



Universidad de Cuenca

Facultad de Ingeniería

Escuela de Electrónica y Telecomunicaciones

Algoritmos de aprendizaje no supervisado para la estimación de la resistencia a la insulina y el síndrome metabólico en el adulto mayor de la ciudad de Cuenca

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN
DEL TÍTULO DE INGENIERO EN ELECTRÓNICA Y
TELECOMUNICACIONES

Autor :

Christian Xavier Vintimilla García. C.I. 0105140438

Director :

PhD. Sara Wong de Balzan. C.I. 081929618

Co-Director :

PhD. Darwin Fabián Astudillo Salinas. C.I. 0103907036

Cuenca - Ecuador

2017



Resumen

En este trabajo de titulación se explora la [Resistencia a la Insulina \(RI\)](#) y el [Síndrome Metabólico \(SM\)](#) en el adulto mayor cuencano desde el punto de vista de métodos de clasificación no supervisada. En el caso de la [RI](#) se analizaron cinco métodos de estimación usando una clasificación de K-medias sobre una población de 119 adultos mayores de la ciudad de Cuenca quienes se sometieron a una [Prueba Oral de Tolerancia a la Glucosa \(POTG\)](#) de dos puntos. El algoritmo de K-medias, con $K=2$ y $K=3$ fue aplicado en experimentos unidimensionales para los métodos Homa-IR, Quicki, Avignon, Matsuda. Los resultados obtenidos permitieron desarrollar una plataforma de ayuda al diagnóstico de [RI](#) y fueron objeto de dos publicaciones (en el IV Congreso de Tecnologías de la Información y Comunicación TIC-EC 2017, y el IEEE ETCM 2017: 2nd IEEE Ecuador Technical Chapters Meeting). Para el estudio del [SM](#) usando [Self-Organizing Map \(SOM\)](#) de Kohonen, se analizaron dos tipos de normalización (binaria y por rangos) para las entradas de la red neuronal usando una población de 387 adultos mayores. Los resultados, usando un pre-procesamiento por rangos permiten una mejor clasificación de la población en todos los casos. Este estudio ha permitido seleccionar el tipo de pre-procesamiento para el diagnóstico de [SM](#) en la población de adultos mayores de la ciudad de Cuenca usando [SOM](#) y fue objeto de una publicación en el V Congreso REDU 2017 y el II Congreso I+D+Ingeniería. Los trabajos futuros están orientados a validar los resultados obtenidos en otras poblaciones de adultos mayores.

Palabras clave : Aprendizaje no supervisado, K-medias, redes neuronales, SOM, Kohonen, Homa-IR, Quicki, resistencia a la insulina, síndrome metabólico, adulto mayor



Abstract

In this degree work, *Insulin Resistance* (IR) and *Metabolic Syndrome* (MS) are explored in the Cuenca's elderly population, from the point of view of unsupervised classification methods. In the case of IR, five estimation methods were analyzed using a K-means classification on a population of 119 people older than 65 years old who underwent a two-point *Oral Glucose Tolerance Test* (OGTT). The K-means algorithm with $K = 2$ and $K = 3$ was applied in one-dimensional experiments for the Homa-IR, Quicki, Avignon, and Matsuda methods. The results obtained allowed the development of a platform to aid in the diagnosis of IR. These findings were object of two publications (IV Congress of Information and Communication Technologies TIC-EC 2017, and IEEE ETCM 2017: 2nd IEEE Ecuador Technical Chapters Meeting). For the study of MS using Kohonen's SOM, two types of normalization (binary and by rank) were analyzed for the inputs of the neural network using a population of 387 elderly people. The results, using a pre-processing by ranges allow a better classification of the population in all cases. This study allowed to select the type of pre-processing for the diagnosis of MS in the elderly population of the city of Cuenca using SOM and was the object of a publication in the V Congress REDU 2017 and the II Congress I+D+Ingeniería. The future work is oriented to validate the results obtained in other elderly populations.

Keywords : Unsupervised learning, K-means, neuronal networks, self-organizing maps, Kohonen, Homa-IR, Quicki, insulin resistance, metabolic syndrome, elderly





Índice general

Resumen	3
Abstract	5
Índice general	7
Índice de figuras	11
Índice de tablas	13
Dedicatoria	19
Agradecimientos	21
Abreviaciones y acrónimos	23
1. Introducción	25
1.1. Antecedentes	26
1.2. Identificación del Problema	27
1.3. Justificación	27
1.4. Alcance	27



1.5. Objetivos	28
1.5.1. Objetivo General	28
2. Marco Teórico: Síndrome Metabólico y la Resistencia a la Insulina	29
2.1. Glosario de términos académico-científicos relacionados al SM y la RI [1]	30
2.2. El Síndrome Metabólico	30
2.2.1. Criterios Diagnósticos	31
2.3. Insulina, Resistencia a la Insulina	34
2.3.1. Metabolismo de la insulina	35
2.3.2. Resistencia a la insulina (RI)	36
2.4. Sensibilidad a la insulina (Sensibilidad a la Insulina (SI))	37
2.4.1. Criterios para el diagnóstico de la SI	37
2.4.2. Cuantificación de la SI a partir de la Prueba Oral de Tolerancia a la Glucosa (POTG)	37
3. Marco Teórico: clasificación no supervisada	39
3.1. Minería de datos	40
3.2. Clasificación supervisada	40
3.3. Clasificación No Supervisada	41
3.3.1. Clustering	41
3.3.2. Algoritmo de agrupamiento de K-medias	42
3.3.3. Modelo de Kohonen	43
3.3.4. Representación del modelo	44
3.3.5. Algoritmo de aprendizaje	46



4. Clasificación no supervisada con el algoritmo de K-medias	49
4.1. Estimación de la RI	50
4.1.1. Base de Datos para la estimación de la RI	50
4.1.2. Algoritmo de K-medias y Experimentos	50
4.1.3. Metodología con $K=2$ y $K=3$	51
4.2. Resultados de la estimación de la Resistencia a la Insulina con $K=2$	53
4.3. Resultados de la estimación de la Resistencia a la Insulina con $K=3$	56
4.4. Diagnóstico con nuevos pacientes	60
4.5. Conclusiones	62
5. Clasificación no supervisada mediante Redes Neuronales de Kohonen	63
5.1. Estimación del SM	64
5.1.1. Base de Datos para la estimación del SM	64
5.1.2. Algoritmo de Kohonen y Experimentos	64
5.2. Uso del software Viscovery SOMine	65
5.2.1. Metodología Viscovery SOMine	66
5.3. Resultados de la estimación del Síndrome Metabólico	69
5.3.1. Entrenamiento de la red de Kohonen	69
5.3.2. Validación de la red de Kohonen	72
5.4. Conclusiones	74
6. Conclusiones y Trabajos Futuros	75
6.1. Conclusiones	76
6.2. Trabajos Futuros	77



A. Artículo: Agrupamiento de K-medias para estimación de insulino-resistencia en Adultos Mayores de Cuenca	81
B. Artículo: An aide Diagnosis System Based on k-means for insulin resistance assessment in elderly people from the Ecuadorian highlands	93
C. Artículo: Normalización de entradas de una red de Kohonen para clasificación de Síndrome Metabólico en Adultos Mayores de la ciudad de Cuenca	101
Bibliografía	111



Índice de figuras

2.1. Sistema Regulatorio Glucosa-Insulina, Fuente: [2, Fig. 1]	35
3.1. Ejemplo del Algoritmo de K -medias. (a) Conjunto de datos original. (b) Se ubican centroides al azar. (c) Se ubica las muestras al centroide del clúster más cercano (color verde o azul). (d) Se mueve el centroide del clúster a la media de los puntos asignados a él. (e)y(f) Agrupamiento luego de 2 iteraciones.	43
3.2. Arquitectura del SOM, Fuente: [3, Fig. 4.4]	44
3.3. Esquema de actualización de pesos para la red SOM , Fuente: [3, Fig. 4.5]	46
3.4. Reducción del radio de la <i>Best Matching Unit</i> (BMU)	47
4.1. Distribución de clústeres C-N y C-RI y puntos de corte para cada experimento unidimensional.	52
4.2. Distribución de clústeres C-N, C-RI y C-L y puntos de corte para cada experimento unidimensional.	52
4.3. Sistema de Diagnóstico. Valores de Homa-IR, Quicki, Avignon y Matsuda se indican para un nuevo paciente representados por el carácter X. El color verde representa valores normales (C-N), el color azul representa el límite entre un sujeto sano y uno con RI (C-L) y el color rojo representa un sujeto con RI (C-RI).	61
5.1. Visualización con el software Viscovery SOMine.	67
5.2. Clasificación de dos clústeres, de acuerdo a un color específico (celeste y rojo).	68



- 5.3. Visualización con el software Viscosity SOMine, se observa que en los dos clústeres hay sujetos con valores normales y alterados de Circunferencia Abdominal, Glicemia y Triglicéridos. 68
- 5.4. a) SOM usando pre-procesamiento binario. b) SOM usando pre-procesamiento por rangos para la base completa. 74



Índice de tablas

2.1. Factores de SM según Adult Treatment Panel III (ATP-III) e International Diabetes Federation (IDF)	33
2.2. Valores de la circunferencia abdominal de acuerdo a la especificidad étnica. . . .	34
2.3. Métodos para cuantificar la Sensibilidad a la Insulina, (I_0 : Insulina basal, I_{120} : Insulina postprandial, G_0 : Glucosa Basal y G_{120} : Glucosa Postprandial).	38
2.4. Valores normales de glucosa basal, insulina basal, glucosa postprandial e insulina postprandial.	38
3.1. Métricas para comparar dos vectores.	45
4.1. Medidas antropométricas y valores de la POTG , calculadas con la media \pm desviación estándar (std) y el rango de cada parámetro.	50
4.2. Número de Adultos Mayores por clústeres ($C-N$ y $C-RI$) y puntos de corte para Homa-IR, Quicki, Matsuda y Avignon.	53
4.3. Valores de Glucosa (mg/dL), Insulina ($\mu UI/mL$) y Circunferencia Abdominal (cm) para cada clúster ($C-N$ y $C-RI$) según cada método.	54
4.4. Características obtenidas para los métodos de SI . (I_0 : Insulina basal, I_{120} : Insulina postprandial, G_0 : Glucosa Basal y G_{120} : Glucosa Postprandial), Normal (N), Elevado (E) y alterado (A).	54
4.5. Número de Adultos Mayores por clústeres ($C-N$, $C-L$ y $C-RI$) y puntos de corte para Homa-IR, Quicki, Matsuda y Avignon.	56



4.6. Valores de Glucosa (mg/dL), Insulina ($\mu UI/mL$), Circunferencia Abdominal (cm) Hemoglobina Glicosilada (mg/dL) para cada clúster ($C-N$, $C-L$ y $C-RI$) según cada método.	57
4.7. Características obtenidas para los métodos de SI . (I_0 : Insulina basal, I_{120} : Insulina postprandial, G_0 : Glucosa Basal y G_{120} : Glucosa Postprandial), Normal (N), Elevado (E) y alterado (A).	58
4.8. Medidas antropométricas y valores de la POTG de otra base de datos de Adultos Mayores.	60
5.1. Factores de SM para 142 hombres mayores a 65 años.	64
5.2. Factores de SM para 245 mujeres mayores a 65 años.	64
5.3. Valores de Corte para la base de datos de SM	65
5.4. Normalización por Intervalo.	65
5.5. Características de los clústeres para toda la población usando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.	69
5.6. Características de los clústeres para la población femenina utilizando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.	70
5.7. Características de los clústeres para la población masculina utilizando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.	70
5.8. Características de los clústeres para toda la población usando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.	71
5.9. Características de los clústeres para la población femenina utilizando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.	72
5.10. Características de los clústeres para la población masculina utilizando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.	72



- 5.11. Características de los clústeres para toda la población usando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL. 73
- 5.12. Características de los clústeres para toda la población usando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL. 73



Christian Xavier Vintimilla García, autor del trabajo de titulación “Algoritmos de aprendizaje no supervisado para la estimación de la resistencia a la insulina y el síndrome metabólico en el adulto mayor de la ciudad de Cuenca”, certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor.

Cuenca, Noviembre de 2017

Christian Xavier Vintimilla García
C.I: 010514043-8



Christian Xavier Vintimilla García en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación “Algoritmos de aprendizaje no supervisado para la estimación de la resistencia a la insulina y el síndrome metabólico en el adulto mayor de la ciudad de Cuenca”, de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, Noviembre de 2017

Christian Xavier Vintimilla García

C.I: 010514043-8



Dedicatoria

A Dios porque sin él no tenemos nada y por haberme permitido culminar este gran logro en mi vida. A mis padres por ese sacrificio incondicional y a mis hermanos por el aliento para seguir, siempre estuvieron en mi vida, me inculcaron honestidad, dignidad y humildad son lo mejor que Dios me ha dado.

Christian Xavier





Agradecimientos

A Dios porque en sus manos está mi vida y el hecho de seguir adelante para ayudar a los demás. A mis padres y hermanos que siempre estuvieron apoyándome. De manera especial a mis directores de Tesis a la Ing. Sara Wong, e Ing. Fabián Astudillo por los ánimos por aportar su tiempo, palabras de aliento y ayuda incondicional en la culminación de mi Tesis. A mis profesores por sus conocimientos impartidos. A mis amigos encontrados en el caminar de la carrera, compartimos momentos inolvidables.

Christian Xavier





Abreviaciones y Acrónimos

AACE *American Association of Clinical Endocrinologists.* [7](#), [8](#)

ATP-III *Adult Treatment Panel III.* [7–9](#)

BMU *Best Matching Unit.* [22](#), [23](#)

EGIR *European Group for the Study of Insulin Resistance.* [7](#)

HDL Lipoproteínas de baja densidad. [40](#), [41](#)

IDF *International Diabetes Federation.* [7–9](#)

LVQ Linear Vector Quantization. [19](#)

POTG Prueba Oral de Tolerancia a la Glucosa. [2](#), [3](#), [13](#), [26](#), [31](#), [35](#), [36](#)

RI Resistencia a la Insulina. [2](#), [3](#), [6–8](#), [12](#), [13](#), [26](#), [27](#), [29](#), [31](#), [32](#), [34–36](#), [38](#), [52](#), [53](#)

RN Redes Neuronales. [2](#)

RNA Redes Neuronales Artificiales. [19](#)

SI Sensibilidad a la Insulina. [7](#), [13](#), [25](#), [27](#), [30](#), [31](#), [34](#), [35](#), [38](#), [52](#), [53](#)

SM Síndrome Metabólico. [2](#), [3](#), [6–9](#), [13](#), [14](#), [40–43](#), [45–50](#), [52](#), [53](#)

SOM *Self-Organizing Map.* [3](#), [19](#), [40–42](#), [50](#), [52](#), [53](#)


WHO *World Health Organization.* [7](#)





Capítulo 1

Introducción



En este capítulo se introduce la Resistencia a la Insulina y el Síndrome Metabólico y la necesidad de evaluar estas condiciones prediabéticas en los Adultos Mayores, buscando otras alternativas que sirvan de respaldo para el diagnóstico médico, como son algoritmos de aprendizajes no supervisados. Adicionalmente se presenta la justificación del problema, alcance y objetivos de este proyecto de tesis.

1.1. Antecedentes

La Resistencia a la Insulina (RI) y el Síndrome Metabólico (SM) son condiciones que pueden favorecer la aparición de la diabetes, por lo que, es importante apoyar los estudios que ayuden al diagnóstico oportuno de estas enfermedades.

La práctica más común para determinar la RI es mediante métodos como, por ejemplo; Homa-IR, Quicki, Matsuda, Avignon [4]. El método indirecto más utilizado es el índice Homa-IR, pero tiene ciertas limitaciones de las cuales podemos citar: i) su incapacidad para detectar la RI en sus primeras etapas y que las alteraciones de la insulina y la glucemia en ayunas se manifiestan cuando las disfunciones metabólicas ya están presentes en el cuerpo, ii) alta sensibilidad en pacientes que ya tienen RI pero no en pacientes diabéticos con disfunción beta pancreática, y iii) la ausencia de un rango de referencia para el diagnóstico de RI. Existen una gran cantidad de métodos para el diagnóstico de la RI, pero ninguna ha sido evaluada en una población de adultos mayores [4].

El SM agrupa un conjunto de factores de riesgo que pueden desarrollar enfermedades cardíacas y diabetes tipo 2. Estos factores son Presión arterial ($\geq 130/85$ mmHg), Triglicéridos (≥ 150 mg/dL), Glucosa sanguínea en ayunas (≥ 110 mg/dL), HDL-Colesterol (Hombres < 40 , Mujeres < 50), Circunferencia abdominal (Hombres > 90 cm, Mujeres > 80 cm). Según el National Cholesterol Education Program-Adult Treatment Panel III (NECP-ADP III) [5, 6], cuando hay tres factores de riesgo se establece un diagnóstico de SM. Sin embargo, no existe un consenso en cuanto a los valores de los límites alterados para estas variables.

En este trabajo se propone diagnosticar la RI y el SM usando algoritmos de clasificación no supervisados que permiten separar en determinadas clases las condiciones patológicas del paciente. Dentro de las alternativas de clasificación se tiene el uso del algoritmo de agrupamiento de k-medias y el uso de Redes Neuronales (RN).

En el 2015, adultos mayores, mayores de 65 años sin diabetes fueron reclutados en el proyecto DIUC-PROMETEO denominado: Variabilidad de la frecuencia cardíaca y sensibilidad a la insulina en la población anciana de Cuenca, Ecuador. A cada sujeto se le realizó una hematología, un perfil lipídico, medidas antropométricas (IMC), cuestionario internacional de actividad física 2 (IPAQ) y una Prueba Oral de Tolerancia a la Glucosa (POTG) de dos puntos (Glucosa e Insulina en ayuno y otra muestra de Glucosa e Insulina tomada en un intervalo de 120 min después de la ingesta de 75 gramos de glucosa). Esta base de datos es utilizada en esta investigación, está compuesta por 120 sujetos, clasificados en dos tipos de población: sujetos con SM y sujetos sin SM según el criterio de la NECP-ATP III [6, 7]. En este trabajo se usaron datos preliminares de la Base de Datos del adulto mayor, sin embargo, aspectos como la evaluación de la RI y el diagnóstico del SM bajo diferentes criterios no fueron explorados.

1.2. Identificación del Problema

La metodología del trabajo comprende cuatro etapas:

- Diseño de experimentos para cada algoritmo: con observaciones unidimensionales (Homa-IR, Quicki, Matsuda, Avignon) y observaciones multidimensionales (POTG con dos puntos). Se considerará los factores asociados al SM.
- Evaluar el algoritmo de agrupamiento de k-medias: se pueden establecer 2 y 3 clases que permitan diferenciar la población con SM y/o RI. En base a factores clínicos y mediante observaciones unidimensionales (de acuerdo al índice de sensibilidad a la insulina).
- Evaluar una Red Neuronal: se tiene una base de datos ampliada de 387 sujetos, los que se dividirán en un conjunto de entrenamiento y un conjunto de prueba. En el algoritmo también podrán ser utilizados factores clínicos y observaciones unidimensionales.
- Análisis de resultados de acuerdo a valores obtenidos en la clasificación: se pretende diferenciar la clasificación obtenida y poder validar las dos clasificaciones.

1.3. Justificación

En esta tesis se propone el uso de los algoritmos de aprendizaje no supervisado, para clasificar sujetos adultos mayores con RI y SM, la misma se enmarca dentro de la línea de investigación prioritaria de la carrera: Desarrollo de hardware/software/firmware para aplicaciones de electromedicina, telemedicina y teleeducación. La aplicación que se desarrollará en la presente tesis puede contribuir a la prestación de servicios públicos de salud para la prevención y diagnóstico de condiciones en el adulto mayor. Los trabajos desarrollados en la presente tesis se enmarcan dentro de la continuación de las investigaciones previamente realizadas por el Ing. Andrés Parra en la Tesis de Ing. Electrónica y Telecomunicaciones de la Facultad de Ingeniería de la Universidad de Cuenca “Desarrollo y validación de algoritmos para el estudio de la variabilidad de la frecuencia cardiaca en adultos mayores 2015”.

1.4. Alcance

Este proyecto de tesis tiene la finalidad de utilizar clasificación no supervisada, específicamente el algoritmo de K -medias para diagnosticar la RI y los *Self-Organizing Map* (SOM) de Kohonen para el diagnóstico de SM utilizando una base de datos de 119 sujetos para el diagnóstico de RI y de 387 sujetos para el diagnóstico de SM.



1.5. Objetivos

Clasificar a la población de adultos mayores con Síndrome Metabólico y/o Resistencia a la Insulina mediante algoritmos de aprendizaje no supervisado.

1.5.1. Objetivo General

- Utilizar métodos indirectos y adaptar pesos a las variables, para implementarlas en el algoritmo de agrupamiento de K -medias y en las redes neuronales.
- Determinar las variables más discriminantes para Síndrome Metabólico y/o Resistencia a la Insulina.
- Evaluar los resultados obtenidos con el algoritmo de agrupamiento de K -medias y con redes neuronales.
- Diseñar e implementar una aplicación que permita evaluar nuevos sujetos en la base de datos.



Capítulo 2

Marco Teórico: Síndrome Metabólico y la Resistencia a la Insulina

UNIVERSIDAD DE CUENCA
desde 1867

En este capítulo se definen los conceptos de los trastornos metabólicos como son el Síndrome Metabólico, la Resistencia a la Insulina y otros relativos a las condiciones pre-diabéticas, de forma de dar a conocer las fisiologías del problema tratado referentes al proyecto de tesis. Adicionalmente, se describen criterios de diagnóstico para determinar la Sensibilidad a la Insulina mediante métodos indirectos.



2.1. Glosario de términos académico-científicos relacionados al SM y la RI [1]

Ayuno: abstinencia de la ingesta de alimentos sólidos y líquidos durante por lo menos 8 horas.

Comorbilidad: la presencia de uno o más trastornos (o enfermedades) además de la enfermedad o trastorno primario.

Factor de riesgo: cualquier rasgo, característica o exposición de un individuo que aumente su probabilidad de sufrir una enfermedad o lesión.

Glucosa alterada en ayunas: valores de glucemia, tomados tras 8 horas de ayuno, que se encuentran entre 100 y 125 mg/dL.

Intolerancia a la glucosa: valores de glucemia, tomados tras 2 horas post ingesta de 75 gramos de glucosa oral, que se encuentran entre 140 y 199 mg/dL.

Prediabetes: término utilizado para individuos con alto riesgo para padecer diabetes mellitus tipo 2, definido por una glucemia alterada en ayunas y/o intolerancia a la glucosa.

Diabetes: enfermedad crónica que se desencadena cuando el organismo pierde su capacidad de producir suficiente insulina o de utilizarla con eficacia.

Diabetes mellitus tipo 2: afectación crónica que se caracteriza por hiperglucemia persistente (elevación del nivel de glucosa en sangre) con alteración del metabolismo de los lípidos, proteínas e hidratos de carbono que se desencadena cuando el organismo pierde su capacidad de producir suficiente insulina o de utilizarla con eficacia, y que produce complicaciones crónicas micro y macro vasculares.

2.2. El Síndrome Metabólico

El proceso biológico de Selección Natural en el pasado, hizo que nuestro cuerpo sea capaz de tener un gran ahorro energético con poco consumo de alimento. En la actualidad se tiene un ambiente de exceso de alimentos y un mínimo gasto energético. Este proceso de alteraciones del metabolismo de reservas de grasas y glúcidos, han causado grandes enfermedades del siglo XXI: obesidad, enfermedades cardíacas, diabetes e hipertensión, todas ellas relacionadas con el SM [5].

No hay una definición precisa de **SM**, se trata de un conjunto de problemas de salud que afectan a un individuo en forma simultánea o secuencial, causada por factores genéticos y factores relacionados al estilo de vida que favorecen a la aparición de la **RI**. Debido al grupo de factores metabólicos el **SM** ha tenido otros nombres a través de la historia como Síndrome Plurimetabólico, Síndrome de Resistencia a la Insulina, Síndrome X o el cuarteto mortífero.

2.2.1. Criterios Diagnósticos

Para el diagnóstico de **SM** se han expuesto varios criterios pertenecientes a la:

- *World Health Organization (WHO)*.
- *European Group for the Study of Insulin Resistance (EGIR)*.
- *Adult Treatment Panel III (ATP-III)*.
- *American Association of Clinical Endocrinologists (AACE)*.
- *International Diabetes Federation (IDF)*.

La **WHO** en 1998 clasificó a los sujetos con **SM** si presentan Diabetes, **RI** y al menos 2 de los siguientes criterios [8]:

- Presión Arterial ≥ 140 mmHg.
- Triglicéridos >150 mg/dL o Colesterol HDL <35 mg/dL en varones y <39 mg/dL en mujeres.
- Índice de masa corporal (IMC) >30
- Cociente cintura-cadera >90 cm en varones y >85 cm en mujeres.
- Microalbuminuria >20 ug/min.

