50%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K5-1	202	38	84.17
K5-2	181	59	75.42
K5-3	204	36	85.00
K5-4	215	25	89.58
K5-5	214	26	89.17
Promedio			84.67%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K6-1	203	27	88.26
K6-2	207	33	86.25
K6-3	206	34	85.83
K6-4	206	34	85.83
K6-5	185	55	77.08
Promedio			84.65%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K7-1	197	43	82.08
K7-2	205	35	85.42
K7-3	164	76	68.33
K7-4	188	52	78.33
K7-5	207	33	86.25
Promedio			80.08%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K8-1	198	42	82.50
K8-2	205	35	85.42
K8-3	196	44	81.67
K8-4	196	44	81.67
K8-5	207	33	86.25
Promedio			83.5%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K9-1	194	46	80.83	K10-1	216	24	90.00
K9-2	185	55	77.08	K10-2	192	48	80.00
K9-3	208	32	86.67	K10-3	206	34	85.83
K9-4	186	54	77.5	K10-4	191	49	79.58
K9-5	207	33	86.25	K10-5	159	81	66.25
Promedio			81.67%	Promedio			80.33%

Tabla 3.4 Porcentaje de efectividad de KFoldCross Validation Base de Datos sin Ambos Campos

Base de Datos Sin Ambos Campos

K de Prueba	Patrones		Efectividad	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K1-1	212	28	88.33	K2-1	211	29	87.92
K1-2	212	28	88.33	K2-2	216	24	90.00
K1-3	215	25	89.58	K2-3	209	31	87.08
K1-4	207	33	86.25	K2-4	145	95	60.42
K1-5	206	33	86.19	K2-5	209	31	87.08
Promedio			87.74%	Promedio			82.50%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K3-1	207	33	86.25	K4-1	196	44	81.67
K3-2	173	67	72.08	K4-2	186	54	77.50
K3-3	188	52	78.33	K4-3	198	42	82.50
K3-4	202	38	84.17	K4-4	201	39	83.75
K3-5	202	38	84.17	K4-5	185	55	77.08
Promedio			81.00%	Promedio			80.50%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K5-1	206	34	85.83
K5-2	202	38	84.17
K5-3	201	39	83.75
K5-4	208	32	86.67
K5-5	183	57	76.25
Promedio			83.33%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K6-1	203	37	84.58
K6-2	191	49	79.58
K6-3	199	41	82.92
K6-4	171	69	71.25
K6-5	203	37	84.58
Promedio			80.58%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K7-1	204	36	85.00
K7-2	191	49	79.58
K7-3	192	48	80.00
K7-4	216	24	90.00
K7-5	175	65	72.92
Promedio			81.5%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K8-1	207	33	86.25
K8-2	205	35	85.42
K8-3	196	44	81.67
K8-4	208	32	86.67
K8-5	194	46	80.83
Promedio			84.17%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K9-1	196	44	81.67
K9-2	185	55	77.08
K9-3	213	24	89.87
K9-4	178	62	74.17
K9-5	212	28	88.33
Promedio			82.22%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K10-1	202	38	84.17
K10-2	213	27	88.75
K10-3	193	47	80.42
K10-4	191	49	79.58
K10-5	200	40	83.33
Promedio			83.25%

Algoritmo 2

Tabla 3.5 Porcentaje de efectividad de KFoldCross Validation Base de Datos Completa

