

UNIVERSIDAD AUTÓNOMA DE SINALOA
MAESTRÍA EN CIENCIAS DE LA INFORMACIÓN



UN ESTUDIO SOBRE LA EFICACIA Y LA EFICIENCIA DE
ALGORITMOS DE CLASIFICACIÓN DE
ELECTROCARDIOGRAMAS DIGITALES EN DIFERENTES
PLATAFORMAS COMPUTACIONALES

Que como requisito parcial para obtener el grado de
MAESTRO EN CIENCIAS
presenta

DANIEL ERNESTO LÓPEZ BARRÓN

DIRECTORES:
INÉS FERNANDO VEGA LÓPEZ
CLAUDIA LERMA GONZÁLEZ

Culiacán, Sinaloa Noviembre 2014

*Hola hijo, ¿Cómo te fue en la escuela?, Hola hermanito lindo lindo!,
Hola pepe-yugui!, Danielon!
— La familia.*

*Así está la choco-aventura, Apreciable joven!,
Dígame, Licenciado!
— Los guías.*

*Pinche cacahuete!, O si no, te mato!, Sufre alimaña!, Hola tú!
— Los amigos.*

*Eres el mejor!, I'm a new soul ..
Hay nos víboras de la mar!
— Los mejores roomies y el INC.*

*Dígalo, sin miedo!
— UPSIN Mazatlán.*

*Licencia?, solo la de Dios, Oh! Happy Day, Happy Day!, Por favor, regresa al párrafo anterior,
Tas' pollo, Oh! mi Danny!, I dare you, I double dare you ..., Eso es obvio!
— Los compañeros de la maestría.*

*Estas son las frases que me acompañaron en el desarrollo
de este trabajo, a todos muchas gracias.*

AGRADECIMIENTOS

Quiero agradecer a mi familia, mis hermanas Ana y Gabriela, mi padre Gabriel y muy especialmente a mi madre Rita que siempre ha estado ahí para mí con su apoyo incondicional. Además, quiero agradecer a mi asesor Inés Fernando Vega López, por su tiempo, guía, apoyo y consejos. Es una persona ejemplar, a quien admiro y respeto mucho.

Aprovecho para agradecer al Departamento de Instrumentación del Instituto Nacional de Cardiología *Ignacio Chávez*, a Óscar Infante, Claudia Lerma, Raúl Memije, Gustavo Sánchez y Brayans Becerra, por su apoyo durante mi estancia de investigación. Así mismo, agradezco a los doctores Luis Mena y Eber Orozco de la Universidad Politécnica de Sinaloa, por su apoyo en mi estancia de investigación.

Agradezco a mis *hermanos mayores académicos* Sandra Lara y Oswaldo Cuen, siempre apreciaré su amistad y experiencia. Otros involucrados en este trabajo son mis amigos Leobardo, Hormandi, Yunibia, Rocío, Alicia y Óscar, a quienes agradezco por hacer de este viaje algo divertido, les debo muchas risas y buenos momentos. Agradezco a mis compañeros de maestría y amigos Marcia, César, Édgar, Miguel, Rosendo, Jorge y Bily, por hacer de la maestría una segunda casa. ¡Unos verdaderos tipazos!.

Finalmente, quiero agradecer al Consejo Nacional de Ciencia y Tecnología por su apoyo financiero en la beca 264829 y en el proyecto de investigación de Ciencia Básica CONACyT 106098. Además, agradezco a la Universidad Autónoma de Sinaloa, así como a la Facultad de Informática Culiacán y a la Facultad de Ciencias de la Tierra y el Espacio, por el apoyo en el desarrollo de este trabajo.

ÍNDICE

Índice de Figuras	III
Índice de Tablas	v
Índice de Algoritmos	x
Resumen	xI
Capítulo 1. Introducción	1
1.1. Descripción del Problema	2
1.2. Motivación	4
Capítulo 2. Clasificación de Electrocardiogramas	7
2.1. Representación del Electrocardiograma (ECG)	8
2.2. Algoritmos de Clasificación de Electrocardiogramas (ECGs)	9
2.3. Clasificación por Vecinos más Cercanos	11
2.3.1. Función de Distancia Euclidiana	14
2.3.2. Coeficiente de Correlación Cruzada	15
2.4. Redes Neuronales Artificiales	16
2.4.1. Cálculo y Clasificación	18
2.4.2. Algoritmo de Entrenamiento	21
2.4.3. Parámetros de Optimización	23
2.5. Plataformas Computacionales	23
2.6. Desempeño Algorítmico	24
Capítulo 3. Implementación de Algoritmos de Clasificación	26
3.1. Datos de Prueba	27
3.1.1. Descripción del Conjunto de Datos	27
3.1.2. Distribución de Datos	28
3.1.3. Selección de Patrones	30
3.2. Desarrollo de los Algoritmos de Clasificación	34
3.2.1. Clasificación por Vecino más Cercano	34
3.2.2. Red Neuronal Artificial	36
3.2.3. Evaluación del Clasificadores	38
3.3. Descripción de las Plataformas Computacionales	38
3.3.1. Esquema Centralizado	39
3.3.2. Esquema Distribuido	39

Capítulo 4. Evaluación Experimental y Resultados	42
4.1. Conformación de Conjuntos de Entrenamiento	43
4.2. Entrenamiento de los Clasificadores	52
4.3. Experimentos sobre una Plataforma Estacionaria	55
4.4. Experimentos sobre una Plataforma Móvil	56
4.5. Resultados de Clasificación	57
4.5.1. Clasificación de Latidos	57
4.5.2. Clasificación Dicotómica	60
4.5.3. Clasificación de Anomalías	64
4.5.4. Eficiencia	98
Capítulo 5. Conclusiones	104
Bibliografía	108
Anexos	116

ÍNDICE DE FIGURAS

1.1.	Trazo de la Actividad Eléctrica.	5
2.1.	Estructura de una Red Neuronal Artificial.	18
3.1.	Ocurrencia de derivaciones en la base de datos MIT-BIH	31
3.2.	Representación Gráfica de Latidos Cardíacos.	33
3.3.	Extracción de una anotación del ECG.	33
4.1.	Desempeño de las RNAs con la Arquitectura 1.	53
4.2.	Desempeño de las RNAs con la Arquitectura 2.	54
4.3.	Desempeño de las RNAs con la Arquitectura 3.	55
4.4.	Eficacia global de clasificadores	59
4.5.	Eficacia centralizada de latidos Normales	61
4.6.	Eficacia centralizada de latidos Anormales	62
4.7.	Eficacia distribuida de latidos Normales	68
4.8.	Eficacia distribuida de latidos Anormales	69
4.9.	Eficacia centralizada de la Clase 1	70
4.10.	Eficacia Distribuida Clase 1	71
4.11.	Eficacia centralizada de la Clase 2	72
4.12.	Eficacia Distribuida Clase 2	73
4.13.	Eficacia centralizada de la Clase 3	74
4.14.	Eficacia Distribuida Clase 3	75
4.15.	Eficacia centralizada de la Clase 4	76
4.16.	Eficacia Distribuida Clase 4	78
4.17.	Eficacia centralizada de la Clase 5	79
4.18.	Eficacia Distribuida Clase 5	81
4.19.	Eficacia centralizada de la Clase 6	82
4.20.	Eficacia Distribuida Clase 6	83
4.21.	Eficacia centralizada de la Clase 8	84
4.22.	Eficacia Distribuida Clase 8	85
4.23.	Eficacia centralizada de la Clase 9	86
4.24.	Eficacia Distribuida Clase 9	88
4.25.	Eficacia centralizada de la Clase 11	89
4.26.	Eficacia Distribuida Clase 11	91
4.27.	Eficacia centralizada de la Clase 12	92
4.28.	Eficacia Distribuida Clase 12	93
4.29.	Eficacia centralizada de la Clase 13	94
4.30.	Eficacia Distribuida Clase 13	95
4.31.	Eficacia centralizada de la Clase 14	96

4.32.	Eficacia Distribuida Clase 14	97
4.33.	Desempeño Tiempo.	98
4.34.	Eficacia de los clasificadores 1-NN	99
4.35.	Desempeño Tiempo.	100

ÍNDICE DE TABLAS

3.1.	Distribución de Anotaciones por Derivación en el Conjunto de Datos	31
3.2.	Distribución de los T.A.s en la Derivación MLII	32
4.1.	Distribución aleatoria uniforme de datos con 14 clases conjuntos de 5K hasta 40K.	46
4.2.	Distribución aleatoria uniforme de datos con 14 clases conjuntos de 45K hasta 80K.	47
4.3.	Distribución aleatoria uniforme de datos con 14 clases conjuntos de 85K hasta el conjunto completo.	47
4.4.	Distribución equi-proporcional de datos con 14 clases conjuntos de 5K hasta 40K.	48
4.5.	Distribución equi-proporcional de datos con 14 clases conjuntos de 45K hasta 80K.	48
4.6.	Distribución equi-proporcional de datos con 14 clases conjuntos de 85K hasta Completa.	49
4.7.	Distribución aleatoria uniforme de datos con dos clases conjuntos de 5K hasta 40K.	50
4.8.	Distribución aleatoria uniforme de datos con dos clases conjuntos de 45K hasta 80K.	50
4.9.	Distribución aleatoria uniforme de datos con dos clases conjuntos de 85K hasta completa.	50
4.10.	Distribución equi-proporcional de datos con dos clases conjuntos de 5K hasta 40K.	51
4.11.	Distribución equi-proporcional de datos con dos clases conjuntos de 45K hasta 80K.	51
4.12.	Distribución equi-proporcional de datos con dos clases conjuntos de 85K hasta completa.	51
4.13.	Tiempo ejecución de los clasificadores 1-NN en los esquemas de trabajo centralizado y distribuido (segs).	101
4.14.	Tiempo ejecución de las clasificación por redes neuronales.	102
4.15.	Cantidad de bytes requeridos por los clasificadores.	103
1.	Porcentaje de eficacia para latidos normales en conjuntos de referencia aleatorios en la clasificación dicotómica.	117
2.	Porcentaje de eficacia para latidos anormales en conjuntos de referencia aleatorios en la clasificación dicotómica.	117
3.	Porcentaje de eficacia para latidos normales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.	118

4.	Porcentaje de eficacia para latidos anormales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.	118
5.	Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia aleatorios en la clasificación de anomalías.	119
6.	Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia aleatorios en la clasificación de anomalías.	119
7.	Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia aleatorios en la clasificación de anomalías.	120
8.	Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia aleatorios en la clasificación de anomalías.	120
9.	Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia aleatorios en la clasificación de anomalías.	120
10.	Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia aleatorios en la clasificación de anomalías.	121
11.	Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia aleatorios en la clasificación de anomalías.	121
12.	Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia aleatorios en la clasificación de anomalías.	121
13.	Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia aleatorios en la clasificación de anomalías.	122
14.	Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia aleatorios en la clasificación de anomalías.	122
15.	Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia aleatorios en la clasificación de anomalías.	122
16.	Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia aleatorios en la clasificación de anomalías.	123
17.	Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia aleatorios en la clasificación de anomalías.	123
18.	Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia aleatorios en la clasificación de anomalías.	123
19.	Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	124
20.	Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	124
21.	Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	125
22.	Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	125
23.	Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	125

24.	Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	126
25.	Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	126
26.	Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	126
27.	Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	127
28.	Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	127
29.	Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	127
30.	Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	128
31.	Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	128
32.	Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	128
33.	Porcentaje de eficacia para latidos normales en conjuntos de referencia aleatorios en la clasificación dicotómica.	129
34.	Porcentaje de eficacia para latidos anormales en conjuntos de referencia aleatorios en la clasificación dicotómica.	129
35.	Porcentaje de eficacia para latidos normales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.	130
36.	Porcentaje de eficacia para latidos anormales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.	130
37.	Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia aleatorios en la clasificación de anomalías.	131
38.	Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia aleatorios en la clasificación de anomalías.	131
39.	Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia aleatorios en la clasificación de anomalías.	132
40.	Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia aleatorios en la clasificación de anomalías.	132
41.	Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia aleatorios en la clasificación de anomalías.	132
42.	Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia aleatorios en la clasificación de anomalías.	133
43.	Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia aleatorios en la clasificación de anomalías.	133

44.	Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia aleatorios en la clasificación de anomalías.	133
45.	Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia aleatorios en la clasificación de anomalías.	134
46.	Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia aleatorios en la clasificación de anomalías.	134
47.	Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia aleatorios en la clasificación de anomalías.	134
48.	Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia aleatorios en la clasificación de anomalías.	135
49.	Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia aleatorios en la clasificación de anomalías.	135
50.	Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia aleatorios en la clasificación de anomalías.	135
51.	Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	136
52.	Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	136
53.	Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	137
54.	Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	137
55.	Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	137
56.	Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	138
57.	Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	138
58.	Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	138
59.	Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	139
60.	Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	139
61.	Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	139
62.	Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	140
63.	Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	140

64.	Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.	140
-----	--	-----

ÍNDICE DE ALGORITMOS

1.	Algoritmo para calcular la salida $\mathcal{Red}(x)$	20
2.	Algoritmo de Backpropagation	22
3.	Algoritmo de clasificación NN	35
4.	Algoritmo de clasificación de Red Neuronal Artificial	37
5.	Selección Equi-Proporcional de Elementos.	45

RESUMEN

El presente estudio realiza una evaluación de algoritmos de clasificación de latidos en electrocardiogramas (ECGs) digitales. Nuestro trabajo determina el aporte de combinaciones de hardware y software (plataformas computacionales) en algoritmos de clasificación, para diferentes tipos de latidos digitales. Los algoritmos evaluados son el algoritmo de Vecinos más Cercanos (conocido también como 1-NN) y la clasificación por Redes Neuronales Artificiales. En la literatura especializada en clasificación de ECGs, se ha demostrado que estos algoritmos son efectivos al clasificar diferentes tipos de latidos digitales. Sin embargo, no ha sido realizado un estudio que compare el aporte que una plataforma computacional otorga a esta eficacia. Parte de nuestro trabajo se enfocó en comparar dos alternativas distintas de clasificadores 1-NN. La diferencia entre los clasificadores 1-NN es el operador que determina la semejanza entre dos latidos distintos. Un clasificador 1-NN implementó la función de la distancia Euclidiana y la otro clasificador correspondió al uso del Coeficiente de Correlación. Otra parte de este estudio evaluó la eficacia de los algoritmos de clasificación al utilizar conjuntos de datos de diferente tamaño y distribución estadística. Además, obtuvimos la eficacia estadística de los algoritmos al distribuir los conjuntos de datos y el procesamiento de los clasificadores, en diferentes unidades de procesamiento.

Nuestros experimentos se realizaron utilizando una plataforma de cómputo estacionaria y una plataforma de cómputo móvil. Así mismo, la plataforma de cómputo estacionaria implementa dos esquemas de trabajo para almacenamiento de los datos: centralizado y distribuido. En el esquema de almacenamiento de datos centralizado, los conjuntos de datos se encuentran en una sola unidad de procesamiento. Por su parte, el esquema de datos distribuido divide, tanto los conjuntos de datos, como el procesamiento de los clasificadores entre las diferentes unidades de procesamiento involucradas.

Los datos experimentales fueron tomados del banco de datos MIT-BIH Arritmia. El banco de datos MIT-BIH Arritmia está compuesto por 48 estudios de electrocardiografías digitales. En este banco de datos se encuentran un total de 225,224 patrones, distribuidos entre 23 tipos de anotaciones distintas. De este total de patrones consideramos 105,056 ejemplos, propios de 14 tipos de alteraciones en la onda eléctrica producida por los latido cardíacos. Cada alteración es identificada por una clase nominal. A partir de los patrones considerados, generamos distintos conjuntos de datos con diferentes cantidades de patrones en cada uno; referidos como *conjuntos de referencia*. El objetivo de los conjuntos de referencia es servir como fuente de información del algoritmo de clasificación. Una vez que el algoritmo consultó los conjuntos de referencia y modificó sus parámetros estadísticos, evaluamos la eficacia del clasificador. Debido a que los datos representan distintos padecimientos cardíacos, la distribución estadística de de estos padecimientos no es constante. La cantidad de patrones entre

cada clase es distinta, esto se debe a que existen alteraciones del ritmo cardíaco que son más comunes que otras. Debido a las diferencias en la cantidad de latidos por clase, se implementaron dos algoritmos de selección de datos. Estos algoritmos permitieron conformar conjuntos de referencia con distintas cardinalidades. Uno de los algoritmos de selección otorga igual probabilidad a cada latido sin importar la clase a la que dicho latido corresponda. La otra alternativa de selección, asegura que todas las clases de latidos están representadas de manera equitativa (a medida de lo posible) en cada conjunto de referencia. Nuestros resultados indican que la eficacia y eficiencia de los algoritmos de clasificación se ve afectada al realizar distribución y partición de datos en los conjuntos de referencia. Igualmente, nuestros resultados exponen la diferencia en la capacidad de las plataformas computacionales para ejecutar los algoritmos de clasificación, al aumentar la cardinalidad de los conjuntos de referencia consultados por el clasificador.

De esta manera, descubrimos que nuestra plataforma de cómputo móvil es capaz de clasificar un latido en un tiempo menor al tiempo de captura de la siguiente muestra en el ECG. Este desempeño en el tiempo de clasificación pertenece a las redes neuronales. Sin importar que arquitectura de red neuronal se utilice para clasificar el latido, de entre las arquitecturas utilizadas en este trabajo. Así mismo, el tiempo de clasificación de las redes neuronales no se ve afectado por el tamaño del conjunto de referencia. Además, comparamos la eficacia de los algoritmos de clasificación cuando los conjuntos de referencia se encuentran distribuidos y centralizados. Concluimos que distribuir los conjuntos de referencia entre diferentes unidades de procesamiento (nodos) mejora (en algunos casos) la eficacia obtenida por los algoritmos de clasificación en el esquema centralizado. El esquema de datos centralizado indica la concentración de la experiencia, representando la clasificación de latidos por un solo experto médico. Ya que, este esquema de almacenamiento requiere de una unidad de procesamiento. En nuestros experimentos de clasificación distribuida, utilizamos distintas cantidades de nodos y repartimos el conjunto de referencia de manera equitativa entre las cantidades de nodos. De esta manera, obtenemos la clase de un latido al procesar dicho latido en todos los nodos. El esquema distribuido asemeja tener múltiples expertos, donde cada experto procesa el latido e indica la clase, es decir vota. Para evaluar la eficacia del este enfoque se realizaron votaciones entre los nodos involucrados. La votación contabiliza el resultado de la clasificación en cada nodo. El resultado de la clasificación distribuida corresponde a la clase con mayor cantidad de votos.

CAPÍTULO 1

INTRODUCCIÓN

El monitoreo de la actividad eléctrica del corazón permite al médico ubicar de manera gráfica puntos de interés, que facilitan la detección temprana de padecimientos cardíacos [12]. De acuerdo con la *Organización Mundial de la Salud* (OMS), en 2008, 17.3 millones de personas fallecieron a nivel mundial por algún padecimiento cardíaco, donde el 80 % de los casos se presentan en países de bajo y mediano ingreso [50]. Esto es debido al menor acceso a servicios eficientes de asistencia sanitaria (*e.g.* estudios de monitoreo cardíaco) que respondan a sus necesidades, comparado con países de mayor desarrollo económico [50]. De igual manera, la OMS señala que cada año 9.4 millones decesos son atribuidos a enfermedades cardíacas a nivel mundial. Para el año 2030, la OMS estima que casi 23.6 millones de personas a nivel mundial morirán por algún padecimiento de este tipo. La detección temprana de padecimientos cardíacos facilita un tratamiento efectivo, además previene que la gravedad de dicho padecimiento aumente [34, 51]. Sin embargo, buscar anomalías en los registros de la actividad eléctrica del corazón consume al experto (*i.e.* cardiólogo), generalmente, largos periodos de tiempo. Debido a que, generalmente los registros pueden tener una duración en tiempo de más de 12 horas, lo que se traduce en millones de observaciones de la actividad eléctrica [12]. Por lo que se requieren alternativas de apoyo computacionales, para automatizar la búsqueda de anomalías en estos registros. Gracias al desarrollo e implementación de las nuevas tecnologías computacionales, los especialistas en el área médica poseen aun mayor apoyo para el análisis y recolección de datos [2, 8]. El uso adecuado de estas tecnologías amplía la cobertura y mejora el servicio médico [13].

Los cardiólogos se han auxiliado de sistemas computacionales para agilizar la detección y clasificación de puntos de interés en electrocardiogramas (ECGs) [3, 6, 49].

Un ECG es una *señal biomédica* que se tiene al observar y capturar la actividad eléctrica del corazón durante un periodo de tiempo. Estos sistemas analizan datos en forma de señales biomédicas. Desde una perspectiva de análisis y almacenamiento de datos, podemos decir que una señal biomédica no posee una estructura con tamaño estándar, a diferencia de tipos de datos convencionales. Generar herramientas computacionales de apoyo a los expertos, presenta retos en el almacenamiento y procesamiento de estos tipos de datos no convencionales [22]. Es necesario considerar un estudio experimental, que aporte información sobre el desempeño de las herramientas de apoyo computarizadas, en el procesamiento de exámenes biomédicos. En el que por medio de la eficiencia y eficacia de las herramientas computacionales, se destaque la capacidad de las nuevas tecnologías para afrontar retos asociados a las señales biomédicas.

1.1. Descripción del Problema

El avance tecnológico en el área médica ha provocado un aumento en la producción y almacenamiento de datos médicos. Analizar estos grandes volúmenes de datos médicos sin apoyo de herramientas computacionales es inviable [14, 59]. Los expertos médicos requieren algoritmos de apoyo que aporten información relevante, como es el caso de la identificación de puntos de interés en señales biomédicas [25]. Un algoritmo es una herramienta computacional con un propósito específico, por ejemplo, la clasificación de un objeto [17, 39].

Sin embargo, al llevar a cabo la implementación de un algoritmo de clasificación debe considerarse una combinación de hardware y software (*i.e.* plataforma computacional) [32]. Entre las plataformas se distinguen la plataforma de cómputo centralizado, la plataforma de cómputo móvil y la plataforma de cómputo distribuido. La plataforma de cómputo centralizado corresponde al cómputo en equipos estacionarios, implementada mayormente por los trabajos en literatura científica de clasificación de

latidos. Por otro lado, la tendencia en el desarrollo de aplicaciones va dirigida hacia la plataforma de cómputo móvil, así como a la plataforma de cómputo distribuido, donde los datos están dispersos entre las estaciones de procesamiento [4, 41, 58]. La variedad de algoritmos y plataformas aporta un abanico de posibilidades para el desarrollo de herramientas de apoyo al experto.

En la literatura de análisis de señales biomédicas, se describen una gran variedad de algoritmos de clasificación de la actividad eléctrica del corazón [49]. Entre los que destacan los algoritmos de clasificación por búsquedas en conjuntos de elementos conocidos [3, 56], ya sea basados en plantillas [16, 40] o utilizando espacios métricos [24, 61]. También se encuentra la clasificación por umbrales, donde dada una serie de reglas se puede identificar un latido o un ritmo cardíaco. Estos algoritmos requieren un amplio conocimiento médico, para determinar el valor de los umbrales [5]. Se encuentra también la clasificación basada en el análisis de señales [16]. Además, la literatura científica menciona la clasificación basada en modelos, que a partir de una etapa de aprendizaje logran la clasificación de latidos [1, 34, 51]. Sin embargo, los algoritmos de clasificación mencionados presentan distintos niveles de precisión [35, 61].

Puntualizar el desempeño de los algoritmos de clasificación requiere de evidencia sólida del comportamiento de estas herramientas de apoyo. La eficiencia de un algoritmo consiste en un estudio teórico, del tiempo consumido por el algoritmo durante su ejecución. Para homologar el cálculo del tiempo de procesamiento, se presupone que el costo computacional de una instrucción es una constante [17, 39]. Por otra parte, la eficacia de un algoritmo de clasificación consiste en contabilizar los aciertos obtenidos por él algoritmo durante su cómputo [10]. La eficiencia y eficacia de un algoritmo de clasificación determinan el tiempo de procesamiento de esta herramienta, así como la capacidad para clasificar correctamente un objeto, respectivamente. El problema que el presente trabajo de tesis aborda es validar si intercambiar de plataforma computacional y variar la cantidad (y distribución estadística) de datos de referencia que

pueden ser procesados por el algoritmo, favorece, o no, a la eficiencia de algunos algoritmos de clasificación y si esto afecta su eficacia. El cómputo de un algoritmo de clasificación puede prolongarse durante días, analizando grandes volúmenes de datos. Donde a pesar de la calidad del resultado del algoritmo, el cardiólogo pudiera requerir una respuesta en un menor lapso de tiempo. De igual manera, existen plataformas de cómputo que no poseen los recursos computacionales suficientes para mantener un periodo de cómputo tan largo. Por otra parte, un algoritmo cuyo cómputo tome una fracción de segundo, pero con una alta incidencia de errores en su respuesta, es una implementación trivial. La implementación de un algoritmo de clasificación debe considerar, tanto la eficiencia, como la eficacia del algoritmo. Los expertos requieren de herramientas computacionales que aprovechen la plataforma computacional en la que son implementadas.

1.2. Motivación

La comunidad científica médica, ha aportado diversos instrumentos para observar el comportamiento del corazón, siendo el Electrocardiograma la elección más utilizada [25, 34, 35]. El *Electrocardiograma* (ECG) es un examen no invasivo, que registra la actividad eléctrica del corazón durante un periodo de tiempo [26]. El ECG contiene una o más señales muestreadas de manera simultanea, llamadas derivaciones [12, 55]. Una derivación permite una vista de la actividad eléctrica del corazón. Debido a que el campo eléctrico generado por el corazón varía espacial y temporalmente, no existe una derivación que aporte un panorama completo de la actividad eléctrica. Es por esto que la cantidad y tipo de derivaciones varía según sea el propósito del ECG. La Figura 1.1 ilustra la actividad eléctrica del corazón en las derivaciones MLII y V1, por un periodo de 30 segundos. Este ejemplo proviene del conjunto de datos MIT-BIH Arritmia [48], cuyo principal propósito es servir como fuente de datos para algoritmos de clasificación de latidos.

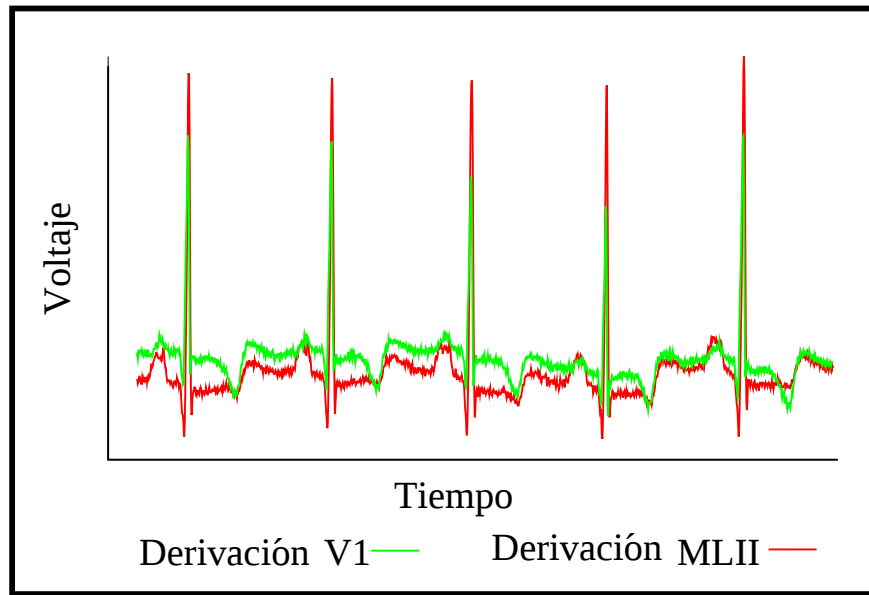


FIGURA 1.1. Trazo de la Actividad Eléctrica.

Motivada por el estudio del corazón, la comunidad científica se ha dedicado a generar almacenes de ECGs [25]. Algunos almacenes de ECGs, además de proporcionar señales biomédicas, aportan información obtenida con ayuda de los expertos [52]. Preservar tanto la información del comportamiento eléctrico, como la interpretación de este por experto, permite a estos almacenes servir como bases de conocimiento para algoritmos de clasificación. Donde un algoritmo de clasificación, aprende a reconocer patrones identificados en el ECG, por expertos [3, 5, 49]. Generar herramientas capaces de reconocer diferentes tipos de latidos cardíacos, aporta al experto en un mayor entendimiento del comportamiento eléctrico del corazón [5].

En el presente estudio, consideramos solamente los algoritmos de clasificación más relevantes en la literatura. Los algoritmos de clasificación presentan los enfoques del área de Cardiología y el área de Ciencias de la Computación. El área de Cardiología busca un clasificador eficaz, interesados en la capacidad de la herramienta computacional para clasificar correctamente [3, 5, 35, 49]. En el caso del área de Ciencias de la Computación, el interés del algoritmo, además de que sea correcto y útil, es en

su eficiencia (*i.e.* uso de recursos computacionales), durante su ejecución [31, 38, 61]. Además, la implementación de los algoritmos de clasificación, se realizó en plataformas computacionales significativas para la comunidad científica. Cada plataforma considerada, posee características computacionales distintas (*e.g.* procesador y memoria) [32], con el objetivo de responder en *cuál* o *cuáles* plataformas es posible aprovechar de mejor manera una herramienta computacional para la clasificación de señales de ECG. Así mismo, estudiamos la eficacia de los clasificadores al variar la cantidad de datos utilizados. Los algoritmos de clasificación implementados, procesan el conjunto de datos MIT-BIH Arritmia [48], el cual es citado con frecuencia por la comunidad científica especializada en la detección y clasificación de anomalías en la actividad eléctrica del corazón.

El resto de esta tesis se organiza de la siguiente manera: En el Capítulo 2 se indica la interpretación que utilizamos para manipular los ECGs. Así mismo, se presenta la revisión de los algoritmos de clasificación reportados en la literatura y sus principales características. Este capítulo describe la evaluación del desempeño de algoritmos de clasificación y presenta las plataformas computacionales más utilizadas en la literatura. El Capítulo 3, presenta las estrategias para formar los conjuntos de datos utilizados en nuestros experimentos. Además de los algoritmos, parámetros y características de los clasificadores y como fueron interpretados y desarrollados. De igual manera este capítulo presenta la implementación y experimentación de las plataformas computacionales en nuestro diseño experimental, que vamos a medir. La experimentación y los resultados obtenidos se describen en el Capítulo 4, donde se observan las capacidades de los clasificadores entre las plataformas computacionales y al variar la cantidad de datos de experimentación. Finalmente nuestra interpretación y conclusiones de los experimentos; así como trabajo futuro es indicada en el Capítulo 5.

CAPÍTULO 2

CLASIFICACIÓN DE ELECTROCARDIOGRAMAS

Analizar el ECG, permite observar patrones de comportamiento específico en la actividad eléctrica del corazón, los cuales representan padecimientos cardíacos (*e.g.* infartos, arritmias). A partir de este análisis, los expertos obtienen una mejor comprensión de padecimientos cardíacos [12, 34, 55, 51]. De igual manera, analizar el ECG puede llevarse a cabo por algoritmos de clasificación de latidos. Un algoritmo de clasificación puede reconocer diferentes anomalías en la actividad eléctrica del corazón a partir de una etapa de aprendizaje. El objetivo del aprendizaje es que al analizar un conjunto de patrones previamente clasificados, el algoritmo ajuste sus parámetros y así establezca una relación entre estos datos y las clases a las que pertenecen [43]. Donde ajustar los parámetros del algoritmo se refiere a modificar los valores de estas variables, buscando encontrar los valores que mejoren la clasificación del algoritmo [3, 6, 49].

El ajuste efectuado en la etapa de aprendizaje se ve afectado por la cantidad de datos utilizados [10, 43]. Un conjunto de datos con pocos ejemplos, difícilmente logra ajustar adecuadamente los parámetros de un algoritmo de clasificación, obteniendo como resultado un clasificador impreciso. Por otro lado, utilizar grandes conjuntos de datos para ajustar los parámetros de un algoritmo de clasificación, puede *sobre-alimentar* al clasificador. Cuando un clasificador es sobre-alimentado, los datos son memorizados y el algoritmo pierde la capacidad de generalizar [37]. Gracias al avance en la identificación de anomalías en el ECG por parte del área de Cardiología [34, 61], así como al almacenamiento de registros de la actividad eléctrica del corazón, es posible conformar conjuntos de datos experimentales. Estos conjuntos de datos se integran de ejemplos de padecimientos cardíacos correctamente clasificados por la

comunidad médica [12, 48, 55]. De esta manera, es posible variar la distribución estadística y cantidad de datos en los conjuntos experimentales, con el fin de obtener un mejor ajuste en los parámetros del algoritmo, evitando un clasificador impreciso o sin capacidad de generalizar.

2.1. Representación del Electrocardiograma (ECG)

Una serie de tiempo es resultado de la observación de un proceso, en cuya ejecución se recolectan mediciones en instantes de tiempo acorde a una frecuencia de muestreo. Formalmente, una serie de tiempo S se define como una secuencia de pares (p_i, t_i) , donde p_i se refiere al i -ésimo valor observado y t_i al i -ésimo índice en el tiempo o *timestamp* para este dato [21, 23].

$$S = ((p_1, t_1), (p_2, t_2), \dots, (p_n, t_n))$$

Así mismo, n indica el total de observaciones o *longitud* de S . La comunidad de series de tiempo destaca entre los retos más significativos el almacenamiento y análisis de estos objetos matemáticos. Entre las propuestas más relevantes para afrontar el almacenamiento de series de tiempo están los vectores característicos [23, 32, 42, 60]. Un vector $\bar{V}$ de n_v dimensiones, caracteriza a una serie S de longitud n , donde $\bar{V}$ se asemeja a S , pero $n_v \ll n$ [21, 42]. Sin embargo, reducir la longitud de una serie de tiempo trae consigo la pérdida información [23, 30, 42].

En el caso del análisis de series de tiempo, la literatura científica presenta amplio trabajo en el análisis de semejanza entre series de tiempo [21, 23]. Este trabajo incluye estrategias para determinar el parecido entre dos series de tiempo a través de una función $Sem(S, T)$ definida para las series S y T , conocida como *función de semejanza* [37, 56, 61]. Además se han implementado algoritmos que calculan la semejanza entre dos series para realizar búsquedas, clasificaciones e indexado, en algunos casos se ha llegado a la implementación en sistemas gestores de base de datos [21, 24, 42, 46].

El ECG es una particularización de una serie de tiempo, donde las mediciones son obtenidas en instantes de tiempo equi-espaciados y el atributo t_i se omite [21]. Por lo tanto, las estrategias de almacenamiento y análisis de series de tiempo pueden ser aplicadas a ECGs. De igual manera, la literatura científica describe una gran variedad de algoritmos para la clasificación de ECGs basados en series de tiempo [6, 30, 35, 42].

2.2. Algoritmos de Clasificación de Electrocardiogramas (ECGs)

El proceso de clasificación relaciona un conjunto de objetos con un conjunto de clases a través de un algoritmo. Cada elemento del conjunto de objetos, es relacionado (*i.e.* clasificado) con al menos un elemento del conjunto de clases. El conjunto de objetos se compone de elementos con diferentes dimensiones, por otra parte, el conjunto de clases está integrado de elementos de naturaleza nominal [10, 15, 37, 43]. La comunidad científica ha propuesto clasificadores que toman ventaja del conocimiento del experto [7, 23, 37, 56], en algunos casos los algoritmos intentan simular este conocimiento [30, 51]. Desarrollar un algoritmo de clasificación, requiere de tres etapas: preprocesamiento, aprendizaje y evaluación [7, 10, 23].

1. *Preprocesamiento*, se refiere a la limpieza de datos. Se definen estructuras para organizar y manipular los datos [22]. De igual manera, se establecen escalas para facilitar el cálculo de valores, *i.e.*, normalización de valores [43].
2. *Aprendizaje*. Esta etapa indica la manera en como se establece la relación entre los datos a clasificar y las clases. El aprendizaje puede categorizarse en dos tipos: *aprendizaje no supervisado* y *aprendizaje supervisado*. Un aprendizaje no supervisado busca patrones en los datos sin establecer *a priori* un conjunto de clases dentro de los datos. Por otro lado, el aprendizaje supervisado requiere de un conjunto de ejemplos donde están identificadas: la clase de cada elemento y las variables que establecen la relación entre el elemento y su clase (*i.e.* variables objetivo) [10, 43].

3. *Evaluación.* En esta etapa se evalúa la relación generada entre un objeto y su clase, establecida en la etapa de aprendizaje. La literatura científica recomienda que el algoritmo debe someterse a una evaluación al finalizar la etapa de aprendizaje, con el objetivo de determinar su eficacia. Para llevar a cabo esta evaluación, debe conformarse un *conjunto de datos de pruebas*, integrado por el mismo tipo de elementos utilizados en la etapa de aprendizaje, pero que no hayan sido analizados previamente [10, 43].

En el caso de la clasificación de latidos en ECGs, los conjuntos de clases y objetos a clasificar corresponden al conjunto de padecimientos cardíacos y al conjunto de ejemplos de latidos cardíacos, respectivamente. La relación entre padecimientos y latidos cardíacos es identificada por un cardiólogo experto. Los clasificadores de ECGs supervisados pueden distinguirse como aquellos basados en instancias y basados en características [7]. Los clasificadores basados en instancias, calculan la clase de un latido, al buscar en un conjunto de latidos clasificados. Los clasificadores basados en instancias implementan una medida de comparación entre el latido (cuya clase es desconocida) y todos aquellos en el conjunto de elementos clasificados [10, 37, 43]. En adelante, nos referimos al latido con clase desconocida como latido desconocido. Definimos a q como el latido desconocido, al conjunto $\mathbb{D}$ el conjunto de referencia del algoritmo y a la función de semejanza $Sem(d_i, q)$ como la función que cuantifica el grado de semejanza entre el latido desconocido q y cada latido $d_i \in \mathbb{D}$. El clasificador determina la clase de q buscando en $\mathbb{D}$ el o los latidos más semejantes d_s a q . Finalmente, q es clasificado con la clase de los latidos d_s .

En cambio, los algoritmos basados en características requieren del conjunto $\mathbb{D}$ y de un algoritmo de entrenamiento que les permita ajustar sus parámetros de clasificación. En una etapa de aprendizaje, el algoritmo realiza un proceso de *entrenamiento* que le permite adquirir conocimiento [10, 15, 37]. El entrenamiento, consiste en ajustar los parámetros del clasificador al analizar cada elemento del conjunto $\mathbb{D}$. El algoritmo de

entrenamiento ajusta los parámetros del clasificador basado en el error obtenido al clasificar cada latido $d_i \in \mathbb{D}$. Donde, grandes errores de clasificación provocan grandes modificaciones y pequeños errores realizan pequeñas modificaciones en los parámetros del algoritmo [9, 10, 43]. Como resultado, los algoritmos basados en características clasifican un latido desconocido q , basado en el conocimiento adquirido en el proceso de entrenamiento.

A pesar de que ambos algoritmos requieren del conjunto $\mathbb{D}$, los algoritmos basados en características no recurren a este conjunto una vez que haya concluido la fase de entrenamiento, es decir, clasifican a q sin consultar $\mathbb{D}$ [9, 10]. Por su parte los clasificadores basados en instancias consultan del conjunto de datos $\mathbb{D}$ cada vez que deba clasificarse un latido cuya categoría se desconoce q [43]. No obstante, los algoritmos basados en instancias no requieren de un proceso de entrenamiento, además de ser necesario pueden modificar la cardinalidad del conjunto de datos $\mathbb{D}$, incorporando o extrayendo elementos sin necesidad de modificar algún parámetro. Por el contrario, los clasificadores basados en características requieren entrenar nuevamente cada que el conjunto de datos $\mathbb{D}$ es modificado [43]. Los ejemplos de mayor relevancia en la literatura para este par de clasificadores son la clasificación por Vecinos más Cercanos y la clasificación por Redes Neuronales Artificiales, como clasificadores basados en instancias y características, respectivamente [23, 37, 49].