El aspecto central según la **WHO** para que un individuo tenga **SM** es la presencia de diabetes y **RI**. Para determinar estas dos condiciones es necesario pruebas de **SI**, mediante modelos de difícil aplicación u otros que no dan resultados precisos. Además se evidenció que el IMC no es una medida fiable de obesidad y esta varía según la edad del individuo, mas bien el mejor indicador es el perímetro de cintura. Por último, la microalbuminuria no aparece siempre en pacientes que tienen diabetes, por lo tanto no contrasta con el síntoma principal según la **WHO** [9].

El **EGIR** corrige las limitaciones que se mostraron en el punto anterior, pero aún estableciendo a la **RI** como factor predominante de **SM** y 2 o más de los siguientes criterios [10]:

- Circunferencia de cintura ≥ 94 cm en hombres y ≥ 80 cm en mujeres.

- Triglicéridos $\geq 150 \text{ mg/dl}$ o colesterol HDL $\leq 39 \text{ mg/dl}$, tanto en hombres como en mujeres.
- Presión arterial $\geq 140/90 \text{ mmHg}$ o estar en tratamiento para presión arterial elevada.
- Intolerancia a la glucosa, pero no diabetes.

En este criterio se excluyeron a los sujetos que tienen diabetes y se estableció un modelo más fácil de aplicar, cuantificando con métodos sencillos la [RI](#) [9].

En 2001 el [ATP-III](#) estableció una definición de [SM](#) donde la [RI](#) no era un criterio de diagnóstico. El [ATP-III](#) establece que un individuo tiene [SM](#) si tiene 3 o más de los siguientes 5 parámetros [11]:

- Circunferencia de cintura $>102 \text{ cm}$ en hombres y >88 en mujeres.
- Triglicéridos $\geq 150 \text{ mg/dl}$.
- Colesterol HDL $<40 \text{ mg/dl}$ en hombres y <50 en mujeres.
- Presión arterial $\geq 130/85 \text{ mmHg}$.
- Glucosa basal $\geq 110 \text{ mg/dl}$.

Este criterio fue aceptado de mejor manera por su fácil aplicación. El [ATP-III](#) sugiere rangos de circunferencia abdominal de acuerdo a la etnia, aunque no existía valores para poblaciones específicas, después salió una versión modificada en la cual se establece que el perímetro abdominal para la población de hombres hispanos es de $\geq 94 \text{ cm}$, y Glucosa basal a $\geq 100 \text{ mg/dl}$ [11].

En la [AACE](#) la [RI](#) es clave para el diagnóstico de [SM](#), siendo la intolerancia a la glucosa o glucosa en ayuno anormal el factor principal más cualquiera de los siguientes parámetros basado en juicio clínico [10]:

- Índice de masa corporal ≥ 25 .
- Hipertrigliceridemia $>150 \text{ mg/dl}$.
- Colesterol HDL $<40 \text{ mg/dl}$ en varones y $<50 \text{ mg/dl}$ en mujeres.
- Presión arterial $\geq 130/85 \text{ mmHg}$.
- Glucosa basal $\geq 110 \text{ mg/dl}$, incluido diabetes.

Este criterio excluía la obesidad central, e indicando a su vez que un individuo que tiene [SM](#) contribuye a que tenga [RI](#) y no lo contrario.

En el 2005 la [IDF](#) propuso otra definición de [SM](#) debido al aumento de personas con obesidad en el mundo y a los riesgos que esto conlleva. El factor más relevante es la perímetro abdominal de acuerdo a la etnia, y 2 o más de los siguientes parámetros [12]:

- Triglicéridos ≥ 150 mg/dl.
- Colesterol HDL < 40 mg/dl en varones y < 50 mg/dl en mujeres.
- Presión sistólica ≥ 130 mmHg y/o diastólica: ≥ 85 mmHg.
- Glucosa basal ≥ 100 mg/dl, incluido diabetes o diabetes tipo 2.

Por último la Asociación Latinoamericana de Diabetes en 2010 publicó criterios obtenidos del Grupo Latinoamericano para el estudio de [SM](#), indicando como factor principal el perímetro abdominal ≥ 94 cm en varones y ≥ 88 cm en mujeres más 2 o más de los siguientes factores:

- Triglicéridos ≥ 150 mg/dl.
- Colesterol HDL < 40 mg/dl en varones y < 50 mg/dl en mujeres.
- Presión sistólica ≥ 130 mmHg y/o diastólica: ≥ 85 mmHg.
- Glucosa basal ≥ 100 mg/dl, incluido diabetes o diabetes tipo 2.

Los criterios de [IDF](#) y [ATP-III](#) son los más usados, ya que permiten el diagnóstico temprano del [SM](#) y debido a la necesidad de obtener resultados sencillos y eficaces.

	ATP-III	IDF
Componentes	Presencia de 3 componentes o mas	C. Abdominal indispensable + 2 componentes
C. Abdominal(cm)	> 102 en hombres y > 88 en mujeres	De acuerdo a la especificidad étnica
Presión Arterial (mmHg)	$\geq 130/85$	$\geq 130/85$
Glucemia (mg/dL)	≥ 110	≥ 100
Triglicéridos (mg/dL)	≥ 150	≥ 150
glsHDL(mg/dL)	< 40 en hombres y < 50 en mujeres	< 40 en hombres y < 50 en mujeres

Tabla 2.1: Factores de [SM](#) según [ATP-III](#) e [IDF](#).

Etnia	Hombres(cm)	Mujeres(cm)
Europeos	94	80
USA	102	88
Asia del sur	90	80
China	90	80
Japón	90	80

Sur y Centro América	Usar recomendaciones de Asia	Usar recomendaciones de Asia
Africa	Usar recomendaciones de Europa	Usar recomendaciones de Europa
Poblaciones Árabes	Usar recomendaciones de Europa	Usar recomendaciones de Europa

Tabla 2.2: Valores de la circunferencia abdominal de acuerdo a la especificidad étnica.

2.3. Insulina, Resistencia a la Insulina

El páncreas secreta dos tipos de células: 1) células exocrinas que producen enzimas que pasan al intestino delgado para acelerar la digestión y 2) células endocrinas que secretan hormonas que pasan a la sangre.

Dos hormonas son esenciales para la regulación del metabolismo de la glucosa, de los lípidos y de las proteínas, éstas son la **insulina** y el **glucagón** [13]. El páncreas está compuesto de dos tejidos, los acinos que secretan jugos digestivos y los islotes de Langerhans. La insulina es secretada por las células beta, éstas representan el 60 % de los islotes, el 25 % de los islotes lo componen las células alfa que secretan glucagón, 10 % de los islotes son las células delta que secretan somastatina, además de otro tipo de célula el polipéptido pancreático que representa el porcentaje restante [13].

Cuando existe glucosa por encima de los niveles normales, el páncreas secreta insulina por medio de las células beta hasta las células de los tejidos para disminuir el nivel de glucosa en la sangre. Lo contrario ocurre cuando los niveles de azúcar están por debajo de lo normal, el páncreas secreta glucagón por medio de las células alfa, se dirigen al hígado, éste libera azúcar al torrente sanguíneo para normalizar el nivel de glucosa en la sangre [2].

En 1922 Banting y Best aislaron por primera vez la insulina del páncreas y el pronóstico de muerte por diabetes dejó de ser algo mortal. "La insulina esta asociada al azúcar en la sangre, de hecho que esta hormona ejerce profundos efectos sobre el metabolismo de los hidratos de carbono. Sin embargo, las causas habituales de muerte de los pacientes diabéticos se deben a alteraciones a nivel del metabolismo lipídico, como es la acidosis y arteriosclerosis. Además de la menor capacidad de síntesis de proteínas de los pacientes con diabetes prolongada generando la atrofia de los tejidos y múltiples alteraciones funcionales"[13].

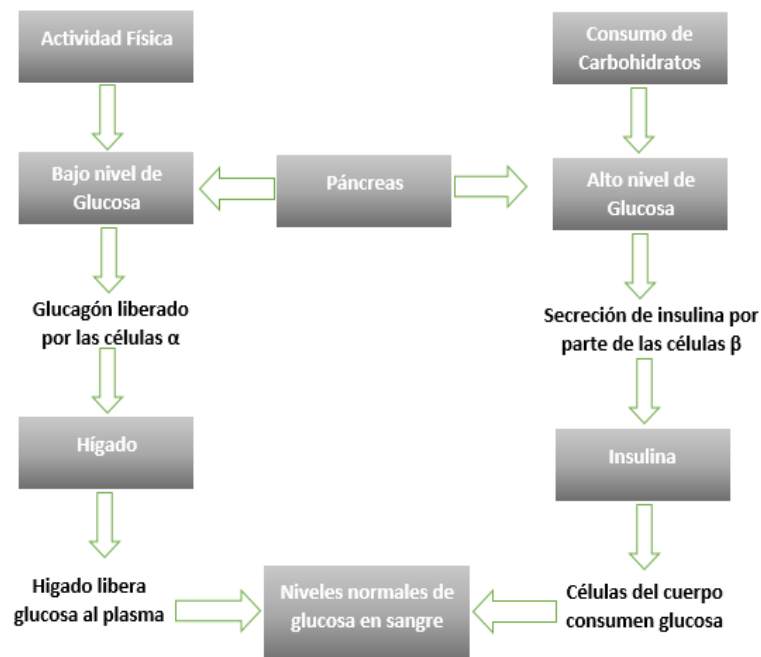


Figura 2.1: Sistema Regulatorio Glucosa-Insulina, Fuente: [2, Fig. 1]

El consumo de alimentos en exceso, particularmente de hidratos de carbono, causa un aumento en la secreción de insulina, por lo tanto, una abundancia energética que se almacena principalmente como glucógeno en el hígado y en los músculos. De la misma manera, el exceso de carbohidratos se almacenan en forma de grasa en el tejido adiposo. En tanto que un exceso de proteínas hace que las células absorban más aminoácidos y las transformen en proteínas interrumpiendo la respuesta de la degradación de las proteínas intracelulares [13].

2.3.1. Metabolismo de la insulina

Después de la digestión de los alimentos, estos se descomponen para crear glucosa. La glucosa pasa a la sangre, donde la insulina permite su ingreso a las células.

Cuando ingerimos mayor cantidad de alimentos, se eleva la concentración de glucosa, en este momento el páncreas estimula un aumento en la secreción de insulina en la sangre, permitiendo la entrada de la glucosa en el interior de las células de los tejidos, en este instante la glucosa es metabolizada, proporcionando energía para mantener las funciones de los órganos y su estructura.

El exceso en la secreción de insulina con el tiempo produce una disminución en la secreción



de la insulina. Este defecto lleva a un exceso en la concentración de la glucosa en la sangre, dañando vasos sanguíneos, los riñones y los nervios y posteriormente a la enfermedad conocida como la Diabetes. No existe cura para esta enfermedad, pero el método de atención es mantener los niveles de la glucosa lo más cercano a los valores normales. Los niveles de glucemia son:

- Hipoglucemia: indica niveles de glucemia debajo de lo normal (55mg/dL o menos).
- Normoglucemia: niveles normales de glicemia (entre 70 y 110mg/dL).
- Hiperglucemia: niveles elevados de glucemia (superior a 110 mg/dL en ayunas).

En personas no diabéticas, los niveles de azúcar de los alimentos se absorben desde que llegan al intestino y pasan a la sangre, elevando los niveles de la misma, que es detectada rápidamente por las células beta. La insulina se trasmite a las células de los músculos, el tejido graso y el hígado disminuyendo el nivel de la glucosa en la sangre. Este proceso se mantiene constante permitiendo que se mantengan la normoglucemia.

En personas con diabetes ocurre lo contrario, la deficiente secreción de insulina como consecuencia produce una debilitación de energía y funciones alteradas del organismo. Entre los cuales se tienen los siguientes síntomas: astenia (sin fuerza o débil), polifagia (sentir hambre), adelgazamiento (el organismo intenta obtener energía a partir de las grasas localizadas en el tejido adiposo), polidipsia (sensación de sed), poliuria (orinar más de lo normal), etc.

2.3.2. Resistencia a la insulina (RI)

La RI existe siempre que las concentraciones normales de insulina produzcan una respuesta biológica menos de lo normal, conlleva a que el páncreas produzca un aumento no controlado de insulina [14].

Una persona con sobrepeso o con obesidad ante este problema tiene valores normales de glucosa por la compensación de altos niveles de insulina en la sangre [14]. Es evidente el aumento proporcional de la RI por los malos hábitos en la comida, por el sobrepeso, la obesidad, y el sedentarismo [15].

Existe relación entre la RI y la Diabetes tipo II cuya causa inicial es la sobreproducción de insulina, que con el tiempo va a alterar el sistema productor de insulina, como consecuencia hay baja y mala producción de insulina, aumentando el diagnóstico de Diabetes tipo II.

2.4. Sensibilidad a la insulina (SI)

La baja SI es la inactividad de reacción de las células del páncreas ante la presencia de la insulina, que conlleva a la hiperglucemia, y causa también la sobreproducción hepática de insulina conocida como hiperinsulinemia [16].

2.4.1. Criterios para el diagnóstico de la SI

El método de referencia, *gold standard*, para medir la SI es el clamp euglicémico hiperinsulinémico, sin embargo, es un método sumamente invasivo, complicado y costoso por lo cual se reserva para fines de investigación clínica.

La práctica más común para estimar la SI es mediante métodos indirectos, los cuales se dividen en dos tipos, aquellos que contemplan los valores de glucosa e insulina durante toda la POTG, entre estos se encuentran el Matsuda y el Avignon, que son más usados pero existen otros. Por otro lado, están los métodos que involucran sólo los datos de glucosa e insulina en la etapa basal, como el Homa-IR, y el Quicki, estos métodos no contemplan el proceso de metabolización de glucosa [4].

El método indirecto más utilizado en clínica es el índice Homa-IR. La mayor limitación de estos métodos es la ausencia o validación de los puntos de corte para el diagnóstico de RI, ya que los puntos de corte han sido determinados en poblaciones específicas. Estos métodos por ejemplo no han sido evaluados en una población de Adultos Mayores.

2.4.2. Cuantificación de la SI a partir de la Prueba Oral de Tolerancia a la Glucosa (POTG)

La POTG son muestras sanguíneas que indican el nivel de glucosa e insulina en un intervalo de tiempo, generalmente estas muestras se toman periódicamente después del consumo oral de una concentración específica de glucosa (generalmente 75 gramos).

Los métodos utilizados para la cuantificación de la SI sirven para establecer de manera aproximada si un sujeto presenta o no RI. En la Tabla 2.1, se describen los métodos para estimar la SI usados en este trabajo [17], [4].

En estudios previos se indica valores de punto de corte para el método de Homa-IR, valor definido en 2.5 [18]. En otro estudio realizado sobre 5518 mayores de 18 años propone como punto de corte para el índice de Quicki, un valor de 0.343 para hombres con SM y de 0.331 en

mujeres con SM [19]. Para los métodos de Avignon y Matsuda no existen puntos de corte de referencia. La Tabla 2.2 presenta los valores de referencia para la glucosa e insulina en niveles basales y postprandiales [20].

Método	Fórmula	Descripción
Homa-IR	$\frac{G_0 * I_0}{405} \quad (2.1)$	Utiliza los valores de glucosa e insulina basal.
Quicki	$\frac{1}{\log G_0 + \log I_0} \quad (2.2)$	Utiliza los valores de glucosa e insulina basal.
Avignon	$\frac{0,137 * Sib + Si2h}{2 * Peso} \quad (2.3)$	Utiliza los valores de glucosa e insulina basal; y, glucosa e insulina a los 120 minutos.
Matsuda	$\frac{10000}{\sqrt{I_0 * G_0 * (G_0 + \frac{G_{120}}{2}) * (I_0 + \frac{I_{120}}{2})}} \quad (2.4)$	Utiliza los valores de glucosa e insulina basal; y, glucosa e insulina a los 120 minutos.
Multidimensional	$G_0, I_0, G_{120}, I_{120}$	Utiliza los valores de los dos puntos de glucosa y dos puntos de insulina.

Tabla 2.3: Métodos para cuantificar la Sensibilidad a la Insulina, (I_0 : Insulina basal, I_{120} : Insulina postprandial, G_0 : Glucosa Basal y G_{120} : Glucosa Postprandial).

Parámetros	Valor Normal
Glucosa Basal G_0	<100 mg/dL
Glucosa postprandial G_{120}	<140 mg/dL
Insulina basal I_0	<20 UI/mL
Insulina postprandial G_{120}	<60 UI/mL

Tabla 2.4: Valores normales de glucosa basal, insulina basal, glucosa postprandial e insulina postprandial.



Capítulo 3

Marco Teórico: clasificación no supervisada

En este capítulo se resumen algunos conceptos de análisis de datos, exploración de datos, clasificación no supervisada, algoritmos empleados en el proyecto de tesis, de forma de utilizar estas herramientas de clasificación no supervisada para clasificar sujetos que tengan o no enfermedades relacionadas al Síndrome Metabólico y la Resistencia a la Insulina.

3.1. Minería de datos

El avance permanente en las Tecnologías de la Información asociadas al internet, específicamente el almacenamiento de datos, han generado gran cantidad de información de diversas índoles, que requieren ser analizadas para descubrir y entender que hay más allá de simples datos mediante técnicas de exploración de datos o minería de datos.

El objetivo de la minería de datos (*Data Mining* en Inglés) es encontrar patrones que sirvan para la solución de un problema. El procedimiento consiste en la obtención de los datos, selección y transformación de los datos, modelamiento de datos, validación del modelo y presentación de resultados.

El modelamiento de los datos llamado extracción del conocimiento como se menciona en [21], es en el que más se centra la comunidad científica gracias a esto se han desarrollado multitud de algoritmos. Siendo los algoritmos de aprendizaje automático los más utilizados.

El aprendizaje automático (*Machine Learning*), es un tipo de Inteligencia Artificial que crea sistemas o algoritmos que pueden aprender a predecir comportamientos. Los sistemas basan su aprendizaje en identificar patrones dentro de un conjunto de datos. Estos algoritmos de aprendizaje automático dividen el conjunto de datos en: 1) datos de entrenamiento para realizar una generalización del modelo y 2) datos de validación para realizar una comprobación del modelo.

Los algoritmos de aprendizaje automático son de tipo descriptivo y predictivo. En el primero se analiza o describe la muestra entre los distintos parámetros que componen la misma, y en el segundo caso se utiliza la muestra de entrenamiento para predecir algo sobre los datos. Las dos clasificaciones generales de aprendizaje automático son: la clasificación supervisada y la clasificación no supervisada.

3.2. Clasificación supervisada

En la clasificación supervisada se busca la clasificación de los datos generalizándolos a una clase o etiqueta. El conjunto de parámetros de entrada son variables independientes y su salida o la clase es su variable dependiente.

El objetivo de esta clasificación es asignar una clase a entradas nuevas o desconocidas. Por este motivo la clasificación supervisada tiene un valor principalmente predictivo. Si los valores a predecir son continuos el procedimiento se conoce como regresión y si los valores son discretos se conoce como clasificación [21].

Existen varios paradigmas de clasificación supervisada como son: Árboles de Decisión, Reglas de Inducción, Métodos bayesianos, Clasificadores basados en vecindad, redes neuronales artificiales, *Support Vector Machines*, Clasificadores múltiples.

3.3. Clasificación No Supervisada

En la clasificación no supervisada, no se establece una clase o etiqueta, los algoritmos tratan de describir de mejor manera el número de clases a las cuales pertenecen los datos de entrada. Esta estructura de clasificación es en la que se centra este trabajo y se hace mención en los algoritmos a utilizar en los siguientes puntos.

3.3.1. Clustering

La técnica más utilizada para clasificación no supervisada es el *Clustering* o Agrupamiento. *Clustering* es el proceso de categorizar datos no etiquetados [22]. Cada elemento de un grupo o clúster generado por el *clustering* debe ser homogéneo y grupos distintos deben ser heterogéneos entre sí [21].

Los elementos homogéneos de un grupo dependen del método de similitud utilizado, como por ejemplo medidas de distancia de una muestra a un clúster, la utilización de conceptos de densidad, entre otras.

Los *clusters* obtenidos dependen del observador ya que los resultados son clústeres no etiquetados obtenidos por el método utilizado. Los datos de 2 o 3 dimensiones se pueden mapear de forma visible, sin embargo, cuando se tienen 4 dimensiones o más, se necesita de algoritmos de *clustering* para agrupar estos datos [23].

El *clustering* se ha utilizado en tres propósitos principales:

- Conocer la estructura de los datos e identificar características principales.
- Identificar el nivel de similitud de diferentes muestras.
- Organizar los datos y resumirlos para mostrar representantes de los clústeres.

Las aplicaciones usadas por medio de *clustering* son numerosas, algunos campos son: biología y estudio genético, procesamiento de imágenes, seguridad informática, psicología, medicina, web *mining*, climatología, procesado de señal, etc. [21].

El *clustering* se divide en dos categorías: **Jerárquicos** y **Particionales**.

En el *clustering* Jerárquico los datos se agrupan en conjuntos más numerosos, pero al final estos se agrupan en un conjunto que abarca cada uno de estos conglomerados. No se especifica la elección del número de clústeres y de parámetros iniciales. Estos algoritmos requieren una medida de disimilaridad entre grupos. Como su nombre lo dice es un método de jerarquías, los niveles inferiores se unen a los niveles superiores y cada uno de estos niveles tiene observaciones hasta el de mayor jerarquía que tiene todas las observaciones.

En el *clustering* basado en particiones los datos se van agrupando en un número de clústeres K fijo. Siendo $1 < K < n$, en el cual se tiene la partición $C = C1, C2, \dots, CK$ de m valores de tamaño distinto o no, en cada grupo se va minimizando la suma de las distancias respecto a su centroide. El objetivo es asignar todas las particiones de n individuos en K grupos.

3.3.2. Algoritmo de agrupamiento de K-medias

El algoritmo de K -medias llamado en inglés *K-means* es un método de aprendizaje no supervisado basado en particiones, utilizado para el agrupamiento de un conjunto de datos en clústeres.

El método consiste en dividir N observaciones (el número de muestras) en K -clústeres, cada muestra pertenece al cluster que esté más cerca, mediante particiones del espacio de datos en celdas Voronoi [24]. El procedimiento es el siguiente:

- Se escoge K (el número de clústeres), Figura 3.1(a)
- Se ingresa aleatoriamente los centros de los K clústeres, Figura 3.1(b).
- Se asigna las muestras al clúster más cercano, mediante la distancia Euclídeana (Figura 3.1(c)).

$$dC_i = \sqrt{\sum_{j=1}^k (C_i^j - x^j)^2} \quad (3.1)$$

C_i : Centroides 1, 2, ..., n

dC_i : Distancia del Centroides C_i : a la entrada x

- Se mueve el centro de cada clúster, calculando la media de sus datos (Figura 3.1(d)).
- Se repiten los pasos 3 y 4 hasta que los centros no cambien.

Se produce una clasificación de datos similares en un clúster y una separación de datos distintos en diferentes clústeres, mediante una métrica de distancia. Aunque este algoritmo siempre converge, la distribución que se alcanza no siempre es la óptima, ya que depende de las condiciones iniciales de los centroides.

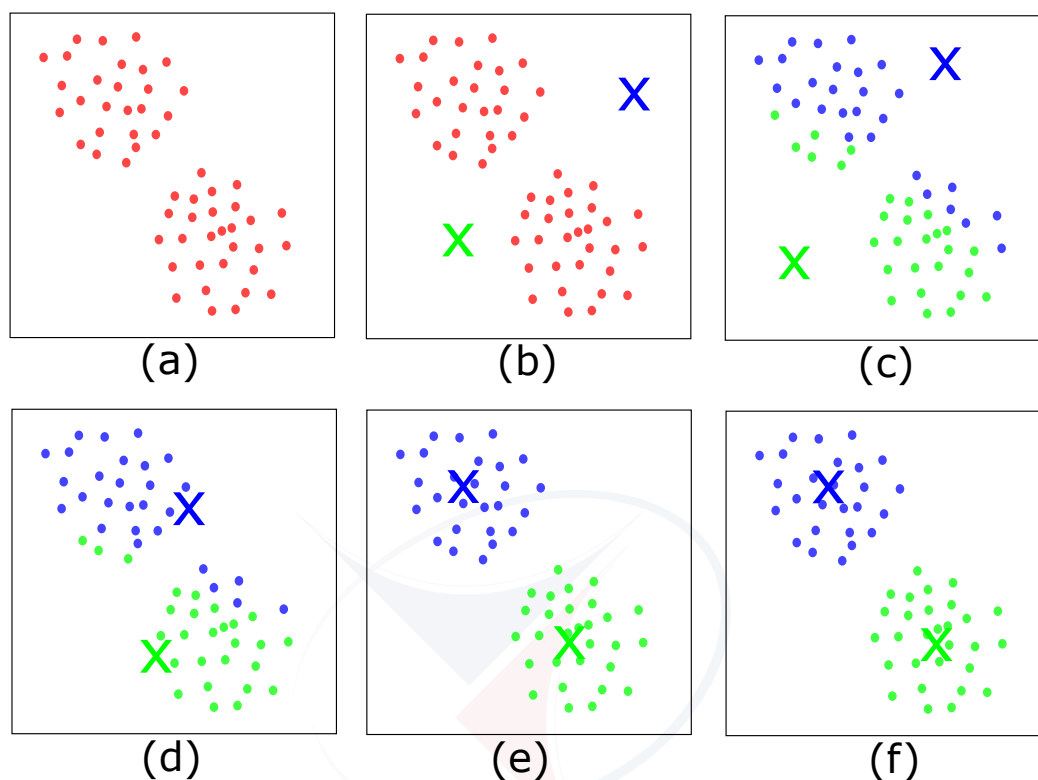


Figura 3.1: Ejemplo del Algoritmo de K -medias. (a) Conjunto de datos original. (b) Se ubican centroides al azar. (c) Se ubica las muestras al centroide del clúster más cercano (color verde o azul). (d) Se mueve el centroide del clúster a la media de los puntos asignados a él. (e) y (f) Agrupamiento luego de 2 iteraciones.