Base de Datos Completa

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K1-1	219	21	91.25	K2-1	223	17	92.92
K1-2	228	12	95.00	K2-2	228	12	95.00
K1-3	217	23	90.42	K2-3	219	21	91.25
K1-4	223	17	92.92	K2-4	217	23	90.42
K1-5	220	20	91.67	K2-5	225	15	93.75
Promedio			92.25%	Promedio			92.67%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K3-1	226	14	94.17	K4-1	227	13	94.58
K3-2	220	20	91.67	K4-2	219	21	91.25
K3-3	217	23	90.42	K4-3	216	24	90.00
K3-4	223	17	92.92	K4-4	220	20	91.67
K3-5	230	10	95.83	K4-5	228	12	95.00
Promedio			93.00%	Promedio			92.50%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K5-1	216	24	90.00	K6-1	214	26	89.17
K5-2	223	17	92.92	K6-2	211	29	87.92
K5-3	223	17	92.92	K6-3	216	24	90.00
K5-4	228	12	95.00	K6-4	225	15	93.75
K5-5	216	24	90.00	K6-5	227	13	94.58
Promedio			92.17%	Promedio			91.08%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K7-1	225	15	93.75	K8-1	224	16	93.33
K7-2	224	16	93.33	K8-2	218	22	90.83
K7-3	230	10	95.83	K8-3	229	11	95.42
K7-4	219	21	91.25	K8-4	230	10	95.83
K7-5	231	9	96.25	K8-5	223	17	92.92
Promedio			94.08%	Promedio			93.67%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K9-1	231	9	96.25	K10-1	227	13	94.58
K9-2	218	22	90.83	K10-2	217	23	90.42
K9-3	225	15	93.75	K10-3	219	21	91.25
K9-4	227	13	94.58	K10-4	228	12	95.00
K9-5	218	22	90.83	K10-5	230	10	95.83
Promedio			93.25%	Promedio			93.42%

Tabla 3.6 Porcentaje de efectividad de KFoldCross Validation Base de Datos sin Jitter Factor

Base de Datos sin Jitter Factor

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K1-1	223	17	92.92	K2-1	208	32	86.67
K1-2	224	16	93.33	K2-2	216	24	90.00
K1-3	223	17	92.92	K2-3	218	22	90.83
K1-4	220	20	91.67	K2-4	227	13	94.58
K1-5	225	15	93.75	K2-5	223	17	92.92
Promedio			92.92%	Promedio			91.00%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K3-1	221	19	92.08	K4-1	223	17	92.92
K3-2	217	23	90.42	K4-2	226	14	94.17
K3-3	219	21	91.25	K4-3	221	19	92.08
K3-4	230	10	95.83	K4-4	219	21	91.25
K3-5	215	25	89.58	K4-5	220	20	91.67
Promedio			91.83%	Promedio			92.42%

K de Prueba	Patrones		Efectividad (%)	K de Prueba	Patrones		Efectividad (%)
	R	P			R	P	
K5-1	218	22	90.83	K6-1	223	17	92.92
K5-2	210	30	87.50	K6-2	225	15	93.75
K5-3	223	17	92.92	K6-3	218	22	90.83
K5-4	227	13	94.58	K6-4	216	24	90.00
K5-5	217	23	90.42	K6-5	217	23	90.42
Promedio			91.25%	Promedio			91.58%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K7-1	218	22	90.83
K7-2	216	24	90.00
K7-3	223	17	92.92
K7-4	219	21	91.25
K7-5	214	26	89.17
Promedio			90.83%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K8-1	216	24	90.00
K8-2	222	18	92.50
K8-3	217	23	90.42
K8-4	220	20	91.67
K8-5	224	16	93.33
Promedio			91.58%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K9-1	225	15	93.75
K9-2	200	40	83.33
K9-3	223	17	92.92
K9-4	216	24	90.00
K9-5	220	20	91.66
Promedio			90.33%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K10-1	228	12	95.00
K10-2	223	17	92.92
K10-3	219	21	91.25
K10-4	224	16	93.33
K10-5	212	28	88.33
Promedio			92.17%

Tabla 3.3 Porcentaje de efectividad de KFoldCross Validation Base de Datos sin PromRelP