2.3. Clasificación por Vecinos más Cercanos

Dado un objeto desconocido b , se conoce como *clasificación por vecinos más cercanos* a la asignación de una clase a b basado en la clasificación de los objetos más cercanos al objeto desconocido, es decir sus vecinos[18]. El algoritmo por excelencia para realizar esta clasificación es k -NN [10, 30, 31, 36, 37, 42, 43, 56]. El algoritmo k -NN es la búsqueda de patrones desconocidos en un conjunto de datos conocidos *a priori* [43]. A partir de encontrar los k objetos más semejantes a un objeto desconoci-

do b en un conjunto de ejemplos $\mathbb{D}$, k -NN clasifica a b por medio de una votación entre las clases de los k objetos [10]. Al realizar la clasificación por vecinos más cercanos se debe tomar en cuenta el número de objetos semejantes al objeto desconocido b , es decir la cantidad de vecinos k [10, 43]. Así como la distribución de los datos en el conjunto de datos conocidos $\mathbb{D}$. Al aumentar k , se corre el riesgo de que exista un sesgo en los resultados del clasificador, un valor de $k = N$; donde N es la cardinalidad de $\mathbb{D}$, provoca sobre-alimentar al algoritmo. Es decir, el algoritmo de clasificación pierde la capacidad de reconocer datos que no se encuentren en $\mathbb{D}$. Lo anterior se debe a que la clase con más elementos en el conjunto de datos $\mathbb{D}$, estará presente entre los k vecinos si el conjunto $\mathbb{D}$ no está equilibrado, *i.e.* cada clase en el conjunto posee el mismo número de elementos [10]. Por el contrario $k = 1$, pudiera llegar a perder capacidad de generalizar sus resultados, ya que solo se considera el elemento más cercano. La literatura científica se refiere como 1-NN a la clasificación considerando solamente un vecino ($k = 1$) [37].

La clasificación de series de tiempo con k -NN, recibe el nombre de *búsqueda por semejanza*. Las búsquedas por semejanza se dividen en: búsquedas completas o búsquedas por subsecuencia [24]. La búsqueda completa se define como: dada una colección de N secuencias $S_1, S_2, \dots, S_N$ y una secuencia de búsqueda Q , encontrar aquellas secuencias que se encuentran a una un umbral de semejanza ϵ de Q . Donde todas las secuencias en la colección, así como la secuencia de búsqueda poseen la misma longitud n . En cambio, las búsquedas por subsecuencias asumen que la longitud de Q es diferente a la longitud de las secuencias en la colección. Para hacer frente a la diferencia en longitud, se analizan las subsecuencias de la secuencia S_i . Definimos una subsecuencia s_{ij} de longitud m como la porción de una serie de tiempo S de longitud n , contemplada por los valores contiguos entre un par de puntos p_i y p_j . Donde $p_i, p_j \in S$, $i \leq j$ y $m = j - i + 1$ [21, 23].

$$s_{ij} = (p_i, \dots, p_j)$$

La búsqueda por subsecuencias se define como: dada una colección de N secuencias $S_1, S_2, \dots, S_N$ y una secuencia de búsqueda Q ; donde $|Q| = m, |S_i| = n, S_i \in S_1, S_2, \dots, S_N, m < n$, encontrar para cada S_i en la colección, aquellas subsecuencias de longitud m que se encuentren a un umbral de semejanza ϵ de Q . Para realizar búsquedas por semejanza, el algoritmo k -NN utiliza una función de semejanza para determina el parecido entre un par de secuencias. Las funciones de semejanza cuantifican el parecido entre dos series de tiempo, lo que permite definir herramientas de análisis *e.g.* búsqueda y clasificación de series de tiempo [24, 37, 46]. Dadas dos series de tiempo T y U , una función de semejanza se define como $Sem(T, U)$, que toma como parámetros dos series de tiempo y obtiene como resultado la semejanza entre estas [15, 21]. La literatura científica del área médica ha calculado la semejanza entre latidos de ECG por medio de funciones estadísticas y por funciones de distancia [11, 16, 21, 37, 40, 44].

Las funciones estadísticas, dependen de un análisis previo de la serie de tiempo para obtener parámetros que les permiten determinar la semejanza con base en la morfología de las series, más allá de su semejanza en escalas [16]. Clasificar latidos de ECG a través de funciones estadísticas es conocido por la comunidad de Cardiología como *comparación por plantillas (template matching)*, utilizando el *Coefficiente de Correlación* como función de semejanza [40, 44]. En el caso de las funciones basadas en distancia, la semejanza entre series de tiempo se calcula por la distancia entre ellas [31, 35, 38]. Al utilizar una función de distancia se asume que la semejanza entre un par de series es inversamente proporcional a la distancia entre estas, es decir, si la distancia entre dos series es 0 estas series son idénticas [24] Donde la función de distancia más citada por la literatura es la Distancia Euclidiana, conocida también como la Norma L_2 [37, 61]. El clasificador 1-NN es la alternativa directa para evaluar la eficiencia entre diferentes funciones de semejanza, ya que a través de la búsqueda de vecinos se refleja la eficacia de la función de semejanza implementada [21, 23, 37].

2.3.1. Función de Distancia Euclidiana

La literatura científica reconoce a la función de distancia Euclidiana como una alternativa eficaz para determinar la semejanza entre dos series de tiempo [7, 23, 24, 37, 56]. La distancia Euclidiana compara un par de vectores $\vec{a}$ y $\vec{b}$ de n dimensiones, calculando la distancia entre ambos como la raíz cuadrada de la suma cuadrada de las diferencias entre la i -ésima dimensión de $\vec{a}$ y la i -ésima dimensión de $\vec{b}$. Dadas dos series T y U con longitud n , esta función se aplica interpretando las series de tiempo como vectores de dimensión n , donde t_i y u_i representan la i -ésima observación de T y U , respectivamente (Ecuación 2.1).

$$E(T, U) = \sqrt{\sum_{i=1}^n (t_i - u_i)^2} \quad (2.1)$$

La función posee un rango que va de 0 hasta ∞ , donde una distancia $E(T, U) = 0$ indica que ambas series son idénticas. En el caso de que $E(T, U) > 0$, las series no son idénticas. La distancia Euclidiana es una métrica, también llamada Norma L_2 [61], se considera métrica debido a que posee tres características: no negatividad, es simétrica y subaditiva [21, 23, 24, 37, 61].

- No negatividad, $E(T, U) \geq 0$, se refiere a que la distancia entre dos series de tiempo no puede ser de signo negativo.
- Simétrica, $E(T, U) = E(U, T)$, la distancia entre un par de series de tiempo T y U será igual en orden inverso.
- Subaditiva, $E(T, U) \leq E(T, V) + E(V, U)$, también llamada desigualdad triangular, considerando tres series de tiempo distintas T , V y U , la distancia entre cualquier par será menor o igual a la suma de las distancias entre las permutaciones restantes.

Las métricas facilitan definir índices en series de tiempo, lo que las hace adecuadas para su uso en consultas en grandes bases de datos [15, 24, 61]. La función de

distancia Euclidiana, presenta ciertas ventajas que la hacen una alternativa eficiente. El cómputo de la distancia Euclidiana es directo y no requiere de algún parámetro estadístico. Así mismo, en una clasificación con k -NN, el cómputo de esta distancia puede omitir calcular la raíz cuadrada (Ecuación 2.1) sin afectar el cálculo de la semejanza [24, 37]. De igual manera, la clasificación k -NN puede implementar un *upper-bound*, para acelerar el cálculo de la semejanza. De manera que se tiene conocimiento de la mejor semejanza encontrada, en caso de superar esa distancia la clasificación es interrumpida y se sigue con el próximo patrón [37].

Por otra parte, la función de distancia Euclidiana debe ser usada considerando que es sensible a ruido y diferencia de escalas entre las series de tiempo, debido a que su cálculo considera las diferencias entre cada i -ésimo valor de T y U . Es recomendable que los valores de las series de tiempo estén bajo la misma escala, como es el caso de las señales de ECG [26, 48].

2.3.2. Coeficiente de Correlación Cruzada

La semejanza entre ECGs ha sido calculada a través de operaciones estadísticas, que toman en cuenta la relación entre los valores en la serie [56]. Entre las operaciones estadísticas más citadas se encuentra el Coeficiente de Correlación Cruzada [16, 40, 44]. El Coeficiente de Correlación Cruzada ($CCI()$, Ecuación (2.2)) requiere de tres parámetros estadísticos (*i.e.* Media aritmética, Covarianza y Varianza). Dadas dos series de tiempo T y U (ambas de longitud n), el índice de correlación entre este par de series $CCI(T, U)$ corresponde a la división de la covarianza entre T y U (Ecuación 2.4), entre la raíz cuadrada del producto de la varianza de T por la varianza de U (Ecuación 2.3). La covarianza entre T y U resulta de la suma del producto de las diferencias entre cada i -ésimo valor de la series con respecto a su media aritmética, dividido entre n . Donde t_i y u_i indican el i -ésimo valor de la serie T y la serie U , respectivamente. Igualmente, μ_T y μ_U corresponden a la media aritmética de T y U ,

respectivamente.

$$CCI(T, U) = \frac{covar(T, U)}{\sqrt{var(T) var(U)}} \quad (2.2)$$

$$var(T) = \frac{\sum_{i=1}^n (t_i - \mu_T)^2}{n} \quad (2.3)$$

$$covar(T, U) = \frac{\sum_{i=1}^n (t_i - \mu_T)(u_i - \mu_U)}{n} \quad (2.4)$$

El Índice de correlación indica el sentido en como dos series están correlacionadas y comprende un rango de valores entre -1 y 1 . Dadas las series de tiempo T y U , si el resultado de $CCI(T, U)$ es igual a -1 , esto indica que cuando un los valores en T incrementan, los valores en U decrementan en igual magnitud. Si el resultado de $CCI(T, U)$ es igual a 1 , esto se interpreta como al incrementar los valores en T , los valores en U a su vez incrementan en igual magnitud. Por otra parte, si el resultado del índice es 0 , esto indica que las series no poseen ninguna correlación entre sus formas [56]. Existen trabajos en la literatura científica que basan su experimentación en el cálculo de CCI , donde se afirman que rebasan la eficacia de otras alternativas [40, 44]. No obstante, a diferencia de la distancia Euclidiana, esta función no puede ser considerada una métrica. A pesar de que $CCI()$ corresponde a una función simétrica y que su rango de valores puede ser normalizado entre 1 y 2 , esta función no cumple con la característica subaditiva [20]. Por lo que no es posible basar un índice en su cálculo, y esto lo vuelve una alternativa poco atractiva para su uso en grandes bases de datos [24].

2.4. Redes Neuronales Artificiales

Una Red Neuronal es un clasificador compuesto por unidades de procesamiento, (nodos) organizados en capas interconectadas, donde cada nodo posee una entrada y salida de datos. La literatura científica especializada en redes neuronales distingue

tres tipos de capas distintas, relacionadas con su funcionalidad: capa de entrada, intermedia y salida. La *capa de entrada* en una red neuronal es el medio por el cual el clasificador obtiene el objeto que será clasificado. La *capa intermedia* (también conocida como *capa oculta*), representa los parámetros ajustables del clasificador. Se conoce como capa oculta debido a que al clasificar un objeto, no se tiene interacción con esta capa en particular. Finalmente, la *capa de salida*, indica el resultado calculado por este clasificador, donde dependiendo de la implementación de la red, puede tratarse de una estimación o una clasificación [9, 10, 43].

Las conexiones entre nodos son denominadas *pesos*, estas ponderan el vínculo entre un par de nodos e indican el tipo de procesamiento de la red neuronal [9, 10, 45]. Al variar la cantidad de capas y nodos en la red neuronal, la precisión del clasificador varía también [30, 51]. La literatura científica indica que establecer una estructura de tres capas (*i.e.*, una capa de entrada, una capa oculta y una capa de salida), al igual que generar conjuntos de datos donde las todas las clases se encuentren representadas, genera una red con eficacia aceptable [6, 34, 51].

En la Figura 2.1 se ilustra un ejemplo de una red neuronal que tiene una estructura de tres capas, capa de entrada (nodos a y b), capa oculta (nodos c , d y e) y capa de salida (nodos f y g). Para relacionar cada capa de la red, se definen pesos entre los nodos, *i.e.* W_{ac} indica la fuerza de la conexión entre el nodo a en la capa de entrada y el nodo c en la capa oculta.

Este clasificador requiere de una normalización de los datos [51]. Decimos que los datos están normalizados cuando sus valores están acotados dentro de un rango establecido entre un valor mayor y menor. Homaeinezhad *et ál.* [30], sugieren que además de normalizar los datos de entrada, deben normalizarse las frecuencias de las series de tiempo, *i.e.* que todas las series de tiempo a analizar hayan sido registradas utilizando la misma cantidad de muestras por segundo, en caso de poseer datos de fuentes distintas [30].

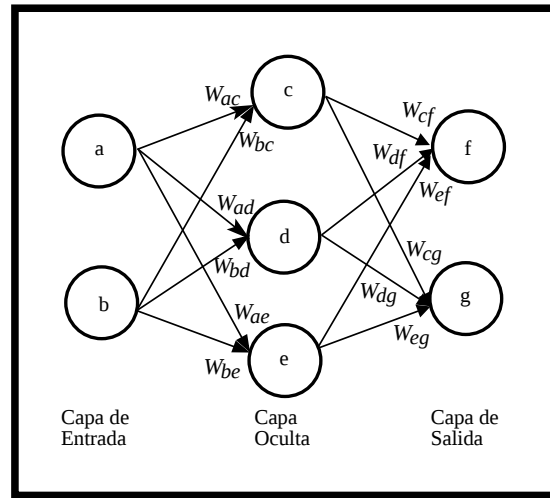


FIGURA 2.1. Estructura de una Red Neuronal Artificial.

2.4.1. Cálculo y Clasificación

Existen diferentes estructuras de redes neuronales, donde la más destacada corresponde a la estructura *Feed-Forward* [10, 43, 45]. El termino *Feed-Forward*, indica la dirección del cálculo realizado por la red neuronal *i.e.* recibe su entrada en la capa de entrada, procesa estos datos a través de las capas intermedias e indica su resultado en la capa de salida. La salida de una red neuronal es calculada al combinar la entrada de la red entre cada una de sus capas, donde cada nodo en una capa posee una conexión con todos los nodos de la capa siguiente.

Inspiradas en la estructura y comportamiento que tienen las redes de neuronas en el cerebro humano, una red neuronal artificial posee una función que concentra las entradas de una neurona y produce una salida, conocida como función de activación [43]. La función de activación determina la colaboración entre una neurona y otra. La literatura ha presentado diferentes funciones de activación, donde cada una determina un comportamiento distinto para el clasificador [6, 30]. La función de activación más usada por la comunidad científica es la Función Sigmoidea (Ecuación (2.5)), esta función toma como entrada un valor real y regresa una salida acotada

entre 0 y 1 [9, 10, 43].

$$h(a) = \frac{1}{1 + e^{-a}} \quad (2.5)$$

En (2.5), a representa un valor real y e se refiere a la base del logaritmo natural. El resultado de una red neuronal es representado por todas las neuronas en la capa de salida. La interpretación de este resultado puede darse al utilizar umbrales, una codificación binaria, una neurona por clase [9, 10, 43].

- Umbrales, Es un alternativa común para realizar identificación de múltiples clases a través de una sola neurona en la capa de salida. El uso de umbrales consiste en acotar el resultado de la neurona de salida Y_i entre distintos rangos para determinar múltiples clases, *e.g.* $0 \leq Y_i < 0.5$ clasifica con clase A y $0.5 < Y_i < 1$ clasifica con clase B .
- Codificación binaria, esto consiste en definir $\log_2(clases)$ neuronas, donde *clases* es la cantidad de clases que la red neuronal clasificará, *e.g.* $Y = [1, 1, 0]$, equivale a la clase $2^0 + 2^1 = 3$ de 2^3 clases, al ser 0 y 1 los bits encendidos. Esta alternativa es común cuando se clasifican tres o más clases.
- Neurona por cada clase, esta alternativa utiliza tantas neuronas como clases se haya que reconocer, *e.g.* $Y = [0, 1, 0]$, reconoce la clase 2 de tres posibles clases.

En el Algoritmo 1 se describe el cálculo de la salida de una red neuronal, que recibe como entrada un objeto desconocido X , interpretado como una secuencia de longitud n que será clasificado, además de una red neuronal $\mathcal{R}$ previamente entrenada. Se copia el contenido de la secuencia X de n elementos, en un vector *entrada* con n dimensiones que representa la capa de entrada de la red $\mathcal{R}$, con n neuronas. Se realiza un ciclo, desde la primer capa oculta de la red $\mathcal{R}$, hasta la capa de salida de esta; donde cada capa será referida como c_j . Por cada capa c_j analizada, se efectúa un ciclo para analizar cada nodo en la capa, donde cada nodo es ubicado como c_{ji} *i.e.* el i -ésimo nodo en la j -ésima capa. Para cada iteración en un nodo c_{ji} , se inicializa la

variable $acum$ en 0 y se realiza una iteración desde $k = 0$ hasta n (elementos en el vector). Para cada elemento en el vector $entrada$ se acumula en $acum$ el resultado de multiplicar el k -ésimo elemento de $entrada$ por el peso entre el nodo en la capa oculta c_{ji} con el k -ésimo nodo de la capa representada por el vector $entrada$, referido como W_{jik} . Al terminar el ciclo para todos los elementos de $entrada$, se calcula la salida del nodo c_{ji} utilizando la función de activación y guardando el resultado como $salida_{ji} = h(acum + W_{ji0})$, donde W_{ji0} indica el aprendizaje del nodo. Al terminar el ciclo para todos los nodos en la capa c_j , se genera un nuevo vector $entrada$, con tantos elementos como nodos en la capa c_j y se copia el contenido de $salida_j$ en $entrada$, el proceso continua, hasta analizar el resto de las capas en $\mathcal{R}$. Una vez obtenido el vector $salida_j$ de la capa de salida en $\mathcal{R}$, se interpreta utilizando alguna de las estrategias de clasificación antes mencionadas, para obtener como resultado la clase del objeto X [43].

Algoritmo 1: Algoritmo para calcular la salida $\mathcal{R}ed(x)$

Entrada: X , vector de entrada de la red,
 $\mathcal{R}$, red neuronal previamente entrenada
Salida: y Clase calculada para x

Guardar $n = dimensiones(X)$
 Crear el vector $entrada$ con n elementos.
 Desde la primer capa oculta c_j hasta la capa de salida,
 Para cada nodo c_{ji} en c_j ,
 Inicializar $acum \leftarrow 0$.
 Para cada elemento x_k en el vector $entrada$,
 Actualizar $acum \leftarrow acum + (x_k \times W_{jik})$
 Fin_Desde
 Calcular $salida_{ji} \leftarrow h(acum + W_{ji0})$.
 Fin_Desde
 Actualizar $n = nodos(c_j)$.
 Crear $entradas$ con n elementos.
 Copiar $salida_j$ a $entrada$.
 Fin_Desde
 Regresar $interpreta(salida_j)$

2.4.2. Algoritmo de Entrenamiento

Como parte del entrenamiento de la red neuronal, la literatura cita el algoritmo de *Retropropagación* (o *Backpropagation* por su nombre en inglés) [10, 34, 43, 45, 51]. Este algoritmo actualiza los pesos entre los nodos a partir de la diferencia entre la salida obtenida por la red y la salida esperada. Es llamado de *Retropropagación* debido a que la modificación de los valores de los pesos inicia en la capa de salida y termina en la capa de entrada de la red. Las entradas del algoritmo son la *RNA* a modificar y el vector q que representa el objeto que debe ser clasificado correctamente.

En el Algoritmo 2, se modifican todos los pesos W de cada capa y cada nodo en la red, desde la última hasta la primera capa. El resultado son la actualización de los pesos W' y la actualización del conocimiento del nodo $bias'$. El ciclo inicia calculando δ_j de cada nodo j . El cálculo de δ_j corresponde a la responsabilidad del nodo j en la salida de la red, donde la salida del nodo j se denota $salida_j$. Así mismo, la clase a la que pertenece el patrón q , se representa como un vector de tantas dimensiones como neuronas en la capa de salida $clase_j(q)$. Para realizar el cálculo de δ_j para todos los nodos de las capas ocultas, el algoritmo se basa en las responsabilidades de los nodos a los que el nodo j se conecta (*i.e.* , nodos dependientes). Una vez calculado el valor de δ para cada nodo, se determina el nuevo peso nuevo W' , resultado de sumar las correcciones del peso ΔW al peso actual W .

$$\delta_j = \begin{cases} salida_j(1 - salida_j) & (clase_j(q) - salida_j) & \text{Nodos en capa de salida.} \\ salida_j(1 - salida_j) \sum_{dependientes} W_{jk} \delta_k & & \text{Nodos en capa oculta.} \end{cases} \quad (2.6)$$

$$\Delta W_{kj} = \delta_j \times salida_k \quad (2.7)$$

Algoritmo 2: Algoritmo de Backpropagation

Entrada: RNA Red neuronal,

q Patrón de aprendizaje

Inicializar W' y $bias'$ basados en la estructura de red , si no han sido inicializados.

Desde c_i igual a la última capa de red , hasta la primer capa oculta de red .

Crear δ_i con la misma cantidad de elementos que c_i .

Colocar en S_i las salidas de c_i .

Para cada nodo j en c_i , hacer:

Si c_i es capa de salida, $\delta_j = S_i[j] \times (1 - S_i[j]) \times (q[j] - S_i[j])$

Si no, entonces

Para todos los nodos k de c_{i+1} , acumular el resultado de $c_{i+1}W[k][j] \times \delta_{i-1}[k]$ en sum .

Calcular $\delta_j = S_i[j] \times (1 - S_i[j]) \times sum$.

Para todos los nodos k de c_{i+1} , $\Delta W_{kj} = \eta \times \delta_{i-1}[k] \times S_i[j]$ y

$W'_{c_{i+1}}[k][j] = c_{i+1}W[k][j] + \Delta W_{kj}$

Fin si.

Calcular $bias'_{c_i}[j] = c_i bias[j] + \Delta W_{j0}$, donde $\Delta W_{j0} = \eta \times \delta_j \times 1$

Fin para.

Hacer $\delta_{i-1} = \delta_i$.

Fin desde.

Obtener S_0 , como las salidas de la capa de entrada c_e .

Para todos los nodos j de la primer capa oculta c_i , hacer:

Calcular $W'_{c_e}[k][j] = c_e W[k][j] + \Delta W_{kj}$, para cada nodo i de c_i

donde $\Delta W_{ji} = \eta \times \delta_{i-1} \times S_0[i]$

Fin para

Copiar W' y $bias'$ a red .

2.4.3. Parámetros de Optimización

La red neuronal, presenta una gran carga computacional en la etapa de entrenamiento, compuesta por gran cantidad iteraciones de operaciones aritméticas realizadas. Debido a que el objetivo de estas operaciones es lograr que el error converja en el menor valor posible; permitiendo que el clasificador aprenda de los datos, la comunidad científica ha propuesto distintos parámetros de optimización, que permiten lograr este objetivo en un menor número de iteraciones.

Los parámetros más utilizados por los investigadores son el coeficiente de aprendizaje (expresado con el símbolo η) y el coeficiente de *momentum* (α). El coeficiente de aprendizaje permite potenciar el aprendizaje de cada neurona en la red, al actualizar el valor del peso. Por otra parte, el coeficiente *momentum*, permite potenciar el aprendizaje de la red basado en su desempeño en iteraciones anteriores. Ambos coeficientes poseen un rango de valores entre 0 y 1. En los trabajos de Ozbay y Karlik [51] y Jadhav *et ál.* [34], se indica el uso de ambos coeficientes para acelerar el proceso de aprendizaje de la *RNA*, en este último los valores de los parámetros fueron $\eta = 1$ y $\alpha = 0.5$.

2.5. Plataformas Computacionales

El punto de referencia de nuestros experimentos corresponde al desempeño de los algoritmos de clasificación en la plataforma computacional centralizada y estacionaria. Que en la literatura científica enfocada a la clasificación de latidos, se refiere a equipos de cómputo tipo Escritorio. De igual manera, estos experimentos generalmente implementan un solo origen de datos, *i.e.* los datos provienen de una sola fuente, sin referirse a problemas de integración o distribución de información. Una plataforma computacional con estas características, representa nuestro caso base para determinar si se observa un cambio significativo en eficacia o eficiencia, en el resto de las plataformas computacionales. Para contrastar los resultados de nuestro punto de referencia,

implementamos la misma experimentación en una plataforma computacional distinta y otro tipo de origen de datos distinto. El cómputo móvil se utiliza debido a su portabilidad. Dicha portabilidad es una ventaja cuando se cuenta con equipo que produce los datos en el momento, lo que nos lleva a un esquema de clasificación en línea. Esta situación se puede dar por un equipo de cómputo móvil (*teléfono inteligente*) y una interfaz bluetooth para su comunicación. No obstante, a diferencia de nuestro punto de referencia, una plataforma móvil no posee una fuente de alimentación constante, en su lugar depende de una batería que tiene un periodo de carga. Además que cada instrucción realizada consume energía. En la plataforma de cómputo móvil es importante considerar algoritmos que realicen la menor cantidad de operaciones de cálculo y de entrada—salida, ya que representan el mayor gasto energético del algoritmo. Estas operaciones corresponden a calcular un resultado y leer y/o escribir datos en algún dispositivo, respectivamente. Algoritmos que con un bajo consumo energético, podrían alargar el uso de la batería del dispositivo donde se encuentren implementados.

Además de la plataforma móvil, evaluamos el uso de un tipo de origen de datos distinto a nuestro punto de referencia, datos distribuidos. Estas plataformas computacionales tienen un margen de contraste, ya que va de la plataforma con mayor capacidad computacional, hasta aquella con los recursos computacionales limitados. Este contraste permite realizar un estudio que evalúe el desempeño de los algoritmos de clasificación de ECGs, en función de develar si existe o no alguna característica; en estas plataformas computacionales, que favorezca a algún algoritmo de clasificación en particular.

2.6. Desempeño Algorítmico

Establecer el desempeño de un algoritmo puede realizarse por medio de un estudio teórico, considerando su *complejidad temporal* y *complejidad espacial*. Estas

características se refieren a la cantidad de instrucciones ejecutadas y la cantidad de memoria requerida por el algoritmo durante su cómputo, respectivamente. Este estudio teórico establece una cota asintótica que describe el consumo de recursos de un algoritmo (*i.e.* tiempo y memoria) al incrementar el tamaño del problema [39, 17]. El cálculo de complejidad presupone que la tecnología de implementación (*e.g.* lenguaje de programación y plataforma computacional) representa un costo computacional constante, con el objetivo de simplificar el cálculo de dicho costo [47]. No obstante, al incrementar el tamaño del problema, estos costos constantes afectan el desempeño en la ejecución del algoritmo. Determinar que tan adecuada es una plataforma con respecto a otra basado en un estudio de complejidad, oculta el efecto que estos costos constantes pudieran tener en la eficiencia del algoritmo. De esta manera, a pesar de ser una clara evidencia del desempeño en un algoritmo, la *complejidad temporal* y la *complejidad espacial* no son criterios suficientes para distinguir las ventajas de una plataforma computacional para un algoritmo de clasificación dado.

Por otro lado, los algoritmos de clasificación han sido evaluados por la literatura en base a su eficacia, determinando como mejor a aquel con mayor eficacia [35, 30, 56]. Evaluar el desempeño de un algoritmo de clasificación a través de un enfoque estadístico (*i.e.* enfocando atención solamente por la precisión observada en la experimentación con el algoritmo [19]), limita la evaluación al ignorar las características operacionales de éste. Debe definirse una evaluación que considere la eficacia del algoritmo de clasificación, así como el aprovechamiento de dicho algoritmo en una plataforma computacional determinada. Donde el desempeño total vendrá dado por el uso combinado de hardware y software [17]. De igual manera, se requiere de un estudio experimental que evalúe la eficacia del algoritmo de clasificación, así como una serie de experimentos con cantidades que permitan evidenciar que las constantes supuestas, en todos los ambientes computacionales, en un estudio de complejidad, pueden eventualmente mermar el cómputo del algoritmo volviéndolo no apto para algún ambiente computacional determinado.

CAPÍTULO 3

IMPLEMENTACIÓN DE ALGORITMOS DE CLASIFICACIÓN

El presente estudio analiza la eficacia de el algoritmo de vecinos más cercanos y redes neuronales artificiales, en la clasificación de latidos cardíacos. Estos algoritmos corresponden a los clasificadores más relevantes en la literatura científica [23, 37, 49]. Los clasificadores son evaluados por la eficacia obtenida al identificar 14 tipos de latidos distintos, así como por su eficacia en una clasificación de latidos dicotómica. Los 14 tipos de latidos están incluidos en el conjunto de datos de prueba que se describen en la siguiente sección. Por otra parte, la clasificación dicotómica agrupa el conjunto de datos en dos clases: latidos *normales* y latidos *anormales*, que corresponden a latidos sin ningún padecimiento cardíaco y latidos con algún tipo de cardiopatía, respectivamente.

La eficacia y eficiencia de los clasificadores son evaluadas a partir de un conjunto de datos reconocido por la literatura y el uso de plataformas computacionales y esquemas de procesamiento distintos . El conjunto de datos corresponde a MIT-BIH Arritmia, una colección de ECGs digitales utilizada por la comunidad científica para la validación de algoritmos de clasificación de latidos cardíacos [48]. Además, consideramos las plataformas de *Cómputo de Escritorio* y *Cómputo Móvil*, así como los esquemas de procesamiento *Centralizado* y *Distribuido*. La evaluación realizada, toma en cuenta la eficacia de los clasificadores al utilizar distintas estrategias para variar la cantidad y distribución de datos en los conjuntos de latidos utilizados.

3.1. Datos de Prueba

Los datos que serán analizados por los algoritmos de clasificación provienen del conjunto de datos MIT-BIH Arritmia [48], que es parte del proyecto PhysioNet [25]. El conjunto de datos MIT-BIH Arritmia posee registros de ECG donde se encuentran anotaciones de distintos ejemplos anomalías en la actividad eléctrica del corazón *e.g.* diferentes tipos de latidos y ritmos cardíacos. Al extraer todos los latidos cardíacos en el conjunto de datos, es posible conformar conjuntos de entrenamiento y pruebas para evaluar la eficacia de los clasificadores. Con el fin de observar cambios en la eficacia y eficiencia de los clasificadores, decidimos variar las clases, así como la distribución y cantidad de elementos al conformar los conjuntos de entrenamiento. Para realizar las variaciones, definimos dos problemas de clasificación considerando dos y 14 clases, al igual que dos estrategias de selección de elementos: Selección Aleatoria y Selección Equi-proporcional.

3.1.1. Descripción del Conjunto de Datos

El conjunto de datos MIT-BIH Arritmia se compone de 48 ECGs digitales provenientes de 47 pacientes, divididos en 25 hombres con edades entre 32 y 89 años y 22 mujeres con edades entre 23 y 89 años. Cada ECG digital en el conjunto de datos posee dos derivaciones, cada una capturada a 360 muestras por segundo, durante un periodo de 30 minutos. Cada ECG digital contiene 650,000 observaciones por derivación, almacenadas en una codificación de 11 bits por observación, con un rango en amplitud de $\pm 5\text{mV}$ [48]. El conjunto de datos MIT-BIH Arritmia requiere un espacio de almacenamiento de 91.7 MB, donde cada uno de los 48 ECGs está relacionado con tres tipos de archivo: archivos de tipo señal, cabecera y anotaciones.

- Archivo de señal (extensión `.dat`), contiene la información del ECG codificada en formato binario con una resolución de 11 bits por valor. Cada uno de estos archivos requiere un espacio de almacenamiento de 1.9 MB.

- Archivo de cabecera (extensión `.hea`), corresponde a información sobre el registro de ECG (*e.g.* cantidad y tipo de derivaciones, número de muestras por segundo, género del paciente). Cada archivo de cabecera requiere 4 KB de espacio de almacenamiento.
- Archivo de anotaciones (extensión `.atr`), almacena el tipo e instante de tiempo donde ocurrió una anomalía cardíaca en el ECG, es decir, un latido o alguna cardiopatía. Estos archivos poseen tamaños distintos, debido a que la cantidad de anomalías detectadas para un ECG varía. El espacio de almacenamiento requerido por este tipo de archivos va desde 4 KB hasta 12 KB.

Por medio de la relación de un ECG en MIT-BIH Arritmia con estos tres tipos de archivos, es posible conformar un conjunto de ejemplos que facilite el aprendizaje de un algoritmo de clasificación. Los conjuntos de ejemplos estarán compuestos de todas las anomalías detectadas en los archivos de anotaciones en cada uno de los ECGs.

3.1.2. Distribución de Datos

Para realizar los experimentos, se utilizó el conjunto de datos MIT-BIH Arritmia descrito anteriormente. No obstante, se requirió de una etapa de preprocesamiento de datos, previo a la experimentación. En la etapa de preprocesamiento se utilizó el paquete de software WaveForm DataBase Software Package (WFDB), disponible en PhysioNet [53]. La etapa de preprocesamiento requirió de las aplicaciones `rdsamp` y `rdann`, la primera permite extraer los valores del ECG a partir de un archivo de señal del conjunto MIT-BIH Arritmia (que se encuentra codificado a 11 bits). Por otra parte la aplicación `rdann`, extrae las anotaciones de un ECG a partir del archivo de señal y el archivo de anotación. La etapa de preprocesamiento se divide en la extracción de las derivaciones y anotaciones del conjunto de datos, así como la transformación de los datos a un formato que facilite la futura experimentación. Finalmente, el último

paso corresponde a la extracción de los patrones de requeridos para los algoritmos de clasificación.

1. *Extracción de derivaciones.* Se utilizó la aplicación **rdsamp** para obtener y almacenar los valores (en mV) del ECG en formato ASCII. Esta aplicación, analiza los archivos de señales del conjunto de datos MIT-BIH Arritmia y extrae los valores en escala de mV de cada una de las observaciones, para cada derivación del ECG. El resultado de este proceso es un par de archivos en formato ASCII, uno para cada derivación en el ECG.
2. *Extracción de anotaciones.* Se utilizó la aplicación **rdann** para extraer todos los índices y tipos de anomalías de todos los ECGs en el conjunto de datos MIT-BIH Arritmia. Este proceso requirió de los archivos de anotación. Mediante los archivos tipo señal y anotación de MIT-BIH Arritmia y la aplicación **rdann**, se extrajeron las anotaciones para cada registro. Por cada registro en MIT-BIH Arritmia, se obtuvo un archivo de anotaciones en formato de texto compuestas del tipo de evento reconocido y su posición en el ECG. La posición presente en el archivo de anotación, corresponde a la parte central del latido identificado.
3. *Transformación de derivaciones.* Tomando los datos cada uno de los archivos de derivaciones extraídas, que tienen un formato de texto, se realizó un proceso de codificación a un formato binario, cuyo resultado es un archivo binario, donde cada observación del ECG es de tipo flotante y ocupa un espacio de 4 bytes. Este formato se utiliza para evitar realizar la decodificación de 11 bits durante la experimentación con los algoritmos de clasificación.
4. *Extracción de patrones a partir de anotaciones.* Debido a que el índice de un evento en el archivo de anotaciones corresponde a la posición central de dicho evento, se obtienen 45 muestras de la parte izquierda y 45 muestras de la parte derecha, donde sumados estos valores representan $\frac{1}{4}$ del total de muestras por

segundo implementadas durante la captura del ECG. Esta cantidad de muestras es suficiente para representar el inicio y fin el evento señalado en el archivo de anotaciones [57]. El resultado de este proceso es un archivo con todas las subsecuencias producidas, donde cada valor se registra como un dato flotante de 4 bytes.

El resultado de la etapa de preprocesamiento es un conjunto de referencia con 105,056 puntos de interés, del cual se crearon el conjunto de datos de entrenamiento $\mathbb{D}$ y el conjunto de datos de prueba $\mathbb{Q}$. Consideramos una muestra de 100,000 elementos del conjunto de datos globales y basándonos en la distribución de las 14 clases en el conjunto de datos global, generamos un conjunto de entrenamiento $\mathbb{D}$ que refleja la naturaleza de los datos. Por otra parte, el conjunto $\mathbb{Q}$ fue integrado con todos los elementos del conjunto global que no se encuentran en el conjunto $\mathbb{D}$. Las cardinalidades son de 97,388 elementos en $\mathbb{D}$ y 7,668 elementos en $\mathbb{Q}$.

3.1.3. Selección de Patrones

Debido a que una derivación de un ECG representa la vista de un plano de la actividad eléctrica del corazón, una anotación observada en diferentes derivaciones posee un comportamiento distinto [26]. Por lo tanto, al desarrollar un algoritmo de clasificación cuya entrada es una serie de tiempo que representa la señal de ECG, debe considerarse un solo tipo de derivación del cual se extraerán los patrones. Cada ECG digital en el conjunto MIT-BIH Arritmia posee dos derivaciones distintas, entre las que se encuentran cuatro derivaciones precordiales (V1, V2, V4, V5) y la derivación modificada II (MLII) [26]. En la Figura 3.1, se ilustra la cantidad de señales de cada una de las derivaciones en el conjunto de datos, aquí es posible observar que la derivación MLII cuenta con el mayor número de señales, seguida por la derivación V1, con 46 y 41 señales, respectivamente.

Sin embargo, la cantidad de anotaciones entre cada señal del conjunto de datos

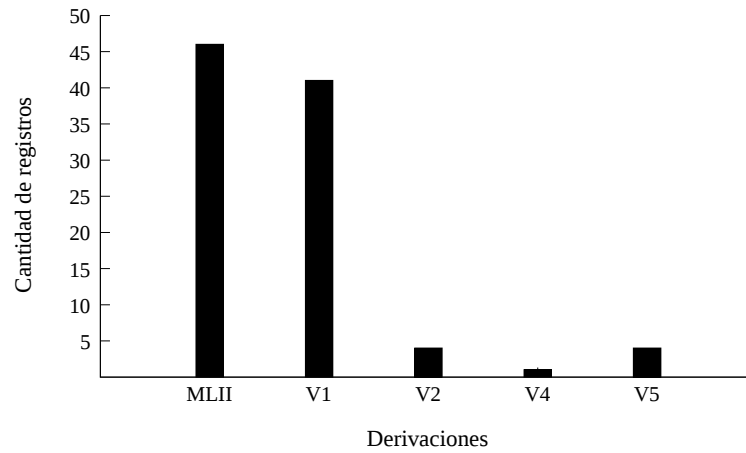


FIGURA 3.1. Ocurrencia de derivaciones en la base de datos MIT-BIH

varía, por lo que fueron obtenidas todas las anotaciones capturadas. El total de anotaciones en el conjunto de datos MIT-BIH Arritmia, suma 225,224 anomalías cardíacas. En la Tabla 3.1, se indica la distribución de las anotaciones entre cada una de las derivaciones, donde la derivación con mayor cantidad de anomalías es MLII con 108,107 ejemplos, que representa el 48 % del total de anomalías detectadas en el conjunto de datos MIT-BIH Arritmia. Basados en la cantidad de señales y anotaciones, elegimos extraer los patrones de la derivación MLII.

TABLA 3.1. Distribución de Anotaciones por Derivación en el Conjunto de Datos

Derivación	MLII	V1	V2	V5	V4	Total
Anotaciones	108,107	97,165	8,131	10,187	1,634	225,224

Al analizar las anotaciones de la derivación MLII, se observaron 23 tipos de anomalías, identificadas por números enteros, con un rango de 1 a 23. La Tabla 3.2, presenta la distribución de cada uno de los tipos de anomalías identificados en la derivación MLII, indicando el identificador para la anomalía (columna *Clase*), el tipo de anomalía en el conjunto de datos (columna *T.A.*) y la cantidad de anotaciones para anomalía *i.e.* el total de ocurrencias de cada anomalía (columna *Total*).

Cada tipo de anomalía identificada representa un evento de interés para los exper-

TABLA 3.2. Distribución de los T.A.s en la Derivación MLII

Clase	T.A.	Total	Clase	T.A.	Total	Clase	T.A.	Total
1	N	74,784	9	F	803	17	]	6
2	L	8,075	10	e	16	18	x	193
3	R	7,259	11	j	229	19	Q	15
4	A	2,546	12	E	106	20		132
5	a	150	13	/	3,620	21	~	578
6	J	83	14	f	260	22	+	1,212
7	S	2	15	[	6	23	”	437
8	V	7,123	16	!	472			

tos [48, 52]. No obstante, de las anomalías identificadas en la Tabla 3.2, se ignoraron aquellas que no indican la clasificación de un latido cardíaco [52]. Las anomalías ignoradas corresponden a las clases $\{15, 16, 17\}$, $\{18, 21, 22, 23\}$, $\{19, 20\}$, separadas en tres grupos distintos de anotaciones que representan marcas de ritmo cardíaco (*i.e.* inicio, fin y duración), tipos de ritmos cardíacos, así como marcas de latidos desconocidos y ruido en la señal, respectivamente. El resultado fue un conjunto de datos de anomalías cardíacas, con 105,056 ejemplos de latidos distintos, divididos en 14 clases, donde cada clase indica una cardiopatía *e.g.* exceptuando la clase 1, que representa los latidos normales [26, 52]. La representación gráfica de los latidos se observa en la Figura 3.2, acompañada con la clase y el tipo de anomalía capturado en el conjunto de datos. El primer ejemplo indica la representación gráfica de un latido cardíaco normal, el indicado por la clase 1 (tipo *N*, el resto de las gráficas corresponden a ejemplos de distintas cardiopatías).