3.3.3. Modelo de Kohonen

A inicios de 1980 Kohonen demostró que un conjunto de datos de entrada puede ordenarse por sí solo, de acuerdo a un modelo de mapas topológicos. Este modelo es llamado **SOM**, y es un tipo de **Redes Neuronales Artificiales (RNA)**.

Una definición clara de **RNA** es la expuesta por *DARPA Neural Network Study (1998, AFCEA International Press, pág.60)*: “Una red neuronal es un sistema compuesto de muchos elementos procesadores simples operando en paralelo, cuya función es determinada por la estructura de la red, la fuerza de las conexiones y el procesamiento realizado por los nodos”.

La red de Kohonen tiene dos variantes en la primera se le suministra el conjunto de datos y la respuesta esperada, esta red es conocida como **Linear Vector Quantization (LVQ)**. La segunda variante son los **SOM**, se le muestra a la red los datos de entrada, sin decirle la salida esperada, este modelo busca establecer una correlación entre los datos de entrada y el espacio de salida

de dos dimensiones (llamado mapa topológico). Esta correlación de entrada-salida se indica mediante la activación de zonas en el mapa de salida.

3.3.4. Representación del modelo

La estructura topológica consta de una capa de entrada y una capa de salida, que consiste en k neuronas de entrada (que es el conjunto de características o vector de parámetros de un dato de entrada). La capa de salida, llamada capa de competencia, se muestra en la Figura 3.2.

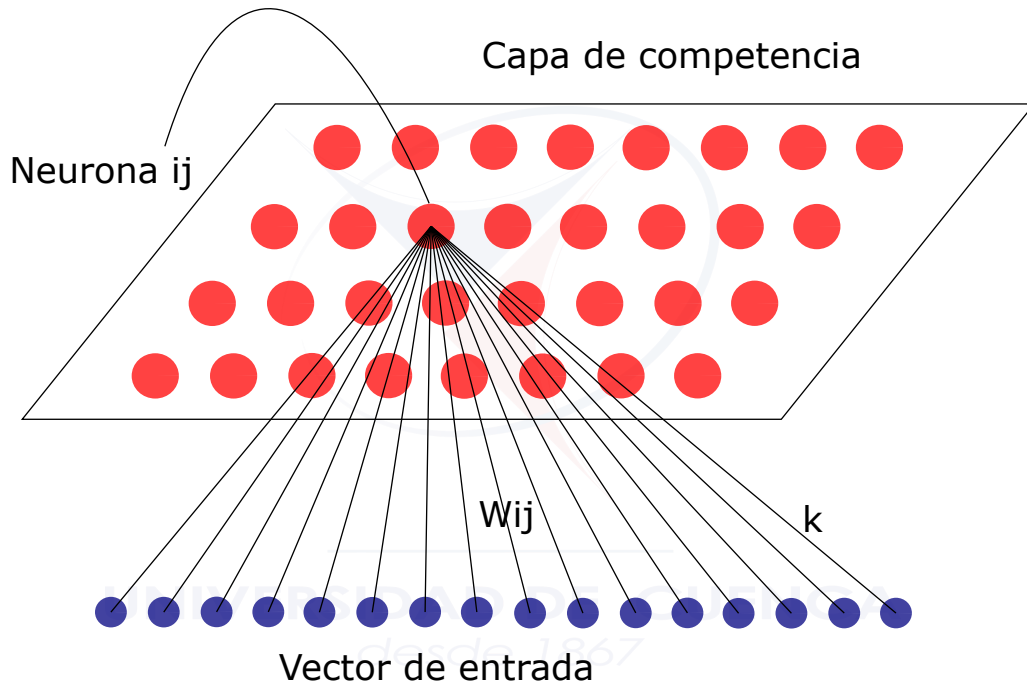


Figura 3.2: Arquitectura del SOM, Fuente: [3, Fig. 4.4]

Cada patrón de entrada está conectado con todas las neuronas de la capa de competencia mediante pesos sinápticos [3]. A cada una de las neuronas de la capa de competencia se le asigna un vector de pesos como se indica en la Figura 3.3.1, de dimensión del vector de entrada.

$$W_{ij} = [W_{ij}^1, W_{ij}^2, \dots, W_{ij}^k] \quad (3.2)$$

La red busca que patrones de entrada similares activen neuronas próximas en la capa de competencia. Esto se realiza cuando los pesos en la capa de competencia asociados a esa entrada son semejantes a ese patrón de entrada. Esta semejanza se consigue con medidas de similitud, en la Tabla 3.1 se indican las más utilizadas:

Distancia Euclídea	$\sqrt{\sum_{s=1}^k (W_{ij}^s - x^s)^2} \quad (3.3)$
Distancia Manhatttan	$\sum_{s=1}^k (W_{ij}^s - x^s) \quad (3.4)$
Distancia Minkowski	$\left[\sum_{s=1}^k (W_{ij}^s - x^s)^\lambda \right]^{\frac{1}{\lambda}} \quad (3.5)$

Tabla 3.1: Métricas para comparar dos vectores.

La neurona vencedora en la capa de competencia será la que tenga mayor grado de similitud con el patrón de entrada. A continuación, se prosigue con la actualización de los pesos de las neuronas de la capa de competencia. Haciendo que la neurona ganadora se aproxime más a la entrada cuando se realice otra iteración, hasta el número de iteraciones determinado. La fórmula de actualización de pesos sináptica es la siguiente [3]:

$$W_{ks}^m = W_{ks}^m + \alpha * (x^m - W_{ks}^m) \quad (3.6)$$

ks : neurona vencedora.

$m = 1, 2, \dots, k$: indica el índice de cada parámetro de un patrón de entrada.

Patrones de entrada similares dan como resultado células ganadoras parecidas, aunque no se garantiza que estén relativamente cerca en la capa de competencia. Por esta razón se incorpora un concepto de interacción lateral entre las neuronas vecinas en la capa de salida para que representen patrones de entrada similares.

Esta interacción lateral se realiza mediante una función de vecindad que determina el grupo de neuronas del vecindario y la intensidad con que éstas deben modificar sus pesos sinápticos.

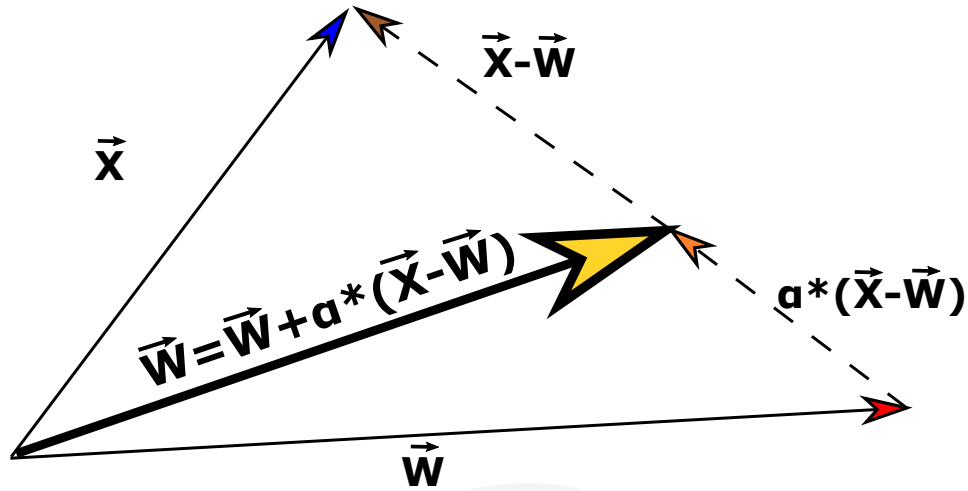


Figura 3.3: Esquema de actualización de pesos para la red SOM , Fuente: [3, Fig. 4.5]

De esta manera la activación de las neuronas se produce también en la proximidad de la neurona vencedora.

3.3.5. Algoritmo de aprendizaje

- 1. Inicializar los pesos W_{ij} con valores aleatorios de 0 a 1, para la entrada: j y salida: i .
- 2. Presentar un vector de entrada $E_k: (e_1, e_2, \dots, e_N)$.
- 3. Determinar la neurona ganadora o **BMU** (mayor similitud del vector de pesos y la entrada), empleando la Distancia Euclídeana u otras funciones.
- 4. Actualizar el vecindario de la **BMU**.

$$r(t) = r_0 * \exp \frac{-t}{\lambda} \quad (3.7)$$

$$\lambda = \frac{\text{numero iteraciones}}{\ln r_0} \quad (3.8)$$

r_0 : radio inicial. λ : constante que permite decrecer el radio cuando aumentan las iteraciones.

- 5. Modificar los pesos de la **BMU** y de las neuronas vecinas.

$$W(t+1) = W(t) + L(t)(E(t) - W(t)) \quad (3.9)$$

$L(t)$: tasa de aprendizaje. Fija la velocidad de cambio de los pesos. Se establece como un

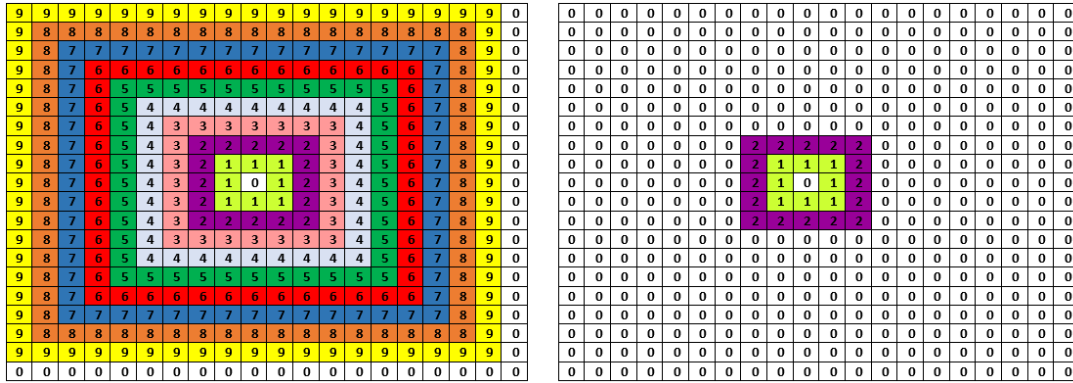


Figura 3.4: Reducción del radio de la BMU

valor constante o en función del número de iteraciones como se indica a continuación:

$$L(t) = l_0 * \exp^{-\frac{t}{\lambda}} \quad (3.10)$$

l_0 : tasa inicial, se ajusta experimentalmente.

Se establece una función $D(t)$ para actualizar más o en mayor valor a las neuronas que estén más cerca a la BMU.

$$W(t+1) = W(t) + D(t)L(t)(E(t) - W(t)) \quad (3.11)$$

Si la función de aproximación es una gaussiana, entonces $D(t)$ es la siguiente:

$$D(t) = \exp^{-\frac{d^2}{2*\sigma(t)^2}} \quad (3.12)$$

- **6.** Volver al paso 2 (presentar todas las entradas), repetir hasta que se cumplan el número de iteraciones ($t=500$, depende de la red).



UNIVERSIDAD DE CUENCA
desde 1867



Capítulo 4

Clasificación no supervisada con el algoritmo de K-medias

En este capítulo se establece la metodología para la estimación de insulino-resistencia en Adultos Mayores con el Agrupamiento de K-medias y se analiza los resultados obtenidos con esta clasificación no supervisada. Adicionalmente se tiene la base de Datos de 119 Adultos Mayores con datos de valores antropométricos y muestras de Glucosa e Insulina necesario para los experimentos de [SI](#).

4.1. Estimación de la RI

4.1.1. Base de Datos para la estimación de la RI

La base de datos está compuesta por 119 Adultos Mayores, no diabéticos, 42 hombres (71 ± 6.692 años) y 77 mujeres (72 ± 6.748 años) reclutados en el proyecto DIUC-PROMETEO denominado: Variabilidad de la frecuencia cardíaca y sensibilidad a la insulina en la población Adulta Mayor de Cuenca, Ecuador [6]. A cada sujeto se le realizó una hematología, un perfil lipídico, medidas antropométricas y POTG de dos puntos. En la POTG se toma una muestra de sangre en ayunas, para medir la glucosa e insulina basal, luego el sujeto ingiere 75 gramos de glucosa y después de dos horas se toma una nueva muestra de sangre para medir la glucosa e insulina postprandial. La Tabla 4.1 presenta las medidas antropométricas y los valores de la POTG de la población de adultos mayores.

Características	Media \pm std	Rango
Peso (<i>kg</i>)	68.03 ± 10.51	[45 – 90]
Altura (<i>m</i>)	1.53 ± 0.098	[1.02 – 1.76]
Circunferencia Abdominal (<i>cm</i>)	93.74 ± 11.14	[69 – 130]
Insulina Basal (μ UI/ <i>mL</i>)	10.67 ± 8.08	[2.00 – 43.60]
Glucosa Basal (<i>mg/dl</i>)	90.51 ± 25.17	[61.30 – 241.10]
Insulina Postprandial (μ UI/ <i>mL</i>)	58.11 ± 44.66	[4.30 – 285.00]
Glucosa Postprandial (<i>mg/dl</i>)	102.31 ± 32.20	[31.10 – 221.60]

Tabla 4.1: Medidas antropométricas y valores de la POTG, calculadas con la media \pm desviación estándar (std) y el rango de cada parámetro.

4.1.2. Algoritmo de K-medias y Experimentos

La RI es una condición prediabética que requiere ser diagnosticada de forma temprana. Sin embargo, no existe consenso en cuanto al método de estimación según la población estudiada. En este trabajo se analizan 5 métodos de estimación de RI usando una clasificación de K-medias con $K=2$ y $K=3$ sobre una población de 119 Adultos Mayores de la ciudad de Cuenca quienes se sometieron a POTG de dos puntos. El algoritmo de K-medias, con $K=2$ y $K=3$, fue aplicado en experimentos unidimensionales para los métodos Homa-IR, Quicki, Avignon, Matsuda y se realizó un experimento multidimensional usando las cuatro observaciones de la POTG (valores de glucosa e insulina basal y postprandial).

4.1.3. Metodología con $K=2$ y $K=3$

Se diseñaron y realizaron los siguientes experimentos:

- **Experimentos unidimensionales:** correspondiente a los cuatro métodos estudiados: Homa-IR, Quicki, Avignon y Matsuda. Cada método fue aplicado a las 119 observaciones, con dimensión $N=1$ y agrupación $K=2$ y $K=3$.
- **Experimento multidimensional:** usando el algoritmo de K -medias con las observaciones correspondiente a $N=4$ dimensiones, correspondientes a los valores de glucosa e insulina basales (G_0 , G_{120}) y postprandiales (I_0 , I_{120}). Estas variables fueron normalizadas para tener media=0 y varianza=1.

Para cada experiencia, el algoritmo fue ejecutado 10 veces usando diferentes inicializaciones para evitar los mínimos locales. Luego, se selecciona el cuadrado de las distancias de cada punto al centroide dentro del grupo (distancia Euclídea) y se elige la menor. El máximo de iteraciones para el algoritmo de K -medias se estableció en 100. El test de Wilcoxon se utilizó para determinar la diferencia entre grupos. Un valor de $p < 0.01$ se consideró estadísticamente significativo.

- **Análisis con $K=2$**

En la Figura 4.1 se observan las asignaciones de cada uno de los experimentos unidimensionales a los grupos C-N y C-RI y los valores de corte entre los clústeres. La población se dividió en dos clústeres: C-N, con valores normales y C-RI con RI. Se validaron los puntos de corte clínicos para Homa-IR y Quicki. La cantidad de individuos clasificados en cada C-RI fue muy diferente según cada método. Los métodos que usan valores basales y postprandiales (Avignon y Matsuda) identificaron una mayor cantidad de población con RI y este grupo correspondió a los Adultos Mayores con una circunferencia abdominal elevada. El método multidimensional parece aportar menos información que los métodos unidimensionales que incluyen valores postprandiales, como se va a observar en los resultados obtenidos en el siguiente capítulo. La pluralidad de los clústeres obtenidos según el método de estimación de RI, revelan que es necesario analizar detenidamente estos resultados desde el punto de vista clínico para establecer estrategias de diagnóstico precoz de RI en la población de adultos mayores cuencana.

- **Análisis con $K=3$**

En la Figura 4.2 se observan las asignaciones de cada uno de los experimentos unidimensionales a los grupos C-N, C-RI y CA. El Algoritmo de K -medias separó la población en tres grupos bien diferenciados de acuerdo con los altos valores de glucosa e insulina utilizando métodos de estimación de SI. Lo interesante es que para cada método el número de individuos que podrían ser diagnosticados como resistentes a la insulina es muy diferente.

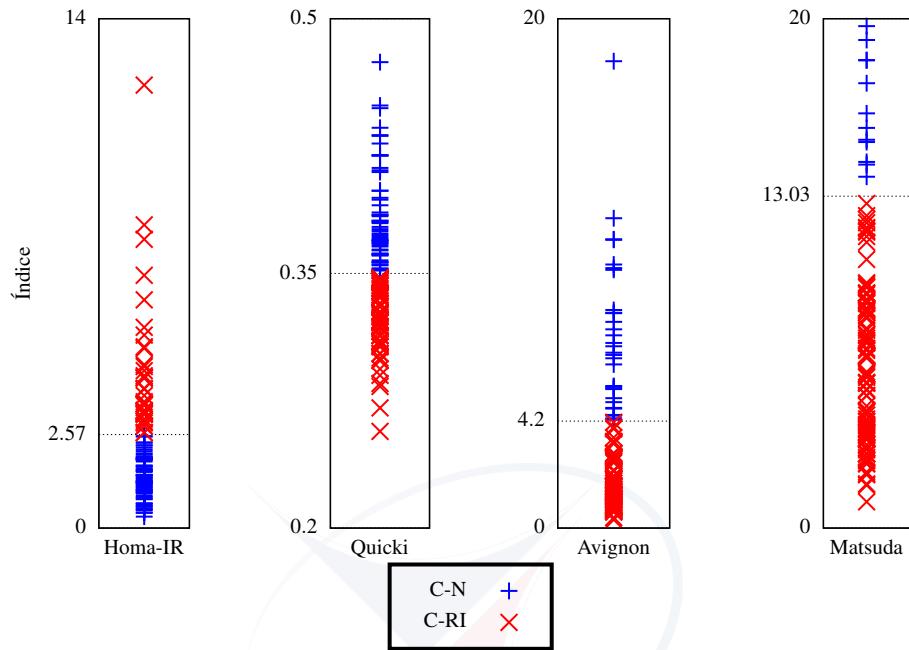


Figura 4.1: Distribución de clústeres C-N y C-RI y puntos de corte para cada experimento unidimensional.

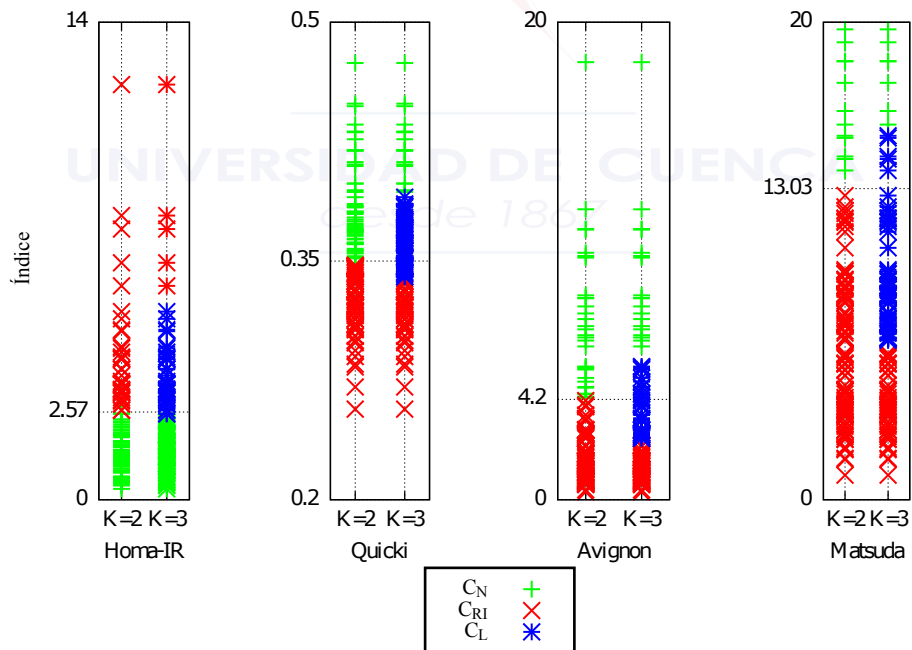


Figura 4.2: Distribución de clústeres C-N, C-RI y C-L y puntos de corte para cada experimento unidimensional.

4.2. Resultados de la estimación de la Resistencia a la Insulina con $K=2$

La clasificación con el algoritmo de K -medias dividió a la base de datos de 119 sujetos en $K=2$ grupos, los cuales denominaremos $C-N$, para el clúster con valores dentro de lo considerado normal y $C-RI$ para el clúster susceptible de presentar características de **RI**, de acuerdo a los puntos de corte obtenidos para cada método como se indica en la Tabla 4.2. En la Tabla 4.3 se presentan las características de cada *cluster* según sus valores de glucosa e insulina basales (G_0 , I_0) y postprandiales (G_{120} , I_{120}) y cintura abdominal. El test de Wilcoxon se usó para determinar la diferencia entre grupos. Se consideró un valor de $p \leq 0.01$ como significativo.

Método	C-N	C-RI	Punto de corte	p
Homa-IR	80 sujetos	39 sujetos	≤ 2.57	1.056e-18
Quicki	59 sujetos	60 sujetos	≥ 0.35	5.213e-21
Matsuda	24 sujetos	95 sujetos	≥ 13.03	4.470e-14
Avignon	31 sujetos	88 sujetos	≥ 4.20	1.515e-16
Multidimensional	80 sujetos	39 sujetos	–	4.2559e-13

Tabla 4.2: Número de Adultos Mayores por clústeres ($C-N$ y $C-RI$) y puntos de corte para Homa-IR, Quicki, Matsuda y Avignon.

Método	Valores	$C-N$ media \pm std	$C-RI$ media \pm std	p
Homa-IR	G_0	84.86 \pm 15.95	102.12 \pm 35.11	0.001
	I_0	6.62 \pm 2.67	18.97 \pm 9.08	0.000
	G_{120}	96.31 \pm 29.37	114.62 \pm 29.37	0.005
	I_{120}	40.63 \pm 24.20	93.96 \pm 24.20	0.000
	CA	92.88 \pm 11.52	95.49 \pm 10.23	N.S
Quicki	G_0	83.16 \pm 15.30	97.75 \pm 30.49	0.001
	I_0	5.51 \pm 2.67	15.74 \pm 8.60	0.000
	G_{120}	96.47 \pm 28.27	108.05 \pm 34.93	0.027
	I_{120}	36.35 \pm 21.46	79.50 \pm 50.96	0.000
	CA	94.72 \pm 11.11	92.74 \pm 11.18	N.S
Matsuda	G_0	79.53 \pm 11.89	93.29 \pm 26.87	0.004
	I_0	3.99 \pm 1.66	12.35 \pm 8.18	0.000
	G_{120}	96.47 \pm 28.27	108.05 \pm 34.93	0.027
	I_{120}	36.35 \pm 21.46	79.50 \pm 50.96	0.000
	CA	88.19 \pm 6.50	95.14 \pm 11.65	0.003

Avignon	G_0	80.55 ± 12.17	94.03 ± 27.58	0.006
	I_0	6.27 ± 7.31	12.22 ± 7.80	0.000
	G_{120}	79.39 ± 23.02	110.39 ± 31.15	0.000
	I_{120}	18.78 ± 8.87	71.96 ± 43.96	0.000
	CA	88.05 ± 7.89	95.74 ± 11.46	0.000
Multidimensional	G_0	85.16 ± 16.26	100.70 ± 34.61	0.012
	I_0	8.21 ± 6.11	15.34 ± 9.30	0.000
	G_{120}	88.58 ± 21.85	128.44 ± 32.73	0.000
	I_{120}	34.90 ± 18.32	102.25 ± 46.79	0.000
	CA	95.36 ± 9.76	93.11 ± 11.62	0.000

Tabla 4.3: Valores de Glucosa (mg/dL), Insulina ($\mu UI/mL$) y Circunferencia Abdominal (cm) para cada clúster ($C-N$ y $C-RI$) según cada método.