Base de Datos Sin PromRelP

K de Prueba	Patrones		Efectividad (%)
	R	P	
K1-1	213	27	88.75
K1-2	229	11	95.42
K1-3	228	12	95.00
K1-4	219	21	91.25
K1-5	225	15	93.75
Promedio			92.83%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K2-1	209	31	87.08
K2-2	210	30	87.50
K2-3	217	23	90.42
K2-4	219	21	91.25
K2-5	226	14	94.17
Promedio			90.08%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K3-1	231	9	96.25
K3-2	224	16	93.33
K3-3	213	27	88.75
K3-4	222	18	92.50
K3-5	219	21	91.25
Promedio			92.42%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K4-1	215	25	89.58
K4-2	230	10	95.83
K4-3	226	14	94.17
K4-4	223	17	92.92
K4-5	219	21	91.25
Promedio			92.75%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K5-1	217	23	90.42
K5-2	229	11	95.42
K5-3	220	20	91.67
K5-4	215	25	89.58
K5-5	214	26	89.17
Promedio			91.25%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K6-1	210	30	87.50
K6-2	216	24	90.00
K6-3	221	19	92.08
K6-4	228	12	95.00
K6-5	216	24	90.00
Promedio			90.92%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K7-1	225	15	93.75
K7-2	205	35	85.42
K7-3	230	10	95.83
K7-4	227	13	94.58
K7-5	207	33	86.25
Promedio			91.17%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K8-1	224	16	93.33
K8-2	220	10	95.65
K8-3	217	23	90.42
K8-4	215	25	89.58
K8-5	217	23	90.42
Promedio			91.88%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K9-1	222	18	92.50
K9-2	218	22	90.83
K9-3	219	21	91.25
K9-4	219	21	91.25
K9-5	220	20	91.66
Promedio			91.50%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K10-1	216	24	90.00
K10-2	223	17	92.92
K10-3	218	22	90.83
K10-4	228	12	95.00
K10-5	225	15	93.75
Promedio			92.50%

Tabla 3.4 Porcentaje de efectividad de KFoldCross Validation Base de Datos sin Ambos Campos

Base de Datos Sin Ambos Campos

K de Prueba	Patrones		Efectividad (%)
	R	P	
K1-1	225	15	93.75
K1-2	217	23	90.42
K1-3	219	21	91.25
K1-4	221	19	92.08
K1-5	218	22	90.83
Promedio			91.67%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K2-1	216	24	90.00
K2-2	215	25	89.58
K2-3	221	19	92.08
K2-4	219	21	91.25
K2-5	217	23	90.42
Promedio			90.67%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K3-1	218	22	90.83
K3-2	228	12	95.00
K3-3	217	23	90.42
K3-4	221	19	92.08
K3-5	231	9	96.25
Promedio			92.92%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K4-1	227	13	94.58
K4-2	215	25	89.58
K4-3	225	15	93.75
K4-4	230	10	95.83
K4-5	219	21	91.25
Promedio			93.00%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K5-1	228	12	95.00
K5-2	218	22	90.83
K5-3	214	26	89.17
K5-4	223	17	92.92
K5-5	231	9	96.25
Promedio			92.83%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K6-1	218	22	90.83
K6-2	216	24	90.00
K6-3	228	12	95.00
K6-4	220	20	91.67
K6-5	226	14	94.17
Promedio			92.33%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K7-1	231	9	96.25
K7-2	217	23	90.42
K7-3	227	13	94.58
K7-4	218	22	90.83
K7-5	219	21	91.25
Promedio			92.66%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K8-1	222	18	92.50
K8-2	220	20	91.67
K8-3	221	19	92.08
K8-4	230	10	95.83
K8-5	229	11	95.42
Promedio			93.50%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K9-1	225	15	93.75
K9-2	227	13	94.58
K9-3	232	8	96.67
K9-4	228	12	95.00
K9-5	224	16	93.33
Promedio			94.67%

K de Prueba	Patrones		Efectividad (%)
	R	P	
K10-1	218	22	90.83
K10-2	221	19	92.08
K10-3	216	24	90.00
K10-4	220	20	91.67
K10-5	222	18	92.50
Promedio			91.41%

5.2 Resultados globales

Una vez obtenidos los resultados: porcentajes y promedios, es necesario hacer una comparación global, con el objetivo de poder analizar la efectividad del clasificador y de esta manera obtener conclusiones.