Extraer cada uno de los 105,056 latidos contenidos en las 46 derivaciones MLII, requiere de definir una subsecuencia a partir del índice en la anotación. La anotación indica el centro de la anomalía cardíaca, a partir de este punto podemos extraer una subsecuencia que abarque todo el latido. En la Figura 3.3, se indica como obtener la subsecuencia que engloba al latido, donde la señal de ECG presenta una anomalía detectada, que es encerrada a partir de dos marcadores cuyo centro es el índice de la

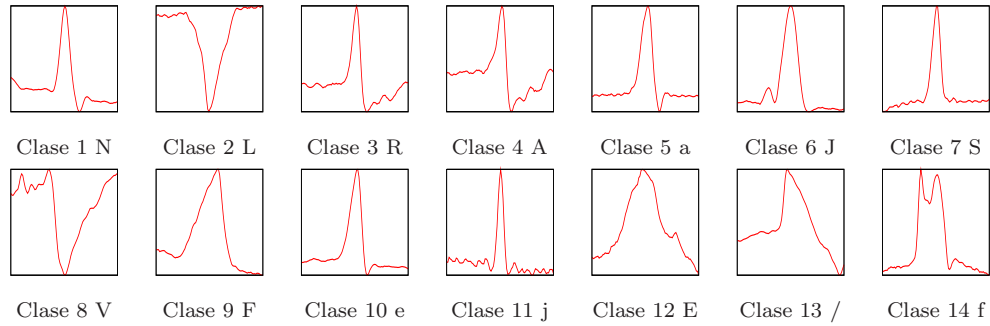


FIGURA 3.2. Representación Gráfica de Latidos Cardíacos.

anomalía.

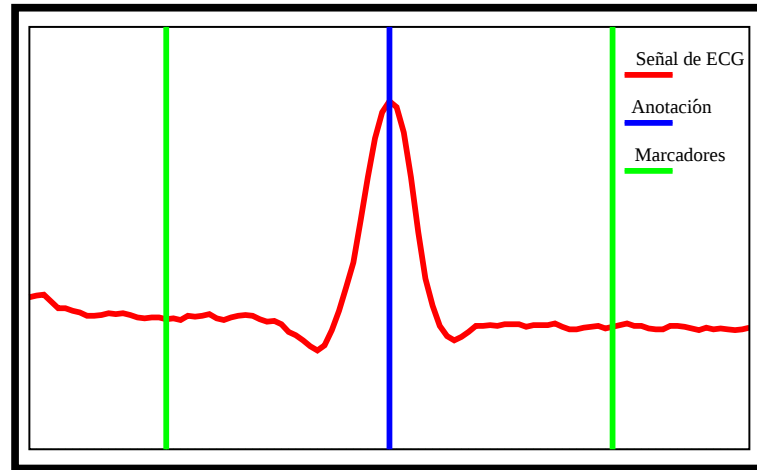


FIGURA 3.3. Extracción de una anotación del ECG.

La subsecuencia s_{kj} de un ECG S que comprende a un latido, se define a partir del índice i de la anotación y la longitud de la subsecuencia l , por medio de los cuales definimos los marcadores $k = i - \frac{l}{2}$ y $j = i + \frac{l}{2} - 1$, que comprenden l elementos contiguos donde k y j indican la posición de inicio y la posición final de la subsecuencia s_{kj} , respectivamente *i.e.* $s_{kj} = (S[k], \dots, S[j])$. La literatura recomienda que $\frac{1}{4}$ de segundo es tiempo suficiente para observar un latido cardíaco completo [57]. De esta manera, definimos $l = 90$, ya que cada ECGs en el conjunto de datos MIT-BIH Arritmia están almacenados utilizando 360 muestras por segundo.

3.2. Desarrollo de los Algoritmos de Clasificación

Del conjunto de datos (subsecuencias) construido como se indica en la sección tenemos que cada latido cuenta con una clasificación, representada por valor nominal o clase. La clasificación de latidos corresponde a la construcción de un modelo, que a partir de un conjunto de referencia $\mathbb{D}_r$. Donde el conjunto $\mathbb{D}$ está compuesto por ejemplos de latidos de ECGs correctamente identificados (*e.g.* latidos normales o anormales). La clase o identificación de los latidos se encuentra en el conjunto $\mathbb{Y}$. Con el objetivo de evaluar su eficacia, el clasificador es puesto a prueba al clasificar un conjunto de latidos $\mathbb{Q}$, *desconocidos* por el clasificador. Los clasificadores desarrollados son el algoritmo de vecinos más cercanos y redes neuronales artificiales. Además, se tomaron en cuenta características de los datos (*e.g.* rango de valores de las observaciones, la cantidad de clases de latidos y la frecuencia de los ECGs) facilitando la implementación de dichos algoritmos.

3.2.1. Clasificación por Vecino más Cercano

La clasificación de vecinos más cercanos, realiza una búsqueda por semejanza de un latido desconocido q en un conjunto de latidos clasificados $\mathbb{D}_r$. Esta búsqueda obtiene como resultado el elemento más semejante a q denotado por d_q y se conoce como vecino más cercano, debido a que el resultado es un único elemento (el de mayor semejanza). Para llevar a cabo los experimentos, implementamos dos funciones de semejanza; distancia Euclidiana e índice de Correlación.

El resultado de la función de distancia Euclidiana representa la distancia entre dos objetos, entendiendo que si la distancia entre estos es 0 los objetos son idénticos. Al efectuar la búsqueda de elemento más semejante con esta función, debemos encontrar el elemento d_q con menor distancia a q . Debido a que esta función de semejanza puede ser modificada sin alterar el significado del clasificador, realizamos tres implementaciones diferentes de esta función de semejanza: una implementación sin ninguna

modificación, una implementación que omite el cálculo de la raíz cuadrada y una última implementación que interrumpe el cálculo si el acumulado de la distancia supera a la mejor distancia global. Las implementaciones corresponden a optimizaciones en el cálculo de la distancia Euclidiana. Por otra parte, el índice de Correlación indica el nivel de relación entre dos objetos y está acotado en un rango de valores entre -1 y 1 . Si el índice de correlación entre dos objetos es 1 ambos objetos se relacionan de manera lineal ascendente. Realizar la búsqueda por semejanza con esta función consiste en encontrar el elemento d_q con mayor índice de correlación para q .

Posterior a identificar d_q , se determina si la clasificación fue correcta o no, al comparar las clases de ambos objetos. Se considera acierto cuando la clase del elemento más semejante encontrado $clase(d_q)$ es idéntica a la clase del elemento de búsqueda $clase(q)$, en caso contrario se considera como un error del clasificador. El Algoritmo 4, presenta al clasificador por Vecino más Cercano.

Algoritmo 3: Algoritmo de clasificación NN

Entrada: $\mathbb{D}$ Conjunto de datos,
 $\mathbb{Q}$ Conjunto de prueba,
 $sem()$ Función de semejanza

Hacer $mejor_semejanza = -\infty$
 Para cada elemento q_i de $\mathbb{Q}$, hacer:
 Iterar para cada elemento d_i en $\mathbb{D}$:
 Calcular $semejanza = sem(q_i, d_i)$.
 Si $es_mejor(semejanza, mejor_semejanza)$, entonces:
 Hacer $mejor_semejanza = semejanza$.
 Hacer $d_q = d_i$.
 Fin si
 Fin iteración

Si $clase(d_q)$ es idéntica que a $clase(q_i)$, entonces:
 Contabilizar como acierto.
 Si no: Contabilizar como error.

Fin para

La búsqueda por vecino más cercano, analiza cada elemento q del conjunto de

pruebas $\mathbb{Q}$, buscando el elemento $d_q \in \mathbb{D}_r$ más semejante a q . Nuestros experimentos buscan todos los elementos del conjunto Q en $\mathbb{D}_r$, donde $\mathbb{D}_r$ es cada uno de los conjuntos de datos conformados en la Sección ???. De igual manera, por cada conjunto $\mathbb{D}_r$ analizado se utilizan las cuatro funciones de semejanza mencionadas.

3.2.2. Red Neuronal Artificial

En el presente estudio se proponen tres arquitecturas *RNA*, Multi-Layer Feed-forward, cada una con distinta cantidad de nodos en la capa oculta. La fase de entrenamiento de estos clasificadores consistió en repasar los conjuntos de referencia 500 veces, donde cada iteración se conoce como *época*. En cada iteración, los pesos de las redes fueron modificados por el algoritmo backpropagation y un coeficiente de aprendizaje $\eta = 0.3$. Debido a las características de los datos utilizados (ver Sección 3.1.3), se define la capa de entrada con 90 nodos, uno para cada valor del patrón de entrada. Por otra parte, la cantidad de nodos en la capa oculta varía para cada una de las arquitecturas, dichas variaciones están basadas en recomendaciones de la literatura científica [9, 43, 45]. A continuación se describen cada una de las arquitecturas:

- Arquitectura 1, posee tres capas, configuradas en: 90, 150, k , que corresponden a la capa de entrada, intermedia y salida respectivamente; k indica el número de clases a clasificar que es 2 o 14.
- Arquitectura 2, posee tres capas, configuradas en: 90, 180, k , que corresponden a la capa de entrada, intermedia y salida respectivamente; k indica el número de clases a clasificar que es 2 o 14.
- Arquitectura 3, posee tres capas, configuradas en: 90, 271, k , que corresponden a la capa de entrada, intermedia y salida respectivamente; k indica el número de clases a clasificar que es 2 o 14.

Finalmente, la capa de salida de las *RNA*'s está definida en dos alternativas: dos y 14 nodos. La primer alternativa permite diferenciar los latidos sanos de los latidos anormales. Así mismo, la alternativa de 14 clases distingue entre 13 diferentes cardiopatías y latidos sanos. Se crea una *RNA* para cada conjunto de datos definido en la Sección 2.4, donde cada conjunto servirá como conjunto de entrenamiento. La finalidad de este experimento es observar la variación en la eficacia de las distintas *RNA*'s desarrolladas. La eficacia de la red fue evaluada considerando su salida. Para el caso de las redes que clasificaban dos clases, la capa de salida se compone de dos nodos. Calcular la salida del j -ésimo nodo s_j en la capa de salida se realiza aplicando un umbral de 0.5 a la salida del nodo, como se indica en la Ecuación (3.1).

$$salida_j = \begin{cases} s_j > 0.5 & 1 \\ s_j \leq 0.5 & 0 \end{cases} \quad (3.1)$$

Una vez normalizadas las salidas de los nodos en la capa de salida, se considera acierto si solamente existe una neurona con salida 1 y el índice de dicha neurona corresponde con la clase del patrón. De encontrarse más de un nodo con salida 1, se considera un error. Así mismo, si solo un nodo fue activado, pero el índice de este no corresponde con la clase, es considerado error.

Algoritmo 4: Algoritmo de clasificación de Red Neuronal Artificial

Entrada: $\mathbb{D}$ Conjunto de datos,
 $\mathbb{Q}$ Conjunto de prueba,
 r Red Neuronal Artificial entrenada,
 k Número de clases

$patron \leftarrow [p_1, \dots, p_k]$

Para cada elemento q_i de $\mathbb{Q}$, hacer:

 Hacer cada p_i en $patron$ 0

$vector_{clase} \leftarrow clasifica(red, q_i)$

$patron[clase(q_i)] \leftarrow 1$

 Si $vector_{clase} = patron$ entonces, contabilizar aciertos

 Si no: Contabilizar como error.

Fin para

3.2.3. Evaluación del Clasificadores

Para determinar el desempeño de los clasificadores, se considera el tiempo de ejecución, la cantidad de memoria utilizada, y la cantidad de aciertos obtenidos por el clasificador. Estas variables permiten identificar la eficiencia del clasificador en tiempo de ejecución y consumo de memoria, con respecto a su eficacia.

El tiempo de ejecución del clasificador, consiste en la cantidad de tiempo que toma el clasificador para clasificar todo el conjunto de datos de prueba $\mathbb{Q}$. Por otra parte, la cantidad de memoria utilizada indica la mayor cantidad de bytes cargada por el clasificador para realizar su cómputo. En el caso de la búsqueda del vecino más cercano, el indicador de memoria indica el tamaño en bytes del conjunto de datos $\mathbb{D}$ utilizado. Sin embargo, esta cantidad en bytes no es posible determinarla para la clasificación con *RNA's*, para este clasificador se indica el tamaño en bytes que posee la red. Esto se refiere a la cantidad de bytes necesarios para almacenar la estructura de la red neuronal como un archivo en la unidad de almacenamiento. Debido a que leer este archivo consiste en la mayor carga que hace el clasificador. Finalmente, para verificar la eficacia del clasificador, se contabilizan los aciertos que este haya obtenido durante la clasificación del conjunto $\mathbb{Q}$.

3.3. Descripción de las Plataformas Computacionales

El presente trabajo requirió una serie de experimentos en distintas plataformas computacionales, con el objetivo de observar la capacidad de una plataforma computacional para alternar la eficacia y eficiencia de los clasificadores. Las plataformas computacionales obedecen dos esquemas de trabajo, el esquema centralizado y el esquema distribuido.

3.3.1. Esquema Centralizado

En este marco de trabajo, los datos utilizados por el clasificador se encuentran integrados en un solo conjunto de datos. Tal alternativa indica que, tanto el conjunto de prueba como el conjunto de datos y el algoritmo de clasificación se encuentran físicamente en la misma unidad de proceso. Con esto, se evitan retrasos en la distribución de trabajo y en la recopilación de los resultados pero requiere de mayores capacidades del equipo de cómputo.

Centralizar los datos fue la alternativa implementada en las plataformas de cómputo de escritorio y cómputo móvil. En el primer caso, todos los diferentes conjuntos $\mathbb{D}$ fueron almacenados en un directorio en el sistema de archivos local del sistema operativo, de igual manera para todas las *RNA's* creadas y entrenadas. Al realizar la clasificación del conjunto $\mathbb{Q}$ por parte del algoritmo, este cargaba los datos del directorio y realizaba la clasificación. Por otra parte, la plataforma de cómputo móvil requirió de una memoria externa para almacenar datos de la experimentación, compuestos por las redes entrenadas y los diferentes conjuntos de datos $\mathbb{D}$ para evaluar la clasificación por vecinos más cercanos. Así mismo, se almacenó el conjunto de patrones a evaluar $\mathbb{Q}$. Fue necesario utilizar una memoria externa (4GB) ya que la memoria interna del dispositivo (2GB) era insuficiente [28].

3.3.2. Esquema Distribuido

El esquema distribuido, a diferencia del esquema centralizado, consiste en repartir el conjunto de datos en distintas unidades de procesamiento, también conocidos como *nodos de trabajo*. Al distribuir la información entre los diferentes nodos, la cantidad de datos que cada uno de estas unidades procesa, es considerablemente menor que el conjunto completo y por lo tanto hay menos requerimientos de capacidad de almacenamiento y procesamiento.

No obstante, involucrar a más de una unidad de procesamiento presenta situacio-

nes que deben ser interpretadas de manera correcta. El procesamiento de los datos es llevado a cabo por más de un nodo, por lo que se presenta una fase de comunicación de datos, así como una fase de integración de resultados. Integrar los resultados de diferentes nodos de procesamiento, puede involucrar discrepancias en los resultados. Por ejemplo, en el caso que los nodos clasifiquen el objeto de manera distinta. Estas diferencias son resultado de la distribución en los datos. Debido a que cada nodo posee una parte diferente del conjunto de datos y a la distribución de las cardiopatías en los conjuntos de datos, un tipo de cardiopatía pudiera no estar representada en el conjunto de datos de algún nodo. Esto provoca que diferentes nodos obtengan diferentes resultados.

De ocurrir discrepancias en los resultados, se elige por dos alternativas: tomar la clasificación con mejor valor de semejanza y realizar una votación entre el resultado de los nodos de procesamiento. La clasificación con mejor valor de semejanza, consiste en ubicar al latido desconocido con la clase del mejor candidato entre todos los nodos. En el caso de 1-NN, se considera como mejor candidato la semejanza y se clasifica al latido desconocido como aquél más semejante en los resultados de los nodos. En el caso de la red neuronal, el mejor candidato consiste en evitar utilizar el umbral para normalizar los resultados y obtener los valores de las neuronas en la capa de salida. La neurona con mayor valor de salida corresponde al mejor candidato, esto es comparado entre los resultados de todos los nodos y se clasifica con la neurona cuyo valor presente menor error. Por otra parte, el proceso de votación en ambos clasificadores es directo, las votaciones entre nodos, corresponden a la ocurrencia de clases. En caso de empate, se suman los resultados de los nodos, donde el resultado con mayor valor acumulado indica la clase del latido desconocido.

Clasificar en base al nodo con mejor resultado permite comparar los resultados con los del enfoque centralizado. El verdadero aporte de estos experimentos es el tiempo de ejecución del clasificador, que se ve beneficiado de distribuir los datos entre los nodos. Por el contrario, votar a través del resultado de los nodos para obtener la clase

resultante sugiere que un conjunto de pequeños clasificadores es mejor que un gran clasificador.

CAPÍTULO 4

EVALUACIÓN EXPERIMENTAL Y RESULTADOS

La experimentación llevada a cabo en el presente estudio, evalúa la eficacia y eficiencia de la clasificación por vecinos más cercanos y por redes neuronales artificiales. Conscientes de que la eficacia de un algoritmo de clasificación depende de los datos de entrenamiento y es invariable de la plataforma computacional, generamos diversos conjuntos de entrenamiento. Los conjuntos de entrenamiento son generados a partir de los datos del conjunto MIT-BIH Arritmia, integrado por ECGs reales [48]. Basados en los datos de MIT-BIH Arritmia, utilizamos dos estrategias de selección de elementos para conformar nuestros conjuntos de entrenamiento: Selección Aleatoria Uniforme y Selección Equi-proporcionada, obteniendo conjuntos donde la cantidad y distribución de datos son distintas. La eficacia de los algoritmos es determinada en función del porcentaje de aciertos obtenidos por estos. En particular se registra la eficacia general del clasificador, así como la eficacia por clase del mismo.

Por otra parte, la evaluación de eficiencia de un clasificador de latidos analiza las diferencias en tiempo de ejecución y cantidad de memoria consumidos por el algoritmo, ya que ejecutar el algoritmo en distintas plataforma computacional pudiese indicar diferentes niveles de eficiencia. Consideramos como tiempo de ejecución, el intervalo de tiempo que requiere un algoritmo para determinar la clasificación de un latido desconocido. Las diferencias observadas en los tiempos de ejecución al cambiar de plataforma computacional, esquema de trabajo y de conjunto de entrenamiento, son evaluadas para cada uno de los clasificadores.

4.1. Conformación de Conjuntos de Entrenamiento

En este estudio se utilizan dos estrategias para la conformación de los conjuntos de entrenamiento. La conformación por *Selección Aleatoria Uniforme* y la conformación por *Selección Equi-proporcional*. El objetivo es formar un nuevo conjunto de entrenamiento $\mathbb{D}_r$, donde $\mathbb{D}_r \subset \mathbb{D}$. Las estrategias utilizadas se describen a continuación:

1. Selección Aleatoria Uniforme, consiste en seleccionar de manera aleatoria *uniforme* n' elementos del conjunto $\mathbb{D}$ (*i.e.* cada elemento de $\mathbb{D}$ tiene la misma probabilidad de ser seleccionado). El resultado es un nuevo conjunto de entrenamiento $\mathbb{D}_r$.
2. Selección Equi-proporcional, el objetivo es producir un conjunto de entrenamiento equilibrado, al seleccionar el mismo número de registros por clase r_C , en la medida que esto sea posible. La selección se realiza del conjunto $\mathbb{D}$ y toma como parámetros el mismo conjunto, el número de elementos a seleccionar n' y las clases de los elementos en $\mathbb{D}$, expresadas como una secuencia y referida como $\mathbb{Y}$.

En ambas estrategias de selección, se consideran los 97,388 elementos del conjunto de entrenamiento. A partir de $\mathbb{D}$, cada estrategia genera 19 conjuntos de entrenamiento. El primer conjunto se conforma por 5,000 elementos y para los siguientes 18 esta cantidad se incrementa en 5,000 por cada nuevo conjunto. El proceso de *Selección Aleatoria Uniforme* consiste en seleccionar de manera aleatoria y sin repetición n' elementos de $\mathbb{D}$ para formar el conjunto $\mathbb{D}_r$.

Por otra parte, la *Selección Equi-proporcional*, que se describe en el **Algoritmo 5**, tiene la finalidad de generar un conjunto de datos equilibrado, donde cada clase del conjunto $\mathbb{Y}$ esté representada. El algoritmo toma como entrada los conjuntos $\mathbb{D}$ y $\mathbb{Y}$, así como la cantidad de elementos a seleccionar n' . Mientras sea posible agregar elementos al conjunto (*i.e.* el tamaño de $\mathbb{D}_r$ es menor a n' y todavía existan clases

sin marcar) se realiza una iteración. En cada iteración se define la representación por clase r_{clase} , que consiste en la proporción de elementos a seleccionar por clase, este valor está dado por la cantidad de elementos a seleccionar $acum$ entre la cantidad de clases no marcadas n_{clases} ; en la primer iteración, los valores corresponden a n' y $|\mathbb{Y}|$, respectivamente. Se realiza una búsqueda por cada clase y_i que no esté marcada, seleccionando todos los elementos del conjunto $\mathbb{D}$ tal que, sean de la clase y_i y no se encuentren en el conjunto $\mathbb{D}_r$. Dichos elementos son definidos por el conjunto S . Después se evalúa si la cantidad de elementos en S es mayor o no a r_{clase} , de ser cierto, se realiza una selección aleatoria uniforme de un elemento x de S , tal que x no se encuentre en $\mathbb{D}_r$. Finalmente x se agrega a $\mathbb{D}_r$ y el proceso se repite r_{clase} veces. En caso de que el resultado de la comparación $|S| > r_{clase}$ sea falso, todos los elementos de S se agregan a $\mathbb{D}_r$, además, se marca la clase y_i para no seleccionarla nuevamente y se reduce en uno la cantidad de clases n_{clases} . Así mismo, se acumula la diferencia entre los elementos encontrados y r_{clase} , esto permite generar una nueva representación a partir de los elementos restantes, que serán tomados de las clases no marcadas. Al cumplir con la cantidad de elementos indicados n' o al tener más clases disponibles, el ciclo termina.

A través de la selección equi-proporcional, se asegura la representación de todas las clases en el conjunto resultante $\mathbb{D}_r$ con al menos un elemento. Esto difiere de los conjuntos de entrenamiento generados por la selección aleatoria uniforme, debido a que todos los elementos poseen la misma probabilidad de ser seleccionados. Las clases compuestas por pocos elementos tienen menor probabilidad de estar representadas en el conjunto $\mathbb{D}_r$, utilizando la selección aleatoria. Aunado a las estrategias para conformar los conjuntos de entrenamiento, consideramos generar conjuntos para clasificar dos y 14 clases. La clasificación de 14 clases corresponde a la correcta identificación de las anomalías indicadas en la Figura 3.2. No obstante, decidimos considerar un problema de clasificación entre dos ejemplos de latidos *Normales* y *Anormales*, el cual será referido como *clasificación dicotómica*. con el objetivo de evaluar las capa-

Algoritmo 5: Selección Equi-Proporcional de Elementos.

Entrada: $\mathbb{D}$, Conjunto de patrones,
 $\mathbb{Y}$, Secuencia de clases,
 n' , Cantidad de registros a seleccionar
Salida: $\mathbb{D}_r$ Nuevo conjunto de patrones
 Crear $\mathbb{D}_r \leftarrow \emptyset$
 Definir $acum \leftarrow n'$, $n_{clases} \leftarrow |\mathbb{Y}|$
 Hasta que $|\mathbb{D}_r| = n'$ y $n_{clases} < 0$
 Definir $acum \leftarrow 0$, $r_{clase} \leftarrow \frac{acum}{n_{clases}}$
 Para cada clase no marcada y_i del conjunto de clases $\mathbb{Y}$
 Crear el conjunto $S = \{x | x \in \mathbb{D}, clase(x) = y_i, x \notin \mathbb{D}_r\}$
 Si $|S| > r_{clase}$ entonces, seleccionar al azar r_{clase} elementos de S
 y colocarlos en $\mathbb{D}_r$
 Si no, $\mathbb{D}_r = \mathbb{D}_r \cup S$, marcar(y_i), $n_{clases} \leftarrow n_{clases} - 1$,
 $acum \leftarrow acum + (r_{clase} - |S|)$
 Fin Para
 Repetir
 Regresar $\mathbb{D}_r$

ciudades de clasificación de los algoritmos. Los latidos normales están representados por la clase 1, por otra parte los latidos anormales están englobados en las 13 clases restantes. Por lo cual, a partir del conjunto $\mathbb{D}$ y $\mathbb{Y}$, así como de las estrategias de selección indicadas anteriormente, generamos conjuntos $\mathbb{D}_r$ donde todas las clases diferentes de la clase 1 (Normal), son homologadas como clase 2, interpretando esta clase como Anormal.

Los conjuntos de entrenamiento conformados poseen diferentes distribuciones de datos, de acuerdo con el tipo de selección de elementos implementada. En las Tablas 4.1, 4.2 y 4.3 observamos la distribución de la Selección Aleatoria Uniforme de datos. Donde la clase 1, que representa el 80 % de los latidos en el conjunto de datos $\mathbb{D}$, cuenta con la mayor cantidad de elementos en cada uno de los conjunto de entrenamiento, a diferencia de la clase 7 que no está representada en los conjuntos con 5, 15, 20, 25, 30, 35, 45 y 55 mil elementos. No obstante, la distribución de datos

es distinta para la Selección Equi-proporcional, donde se presentan conjuntos de entrenamiento equilibrados. En las Tablas 4.4, 4.5 y 4.6 se indica la distribución de datos para cada clase, observando que la equi-proporcionalidad de las clase se cumple hasta donde la cantidad de elementos por clase lo permite. La Tabla 4.4 indica como la distribución equi-proporcional del conjunto con 5,000 latidos solo es obtenida en 7 de las 14 clases. Además dicha proporción de clases va disminuyendo conforme se incrementa la cantidad de elementos en el conjunto, de manera que en el conjunto con 40,000 latidos la clase 1 posee la mayor cantidad de elementos.

TABLA 4.1. Distribución aleatoria uniforme de datos con 14 clases conjuntos de 5K hasta 40K.

Clase	Total de Registros							
	5K	10K	15K	20K	25K	30K	35K	40K
1	3,516	7,097	10,738	14,106	17,819	21,218	24,852	28,363
2	411	773	1,164	1,504	1,918	2,322	2,677	3,050
3	348	693	963	1,439	1,726	2,106	2,462	2,787
4	134	256	378	501	597	720	822	986
5	7	14	24	29	42	38	57	60
6	3	7	19	11	26	28	23	34
7	0	1	0	0	0	0	0	1
8	344	679	993	1,403	1,644	2,070	2,374	2,706
9	32	70	102	145	204	218	247	303
10	1	1	1	2	2	6	4	8
11	8	18	34	44	57	69	81	87
12	5	8	16	27	29	24	41	36
13	178	367	526	735	889	1,103	1,274	1,474
14	13	16	42	54	47	78	86	105

TABLA 4.2. Distribución aleatoria uniforme de datos con 14 clases conjuntos de 45K hasta 80K.

Clase	Total de Registros							
	45K	50K	55K	60K	65K	70K	75K	80K
1	31,938	35,568	39,008	42,514	46,036	49,617	53,268	56,909
2	3,475	3,806	4,156	4,591	5,028	5,374	5,751	6,097
3	3,139	3,436	3,842	4,150	4,549	4,891	5,179	5,494
4	1,088	1,205	1,334	1,478	1,583	1,652	1,827	1,922
5	60	78	75	100	87	100	112	108
6	39	28	44	40	55	52	61	65
7	0	0	1	1	1	1	1	1
8	3,089	3,330	3,812	4,098	4,405	4,792	5,094	5,443
9	366	384	409	468	486	544	570	612
10	4	10	5	11	12	12	11	13
11	96	109	123	132	152	155	164	171
12	25	45	53	61	72	71	77	82
13	1,584	1,868	1,996	2,209	2,375	2,559	2,702	2,890
14	97	133	142	147	159	180	183	193

TABLA 4.3. Distribución aleatoria uniforme de datos con 14 clases conjuntos de 85K hasta el conjunto completo.

Clase	Total de Registros			
	85K	90K	95K	97K
1	60,427	63,857	67,539	69,180
2	6,506	6,898	7,262	7,470
3	5,876	6,219	6,550	6,715
4	2,044	2,173	2,290	2,355
5	121	128	135	140
6	70	73	76	77
7	0	1	1	1
8	5,757	6,120	6,410	6,590
9	633	685	718	743
10	13	14	15	15
11	189	195	209	212
12	85	95	98	100
13	3,075	3,311	3,463	3,550
14	204	231	234	240

TABLA 4.4. Distribución equi-proporcional de datos con 14 clases conjuntos de 5K hasta 40K.

[illegible]

TABLA 4.5. Distribución equi-proporcional de datos con 14 clases conjuntos de 45K hasta 80K.

[illegible]

TABLA 4.6. Distribución equi-proporcional de datos con 14 clases conjuntos de 85K hasta Completa.

Clase	Total de Registros			
	85K	90K	95K	97K
1	56,792	61,792	66,792	69,180
2	7,470	7,470	7,470	7,470
3	6,715	6,715	6,715	6,715
4	2,355	2,355	2,355	2,355
5	140	140	140	140
6	77	77	77	77
7	1	1	1	1
8	6,590	6,590	6,590	6,590
9	743	743	743	743
10	15	15	15	15
11	212	212	212	212
12	100	100	100	100
13	3,550	3,550	3,550	3,550
14	240	240	240	240

En el caso de la clasificación dicotómica, que comprende latidos sanos y no sanos, las distribuciones de datos en los conjuntos de entrenamiento obedecen el mismo comportamiento de los conjuntos con 14 clases. Las Tablas 4.7, 4.8 y 4.9 indican que la Selección Aleatoria Uniforme favorece en cada conjunto a la clase de latidos sanos, debido a que posee mayor cantidad de ejemplos que la clase de latidos no sanos. Así mismo, las Tablas 4.10, 4.11 y 4.12, ilustran que la Selección Equi-proporcional obedece una distribución de datos equilibrada hasta el conjunto con 55,000 latidos. Los conjuntos de entrenamiento con más de 55,000 elementos superan la cantidad de latidos no sanos, por lo que la clase 1 completa el resto de los elementos necesarios.

TABLA 4.7. Distribución aleatoria uniforme de datos con dos clases conjuntos de 5K hasta 40K.

Clase	Total de Registros							
	5K	10K	15K	20K	25K	30K	35K	40K
1	3,585	7,086	10,688	14,189	17,695	21,281	24,808	28,322
2	1,415	2,914	4,312	5,811	7,305	8,719	10,192	11,678

TABLA 4.8. Distribución aleatoria uniforme de datos con dos clases conjuntos de 45K hasta 80K.

Clase	Total de Registros							
	45K	50K	55K	60K	65K	70K	75K	80K
1	31,886	35,569	39,025	42,624	46,226	49,705	53,155	56,833
2	13,114	14,431	15,975	17,376	18,774	20,295	21,845	23,167

TABLA 4.9. Distribución aleatoria uniforme de datos con dos clases conjuntos de 85K hasta completa.

Clase	Total de Registros			
	85K	90K	95K	97K
1	60,367	63,924	67,492	69,180
2	24,633	26,076	27,508	28,208

TABLA 4.10. Distribución equi-proporcional de datos con dos clases conjuntos de 5K hasta 40K.

Clase	Total de Registros							
	5K	10K	15K	20K	25K	30K	35K	40K
1	2,500	5,000	7,500	10,000	12,500	15,000	17,500	20,000
2	2,500	5,000	7,500	10,000	12,500	15,000	17,500	20,000

TABLA 4.11. Distribución equi-proporcional de datos con dos clases conjuntos de 45K hasta 80K.

Clase	Total de Registros							
	45K	50K	55K	60K	65K	70K	75K	80K
1	22,500	25,000	27,500	31,792	36,792	41,792	46,792	51,792
2	22,500	25,000	27,500	28,208	28,208	28,208	28,208	28,208

TABLA 4.12. Distribución equi-proporcional de datos con dos clases conjuntos de 85K hasta completa.

Clase	Total de Registros			
	85K	90K	95K	97K
1	56,792	61,792	66,792	69,180
2	28,208	28,208	28,208	28,208

4.2. Entrenamiento de los Clasificadores

Para llevar a cabo nuestros experimentos consideramos dos clasificadores de latidos distintos: Vecinos más cercanos y Redes neuronales. La clasificación de Vecinos más cercanos implementó el algoritmo 1-NN, además de dos funciones de semejanza distintas el Coeficiente de Correlación y la Función de Distancia Euclidiana. Este algoritmo de clasificación no requiere de una etapa de entrenamiento para ajustar parámetros, ya que el único parámetro a definir es la cantidad de vecinos, la cual se establece al desarrollar el algoritmo. Por otra parte, en la clasificación con Redes neuronales se implementó una arquitectura de tres capas, donde se desarrollaron tres variantes en la capa oculta del clasificador, como se indica en la **Sección 3.2.2**.

A diferencia del algoritmo 1-NN, la red neuronal lleva a cabo una fase de entrenamiento, que les permite actualizar los pesos entre los nodos y así reconocer los patrones. Las redes neuronales fueron entrenadas iterando 500 épocas con el algoritmo backpropagation, además se estableció el coeficiente de entrenamiento $\eta = 0.3$. Durante esta etapa se evalúa la eficacia de cada red para reconocer los patrones en el conjunto de entrenamiento durante cada época. Después de repasar el conjunto de entrenamiento y actualizar los pesos de la red, se calcula el error obtenido en la época. Este error corresponde a repasar nuevamente el conjunto de entrenamiento y contabilizar los aciertos obtenidos por el clasificador, esto nos permite observar el aprendizaje de la red neuronal.

En las **Figuras 4.1, 4.2 y 4.3** se aprecia la convergencia hacia un valor de error, durante el entrenamiento de las redes neuronales para las Arquitecturas 1, 2 y 3, respectivamente. Donde el error observado proviene del entrenamiento con los conjuntos de entrenamiento de 5,000, 50,000, 95,000 y 97,388 latidos; que corresponden a las etiquetas 5K, 50K, 95K y 97K. En las figuras se indican los términos: SAU y SEP, que se refieren al tipo de selección implementada para formar los conjuntos de referencia *i.e.* **Selección Aleatoria Uniforme SAU y Selección Equi-proporcional SEP**.

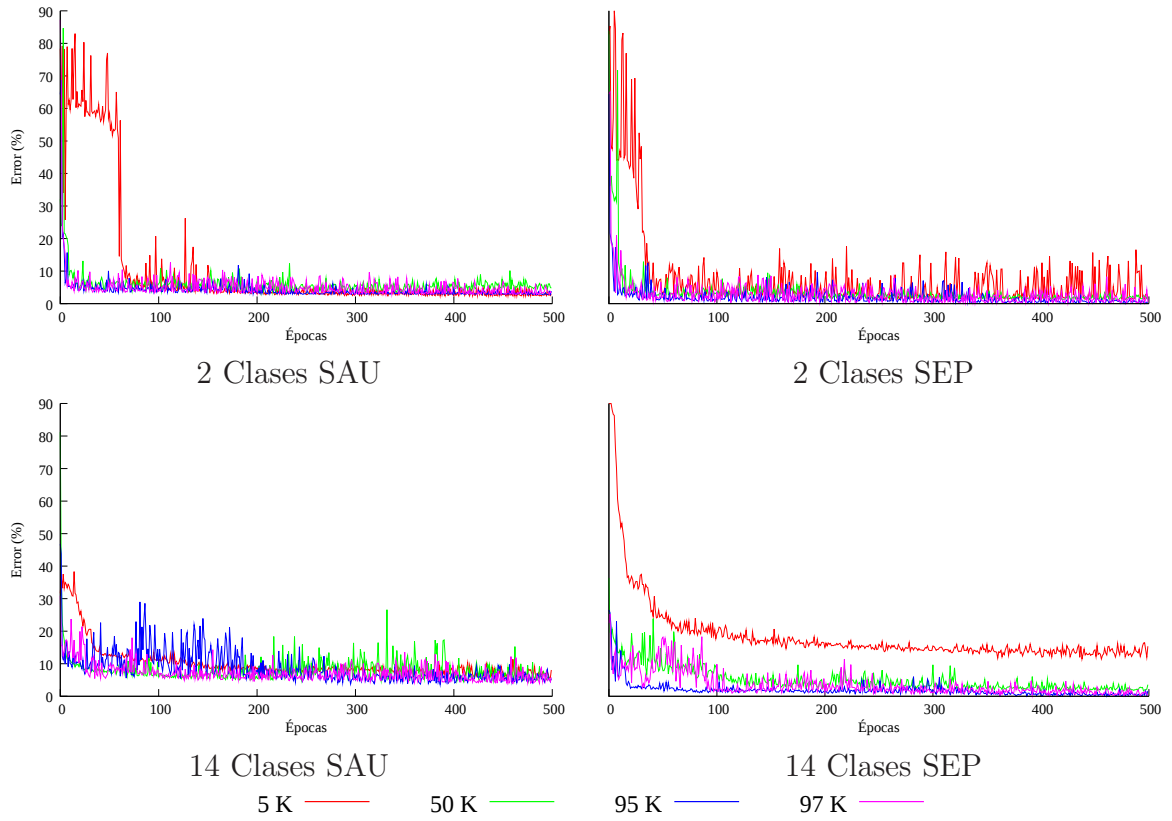


FIGURA 4.1. Desempeño de las RNAs con la Arquitectura 1.

La Figura 4.1, indica la evolución del porcentaje de error con la Arquitectura Uno, se indica como el conjunto de entrenamiento con 5,000 elementos presenta menor convergencia de error al implementar la selección equi-proporcional. Así mismo, en el caso de la clasificación de 14 cardiopatías e implementando una selección equi-proporcional, la red neuronal entrenada con un conjunto de 5,000 latidos estabiliza su error en un 20 %, el error más alto para todos los casos de la Arquitectura 1. Este caso se repite en la Arquitectura 2 y 3, como se indica en las Figuras 4.2 y 4.3, el error del clasificador al entrenar contemplando 14 clases distintas y el conjunto de 5,000 latidos seleccionados equi-proporcionalmente es el mayor. No obstante, a diferencia de la Arquitectura 1, el error de estos clasificadores desciende, lo que nos lleva a suponer que un mayor número de épocas ayudaría a estabilizar dicho error.

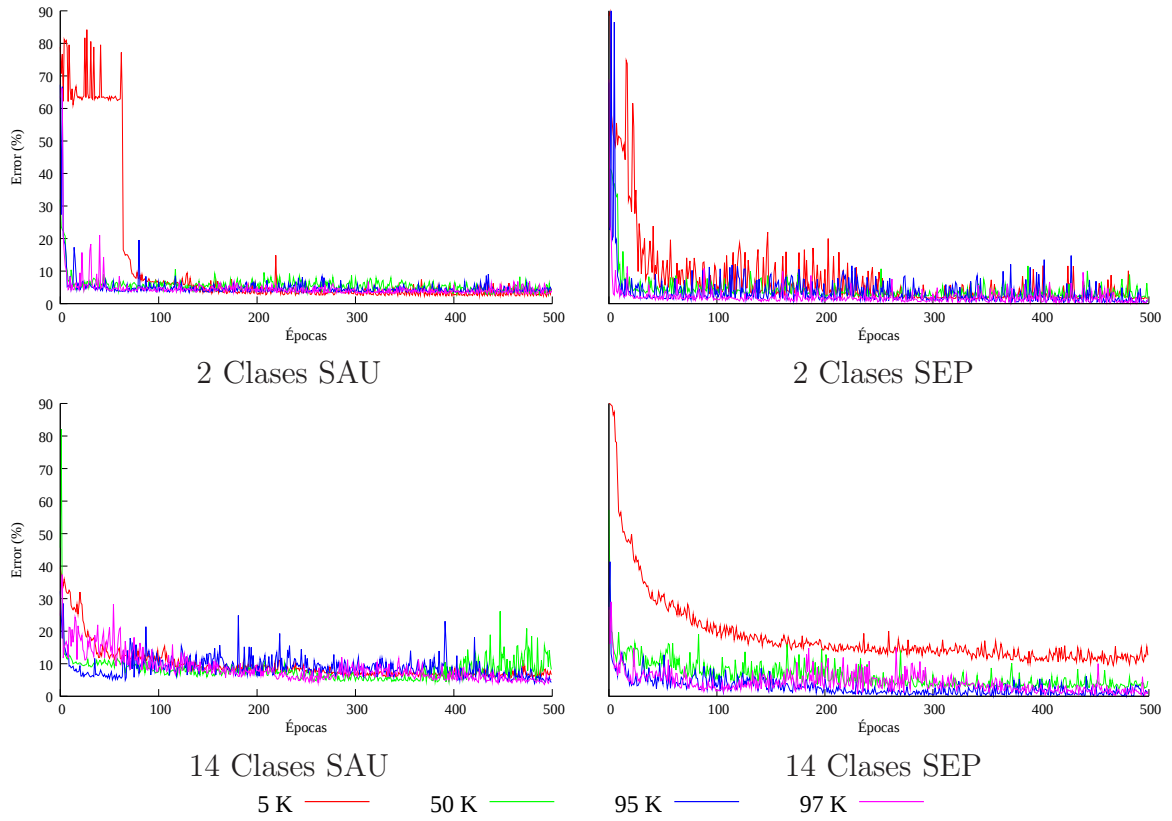


FIGURA 4.2. Desempeño de las RNAs con la Arquitectura 2.