En la Tabla 4.4, se presenta una síntesis de las características de los clústeres, según el siguiente criterio N indica un valor normal, A un valor alterado y E un valor elevado, según las bandas de criterio clínico.

	clúster	G_0	I_0	G_{120}	I_{120}
Homa-IR	$C-N$	N	N	N	N
	$C-RI$	A	E	N	A
Quicki	$C-N$	N	N	N	N
	$C-RI$	E	E	N	A
Matsuda	$C-N$	N	N	N	N
	$C-RI$	N	E	N	A
Avignon	$C-N$	N	N	N	N
	$C-RI$	N	E	N	A
Multidimensional	$C-N$	N	N	N	N
	$C-RI$	A	E	E	A

Tabla 4.4: Características obtenidas para los métodos de [SI](#). (I_0 : Insulina basal, I_{120} : Insulina postprandial, G_0 : Glucosa Basal y G_{120} : Glucosa Postprandial), Normal (N), Elevado (E) y alterado (A).

Los puntos de corte obtenidos con el Homa-IR y el Quicki son bastante cercanos a los reportados por otros estudios[18],[19]. Al analizar las poblaciones según los valores promedios

de glucosa e insulina se desprende de la Tabla 4.3 lo siguiente:

1. El Homa-IR y el Quicki diferencian los clústeres según los valores de G_0 , I_0 alterados y elevados respectivamente, y el $C\text{-}RI$ también presenta I_{120} alterado, aunque este valor no es considerado por el método. Sin embargo, los valores aumentados de I_{120} son de esperarse si los basales son alterados [7], esto podría indicar que el sujeto lleva algún tiempo desarrollando RI; por tanto, tiene continuamente valores alterados de insulina. Del mismo modo, el Quicki diferencia los clústeres según los valores elevados de G_0 , I_0 y el $C\text{-}RI$, también presenta I_{120} alterado, aunque este valor no es considerado en el método.
2. Matsuda y Avignon diferencian los clústeres por sus valores de I_{120} alterados y valores de I_0 elevados, aunque también se observan que todos los valores de la POTG son superiores para el clúster $C\text{-}RI$. Pareciera que el agrupamiento para estos métodos incluye como $C\text{-}RI$ a aquellos individuos que presentan valores alterados de I_{120} , estos individuos no necesariamente presentan valores basales alterados, y en los métodos anteriores (Quicki y Homa-IR) no son detectados como $C\text{-}RI$. Los métodos que toman en cuenta los valores de I_{120} e G_{120} pudieran ser más efectivos detectando grupos con riesgo a padecer de RI.
3. Usando 4 observaciones se evidencia que este método está separando los clústeres en base a G_0 y I_{120} alterados; adicionalmente, éste es el único método en donde para los clústeres en base a I_0 y G_{120} presenta valores elevados.

El K -medias permite separar en dos clústeres; según los valores elevados de glucosa e insulina usando métodos de estimación de SI, o directamente las variables medidas de la POTG. Lo interesante es que para cada método la cantidad de individuos que podrían diagnosticarse como insulino-resistentes o con RI es diferente.

Los métodos que incluyen todos los valores de glucosa e insulina ubican un mayor número de individuos en el grupo con valores alterados o valores más altos ($C\text{-}RI$). Esto pudiera deberse a que estos métodos al incluir los valores basales y postprandiales, detectan aquellos casos en donde los valores basales se encuentran normales, pero tienen valores postprandiales alterados. Estos casos en particular, son personas que tienen una disfunción metabólica que todavía no se ve reflejada en los valores basales, hay estudios que indican que los valores basales son los últimos en alterarse cuando se presenta la resistencia a la insulina [7].

Adicionalmente, para estos dos métodos se observa que los individuos con $C\text{-}RI$ presentan circunferencias abdominales mayores que los individuos del $C\text{-}N$, lo cual no se observa con los métodos que solo toman en cuenta los valores basales, ni con el método multidimensional. Lo cual puede indicar que los métodos de SI que usan los valores de glucosa e insulina postprandial podrían detectar valores alterados de circunferencia abdominal; por tanto, aportan información fisiológica adicional que no se observa en la clasificación realizada por los métodos que solo

toman en cuenta valores basales de glucosa e insulina.

4.3. Resultados de la estimación de la Resistencia a la Insulina con $K=3$

La clasificación con el algoritmo de K -medias dividió a la base de datos de 119 sujetos en $K=3$ grupos, los cuales denominaremos $C-N$, para el clúster con valores dentro de lo considerado normal, $C-L$ para el clúster con valores límite entre deteriorado y normal, y $C-RI$ para el clúster susceptible de presentar características de [RI](#) de acuerdo a los puntos de corte obtenidos para cada método. En la tabla 4.5, se presentan los puntos de corte obtenidos de cada método para $K=3$.

Método	C-N	C-L	C-RI	Punto de corte 1	Punto de corte 2	$p12$	$p13$	$p23$
Homa-IR	78	35	6	≤ 2.35	≤ 5.51	2.4E-17	5E-05	1.2E-04
Quicki	17	56	46	> 0.39	≥ 0.34	5.4E-10	1.5E-09	4.7E-18
Matsuda	19	51	49	≥ 15.25	≥ 6.14	1.6E-10	2.1E-10	7.2E-18
Avignon	19	33	67	≥ 5.28	≥ 2.27	2.7E-09	3.6E-11	5.5E-16
Multi	65	25	29	–	–	9.3E-09	1.8E-12	3.8E-04

Tabla 4.5: Número de Adultos Mayores por clústeres ($C-N$, $C-L$ y $C-RI$) y puntos de corte para Homa-IR, Quicki, Matsuda y Avignon.

Método	Valor	$C-N$ media \pm std	$C-L$ media \pm std	$C-RI$ media \pm std	$p12$	$p123$	$p23$
Homa-IR	G_0	84.63 \pm 16.09	98.70 \pm 31.46	119.23 \pm 47.27	0.001	0.080	0.440
	I_0	6.51 \pm 2.61	15.77 \pm 5.32	34.95 \pm 9.07	0.000	0.000	0.000
	G_{120}	95.96 \pm 29.56	115.54 \pm 34.31	107.68 \pm 33.37	0.003	0.460	0.520
	I_{120}	38.97 \pm 21.59	96.51 \pm 50.63	82.93 \pm 75.37	0.000	0.140	0.340
	CA	92.42 \pm 11.20	96.49 \pm 11.23	94.83 \pm 7.65	0.023	0.380	0.530
	HG	7.48 \pm 9.79	5.43 \pm 0.88	5.93 \pm 0.59	0.240	0.450	0.170
	IMC	28.55 \pm 4.48	30.60 \pm 6.72	28.52 \pm 4.25	0.110	0.830	0.250
	G_0	81.06 \pm 10.60	85.13 \pm 17.84	100.56 \pm 32.55	0.700	0.004	0.000
	I_0	3.02 \pm 0.60	7.24 \pm 1.96	17.67 \pm 8.91	0.000	0.000	0.000
	G_{120}	93.89 \pm 25.24	97.54 \pm 30.84	111.23 \pm 34.54	0.540	0.030	0.030

Quicki	I_{120}	23.54 ± 13.18	43.22 ± 22.05	89.00 ± 53.58	0.000	0.000	0.000
	CA	87.82 ± 6.04	93.72 ± 12.41	95.93 ± 10.30	0.070	0.002	0.150
	HG	9.33 ± 14.37	7.13 ± 8.52	5.47 ± 0.83	0.310	0.080	0.440
	IMC	26.92 ± 4.43	29.24 ± 4.32	29.88 ± 6.37	0.040	0.030	0.800
Matsuda	G_0	78.06 ± 12.35	87.10 ± 32.41	98.89 ± 32.41	0.030	0.000	0.040
	I_0	3.86 ± 1.69	7.16 ± 2.35	16.96 ± 9.04	0.000	0.000	0.000
	G_{120}	79.00 ± 18.84	95.81 ± 26.72	118.12 ± 34.00	0.007	0.000	0.000
	I_{120}	16.35 ± 7.46	40.13 ± 16.69	93.01 ± 48.10	0.000	0.000	0.000
	CA	89.03 ± 6.17	91.82 ± 12.00	97.55 ± 10.64	0.490	0.000	0.008
	HG	11.68 ± 18.63	6.22 ± 3.70	5.52 ± 0.88	0.870	0.710	0.350
	IMC	27.02 ± 3.98	28.97 ± 4.78	30.17 ± 5.97	0.090	0.010	0.410
Avignon	G_0	78.73 ± 11.87	89.35 ± 20.44	94.43 ± 28.91	0.040	0.005	0.550
	I_0	7.16 ± 9.06	6.64 ± 4.82	13.64 ± 7.90	0.210	0.000	0.000
	G_{120}	73.38 ± 19.34	91.06 ± 23.42	116.06 ± 31.45	0.004	0.000	0.000
	I_{120}	14.24 ± 6.92	32.78 ± 11.24	83.02 ± 44.57	0.000	0.000	0.000
	CA	87.97 ± 7.27	88.82 ± 10.94	97.79 ± 10.54	0.990	0.000	0.000
	HG	11.81 ± 18.59	6.35 ± 4.56	5.60 ± 0.94	0.510	0.510	0.940
	IMC	28.97 ± 3.85	27.50 ± 3.80	30.93 ± 5.54	0.070	0.000	0.001
Multi	G_0	84.28 ± 17.43	97.52 ± 33.30	98.45 ± 28.73	0.003	0.005	0.850
	I_0	7.17 ± 5.39	8.69 ± 2.82	20.20 ± 8.70	0.005	0.000	0.000
	G_{120}	83.45 ± 18.53	145.23 ± 25.28	107.60 ± 23.15	0.000	0.000	0.000
	I_{120}	32.29 ± 18.51	66.13 ± 20.08	109.06 ± 54.61	0.000	0.000	0.000
	CA	91.22 ± 11.38	98.24 ± 10.68	95.48 ± 9.65	0.006	0.030	0.640
	HG	7.83 ± 10.70	5.62 ± 0.97	5.51 ± 0.90	0.660	0.480	0.830
	IMC	28.55 ± 4.80	29.17 ± 3.08	30.48 ± 7.31	0.380	0.270	0.760

Tabla 4.6: Valores de Glucosa (mg/dL), Insulina ($\mu UI/mL$), Circunferencia Abdominal (cm) Hemoglobina Glicosilada (mg/dL) para cada clúster ($C-N$, $C-L$ y $C-RI$) según cada método.

Se analizan los grupos de acuerdo con los valores medios de glucosa (G_0 , G_{120}) e insulina (I_0 , I_{120}) de la tabla 4.7, se observa lo siguiente:

	clúster	núm.	G_0	I_0	G_{120}	I_{120}
Homa-IR	$C-N$	78	N	N	N	N
	$C-L$	35	E	E	N	A
	$C-RI$	6	A	A	N	A
Quicki	$C-N$	17	N	N	N	N
	$C-L$	56	N	N	N	N

	<i>C-RI</i>	46	A	E	N	A
Matsuda	<i>C-N</i>	19	N	N	N	N
	<i>C-L</i>	51	N	N	N	N
	<i>C-RI</i>	49	E	E	N	A
Avignon	<i>C-N</i>	19	N	N	N	N
	<i>C-L</i>	33	N	N	N	N
	<i>C-RI</i>	67	E	E	N	A
Multidimensional	<i>C-N</i>	65	N	N	N	N
	<i>C-L</i>	25	E	N	A	A
	<i>C-RI</i>	29	E	A	N	A

Tabla 4.7: Características obtenidas para los métodos de [SL](#). (I_0 : Insulina basal, I_{120} : Insulina postprandial, G_0 : Glucosa Basal y G_{120} : Glucosa Postprandial), Normal (N), Elevado (E) y alterado (A).

Analizando los grupos de acuerdo con los valores medios de glucosa (G_0 , G_{120}) e insulina (I_0 , I_{120}) de la Tabla 4.6 se observa lo siguiente:

- **Homa-IR:** En este método, los valores promedio de $C-N$ están dentro de los rangos normales, para $C-L$ los valores promedio de G_{120} y I_{120} están alterados y en el grupo $C-RI$, todos los valores están alterados, pero este clúster solo tiene 6 sujetos. También se observa que los valores de CA e IMC están alterados en el clúster $C-L$. Con respecto a la clasificación con $K=2$, se puede considerar que el clúster $C-L$ está más cerca de los sujetos con [RI](#) que los sujetos normales.
- **Quicki:** Los valores promedio de $C-N$ y $C-L$ están dentro de los rangos normales, aunque los valores de $C-L$ son más altos. Para el clúster $C-RI$, todos los valores están alterados excepto para I_0 . El grupo $C-RI$ tiene valores diferentes significativamente para glucosa (G_0 y G_{120}) y para insulina (I_0 y I_{120}) con respecto a los grupos $C-N$ y $C-L$: aunque los valores postprandiales no están considerados para la estimación de este método. En este caso el grupo $C-L$ puede ser considerado con un grupo de riesgo de $C-RI$. La CA del grupo $C-RI$ es significativamente más grande que el grupo $C-N$.
- **Avignon:** Los valores promedio de $C-N$ y $C-L$ están dentro de los rangos normales, aunque los valores de $C-L$ son más altos. Para el clúster $C-RI$, G_{120} y I_{120} están alterados. Los valores postprandiales (G_0 y I_{120}) permiten separar significativamente los tres grupos. Adicionalmente, el grupo $C-RI$ tiene valores de CA y IMC más grandes que los grupos $C-N$ y $C-L$, en cuanto al caso anterior, el grupo $C-L$ puede ser considerado como grupo

de riesgo de padecer [RI](#).

- **Matsuda:** Como con el método Avignon, los valores promedios de $C-N$ y $C-L$ están dentro de los rangos normales, aunque los valores de $C-L$ son más altos. Para el clúster $C-RI$, solo el promedio de I_{120} está alterado. La insulina basal I_0 y los valores postprandiales (I_{120} y G_{120}) permiten separar significativamente los tres grupos. La CA del grupo $C-RI$ es significativamente más alta con respecto a los otros dos clústers. Al igual que los métodos Quicki y Avignon, el grupo $C-L$ puede considerarse como un grupo de riesgo de [RI](#).

Al igual que con $K=2$, el K -medias para $K=3$, permite separar en tres *clusters*; según los valores elevados de glucosa e insulina usando métodos de estimación de [SI](#), o directamente las variables medidas de la [POTG](#). Lo interesante es que para cada método la cantidad de individuos que podrían diagnosticarse como insulino-resistentes o con [RI](#) es diferente.

Aunque en todos los casos el grupo $C-L$ parece un grupo que se arriesga a presentar características de sufrimiento de [RI](#) en el caso del método Homa-IR la distribución de los sujetos en cada grupo es bastante diferente del resto de los métodos. En particular, los métodos Homa-IR clasifican el mayor número (65,54 %) de sujetos en el grupo $C-N$ a diferencia de los otros métodos en los que el número de individuos en esta clase no excede el 16 %. También en relación con $K=2$, vemos que la distribución sigue siendo prácticamente la misma, para $C-N$ hay 80 sujetos y 78 con $K=3$; y sólo hay 6 sujetos en el $C-RI$ (Tabla 4.7).

Para los otros tres métodos el *cluster* $C-L$ está compuesto de sujetos que fueron clasificados como $C-N$ y $C-RI$ usando $K=2$.

Matsuda y Avignon diferencian los tres clústeres por sus valores postprandiales (I_{120} y G_{120}), aunque esto es también observado que todos los valores de la [POTG](#) son más altos para el clúster $C-RI$. Parece que el agrupamiento para estos métodos incluye como $C-RI$ aquellos individuos que tienen valores alterados de I_{120} , estos individuos no tienen necesariamente valores de ayuno alterados, y en métodos anteriores (Quicki y Homa-IR) no se detectan como ($C-RI$). Los métodos que toman en cuenta los valores de I_{120} y G_{120} podrían ser más efectivos detectando grupos con riesgo de sufrir de [RI](#).

Métodos que incluyen todos los valores de glucosa e insulina colocan un mayor número de individuos en el grupo con valores alterados ($C-RI$). Esto puede deberse al hecho de que estos

métodos, al incluir los valores basales y postprandiales, detectan aquellos casos en los que los valores basales son normales, pero tienen valores postprandiales alterados. Estos casos particulares son personas que tienen una disfunción metabólica que aún no se refleja en los valores de ayuno, hay estudios que indican que los valores de ayuno son los últimos en ser alterados cuando la [RI](#) está presente.

4.4. Diagnóstico con nuevos pacientes

Se ingresaron nuevos pacientes de una población de 319 sujetos Adultos Mayores de 65 años.

Características	Media	Rango
Peso (kg)	58.34	[29.90 - 108.5]
Altura (m)	1.55	[1.35 - 1.79]
Circunferencia Abdominal (cm)	85.92	[56 - 120]
Insulina Basal ($\mu UI/mL$)	7.15	[3.42 - 12.41]
Glucosa Basal (mg/dl)	91.31	[26 - 411]
Insulina Postprandial ($\mu UI/mL$)	40.48	[24.65 - 59.39]
Glucosa Postprandial (mg/dl)	101.94	[71.04 - 525.97]

Tabla 4.8: Medidas antropométricas y valores de la [POTG](#) de otra base de datos de Adultos Mayores.

En la Figura 4.3, se muestra las características de un paciente como son: glucosa e insulina (G_0 , G_{120} , I_0 y I_{120}), peso, circunferencia abdominal e índice de masa corporal. Los resultados para $K=2$, indican que el sujeto tiene valores de Homa-IR normal, pero para los otros experimentos Quicki, Avignon y Matsuda están en presencia de [RI](#).

Adicionalmente en $K=3$, el sujeto está en el grupo de $C-L$ (límite entre Normal y con [RI](#)) y para los demás métodos el sujeto tiene [RI](#). También él, presenta G_0 alterada y valores normales de G_{120} , I_0 y I_{120} esto indica que la G_0 es representativo de la pre-diabetes. Por lo tanto, además de los valores numéricos es necesario incorporar el análisis clínico de cada método para realizar el diagnóstico de [RI](#).

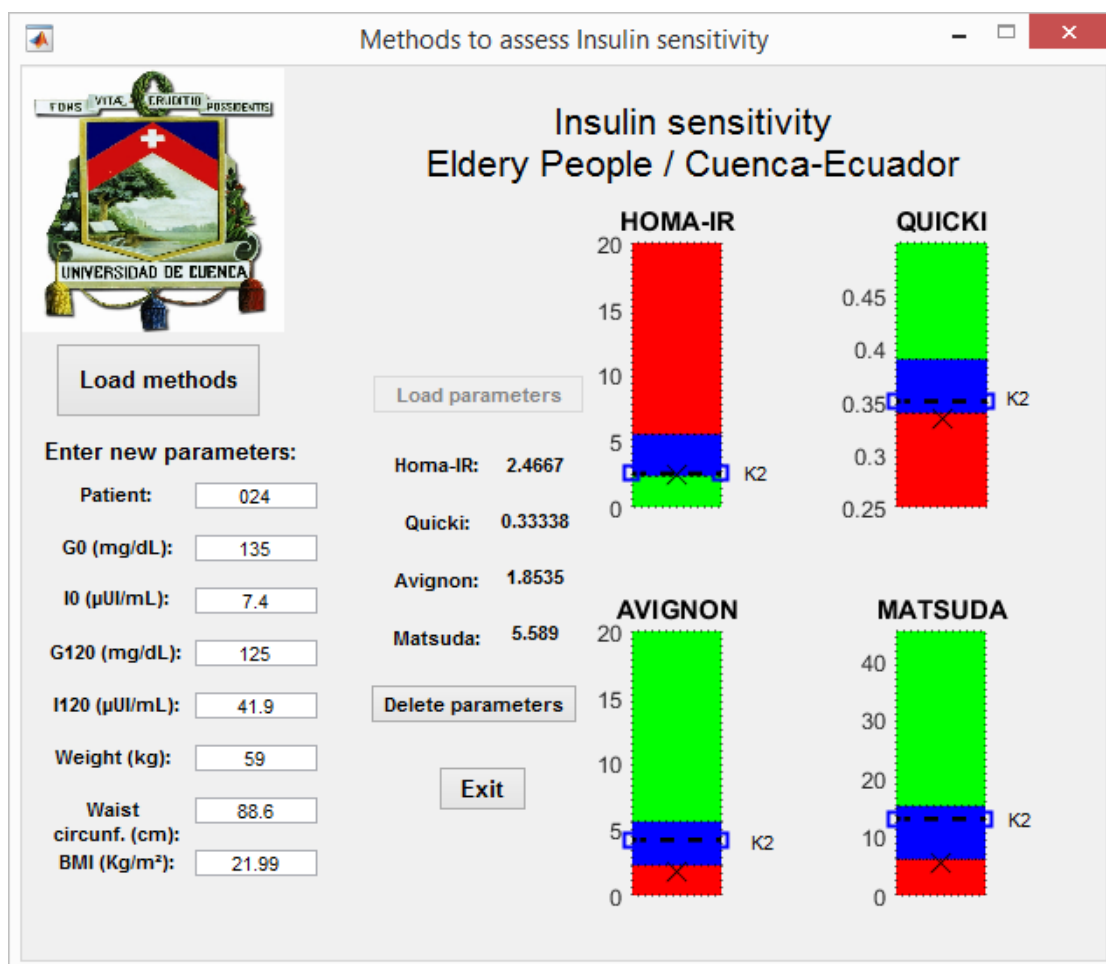


Figura 4.3: Sistema de Diagnóstico. Valores de Homa-IR, Quicki, Avignon y Matsuda se indican para un nuevo paciente representados por el carácter X. El color verde representa valores normales (C-N), el color azul representa el límite entre un sujeto sano y uno con RI (C-L) y el color rojo representa un sujeto con RI (C-RI).



4.5. Conclusiones

El método de K -medias ha permitido estudiar cinco métodos para estimar la **SI** en una población adulta mayor. Los agrupamientos obtenidos usando los métodos empleados en clínica como el Homa-IR y el Quicki, presentan valores de corte similares a otros reportados en la literatura. Lo cual constituye un hallazgo importante para la evaluación de la población adulta mayor de Cuenca, ya que valida estos límites.

Los métodos que usan valores basales y postprandiales identifican una mayor cantidad de población con **RI** y este grupo (C -**RI**) corresponde a los adultos mayores con una circunferencia abdominal elevada, lo cual describe una dislipidemia frecuente en sujetos con **RI** [25]; y, por tanto, una población con riesgo a desarrollar enfermedades cardiovasculares.

El método multidimensional, aunque separa dos clústeres bien definidos, parece aportar menos información que los métodos que incluyen información postprandial.

Las diferencias de diagnóstico obtenidas según el método para la estimación de la **SI** usado, indican que es necesario explorar con más profundidad desde el punto de vista clínico, la relación de la **SI** con otros parámetros de laboratorio, como por ejemplo con la hemoglobina glucosilada. Todo esto para mejorar el diagnóstico de adultos mayores de nuestra región sensibles de padecer **RI**.



Capítulo 5

Clasificación no supervisada mediante Redes Neuronales de Kohonen

UNIVERSIDAD DE CUENCA
desde 1867

En este capítulo se da a conocer la metodología para la estimación del Síndrome Metabólico de Adultos Mayores mediante Redes Neuronales de Kohonen utilizando el software Viscovery SOMine. Adicionalmente se tiene una base de datos de 387 Adultos Mayores para la estimación de Síndrome Metabólico.

5.1. Estimación del SM

5.1.1. Base de Datos para la estimación del SM

La base de datos está compuesta por 387 adultos mayores, no diabéticos, 142 hombres (72.32 ± 7.26 años) y 245 mujeres (73.11 ± 7.64 años) reclutados en el proyecto DIUC-PROMETEO denominado: Variabilidad de la frecuencia cardíaca y sensibilidad a la insulina en la población Adulta Mayor de Cuenca, Ecuador [6]. Para cada sujeto se utilizan como parámetros de entrada las seis variables usadas para diagnosticar SM, la presión arterial se consideró como una variable doble. Las Tablas 5.1 y 5.2 presentan ésta población de adultos mayores clasificados por género.

Características	Media \pm std	Rango
Presión Sistólica (<i>mmHg</i>)	128.20 ± 14.27	[94 – 184]
Presión Diastólica (<i>mmHg</i>)	79.92 ± 9.71	[50 – 110]
Circunferencia Abdominal (<i>cm</i>)	94.35 ± 11.68	[68 – 126]
Glucosa Basal (<i>mg/dl</i>)	86.38 ± 21.27	[49.7 – 220.04]
Triglicéridos (<i>mg/dl</i>)	147.64 ± 80.95	[54.6 – 518.8]
Colesterol Lipoproteínas de baja densidad (HDL) (<i>mg/dl</i>)	39.80 ± 11.93	[10.9 – 129.8]

Tabla 5.1: Factores de SM para 142 hombres mayores a 65 años.