Tabla 4.1 Efectividad de Leave One Out

Algoritmo 1

Leave One Out			
Efectividad Base de Datos Completa	Efectividad Base de Datos Sin Jitter Factor	Efectividad Base de Datos Sin PromRelP	Efectividad Base de Datos Sin Ambos Campos
90.47%	90.07%	91.26%	90.87%

Algoritmo 2

Leave One Out			
Efectividad Base de Datos Completa	Efectividad Base de Datos Sin Jitter Factor	Efectividad Base de Datos Sin PromRelP	Efectividad Base de Datos Sin Ambos Campos
95.63%	95.23%	96.82%	96.03%

Tabla 4.2 Efectividad de Hold Out

Algoritmo 1

Hold Out					
Porcentaje de Aprendizaje	Porcentaje de Prueba	Efectividad Base de Datos Completa	Efectividad Base de Datos Sin Jitter Factor	Efectividad Base de Datos Sin PromRelP	Efectividad Base de Datos Sin Ambos Campos
10%	90%	83.94%	88.05%	86.34%	88.28%
20%	80%	90.99%	89.30%	89.06%	90.14%
30%	70%	86.68%	87.58%	87.19%	89.83%
40%	60%	88.02%	87.69%	88.82%	89.21%
50%	50%	87.61%	89.12%	89.68%	90.07%
60%	40%	88.90%	88.33%	88.23%	89.80%
70%	30%	90.39%	90.26%	89.61%	91.05%
80%	20%	92.30%	90.96%	90.77%	93.46%
90%	10%	90.00%	88.46%	89.62%	90.28%

Algoritmo 2

Hold Out					
Porcentaje de Aprendizaje	Porcentaje de Prueba	Efectividad Base de Datos Completa	Efectividad Base de Datos Sin Jitter Factor	Efectividad Base de Datos Sin PromRelP	Efectividad Base de Datos Sin Ambos Campos
10%	90%	91.97%	93.35%	92.33%	95.13%
20%	80%	95.17%	94.10%	95.25%	96.03%
30%	70%	91.06%	92.35%	93.15%	96.09%
40%	60%	93.35%	93.15%	94.34%	95.78%
50%	50%	92.38%	94.28%	95.63%	96.03%
60%	40%	93.03%	93.33%	94.31%	95.09%
70%	30%	95.13%	95.13%	95.39%	96.18%
80%	20%	97.30%	95.57%	96.35%	98.07%
90%	10%	95%	93.46%	95.50%	96.15%

Tabla 4.3 Efectividad de k-Fold Cross Validation

Algoritmo 1

K-Fold Cross Validation				
No. de Prueba	Efectividad Base de Datos Completa	Efectividad Base de Datos Sin Jitter Factor	Efectividad Base de Datos Sin PromRelP	Efectividad Base de Datos Sin Ambos Campos
K -1	81.42%	78.75%	79.75%	87.74%
K-2	78.17%	75.92%	81.92%	82.50%
K-3	74.00%	79.96%	82.17%	81.00%
K -4	79.00%	83.00%	82.50%	80.50%
K -5	79.67%	76.42%	84.67%	83.33%
K -6	81.00%	75.16%	84.65%	80.58%
K -7	78.33%	76.17%	80.08%	81.5%
K-8	81.33%	81.75%	83.5%	84.17%
K-9	75.50%	75.83%	81.67%	82.22%
K-10	75.33%	81.17%	80.33%	83.25%

Algoritmo 2

K-Fold Cross Validation				
No. de Prueba	Efectividad Base de Datos Completa	Efectividad Base de Datos Sin Jitter Factor	Efectividad Base de Datos Sin PromRelP	Efectividad Base de Datos Sin Ambos Campos
K -1	92.25%	92.92%	92.83%	91.67%
K-2	92.67%	91.00%	90.08%	90.67%
K-3	93.00%	91.83%	92.42%	92.92%
K -4	92.50%	92.42%	92.75%	93.00%
K -5	92.17%	91.25%	91.25%	92.83%
K -6	91.08%	91.58%	90.92%	92.33%
K -7	94.08%	90.83%	91.17%	92.66%
K-8	93.67%	91.58%	91.88%	93.50%
K-9	93.25%	90.33%	91.50%	94.67%
K-10	93.42%	92.17%	92.50%	91.41%

Algoritmo 1

En la prueba Leave One Out el mejor resultado se obtuvo **91.26%** de efectividad con la Base de Dato Sin PromRelP.