Para entrenar las redes neuronales, generamos 78 conjuntos de entrenamiento distintos a partir del conjunto de datos $\mathbb{D}$, obteniendo 19 conjuntos para cada tipo de selección de elementos implementada e incluyendo el conjunto $\mathbb{D}$. Esto se llevó a cabo para cada tipo de clasificación a realizar *e.g.* se generaron 39 conjuntos de entrenamiento para realizar la clasificación dicotómica y otros 39 conjuntos de entrenamiento para realizar la clasificación de 14 cardiopatías. Por cada uno de los 78 conjuntos de entrenamiento se crearon tres redes neuronales, una para cada arquitectura, dando un total de 234 redes.

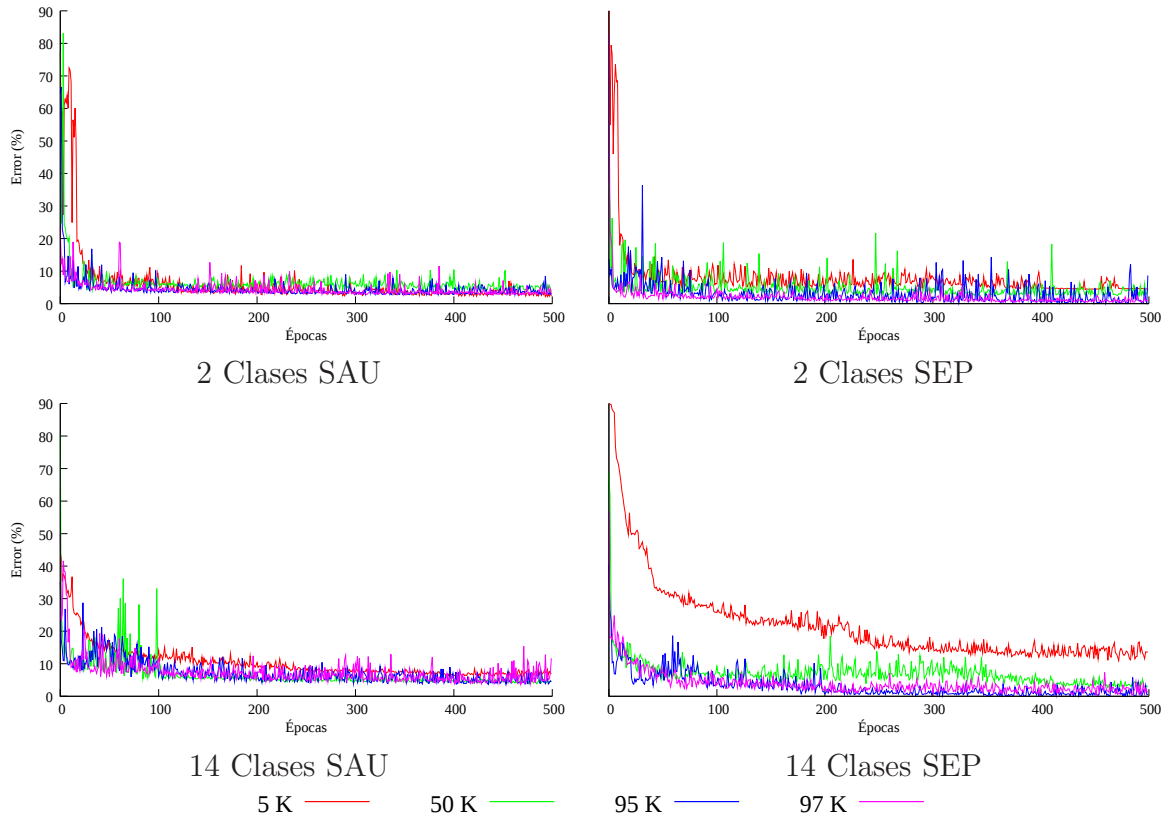


FIGURA 4.3. Desempeño de las RNAs con la Arquitectura 3.

4.3. Experimentos sobre una Plataforma Estacionaria

Una plataforma computacional consiste en una combinación de hardware y software que presenta un destino de ejecución de un programa. Este trabajo considera dos plataformas computacionales: plataforma de cómputo estacionaria y plataforma de cómputo móvil. De igual manera, se consideran dos esquemas de trabajo: el esquema centralizado y el esquema distribuido. Tradicionalmente, la literatura utiliza un esquema de trabajo centralizado para evaluar la eficacia de un algoritmo de clasificación de latidos, de igual manera las plataformas computacionales representan equipo de cómputo estacionario [6, 5, 31, 37]. Los datos utilizados (ambos conocimiento y datos de consulta) en este esquema de trabajo se encuentran almacenados en un solo lugar (*i.e.* se encuentran centralizados), evitando realizar comunicación por alguna red de

computadora para solicitar o integrar información. Para este tipo de plataformas, los detalles de uso de energía son triviales. Este estudio ubica al cómputo estacionario, con un esquema de trabajo centralizado como el *benchmark* de nuestra experimentación, cuya eficiencia servirá de base para suponer cambios significativos entre el resto de las plataformas computacionales.

Por otra parte, consideramos el esquema de trabajo distribuido, donde el conocimiento representado en el algoritmo de clasificación está distribuido entre un grupo de expertos [29]. La ventaja de este esquema de trabajo se encuentra cuando se cuenta con grandes volúmenes de datos, en los que realizar un análisis de datos de manera concentrada o en un solo proceso, implica mucha carga computacional. El análisis de conjuntos de datos masivos conocidos como *Big Data*, implementa una distribución de los datos para reducir la carga computacional, esto se traduce en tiempos de cómputo menores, ya que se resuelven pequeñas partes del problema de forma distribuida. La solución global del problema resulta de la integración de estas soluciones parciales [54].

4.4. Experimentos sobre una Plataforma Móvil

La plataforma de cómputo móvil, equivale a un cambio en el enfoque de clasificación, *clasificación en tiempo real* [27, 33]. Este esquema se ve representado por la plataforma computacional de dispositivos móviles o portables (*e.g.* celulares y tablets). Para implementar la clasificación en línea, incorporamos la información en el dispositivo, evitando la distribución de los datos. Esta plataforma computacional, que comparada con una plataforma tradicional posee menos recursos computacionales, utiliza de un suministro agotable de energía (*i.e.* pila). La ejecución de un algoritmo de cómputo en esta plataforma puede agotar la energía del dispositivo, donde el nivel de reducción de este recurso pudiera calificar al algoritmo como inviable para implementarse. Esta plataforma computacional ofrece una compensación entre sus

dimensiones físicas (que la vuelven portable) y poder de cómputo. Aprovechando la portabilidad de esta plataforma, se evalúa si es posible incorporar los algoritmos de clasificación sin perder eficacia, comparando los resultados con el *benchmark*. Implementar *clasificación en línea*, permite validar si la eficacia se mantiene, considerando que los datos de consulta son producidos en el momento.

4.5. Resultados de Clasificación

A continuación presentamos los resultados obtenidos por los clasificadores, agrupados por la eficacia en la detección de latidos: eficacia global (involucrando todas las clases), eficacia para la clasificación dicotómica (clasificación entre latidos normales y latidos anormales) y la eficacia para la clasificación de anomalías (la clasificación de las 14 clases de latidos en el conjunto de datos). Así mismo, en la eficacia en la clasificación dicotómica y de anomalías se presenta en los esquemas de trabajo centralizado y distribuido. El esquema centralizado fue implementado en las plataformas computacionales de escritorio y móvil con resultados equivalentes, por lo que la eficacia centralizada es la misma en ambas plataformas. Por otra parte, la eficacia distribuida fue implementada solamente en la plataforma computacional de escritorio.

Los resultados obtenidos la fase experimental son añadidos al final del documento a manera de anexos, donde el Anexo I corresponde a los resultados de la clasificación dicotómica en el esquema de trabajo centralizado. El Anexo II, indica los resultados obtenidos en la clasificación de anomalías, utilizando el esquema de trabajo centralizado y finalmente el Anexo III cubre los resultados del esquema de trabajo distribuido.

4.5.1. Clasificación de Latidos

Se reporta la eficacia global ($eficacia_g$) de los clasificadores, está corresponde al porcentaje de aciertos que el clasificador obtuvo. El porcentaje se obtiene de dividir el total de aciertos obtenidos por el clasificador al clasificar el conjunto de pruebas

(*Aciertos*) entre la cantidad de elementos en el conjunto de prueba (*Elementos*), multiplicando el resultado por 100.

$$eficacia_g = \frac{Aciertos}{Elementos} \times 100 \quad (4.1)$$

Donde podemos observar la eficacia de los algoritmos para clasificar dos y 14 clases. En la **Figura 4.4.I.A**), se observa la variación en la eficacia global de los algoritmos al clasificar entre latidos normales y anormales. El menor porcentaje de eficacia global obtenido proviene de la red neuronal con Arquitectura 1, cuyo porcentaje es de 87.8 %, entrenada con un conjunto de referencia de 30,000 latidos con selección aleatoria uniforme. Además, en la gráfica se observa que el clasificador con mayor eficacia en todos los casos es 1-NN-CCI (Coeficiente de Correlación), seguido por 1-NN-E (distancia Euclidiana). En el caso de la clasificación dicotómica por redes neuronales, la mejor eficacia entre las arquitecturas no es clara, esto se atribuye al poco entrenamiento de los clasificadores. Así mismo, en la **Figura 4.4.I.B**), que indica la variación en el porcentaje de eficacia global con conjuntos de referencia con selección equi-proporcional, se observa el mismo comportamiento de los clasificadores 1-NN-CCI y 1-NN-E. No obstante, la clasificación con redes neuronales indica mayor variación en la eficacia al utilizar conjuntos de referencia con datos seleccionados de manera equi-proporcional. El menor porcentaje de eficacia global obtenido, corresponde a la red neuronal con Arquitectura 3 con un valor de 31.95 %, entrenada con un conjunto de referencia de 55,000 latidos. Este bajo porcentaje de eficacia corresponde a un promedio entre una cantidad de aciertos baja, al clasificar latidos normales, y alta al clasificar de latidos anormales. La Tabla 4.11 indica la distribución de latidos de dicho conjunto de referencia, donde se deja de observar la equi-proporcionalidad de las clases en aquellos conjuntos con cardinalidades superiores a 55 mil latidos. Por lo tanto, asociamos este resultado al entrenamiento del clasificador, ya que solo en este conjunto se obtuvo un porcentaje bajo, a diferencia de conjuntos de mayor cardinalidad.

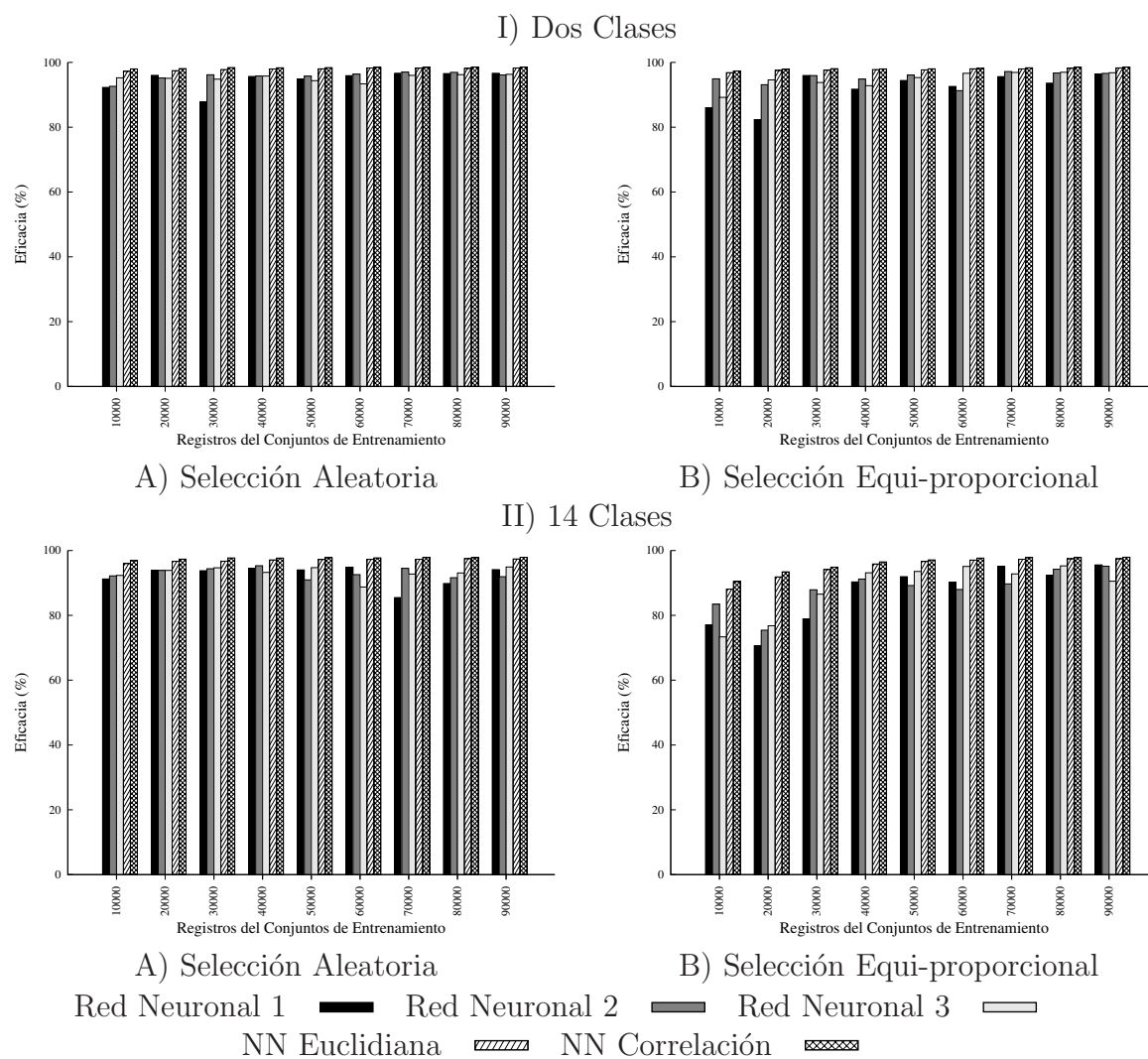


FIGURA 4.4. Eficacia global de clasificadores

De igual manera, en la **Figura 4.4.II**, se observa que la eficacia global de la clasificación de 14 clases, donde el mayor porcentaje de eficacia global proviene del clasificador 1-NN-CCI, seguido de 1-NN-E, al igual que en la clasificación dicotómica. No obstante, a diferencia de la clasificación dicotómica donde el porcentaje de eficacia global no se ve afectado al incrementar la cantidad de datos en los conjuntos de referencia, la clasificación de 14 clases presenta un comportamiento distinto. Donde la **Figura 4.4.II.B)** indica que el porcentaje de eficacia global de los clasificadores con conjuntos de referencia con selección equi-proporcional, incrementa a medida que la cardinalidad del conjunto de referencia aumenta. Esto se debe, a que en los datos en conjuntos de referencia aleatorios uniformes existen clases de latidos no representadas, a diferencia de la selección equi-proporcional, que asegura la representación de todas las clases en el conjunto.

4.5.2. Clasificación Dicotómica

La clasificación dicotómica considera dos clases solamente, identificando entre latidos sanos y latidos con cardiopatías. Para efectuar esta clasificación se dividieron los conjuntos de latidos entre latidos normales (clase 1 de las 14 clases del conjunto) y latidos anormales (el resto de clases en el conjunto de datos). El resultado indica que los clasificadores tienen un desempeño similar cuando los conjuntos de datos se encuentran bien establecidos. El porcentaje de eficacia de los clasificadores, es obtenido al dividir la cantidad de latidos correctamente identificados de una clase particular entre la latidos de dicha clase en el conjunto. En la Ecuación (4.2), $aciertos_c$ y $errores_c$ corresponden a la cantidad de latidos correctamente clasificados y la cantidad de latidos mal clasificados para la clase c , respectivamente. De manera que c puede ser la clase que represente a los latidos sanos o con cardiopatías, así como cualquiera de las 14 clases en la clasificación de anomalías.

El porcentaje eficacia de clasificación es calculado para cada clase involucrada. Es decir, se calcula el porcentaje de eficacia para la clase de latidos normales y de latidos anormales por separado.

$$eficacia_n = \frac{aciertos_c}{errores_c + aciertos_c} \times 100 \quad (4.2)$$

En la **Figura 4.5**, se observa el porcentaje de eficacia de los clasificadores en la clasificación de latidos sanos. Así mismo, la **Figura 4.5.A)** indica el porcentaje de eficacia obtenido de clasificadores con conjuntos de referencia con selección aleatoria uniforme. Observando que la clasificación con redes neuronales, supera a la clasificación con 1-NN, para los conjuntos de referencia con 10, 15, 25, 35, 40, 45, 50, 65, 70, 75, 80, 90, 95 y 97 mil latidos. Esto se debe a que la distribución de datos en dichos conjuntos favorece a la clase de latidos normales, como se indica en las **Ta-blas 4.7, 4.8 y 4.9**. De igual manera, la **Figura 4.5.B)** ilustra el porcentaje de eficacia, utilizando conjuntos de referencia con selección equi-proporcional.

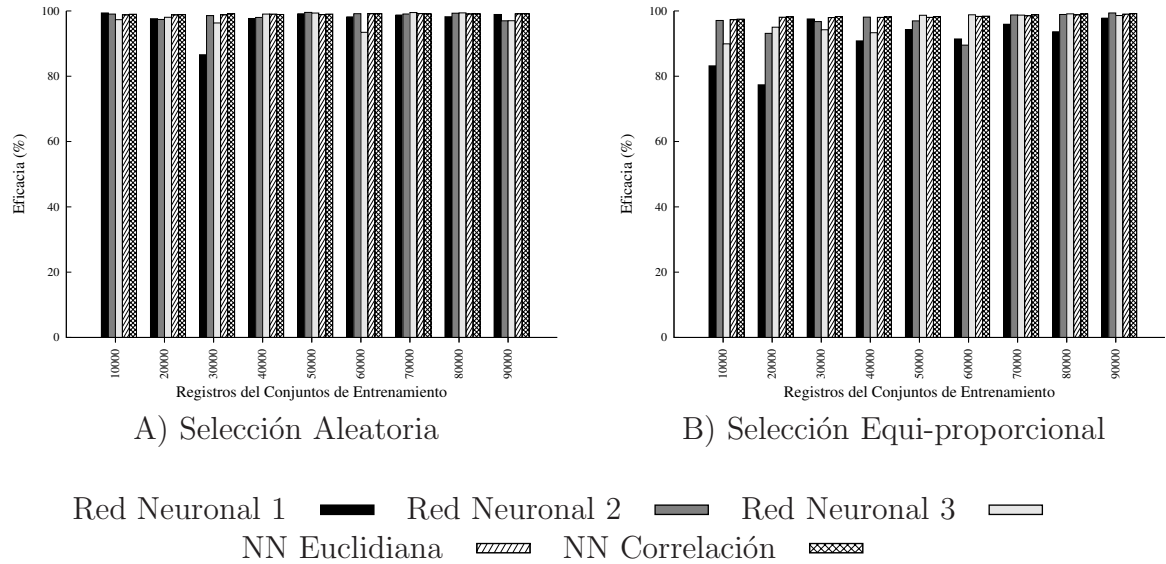


FIGURA 4.5. Eficacia centralizada de latidos Normales

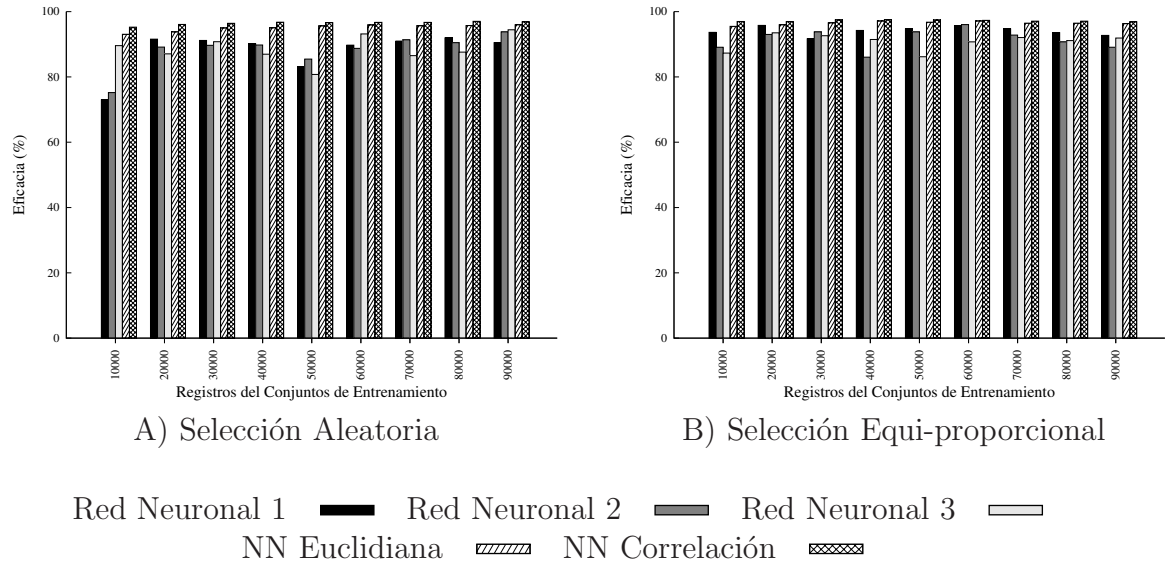


FIGURA 4.6. Eficacia centralizada de latidos Anormales

El porcentaje de eficacia para la clasificación de latidos anormales, se presenta en la **Figura 4.5**. El porcentaje de eficacia para la clasificación de latidos anormales, es menor que el porcentaje de eficacia para la clasificación de latidos normales, utilizando conjuntos de referencia con selección aleatoria uniforme es menor, como se muestra en la **Figura 4.5.A**). Este comportamiento se debe a que la clase de latidos anormales, tiene menor representación en la selección aleatoria uniforme. Al utilizar conjuntos de referencia donde la representación de las clases se encuentra equilibrada se presenta una mejor eficacia para la clasificación de latidos anormales. Lo anterior se observa en la **Figura 4.5.B**), donde el mayor porcentaje de eficacia corresponde con el conjunto de referencia con 55 mil latidos en el cual las clases de latidos están equi-proporcionalmente representadas en el conjunto de referencia de acuerdo con las **Tablas 4.7, 4.8 y 4.9**. Así mismo, observamos como conjuntos de referencia con cardinalidades mayores a 55 mil latidos sobre-alimentan al clasificador y no aportan a la eficacia de este.

En el esquema de trabajo distribuido, los conjuntos de referencia fueron repartidos entre distintas unidades de procesamiento (nodos), de acuerdo con la cardinalidad de

estos. Utilizamos los conjuntos con 5, 10, 20, 40 y 80 mil latidos, distribuidos en 1, 2, 4, 8 y 16 nodos, respectivamente. La clasificación dicotómica a través de un esquema de trabajo distribuido se llevó a cabo considerando dos alternativas de clasificación: Mejor candidato y Votación. La clasificación por mejor candidato, en el caso de 1-NN consiste en, buscar un latido desconocido en cada uno de los nodos y clasificar al latido con la mayor semejanza encontrada entre los nodos. En el caso de la clasificación por votación, se clasifica por la clase con mayor presencia entre los nodos. Donde de existir un empate se suma la semejanza obtenida por cada una de las partes y se asigna la clase con mayor semejanza. La clasificación distribuida por mejor candidato de 1-NN, es equivalente a la clasificación utilizando un esquema de trabajo centralizado con este mismo clasificador. No obstante, para el caso de la clasificación por redes neuronales, para cada partición del conjunto de referencia en un nodo se creó y entrenó una red neuronal. La clasificación por mejor candidato con redes neuronales está basada en la neurona de la capa de salida con valor mayor. Cada nodo indica la clase y el valor en la neurona de salida, donde la clasificación final es dada a través de el mayor valor. Por otra parte, la clasificación por votación con redes neuronales realiza considera a los nodos involucrados a través de un conteo entre las clases. La clasificación corresponde a la clase con más votos y en caso de empate se pondera la salida de las neuronas de salida de las clases.

En la **Figura 4.7.I)** se ilustra el comportamiento de la eficacia en los clasificadores distribuidos al clasificar correctamente los latidos normales, utilizando conjuntos de referencia con selección aleatoria. Se observa una alta eficacia (arriba del 90 %) para clasificar correctamente latidos normales, debido a que en estos conjuntos, los latidos normales tienen mayor representación que los latidos anormales. Así mismo, las redes neuronales utilizadas no son sobre-alimentadas como en el caso del esquema de datos centralizado, donde las redes fueron entrenadas con conjuntos de referencia de cardinalidades distintas. La clasificación distribuida, al incrementar la cardinalidad del conjunto de referencia, también realiza un incremento en las unidades de

procesamiento, de manera que cada unidad de procesamiento es entrenada con un subconjunto de 5 mil elementos. Como resultado de la división de datos y la clasificación por votación por parte del esquema distribuido, se obtuvo mayor porcentaje de eficacia con las redes neuronales que en el esquema centralizado. Además, la **Figura 4.7.I.B** indica que el porcentaje de eficacia de los clasificadores 1-NN-E mejora al utilizar la clasificación por votación entre los nodos, superando a 1-NN-C con 0.143, 0.107, 0.161, 0.268 y 0.428 %, en 1, 2, 4, 8 y 16 nodos respectivamente.

No obstante, en la **Figura 4.7.II**), se ilustra como las redes neuronales aprovechan conjuntos de referencia conformados por la selección equi-proporcional. En el caso de la clasificación por mejor candidato, las redes neuronales superaron la eficacia de los clasificadores 1-NN, al igual que la clasificación de latidos normales por votación utilizando dos nodos, como se indica en las **Figuras 4.7.II.A** y 4.7.II.B. En el caso de 4, 8 y 16 nodos, el mayor porcentaje de eficacia lo obtuvo el algoritmo 1-NN-E.

En el caso de la clasificación de latidos anormales, observamos en la **Figura 4.8** que el clasificador 1-NN-C presenta el mayor porcentaje de eficacia seguido por 1-NN-E, para las selecciones aleatoria y equi-proporcional así como la clasificación por mejor candidato y votación. La clasificación por redes neuronales que mejor resultados obtuvo en la selección aleatoria fue la Arquitectura 2, como se indica en la **Figura 4.8.I.A**. Esta misma red neuronal superó al resto de las redes en la clasificación de latidos anormales con mejor candidato selección equi-proporcional, esto es visto en la **Figura 4.8.II.A**. Sin embargo la clasificación por votación no presenta una red neuronal que supere al resto, donde solo en ciertos casos una red supera al resto como se indica en la **Figura 4.8.II.B**.

4.5.3. Clasificación de Anomalías

Esta clasificación representa todas las anomalías, es decir, detectar las 14 clases distintas de anomalías. Este problema requiere necesariamente de conjuntos donde

se encuentren bien representadas cada una de las clases a clasificar. Se observa que el porcentaje de eficacia de los clasificadores es mayor al utilizar conjuntos de datos equilibrados (selección equi-proporcional), que al utilizar conjuntos de latidos conformados por selección aleatoria uniforme. Los resultados presentados a continuación omiten las clases 7 y 10, debido a que ninguno de los clasificadores logró reconocer estas clases, a pesar de utilizar el esquema de trabajo distribuido.

En la **Figura 4.9**, se observa el porcentaje de eficacia de los algoritmos, al clasificar la Clase 1, utilizando un esquema de trabajo centralizado. Así mismo, se distingue que el mayor porcentaje de eficacia al clasificar esta clase es obtenido por los algoritmos 1-NN, sin importar que tipo de conjunto de referencia se utilice, ya sean conjuntos de referencia con selección aleatoria uniforme o con selección equi-proporcional, en un esquema de trabajo centralizado. Así mismo, el porcentaje de eficacia obtenido con redes neuronales y conjuntos de referencia aleatorios, es comparable con los algoritmos 1-NN en la mayoría de los casos, como se indica en la **Figura 4.9.A**

No obstante, al utilizar conjuntos de referencia equi-proporcionales, *i.e.* con latidos seleccionados de manera equi-proporcional, la eficacia de los algoritmos va incrementando a medida que el conjunto de datos aumenta. Además, en la **Figura 4.9.B**, se indica como a partir de conjuntos de referencia equi-proporcionales con 65 mil latidos, la eficacia de las redes neuronales y la búsqueda 1-NN se asemejan. Al comparar los resultados de la clasificación utilizando conjuntos de referencia aleatorios (con un porcentaje de eficacia de entre 82.65 a 99.19%, referente a la red neuronal con Arquitectura 1 y a 1-NN-C), frente al uso de conjuntos de equi-proporcionados (con un porcentaje de eficacia de entre 53.17 a 99.19%, correspondientes a la red neuronal con Arquitectura 2 y 1-NN-C), se observa un rango de eficacia mayor en la clasificación con conjuntos de referencia aleatorios. Este resultado se debe a que la Clase 1 posee siempre la mayor proporción en dichos conjuntos de referencia, a diferencia de los conjuntos equi-proporcionados donde esto ocurre en conjuntos con una cardinalidad mayor a 40 mil latidos.

Por otra parte, al utilizar el esquema de trabajo distribuido, el porcentaje de eficacia de las redes neuronales, en la clasificación de la Clase 1, es mayor que el obtenido por los algoritmos 1-NN, como se observa en la **Figura 4.10**. Al llevar a cabo la clasificación utilizando conjuntos de referencia aleatorios, en donde la distribución de latidos de Clase 1 siempre es mayor que el resto de las clases, se indica un porcentaje de eficacia mayor al 90% en la clasificación por mejor candidato y por votación, representadas por las **Figura 4.10.I.A** y **Figura 4.10.I.B**, respectivamente. No obstante, en la **Figura 4.10.II** se indica como al utilizar conjuntos de referencia equi-proporcionados para clasificar la Clase 1 presenta diferentes resultados. Donde el mayor porcentaje de eficacia obtenido al clasificar la Clase 1 con conjuntos de referencia equi-proporcionales proviene del clasificador 1-NN-C, en la clasificación de mejor candidato, observado en la **Figura 4.10.II.A**. Sin embargo, como se ilustra en la **Figura 4.10.II**, al utilizar la clasificación por votación e incrementar la cantidad de unidades de procesamiento (y el tamaño de los conjuntos de referencia), la eficacia de las redes neuronales se incrementa.

El porcentaje de eficacia de los algoritmos al clasificar la Clase 2, utilizando el esquema de trabajo centralizado es observado en la **Figura 4.11**. Los algoritmos con mayor porcentaje de eficacia en la clasificación de la Clase 2, pertenecen a las búsquedas 1-NN-C y 1-NN-E, siendo superior la clasificación con 1-NN-C. En los resultados obtenidos, no se observa una variación entre el uso de conjuntos aleatorios (**Figura 4.11.A**) y conjuntos equi-proporcionados (**Figura 4.11.B**). A pesar de que la distribución de latidos de los conjuntos de referencia aleatorios, para la Clase 2, indica la representación más alta con 7,262 latidos en el conjunto de referencia con una cardinalidad de 95 mil. Por otra parte, los conjuntos de referencia equi-proporcionados mantienen 7,470 latidos a partir de cardinalidades iguales o mayores a 45 mil elementos.

Así mismo, al utilizar el esquema de trabajo distribuido observamos que el mayor porcentaje de eficacia de los algoritmos pertenece a los clasificadores 1-NN, rebasando

a las redes neuronales, como se indica en la **Figura 4.12**. No obstante, observamos una menor diferencia en el porcentaje de eficacia entre los algoritmos de clasificación utilizando conjuntos de referencia equi-proporcionados (como lo indica la **Figura 4.12.II**), que aquella resultado de utilizar conjuntos de referencia aleatorios (visto en la **Figura 4.12.I**).

Los resultados de la clasificación de la Clase 3, llevada a cabo bajo el esquema de trabajo centralizado se observan en la **Figura 4.13**. Los conjuntos de referencia aleatorios indican que la mayor representación de esta clase se presenta en el conjunto con 95 latidos, de los cuales 6,550 pertenecen a dicha clase. Por otra parte, los conjuntos de referencia equi-proporcionados indican que la mayor representación de la Clase 3 se obtiene en conjuntos de referencia con cardinalidades iguales o mayores a 35 mil latidos, donde 6,715 pertenecen a esta clase. Al igual que los resultados de la clase anterior, los mayores porcentajes de eficacia fueron obtenidos por los clasificadores 1-NN-C y 1-NN-E, a excepción de los resultados obtenidos con conjuntos de referencia aleatorios de 65 y 75 mil latidos, donde las redes neuronales con Arquitecturas 1 y 3, respectivamente, superaron a los algoritmos 1-NN, como se indica en la **Figura 4.13.A**. En el caso de conjuntos de referencia equi-proporcionados, las redes neuronales superaron a los algoritmos 1-NN, únicamente en el conjunto con una cardinalidad de 75 mil latidos. Esto se observa en la **Figura 4.13.B**.

En el caso de la clasificación utilizando el esquema de trabajo distribuido, se observa que el porcentaje de eficacia de los clasificadores aumenta al utilizar la clasificación por votación, ya sea con conjuntos de referencia aleatorios o equi-proporcionados, como es ilustrado en la **Figura 4.14**.