Características	Media \pm std	Rango
Presión Sistólica (<i>mmHg</i>)	128.95 ± 14.63	[90 – 190]
Presión Diastólica (<i>mmHg</i>)	78.42 ± 9.21	[58 – 108]
Circunferencia Abdominal (<i>cm</i>)	94.15 ± 10.02	[68 – 130]
Glicemia (<i>mg/dl</i>)	88.15 ± 19.16	[49.1 – 197.8]
Triglicéridos (<i>mg/dl</i>)	154.92 ± 73.85	[59.1 – 575.6]
Colesterol HDL (<i>mg/dl</i>)	41.24 ± 11.25	[16 – 125.8]

Tabla 5.2: Factores de SM para 245 mujeres mayores a 65 años.

5.1.2. Algoritmo de Kohonen y Experimentos

Uno de los desafíos al usar SOM es el pre-procesamiento o normalización de las variables de entrada. Existen varias metodologías: usando categorías o rangos, normalizando restando la media y dividiendo por la desviación estándar, aplicando logaritmos cuando el rango de

variación ocupa varios ordenes de magnitud o la codificación binaria [26],[27].

En este trabajo se exploran el uso de rangos y la codificación binaria en el pre-procesamiento de las variables para clasificar SM en adultos mayores de las parroquias urbanas de Cuenca usando el algoritmo de Kohonen.

- **Normalización por rangos:** se asigna un valor [0,2 0,4 0,6 0,8 1] en función de la clase en el cual se encuentre la variable, tal como se indica en la Tabla 5.4.
- **Normalización Binaria:** las entradas son vectores de 5 elementos, se asigna uno a la posición de la clase de la variable.

Valores de Corte	Mínimo	1	2	3	4	Máximo
Presión Sistólica (<i>mmHg</i>)	90	110	130	150	170	190
Presión Diastólica (<i>mmHg</i>)	50	62	74	86	98	110
Circunferencia Abdominal (<i>cm</i>)	68	80.4	92.8	105.2	117.6	130
Glicemia (<i>mg/dl</i>)	49.1	83.28	117.47	151.66	185.85	220.04
Triglicéridos (<i>mg/dl</i>)	54.6	158.8	263	367.2	471.4	575.6
Colesterol HDL (<i>mg/dl</i>)	10.9	34.68	58.46	82.24	106.02	129.8

Tabla 5.3: Valores de Corte para la base de datos de SM.

Valor de entrada	Mín - 1	1 - 2	2 - 3	3 - 4	4 - Máx
Valor normalizado	0.2	0.4	0.6	0.8	1

Tabla 5.4: Normalización por Intervalo.

5.2. Uso del software Viscovery SOMine

Viscovery SOMine es una herramienta que se ha utilizado con éxito en el análisis de datos. En el cual se realiza un análisis predictivo en la investigación de la minería de datos y su uso se ha empleado en aplicaciones industriales.

Este software utiliza SOM para la representación y visualización de datos y también métodos estadísticos para el análisis de datos obtenidos. Los clústeres obtenidos se obtienen con algoritmos de agrupación que aprovechan el ordenamiento del SOM. A continuación, se menciona un ejemplo ilustrativo que indica la idea de representación y visualización de los datos

con [SOM](#).

Ejemplo: En un campo de fútbol están concentradas 1000 personas. Se supone un conjunto de atributos para cada persona estos son: genero, edad estado familiar e ingresos. Ahora se pide que cada persona en el campo se acerque a personas que son parecidas a ellas de acuerdo a los atributos antes mencionados. Se puede decir que las personas están rodeadas de otras con características similares. Esto indica una configuración bidimensional de datos multidimensionales.

A continuación, se pide a la multitud que levante una bandera de colores de acuerdo a su edad (azul para <20 , verde entre 20 a 29, amarillo entre 30 a 39, naranja entre 40 y 49 y rojo para >50). En la Figura 4.3 se observa la distribución del atributo edad en el campo de fútbol. De la misma manera se pide a las personas que levanten la mano de acuerdo a los demás atributos. Para cada atributo se tiene una distribución de color en el campo. Este patrón de color es codificado por colores y visualizadas en el software Viscovery SOMine. Se pueden poner todos los mapas juntos e inspeccionar los resultados. Por ejemplo, se puede estar clasificando personas que pertenecen a un clúster de personas jóvenes y con ingresos menores, por lo tanto, existe correlación entre la edad y el nivel de ingreso. De esta manera se encuentran relaciones definidas entre personas de mismos clústeres o grupos.

5.2.1. Metodología Viscovery SOMine

Es necesario la normalización de los datos, se emplea Matlab para esto, y se obtiene: una base completa, una base de datos de hombres y una base de datos de mujeres que se van a dividir en dos conjuntos, para el entrenamiento y la validación de los resultados.

Estos datos se guardan en formato de excel para exportarlos a Viscovery SOMine, y se establece en este caso los 6 atributos que son: Presión Arterial Sistólica, Presión Arterial Diastólica, Circunferencia Abdominal, Glicemia, Triglicéridos y Colesterol HDL y también se define un código para identificar que entrada se ingresó que no es una etiqueta, es decir no se está identificando a un sujeto con o sin [SM](#).

En las configuraciones se define el número de neuronas en la capa de competición en este caso se establecieron 400 neuronas siendo el máximo de 10000 y se establece el método de agrupamiento SOM-Ward-Clusters que define la relación entre neuronas que pertenecen a un mismo clúster en la capa de competición, el número de clústeres se define o se establece como predeterminado.

El resultado es un mapa de colores de acuerdo al número de clústeres y en cada clúster están los sujetos que pertenecen al mismo, en el cual existe un registro del porcentaje de sujetos.

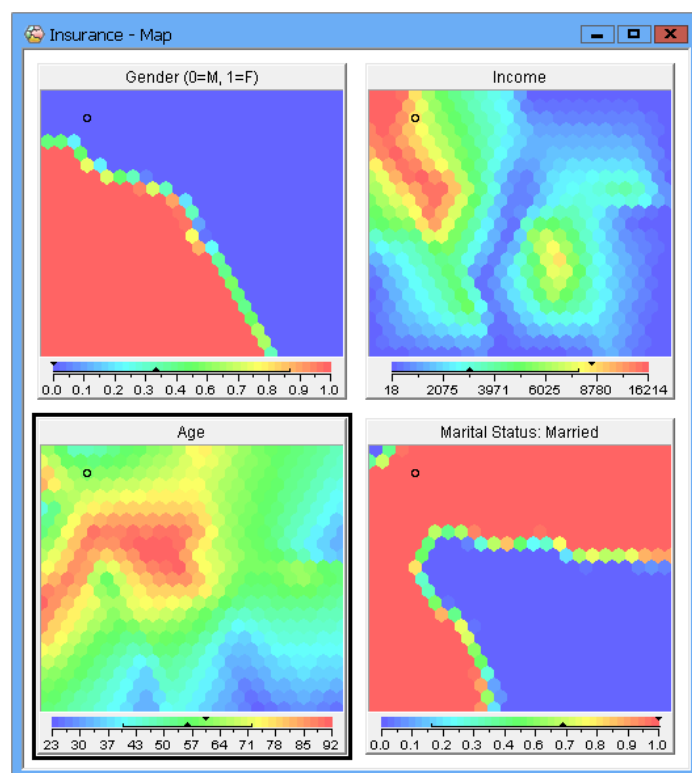


Figura 5.1: Visualización con el software Viscosity SOMine.

El código de cada sujeto agrupado en un clúster se compara en Matlab para determinar la media, distribución estándar, correlación entre clústeres y cuantos sujetos tienen o no SM. La Figura 5.3 muestra los valores correspondientes a 6 parámetros de la base de datos agrupados en dos clústeres, estos corresponden a los factores de riesgo de SM (Presión Arterial Sistólica, Presión Arterial Diastólica, Circunferencia Abdominal, Glucosa en Ayunas, Colesterol HDL y Triglicéridos), los valores tienen definidos puntos de corte que concuerdan con los puntos de color siendo un color azul a verde correspondiente a valores normales y un tono más verde a rojo correspondiente a valores alterados.

Hay que tener en cuenta que los valores de colesterol HDL >40 significa colesterol bueno, por tal motivo este corresponde a que el tono de colores de azul a celeste claro corresponde a valores malos y un tono verde a rojo valores normales, en los resultados se puede observar que ciertos atributos específicamente la Presión Sistólica y la Presión Diastólica se parecen pero difieren con todos los demás atributos, por lo tanto, no es posible obtener una buena correlación de los datos que hay en un mismo clúster que permita diferenciar a un sujeto que tiene o no SM, por este motivo se emplean otras normalizaciones para establecer semejanzas.

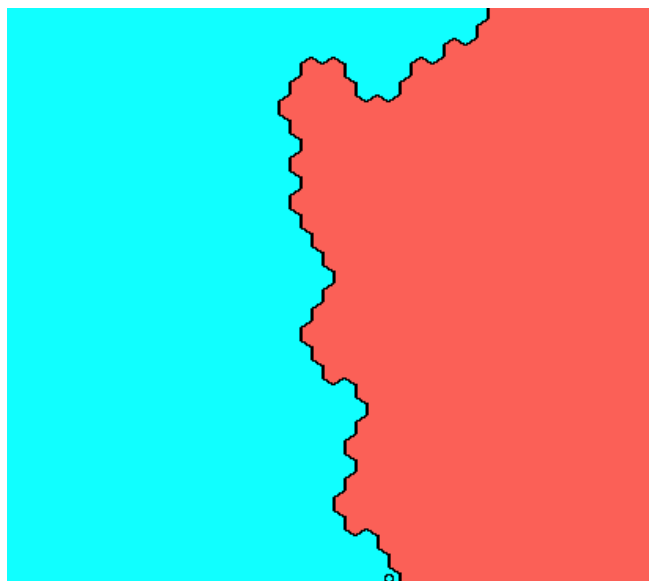


Figura 5.2: Clasificación de dos clústeres, de acuerdo a un color específico (celeste y rojo).

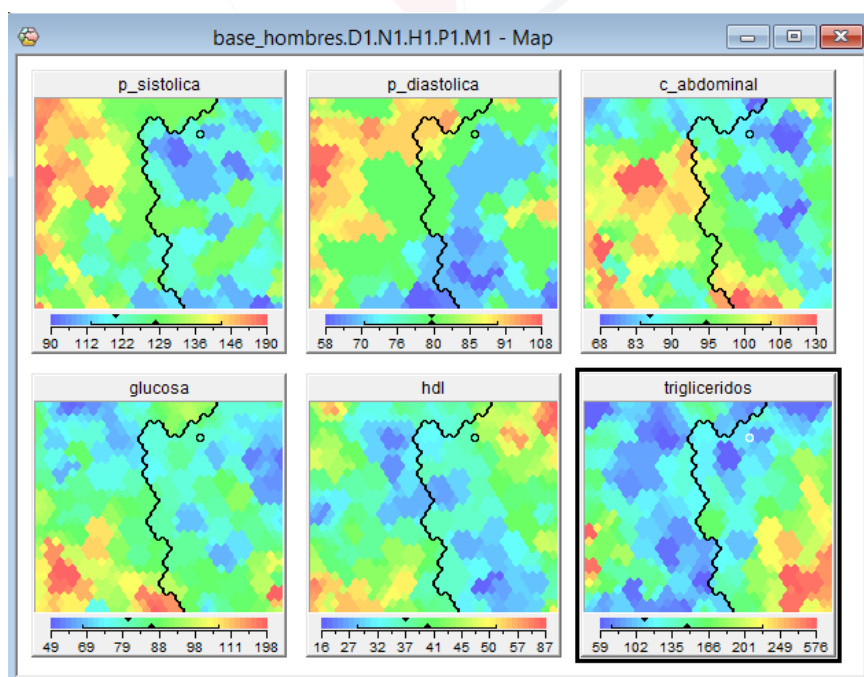


Figura 5.3: Visualización con el software Viscovery SOMine, se observa que en los dos clústeres hay sujetos con valores normales y alterados de Circunferencia Abdominal, Glicemia y Triglicéridos.

5.3. Resultados de la estimación del Síndrome Metabólico

Se realizaron tres experimentos, considerando en cada caso los dos tipos de pre-procesamiento y las seis variables del SM, en cada experiencia se definieron tres clústeres.

- **Toda la población:** 270 adultos mayores para entrenamiento y 117 para validación.
- **Solo población femenina:** 170 para entrenamiento y 75 para validación.
- **Solo población masculina:** 100 para entrenamiento y 42 para validación.

5.3.1. Entrenamiento de la red de Kohonen

Normalización Binaria

La base de entrenamiento tiene 270 sujetos entre hombres y mujeres. Los 6 parámetros de SM, se dividen en tres agrupaciones. El clúster 1 tiene mayor cantidad de sujetos con SM, mientras que la Clase 2 y 3 tienen mayor cantidad de sujetos sanos, como indican los resultados en la Tabla 5.5.

Núm. clúster	clúster 1	clúster 2	clúster 3			
Entrenamiento	N=121	N=112	N=37			
Con SM	90(74.38 %)	34(30.35 %)	10(27.02 %)			
Características	Media±std	Media±std	Media±std	<i>p12</i>	<i>p13</i>	<i>p23</i>
PAS (<i>mmHg</i>)	140,70±12,30	124,70±6,61	111,41±5,90	0,000	0,000	0,000
PAD (<i>mmHg</i>)	85,20±8,94	76,95±6,18	67,30±4,71	0,000	0,000	0,000
CA (<i>cm</i>)	96,43±11,83	91,60±9,57	93,00±8,09	0,009	0,188	0,504
Glicemia (<i>mg/dl</i>)	84,26±13,58	85,16±17,32	87,87±25,46	0,667	0,675	0,477
Triglicéridos (<i>mg/dl</i>)	144,40±68,39	153,48±72,79	135,62±65,25	0,231	0,021	0,308
HDL (<i>mg/dl</i>)	39,65±9,40	39,95±10,35	41,76±12,88	0,840	0,039	0,408

Tabla 5.5: Características de los clústeres para toda la población usando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.

La base de entrenamiento consta de 170 mujeres. La Clase 1 tiene mayor cantidad de sujetos con SM, la Clase 2 indica una buena clasificación de sujetos sin SM (35 de los 36 sujetos no tienen SM) y la Clase 3 no es muy clara ya que tiene 3 sujetos como se indica en la Tabla 5.6.

Núm. clúster	clúster 1	clúster 2	clúster 3			
Entrenamiento	N=131	N=36	N=3			
Con SM	79(60,31 %)	1(2,77 %)	1(33,33 %)			
Características	Media±std	Media±std	Media±std	<i>p12</i>	<i>p13</i>	<i>p23</i>
PAS (<i>mmHg</i>)	133,40±12,68	117,42±11,66	146,67±11,55	0,000	0,049	0,005
PAS (<i>mmHg</i>)	81,30±7,80	67,33±5,78	103,33±5,77	0,000	0,002	0,002
CA (<i>cm</i>)	94,05±12,52	93,33±10,85	93,00±12,17	0,873	0,775	1
Glicemia (<i>mg/dl</i>)	84,48±13,20	87,63±25,82	80,66±7,02	0,632	0,609	0,732
Triglicéridos (<i>mg/dl</i>)	149,35±72,94	154,13±74,25	178,63±14,38	0,703	0,108	0,304
HDL (<i>mg/dl</i>)	40,28±10,51	42,74±12,32	51,23±20,08	0,196	0,336	0,772

Tabla 5.6: Características de los clústeres para la población femenina utilizando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.

La base de entrenamiento consta de 100 hombres. La Clase 1 y 2 tienen igual cantidad de sujetos con SM y la Clase 3 tiene 4 sujetos con valores normales de SM, como se indica en la Tabla 5.7.

Núm. clúster	clúster 1	clúster 2	clúster 3			
Entrenamiento	N=60	N=36	N=4			
Con SM	18(30 %)	18(50 %)	0(0 %)			
Características	Media±std	Media±std	Media±std	<i>p12</i>	<i>p13</i>	<i>p23</i>
PAS (<i>mmHg</i>)	128,62±14,61	133,58±13,95	111,25±2,50	0,062	0,007	0,002
PAD (<i>mmHg</i>)	81,47±8,19	80,56±9,92	61,25±2,50	0,335	0,001	0,002
CA (<i>cm</i>)	91,90±6,62	97,69±11,11	94,50±9,98	0,000	0,813	0,484
Glicemia (<i>mg/dl</i>)	81,23±17,77	87,39±9,84	125,38±14,99	0,003	0,001	0,001
Triglicéridos (<i>mg/dl</i>)	145,21±66,70	135,78±54,45	107,55±32,39	0,652	0,207	0,367
HDL (<i>mg/dl</i>)	37,77±7,98	39,57±8,52	39,38±2,84	0,120	0,454	0,882

Tabla 5.7: Características de los clústeres para la población masculina utilizando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.

Normalización por rangos

La base de entrenamiento tiene 270 sujetos entre hombres y mujeres. Los 6 parámetros de **SM**, se dividen en tres agrupaciones. La Clase 2 tiene mayor cantidad de sujetos con **SM**, mientras que la Clase 1 y 3 tienen mayor cantidad de sujetos con valores normales de **SM**, como indican los resultados en la Tabla 5.8.

Núm. clúster Entrenamiento Con SM Características	clúster 1 N=112 51(41 %) Media±std	clúster 2 N=112 76(62 %) Media±std	clúster 3 N=46 10(19 %) Media±std	<i>p12</i>	<i>p13</i>	<i>p23</i>
PAS (<i>mmHg</i>)	122,87±10,40	139,46±13,51	124,61±11,58	0,000	0,380	0,000
PAD (<i>mmHg</i>)	75,05±8,42	85,50±8,64	74,67±6,91	0,000	0,927	0,000
CA (<i>cm</i>)	92,29±9,00	96,01±12,75	93,02±10,47	0,395	0,457	0,421
Glicemia (<i>mg/dl</i>)	86,28±21,29	85,38±13,37	81,73±11,40	0,650	0,380	0,166
Triglicéridos (<i>mg/dl</i>)	170,76±84,47	131,62±50,21	126,39±44,64	0,001	0,002	0,498
HDL (<i>mg/dl</i>)	35,00±7,58	40,09±8,18	52,33±9,64	0,001	0,000	0,000

Tabla 5.8: Características de los clústeres para toda la población usando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAD: Presión Arterial diastólica, Colesterol HDL).

La base de entrenamiento consta de 170 mujeres. La Clase 1 tiene mayor cantidad de sujetos con **SM**, la Clase 2 tiene un porcentaje de Sujetos con y sin **SM** y la Clase 3 clasifica de mejor manera a los sujetos sin **SM** (Tabla 5.9).

Núm. clúster Entrenamiento Con SM Características	clúster 1 N=66 57(86,36 %) Media±std	clúster 2 N=59 25(42,37 %) Media±std	clúster 3 N=45 0(0 %) Media±std	<i>p12</i>	<i>p13</i>	<i>p23</i>
PAS (<i>mmHg</i>)	134,24±15,76	134,20±10,58	119,20±9,37	0,936	0,000	0,000
PAD (<i>mmHg</i>)	78,53±10,32	84,88±7,30	70,96±5,91	0,000	0,000	0,000
CA (<i>cm</i>)	102,62±11,77	84,68±7,17	93,13±7,94	0,000	0,000	0,000
Glicemia (<i>mg/dl</i>)	88,53±20,77	83,04±14,362	82,70±10,80	0,179	0,482	0,477
Triglicéridos (<i>mg/dl</i>)	187,26±94,00	126,12±36,85	129,99±45,64	0,000	0,002	0,660
HDL (<i>mg/dl</i>)	37,20±9,52	43,67±12,52	43,04±10,04	0,008	0,004	0,896

Tabla 5.9: Características de los clústeres para la población femenina utilizando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.

La base de entrenamiento consta de 100 hombres. La Clase 1 y 2 tiene mayor cantidad de sujetos con SM, la Clase 3 indica una buena clasificación de sujetos sin SM como se indica en la Tabla 5.10.

Núm. clúster Entrenamiento Con SM Características	clúster 1 N=57 16(28,07 %) Media±std	clúster 2 N=31 18(58,06 %) Media±std	clúster 3 N=12 1(8,33 %) Media±std			
				<i>p12</i>	<i>p13</i>	<i>p23</i>
PAS (<i>mmHg</i>)	126,42±13,27	137,94±15,21	124,08±11,68	0,001	0,543	0,008
PAD (<i>mmHg</i>)	77,96±8,13	87,87±7,02	72,08±8,95	0,000	0,065	0,000
CA (<i>cm</i>)	90,54±6,64	99,98±10,01	95,75±7,70	0,000	0,073	0,249
Glicemia (<i>mg/dl</i>)	79,18±11,96	86,09±7,50	111,63±30,32	0,011	0,000	0,003
Triglicéridos (<i>mg/dl</i>)	147,76±69,70	139,40±50,87	107,23±29,91	0,948	0,047	0,051
HDL (<i>mg/dl</i>)	37,71±8,35	36,79±5,93	46,54±7,01	0,803	0,001	0,001

Tabla 5.10: Características de los clústeres para la población masculina utilizando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, Colesterol HDL.

5.3.2. Validación de la red de Kohonen

La Figura 6.1. muestra la distribución de los clústeres según el tipo de pre-procesamiento para la población que se utilizó para el entrenamiento de la red (base de datos de 270 adultos mayores). En las Tablas 5.11 y 5.12 se presentan las características de los tres clústeres obtenidos para la normalización binaria y por rangos respectivamente para el caso del experimento 1 (N=117 sujetos).

Núm. clúster Validación Con SM	clúster 1 N=33 28(86,36 %)	clúster 2 N=63 27(42,37 %)	clúster 3 N=21 10(42,37 %)			
--------------------------------------	----------------------------------	----------------------------------	----------------------------------	--	--	--

Características	Media±std	Media±std	Media±std	<i>p12</i>	<i>p13</i>	<i>p23</i>
PAS (<i>mmHg</i>)	137,48±14,34	123,46±8,73	112,57±12,04	0,001	0,001	0,001
PAD (<i>mmHg</i>)	84,73±8,54	77,67±6,69	69,29±7,79	0,001	0,001	0,001
CA (<i>cm</i>)	95,71±11,41	94,57±11,66	94,67±9,87	0,841	0,972	0,832
Glicemia (<i>mg/dl</i>)	92,26±23,41	93,47±27,76	94,67±9,87	0,951	0,716	0,844
Triglicéridos (<i>mg/dl</i>)	147,65±79,66	176,90±109,78	155,17±71,28	0,247	0,546	0,642
HDL (<i>mg/dl</i>)	38,04±6,67	43,71±18,39	43,91±8,34	0,157	0,005	0,173

Tabla 5.11: Características de los clústeres para toda la población usando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAD: Presión Arterial diastólica, Colesterol HDL).

Núm. clúster Validación Con SM Características	clúster 1 N=85 41(48,23 %) Media±std	clúster 2 N=31 22(70,97 %) Media±std	<i>p12</i>
PAS (<i>mmHg</i>)	120,14±11,01	140,35±10,51	0,000
PAD (<i>mmHg</i>)	75,65±8,05	84,65±8,15	0,000
CA (<i>cm</i>)	93,54±10,32	98,82±12,68	0,076
Glicemia (<i>mg/dl</i>)	90,47±25,22	100,52±28,61	0,040
Triglicéridos (<i>mg/dl</i>)	168,55±95,88	154,02±99,82	0,295
HDL (<i>mg/dl</i>)	39,77±8,67	46,23±18,51	0,080

Tabla 5.12: Características de los clústeres para toda la población usando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAD: Presión Arterial diastólica, Colesterol HDL).

En el caso del pre-procesamiento binario (Tabla 5.11) se observa que las poblaciones son diferentes entre ellas en relación a la presión arterial. Se observan tres clústeres con valores de presión arterial diastólica y sistólica bien diferenciados. Para el resto de las características del SM no se presentan diferencias entre los clústeres.