En las pruebas Hold Out se obtuvo **93.07%** con 80% de Aprendizaje y 20% de Prueba con la Base de Datos Sin Ambos Campos.

El promedio de las pruebas en K-FoldCross Validation más alto es del **87.74%** con la Base de Datos Sin Ambos Campos.

Algoritmo 2

En la prueba Leave One Out el mejor resultado se obtuvo **96.82%** de efectividad con la Base de Dato Sin PromRelP.

En las pruebas Hold Out se obtuvo **98.08%** con 80% de Aprendizaje y 20% de Prueba con la Base de Datos Sin Ambos Campos..

El promedio de las pruebas en K-FoldCross Validation más alto es del **94.67%** con la Base de Datos Sin Ambos Campos.

Conclusiones

Se ha diseñado dos clasificadores de pacientes con hipoacusia usando memorias asociativas morfológicas, el funcionamiento de estos algoritmos ha sido satisfactorio para realizar pruebas de efectividad en la clasificación de pacientes, discriminando los pacientes hipacusicos y no hipoacusicos.

Después de realizar pruebas de efectividad, los resultados que se han obtenido para *Leave One Out* del algoritmo 1 es de 91.26% con la Base de Datos sin el Campo *PromRelP* , con Hold Out ha sido de 93.07% usando 80% como conjunto de aprendizaje y 20% como conjunto de prueba con la Base de Datos Sin Ambos Registros, y 87.74% para K-FoldCross Validation con la Base de Datos Sin Ambos Registros.

Para el algoritmo 2 los resultados que se han obtenido para *Leave One Out* es de 96.82% con la Base de Datos sin el Campo *PromRelP* , con Hold Out ha sido de 98.08% usando 80% como conjunto de aprendizaje y 20% como conjunto de prueba con la Base de Datos Sin Ambos Registros, y 94.67% para K-FoldCross Validation con la Base de Datos Sin Ambos Registros.

Haber realizado este trabajo implico hacer dos clasificadores debido a que no hay clasificadores que estén basados específicamente a pacientes con hipoacusia.

Los modelos asociativos superan en efectividad unos a otros modelos de clasificación de patrones, como en el primer clasificador con memorias asociativas morfológicas que se propuso en este trabajo, ha sido superado por el segundo clasificador con memorias asociativas morfológicas implementando el método Johnson-Môbuis modificado. Sin embargo, el segundo clasificador a pesar de tener un alto porcentaje de efectividad es muy lento en la obtención de la clasificación debido a la gran cantidad de información que se va procesando.

Trabajos Futuros

Implementación de clasificadores del trabajo presentado para obtención de resultados con alta efectividad y sobre todo obtener una mejor eficacia en el procesamiento de datos.