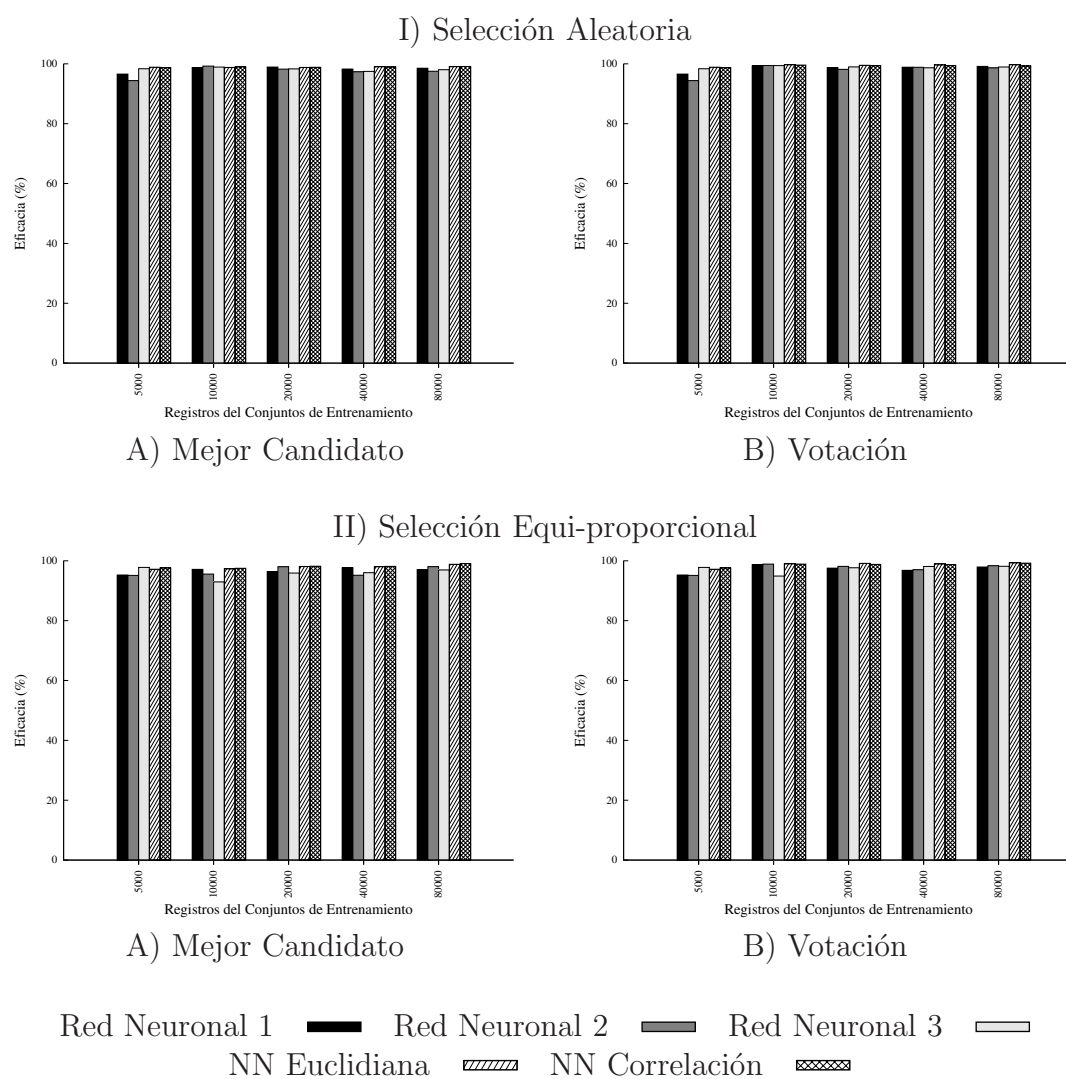


FIGURA 4.7. Eficacia distribuida de latidos Normales

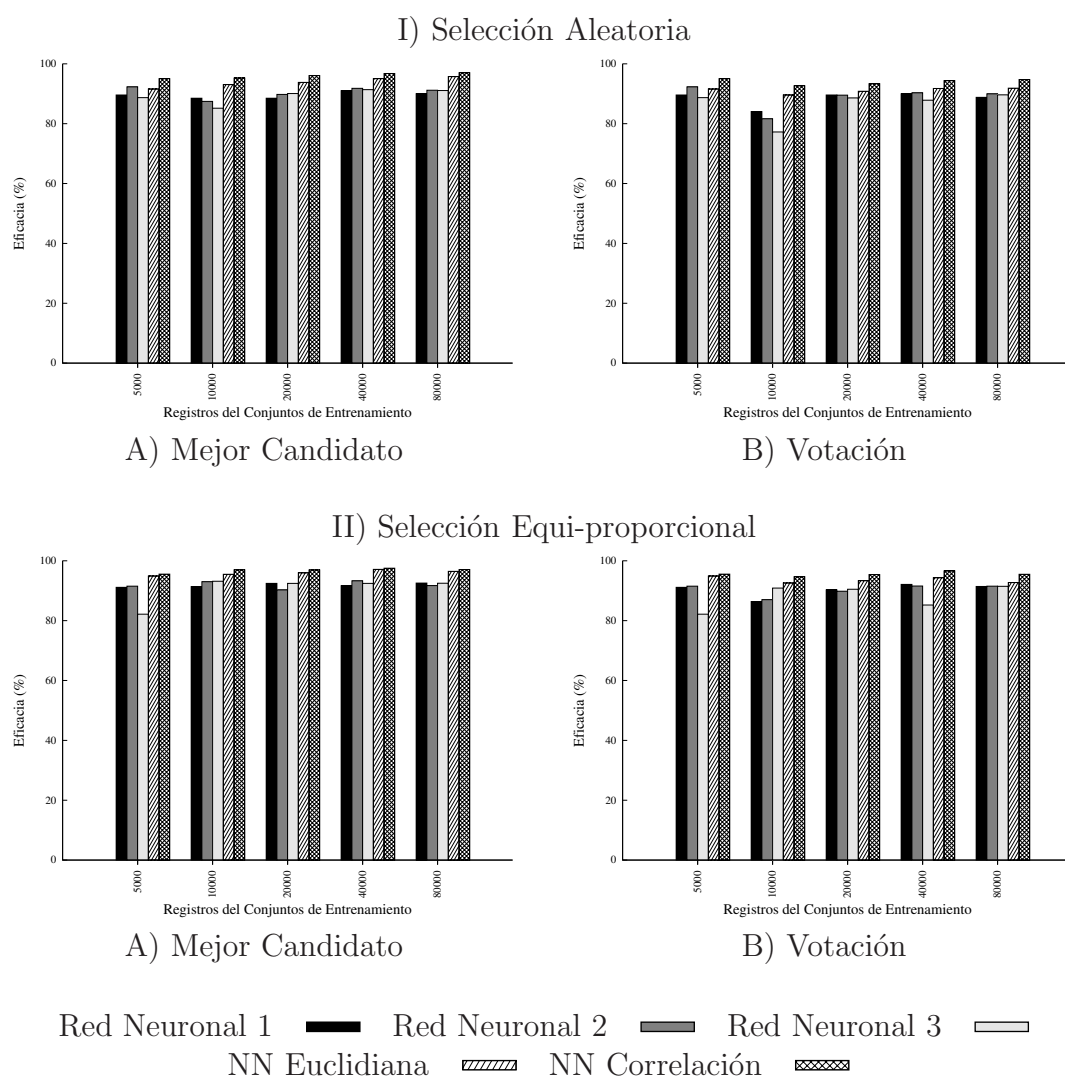


FIGURA 4.8. Eficacia distribuida de latidos Anormales

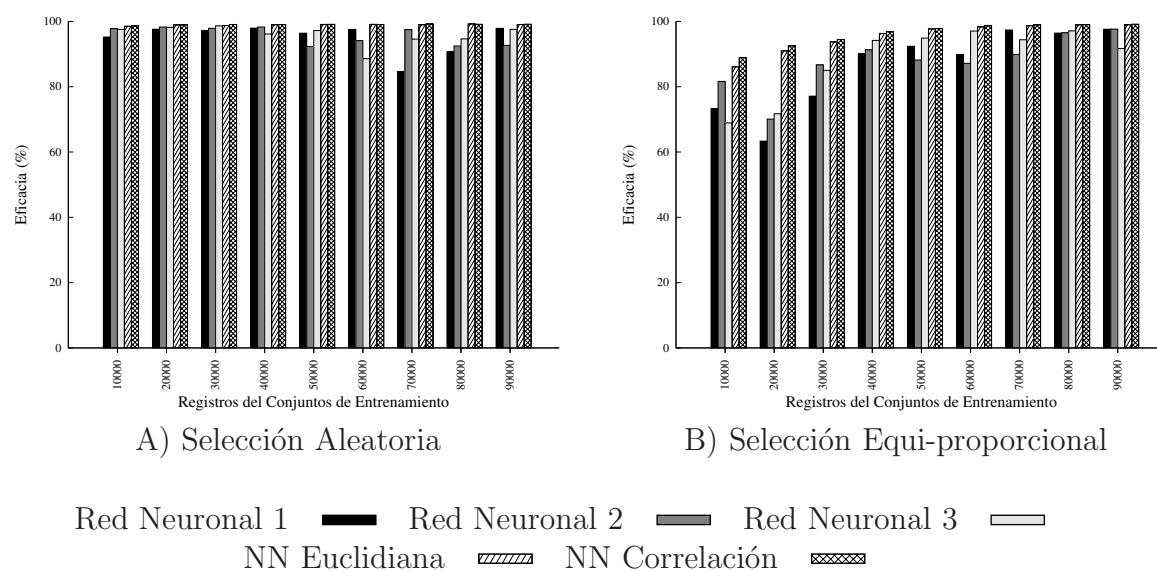


FIGURA 4.9. Eficacia centralizada de la Clase 1

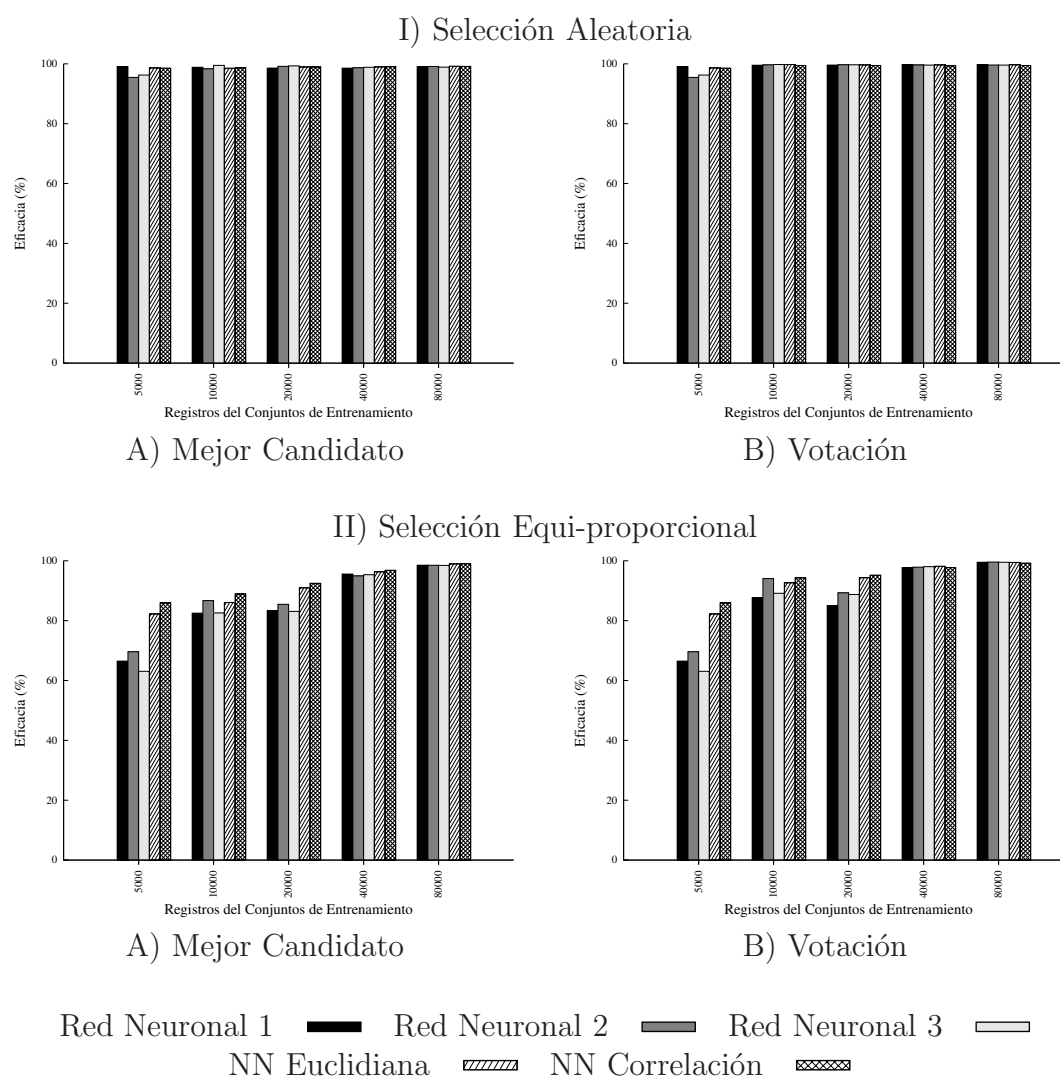


FIGURA 4.10. Eficacia Distribuida Clase 1

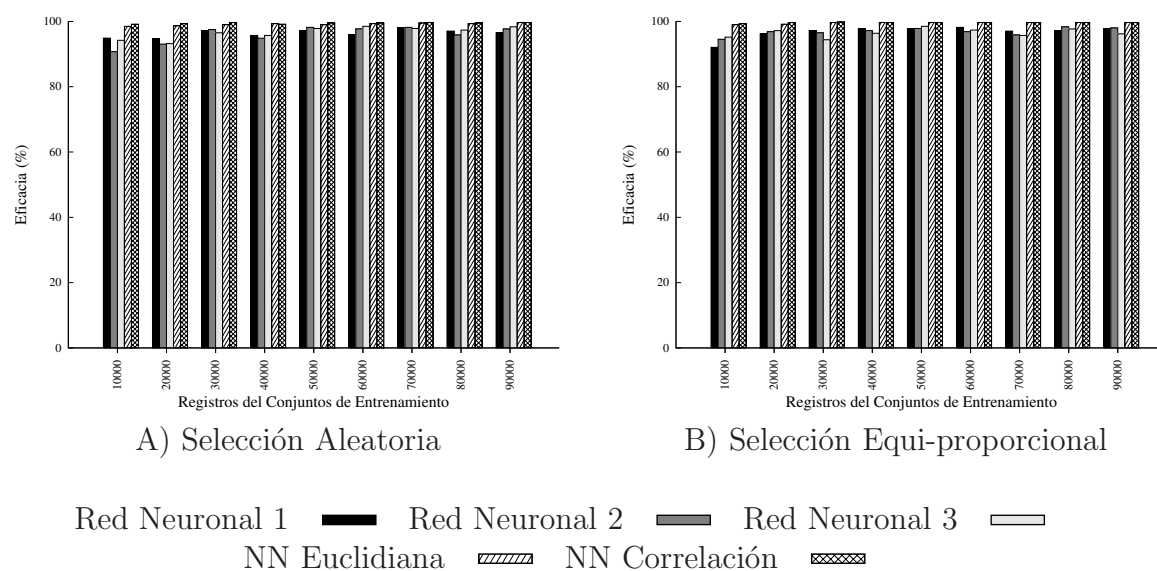


FIGURA 4.11. Eficacia centralizada de la Clase 2

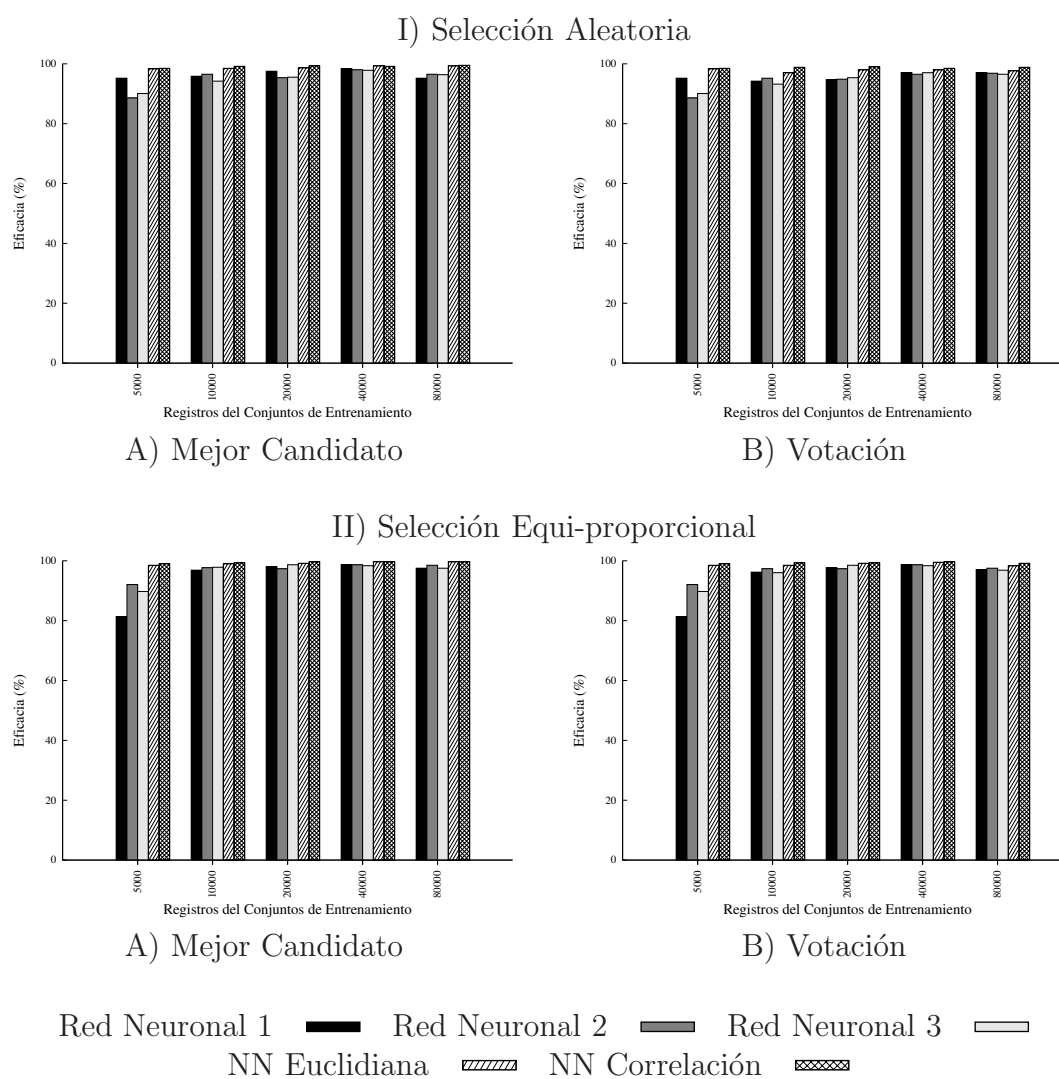


FIGURA 4.12. Eficacia Distribuida Clase 2

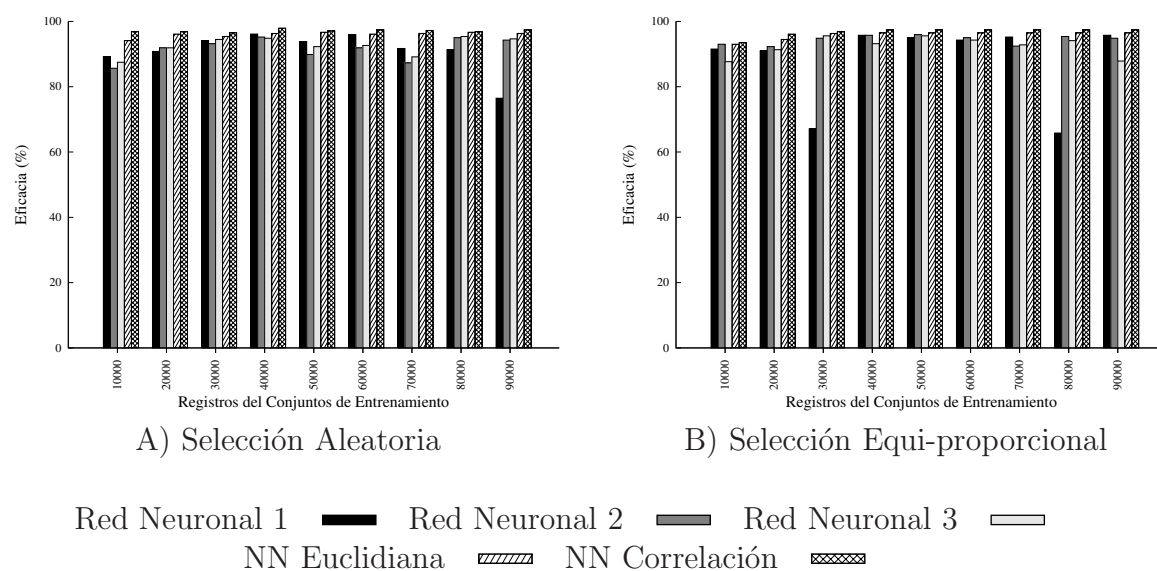


FIGURA 4.13. Eficacia centralizada de la Clase 3

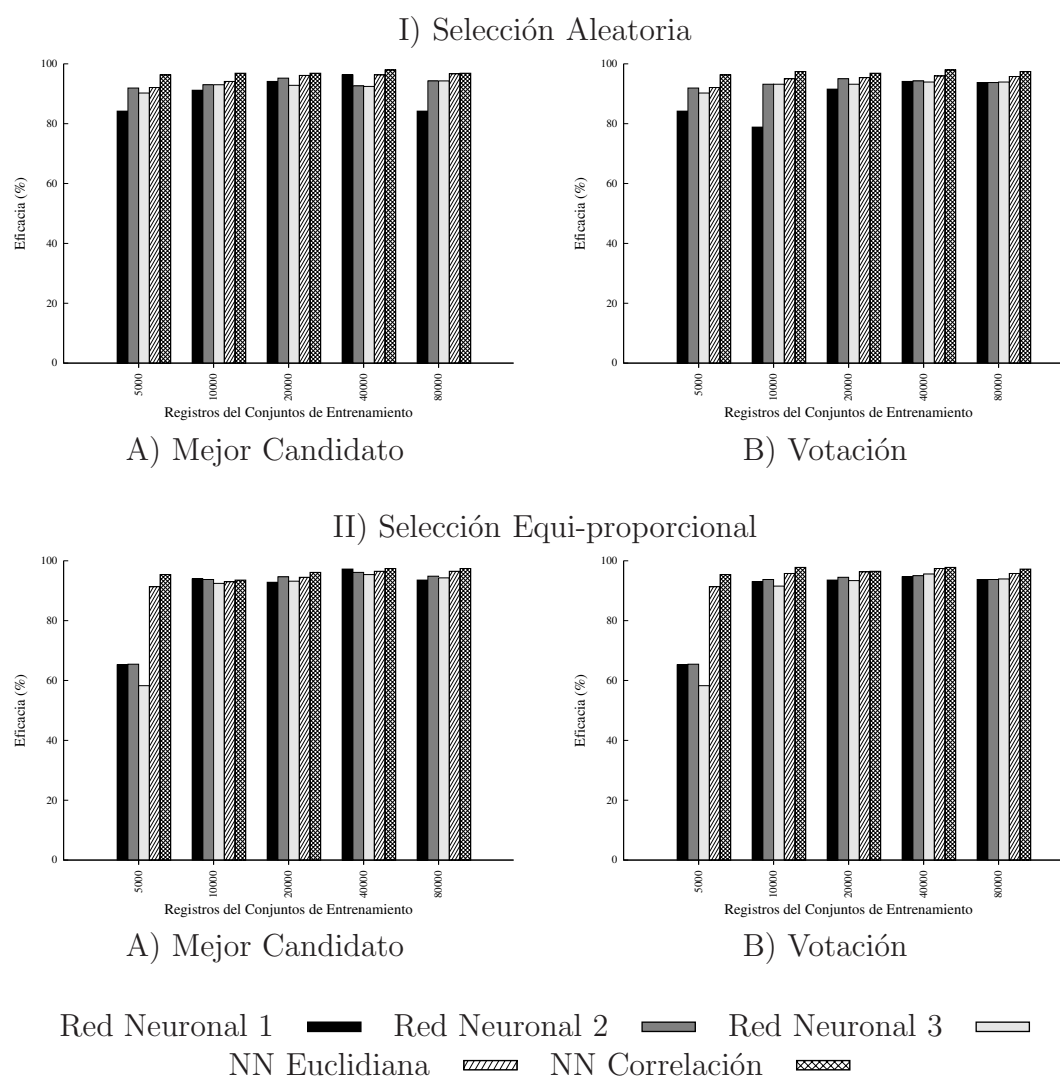


FIGURA 4.14. Eficacia Distribuida Clase 3

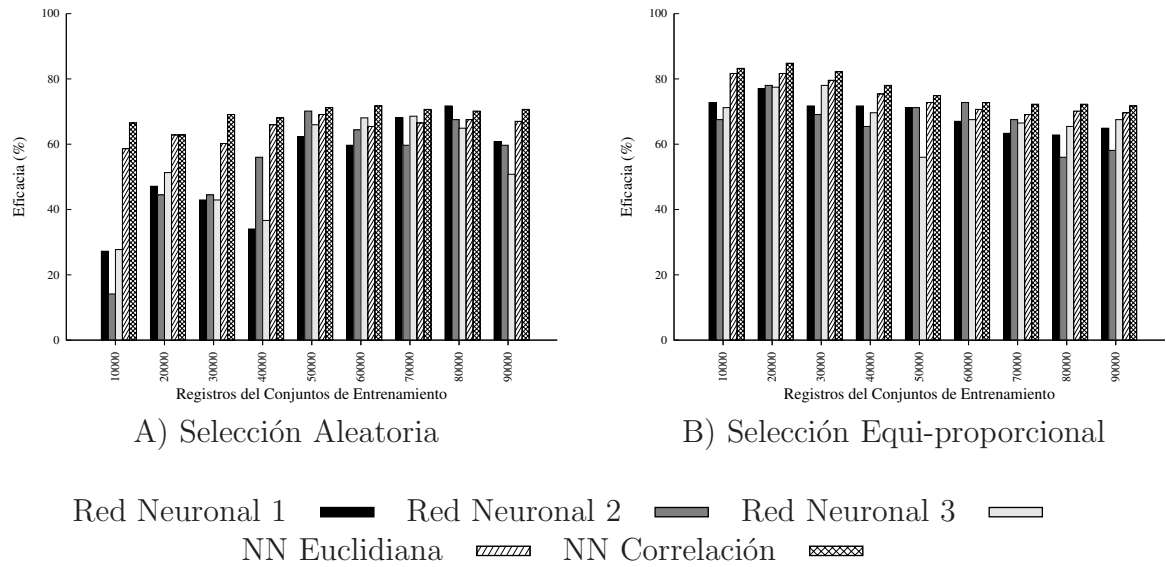


FIGURA 4.15. Eficacia centralizada de la Clase 4

Los resultados de clasificar la Clase 4, se observan en la **Figura 4.15**. De acuerdo con la distribución de esta clase en los conjuntos de referencia aleatorios, esta clase posee mayor representación en el conjunto de 95 mil latidos, con 2, 290 elementos en el conjunto. Esta misma clase, en conjuntos de referencia equi-proporcionados, encuentra su mayor representación en los conjuntos con 20 mil o más elementos. El algoritmo que mejor reconoce a esta clase, bajo un esquema de trabajo centralizado y utilizando conjuntos de referencia aleatorios, es 1-NN-C. En la mayoría de estos resultados, la eficacia lograda por 1-NN-C supera a la eficacia de las redes neuronales, como se indica en la **Figura 4.15.A**. No obstante, al utilizar el conjunto de referencia con 45 mil elementos observamos que la red neuronal con Arquitectura 1 supera a 1-NN, además en el experimento con un conjunto de referencia de 85 mil elementos, la red neuronal con Arquitectura 2 igualó la eficacia de 1-NN-C. Así mismo, el porcentaje de eficacia de los clasificadores utilizando conjuntos de referencia equi-proporcionados, presenta el mayor valor al utilizar el conjunto con 20 mil latidos. En la **Figura 4.15.B**, observamos como al incrementar la cardinalidad de datos, el clasificador es sobre-alimentado.

Utilizando el esquema de trabajo distribuido, observamos un porcentaje de eficacia semejante al obtenido en el esquema centralizado, donde al utilizar conjuntos de referencia aleatorios se alcanza menor eficacia que al utilizar conjuntos equi-proporcionados, como se ilustra en la **Figura 4.16**. Bajo el esquema de trabajo distribuido, observamos que el mayor porcentaje de eficacia en clasificación utilizando conjuntos de referencia aleatorios pertenece a los algoritmos 1-NN, para la clasificación por mejor candidato y por votación, indicado en la **Figura 4.16.I**. En cambio, al utilizar conjuntos de referencia equi-proporcionados el porcentaje de eficacia de las redes neuronales se asemeja al de los algoritmos 1-NN, como se observa en la **Figura 4.16.II**. Además se ilustra como, con conjuntos de referencia de cardinalidad mayor a 20 mil latidos, las redes neuronales son sobre-alimentadas. Así mismo, el porcentaje de eficacia obtenido con el conjunto de referencia de 20 mil latidos, se iguala la eficacia de 1-NN-C (en el caso de la clasificación por mejor candidato) y se supera en el caso de clasificación por votación.

La clasificación bajo un esquema centralizado de la Clase 5, se indica en la **Figura 4.17**. La distribución de latidos de los conjuntos de referencia aleatorios indica que la mayor cantidad de latidos de Clase 5 están presentes en el conjunto con 95 mil latidos, de los cuales 135 pertenecen a esta clase. Por el contrario, la distribución de latidos en conjuntos de referencia equi-proporcionados indica que todos los conjuntos de referencia poseen la mayor representación de esta clase (140 latidos). En la **Figura 4.17.A** se observa como, debido a la poca representación de esta clase en los conjuntos de referencia aleatorios, los clasificadores poseen menor eficacia que la observada en clases mejor representadas *e.g.* Clases 1, 2 y 3. Así mismo, nuestros resultados señalan que el porcentaje de eficacia de los algoritmos 1-NN-C y 1-NN-E es idéntico, en la mayoría de los conjuntos de referencia. De igual manera, los resultados de la clasificación con conjuntos de referencia equi-proporcionados, indican que el incrementar la cantidad de latidos en el conjunto, provoca sobre-alimentar las redes neuronales. Ya que, como se indica en la **Figura 4.17.B**, el mayor porcentaje

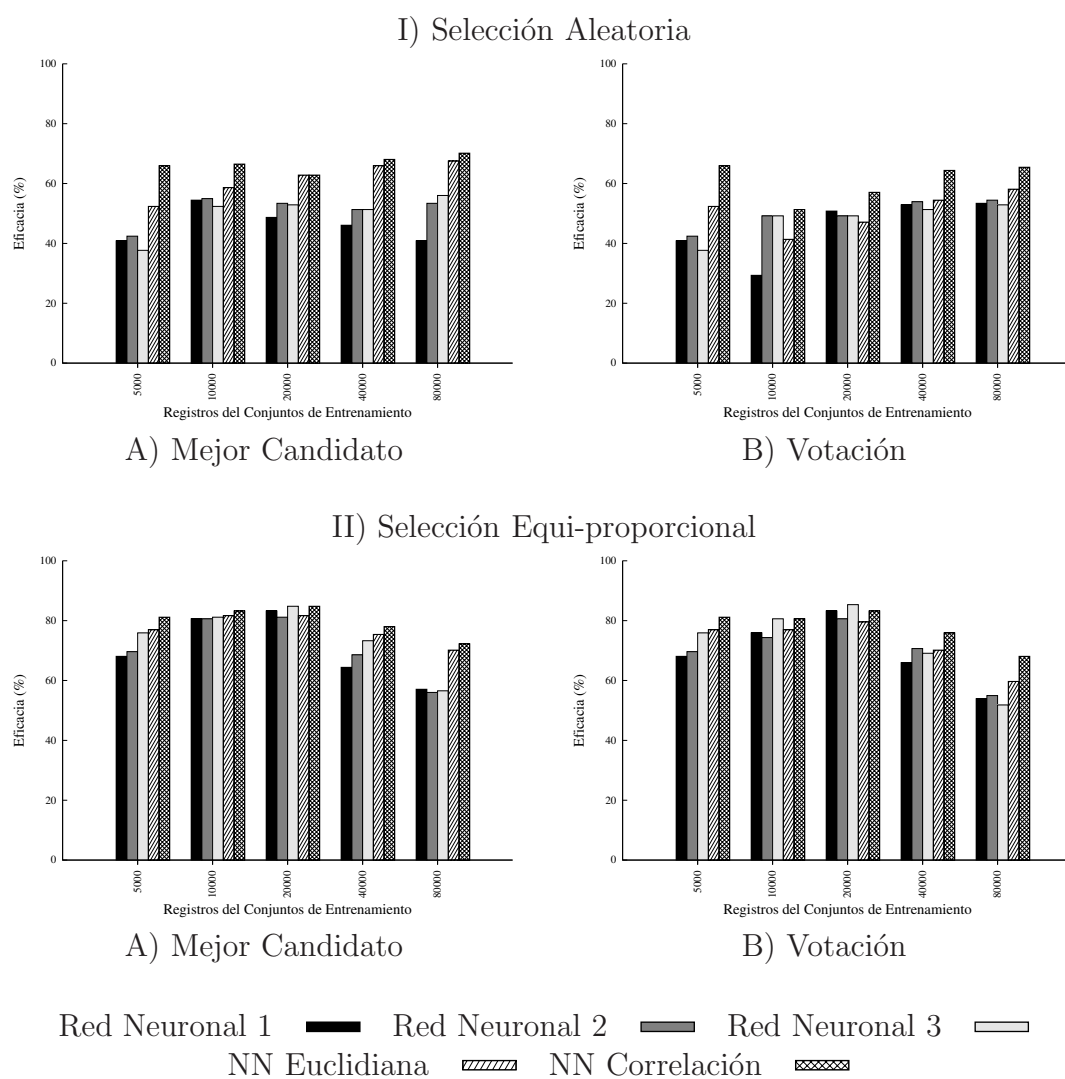


FIGURA 4.16. Eficacia Distribuida Clase 4

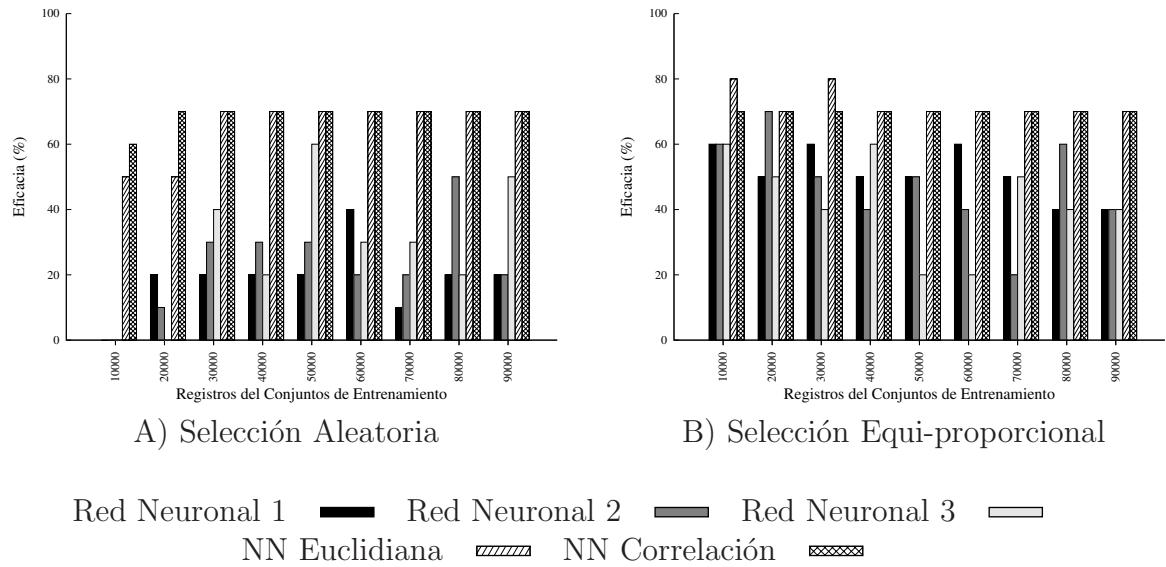


FIGURA 4.17. Eficacia centralizada de la Clase 5

de eficacia de las redes neuronales es obtenido en el conjunto con cardinalidad de 5 mil elementos, donde al incrementar la cardinalidad no se supera la eficacia obtenida.

La ventaja de la selección equi-proporcional también se observa en el esquema de trabajo distribuido, ilustrado en la **Figura 4.18**. Como se indicó anteriormente, la distribución de la Clase 5 no alcanza su máxima representación en los conjuntos de referencia aleatorios, por lo que, al dividir los datos entre los nodos de trabajo, la representación de esta clase se reduce. Esto se observa al comparar la eficacia de la clasificación por mejor candidato (representada por la **Figura 4.18.I.A**), frente a la la eficacia de la clasificación por votación (indicada en la **Figura 4.18.I.B**). Observamos que por medio de una selección aleatoria, la red neuronal con Arquitectura 1 reconoce a la Clase 5. No obstante, al utilizar la clasificación por votación con más de dos nodos de procesamiento (*e.g.* 4, 8 y 16 nodos), observamos que las redes neuronales no son capaces de reconocer esta clase. Por el contrario, al utilizar conjuntos de referencia equi-proporcionados, las redes son capaces de clasificar la Clase 5, ya sea por medio de la clasificación por mejor candidato (indicada en la la **Figura 4.18.II.A**) o por votación (indicada en la la **Figura 4.18.II.B**). Además, en ambos casos se

ilustra como los clasificadores son sobre-alimentados con conjuntos de referencia de cardinalidades mayores a 20 mil latidos.

Los resultados de la clasificación de los latidos de Clase 6 están ilustrados en la **Figuras 4.19 y 4.20**, para los esquemas de trabajo centralizado y distribuido, respectivamente. El mayor porcentaje de eficacia observado corresponde al clasificador 1-NN-C, en ambos esquemas de trabajo. La distribución de los conjuntos aleatorios indica que la mayor cantidad de elementos de Clase 6 se encuentra en el conjunto con 95 mil latidos, de los cuales 76 pertenecen a dicha clase. A diferencia de la distribución de datos en los conjuntos de referencia equi-proporcionados, donde todos los conjuntos poseen 77 latidos de dicha clase; siendo esta la cantidad máxima de latidos disponibles. El porcentaje de eficacia para la clasificación con conjuntos aleatorios (observado en la **Figura 4.19.A**), indica que debido a la baja representación de esta clase en los conjuntos, las redes neuronales difícilmente son capaces de reconocer latidos de Clase 6. Sin embargo, al utilizar conjuntos equi-proporcionados las redes neuronales incrementan su eficacia, como se muestra en la **Figura 4.19.B**. Los resultados obtenidos en la clasificación de la Clase 6, señalan que conjuntos de cardinalidades mayores a 5 mil latidos sobre-alimentan los clasificadores.

Al igual que la Clase 5, donde en el esquema distribuido observamos un bajo porcentaje de eficacia por parte de conjuntos de referencia aleatorios, la eficacia en de los clasificadores es baja comparada con otras clases, como se indica en la **Figura 4.20**. Tanto en la clasificación por mejor candidato (**Figura 4.20.I.A**), así como en la clasificación por votación (**Figura 4.20.I.B**), el porcentaje de eficacia de las redes neuronales es 0, ninguna red neuronal fue capaz de reconocer esta clase. Así mismo, al utilizar conjuntos de referencia equi-proporcionados (como se indica en la **Figura 4.20.II**), las redes neuronales poseen una cantidad mínima de ejemplos que les permite identificar la Clase 6. No obstante, al distribuir los conjunto de referencia equi-proporcionados, para generar las redes neuronales en los nodos, se generan subconjuntos con pocos ejemplos de la Clase 6. Provocando que las redes neuronales; en

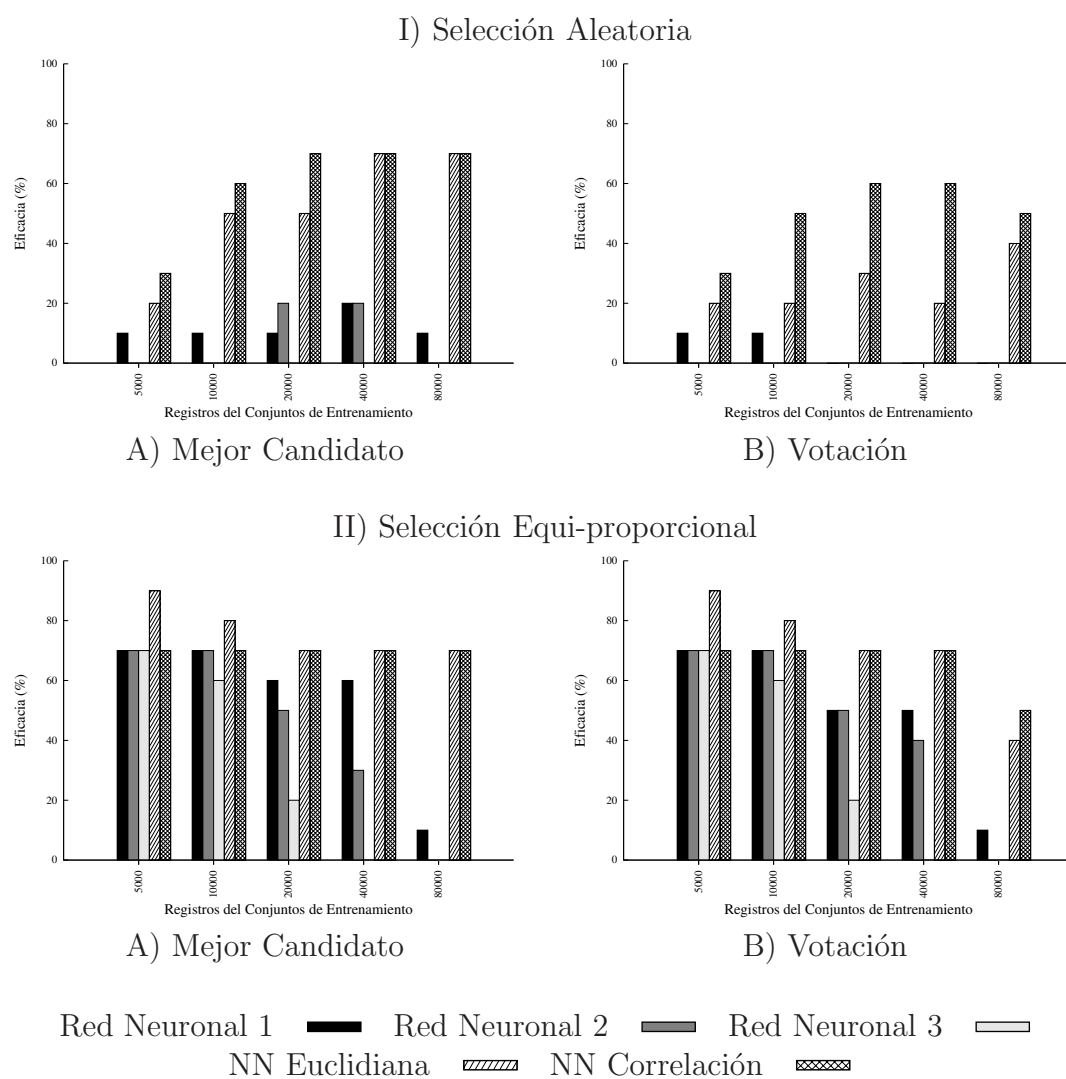


FIGURA 4.18. Eficacia Distribuida Clase 5

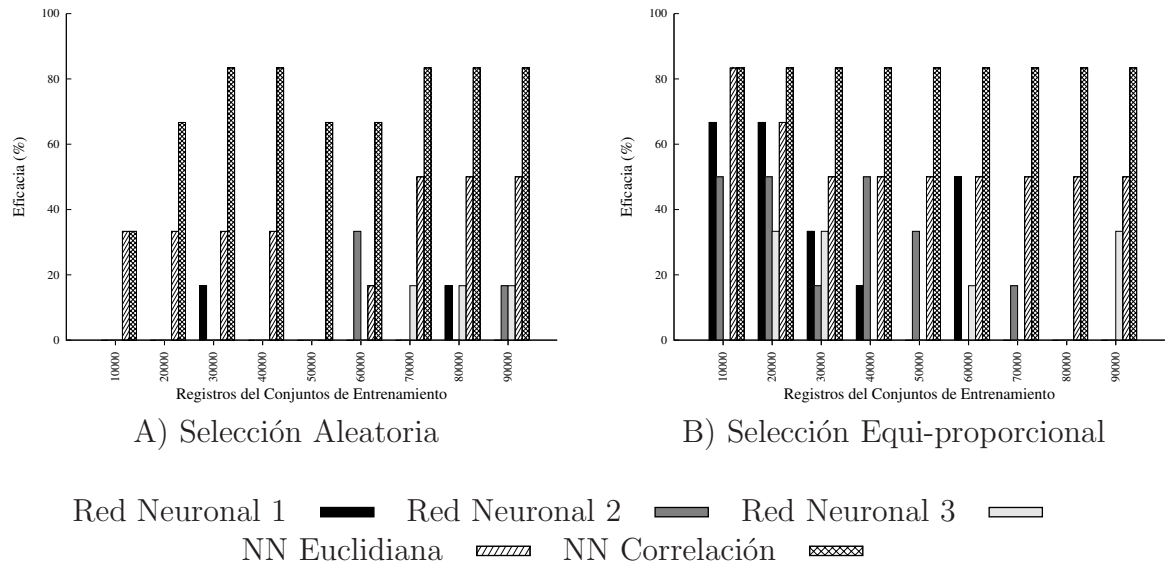


FIGURA 4.19. Eficacia centralizada de la Clase 6

los diferentes nodos, no sean capaces de reconocer dicha clase.

El porcentaje de eficacia de los clasificadores al identificar la Clase 8 se describen en la **Figura 4.21**. Esta clase cuenta con un total de 6,590 latidos disponibles, de los cuales solo 6,410 son seleccionados de manera aleatoria para el conjunto de referencia de 95 mil latidos. No obstante, los conjuntos de referencia equi-proporcionados cuentan con el máximo de latidos en cardinalidades iguales o mayores a 35 mil latidos. Al utilizar el esquema de trabajo centralizado, observamos que el clasificador con mayor eficacia es 1-NN, donde el porcentaje de eficacia utilizando el coeficiente de correlación (1-NN-C) como función de semejanza supera al porcentaje de eficacia utilizando la función de distancia Euclidiana. Este comportamiento se observa en los conjuntos aleatorios (**Figura 4.21.A**) y en conjuntos equi-proporcionales (**Figura 4.21.B**). Así mismo, la **Figura 4.21.B** ilustra como a partir de conjuntos de referencia con cardinalidades mayores a 35 mil latidos las redes neuronales son sobre-alimentadas.