En la Tabla 5.12, se observa que la mayor parte de la población se divide entre los clústeres 1 y 2. El clúster 3 solo tiene un adulto mayor. El clúster 1 presenta valores promedios normales

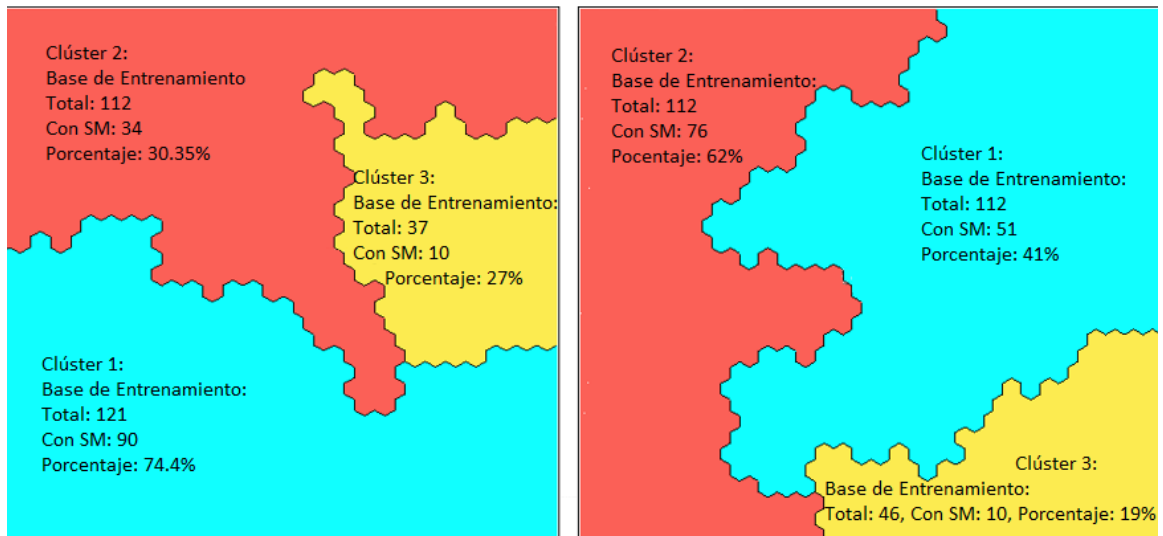


Figura 5.4: a) SOM usando pre-procesamiento binario. b) SOM usando pre-procesamiento por rangos para la base completa.

para todas las variables a excepción de los triglicéridos. La mayoría de valores promedios del clúster 2 se encuentran alterados o en el límite del valor corte de diagnóstico. Se observa que los sujetos agrupados en este grupo presentan características más relacionadas con el SM y por ende un porcentaje mayor de adultos mayores con SM.

5.4. Conclusiones

Los resultados usando un pre-procesamiento por rangos permiten una mejor clasificación de la población con SM. La selección del método de pre-procesamiento de los datos es de suma importancia para la utilización de SOM.

Este estudio ha permitido escoger el tipo de pre-procesamiento para el diagnóstico de SM en la población de adultos mayores de la ciudad de Cuenca. Los trabajos futuros están orientados a experimentar con diferentes números de clases para la normalización por rangos y de clústeres para el diagnóstico de SM en adultos mayores usando SOM.



Capítulo 6

Conclusiones y Trabajos Futuros



UNIVERSIDAD DE CUENCA
desde 1867

6.1. Conclusiones

En este trabajo se evaluó el uso de los algoritmos de aprendizaje no supervisados para el diagnóstico de la [Resistencia a la Insulina \(RI\)](#) y el [Síndrome Metabólico \(SM\)](#) en adultos mayores.

Para el estudio de la [RI](#), se analizaron los algoritmos de clasificación no supervisada, para agrupar sujetos en dos y tres clústeres, es decir clasificar sujetos que tengan [RI](#) e identificar los adultos mayores en riesgo de padecer [RI](#).

El método de *K*-medias permitió estudiar cuatro métodos de estimación de la [SI](#): Homa-IR, Quicki, Matsuda y Avignon. Los agrupamientos obtenidos usando los métodos empleados en clínica como el Homa-IR y el Quicki, presentaron valores de corte similares a otros reportados en la literatura. Lo cual constituyó un hallazgo importante para la evaluación de la población adulta mayor de Cuenca, ya que valida estos límites.

Los métodos que usan valores basales y postprandiales identifican una mayor cantidad de población con [RI](#) y este grupo correspondió a los adultos mayores con una circunferencia abdominal elevada, lo cual describe una dislipidemia frecuente en sujetos con [RI](#) [25] y, por tanto, una población con riesgo a desarrollar enfermedades cardiovasculares.

Las diferencias de diagnóstico que se obtuvieron según el método para la estimación de la [SI](#) usado, indicó que es necesario explorar con más profundidad desde el punto de vista clínico, la relación de la [SI](#) con otros parámetros de laboratorio. Todo esto para mejorar el diagnóstico de adultos mayores de nuestra región sensibles de padecer [RI](#).

Como consecuencia de la clasificación de los 119 adultos mayores se elaboró una plataforma para ingresar pacientes, los parámetros que se ingresan son valores basales y postprandiales de insulina y glucosa, peso, circunferencia abdominal e índice de masa corporal. En esta plataforma se determinó si un sujeto pertenece a uno de los tres clústeres obtenidos con el algoritmo de *K*-medias en función de los cuatro métodos de estimación estudiados. Esta plataforma de clasificación permite cuantificar los riesgos de padecer [RI](#) y puede establecerse como un sistema de ayuda al diagnóstico médico.

Para el diagnóstico del [SM](#) se usaron redes de Kohonen de tipo no supervisado, se exploró el uso de rangos y la codificación binaria en el preprocesamiento de las variables para clasificar [SM](#) en adultos mayores de las parroquias urbanas de Cuenca. Los resultados usando un preprocesamiento por rangos permitió una mejor clasificación de la población con [SM](#). La selección del método de preprocesamiento de los datos es de suma importancia para la utilización de [SOM](#). Este estudio ha permitido escoger el tipo de preprocesamiento para el diagnóstico de [SM](#).



en la población de adultos mayores de la ciudad de Cuenca.

Los resultados obtenidos permitieron desarrollar una plataforma de ayuda al diagnóstico de **RI** y fueron objeto de dos publicaciones en el IV Congreso de Tecnologías de la Información y Comunicación TIC-EC 2017 y el IEEE ETCM 2017: 2nd IEEE Ecuador Technical Chapters Meeting. Asimismo, los resultados del diagnóstico de **SM** se publicaron en el V Congreso REDU 2017 y el II Congreso I+D+Ingeniería.

6.2. Trabajos Futuros

Los resultados obtenidos en este trabajo, los valores de corte de **SI** para cada uno de los métodos explorados permitirán estudiar una población de 319 adultos mayores de Venezuela.

Se analizará si es válido realizar el procedimiento de clasificación de K -medias para determinar si las dos poblaciones dan los mismos puntos de corte, estableciendo criterios de **SI** de manera general en la población de adultos mayores de Sudamerica. Los trabajos futuros en cuanto al **SM** están orientados a experimentar con diferentes números de clases para la normalización por rangos y de clústeres para el diagnóstico de **SM** en adultos mayores usando **SOM**.

También se utilizará los **SOM** con clasificación supervisada (salida esperada), para obtener 2 clústeres con adultos mayores que tengan o no **SM**. Se analizará el clúster sin **SM** ya que hay sujetos que tienen factores normales pero elevados que pueden desencadenar en **SM**.

UNIVERSIDAD DE CUENCA
desde 1867



UNIVERSIDAD DE CUENCA
desde 1867

El logo de la Universidad de Cuenca, que consiste en un círculo azul claro con una flecha roja y blanca que atraviesa el círculo.

Anexos

UNIVERSIDAD DE CUENCA
desde 1867



Apéndice A

Artículo: Agrupamiento de K-medias para estimación de insulino-resistencia en Adultos Mayores de Cuenca

UNIVERSIDAD DE CUENCA
desde 1867

Agrupamiento de K-medias para estimación de insulino-resistencia en Adultos Mayores de Cuenca

Christian Vintimilla¹, Darwin Astudillo¹, Erika Severeyn^{2,3}, Lorena Encalada⁴, Sara Wong^{1,2}

¹ Departamento de Ingeniería Eléctrica, Electrónica y Telecomunicaciones, Universidad de Cuenca, Av. 12 de Abril y Agustín Cueva, Cuenca, Ecuador, 010201

² Grupo de Bioingeniería y Biofísica Aplicada, Universidad Simón Bolívar, Venezuela

³ Departamento de Tecnología de Procesos Biológicos y Bioquímicos, Universidad Simón Bolívar, Venezuela

⁴ Facultad de Ciencias Médicas, Universidad de Cuenca, Av. 12 de Abril y Paraíso, Cuenca, Ecuador, 010204

Autores para correspondencia: xavier.vintimilla92@ucuenca.ec, {fabian.astudillos, lorena.encalada, sara.wong}@ucuenca.edu.ec, {severeynrika, swong}@usb.ve

Fecha de recepción: 16 de mayo 2017 - Fecha de aceptación: 19 de junio 2017

ABSTRACT

Insulin resistance (IR) is a prediabetic condition that requires early diagnosis. However, there is no consensus regarding the estimation method according to the population studied. In this paper, we analyze 5 methods of IR estimation using a K-means clustering algorithm on a population of 119 older adults from Cuenca who underwent an Oral Test of Tolerance to the Glucose (OTTG) of two points. The K-means clustering algorithm, with $K = 2$, was applied in one-dimensional experiments for the Homa-IR, Quicki, Avignon, Matsuda and a multidimensional experiment using the four observations of the OGTT (fasting and postprandial values of glucose and insulin). The population was divided into two clusters: C-N, with normal values and C-IR with IR. Clinical cutoff points for Homa-IR and Quicki were validated. The number of individuals classified in each C-IR was very different according to each method. Methods using fasting and postprandial values (Avignon and Matsuda) identify a larger population with IR and this group corresponds to the Elderly with a high abdominal circumference. The multidimensional method seems to provide less information than one-dimensional methods that include postprandial values. The variety of clusters obtained according to the IR estimation method indicate that it is necessary to carefully analyze these results from the clinical point of view to establish strategies for early diagnosis of IR in the Elderly population of Cuenca.

Keywords: K-means, unsupervised learning, insulin resistance, elderly, Homa-IR, Quicki

RESUMEN

La Resistencia a la Insulina (RI) es una condición prediabética que requiere ser diagnosticada de forma temprana. Sin embargo, no existe consenso en cuanto al método de estimación según la población estudiada. En este trabajo se analizan 5 métodos de estimación de RI usando una clasificación de K-medias sobre una población de 119 Adultos Mayores de la ciudad de Cuenca quienes se sometieron a una Prueba oral de tolerancia a la Glucosa (POTG) de dos puntos. El algoritmo de K-medias, con $K=2$, fue aplicado en experimentos unidimensionales para los métodos Homa-IR, Quicki, Avignon, Matsuda y se realizó un experimento multidimensional usando las cuatro observaciones de la POTG (valores de glucosa e insulina basal y postprandial). La población se dividió en dos clústeres: C-N, con valores normales y C-RI con RI. Se validaron los puntos de corte clínicos para Homa-IR y Quicki. La cantidad de individuos clasificados en cada C-RI fue muy diferente según cada método. Los métodos que usan valores basales y postprandiales (Avignon y Matsuda) identificaron una mayor cantidad de población con RI y este grupo correspondió a los Adultos Mayores con una circunferencia abdominal elevada. El método multidimensional parece aportar menos información que los métodos unidimensionales que incluyen valores postprandiales. La pluralidad de los clústeres obtenidos según el método de estimación de RI, revelan que es necesario analizar detenidamente estos resultados desde el punto de vista clínico para establecer estrategias de diagnóstico precoz de RI en la población de Adultos Mayores cuencana.

Palabras clave: K-medias, aprendizaje no supervisado, resistencia a la insulina, adulto mayor, Homa-IR, Quicki

1. INTRODUCCIÓN

La baja Sensibilidad a la Insulina (SI) es la inactividad de reacción de las células del páncreas ante la presencia de la insulina, que conlleva al aumento no controlado de los niveles de glucosa en la sangre (hiperglucemia), y causa también la sobreproducción hepática de insulina conocida como hiperinsulinemia (Ronald Kahn, 1978). La Resistencia a la Insulina (RI) es una condición que puede favorecer la aparición de la diabetes.

El método de referencia, *gold standard*, para medir la SI es el clamp euglicémico-hiperinsulinémico sin embargo, es un método sumamente invasivo, complicado y costoso por lo cual se reserva para fines de investigación clínica. La práctica más común para estimar la SI es mediante métodos indirectos, los cuales se dividen en dos tipos, aquellos que contemplan los valores de glucosa e insulina durante toda la POTG, entre estos se encuentran el Matsuda y el Avignon entre otros. Por otro lado, están los métodos que involucran sólo los datos de glucosa e insulina en la etapa basal, como el Homa-IR, y el Quicki, estos métodos no contemplan el proceso de metabolización de glucosa (Severeyn, Wong, Passariello, Cevallos, & Almeida, 2012).

El método indirecto más utilizado en clínica es el índice Homa-IR. La mayor limitación de estos métodos es la ausencia o validación de los puntos de corte para el diagnóstico de RI, ya que los puntos de corte han sido determinados en poblaciones específicas. Estos métodos por ejemplo no han sido evaluados en una población de Adultos Mayores.

La clasificación no supervisada con el algoritmo de agrupamiento de k-medias ha sido utilizada en un trabajo previo para clasificar sujetos con Síndrome Metabólico (Altuve, Severeyn, & Wong, 2015). En este trabajo se tomó como referencia las observaciones unidimensionales (índice Homa-IR) y observaciones multidimensionales durante una Prueba Oral de Tolerancia a la Glucosa (POTG) luego se agruparon los sujetos empleando clústeres, con el objetivo de relacionar los grupos con RI, los resultados indicaron que es probable diagnosticar resistencia a la insulina usando un punto de corte de HOMA-IR diferente al usado en la clínica. En este trabajo se propone el uso del algoritmo de agrupamiento no supervisado de k-medias, para analizar cinco métodos de estimación de SI y sus puntos de corte para diferenciar entre Adultos Mayores con RI.

2. MATERIALES Y MÉTODOS

2.1 Base de Datos

La base de datos está compuesta por 119 Adultos Mayores, no diabéticos, 42 hombres (71 ± 6.692 años) y 77 mujeres (72 ± 6.748 años) reclutados en el proyecto DIUC-PROMETEO denominado: Variabilidad de la frecuencia cardíaca y sensibilidad a la insulina en la población Adulta Mayor de Cuenca, Ecuador (Parra et al., 2015). A cada sujeto se le realizó una hematología, un perfil lipídico, medidas antropométricas y una Prueba Oral de tolerancia a la Glucosa (POTG) de dos puntos.

En la POTG se toma una muestra de sangre en ayunas, para medir la glucosa e insulina basal, luego el sujeto ingiere 75 gramos de glucosa y después de dos horas se toma una nueva muestra de sangre para medir la glucosa e insulina postprandial. La tabla 1 presenta las medidas antropométricas y los valores de la POTG de la población de Adultos Mayores.

Tabla 1. Medidas antropométricas y valores de la POTG.

Características	Media \pm std	Rango
Peso (kg)	68.03 ± 10.51	[45 – 90]
Altura (m)	1.53 ± 0.098	[1.02 – 1.76]
Circunferencia Abdominal (cm)	93.74 ± 11.14	[69 - 130]
Insulina Basal(μ UI/mL)	10.67 ± 8.08	[2.00 – 43.60]
Glucosa Basal (mg/dl)	90.51 ± 25.17	[61.30 – 241.10]
Insulina Postprandial(μ UI/mL)	58.11 ± 44.66	[4.30 – 285.00]
Glucosa Postprandial (mg/dl)	102.31 ± 32.20	[31.10 – 221.60]

2.2 Cuantificación de la SI a partir de la POTG.

Los métodos utilizados para la cuantificación de la SI sirven para establecer de manera aproximada si un sujeto presenta o no resistencia a la insulina. En la tabla 2, se describen los métodos para estimar la SI usados en este trabajo (Brown & Yanovski, 2014; Conwell, Trost, Brown, & Batch, 2004; Severeyn et al., 2012). En estudios previos se indica valores de punto de corte para el método de Homa-IR, valor definido en 2.5 (Friedrich et al., 2012). En otro estudio realizado sobre 5518 mayores de 18 años propone

como punto de corte para el índice de Quicki, un valor de 0.343 para hombres con SM y de 0.331 en mujeres con SM (Motamed et al., 2016). Para los métodos de Avignon y Matsuda no existen puntos de corte de referencia. La tabla 3 presenta los valores de referencia para la glucosa e insulina en niveles basales y postprandiales (Ronald Kahn, 2003).

Tabla 2. Métodos para cuantificar la Sensibilidad a la Insulina, (I_0 : Insulina basal, I_{120} : Insulina postprandial, G_0 : Glucosa Basal y G_{120} : Glucosa Postprandial).

Método	Fórmula	Descripción
Homa-IR	$\frac{G_0 I_0}{405}$	Utiliza los valores de glucosa e insulina basal.
Quicki	$\frac{1}{\log G_0 + \log I_0}$	
Avignon	$\frac{0.137 * Sib + Si2h}{2 * \frac{Peso}{10^8}}$ $Sib = \frac{I_0 G_0 * 150 \text{ mL/kg}}{10^8}$ $Si2h = \frac{I_{120} G_{120} * 150 \text{ mL/kg}}{10^8}$	Utiliza los valores de glucosa e insulina basal; y, glucosa e insulina a los 120 minutos.
Matsuda-Modificado	$\frac{10000}{\sqrt{I_0 * G_0 * \left(\frac{G_0 + G_{120}}{2} \right) * \left(\frac{I_0 + I_{120}}{2} \right)}}$	
Multi-dimensional	G_0 : (mg/dL) I_0 : (μUI/mL) G_{120} : (mg/dL) I_{120} : (μUI/mL) --	Utiliza los valores de los dos puntos de glucosa y dos puntos de insulina

Tabla 3. Valores normales de glucosa basal, insulina basal, glucosa postprandial e insulina postprandial.

Parámetros	Valor Normal
Glucosa basal (G_0)	<100 mg/dL
Glucosa postprandial (G_{120})	<140 mg/dL
Insulina basal (I_0)	<20 μUI/mL
Insulina postprandial (I_{120})	<60 μUI/mL

2.3 Algoritmo de K-medias y Experimentos

El algoritmo de k -medias, agrupa o predice los datos con un aprendizaje no supervisado. El usuario especifica el número de agrupaciones k . Cada agrupación tiene un vector medio (centroide). La distancia Euclídea se utiliza para encontrar el clúster más cercano al objeto. El agrupamiento se realiza iterativamente para asignar el objeto al clúster más cercano. La media μC con nC el número de objetos

del clúster C contiene los valores medios de sus objetos y se define para el atributo $i = \sum_{j=1}^{n_C} X_{ij} / n_C$. La iteración continúa hasta que los objetos que cambian los clústeres estén por debajo de un umbral especificado por el usuario. Los objetos cambian de clúster mientras el agrupamiento de k -medias tiene complejidad $O(tkn)$, donde t es el número de iteraciones. Típicamente $k, t \ll n$, de manera que los k -medios son esencialmente lineales en el número de objetos n (Andreopoulos, An, Wang, & Schroeder, 2009). Los k -medios pueden terminar en un óptimo local, más que global. La finalidad del algoritmo es dividir M observaciones (número de objetos) de N dimensiones (atributos de un objeto) en k grupos.

Se diseñaron y realizaron los siguientes experimentos:

1. **Experimentos unidimensionales:** correspondiente a los cuatro métodos estudiados (Homa-IR, Quicki, Avignon y Matsuda). Cada método fue aplicado a las 119 observaciones, con dimensión $N = 1$ y agrupación $K = 2$.
2. **Experimento multidimensional:** usando el algoritmo de k -medias con las observaciones correspondiente a $N = 4$ dimensiones, correspondientes a los valores de Glucosa e Insulina basales G_0, G_{120} y Postprandiales (I_0, I_{120}) . Estas variables fueron normalizadas para tener media=0 y varianza=1.

Para cada experiencia, el algoritmo fue ejecutado 10 veces usando diferentes inicializaciones para evitar los mínimos locales. Luego, se selecciona el cuadrado de las distancias de cada punto al centroide dentro x|del grupo (distancia Euclídea) y se elige la menor. El máximo de iteraciones para el algoritmo de k -medias se estableció en 100. En la figura 1 se observan las asignaciones de cada uno de los experimentos unidimensionales a los grupos C-N y C-RI y sus los valores de corte entre los clústeres.

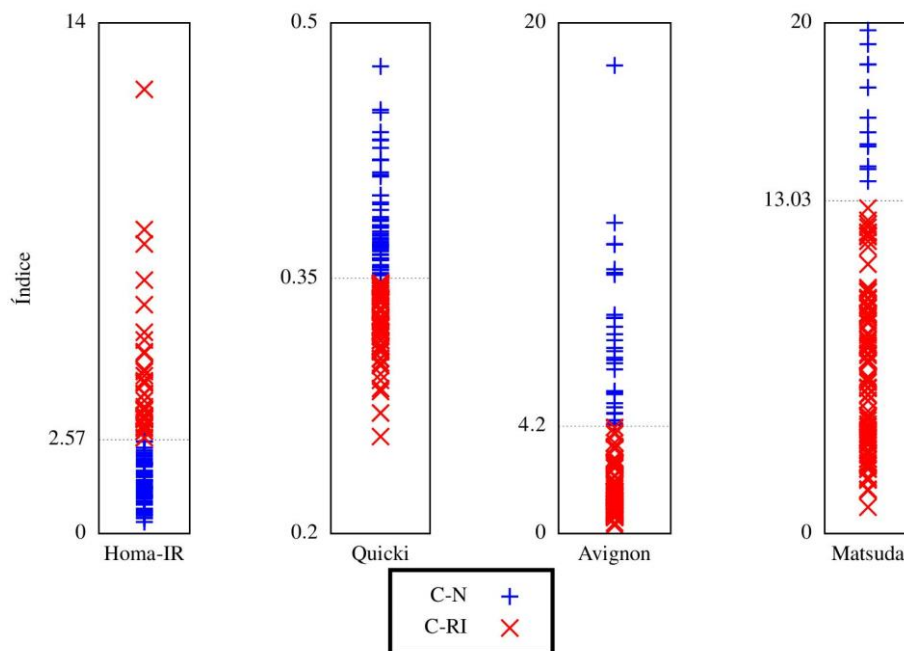


Figura 1. Distribución de clústeres C-N y C-RI y puntos de corte para cada experimento unidimensional.

3. RESULTADOS

Agrupamiento de *k-medias*

La clasificación con el algoritmo de *k-medias* dividió a la base de datos de 119 sujetos en K=2 grupos, los cuales denominaremos C-N, para el clúster con valores dentro de lo considerado normal y C-RI para el clúster susceptible de presentar características de RI, de acuerdo a los puntos de corte obtenidos para cada método como se indica en la tabla 4. En la tabla 5 se presentan las características de cada clúster según sus valores de glucosa e insulina basales (G_0 , I_0) y postprandiales (G_{120} , I_{120}) y cintura abdominal. El *test* de Wilcoxon se usó para determinar la diferencia entre grupos. Se consideró un *p* valor ≤ 0.01 como significativo.

Tabla 4. Número de Adultos Mayores por clústeres (C-N y C-RI) y puntos de corte para Homa-IR, Quicki, Matsuda y Avignon

Método	C-N	C-RI	Punto de Corte	<i>p</i>
Homa-IR	80 sujetos	39 sujetos	≤ 2.57	1.056e-18
Quicki	59 sujetos	60 sujetos	≥ 0.35	5.213e-21
Matsuda	24 sujetos	95 sujetos	≥ 13.03	4.470e-14
Avignon	31 sujetos	88 sujetos	≥ 4.20	1.515e-16
Multidimensional	86 sujetos	33 sujetos	--	4.2559e-13

Tabla 5. Valores de Glucosa (mg/dL), Insulina (μ UI/mL) y Circunferencia Abdominal (cm) para cada clúster (C-N y C-RI) según cada método.

Método	valores	C-N (media \pm SD)		C-RI (media \pm SD)		<i>p</i>
Homa-IR	G_0	84.86	15.95	102.12	35.11	0.001
	I_0	6.62	2.67	18.97	9.08	0.000
	G_{120}	96.31	29.37	114.62	29.37	0.005
	I_{120}	40.63	24.20	93.96	24.20	0.000
	CA	92.88	11.52	95.49	10.23	N.S.
Quicki	G_0	83.16	15.30	97.75	30.49	0.000
	I_0	5.51	1.98	15.74	8.60	0.000
	G_{120}	96.47	28.27	108.05	34.93	0.027
	I_{120}	36.35	21.46	79.50	50.96	0.000
	CA	94.72	11.11	92.74	11.18	N.S.
Matsuda	G_0	79.53	11.89	93.29	26.87	0.004
	I_0	3.99	1.66	12.35	8.18	0.000
	G_{120}	96.47	28.27	108.05	34.94	0.000
	I_{120}	36.35	21.46	79.50	50.96	0.000
	CA	88.19	6.50	95.14	11.65	0.003
Avignon	G_0	80.55	12.17	94.03	27.58	5.55E-03
	I_0	6.27	7.31	12.22	7.80	8.96E-09
	G_{120}	79.39	23.02	110.39	31.15	5.35E-07
	I_{120}	18.78	8.87	71.96	43.96	7.78E-15
	CA	88.05	7.89	95.74	11.46	4.08E-04
	G_0	85.16	16.26	100.70	34.61	1.22E-02

	I ₀	8.21	6.11	15.34	9.30	1.18E-13
	G ₁₂₀	88.58	21.85	128.44	32.73	6.25E-05
Multidimensional	I ₁₂₀	34.90	18.32	102.25	46.79	4.26E-13
	CA	95.36	9.76	93.11	11.62	N.S

En la tabla 6, se presenta una síntesis de las características de los clústeres, según el siguiente criterio N indica un valor normal, A un valor alterado y E un valor elevado, según las bandas de criterio clínico.