REFERENCIAS

- [1] Hassoun, M. H. 1993, *Associative Neural Memories*, Oxford University Press, New York.
- [2] <http://widexvenezuela.com/perdidas-auditivas/>
- [3] http://health.kernan.org/esp_ency/article/001045.htm
- [4] <http://www.oticon.es/hearing/facts/hearing/how-hearing-works.aspx>
- [5] <http://www.webconsultas.com/hipoacusia/hipoacusia-761>
- [6] http://www.medicasur.com.mx/es_mx/ms/ms_oir_Hipoacusia_o_disminucion_de_la_audicion
- [7] <http://www.oticon.es/hearing/facts/hearing/how-hearing-works.aspx>
- [8] <http://hyperphysics.phy-astr.gsu.edu/hbasees/sound/earsens.html>
- [9] http://kidshealth.org/parent/en_espanol/general/ears_esp.html
- [10] <http://www.neurowikia.es/content/epidemiologi>
- [11] <http://www.who.int/mediacentre/factsheets/fs300/es/>
- [12] <http://chemocare.com/es/chemotherapy/side-effects/problemas-de-la-audioacuten.aspx>
- [13] <http://www.maneyonline.com/doi/abs/10.1179/136132899807557475>
- [14] <http://www.sciencedirect.com/science/article/pii/S0892199704001213>

[15] <http://www.ncvs.org/freebooks/summary-statement.pdf>

[16] *Consortio de datos lingüísticos, 441 Williams Hall, Universidad de Pennsylvania, Philadelphia, PA 19104, World Wide Web: Ftp \\ www.cis.upenn. edu.*

[17] <http://www.clinicadam.com/salud/5/003044.html>

[18] Kohonen, T. 1989, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin.

[19] Yáñez, C. 2002, *Memorias Asociativas basadas en Relaciones de Orden y Operadores Binarios*, Tesis de Doctorado, Centro de Investigación en Computación, México.

[20] Minsky, M. & Papert, S. 1969, *Perceptrons*, MIT Press ,Cambridge.

[21] Steinbuch, K. & Frank, H. 1961, "Nichtdigitale Lernmatrizen als Perzeptoren", *Kybernetik*, vol. 1, no.3, pp. 117-124.

[22] Kohonen, T. 1972, "Correlation matrix memories", *IEEE Transactions on Computers*, C-21, vol. 4, pp. 353-359.

[23] Papadomanolakis, K., Kakarountas, A., Sklavos, N. & Goutis C. E., 2002, A Fast Johnson-Mobius Encoding Scheme for Fault Secure Binary Counters, *Proceedings of Design, Automation and Test in Europe*, (DATE '02), pp. 1-7.

[24] Willshaw, D., Buneman, O. & Longuet-Higgins, H. 1969, "Non-holographic associative memory", *Nature*, no. 222, pp. 960-962.

[25] Anderson, J. A. 1972, "A simple neural network generating an interactive memory", *Mathematical Biosciences*, vol. 14, pp. 197-220.

[26] Hopfield, J.J. 1982, "Neural networks and physical systems with emergent collective computational abilities", *Proceedings of the National Academy of Sciences*, vol. 79, pp. 2554-2558.

[27] Ritter, G. X., Sussner, P. & Diaz-de-Leon, J. L. 1998, "Morphological associative memories", *IEEE Transactions on Neural Networks*, vol. 9, pp. 281-293.

[28] Sossa, H., Barrón, R. & Vázquez, R. 2004, "New Associative Memories to Recall Real-Valued Patterns", *CIARP, LNCS 3287*, pp. 195-202.

[29] Simpson, P. K. 1990, *Artificial Neural Systems*, Pergamon Press , New York.

[30] Abu-Mostafa, Y. & St. Jacques, J. 1985, "Information capacity of the Hopfield model", *IEEE Transactions on Information Theory*, IT-31, no. 4, pp. 461-464.

[31] Willshaw, D., Buneman, .& Longuet-Higgins, H. 1969. Non-holographic associative memory, *Nature*. No.222. Pág. 960-962.

[32]<http://recursostic.educacion.es/observatorio/web/es/component/content/article/391-monografico-sordera-y-nuevas-tecnologias?start=5>.

[33] Guyon, I. & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.

[34] Blum, A & Langley P.(1997), Selection of Relevant Features and examples in Machine Learning *Artificial Intelligence*, 97, 245-271.

[35] Aldape-Pérez, M., Yáñez-Márquez, C., & Argüelles-Cruz, A. J. (2007). Optimized Associative Memories for Feature Selection. *Lecture Notes in Computer Science*, 4477, 435-442.