De igual manera, el porcentaje de eficacia de los algoritmos utilizando el esquema de trabajo distribuido se ilustra en la **Figura 4.22**. Al utilizar conjuntos de referencia aleatorios, el clasificador con mayor porcentaje de eficacia es 1-NN-C, como se indica

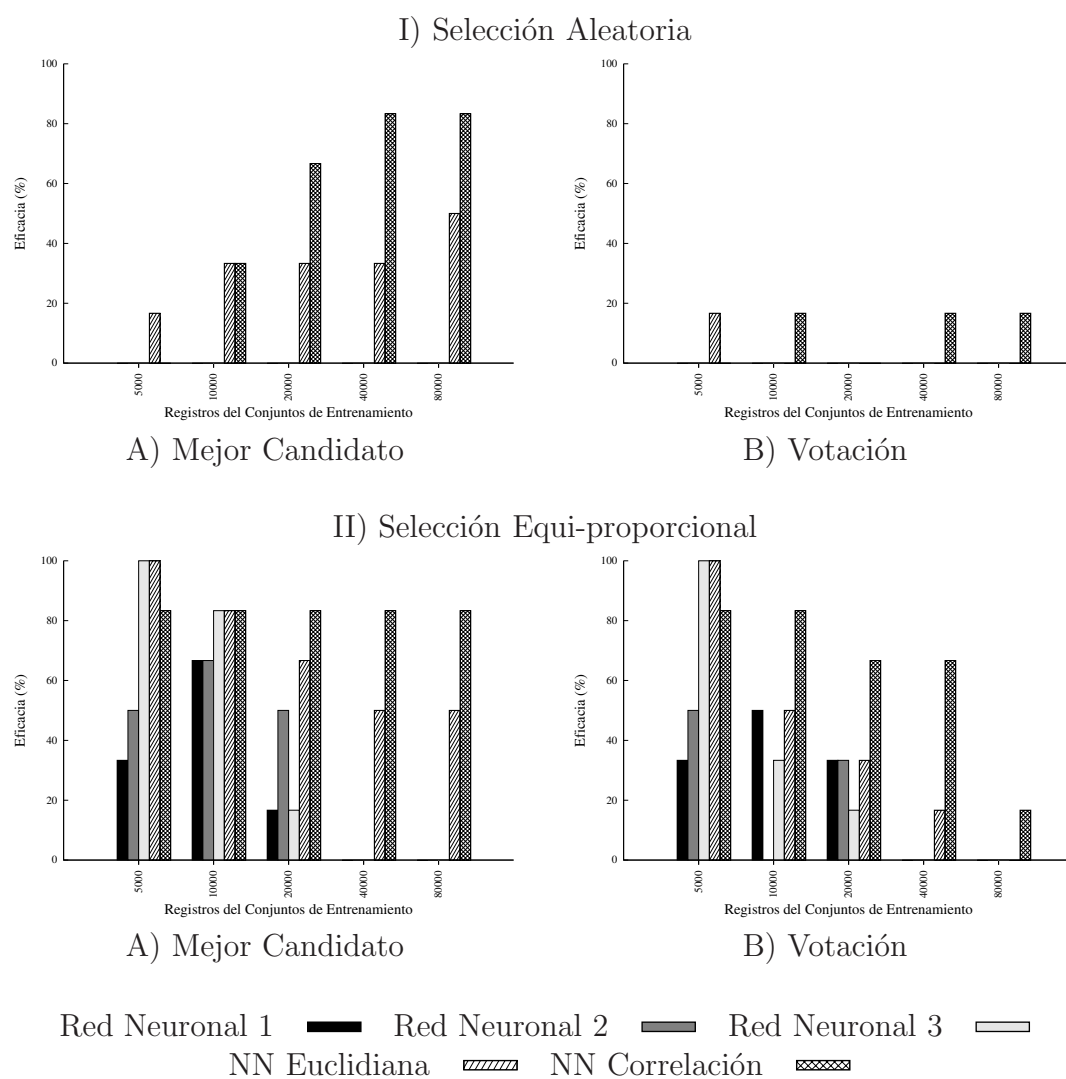


FIGURA 4.20. Eficacia Distribuida Clase 6

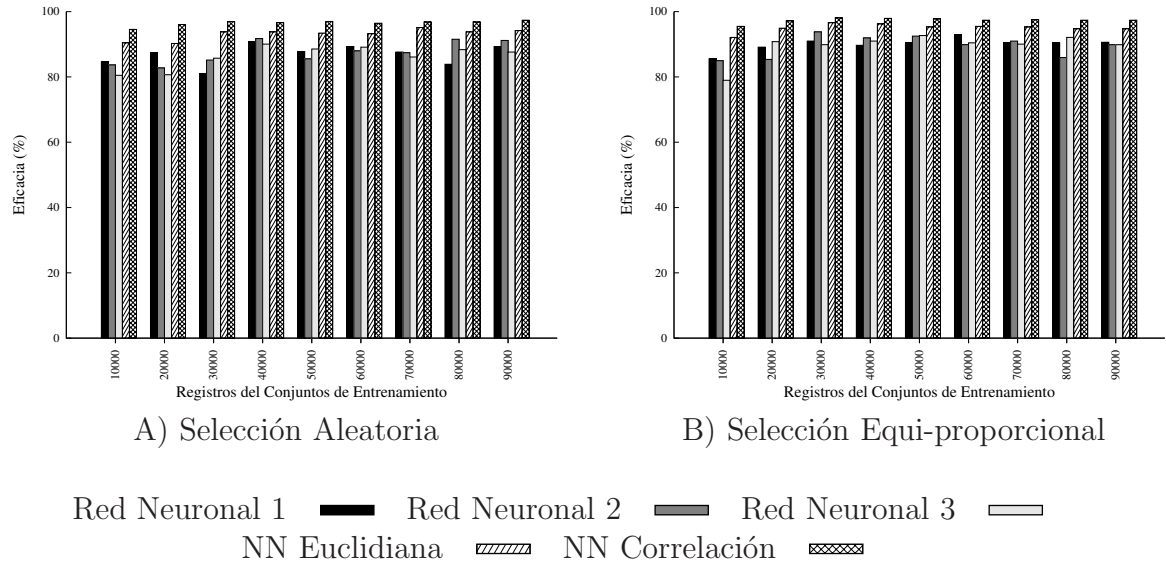


FIGURA 4.21. Eficacia centralizada de la Clase 8

en la **Figura 4.22.I**. En la clasificación por mejor candidato, ilustrada en la **Figura 4.22.I.A**, se observa que el clasificador 1-NN implementando la función de distancia Euclidiana (1-NN-E) presenta poca diferencia con respecto de 1-NN-C. Sin embargo al realizar la clasificación por votación, las redes neuronales artificiales superan a 1-NN-E, como se indica en la **Figura 4.22.I.B**. Por otra parte, la **Figura 4.22.II** ilustra la clasificación utilizando conjuntos de referencia equi-proporcionados, en la cual se indica un mayor porcentaje de eficacia para las redes neuronales. Donde en la clasificación por mejor candidato, la eficacia de las redes neuronales se asemeja a la eficacia del algoritmo 1-NN-E, como se indica en la **Figura 4.22.II.A**. Además, la **Figura 4.22.II.B** muestra como al realizar la clasificación por votación, las redes neuronales superan el porcentaje de eficacia de 1-NN-E, con 4, 8 y 16 nodos.

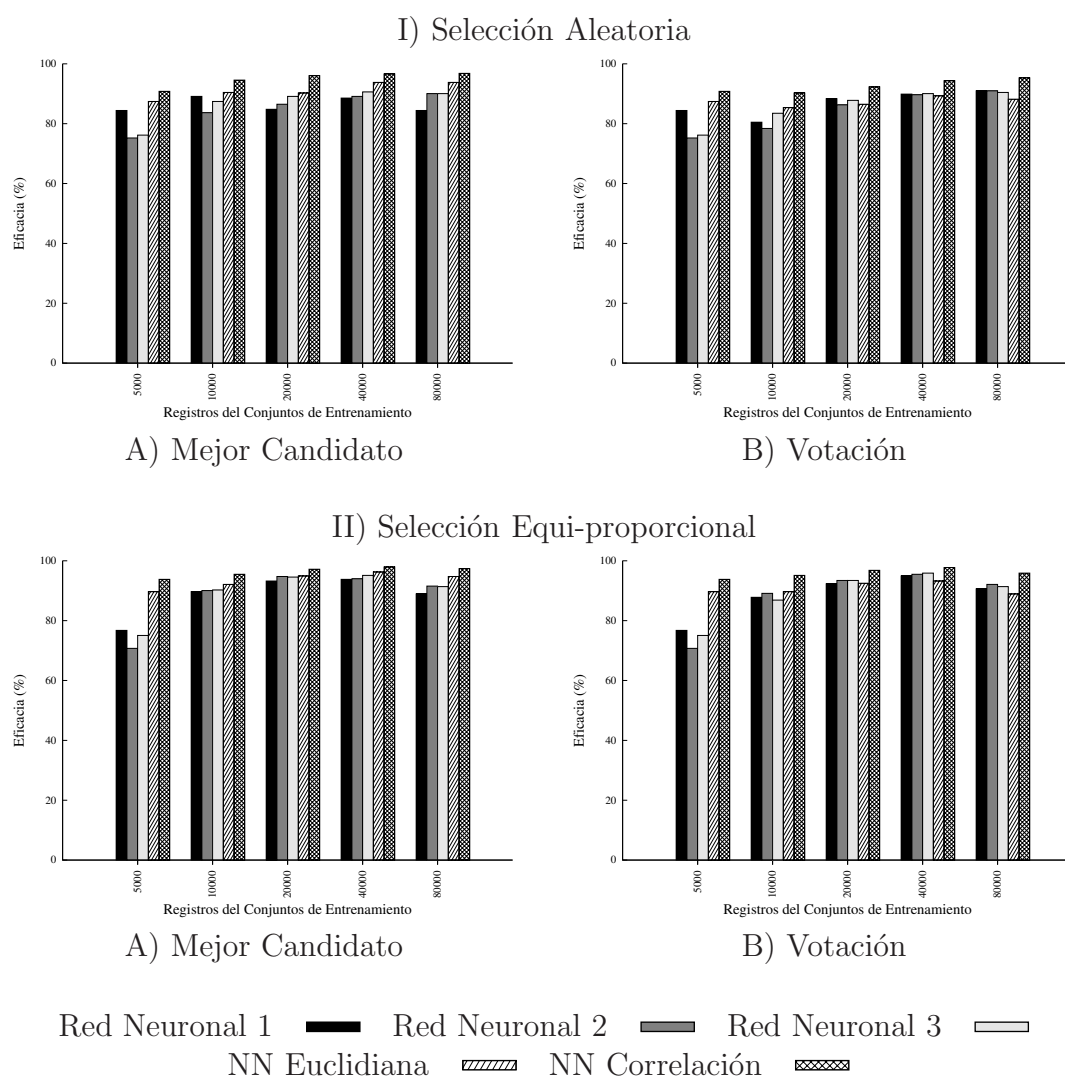


FIGURA 4.22. Eficacia Distribuida Clase 8

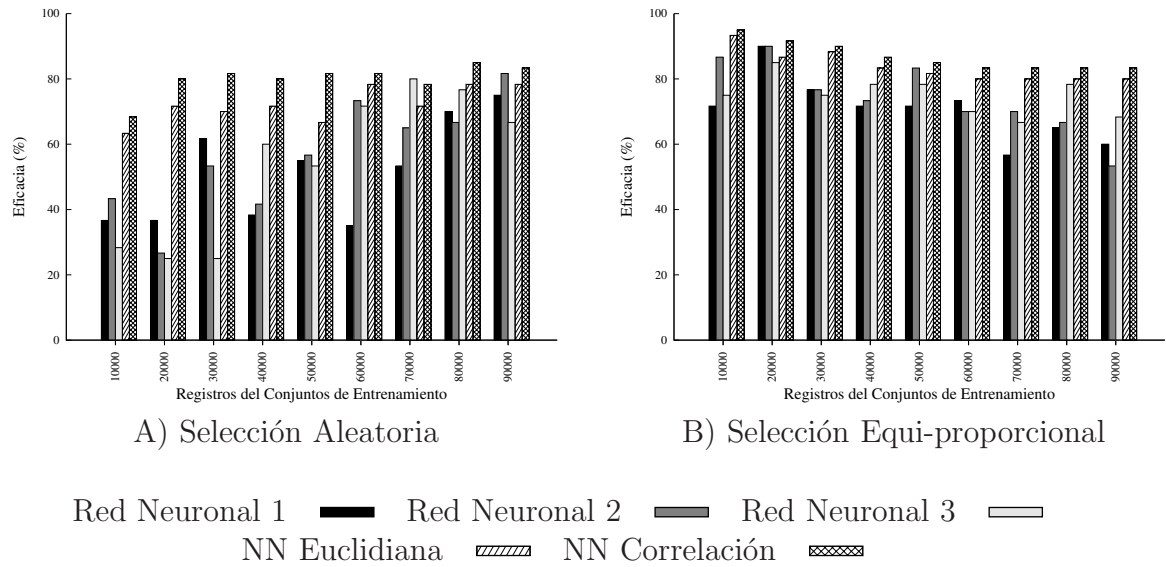


FIGURA 4.23. Eficacia centralizada de la Clase 9

La clasificación de la Clase 9, utilizando el esquema de trabajo centralizado se ilustra en la **Figura 4.23**. Utilizar conjuntos de referencia aleatorios, donde la máxima cantidad de esta clase de latidos se encuentra en el conjunto con cardinalidad de 95 latidos, de los cuales 718 son de clase 9. A diferencia de los conjuntos equi-proporcionados, donde a partir de conjuntos con cardinalidades iguales o mayores a 10 mil latidos estos contienen 743 latidos de Clase 9, que representa el total de latidos de dicha clase. El clasificador con mayor porcentaje de eficacia (en la mayoría de los casos) corresponde a 1-NN-C. Sin embargo, como lo indica la **Figura 4.23.A**, la eficacia del clasificador 1-NN-C es superada por las redes neuronales en los conjuntos de referencia con cardinalidades de 45 y 70 latidos, por las Arquitecturas 1 y 3, respectivamente.

Así mismo, la eficacia obtenida por los clasificadores con conjuntos de referencia equi-proporcionados presenta como estos algoritmos son sobre-alimentados al incrementar la cardinalidad del conjunto. Como la **Figura 4.23.B** indica, a partir de conjuntos con cardinalidad mayor a 10 latidos, el porcentaje de eficacia de los clasificadores 1-NN desciende. En el caso de las redes neuronales, a partir de conjuntos de

referencia con cardinalidades mayores a 20 latidos se observa como los clasificadores son sobre-alimentados, provocando que el porcentaje de eficacia de dichos clasificadores no supere aquél obtenido en este conjunto de referencia.

El porcentaje de eficacia de los clasificadores utilizando el esquema de trabajo distribuido se observa en la **Figura 4.24**. Como resultado de la poca representación de esta clase en los conjuntos de referencia aleatorios, se obtiene un mayor porcentaje de eficacia al realizar la clasificación por mejor candidato, comparada con la clasificación por votación, como se indica en la **Figura 4.24.I**. En el caso de la clasificación por mejor candidato, se observa como al incrementar la cardinalidad del conjunto y la cantidad de nodos el porcentaje de eficacia del clasificador aumenta (como se indica en la **Figura 4.24.I.A**), a diferencia de la clasificación por votación, donde se observa que a partir del conjunto con 20 mil latidos los clasificadores 1-NN son sobre-alimentados, observado en la **Figura 4.24.I.B**.

No obstante, al utilizar conjuntos de referencia equi-proporcionados con 4 y 8 nodos, se observa que las redes neuronales presentan un porcentaje de eficacia cercano a los algoritmos 1-NN. Inclusive, en la clasificación por mejor vecino observamos que las redes neuronales con Arquitecturas 1 y 2 superan al resto de los clasificadores utilizando 2 y 8 nodos respectivamente, esto es ilustrado por la **Figura 4.24.II.A**. De igual manera, la **Figura 4.24.II.B** ilustra como al aumentar la cantidad de nodos en 4, 8 y 16, los clasificadores son sobre-alimentados y el porcentaje de eficacia de los clasificadores disminuye, con conjuntos de referencia equi-proporcionados.

El porcentaje de eficacia obtenido en la clasificación de la Clase 11 a través de un esquema centralizado, se ilustra en la **Figura 4.25**. El bajo porcentaje de eficacia obtenido en conjuntos de referencia aleatorios es causado por la poca representación de esta clase en dicho conjunto, donde de los 212 latidos disponibles de esta clase, los conjuntos de referencia obtienen 209 en el conjunto de referencia con 95 mil latidos. Por el contrario, todos los conjuntos de referencia equi-proporcionados poseen los 212 latidos de esta clase. El clasificador con mayor porcentaje de eficacia (en la mayoría

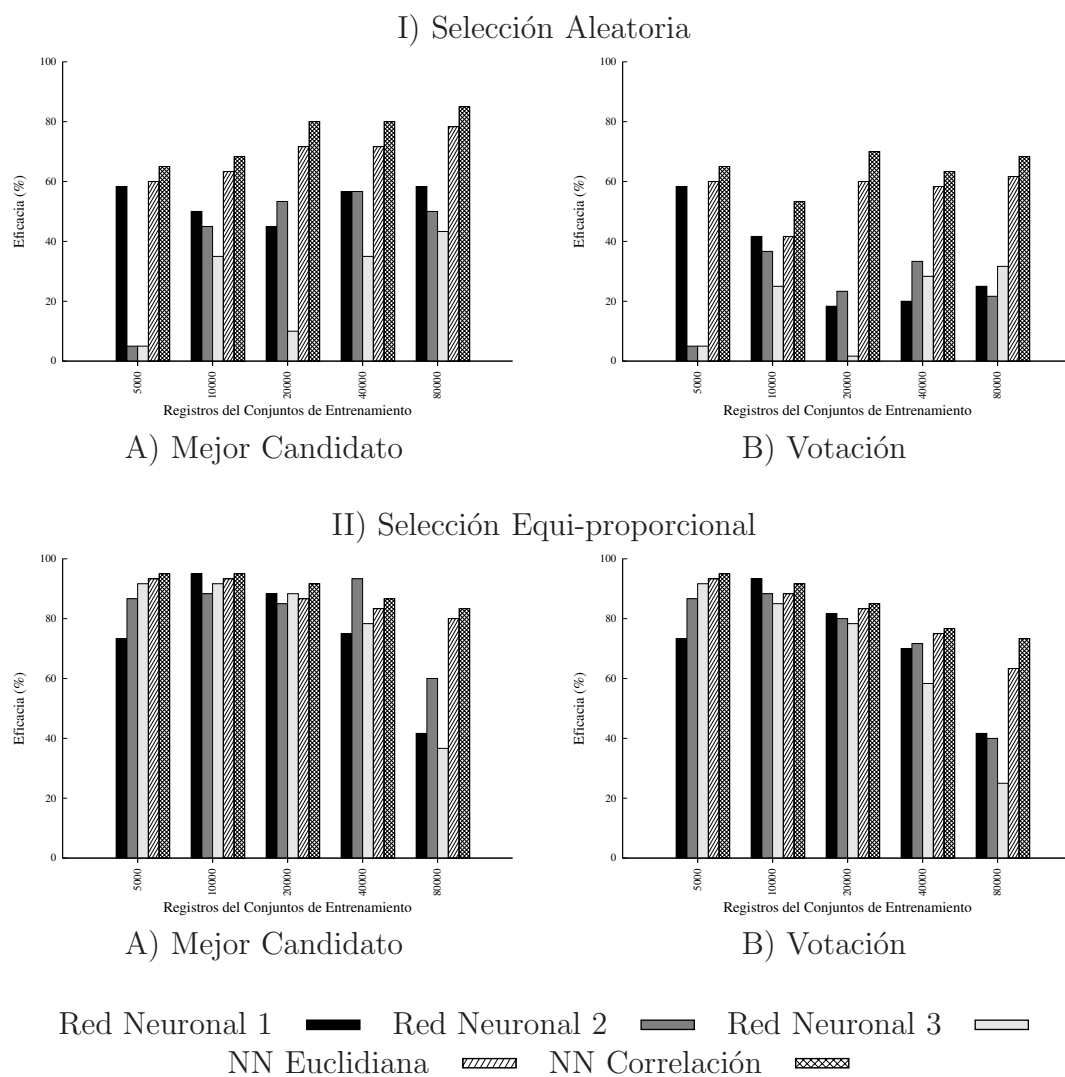


FIGURA 4.24. Eficacia Distribuida Clase 9

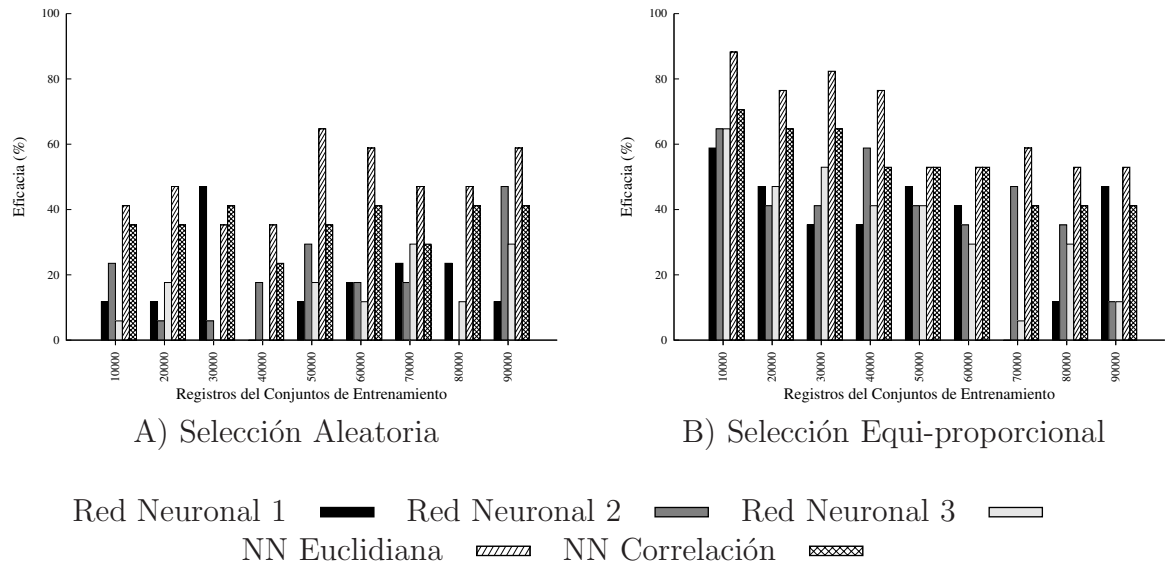


FIGURA 4.25. Eficacia centralizada de la Clase 11

de los casos) corresponde a 1-NN utilizando la función de distancia Euclidiana. Como se ilustra en la **Figura 4.25.A**, al utilizar conjuntos de referencia aleatorios, donde la representación de la clase es baja, las redes neuronales obtienen un bajo porcentaje de eficacia comparado con los clasificadores 1-NN. En cambio, en conjuntos de referencia equi-proporcionados, donde la clases (por menor cantidad de latidos que posean) siempre están representadas observamos que la eficacia de las redes neuronales supera a aquella obtenida a través de conjuntos de referencia aleatorios. Además, como es visto en la **Figura 4.25.B**, se ilustra como la incrementar la cardinalidad del conjunto de referencia equi-proporcionado los clasificadores son sobre-alimentados. En el caso de los algoritmos 1-NN y la red neuronal con Arquitectura 1, la mayor eficacia corresponde a los conjuntos de 5 mil elementos. De igual manera, las redes neuronales con Arquitecturas 2 y 3 son sobre-alimentadas con los conjuntos de 10 mil elementos.

Así mismo, al utilizar el esquema de trabajo distribuido observamos que la baja representación de la Clase 11 en los conjuntos de referencia aleatorios provoca que al distribuir el conjunto entre los nodos existan nodos que no sean capaces de reconocer la clase. Como se indica en la **Figura 4.26.II**, ya sea utilizando la clasificación

por mejor candidato (**Figura 4.26.II.A**) o por votación (**Figura 4.26.II.B**), la eficacia de los clasificadores se encuentra por debajo del 60 %. De igual manera, utilizar conjuntos de referencia equi-proporcionados en un esquema de trabajo distribuido, presenta mayor porcentaje de eficacia al clasificar la Clase 11, como se indica en la **Figura 4.26.II**. No obstante, a pesar de que la clase se encuentra representada, observamos que utilizar la clasificación por votación involucrando todos los nodos, el porcentaje de eficacia se reduce a medida que la cantidad de nodos aumenta, como se observa en la **Figura 4.26.II.B**.

El porcentaje de eficacia para la Clase 12, obtenido por los clasificadores utilizando un esquema de trabajo centralizado se observa en el **Figura 4.27**. Esta clase cuenta con 100 latidos, de los cuales 98 fueron seleccionados para el conjunto de referencia aleatorio con 95 mil latidos, a diferencia de los conjuntos de referencia equi-proporcionados, donde todos los conjuntos poseen todos los latidos de esta clase. Debido a la distribución de los datos en los conjuntos de referencia aleatorios, observamos que el porcentaje de eficacia de las redes neuronales, utilizando dichos conjuntos, es menor que el porcentaje obtenido por los algoritmos 1-NN, como se indica en la **Figura 4.27.A**. No obstante, al utilizar conjuntos de referencia equi-proporcionados el porcentaje de eficacia de las redes neuronales iguala a los algoritmos 1-NN en la mayoría de los casos, obteniendo un 100 % de eficacia al clasificar latidos de Clase 12. Así mismo, se observa en la **Figura 4.27.B** que a partir del conjunto de referencia con cardinalidad de 5 mil latidos, todas las arquitecturas de redes neuronales alcanzan el mayor porcentaje de eficacia.

Al utilizar el esquema de trabajo distribuido, como se indica en la **Figura 4.28**, observamos que los algoritmos 1-NN presentan el mayor porcentaje de eficacia al utilizar conjuntos de referencia aleatorios. La clasificación por mejor candidato, ilustrada en la **Figura 4.28.I.A**, indica que al aumentar la cantidad de nodos, ambas funciones de semejanza (distancia Euclidiana y Coeficiente de Correlación) presentan el mismo porcentaje de eficacia. Sin embargo, involucrar a todos los nodos por medio de la cla-

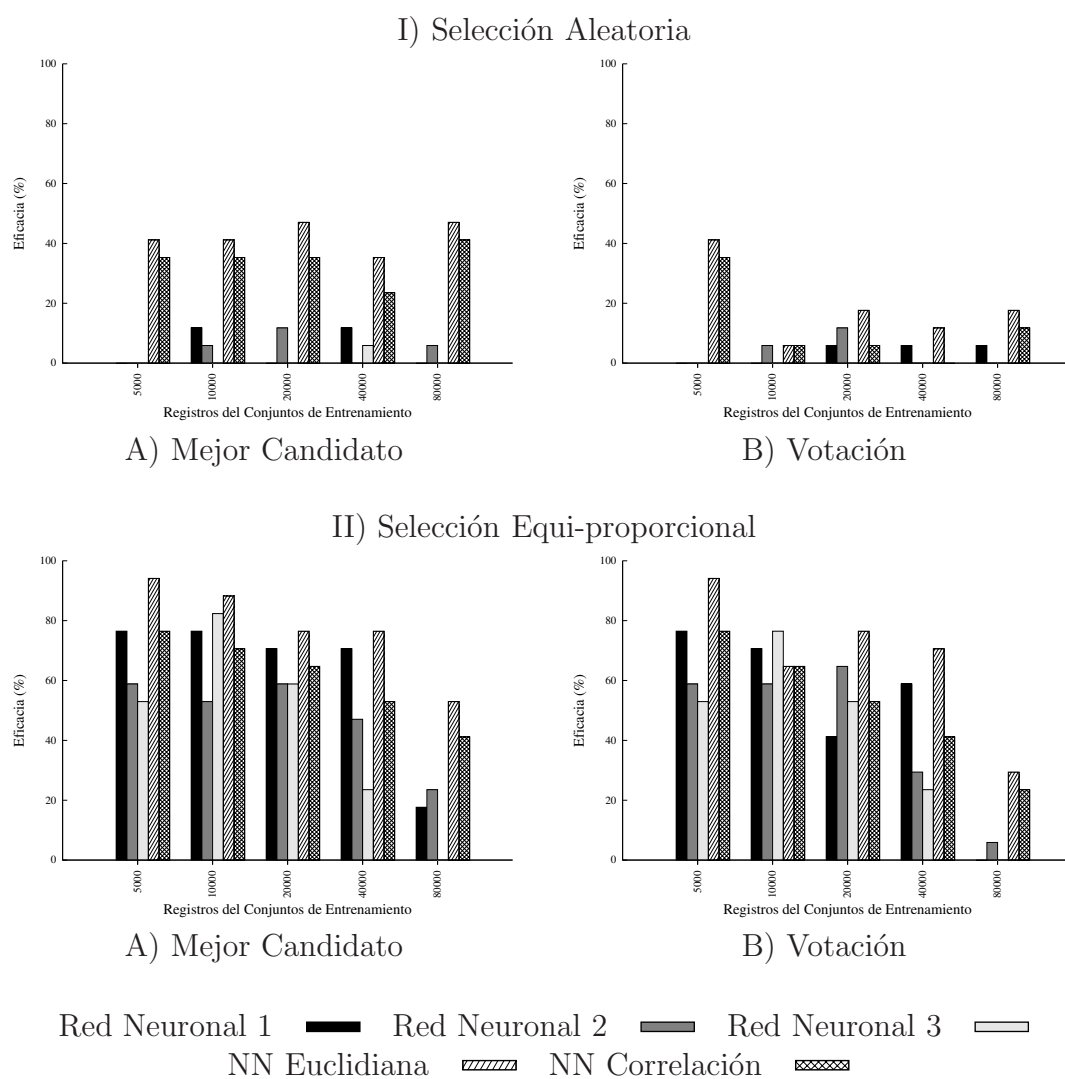


FIGURA 4.26. Eficacia Distribuida Clase 11

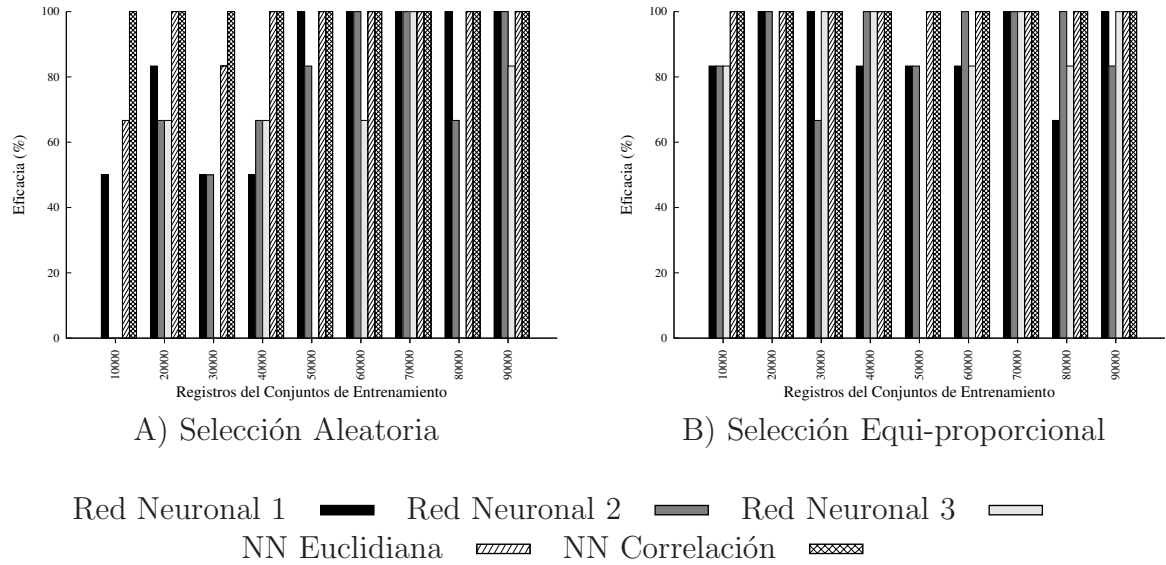


FIGURA 4.27. Eficacia centralizada de la Clase 12

sificación por votación señala que el Coeficiente de Correlación supera a la distancia Euclidiana utilizando 2, 4 y 8 nodos, como se indica en la **Figura 4.28.I.B.**. Por otra parte, se observa que las redes neuronales no igualan el porcentaje de eficacia de los algoritmos 1-NN, debido a la distribución de latidos entre los nodos involucrados, ya sea clasificando con el mejor candidato o por medio de votación.

La Clase 13 cuenta con 3,550 latidos disponibles, de los cuales los conjuntos de referencia aleatorios poseen como máximo 3,463 latidos de esta clase en el conjunto con cardinalidad de 95 mil latidos. No obstante, en conjuntos de referencia equi-proporcionados, podemos observar que la clase presenta su mayor representación a partir de conjuntos con cardinalidades superiores a los 20 mil latidos. El porcentaje de eficacia obtenido por al utilizar conjuntos de referencia aleatorios supera el 80 % en todos los casos, logrando inclusive el 100 % de eficacia, como se observa en la **Figura 4.29.A**. Sin embargo esto solamente sucede en el conjunto con cardinalidad de 65 mil latidos. Así mismo, el porcentaje de eficacia utilizando conjuntos equi-proporcionados alcanza valores de 100 %, a partir del conjunto de referencia con 30 mil latidos, como lo indica **Figura 4.29.B**. Observando que para conjuntos de referencia

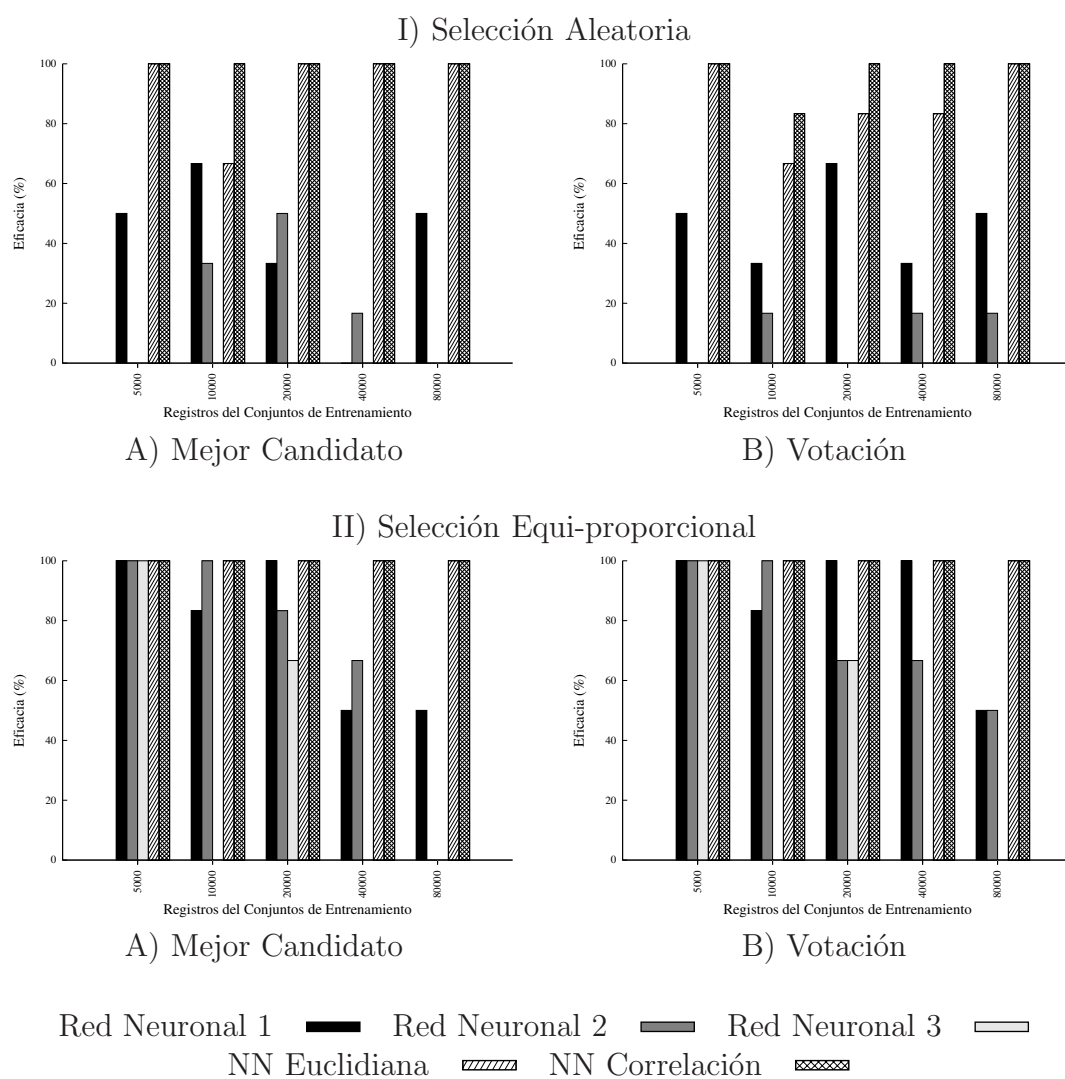


FIGURA 4.28. Eficacia Distribuida Clase 12

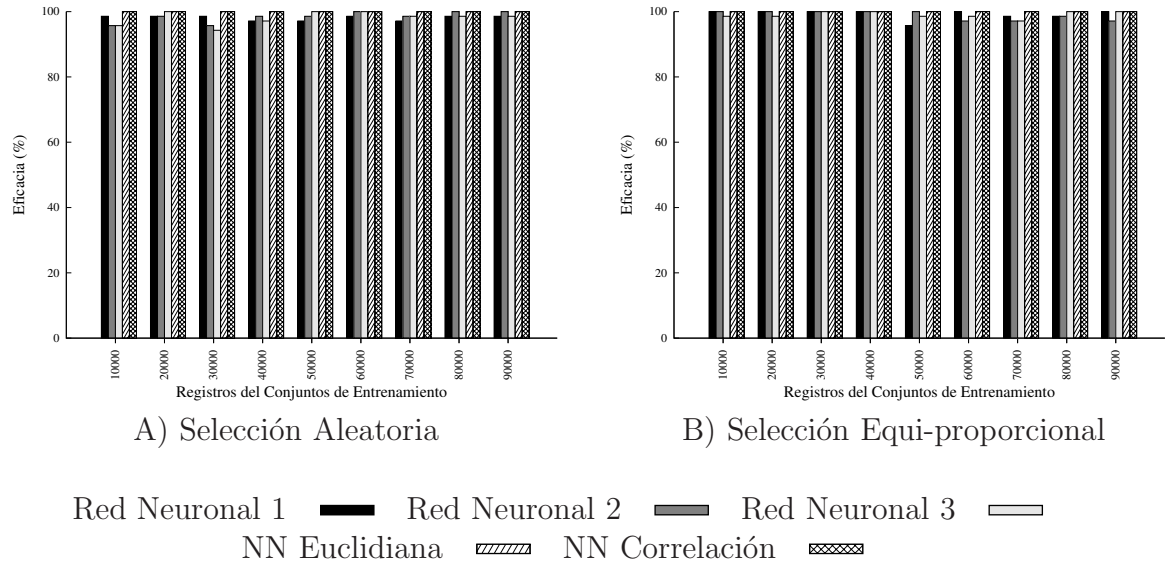


FIGURA 4.29. Eficacia centralizada de la Clase 13

con cardinalidades superiores a 40 mil latidos, los clasificadores son sobre-alimentados.

De igual manera, al utilizar el esquema de trabajo distribuido y conjuntos de referencia aleatorios, observamos que la **Figura 4.30.I** sugiere a los algoritmos 1-NN como los clasificadores con mayor eficacia. Por el contrario, realizar la clasificación utilizando conjuntos de referencia equi-proporcionados, las redes neuronales incrementan su porcentaje de eficacia, como es visto en la **Figura 4.30.II**. Además, se observa que los algoritmos implementando la clasificación por votación las redes neuronales alcanzan los niveles de eficacia de los algoritmos 1-NN.

Finalmente, la Clase 14 tiene un total de 240 latidos, de los cuales 234 es la máxima cantidad seleccionada de manera aleatoria para el conjunto de referencia con 95 mil latidos. A diferencia de los conjuntos de referencia equi-proporcionados, donde todos los conjuntos poseen los 240 latidos. Los clasificadores 1-NN, presentan los mayores porcentajes de eficacia, donde 1-NN-C supera en todos los casos a 1-NN-E, como se indica en la **Figura 4.31**. Así mismo, la **Figura 4.31.B** indica que al utilizar conjuntos de referencia equi-proporcionados observamos que el porcentaje de eficacia de las redes neuronales es mayor con respecto al obtenido con conjuntos aleatorios.