Tabla 6. Características obtenidas para los métodos de SI. (I₀: Insulina basal, I₁₂₀: Insulina postprandial, G₀: Glucosa Basal y G₁₂₀: Glucosa Postprandial), Normal (N), Elevado (E) y alterado (A).

	clúster	G ₀	I ₀	G ₁₂₀	I ₁₂₀
Homa-IR	C-N	N	N	N	N
	C-RI	A	E	N	A
Quicki	C-N	N	N	N	N
	C-RI	E	E	N	A
Matsuda	C-N	N	N	N	N
	C-RI	N	E	N	A
Avignon	C-N	N	N	N	N
	C-RI	N	E	N	A
Multidimensional	C-N	N	N	N	N
	C-RI	A	E	E	A

Los puntos de corte obtenidos con el Homa-IR y el Quicki son bastante cercanos a los reportados por otros estudios (Friedrich et al., 2012; Motamed et al., 2016).

Al analizar las poblaciones según los valores promedios de glucosa e insulina se desprende de la tabla 1 lo siguiente:

- 1) El Homa-IR y el Quicki diferencian los clústeres según los valores de G₀, I₀ alterados y elevados respectivamente, y el C-RI también presenta I₁₂₀ alterado, aunque este valor no es considerado por el método. Sin embargo, los valores aumentados de I₁₂₀ son de esperarse si los basales son alterados (Velásquez, Severeyn, Herrera, Encalada, & Wong, 2017), esto podría indicar que el sujeto lleva algún tiempo desarrollando RI; por tanto, tiene continuamente valores alterados de insulina. Del mismo modo, el Quicki diferencia los clústeres según los valores elevados de G₀, I₀ y el C-RI, también presenta I₁₂₀ alterado, aunque este valor no es considerado en el método
- 2) Matsuda y Avignon diferencian los clústeres por sus valores de I₁₂₀ alterados y valores de I₀ elevados, aunque también se observan que todos los valores de la POTG son superiores para el clúster C-RI. Pareciera que el agrupamiento para estos métodos incluye como C-RI a aquellos individuos que presentan valores alterados de I₁₂₀, estos individuos no necesariamente presentan valores basales alterados, y en los métodos anteriores (Quicki y Homa-IR) no son detectados

como C-IR. Los métodos que toman en cuenta los valores de I_{120} e G_{120} pudieran ser más efectivos detectando grupos con riesgo a padecer de IR.

- 3) Usando 4 observaciones se evidencia que este método está separando los clústeres en base a G_0 y I_{120} alterados; adicionalmente, éste es el único método en donde para los clústeres en base a I_0 y G_{120} presenta valores elevados.

El k-medias permite separar en dos clústeres; según los valores elevados de glucosa e insulina usando métodos de estimación de SI, o directamente las variables medidas de la POTG. Lo interesante es que para cada método la cantidad de individuos que podrían diagnosticarse como insulino-resistentes o con RI es diferente.

Los métodos que incluyen todos los valores de glucosa e insulina ubican un mayor número de individuos en el grupo con valores alterados o valores más altos (C-RI). Esto pudiera deberse a que estos métodos al incluir los valores basales y postprandiales, detectan aquellos casos en donde los valores basales se encuentran normales, pero tienen valores postprandiales alterados. Estos casos en particular, son personas que tienen una disfunción metabólica que todavía no se ve reflejada en los valores basales, hay estudios que indican que los valores basales son los últimos en alterarse cuando se presenta la resistencia a la insulina (Velásquez et al., 2017).

Adicionalmente, para estos dos métodos se observa que los individuos con RI presentan circunferencias abdominales mayores que los individuos del C-N, lo cual no se observa con los métodos que solo toman en cuenta los valores basales, ni con el método multidimensional. Lo cual puede indicar que los métodos de SI que usan los valores de glucosa e insulina postprandial podrían detectar valores alterados de circunferencia abdominal; por tanto, aportan información fisiológica adicional que no se observa en la clasificación realizada por los métodos que solo toman en cuenta valores basales de glucosa e insulina.

4. CONCLUSIONES

El método de *K-medias* ha permitido estudiar cinco métodos para estimar la SI en una población adulta mayor. Los agrupamientos obtenidos usando los métodos empleados en clínica como el Homa-IR y el Quicki, presentan valores de corte similares a otros reportados en la literatura. Lo cual constituye un hallazgo importante para la evaluación de la población adulta mayor de Cuenca, ya que valida estos límites. Los métodos que usan valores basales y postprandiales identifican una mayor cantidad de población con RI y este grupo (C-RI) corresponde a los Adultos Mayores con una circunferencia abdominal elevada, lo cual describe una dislipidemia frecuente en sujetos con RI (Kawada, Andou, & Fukumitsu, 2016); y, por tanto, una población con riesgo a desarrollar enfermedades cardiovasculares. El método multidimensional, aunque separa dos clústeres bien definidos, parece aportar menos información que los métodos que incluyen información postprandial.

Las diferencias de diagnóstico obtenidas según el método para la estimación de la SI usado, indican que es necesario explorar con más profundidad desde el punto de vista clínico, la relación de la SI con otros

parámetros de laboratorio, como por ejemplo con la hemoglobina glucosilada. Todo esto para mejorar el diagnóstico de Adultos Mayores de nuestra región sensibles de padecer RI. Como trabajo futuro, se analizarán la relación de la RI con el Síndrome Metabólico en la población adulta mayor de Cuenca.

AGRADECIMIENTOS

Este trabajo fue realizado gracias al financiamiento de la Dirección de Investigación de la Universidad de Cuenca (DIUC).

REFERENCIAS

- Altuve, M., Severein, E., & Wong, S. (2015). Unsupervised subjects classification using insulin and glucose data for insulin resistance assessment. In *2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA)* (pp. 1–7). IEEE.
<https://doi.org/10.1109/STSIVA.2015.7330444>
- Andreopoulos, B., An, A., Wang, X., & Schroeder, M. (2009). A roadmap of clustering algorithms: finding a match for a biomedical application. *Briefings in Bioinformatics*, 10(3), 297–314.
<https://doi.org/10.1093/bib/bbn058>
- Brown, R. J., & Yanovski, J. A. (2014). Estimation of insulin sensitivity in children: methods, measures and controversies. *Pediatric Diabetes*, 15(3), 151–161.
<https://doi.org/10.1111/pedi.12146>
- Conwell, L. S., Trost, S. G., Brown, W. J., & Batch, J. A. (2004). Indexes of Insulin Resistance and Secretion in Obese Children and Adolescents. *Diabetes Care*, 27(2). Retrieved from <http://care.diabetesjournals.org/content/27/2/314.short>
- Friedrich, N., Thuesen, B., Jørgensen, T., Juul, A., Spielhagen, C., Wallaschofski, H., & Linneberg, A. (2012). The Association Between IGF-I and Insulin Resistance. *Diabetes Care*, 35(4). Retrieved from <http://care.diabetesjournals.org/content/35/4/768.short>
- Kawada, T., Andou, T., & Fukumitsu, M. (2016). Waist circumference, visceral abdominal fat thickness and three components of metabolic syndrome. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 10(1), 4–6. <https://doi.org/10.1016/j.dsx.2015.08.013>
- Motamed, N., Miresmail, S. J. H., Rabiee, B., Keyvani, H., Farahani, B., Maadi, M., & Zamani, F. (2016). Optimal cutoff points for HOMA-IR and QUICKI in the diagnosis of metabolic syndrome and non-alcoholic fatty liver disease: A population based study. *Journal of Diabetes and Its Complications*, 30(2), 269–274. <https://doi.org/10.1016/j.jdiacomp.2015.11.019>
- Parra, F., Andrade, D., Cruz, J., Solano-Quinde, L., Palacio-Baus, K., Encalada, L., & Wong, S. (2015). Plataforma basada en ecgML para el estudio de las complicaciones cardiovasculares en el adulto mayor con síndrome metabólico. *Maskana*, 6(Supl.), 157–164. Retrieved from <https://www.ucuenca.edu.ec/ojs/index.php/maskana/article/view/709>

Ronald Kahn, C. (1978). Insulin resistance, insulin insensitivity, and insulin unresponsiveness: A necessary distinction. *Metabolism*, 27(12), 1893–1902. [https://doi.org/10.1016/S0026-0495\(78\)80007-9](https://doi.org/10.1016/S0026-0495(78)80007-9)

Ronald Kahn, C. (2003). Follow-up report on the diagnosis of diabetes mellitus: the expert committee on the diagnosis and classifications of diabetes mellitus. *Diabetes care*, 26(11), 3160.

Severeyn, E., Wong, S., Passariello, G., Cevallos, J. L., & Almeida, D. (2012). Methodology for the study of metabolic syndrome by heart rate variability and insulin sensitivity. *Revista Brasileira de Engenharia Biomédica*, 28(3), 272–277. <https://doi.org/10.4322/rbeb.2012.026>

Velásquez, J., Severeyn, E., Herrera, H., Encalada, L., & Wong, S. (2017). Anthropometric index for insulin sensitivity assessment in older adults from Ecuadorian highlands. In E. Romero, N. Lepore, J. Brieva, & I. Larrabide (Eds.) (p. 101600S). International Society for Optics and Photonics. <https://doi.org/10.1117/12.2257207>





UNIVERSIDAD DE CUENCA
desde 1867



Apéndice B

**Artículo: An aide Diagnosis
System Based on k-means for
insulin resistance assessment in
eldery people from the
Ecuadorian highlands**

UNIVERSIDAD DE CUENCA
desde 1867

An aide Diagnosis System Based on k-means for insulin resistance assessment in elderly people from the Ecuadorian highlands

Christian Vintimilla¹, Sara Wong¹, Fabián Astudillo-Salinas¹, Lorena Encalada², Erika Severeñn³

¹Facultad de Ingeniería/DEET, Universidad de Cuenca, Ecuador

²Facultad de Ciencias Médicas, Universidad de Cuenca, Ecuador

³Departamento de Tecnología de Procesos Biológicos y Bioquímicos, Universidad Simón Bolívar, Venezuela

Abstract—The lack of standardized cut-off values for the surrogate methods to diagnose Insulin resistance (IR) and the fact that the sensitivity of these methods have been studied in specific populations have limited their use in clinical routine. We developed a system that could aide to diagnosis IR in elderly people, analyzing four surrogate methods of IR estimation using a k-means clustering algorithm. Study subjects included 119 nondiabetic participants over 65 year old from Ecuadorian highlands. Blood tests included a two-point oral glucose test tolerance. The k-means clustering algorithm, was applied in one-dimensional experiments for the Homa-IR, Quicki, Avignon and Matsuda. The population was divided into three clusters: C_N with normal values, C_{IR} with IR and C_A with values in between. The number of individuals classified in each C_{IR} was very different according to each method. With the cut-off values obtained, for each method, the system for the evaluation of IR in elderly people was developed. Our work is intended to aid physicians in the early detection of IR by using information from diverse methods.

Index Terms—K-means, unsupervised learning, insulin resistance, elderly, Homa-IR, Quicki

I. INTRODUCTION

The proportion of elderly people in Ecuador and around the world has increased considerably. In Latin America it is expected that by 2030, 30% of the population will be composed of seniors aged 65 years and over. The Ecuadorian population has improved their life expectancy from 48.3 years in 1950–55 to 75.6 years in 2010–15 [1]. This allowed many Ecuadorians to reach older ages, but this increase in age is also accompanied by an increase in chronic diseases like diabetes that affects the quality of life [2].

A prediabetic condition is Insulin Resistance (IR), which is the diminished biological response to the biological effects of this hormone [3]. IR is a condition that may favor the onset of diabetes and cardiovascular disease. Conversely, Insulin Sensitivity (IS) is defined as the ability of cells to react to the presence of glucose. Low IS is the reaction inactivity of pancreas cells in presence of insulin, which leads to uncontrolled increase in blood glucose levels (hyperglycemia), and also causes insulin overproduction known as hyperinsulinemia [4].

The gold standard reference method for measuring IS is the euglycemic-hyperinsulinemic clamp, however, it is extremely invasive, complicated and expensive and therefore reserved for clinical research purposes. The most common practice to estimate IS is using indirect methods, which are divided into two types: i) those that consider only glucose and insulin data in the basal stage (fasting values), such as Homa-IR, and Quicki and ii) the methods using the glucose and insulin values during the Oral Glucose Tolerance Test (OGTT), among which are Matsuda and Avignon, these methods do not contemplate the process of glucose metabolism [5].

Homa-IR is the standard method used routinely in clinical practice, although its performance is limited since patients with impaired carbohydrates metabolism may be diagnosed as false negative if the fasting values do not reflect the metabolic alterations. The greatest limitation of indirect methods to estimate IS is the absence or validation of cut-off points for the diagnosis of IR, since cut-off points have been determined in specific populations [6], [7].

The unsupervised classification with the clustering algorithm of k-means has been used in previous work, to classify 40 subjects with Metabolic Syndrome, Sportsmen and Sedentaries [8], [9], and more recently in a population of older adults in the Ecuadorian highlands [10]. In this work the k-means clustering algorithm, with $K = 2$, was applied in one-dimensional experiments for the Homa-IR, Quicki, Avignon, Matsuda and a multidimensional experiment using the four observations of the OGTT (fasting and postprandial values of glucose and insulin). The population was divided into two clusters: C_N , with normal values and C_{IR} with IR. Clinical cutoff points for Homa-IR and Quicki were validated. The number of individuals classified in each C_{IR} was very different according to each method. Methods using fasting and postprandial values (Avignon and Matsuda) identify a larger population with IR and this group corresponds to the Elderly with a high abdominal circumference. The multidimensional method seems to provide less information than one-dimensional methods that include postprandial values. The variety of clusters obtained according to the IR estimation

method indicate that it is necessary to carefully analyze these results from the clinical point of view to establish strategies for early diagnosis of IR in the Elderly population of Cuenca.

This study was designed to investigate whether the k-means clustering algorithm, could be applied to classify elderly people into groups that are associated with IR. We investigate if the clusters obtained are associated with IR prognosis that could lead to a better understanding of this mechanism in the elderly people.

The importance of early diagnosis of impaired glucose metabolism is based on the need to incorporate patients into prevention or treatment programs in the early stages of the disease. In order to keep elderly people with normal blood glucose levels and to avoid any possible complications. It is for this reason that the objective of this paper is to use the non-supervised k-means clustering algorithm to analyze four methods of estimation of IS and use its cut-off obtained points to develop a Diagnosis System Based on k-Means for IR assesment in eldery people.

II. METHODS

A. Insulin Resistance Methods

The methods used for the quantification of IS are used to establish approximately whether or not a subject has IR. In this work four methods are considered. Two that consider only fasting values (Homa-IR and Quicki) and two that consider glucose metabolism (Avignon and Matsuda). Equations (1) and (2) describe Homa-IR and Quicki methods respectively. These methods use only fasting plasma glucose G_0 and fasting insulin I_0 .

$$\text{Homa-IR} = \frac{G_0 * I_0}{405} \quad (1)$$

$$\text{Quicki} = \frac{1}{\log G_0 + \log I_0} \quad (2)$$

In the same way, equations 3 and 4 define Avignon and Matsuda methods. These methods use fasting and postprandial values for glucose G_{120} and insulin I_{120} .

$$\text{Avignon} = \frac{0.135 * Sib + Si2h}{2 * weight} \quad (3)$$

$$Sib = \frac{10^8}{I_0 * G_0 * 150 \text{ mL/Kg}}$$

$$Si2h = \frac{10^8}{I_{120} * G_{120} * 150 \text{ mL/Kg}}$$

$$\text{Matsuda}_m = \frac{10000}{\sqrt{G_0 * I_0 * (G_0 + \frac{G_{120}}{2}) * (I_0 + \frac{I_{120}}{2})}} \quad (4)$$

Previous studies indicate the cut-off point for the Homa-IR method in 2.5 [6]. Another study carried out on 5518 subjects proposes as a cut-off point for the Quicki index, a value of 0.343 for men and of 0.331 for women with Metabolic syndrome [11]. For the Avignon and Matsuda methods there are not reference cut-off points. Table I presents the reference

values for glucose and insulin at fasting and postprandial levels [4].

TABLE I. CUT-OFF FOR FASTING/POSTPRANDIAL GLUCOSE/ INSULIN IMPAIRED VALUES

Parameter	Cut-off value
Fasting Glucose (G_0)	$< 100 \text{ mg/dL}$
Fasting Insulin (I_0)	$< 20 \mu\text{UI/mL}$
Postprandial Glucose (G_{120})	$< 140 \text{ mg/dL}$
Postprandial Insulin (I_{120})	$< 60 \mu\text{UI/mL}$

B. Database

The sample was taken from participants in the DIUC-PROMETEO project: Heart Rate Variability and Insulin Sensitivity in the elderly people of Cuenca, Ecuador [12]. Study subjects included 119 nondiabetic participants: 42 men ($71.45 \pm 6.69 \text{ yr}$) and 77 women ($71.87 \pm 6, 75 \text{ yr}$). All subjects were carefully interviewed to obtain a detailed social and family history. Weight, height and waist circumference (WC) were measured in subjects wearing light clothes only, and the body mass index (BMI), was calculated as weight divided by height squared (kg/m^2). All procedures were conducted in the morning after a 10 – h fast. Blood tests included a routine biochemistry and two-point OGTT. In the two-point OGTT, a fasting blood sample is taken to measure glucose and basal insulin (G_0, I_0), then the subject ingests 75 grams of glucose and after two hours a new blood sample is taken to measure glucose and insulin postprandial (G_{120}, I_{120}). Baseline characteristics are shown in table II.

TABLE II. DATABASE CLINICAL CHARACTERISTICS

Characteristic $N = 119$	Mean	Range
Height (m)	1.53	[1.02, 1.76]
Weight(Kg)	68.03	[45, 90]
Waist circumference (cm)	93.74	[69, 130]
Body Mass Index (Kg/m^2)	29.15	[19.36, 63.90]
Glucose basal (mg/dL)	90.51	[61.30, 241.10]
Insulin basal ($\mu\text{UI/mL}$)	10.67	[2.00, 43.60]
Glucose postprandial(mg/dL)	102.31	[31.10, 221.60]
Insulin postprandial ($\mu\text{UI/mL}$)	58.11	[4.30, 285.00]

C. K-means clustering

The k-means algorithm groups or predicts data with unsupervised learning. The user specifies the number of clusters k . Each cluster has an average vector (centroid). The Euclidean distance is used to find the cluster closest to the object. Clustering is done iteratively to allocate the object to the nearest cluster. The mean μC with nC number of objects in cluster C contains the mean values of its objects and is defined by the attribute i (equation 5). The iteration stays until the number of objects shifting clusters is lower the threshold specified [13].

$$i = \sum_{j=1}^{nC} \frac{X_{ij}}{nC} \quad (5)$$

The purpose of the algorithm is to divide M observations (number of objects,) from N dimensions (attributes of an object) into K groups. The classification with the k-means algorithm divided the database of 119 subjects into $K = 3$ groups, which we will call C_N , for the cluster with values within what is considered normal, C_A for the cluster with boundary values between impaired and normal and C_{IR} for the cluster susceptible to present IR characteristics, according to the cut-off points obtained for each method as indicated in the table I.

Four experiments were performed corresponding to each method (Homa-IR, Quicki, Avignon and Matsuda). Each method was applied to 119 observations, with dimension $N = 1$ and grouping $K = 3$. The Wilcoxon test was used to determine the difference between groups. A value of $p < 0.01$ was considered to indicate statistical significance.

D. Diagnosis System

Using cut-off values obtained in [10] for $K = 2$ and $K = 3$, an application was developed using Matlab.

III. RESULTS

A. Cut-off values

The numbers of subjects in each clusters and cut-off values are shown in table III. The characteristics of each cluster according to their fasting (G_0, I_0) and postprandial (G_{120}, I_{120}) values and waist circumference (WC) are presented in table IV. Figure 1 shows the cluster assignments and centroids for $K = 3$ for each method considered.

In the table IV the cut-off points obtained for each method are presented for $K = 3$, also the cut-off and population distribution with $K = 2$ are shown as reference [10]. Analyzing the clusters according to the average values of glucose (G_0, G_{120}) and insulin (I_0, I_{120}) from the table IV the following is observed:

Homa-IR

In this method, the average values of C_N are within the normal ranges, for C_A the average values of G_{120} and I_{120} are altered and in the group C_{IR} , all values are altered, but this last cluster only groups 6 subjects. It is also observed that the values of WC and BMI are altered in the Cluster C_A . With respect to the classification with $K = 2$, it can be considered that the C_A is closer to the subjects with IR than the normal subjects.

Quicki

The average values of C_N and C_A are within the normal ranges, although the values of C_A are higher. For the cluster C_{IR} , all values are altered except for I_0 . The group C_{IR} has significantly different values for glucose (G_0, G_{120}) and for insulin (I_0, I_{120}) with respect to the groups C_N and C_A : although postprandial values are not considered for the estimation of this method. In this case the group

C_A can be considered as a risk group of C_{IR} . The WC of the group C_{IR} is significantly greater than that of the group C_N .

Avignon

The average values of C_N and C_A are within the normal ranges, although the values of C_A are higher. For the cluster C_{IR} , G_{120} and I_{120} are altered. The postprandial values (G_{120}, I_{120}) allow to significantly separate the three groups. In addition, the group C_{IR} has values of WC and BMI greater than the groups C_N and C_A . As for the above case, the group C_A can be considered as a risk group suffering IR.

Matsuda

As with the Avignon method, average values of C_N and C_A are within the normal ranges, although the values of C_A are higher. For the cluster C_{IR} , only the average of I_{120} is altered. The basal insulin I_0 and the postprandial values (I_{120}, G_{120}) allow to significantly separate the three groups. The WC of the group C_{IR} is significantly higher with respect to the other two clusters. Like the Quicki and Avignon methods, the C_A group can be considered as a risk group of IR.

The k-means allows to separate the population in three well differentiated clusters according to the high values of glucose and insulin using methods of IS estimation. The interesting thing is that for each method the number of individuals that could be diagnosed as insulin resistant or with IR is very different.

Although in all cases the group C_A seems like a group risking to presents characteristics of suffering IR in the case of the Homa-IR method the distribution of the subjects in each group is quite different from the rest of the methods. In particular Homa-IR methods classifies the largest number (65.54%) of subjects in the group C_N unlike the others methods where the number of individuals in this class does not exceed 16 %. Also in relation to $K = 2$, we see that the distribution remains practically the same, for C_N there are 80 subjects and 78 with in $K = 3$; and there are only 6 subjects in the C_{IR} (table III).

For the other three methods C_A cluster are composed of subjects who were classified as C_N and C_{IR} using $K = 2$ (Figure 1).

Matsuda and Avignon differentiate the three clusters by their postprandial values (I_{120}, G_{120}), although it is also observed that all values of the OGTT are higher for the cluster C_{IR} . It appears that clustering for these methods includes as C_{IR} those individuals who have altered values of I_{120} , these individuals do not necessarily have altered fasting values, and in previous methods (Quicki and Homa-IR) are not detected as C_{IR} . The methods that take into account the I_{120} and G_{120}

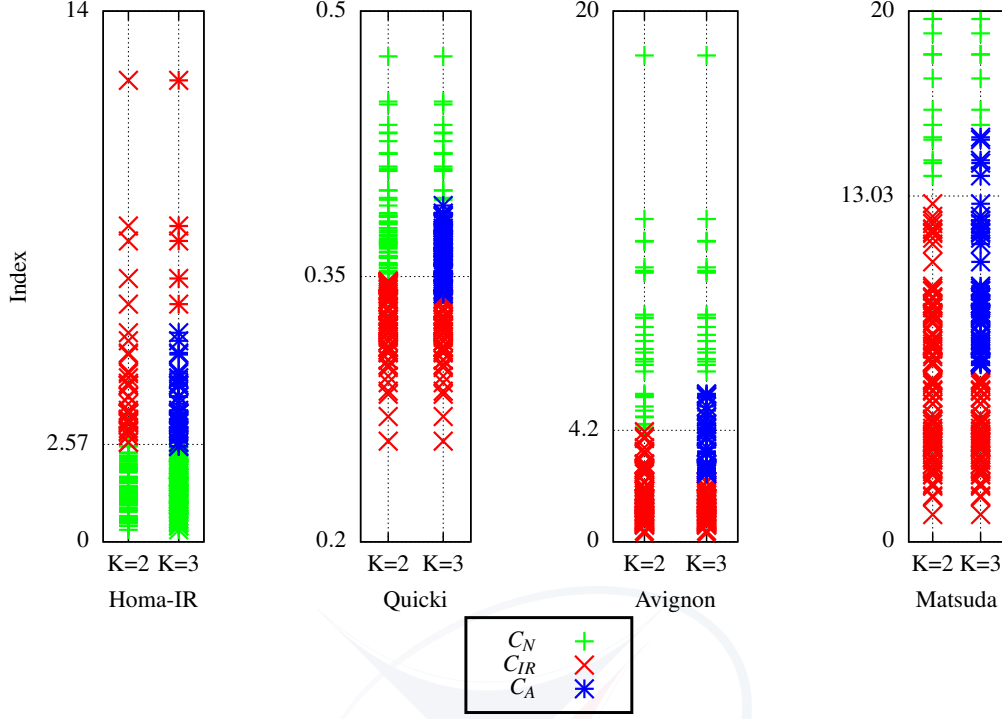


Fig. 1. Assignment of individuals to clusters for $K = 2$ and $K = 3$, using Homa-IR, Quicki, Avignon and Matsuda's observations. Cluster centroids are represented by the character X

TABLE III. NUMBER OF SUBJECT IN EACH CLUSTERS (C_N , C_A AND C_{IR}) AND CUT-OFF VALUES

Method	$K = 2$	C_N $n(\%)$	C_{IR} $n(\%)$	$K = 3$		C_N $n(\%)$	C_A $n(\%)$	C_{IR} $n(\%)$
	cut-off			cut-off ₁	cut-off ₂			
Homa-IR	≥ 2.57	80(67.27)	39(32.77)	≥ 2.35	> 5.51	78(65.54)	35(29.41)	6(5.04)
Quicki	≤ 0.34	59(49.38)	60(50.42)	≤ 0.34	< 0.39	17(14.29)	56(47.06)	46(38.66)
Avignon	≤ 4.20	24(20.17)	95(79.83)	≤ 2.27	< 5.58	19(15.97)	51(42.86)	49(41.18)
Matsuda	≤ 13.03	31(26.05)	88(73.95)	≤ 6.14	< 15.25	19(15.97)	33(27.73)	67(56.30)

values could be more effective detecting groups with risk of suffering from IR.