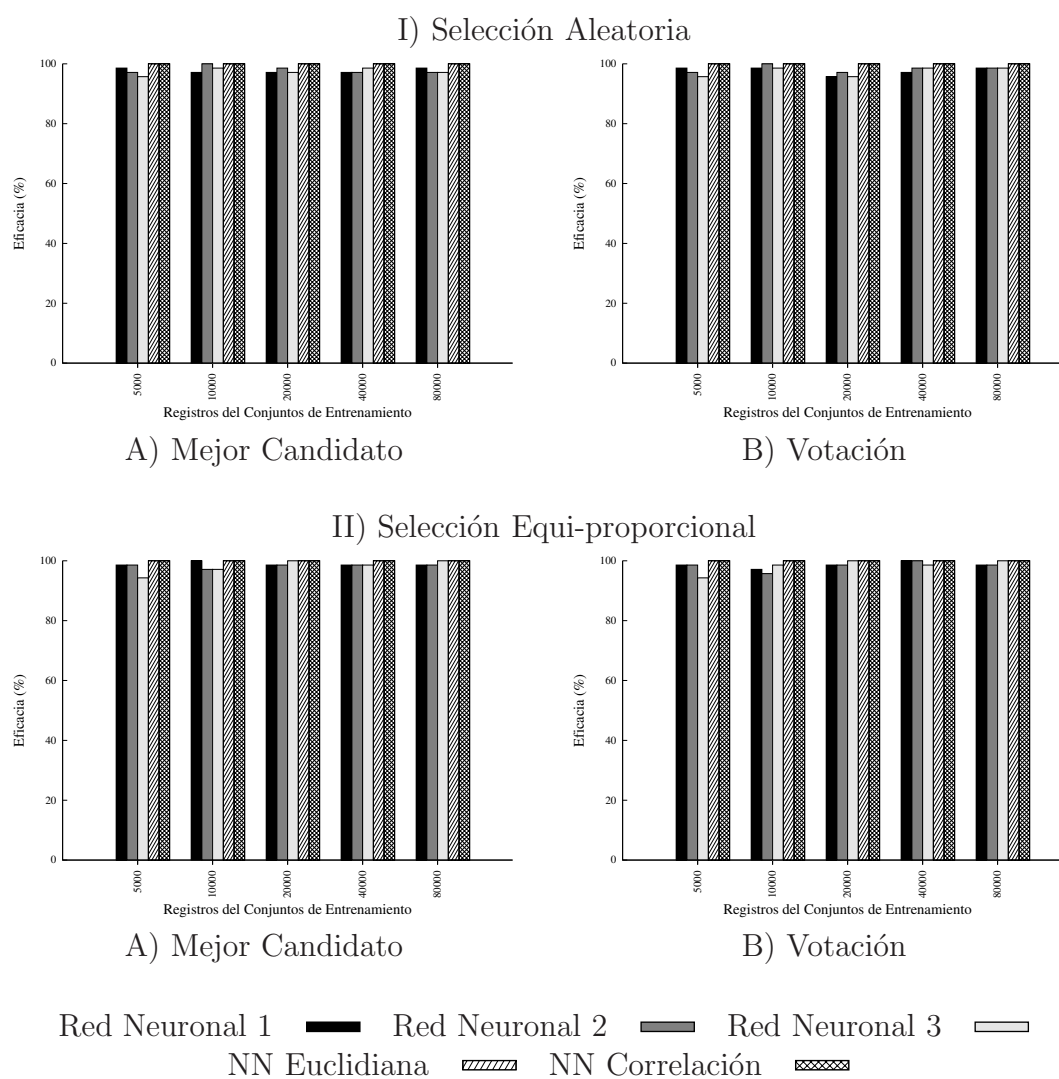


FIGURA 4.30. Eficacia Distribuida Clase 13

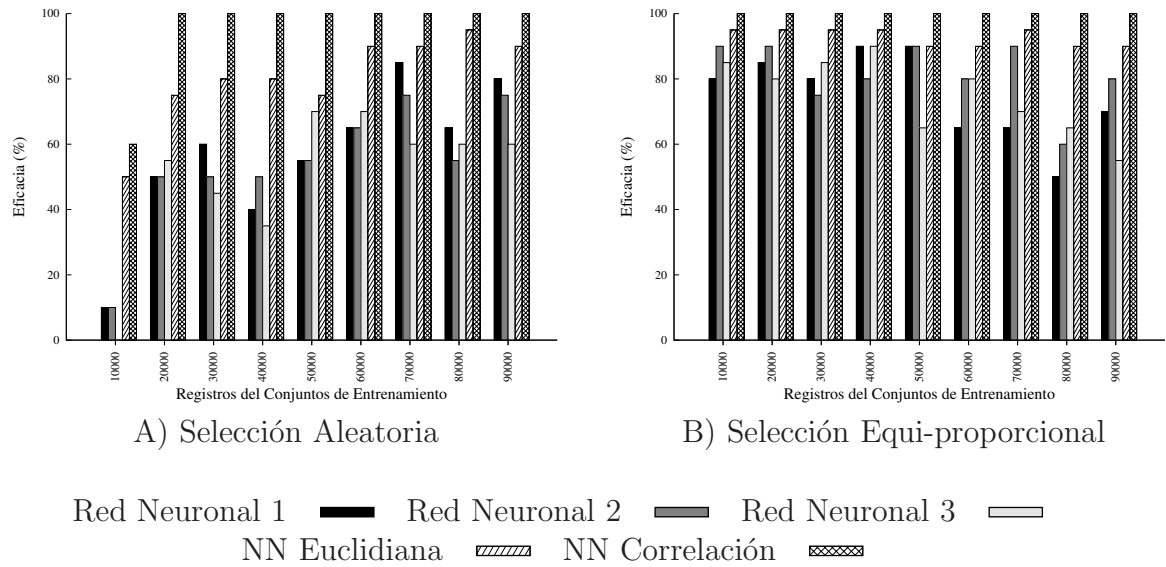


FIGURA 4.31. Eficacia centralizada de la Clase 14

El esquema de trabajo distribuido es ilustrado por la **Figura 4.32**. Los resultados indican que debido a la representación de la Clase 14 y a la distribución entre los nodos de los conjuntos de referencia aleatorios, las redes neuronales presentan baja eficacia comparada con la obtenida por los clasificadores 1-NN. Así mismo, observamos en la **Figura 4.32.I.A**, que al incrementar la cantidad de nodos y utilizando la clasificación por mejor candidato, la diferencia en el porcentaje de eficacia entre 1-NN-C y 1-NN-E se reduce. Por el contrario, utilizar la clasificación por votación provoca que tanto la eficacia de 1-NN-C y 1-NN-E se reduzca, además de no favorecer a este último, como se indica en la **Figura 4.32.I.B**. Al igual que en el esquema centralizado, las redes neuronales obtienen un mayor porcentaje de eficacia al clasificar la Clase 14 utilizando conjuntos de referencia equi-proporcionados. Sin embargo, como la **Figura 4.32.II** lo indica, una cantidad mayor a dos nodos provoca que la eficacia de las redes decaiga.

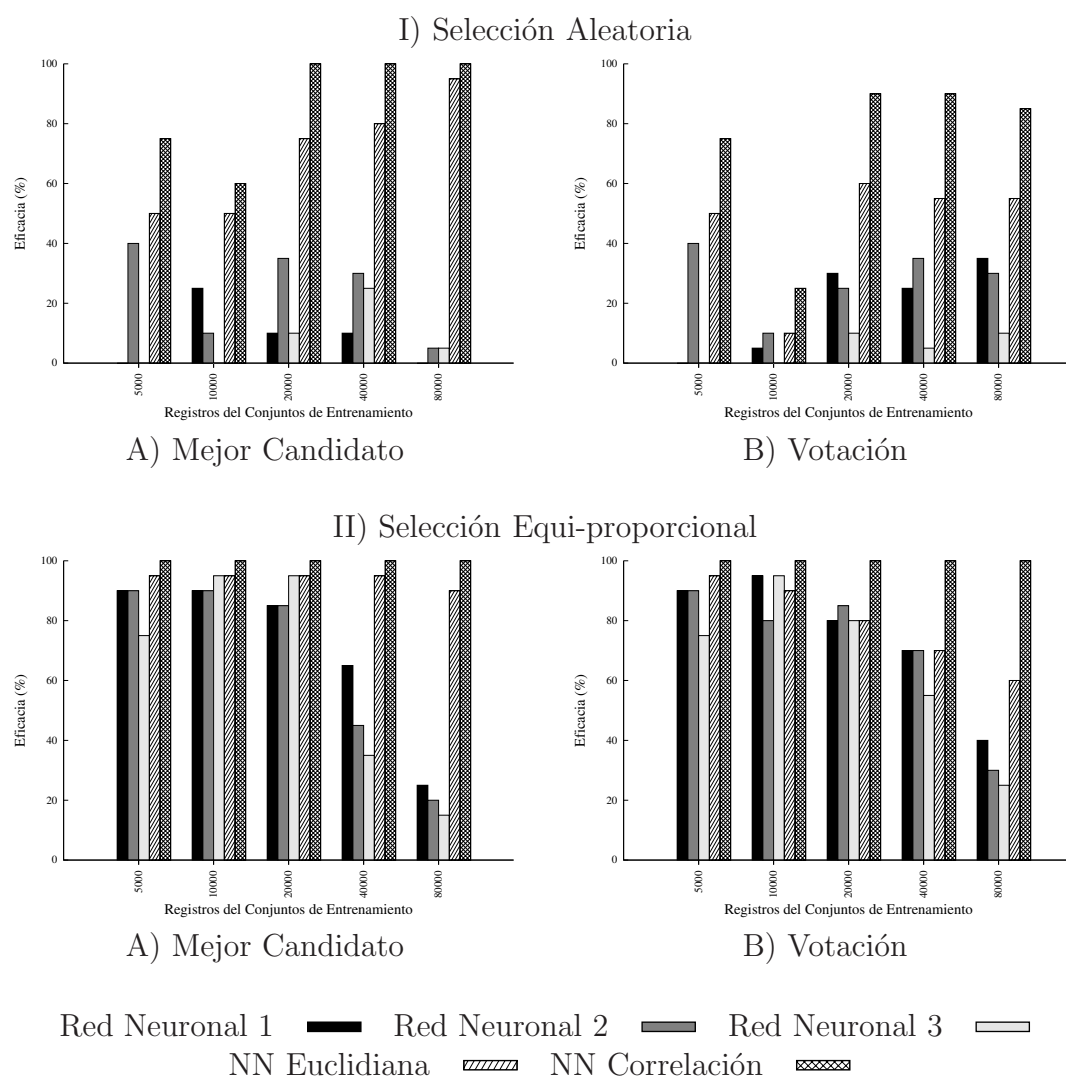


FIGURA 4.32. Eficacia Distribuida Clase 14

4.5.4. Eficiencia

Los tiempos de cómputo de los clasificadores se encuentran expresados en la **Figura 4.33**. Esta gráfica corresponde al tiempo que le toma a los clasificadores identificar un latido (cabe mencionar que los resultados de esta gráfica están expresado en escala logarítmica). Debido a las diferencias en los tiempos de cómputo, la gráfica solo permite apreciar la eficiencia de los algoritmos 1-NN, en los esquemas de trabajo distribuido y centralizado. Así mismo, se observa que al implementar un esquema de trabajo distribuido e incrementar la cantidad de unidades de procesamiento, el tiempo es menor que el observado en el esquema centralizado.

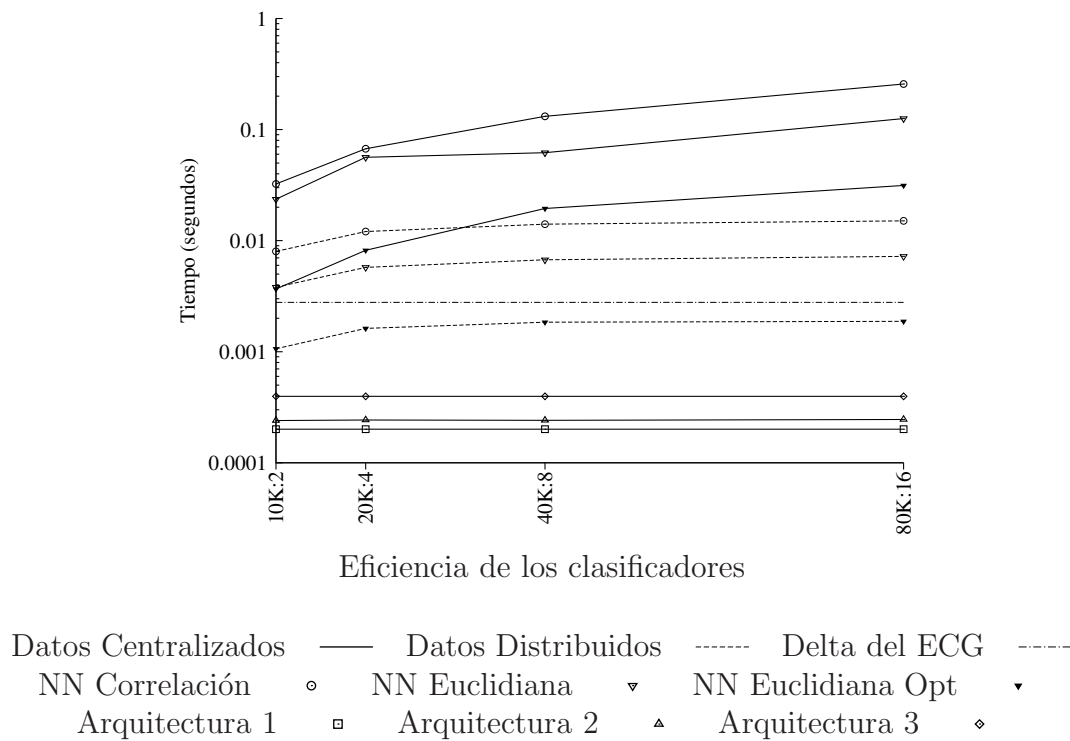


FIGURA 4.33. Desempeño Tiempo.

No obstante, el tiempo de cómputo de las alternativas 1-NN (a excepción de el algoritmo 1-NN implementando la distancia Euclidiana optimizada) es mayor que el tiempo transcurrido entre cada captura de valores por el equipo de electrocardiografía,

referido en la **Figura 4.33** como *Delta ECG*. El tiempo de cómputo requerido por las redes neuronales en la clasificación de latidos es menor que el requerido *Delta ECG*, como lo indica la **Figura 4.33**. De manera que, las redes neuronales, permiten realizar la clasificación de latidos en tiempo real, mientras el equipo de captura registra la actividad eléctrica del corazón. El cómputo de las redes neuronales consta realizar un número constante de multiplicaciones (determinadas por la arquitectura de la red).

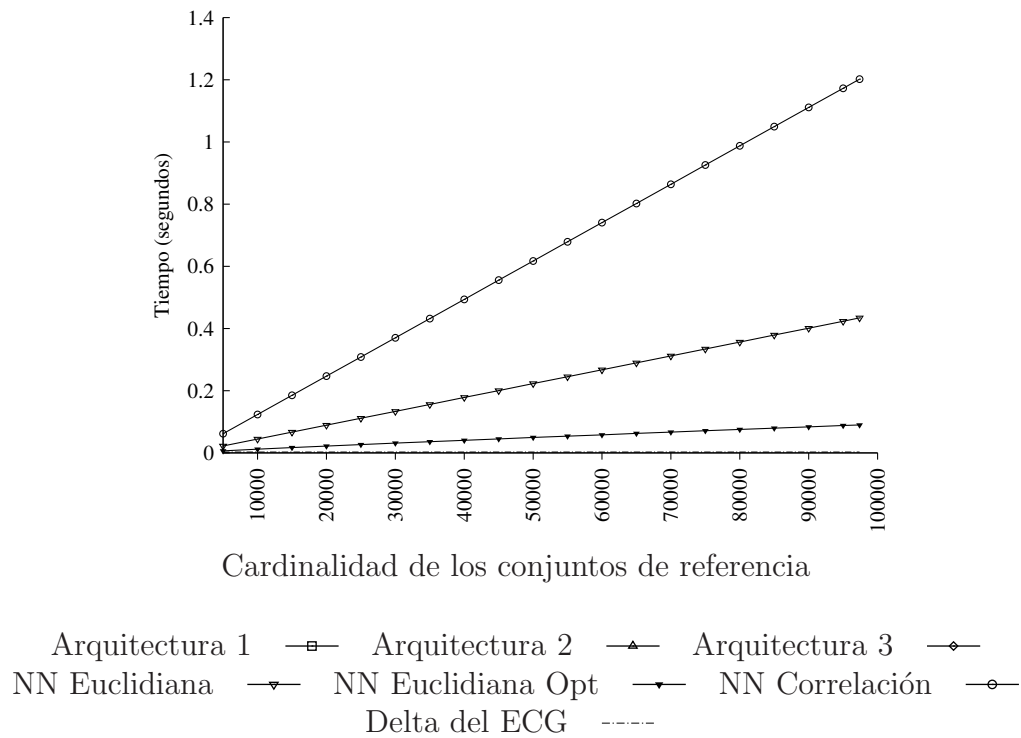


FIGURA 4.34. Eficacia de los clasificadores 1-NN

Por lo que el esquema de trabajo distribuido y el esquema centralizado reportan la misma eficiencia. Así mismo, la eficiencia obtenida en la plataforma móvil (por medio del esquema de trabajo centralizado) se observan en las **Figuras 4.34** y **4.35**. Como se indica en la **Figuras 4.34**, al incrementar la cardinalidad del conjunto de referencia, el tiempo de cómputo requerido por los algoritmos 1-NN incrementa. Además, el tiempo transcurrido durante la clasificación con 1-NN (utilizando cualquier conjunto de referencia y cualquier función de semejanza), supera el tiempo de cómputo de las

redes neuronales y el *Delta ECG* del equipo de electrocardiografía. Por lo que en la **Figuras 4.34** solamente se observa la eficacia de los algoritmos 1-NN. Por el contrario, el tiempo que toman las redes neuronales en identificar un latido permanece constante al incrementar la cardinalidad del conjunto de referencia, como se indica en la **Figuras 4.35**. Donde, el mayor tiempo corresponde a la Arquitectura 3 y el menor a la Arquitectura 1, ya que corresponden a la mayor y menor cantidad de nodos, respectivamente.

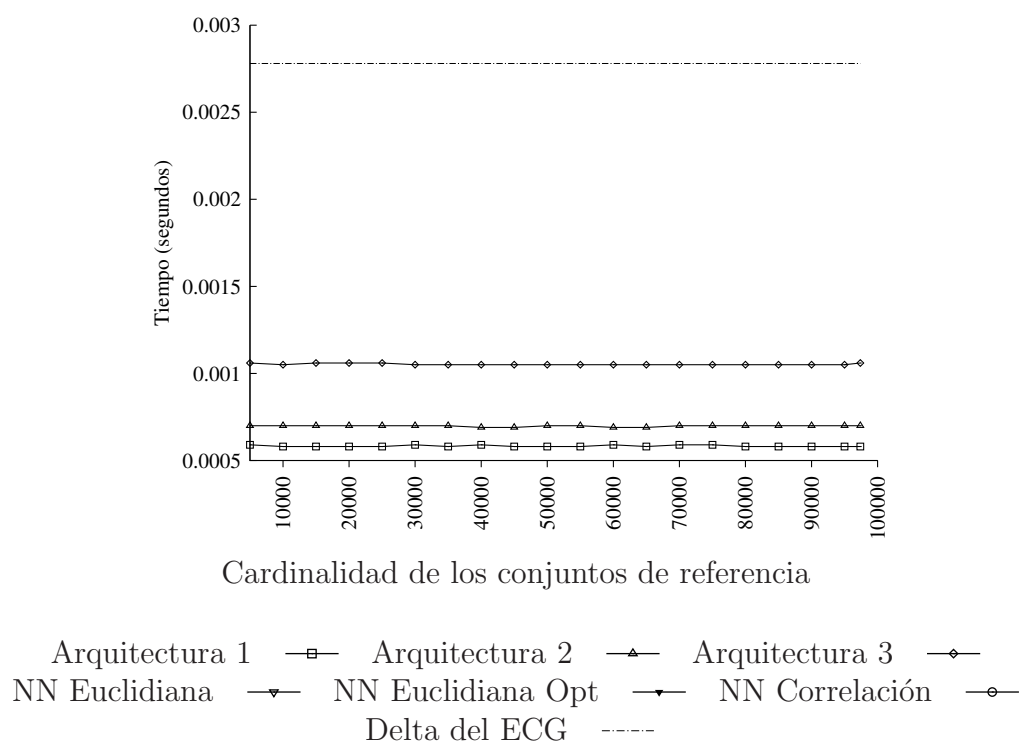


FIGURA 4.35. Desempeño Tiempo.

Los tiempos de ejecución de los algoritmos 1-NN, se indican en la **Tabla 4.13**. En esta tabla se indican los resultados obtenidos al utilizar los esquemas de trabajo centralizado y distribuido, en una plataforma estacionaria. La clasificación distribuida, dividió los conjuntos de referencia con cardinalidades de 10, 20, 40 y 80 mil latidos entre 2, 4, 8 y 16 unidades de procesamiento, respectivamente. De esta manera, cada

unidad de procesamiento realizó la clasificación de latidos consultando un conjunto de referencia de 5 mil latidos. Por lo que nuestros resultados indican que al clasificar latidos con el algoritmo 1-NN e implementando el esquema de trabajo distribuido, el tiempo de ejecución de los clasificadores disminuye, comparado con el esquema de trabajo centralizado. En el caso de la clasificación por 1-NN con el coeficiente de correlación (CCI) como función de semejanza, el tiempo de ejecución obtenido en el esquema centralizado, disminuyó utilizando el esquema distribuido 6.81, 8.67, 13.49 y 27.26 veces, con 2, 4, 8 y 16 nodos, respectivamente.

TABLA 4.13. Tiempo ejecución de los clasificadores 1-NN en los esquemas de trabajo centralizado y distribuido (segs).

Total de registros	Esquema de trabajo centralizado			Esquema de trabajo distribuido		
	CCI	Euc	Euc Opt	CCI	Euc	Euc Opt
5,000	0.02497	0.01471	0.00200			
10,000	0.05455	0.02728	0.00373	0.00801	0.00381	0.00106
15,000	0.08116	0.03121	0.00626			
20,000	0.10459	0.03242	0.00729	0.01207	0.00575	0.00162
25,000	0.08792	0.06155	0.00994			
30,000	0.14680	0.07246	0.01037			
35,000	0.16882	0.07230	0.01699			
40,000	0.18979	0.07799	0.02101	0.01407	0.00673	0.00184
45,000	0.16287	0.10574	0.02330			
50,000	0.17086	0.12742	0.02552			
55,000	0.17805	0.13238	0.02424			
60,000	0.19737	0.09655	0.02996			
65,000	0.27112	0.19007	0.03215			
70,000	0.32041	0.17657	0.02715			
75,000	0.40987	0.19434	0.03029			
80,000	0.41058	0.17250	0.02518	0.01506	0.00721	0.00188
85,000	0.37454	0.23193	0.02921			
90,000	0.32047	0.24997	0.02802			
95,000	0.38594	0.26466	0.03480			
97,388	0.46257	0.23327	0.04586			

Así mismo, la **Tabla 4.13**, indica el tiempo de ejecución del algoritmo 1-NN utilizando la distancia Euclidiana optimizada como función de semejanza e implementado en el esquema de trabajo centralizado, disminuyó al utilizar el esquema de trabajo distribuido 3.52, 4.50, 11.42 y 13.39 veces con 2, 4, 8 y 16 nodos, respectivamente. La **Tabla 4.14**, indica el tiempo de ejecución (en segundos) de las redes neuronales artificiales, el cual permanece constante. Esto debido a que a pesar de que la cardinalidad del conjunto de referencia aumente, la carga computacional (para una arquitectura en particular) en la clasificación de un latido es constante. De igual manera, el tiempo de ejecución de estos algoritmos al utilizar el esquema distribuido no cambia, ya que la ventaja de este esquema es la división del conjunto de referencia entre los nodos. Al comparar el menor tiempo de ejecución de la red neuronal más eficiente (Arquitectura 1) contra el algoritmo 1-NN más eficiente (obtenido al utilizar la distancia Euclidiana optimizada como función de semejanza y el esquema de datos distribuido), observamos que el tiempo de ejecución de la red neuronal es 5.3, 8.1, 9.2 y 9.4 veces menor, que al utilizar 2, 4, 8 y 16 nodos, respectivamente.

TABLA 4.14. Tiempo ejecución de las clasificación por redes neuronales.

Arquitecturas	Tiempo de ejecución
Arq. 1	0.00020
Arq. 2	0.00024
Arq. 3	0.00040

En la **Tabla 4.15**, indicamos el espacio de almacenamiento utilizado al grabar los clasificadores en una unidad de almacenamiento (*e.g.* disco duro, memoria del teléfono celular), a partir de un esquema de trabajo centralizado. El espacio reportado por las alternativas 1-NN se refiere al espacio utilizado para almacenar el conjunto de referencia, debido a que representa el conocimiento del algoritmo. El resultado reportado para las redes neuronales artificiales es el espacio de ocupado al que requiere alma-

cenar el clasificador en forma de archivo binario, una vez que este ha sido entrenado. Se observa que el espacio de almacenamiento para las alternativas 1-NN es mayor a medida que la cardinalidad del conjunto de referencia incrementa. A diferencia de las redes neuronales, donde el espacio de almacenamiento no está en función de la cardinalidad del conjunto de referencia, en su lugar el espacio incrementa al aumentar el número de nodos en la capa oculta, así como al incrementar la cantidad de clases a identificar.

TABLA 4.15. Cantidad de bytes requeridos por los clasificadores.

Total de registros	Bytes utilizados						
	1-NN	Arq-1		Arq-2		Arq-3	
		2	14	2	14	2	14
5000	1805016	56440	63736	67720	76456	101936	115040
10000	3610016	56441	63737	67721	76457	101937	115041
15000	5415016	56442	63738	67722	76458	101938	115042
20000	7220016	56443	63739	67723	76459	101939	115043
25000	9025016	56444	63740	67724	76460	101940	115044
30000	10830016	56445	63741	67725	76461	101941	115045
35000	12635016	56446	63742	67726	76462	101942	115046
40000	14440016	56447	63743	67727	76463	101943	115047
45000	16245016	56448	63744	67728	76464	101944	115048
50000	18050016	56449	63745	67729	76465	101945	115049
55000	19855016	56450	63746	67730	76466	101946	115050
60000	21660016	56451	63747	67731	76467	101947	115051
65000	23465016	56452	63748	67732	76468	101948	115052
70000	25270016	56453	63749	67733	76469	101949	115053
75000	27075016	56454	63750	67734	76470	101950	115054
80000	28880016	56455	63751	67735	76471	101951	115055
85000	30685016	56456	63752	67736	76472	101952	115056
90000	32490016	56457	63753	67737	76473	101953	115057
95000	34295016	56458	63754	67738	76474	101954	115058
97388	35157084	56459	63755	67739	76475	101955	115059

CAPÍTULO 5

CONCLUSIONES

La experimentación realizada en este trabajo de investigación, aporta evidencia del desempeño de realizar clasificación de latidos con los algoritmos de 1-NN y Redes Neuronales. El desempeño de estos clasificadores es observado en las plataformas de cómputo móvil y cómputo de escritorio. En esta última, se evaluaron los esquemas de procesamiento centralizado y distribuido. Se observaron diferencias en la eficacia de los algoritmos, asociadas a la cantidad y distribución de datos utilizados, así como a las características de diseño de los clasificadores. Aunado a la eficacia de los algoritmos, nuestros resultados indican la eficiencia de los clasificadores entre las plataformas computacionales. Situando a las redes neuronales artificiales como los algoritmos más eficientes temporal y espacialmente. Donde el tiempo de ejecución de la red neuronal con Arquitectura-1 es hasta 9.4 veces menor, con respecto al menor tiempo obtenido en la clasificación con el algoritmo 1-NN (utilizando la distancia Euclidiana como función de semejanza y el esquema de trabajo distribuido). Así mismo, la cantidad de bytes necesarios para almacenar la red neuronal con Arquitectura-1 es hasta 622.71 veces menor que el requerido para almacenar el conjunto de referencia con 97 mil latidos. El cómputo que una red neuronal realiza para clasificar un latido desconocido varía en función de la estructura del clasificador. Es decir, redes con la misma estructura ejecutan un número constante de instrucciones para clasificar un latido. De igual manera, el espacio ocupado por la red varía si la estructura de la red varía, ya que una ventaja de este clasificador es que solo requiere de un conjunto de datos de ejemplo durante la fase de entrenamiento. Sin embargo, los clasificadores más eficaces corresponden a las implementaciones de 1-NN. Con resultados cercanos a 100 % de eficacia, estos algoritmos superaron a las distintas implementaciones de

redes neuronales.

Nuestras conclusiones están divididas por las plataformas de cómputo utilizadas, por el tipo de almacenamiento implementado y por la cantidad de clases consideradas. Finalmente indicamos nuevas preguntas resultado de esta investigación, las cuales surgieron durante el proyecto y dan pie a continuar con el trabajo utilizando otras herramientas.

A pesar de que los mejores resultados de clasificación en el esquema de almacenamiento centralizado corresponden a 1-*NN*, las redes neuronales desarrolladas demostraron ser una alternativa eficiente para los datos de MIT-BIH Arritmia. Esto debido al poco tiempo que requieren las redes para clasificar un latido, demostrando un menor tiempo, que el intervalo de tiempo entre cada valor capturado en el ECG (*i.e.* frecuencia de muestreo) [48]. De manera que la clasificación de latidos, utilizando la red neuronal puede aplicarse en aplicaciones de detección en línea. A diferencia del algoritmo 1-*NN*, que supera el intervalo entre los valores del ECG. Así mismo, los resultados obtenidos en la clasificación con redes neuronales se relacionan con el entrenamiento de 500 épocas de los clasificadores, donde el error no se estabiliza a través de las épocas. Por medio de experimentación con redes neuronales de Arquitectura 2 y conjuntos de referencia con cardinalidades de 5, 50 y 97 mil latidos (este último corresponde al conjunto completo), observamos el error del clasificador desciende al incrementar el número de épocas a 20,000, llegando a estabilizarse. No obstante, el tiempo de cómputo requerido durante el entrenamiento de las redes fue de 9, 99 y 193 horas para los conjuntos de referencia con 5, 55 y 97 mil latidos, respectivamente. Debido al tiempo que tomaba realizar el entrenamiento de la Arquitectura 2, con los conjuntos de referencia mencionados, consideramos que realizar el entrenamiento de todas las redes neuronales definidas, para todos los conjuntos establecidos aplazaría el desarrollo del presente trabajo. Este trabajo recomienda el uso de redes neuronales en aplicaciones con una alta prioridad de tiempo de respuesta, como es la plataforma de cómputo móvil.

El trabajo consideró dos plataformas computacionales, nuestro punto de referencia y una plataforma de cómputo móvil. La plataforma de cómputo móvil, presentó los mismos resultados en eficiencia que nuestro punto de referencia. Esto indica que podemos llegar a tener la misma eficacia al pasar de una plataforma a otra. En el caso de la eficiencia, donde la plataforma móvil posee restricciones en la alimentación eléctrica, debemos considerar los clasificadores basados en instancias agotan drásticamente la carga de batería del dispositivo, debido a la complejidad que estos tienen. Por el contrario, los clasificadores basados en características como las redes neuronales, realizan un mismo número de instrucciones para llevar a cabo la clasificación de latidos, lo que reduce en menor proporción la carga de la batería en el dispositivo. Además, esta etapa de entrenamiento se puede realizar en un dispositivo sin restricciones de energía o poder de cómputo.

La distribución de los datos entre distintas unidades de procesamiento, demostró agilizar la clasificación frente a una alternativa centralizada. Debido a que la distribución de datos expone una cantidad menor de datos para cada nodo de procesamiento y con esto se reduce el tiempo de clasificación. La clasificación por 1- NN , aprovecha de que el conjunto de datos se reduce, considerando una menor cantidad de elementos en la búsqueda del elemento más semejante. Clasificar a través de redes neuronales en el esquema distribuido, consistió en distribuir los datos entre los nodos y después entrenar una red neuronal para cada nodo utilizado. En los casos de redes neuronales y 1- NN , los mejores resultados son generados al combinar el conocimiento entre cada nodo, mediante un proceso de votación. El aporte de distribuir los datos a los clasificadores es el proceso de votación, donde se observa una eficacia semejante entre los clasificadores que difiere del esquema centralizado y se concluye que la mayor eficacia de los clasificadores fue observada en el problema de clasificación dicotómica (distinguir entre latidos enfermos y sanos), debido a se engloban 13 clases distintas en una sola, reduciendo la probabilidad de error. En el caso de la clasificación de 14 clases, el algoritmo 1- NN utilizando el Coeficiente de Correlación como función de

semejanza, indica mayor eficacia, reconociendo la mayoría de las clases. La eficacia obtenida en la clasificación de anomalías (14 clases), se debe a que existen clases poco representadas en el conjunto de datos, como es el caso de la clase 7 y 10.

Como parte del desarrollo de aplicaciones médicas, deseamos extender este trabajo e incorporar herramientas de obtención de datos que nos permitan realizar una clasificación en línea con dispositivos móviles, obteniendo el ECG y clasificarlo en el mismo instante. Además de realizar un estudio más profundo redes neuronales, utilizando otras características en el clasificador y evaluar; su eficiencia y eficacia, frente a otros clasificadores de latidos. Para esto, deseamos incorporar datos de otras fuentes, incrementando la cantidad de ejemplos entre las clases a clasificar.

BIBLIOGRAFÍA

- [1] Adams E.R. y Choi A. Using Neural Networks to Predict Cardiac Arrhythmias. In *Proceedings of the IEEE-SMC International Conference*, pages 402–407, Danvers, MA, USA, Octubre 2012. IEEE.
- [2] Ananthanarayan S., Lapinski N., Siek K., Eisenberg M. Towards the crafting of personal health technologies. In *Proceedings of the 2014 Conference on Designing Interactive Systems*, pages 587–596, NY, USA, Junio 2014. ACM.
- [3] Arif M., Malagore I. y Afsar F. Detection and Localization of Myocardial Infarction using K-nearest Neighbor Classifier. *Journal of Medical Systems*, 36:279–289, Noviembre 2012.
- [4] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. A view of cloud computing. *Commun. ACM*, 53(4):50–58, April 2010.
- [5] Moustafa A Bani-Hasan, Yasser M Kadah, and Fatma M El-Hefnawi. Identification of cardiac arrhythmias using natural resonance complex frequencies. *International Journal of Biological and Life Sciences*, 6(3), 2010.
- [6] Bani-Hasan A.M, El-Hefnawi M.F. y Kadah M.Y. Model-based Parameter Estimation applied on Electrocardiogram Signal. *Journal of Computational Biology and Bioinformatics Research*, 3(2):25–28, Febrero 2011.
- [7] Baydogan M.G., Runger G. y Tuv E. A Bag-of-Features Framework to Classify Time Series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2796–2802, 2013.

- [8] Becerra-Luna, B., Dávila-García R., Salgado-Rodríguez P., Martínez-Memije R., Infante-Vázquez O. Monitor de señales de electrocardiografía y frecuencia cardiaca mediante un teléfono móvil con el protocolo de comunicación Bluetooth. *Archivos de cardiología de México*, 82:197 – 203, Septiembre 2012.
- [9] C.M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press Oxford, 1995.
- [10] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006.
- [11] G. Bortolan, I. Jekova, and I. Christov. Comparison of four methods for premature ventricular contraction and normal beat clustering. In *Computers in Cardiology, 2005*, pages 921 –924, sept. 2005.
- [12] Cabrera E. y Gaxiola A. *Teoría y Práctica de la Electrocardiografía*. La Prensa Médica Mexicana, 1966.
- [13] Cantrill S.V. Computers in Patient Care: The Promise and the Challenge. *Communications of the ACM*, 53(9):42–47, Septiembre 2010.
- [14] Chee J., Acharya U.R., Er K., Tan W. y Kuang-Chua C. Visualization of Cardiac Health using Vector Cardiogram. *IRBM*, 29(4):245 – 254, 2008.
- [15] Chen M.S., Han J. y Yu P.S. Data Mining: An Overview from a Database Perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):866 –883, Diciembre 1996.
- [16] Chin, F.J., Fang, Q., Zhang, T. y Cosic, I. A Fast Critical Arrhythmic ECG Waveform Identification Method Using Cross-Correlation and Multiple Template Matching. In *Proceedings of the IEEE International Conference on Engineering in Medicine and Biology Society*, pages 1922–1925, NJ, USA, Septiembre 2010. IEEE.

- [17] Cormen T.H., Leiserson C.E., Rivest R.L. y Stein C. *Introduction to Algorithms*. MIT Press, 2009.
- [18] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.*, 13(1):21–27, Septiembre 1967.
- [19] Darrington J. y Hool L. A New Methodology for Assessment of the Performance of Heartbeat Classification Systems. *BMC Medical Informatics and Decision Making*, 8(1):7, 2008.
- [20] M. de Hoon, S. Imoto, and S. Miyano. The C Clustering Library. *Institute of Medical Science, Human Genome Center, University of Tokyo*, 2003.
- [21] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. *Proceedings of the VLDB Endowment*, 1(2):1542–1552, 2008.
- [22] Ebadollahi S., Chang S., Syeda-Mahmood T., Coden A.R., Amir A. y Tanenblatt M.A. Concept-Based Electronic Health Records: Opportunities and Challenges. In *Proceedings of the 14th ACM-MM Conference*, pages 997–1006. ACM, 2006.
- [23] P. Esling and C. Agon. Time-Series Data Mining. *ACM Computing Surveys (CSUR)*, 45(1):12, 2012.
- [24] Faloutsos C., Ranganathan M. y Manolopoulos Y. Fast Subsequence Matching in Time-series Databases. *SIGMOD Rec.*, 23(2):419–429, Mayo 1994.
- [25] Goldberger A.L., Amaral L.A.N., Glass L., Hausdorff J.M., Ivanov P.C., Mark R.G., Mietus J.E., Moody G.B., Peng C.K. y Stanley H.E. PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, 101(23):e215–e220, Junio 2000.

- [26] Goldschlager N.F., Goldman M.J. y Testelli M. *Principios de electrocardiografía Clínica*. Manual Moderno, 1992.
- [27] Gradl, S., Kugler, P., Lohmuller, C., y Eskofier, B. Real-time ECG Monitoring and Arrhythmia Detection using Android-based Mobile Devices. In *Proceedings on the Annual International Conference of Engineering in Medicine and Biology Society*, pages 2452–2455, NJ, USA, Septiembre 2012. IEEE.
- [28] GSM Arena. LG Optimus Black (White version). http://www.gsmarena.com/lg_optimus_black_%28white_version%29-4037.php, Febrero 2013.
- [29] Hoffmann L. Implementing Electronic Medical Records. *Communications of the ACM*, 52(11):18–20, 2009.
- [30] Homaeinezhad M.R., Atyabi S.A., Tavakkoli E., Toosi H.N., Ghaffari A. y Ebrahimpour R. ECG Arrhythmia Recognition via a Neuro-SVM-KNN Hybrid Classifier with Virtual QRS Image-based Geometrical Features. *Expert Systems with Applications*, 39(2):2047–2058, Febrero 2012.
- [31] Hu Bing., Chen Y. y Keogh E. Time Series Classification Under more Realistic Assumptions. In *Proceedings of the thirteenth SIAM conference on data mining (SDM)*, 2013.
- [32] Hyejung K., Yazicioglu R.F., Merken P., Van Hoof C. y Hoi-Jun Y. ECG Signal Compression and Classification Algorithm With Quad Level Vector for ECG Holter System. *IEEE Transactions on Information Technology in Biomedicine*, 14(1):93–100, 2010.
- [33] Iliopoulos, C.S. y Michalakopoulos, S. Combinatorial ECG Analysis for Mobile Devices. In *Proceedings of the International Conference on Multimedia information Retrieval*, pages 409–412, NY, USA, Julio 2010. ACM.

- [34] Jadhav S.M., Nalbalwar S.L. y Ghatol A. Artificial Neural Network Based Cardiac Arrhythmia Classification using ECG Signal Data. In *Proceedings of the IEEE-ICEIE Conference*, volume 1, pages V1–228 –V1–231, Agosto 2010.
- [35] Jekova I., Bortolan G. y Christov I. Assessment and Comparison of Different Methods for Heartbeat Classification. *Medical Engineering & Physics*, 30(2):248–257, 2008.
- [36] Karimifard, S., Ahmadian, A., Khoshnevisan, M., y Nambakhsh, M.S. Morphological Heart Arrhythmia Detection using Hermitian Basis Functions and kNN Classifier. In *Proceedings of the Annual International Conference on Engineering in Medicine and Biology Society*, pages 1367–1370, NJ, USA, Septiembre 2006. IEEE.
- [37] E. Keogh and S. Kasetty. On the need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. *Data Mining and Knowledge Discovery*, 7(4):349–371, 2003.
- [38] Keogh E. y Pazzani M. Scaling up Dynamic Time Warping for Datamining Applications. pages 285–289, Boston, MA, August 2000.
- [39] Kleinberg J. y Tardos E. *Algorithm Design*. Alternative Etext Formats. Addison-Wesley, 2006.
- [40] Vessela Krasteva and Irena Jekova. Qrs template matching for recognition of ventricular ectopic beats. *Annals of biomedical engineering*, 35(12):2065–2076, 2007.
- [41] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

- [42] Lara-Dévora S.L., Vega-López I.F. y López-Barrón D.E. Una Evaluación Experimental de Técnicas de Caracterización de Series de Tiempo Aplicadas a Electrocardiogramas. In *Memorias del VIII Congreso Internacional sobre Innovación y Desarrollo Tecnológico*, Noviembre 2010.
- [43] D.T. Larose. *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley-Interscience, 2005.
- [44] Thorsten Last, Chris Nugent, and Frank Owens. Multi-component based cross correlation beat detection in electrocardiogram analysis. *BioMedical Engineering OnLine*, 3(1):26, 2004.
- [45] R.P. Lippmann. An introduction to computing with neural nets. *ASSP Magazine, IEEE*, 4(2):4–22, Abril 1987.
- [46] D.E. Lopez-Barron, I.F. Vega-Lopez, and O. Cuen-Tellez. Supporting Query By Content on ECG Data with User Defined Functions. In *Proceedings of the IASTED International Conference Advances in Computer Science*, pages 392–399. ACTA-Press, April 2013.
- [47] Sipser M. *Introduction To The Theory Of Computation*. Computer Science Series. Course Technology Ptr, 2006.
- [48] G.B. Moody and R.G. Mark. The impact of the MIT-BIH Arrhythmia Database. *Engineering in Medicine and Biology Magazine, IEEE*, 20(3):45 –50, may-june 2001.
- [49] Nasehi S. y Pourghassem H. Seizure Detection Algorithms Based on Analysis of EEG and ECG Signals: a Survey. *Neurophysiology*, 44:174–186, 2012.
- [50] Organización Mundial de la Salud. Enfermedades Cardiovasculares. http://www.who.int/cardiovascular_diseases/es/index.html, Febrero 2013.

- [51] Ozbay Y. y Karlik B. A Recognition of ECG Arrhythmias using Artificial Neural Networks. In *Proceedings of the 23rd IEEE-EMBS Conference*, volume 2, pages 1680 – 1683 vol.2, Octubre 2001.
- [52] PhysioNet. PhysioBank Annotations. <http://www.physionet.org/physiobank/annotations.shtml>, Diciembre 2013. Visitado 21/12/2013.
- [53] PhysioNet. The WFDB Software Package. <http://www.physionet.org/physiotools/wfdb.shtml#applications>, Diciembre 2013. Visitado 21/12/2013.
- [54] Piatetsky-Shapiro G. y Fayyad U. An Introduction to SIGKDD and a Reflection on the term 'Data Mining'. *SIGKDD Explor. Newsl.*, 13(2):102–103, Mayo 2012.
- [55] Pérez-Nava M. y Vezsco-Cruz Ma. *Electrocardiografía Básica*. Instituto Nacional de Cardiología Ignacio Chávez, Dirección de Enfermería, Departamento de Enseñanza, 2005.
- [56] Saini I., Singh D. y Khosla A. Classification of RR-Interval and Blood Pressure for Different Postures using KNN Algorithm. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, pages 13–20, 2012.
- [57] Sánchez-Torres G. *Tecnología en Desarrollo. Puntal en la Evolución de la Cardiología 1986–2009*. Prado, 2009.
- [58] Luis M. Vaquero, Luis Rodero-Merino, Juan Caceres, and Maik Lindner. A break in the clouds: towards a cloud definition. *SIGCOMM Comput. Commun. Rev.*, 39(1):50–55, December 2008.
- [59] Vega-Lopez, I.F. e Infante-Vazquez, O. Diseño Conceptual de una Base de Datos para el Almacenamiento Estructurado de Electrocardiogramas Digitales en el INCICH. In *Memorias del Congreso Nacional de Ingeniería Biomédica*, León, Guanajuato, México, Octubre 2010. SOMIB.

- [60] JINFEI Xie and WY Yan. Pattern-based characterization of time series. *International Journal of Information and Systems Science*, 3(3):479–491, 2007.
- [61] Yi B.K. y Faloutsos C. Fast Time Sequence Indexing for Arbitrary Lp Norms. In *Proceedings of the 26th International Conference on Very Large Data Bases*, VLDB '00, pages 385–394, San Francisco, CA, USA, Septiembre 2000. Morgan Kaufmann Publishers Inc.

Anexos

ANEXO III: RESULTADO DE LA CLASIFICACIÓN DICOTÓMICA DISTRIBUIDA

TABLA 1. Porcentaje de eficacia para latidos normales en conjuntos de referencia aleatorios en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.73	98.88	96.57	94.38	98.36	102.00	98.88	96.57	94.38	98.36
10,000	99.00	98.80	98.77	99.23	98.93	1913.00	99.70	99.32	99.45	99.41
20,000	98.84	98.80	98.91	98.23	98.30	5568.00	99.52	98.70	98.16	98.98
40,000	98.93	99.04	98.25	97.36	97.48	8.00	99.64	98.88	98.84	98.68
80,000	99.13	99.07	98.54	97.52	98.04	80000.00	99.71	99.14	98.64	98.95

TABLA 2. Porcentaje de eficacia para latidos anormales en conjuntos de referencia aleatorios en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	95.06	91.57	89.58	92.30	88.71	98.73	91.57	89.58	92.30	88.71
10,000	95.25	93.07	88.52	87.45	85.17	23.00	89.58	84.06	81.64	77.23
20,000	96.08	93.80	88.52	89.78	90.07	1927.00	90.84	89.58	89.53	88.61
40,000	96.75	95.06	91.04	91.81	91.38	5569.00	91.76	89.97	90.31	87.84
80,000	97.04	95.74	90.02	91.18	91.09	16.00	91.86	88.81	89.97	89.63

TABLA 3. Porcentaje de eficacia para latidos normales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.73	98.88	96.57	94.38	98.36	102.00	98.88	96.57	94.38	98.36
10,000	99.00	98.80	98.77	99.23	98.93	1913.00	99.70	99.32	99.45	99.41
20,000	98.84	98.80	98.91	98.23	98.30	5568.00	99.52	98.70	98.16	98.98
40,000	98.93	99.04	98.25	97.36	97.48	8.00	99.64	98.88	98.84	98.68
80,000	99.13	99.07	98.54	97.52	98.04	80000.00	99.71	99.14	98.64	98.95

TABLA 4. Porcentaje de eficacia para latidos anormales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	95.06	91.57	89.58	92.30	88.71	98.73	91.57	89.58	92.30	88.71
10,000	95.25	93.07	88.52	87.45	85.17	23.00	89.58	84.06	81.64	77.23
20,000	96.08	93.80	88.52	89.78	90.07	1927.00	90.84	89.58	89.53	88.61
40,000	96.75	95.06	91.04	91.81	91.38	5569.00	91.76	89.97	90.31	87.84
80,000	97.04	95.74	90.02	91.18	91.09	16.00	91.86	88.81	89.97	89.63

ANEXO IV: RESULTADO DE LA CLASIFICACIÓN DE ANOMALÍAS DISTRIBUIDA

TABLA 5. Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.57	98.63	99.05	95.47	96.25	5.00	98.63	99.05	95.47	96.25
10,000	98.66	98.55	98.86	98.32	99.48	0.00	99.75	99.54	99.66	99.79
20,000	98.93	98.91	98.57	99.16	99.32	0.00	99.64	99.57	99.71	99.68
40,000	99.05	98.97	98.55	98.70	98.86	17.00	99.63	99.66	99.64	99.57
80,000	99.14	99.21	99.05	99.13	98.89	1.00	99.68	99.70	99.61	99.57

TABLA 6. Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.51	98.35	95.21	88.60	90.08	98.57	98.35	95.21	88.60	90.08
10,000	99.17	98.51	95.87	96.53	94.21	15.00	97.02	94.21	95.21	93.22
20,000	99.34	98.68	97.52	95.37	95.54	0.00	98.02	94.71	94.88	95.37
40,000	99.17	99.34	98.35	98.02	97.85	0.00	98.02	97.02	96.53	97.02
80,000	99.50	99.34	95.21	96.53	96.36	15.00	97.69	97.02	96.86	96.53

TABLA 7. Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	96.32	92.10	84.19	91.91	90.26	98.51	92.10	84.19	91.91	90.26
10,000	96.88	94.12	91.18	93.01	93.01	99.46	95.04	78.86	93.20	93.20
20,000	96.88	96.14	94.12	95.22	92.83	2.00	95.40	91.54	95.04	93.20
40,000	97.98	96.32	96.32	92.65	92.46	0.00	95.96	94.12	94.30	93.93
80,000	96.88	96.69	84.19	94.30	94.30	0.00	95.77	93.75	93.75	93.93

TABLA 8. Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	65.97	52.36	40.84	42.41	37.70	96.32	52.36	40.84	42.41	37.70
10,000	66.49	58.64	54.45	54.97	52.36	98.84	41.36	29.32	49.21	49.21
20,000	62.83	62.83	48.69	53.40	52.88	99.45	47.12	50.79	49.21	49.21
40,000	68.06	65.97	46.07	51.31	51.31	2.00	54.45	52.88	53.93	51.31
80,000	70.16	67.54	40.84	53.40	56.02	0.00	58.12	53.40	54.45	52.88

TABLA 9. Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	30.00	20.00	10.00	0.00	0.00	65.97	20.00	10.00	0.00	0.00
10,000	60.00	50.00	10.00	0.00	0.00	97.43	20.00	10.00	0.00	0.00
20,000	70.00	50.00	10.00	20.00	0.00	99.01	30.00	0.00	0.00	0.00
40,000	70.00	70.00	20.00	20.00	0.00	99.36	20.00	0.00	0.00	0.00
80,000	70.00	70.00	10.00	0.00	0.00	3.00	40.00	0.00	0.00	0.00

TABLA 10. Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	16.67	0.00	0.00	0.00	30.00	16.67	0.00	0.00	0.00
10,000	33.33	33.33	0.00	0.00	0.00	51.31	0.00	0.00	0.00	0.00
20,000	66.67	33.33	0.00	0.00	0.00	96.88	0.00	0.00	0.00	0.00
40,000	83.33	33.33	0.00	0.00	0.00	98.51	0.00	0.00	0.00	0.00
80,000	83.33	50.00	0.00	0.00	0.00	99.43	0.00	0.00	0.00	0.00

TABLA 11. Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10,000	100.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00
20,000	0.00	0.00	0.00	0.00	0.00	57.07	0.00	0.00	0.00	0.00
40,000	0.00	0.00	0.00	0.00	0.00	97.98	0.00	0.00	0.00	0.00
80,000	0.00	0.00	0.00	0.00	0.00	98.84	0.00	0.00	0.00	0.00

TABLA 12. Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	90.81	87.43	84.43	75.23	76.17	0.00	87.43	84.43	75.23	76.17
10,000	94.56	90.43	89.12	83.68	87.43	16.67	85.37	80.49	78.42	83.49
20,000	96.06	90.24	84.80	86.49	89.12	60.00	86.49	88.37	86.30	87.80
40,000	96.62	93.81	88.56	89.12	90.62	64.40	89.31	89.87	89.68	90.06
80,000	96.81	93.81	84.43	90.06	90.06	97.43	88.18	90.99	90.99	90.43

TABLA 13. Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	65.00	60.00	58.33	5.00	5.00	90.81	60.00	58.33	5.00	5.00
10,000	68.33	63.33	50.00	45.00	35.00	0.00	41.67	41.67	36.67	25.00
20,000	80.00	71.67	45.00	53.33	10.00	0.00	60.00	18.33	23.33	1.67
40,000	80.00	71.67	56.67	56.67	35.00	60.00	58.33	20.00	33.33	28.33
80,000	85.00	78.33	58.33	50.00	43.33	65.45	61.67	25.00	21.67	31.67

TABLA 14. Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	0.00	0.00	0.00	0.00	65.00	0.00	0.00	0.00	0.00
10,000	0.00	0.00	0.00	0.00	0.00	90.24	0.00	0.00	0.00	0.00
20,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40,000	0.00	0.00	0.00	0.00	0.00	16.67	0.00	0.00	0.00	0.00
80,000	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00

TABLA 15. Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	35.29	41.18	0.00	0.00	0.00	0.00	41.18	0.00	0.00	0.00
10,000	35.29	41.18	11.76	5.88	0.00	53.33	5.88	0.00	5.88	0.00
20,000	35.29	47.06	0.00	11.76	0.00	92.31	17.65	5.88	11.76	0.00
40,000	23.53	35.29	11.76	0.00	5.88	0.00	11.76	5.88	0.00	0.00
80,000	41.18	47.06	0.00	5.88	0.00	16.67	17.65	5.88	0.00	0.00

TABLA 16. Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	100.00	100.00	50.00	0.00	0.00	35.29	100.00	50.00	0.00	0.00
10,000	100.00	66.67	66.67	33.33	0.00	0.00	66.67	33.33	16.67	0.00
20,000	100.00	100.00	33.33	50.00	0.00	70.00	83.33	66.67	0.00	0.00
40,000	100.00	100.00	0.00	16.67	0.00	94.37	83.33	33.33	16.67	0.00
80,000	100.00	100.00	50.00	0.00	0.00	0.00	100.00	50.00	16.67	0.00

TABLA 17. Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	100.00	100.00	98.57	97.14	95.71	100.00	100.00	98.57	97.14	95.71
10,000	100.00	100.00	97.14	100.00	98.57	5.88	100.00	98.57	100.00	98.57
20,000	100.00	100.00	97.14	98.57	97.14	0.00	100.00	95.71	97.14	95.71
40,000	100.00	100.00	97.14	97.14	98.57	63.33	100.00	97.14	98.57	98.57
80,000	100.00	100.00	98.57	97.14	97.14	95.31	100.00	98.57	98.57	98.57

TABLA 18. Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	75.00	50.00	0.00	40.00	0.00	100.00	50.00	0.00	40.00	0.00
10,000	60.00	50.00	25.00	10.00	0.00	83.33	10.00	5.00	10.00	0.00
20,000	100.00	75.00	10.00	35.00	10.00	5.88	60.00	30.00	25.00	10.00
40,000	100.00	80.00	10.00	30.00	25.00	0.00	55.00	25.00	35.00	5.00
80,000	100.00	95.00	0.00	5.00	5.00	68.33	55.00	35.00	30.00	10.00

TABLA 19. Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.57	98.63	99.05	95.47	96.25	5.00	98.63	99.05	95.47	96.25
10,000	98.66	98.55	98.86	98.32	99.48	0.00	99.75	99.54	99.66	99.79
20,000	98.93	98.91	98.57	99.16	99.32	0.00	99.64	99.57	99.71	99.68
40,000	99.05	98.97	98.55	98.70	98.86	17.00	99.63	99.66	99.64	99.57
80,000	99.14	99.21	99.05	99.13	98.89	1.00	99.68	99.70	99.61	99.57

TABLA 20. Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.51	98.35	95.21	88.60	90.08	98.57	98.35	95.21	88.60	90.08
10,000	99.17	98.51	95.87	96.53	94.21	15.00	97.02	94.21	95.21	93.22
20,000	99.34	98.68	97.52	95.37	95.54	0.00	98.02	94.71	94.88	95.37
40,000	99.17	99.34	98.35	98.02	97.85	0.00	98.02	97.02	96.53	97.02
80,000	99.50	99.34	95.21	96.53	96.36	15.00	97.69	97.02	96.86	96.53

TABLA 21. Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	96.32	92.10	84.19	91.91	90.26	98.51	92.10	84.19	91.91	90.26
10, 000	96.88	94.12	91.18	93.01	93.01	99.46	95.04	78.86	93.20	93.20
20, 000	96.88	96.14	94.12	95.22	92.83	2.00	95.40	91.54	95.04	93.20
40, 000	97.98	96.32	96.32	92.65	92.46	0.00	95.96	94.12	94.30	93.93
80, 000	96.88	96.69	84.19	94.30	94.30	0.00	95.77	93.75	93.75	93.93

TABLA 22. Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	65.97	52.36	40.84	42.41	37.70	96.32	52.36	40.84	42.41	37.70
10, 000	66.49	58.64	54.45	54.97	52.36	98.84	41.36	29.32	49.21	49.21
20, 000	62.83	62.83	48.69	53.40	52.88	99.45	47.12	50.79	49.21	49.21
40, 000	68.06	65.97	46.07	51.31	51.31	2.00	54.45	52.88	53.93	51.31
80, 000	70.16	67.54	40.84	53.40	56.02	0.00	58.12	53.40	54.45	52.88

TABLA 23. Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	30.00	20.00	10.00	0.00	0.00	65.97	20.00	10.00	0.00	0.00
10, 000	60.00	50.00	10.00	0.00	0.00	97.43	20.00	10.00	0.00	0.00
20, 000	70.00	50.00	10.00	20.00	0.00	99.01	30.00	0.00	0.00	0.00
40, 000	70.00	70.00	20.00	20.00	0.00	99.36	20.00	0.00	0.00	0.00
80, 000	70.00	70.00	10.00	0.00	0.00	3.00	40.00	0.00	0.00	0.00

TABLA 24. Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	0.00	16.67	0.00	0.00	0.00	30.00	16.67	0.00	0.00	0.00
10, 000	33.33	33.33	0.00	0.00	0.00	51.31	0.00	0.00	0.00	0.00
20, 000	66.67	33.33	0.00	0.00	0.00	96.88	0.00	0.00	0.00	0.00
40, 000	83.33	33.33	0.00	0.00	0.00	98.51	0.00	0.00	0.00	0.00
80, 000	83.33	50.00	0.00	0.00	0.00	99.43	0.00	0.00	0.00	0.00

TABLA 25. Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10, 000	100.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00
20, 000	0.00	0.00	0.00	0.00	0.00	57.07	0.00	0.00	0.00	0.00
40, 000	0.00	0.00	0.00	0.00	0.00	97.98	0.00	0.00	0.00	0.00
80, 000	0.00	0.00	0.00	0.00	0.00	98.84	0.00	0.00	0.00	0.00

TABLA 26. Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	90.81	87.43	84.43	75.23	76.17	0.00	87.43	84.43	75.23	76.17
10, 000	94.56	90.43	89.12	83.68	87.43	16.67	85.37	80.49	78.42	83.49
20, 000	96.06	90.24	84.80	86.49	89.12	60.00	86.49	88.37	86.30	87.80
40, 000	96.62	93.81	88.56	89.12	90.62	64.40	89.31	89.87	89.68	90.06
80, 000	96.81	93.81	84.43	90.06	90.06	97.43	88.18	90.99	90.99	90.43

TABLA 27. Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	65.00	60.00	58.33	5.00	5.00	90.81	60.00	58.33	5.00	5.00
10,000	68.33	63.33	50.00	45.00	35.00	0.00	41.67	41.67	36.67	25.00
20,000	80.00	71.67	45.00	53.33	10.00	0.00	60.00	18.33	23.33	1.67
40,000	80.00	71.67	56.67	56.67	35.00	60.00	58.33	20.00	33.33	28.33
80,000	85.00	78.33	58.33	50.00	43.33	65.45	61.67	25.00	21.67	31.67

TABLA 28. Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	0.00	0.00	0.00	0.00	65.00	0.00	0.00	0.00	0.00
10,000	0.00	0.00	0.00	0.00	0.00	90.24	0.00	0.00	0.00	0.00
20,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40,000	0.00	0.00	0.00	0.00	0.00	16.67	0.00	0.00	0.00	0.00
80,000	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00

TABLA 29. Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	35.29	41.18	0.00	0.00	0.00	0.00	41.18	0.00	0.00	0.00
10,000	35.29	41.18	11.76	5.88	0.00	53.33	5.88	0.00	5.88	0.00
20,000	35.29	47.06	0.00	11.76	0.00	92.31	17.65	5.88	11.76	0.00
40,000	23.53	35.29	11.76	0.00	5.88	0.00	11.76	5.88	0.00	0.00
80,000	41.18	47.06	0.00	5.88	0.00	16.67	17.65	5.88	0.00	0.00

TABLA 30. Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	100.00	100.00	50.00	0.00	0.00	35.29	100.00	50.00	0.00	0.00
10, 000	100.00	66.67	66.67	33.33	0.00	0.00	66.67	33.33	16.67	0.00
20, 000	100.00	100.00	33.33	50.00	0.00	70.00	83.33	66.67	0.00	0.00
40, 000	100.00	100.00	0.00	16.67	0.00	94.37	83.33	33.33	16.67	0.00
80, 000	100.00	100.00	50.00	0.00	0.00	0.00	100.00	50.00	16.67	0.00

TABLA 31. Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	100.00	100.00	98.57	97.14	95.71	100.00	100.00	98.57	97.14	95.71
10, 000	100.00	100.00	97.14	100.00	98.57	5.88	100.00	98.57	100.00	98.57
20, 000	100.00	100.00	97.14	98.57	97.14	0.00	100.00	95.71	97.14	95.71
40, 000	100.00	100.00	97.14	97.14	98.57	63.33	100.00	97.14	98.57	98.57
80, 000	100.00	100.00	98.57	97.14	97.14	95.31	100.00	98.57	98.57	98.57

TABLA 32. Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	75.00	50.00	0.00	40.00	0.00	100.00	50.00	0.00	40.00	0.00
10, 000	60.00	50.00	25.00	10.00	0.00	83.33	10.00	5.00	10.00	0.00
20, 000	100.00	75.00	10.00	35.00	10.00	5.88	60.00	30.00	25.00	10.00
40, 000	100.00	80.00	10.00	30.00	25.00	0.00	55.00	25.00	35.00	5.00
80, 000	100.00	95.00	0.00	5.00	5.00	68.33	55.00	35.00	30.00	10.00

ANEXO III: RESULTADO DE LA CLASIFICACIÓN DICOTÓMICA DISTRIBUIDA

TABLA 33. Porcentaje de eficacia para latidos normales en conjuntos de referencia aleatorios en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	98.73	98.88	96.57	94.38	98.36	102.00	98.88	96.57	94.38	98.36
10, 000	99.00	98.80	98.77	99.23	98.93	1913.00	99.70	99.32	99.45	99.41
20, 000	98.84	98.80	98.91	98.23	98.30	5568.00	99.52	98.70	98.16	98.98
40, 000	98.93	99.04	98.25	97.36	97.48	8.00	99.64	98.88	98.84	98.68
80, 000	99.13	99.07	98.54	97.52	98.04	80000.00	99.71	99.14	98.64	98.95

TABLA 34. Porcentaje de eficacia para latidos anormales en conjuntos de referencia aleatorios en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	95.06	91.57	89.58	92.30	88.71	98.73	91.57	89.58	92.30	88.71
10, 000	95.25	93.07	88.52	87.45	85.17	23.00	89.58	84.06	81.64	77.23
20, 000	96.08	93.80	88.52	89.78	90.07	1927.00	90.84	89.58	89.53	88.61
40, 000	96.75	95.06	91.04	91.81	91.38	5569.00	91.76	89.97	90.31	87.84
80, 000	97.04	95.74	90.02	91.18	91.09	16.00	91.86	88.81	89.97	89.63

TABLA 35. Porcentaje de eficacia para latidos normales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.73	98.88	96.57	94.38	98.36	102.00	98.88	96.57	94.38	98.36
10,000	99.00	98.80	98.77	99.23	98.93	1913.00	99.70	99.32	99.45	99.41
20,000	98.84	98.80	98.91	98.23	98.30	5568.00	99.52	98.70	98.16	98.98
40,000	98.93	99.04	98.25	97.36	97.48	8.00	99.64	98.88	98.84	98.68
80,000	99.13	99.07	98.54	97.52	98.04	80000.00	99.71	99.14	98.64	98.95

TABLA 36. Porcentaje de eficacia para latidos anormales en conjuntos de referencia equi-proporcionados en la clasificación dicotómica.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	95.06	91.57	89.58	92.30	88.71	98.73	91.57	89.58	92.30	88.71
10,000	95.25	93.07	88.52	87.45	85.17	23.00	89.58	84.06	81.64	77.23
20,000	96.08	93.80	88.52	89.78	90.07	1927.00	90.84	89.58	89.53	88.61
40,000	96.75	95.06	91.04	91.81	91.38	5569.00	91.76	89.97	90.31	87.84
80,000	97.04	95.74	90.02	91.18	91.09	16.00	91.86	88.81	89.97	89.63

ANEXO IV: RESULTADO DE LA CLASIFICACIÓN DE ANOMALÍAS DISTRIBUIDA

TABLA 37. Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.57	98.63	99.05	95.47	96.25	5.00	98.63	99.05	95.47	96.25
10,000	98.66	98.55	98.86	98.32	99.48	0.00	99.75	99.54	99.66	99.79
20,000	98.93	98.91	98.57	99.16	99.32	0.00	99.64	99.57	99.71	99.68
40,000	99.05	98.97	98.55	98.70	98.86	17.00	99.63	99.66	99.64	99.57
80,000	99.14	99.21	99.05	99.13	98.89	1.00	99.68	99.70	99.61	99.57

TABLA 38. Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.51	98.35	95.21	88.60	90.08	98.57	98.35	95.21	88.60	90.08
10,000	99.17	98.51	95.87	96.53	94.21	15.00	97.02	94.21	95.21	93.22
20,000	99.34	98.68	97.52	95.37	95.54	0.00	98.02	94.71	94.88	95.37
40,000	99.17	99.34	98.35	98.02	97.85	0.00	98.02	97.02	96.53	97.02
80,000	99.50	99.34	95.21	96.53	96.36	15.00	97.69	97.02	96.86	96.53

TABLA 39. Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	96.32	92.10	84.19	91.91	90.26	98.51	92.10	84.19	91.91	90.26
10,000	96.88	94.12	91.18	93.01	93.01	99.46	95.04	78.86	93.20	93.20
20,000	96.88	96.14	94.12	95.22	92.83	2.00	95.40	91.54	95.04	93.20
40,000	97.98	96.32	96.32	92.65	92.46	0.00	95.96	94.12	94.30	93.93
80,000	96.88	96.69	84.19	94.30	94.30	0.00	95.77	93.75	93.75	93.93

TABLA 40. Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	65.97	52.36	40.84	42.41	37.70	96.32	52.36	40.84	42.41	37.70
10,000	66.49	58.64	54.45	54.97	52.36	98.84	41.36	29.32	49.21	49.21
20,000	62.83	62.83	48.69	53.40	52.88	99.45	47.12	50.79	49.21	49.21
40,000	68.06	65.97	46.07	51.31	51.31	2.00	54.45	52.88	53.93	51.31
80,000	70.16	67.54	40.84	53.40	56.02	0.00	58.12	53.40	54.45	52.88

TABLA 41. Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	30.00	20.00	10.00	0.00	0.00	65.97	20.00	10.00	0.00	0.00
10,000	60.00	50.00	10.00	0.00	0.00	97.43	20.00	10.00	0.00	0.00
20,000	70.00	50.00	10.00	20.00	0.00	99.01	30.00	0.00	0.00	0.00
40,000	70.00	70.00	20.00	20.00	0.00	99.36	20.00	0.00	0.00	0.00
80,000	70.00	70.00	10.00	0.00	0.00	3.00	40.00	0.00	0.00	0.00

TABLA 42. Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	16.67	0.00	0.00	0.00	30.00	16.67	0.00	0.00	0.00
10,000	33.33	33.33	0.00	0.00	0.00	51.31	0.00	0.00	0.00	0.00
20,000	66.67	33.33	0.00	0.00	0.00	96.88	0.00	0.00	0.00	0.00
40,000	83.33	33.33	0.00	0.00	0.00	98.51	0.00	0.00	0.00	0.00
80,000	83.33	50.00	0.00	0.00	0.00	99.43	0.00	0.00	0.00	0.00

TABLA 43. Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10,000	100.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00
20,000	0.00	0.00	0.00	0.00	0.00	57.07	0.00	0.00	0.00	0.00
40,000	0.00	0.00	0.00	0.00	0.00	97.98	0.00	0.00	0.00	0.00
80,000	0.00	0.00	0.00	0.00	0.00	98.84	0.00	0.00	0.00	0.00

TABLA 44. Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	90.81	87.43	84.43	75.23	76.17	0.00	87.43	84.43	75.23	76.17
10,000	94.56	90.43	89.12	83.68	87.43	16.67	85.37	80.49	78.42	83.49
20,000	96.06	90.24	84.80	86.49	89.12	60.00	86.49	88.37	86.30	87.80
40,000	96.62	93.81	88.56	89.12	90.62	64.40	89.31	89.87	89.68	90.06
80,000	96.81	93.81	84.43	90.06	90.06	97.43	88.18	90.99	90.99	90.43

TABLA 45. Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	65.00	60.00	58.33	5.00	5.00	90.81	60.00	58.33	5.00	5.00
10,000	68.33	63.33	50.00	45.00	35.00	0.00	41.67	41.67	36.67	25.00
20,000	80.00	71.67	45.00	53.33	10.00	0.00	60.00	18.33	23.33	1.67
40,000	80.00	71.67	56.67	56.67	35.00	60.00	58.33	20.00	33.33	28.33
80,000	85.00	78.33	58.33	50.00	43.33	65.45	61.67	25.00	21.67	31.67

TABLA 46. Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	0.00	0.00	0.00	0.00	65.00	0.00	0.00	0.00	0.00
10,000	0.00	0.00	0.00	0.00	0.00	90.24	0.00	0.00	0.00	0.00
20,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40,000	0.00	0.00	0.00	0.00	0.00	16.67	0.00	0.00	0.00	0.00
80,000	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00

TABLA 47. Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	35.29	41.18	0.00	0.00	0.00	0.00	41.18	0.00	0.00	0.00
10,000	35.29	41.18	11.76	5.88	0.00	53.33	5.88	0.00	5.88	0.00
20,000	35.29	47.06	0.00	11.76	0.00	92.31	17.65	5.88	11.76	0.00
40,000	23.53	35.29	11.76	0.00	5.88	0.00	11.76	5.88	0.00	0.00
80,000	41.18	47.06	0.00	5.88	0.00	16.67	17.65	5.88	0.00	0.00

TABLA 48. Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	100.00	100.00	50.00	0.00	0.00	35.29	100.00	50.00	0.00	0.00
10,000	100.00	66.67	66.67	33.33	0.00	0.00	66.67	33.33	16.67	0.00
20,000	100.00	100.00	33.33	50.00	0.00	70.00	83.33	66.67	0.00	0.00
40,000	100.00	100.00	0.00	16.67	0.00	94.37	83.33	33.33	16.67	0.00
80,000	100.00	100.00	50.00	0.00	0.00	0.00	100.00	50.00	16.67	0.00

TABLA 49. Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	100.00	100.00	98.57	97.14	95.71	100.00	100.00	98.57	97.14	95.71
10,000	100.00	100.00	97.14	100.00	98.57	5.88	100.00	98.57	100.00	98.57
20,000	100.00	100.00	97.14	98.57	97.14	0.00	100.00	95.71	97.14	95.71
40,000	100.00	100.00	97.14	97.14	98.57	63.33	100.00	97.14	98.57	98.57
80,000	100.00	100.00	98.57	97.14	97.14	95.31	100.00	98.57	98.57	98.57

TABLA 50. Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia aleatorios en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	75.00	50.00	0.00	40.00	0.00	100.00	50.00	0.00	40.00	0.00
10,000	60.00	50.00	25.00	10.00	0.00	83.33	10.00	5.00	10.00	0.00
20,000	100.00	75.00	10.00	35.00	10.00	5.88	60.00	30.00	25.00	10.00
40,000	100.00	80.00	10.00	30.00	25.00	0.00	55.00	25.00	35.00	5.00
80,000	100.00	95.00	0.00	5.00	5.00	68.33	55.00	35.00	30.00	10.00

TABLA 51. Porcentaje de eficacia para latidos de clase 1 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.57	98.63	99.05	95.47	96.25	5.00	98.63	99.05	95.47	96.25
10,000	98.66	98.55	98.86	98.32	99.48	0.00	99.75	99.54	99.66	99.79
20,000	98.93	98.91	98.57	99.16	99.32	0.00	99.64	99.57	99.71	99.68
40,000	99.05	98.97	98.55	98.70	98.86	17.00	99.63	99.66	99.64	99.57
80,000	99.14	99.21	99.05	99.13	98.89	1.00	99.68	99.70	99.61	99.57

TABLA 52. Porcentaje de eficacia para latidos de clase 2 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	98.51	98.35	95.21	88.60	90.08	98.57	98.35	95.21	88.60	90.08
10,000	99.17	98.51	95.87	96.53	94.21	15.00	97.02	94.21	95.21	93.22
20,000	99.34	98.68	97.52	95.37	95.54	0.00	98.02	94.71	94.88	95.37
40,000	99.17	99.34	98.35	98.02	97.85	0.00	98.02	97.02	96.53	97.02
80,000	99.50	99.34	95.21	96.53	96.36	15.00	97.69	97.02	96.86	96.53

TABLA 53. Porcentaje de eficacia para latidos de clase 3 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	96.32	92.10	84.19	91.91	90.26	98.51	92.10	84.19	91.91	90.26
10, 000	96.88	94.12	91.18	93.01	93.01	99.46	95.04	78.86	93.20	93.20
20, 000	96.88	96.14	94.12	95.22	92.83	2.00	95.40	91.54	95.04	93.20
40, 000	97.98	96.32	96.32	92.65	92.46	0.00	95.96	94.12	94.30	93.93
80, 000	96.88	96.69	84.19	94.30	94.30	0.00	95.77	93.75	93.75	93.93

TABLA 54. Porcentaje de eficacia para latidos de clase 4 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	65.97	52.36	40.84	42.41	37.70	96.32	52.36	40.84	42.41	37.70
10, 000	66.49	58.64	54.45	54.97	52.36	98.84	41.36	29.32	49.21	49.21
20, 000	62.83	62.83	48.69	53.40	52.88	99.45	47.12	50.79	49.21	49.21
40, 000	68.06	65.97	46.07	51.31	51.31	2.00	54.45	52.88	53.93	51.31
80, 000	70.16	67.54	40.84	53.40	56.02	0.00	58.12	53.40	54.45	52.88

TABLA 55. Porcentaje de eficacia para latidos de clase 5 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	30.00	20.00	10.00	0.00	0.00	65.97	20.00	10.00	0.00	0.00
10, 000	60.00	50.00	10.00	0.00	0.00	97.43	20.00	10.00	0.00	0.00
20, 000	70.00	50.00	10.00	20.00	0.00	99.01	30.00	0.00	0.00	0.00
40, 000	70.00	70.00	20.00	20.00	0.00	99.36	20.00	0.00	0.00	0.00
80, 000	70.00	70.00	10.00	0.00	0.00	3.00	40.00	0.00	0.00	0.00

TABLA 56. Porcentaje de eficacia para latidos de clase 6 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	0.00	16.67	0.00	0.00	0.00	30.00	16.67	0.00	0.00	0.00
10, 000	33.33	33.33	0.00	0.00	0.00	51.31	0.00	0.00	0.00	0.00
20, 000	66.67	33.33	0.00	0.00	0.00	96.88	0.00	0.00	0.00	0.00
40, 000	83.33	33.33	0.00	0.00	0.00	98.51	0.00	0.00	0.00	0.00
80, 000	83.33	50.00	0.00	0.00	0.00	99.43	0.00	0.00	0.00	0.00

TABLA 57. Porcentaje de eficacia para latidos de clase 7 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10, 000	100.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00
20, 000	0.00	0.00	0.00	0.00	0.00	57.07	0.00	0.00	0.00	0.00
40, 000	0.00	0.00	0.00	0.00	0.00	97.98	0.00	0.00	0.00	0.00
80, 000	0.00	0.00	0.00	0.00	0.00	98.84	0.00	0.00	0.00	0.00

TABLA 58. Porcentaje de eficacia para latidos de clase 8 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	90.81	87.43	84.43	75.23	76.17	0.00	87.43	84.43	75.23	76.17
10, 000	94.56	90.43	89.12	83.68	87.43	16.67	85.37	80.49	78.42	83.49
20, 000	96.06	90.24	84.80	86.49	89.12	60.00	86.49	88.37	86.30	87.80
40, 000	96.62	93.81	88.56	89.12	90.62	64.40	89.31	89.87	89.68	90.06
80, 000	96.81	93.81	84.43	90.06	90.06	97.43	88.18	90.99	90.99	90.43

TABLA 59. Porcentaje de eficacia para latidos de clase 9 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	65.00	60.00	58.33	5.00	5.00	90.81	60.00	58.33	5.00	5.00
10,000	68.33	63.33	50.00	45.00	35.00	0.00	41.67	41.67	36.67	25.00
20,000	80.00	71.67	45.00	53.33	10.00	0.00	60.00	18.33	23.33	1.67
40,000	80.00	71.67	56.67	56.67	35.00	60.00	58.33	20.00	33.33	28.33
80,000	85.00	78.33	58.33	50.00	43.33	65.45	61.67	25.00	21.67	31.67

TABLA 60. Porcentaje de eficacia para latidos de clase 10 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	0.00	0.00	0.00	0.00	0.00	65.00	0.00	0.00	0.00	0.00
10,000	0.00	0.00	0.00	0.00	0.00	90.24	0.00	0.00	0.00	0.00
20,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40,000	0.00	0.00	0.00	0.00	0.00	16.67	0.00	0.00	0.00	0.00
80,000	0.00	0.00	0.00	0.00	0.00	50.00	0.00	0.00	0.00	0.00

TABLA 61. Porcentaje de eficacia para latidos de clase 11 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5,000	35.29	41.18	0.00	0.00	0.00	0.00	41.18	0.00	0.00	0.00
10,000	35.29	41.18	11.76	5.88	0.00	53.33	5.88	0.00	5.88	0.00
20,000	35.29	47.06	0.00	11.76	0.00	92.31	17.65	5.88	11.76	0.00
40,000	23.53	35.29	11.76	0.00	5.88	0.00	11.76	5.88	0.00	0.00
80,000	41.18	47.06	0.00	5.88	0.00	16.67	17.65	5.88	0.00	0.00

TABLA 62. Porcentaje de eficacia para latidos de clase 12 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	100.00	100.00	50.00	0.00	0.00	35.29	100.00	50.00	0.00	0.00
10, 000	100.00	66.67	66.67	33.33	0.00	0.00	66.67	33.33	16.67	0.00
20, 000	100.00	100.00	33.33	50.00	0.00	70.00	83.33	66.67	0.00	0.00
40, 000	100.00	100.00	0.00	16.67	0.00	94.37	83.33	33.33	16.67	0.00
80, 000	100.00	100.00	50.00	0.00	0.00	0.00	100.00	50.00	16.67	0.00

TABLA 63. Porcentaje de eficacia para latidos de clase 13 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	100.00	100.00	98.57	97.14	95.71	100.00	100.00	98.57	97.14	95.71
10, 000	100.00	100.00	97.14	100.00	98.57	5.88	100.00	98.57	100.00	98.57
20, 000	100.00	100.00	97.14	98.57	97.14	0.00	100.00	95.71	97.14	95.71
40, 000	100.00	100.00	97.14	97.14	98.57	63.33	100.00	97.14	98.57	98.57
80, 000	100.00	100.00	98.57	97.14	97.14	95.31	100.00	98.57	98.57	98.57

TABLA 64. Porcentaje de eficacia para latidos de clase 14 en conjuntos de referencia equi-proporcionados en la clasificación de anomalías.

Total de registros	Mejor Candidato					Votación				
	1-NN		Redes neuronales			1-NN		Redes neuronales		
	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3	<i>CCI</i>	<i>Euc</i>	Arq-1	Arq-2	Arq-3
5, 000	75.00	50.00	0.00	40.00	0.00	100.00	50.00	0.00	40.00	0.00
10, 000	60.00	50.00	25.00	10.00	0.00	83.33	10.00	5.00	10.00	0.00
20, 000	100.00	75.00	10.00	35.00	10.00	5.88	60.00	30.00	25.00	10.00
40, 000	100.00	80.00	10.00	30.00	25.00	0.00	55.00	25.00	35.00	5.00
80, 000	100.00	95.00	0.00	5.00	5.00	68.33	55.00	35.00	30.00	10.00