Methods that include all glucose and insulin values place a greater number of individuals in the group with altered values (C_{IR}). This may be due to the fact that these methods, when including the basal and postprandial values, detect those cases where the basal values are normal but have altered postprandial values. These particular cases are people who have a metabolic dysfunction that is not yet reflected in the fasting values, there are studies indicating that fasting values are the last to be altered when IR is present [14].

B. Diagnosis System

Taking the cut-off values for $K = 2$ and $K = 3$, the aide Diagnosis System for IR assessments was designed. In the Figure 2 the system is shown. At the left the user enters the patient data. On the right side, the patient, is classified according to the four methods. The IS values for each method are shown on the lower left side. The area of each defined cluster is marked in a color: red, blue and green for C_{IR} , C_A

and C_N respectively. The cut-off value with $K = 2$ is marked by the black dotted line.

In this example, the Diagnosis System shows the results for 80 years old male patient. It can be seen that the subject presents normal IS values for Homa-IR ($IS = 2.47$) but for the others three methods (Quicki, Avignon y Matsuda) this patient presents IR in the normal/risk regions, also he present G_0 altered. There is a ratio of 3 to 1 methods that specify RI and the value of G_0 is representative of pre-diabetes. So, in addition to numerical values it is necessary to incorporate the clinical analysis of each method (section III-A) to perform the diagnosis of IR. The Diagnosis System is under evaluation by our clinical team.

IV. CONCLUSION

The k-means method has allowed us to study four surrogate methods to estimate IS in an elderly population. With the cut-off values obtained, for each method, the system for the evaluation of IR in elderly people were developed. Our work is intended to aid physicians in the early detection of IR

TABLE IV. GLUCOSE (mg/dL), INSULINE (UI/mL) AND WAIST CIRCUMFERENCE (WC, cm) FOR EACH CLUSTER (C_N , C_A AND C_{IR}) BY EACH METHOD

Method	Glucose, insulin WC	C_N	Clusters C_A	C_{IR}	(C_N, C_A)	p value (C_N, C_{IR})	(C_A, C_{IR})
Homa-IR	G_0	84.63 ± 16.09	98.70 ± 31.46	119.23 ± 47.27	$< 10^{-3}$	N.S.	N.S.
	I_0	6.51 ± 2.61	15.77 ± 5.32	34.95 ± 9.07	$< 10^{-16}$	$< 10^{-5}$	$< 10^{-4}$
	G_{120}	95.96 ± 29.56	115.54 ± 34.31	107.68 ± 33.37	< 0.02	N.S.	N.S.
	I_{120}	38.97 ± 21.59	96.51 ± 50.63	82.93 ± 75.37	$< 10^{-11}$	N.S.	N.S.
	WC(cm)	92.42 ± 11.20	96.49 ± 11.23	98.83 ± 7.65	N.S.	N.S.	N.S.
	BMI(kg/m ²)	28.55 ± 4.48	30.60 ± 6.72	28.52 ± 4.25	N.S.	N.S.	N.S.
Quicki	G_0	81.06 ± 10.60	85.13 ± 17.84	100.56 ± 32.33	N.S.	$< 10^{-3}$	$< 10^{-4}$
	I_0	3.02 ± 0.60	7.24 ± 1.96	17.67 ± 8.91	$< 10^{-10}$	$< 10^{-9}$	$< 10^{-16}$
	G_{120}	93.89 ± 25.24	97.54 ± 30.84	111.23 ± 34.54	N.S.	N.S.	N.S.
	I_{120}	23.54 ± 13.18	43.22 ± 22.05	89.00 ± 53.58	$< 10^{-4}$	$< 10^{-7}$	$< 10^{-7}$
	WC(cm)	87.82 ± 6.04	93.72 ± 12.41	95.93 ± 10.30	N.S.	$< 10^{-3}$	N.S.
	BMI(kg/m ²)	26.92 ± 4.43	29.24 ± 4.32	29.88 ± 6.37	0.04	0.03	N.S.
Avignon	G_0	78.73 ± 11.87	89.35 ± 20.44	94.43 ± 28.91	N.S.	$< 10^{-3}$	N.S.
	I_0	7.16 ± 2.61	6.64 ± 4.82	13.64 ± 7.90	N.S.	$< 10^{-6}$	$< 10^{-9}$
	G_{120}	73.38 ± 19.34	91.06 ± 23.42	116.06 ± 31.45	$< 10^{-3}$	$< 10^{-7}$	$< 10^{-4}$
	I_{120}	14.24 ± 6.92	32.78 ± 11.24	83.02 ± 44.57	$< 10^{-7}$	$< 10^{-11}$	$< 10^{-13}$
	WC(cm)	87.97 ± 7.27	88.82 ± 10.94	97.79 ± 10.54	N.S.	$< 10^{-4}$	$< 10^{-5}$
	BMI(kg/m ²)	28.97 ± 3.85	27.50 ± 3.80	30.93 ± 5.54	N.S.	$< 10^{-5}$	$< 10^{-4}$
Matsuda	G_0	78.06 ± 12.35	87.10 ± 17.10	98.89 ± 32.41	N.S.	$> 10^{-4}$	N.S.
	I_0	3.86 ± 1.69	7.16 ± 2.35	16.96 ± 9.04	$< 10^{-7}$	$< 10^{-10}$	$< 10^{-14}$
	G_{120}	79.00 ± 18.84	95.81 ± 26.72	118.12 ± 34.00	$< 10^{-3}$	$< 10^{-6}$	$< 10^{-4}$
	I_{120}	16.35 ± 7.46	40.13 ± 16.69	93.01 ± 48.10	$< 10^{-8}$	$< 10^{-9}$	$< 10^{-13}$
	WC(cm)	89.03 ± 6.17	91.82 ± 12.00	97.55 ± 10.64	N.S.	$< 10^{-4}$	$< 10^{-3}$
	BMI(kg/m ²)	27.02 ± 3.98	28.97 ± 4.78	30.17 ± 5.97	N.S.	0.01	N.S.

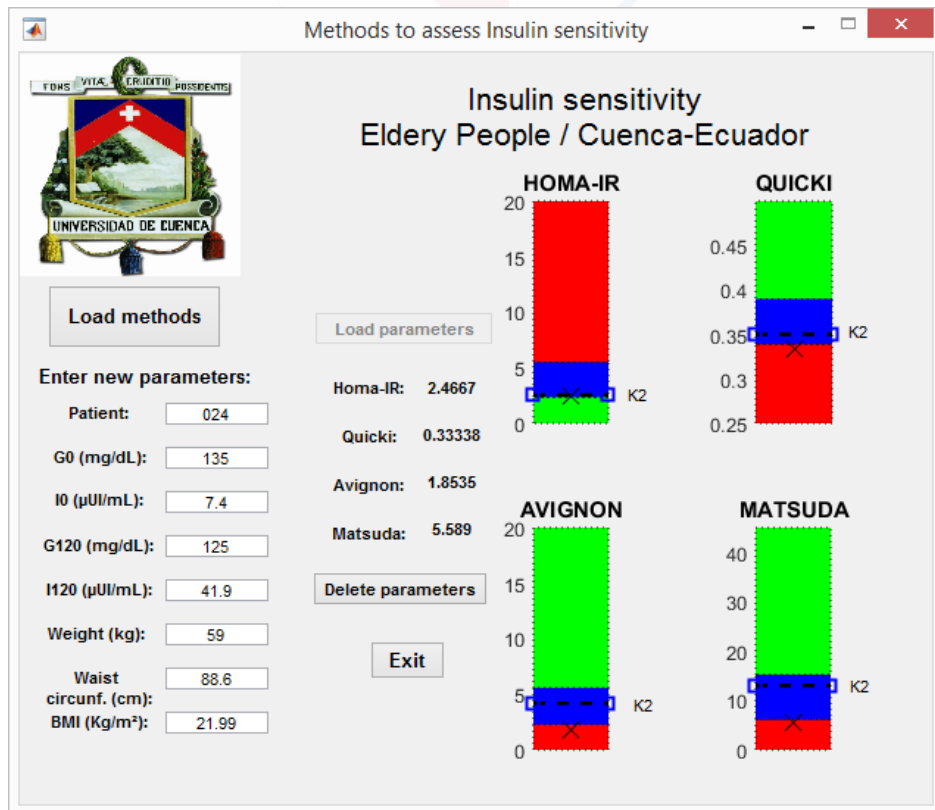


Fig. 2. Diagnosis System. Homa-IR, Quicki, Avignon and Matsudas values for a new patient are represented by the character X

by using information from diverse methods. The developed system allows the physician to diagnose IR based on four methods.

A system that appropriately classifies individuals with IR is paramount from the point of view of public health, IR increases the risk of suffering type 2 diabetes and its early identification could prevent and delay the appearance of irreversible pathological entities.

ACKNOWLEDGMENT

This work has been partially funded by the Research Direction of the University of Cuenca, Ecuador (DIUC – Dirección de Investigación de la Universidad de Cuenca, Ecuador).

REFERENCES

- [1] Crecimiento acelerado de la población adulta de 60 años y más de edad: Reto para la salud pública. [Online]. Available: http://www.paho.org/hq/index.php?option=com_content&view=article&id=2796poblacion-adulta-60-anos-mas-edad-retosaludpublica
- [2] T. G. Heck, M. S. Ludwig, A. B. dos Santos, and P. B. Goettems-Fiorin, "Lifestyle and aging effects in the development of insulin resistance activating the muscle as strategy against insulin resistance by modulating cytokines and hsp70," in *Muscle Cell and Tissue*. InTech, 2015.
- [3] E. M. Morales and M. V. V. Fornieri, "Consideraciones clínicas sobre la importancia de la sensibilidad a la insulina, su resistencia y la intolerancia a la glucosa," *Revista de la ALAD*, 2007.
- [4] C. R. Kahn, "Insulin resistance, insulin insensitivity, and insulin unresponsiveness: a necessary distinction," *Metabolism*, vol. 27, no. 12, pp. 1893–1902, 1978.
- [5] E. Severeyn, S. Wong, G. Passariello, J. L. Cevallos, and D. Almeida, "Methodology for the study of metabolic syndrome by heart rate variability and insulin sensitivity," *Revista Brasileira de Engenharia Biomédica*, vol. 28, no. 3, pp. 272–277, 2017.
- [6] N. Friedrich, B. Thuesen, T. Jørgensen, A. Juul, C. Spielhagen, H. Wallaschofski, and A. Linneberg, "The association between igf-i and insulin resistance," *Diabetes care*, vol. 35, no. 4, pp. 768–773, 2012.
- [7] H.-Q. Qu, Q. Li, A. R. Rentfro, S. P. Fisher-Hoch, and J. B. McCormick, "The definition of insulin resistance using homa-ir for americans of mexican descent using machine learning," *PloS one*, vol. 6, no. 6, p. e21041, 2011.
- [8] M. Altuve, E. Severeyn, and S. Wong, "Adaptation of five indirect insulin sensitivity evaluation methods to three populations: Metabolic syndrome, athletic and normal subjects," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*. IEEE, 2014, pp. 4555–4558.
- [9] —, "Unsupervised subjects classification using insulin and glucose data for insulin resistance assessment," in *Signal Processing, Images and Computer Vision (STSIVA), 2015 20th Symposium on*. IEEE, 2015, pp. 1–7.
- [10] C. Vintimilla, D. Astudillo, E. Severeyn, L. Encalada, and S. Wong, "Agrupamiento de k-medias para estimación de insulino-resistencia en adultos mayores de cuenca," *Maskana*, no. In press., 2017.
- [11] N. Motamed, S. J. H. Miresmail, B. Rabiee, H. Keyvani, B. Farahani, M. Maadi, and F. Zamani, "Optimal cutoff points for homa-ir and quicki in the diagnosis of metabolic syndrome and non-alcoholic fatty liver disease: A population based study," *Journal of diabetes and its complications*, vol. 30, no. 2, pp. 269–274, 2016.
- [12] F. Parra, D. Andrade, J. Cruz, L. Solano-Quinde, K. Palacio-Baus, L. Encalada, and S. Wong, "Plataforma basada en ecgml para el estudio de las complicaciones cardiovasculares en el adulto mayor con síndrome metabólico," *Maskana*, vol. 6, no. Supl., pp. 157–164, 2015.
- [13] B. Andreopoulos, A. An, X. Wang, and M. Schroeder, "A roadmap of clustering algorithms: finding a match for a biomedical application," *Briefings in Bioinformatics*, vol. 10, no. 3, pp. 297–314, 2009.
- [14] J. Velásquez, E. Severeyn, H. Herrera, L. Encalada, and S. Wong, "Anthropometric index for insulin sensitivity assessment in older adults from ecuadorian highlands," in *12th International Symposium on Medical Information Processing and Analysis*. International Society for Optics and Photonics, 2017, pp. 101 600S–101 600S.



UNIVERSIDAD DE CUENCA
desde 1867



Apéndice C

Artículo: Normalización de entradas de una red de Kohonen para clasificación de Síndrome Metabólico en Adultos Mayores de la ciudad de Cuenca

UNIVERSIDAD DE CUENCA
desde 1867

Normalización de entradas de una red de Kohonen para clasificación de Síndrome Metabólico en Adultos Mayores de la ciudad de Cuenca

Christian Vintimilla¹, Darwin Astudillo¹, Erika Severeyn^{2,3}, Lorena Encalada⁴, Sara Wong^{1,2}

¹ Departamento de Ingeniería Eléctrica, Electrónica y Telecomunicaciones, Universidad de Cuenca, Av. 12 de Abril y Agustín Cueva, Cuenca, Ecuador, 010201

² Grupo de Bioingeniería y Biofísica Aplicada, Universidad Simón Bolívar, Venezuela

³ Departamento de Tecnología de Procesos Biológicos y Bioquímicos, Universidad Simón Bolívar, Venezuela

⁴ Facultad de Ciencias Médicas, Universidad de Cuenca, Av. 12 de Abril y Paraíso, Cuenca, Ecuador, 010204

Autores para correspondencia: xavier.vintimilla92@ucuenca.ec, {fabian.astudillos, lorena.encalada, sara.wong}@ucuenca.edu.ec, {severeynrika, swong}@usb.ve

Fecha de recepción: 1 de septiembre 2017 - Fecha de aceptación: 1 de septiembre 2017

ABSTRACT

One of the challenges in using Kohonen self-Organizing maps (SOM) is the pre-processing or normalization of inputs variables. In the present work two pre-processing techniques (binary and by ranges) of the variables to diagnose Metabolic Syndrome (MS) in older adults of the urban districts of Cuenca are explored. Three experiments were carried out considering the entire population ($N = 387$) and dividing the population by sex, in each experiment 3 clusters were defined. The results, using pre-processing by ranges allow a better classification of the population in all cases. This study allowed us to select the type of pre-processing for the diagnosis of MS in the Elderly population of the city of Cuenca using SOM.

Keywords: Neuronal Networks, Kohonen, Metabolic Syndrome, pre-processing, SOM

RESUMEN

Uno de los desafíos al usar mapas auto-organizativos de Kohonen (SOM) es el pre-procesamiento o normalización de las variables de entrada. En el presente trabajo se exploran dos técnicas pre-procesamiento (binaria y por rangos) de las variables para diagnosticar Síndrome Metabólico (SM) en adultos mayores de las parroquias urbanas de Cuenca. Se realizaron tres experimentos: considerando toda la población ($N=387$) y dividiendo la población por sexos, en cada experiencia se definieron 3 clústeres. Los resultados, usando un pre-procesamiento por rangos permiten una mejor clasificación de

la población en todos los casos. Este estudio ha permitido seleccionar el tipo de pre-procesamiento para el diagnóstico de SM en la población de Adultos Mayores (AM) de la ciudad de Cuenca usando SOM.

Palabras clave: Redes Neuronales, Kohonen, Síndrome Metabólico, pre-procesamiento, SOM

1. INTRODUCCIÓN

El algoritmo de los mapas auto-organizativos de Kohonen SOM (Self-Organizing Map) se basa en un proceso iterativo de comparación con un conjunto de datos y cambios para aproximarse a los mismos, crea un modelo de esos mismos datos para agruparlos por criterios de similitud (Kohonen, & Honkela, 2007). Los mapas de Kohonen han sido usados para numerosas aplicaciones médicas, entre ellas para el diagnóstico de diabetes y síndrome metabólico (SM) (Kohonen, & Honkela, 2007; Isasi & Galván, 2004)

El síndrome metabólico es conocido como precursor de la Diabetes tipo 2 y de enfermedades cardiovasculares (Klein BE, Klein R, Lee, 2002). Según los criterios del Programa Nacional de Educación sobre el Colesterol y el Panel de Tratamiento del Adulto (Grundy et al., 2005) se diagnostica SM, cuando un sujeto presenta al menos tres de las siguientes cinco condiciones:

- i) Glucosa en ayunas alterada (> 100 mg/dl),
- ii) Altos valores de triglicéridos (≥ 150 mg/dl),
- iii) Elevada presión arterial sistólica y diastólica ($\geq 130/85$)
- iv) Bajos niveles de C-HDL, C-HDL < 40 mg/dl para los hombres y < 50 mg/dl para mujeres y
- v) Circunferencia Abdominal (CA) aumentada (CA > 90 cm en hombres y > 80 cm en mujeres).

Uno de los desafíos al usar SOM es el pre-procesamiento o normalización de las variables de entrada. Existen varias metodologías: usando categorías o rangos, normalizando restando la media y dividiendo por la desviación estándar, aplicando logaritmos cuando el rango de variación ocupa varios ordenes de magnitud o la codificación binaria (Montaño Moreno, 2002).

En este trabajo se exploran el uso de rangos y la codificación binaria en el pre-procesamiento de las variables para clasificar SM en adultos mayores de las parroquias urbanas de Cuenca usando mapas auto-organizativos de Kohonen.

2. MATERIALES Y MÉTODOS

2.1 Modelo de Kohonen

A inicios de 1980 Kohonen demostró que un conjunto de datos de entrada puede ordenarse por sí solo, de acuerdo a un modelo de mapas topológicos. Este modelo busca establecer una correlación entre los datos de entrada y el espacio de salida de dos dimensiones (llamado mapa topológico). Esta correlación de entrada/salida se indica mediante la activación de zonas en el mapa de salida. La estructura topológica consta de una capa de entrada y una capa de salida, que consiste en k neuronas de entrada

(que es el conjunto de características o vector de parámetros de un dato de entrada), mientras en la capa de salida llamada capa de competición como se indica en la Figura 1.

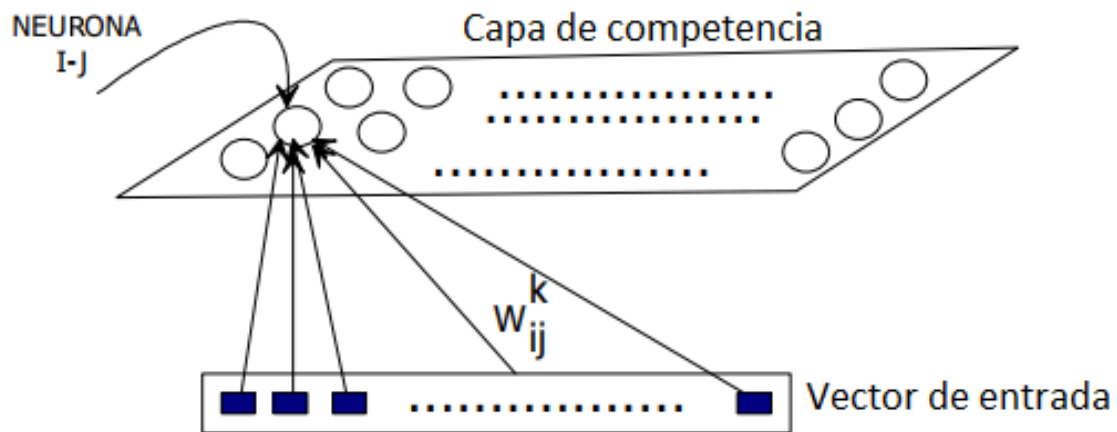


Figura 1. Arquitectura del Modelo de Kohonen (Serrano, Soria, Martin, 2009)

Cada patrón de entrada está conectado con todas las neuronas de la capa de competencia mediante pesos sinápticos (Serrano, Soria, Martin, 2009). A cada una de las neuronas de la capa de competencia se le asigna un vector de pesos como se indica en la Figura 1, de dimensión del vector de entrada.

$$W_{ij} = [W_{ij}^1 \quad W_{ij}^2 \quad W_{ij}^3 \quad \dots \quad W_{ij}^k]$$

La red busca que patrones de entrada similares activen neuronas próximas en la capa de competencia. Esto se realiza cuando los pesos en la capa de salida asociados a esa entrada son semejantes a ese patrón de entrada. Esta semejanza se consigue con medidas de similitud, las más utilizada es la Distancia Euclídeana.

El algoritmo del modelo de Kohonen es el siguiente:

- Paso 1)** Inicializar los pesos, se inicializan aleatoriamente.
- Paso 2)** Presentar un patrón de entrada en cada iteración.
- Paso 3)** Determinar la neurona ganadora (mayor similitud del vector de pesos y la entrada), empleando la Distancia Euclídeana u otras funciones.

$$\text{Ecuacion 1:} \quad d(W_{ij}, W_R) = \sqrt{\sum_{k=1}^M (W_{ij}^k - x^k)^2}$$

- Paso 4)** Actualizar los pesos sinápticos, en la neurona ganadora y el vecindario:

$$\text{Ecuacion 2:} \quad w_{ij}(n+1) = w_{ij}(n) + \alpha(n) h(n) (x_j - w_{ij})$$

$\alpha(n)$: es la tasa de aprendizaje. Fija la velocidad de cambio de los pesos. Se establece en función del número de iteraciones o se establece como un valor constante.

$h(n)$: es la función de vecindad (con valor máximo en la neurona ganadora).

Paso 5) Volver al paso 2, si no se han realizado todas las iteraciones.

2.2 Base de Datos

Se estudiaron 387 adultos mayores de 65 años de las parroquias urbanas de Cuenca. Se excluyeron sujetos con deficiencia mental, alteración del estado de conciencia, impedimento físico para movilizarse y diabéticos, una descripción más amplia de la población se encuentra disponible en el artículo sobre Actividad Física, en Adultos Mayores de las parroquias urbanas de Cuenca-Ecuador, 2015 (Chimbo, Chuchuca, Encalada, Wong, 2015).

Para cada sujeto se utilizan como parámetros de entrada las seis variables usadas para diagnosticar SM, la presión arterial se consideró como una variable doble.

2.3 Pre-procesamiento de los datos

Se han considerado dos tipos de normalización: binaria y por rangos, en cada caso se consideran 5 clases determinadas a partir de los quintiles del rango de la variable. Los valores de corte para cada variable del SM se muestran en la tabla 1.

1. **Normalización Binaria:** las entradas son vectores de 5 elementos, se asigna un uno a la posición de la clase de la variable.
2. **Normalización por rangos:** se asigna un valor [0,2 0,4 0,6 0,8 1] en función de la clase en el cual se encuentre la variable, tal como se indica en la tabla 2.

Tabla 1. Valores de Corte para la base de datos de SM (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAD: Presión Arterial diastólica, HDL: Lipoproteínas de alta densidad).

Características	Valores de Corte					
	Mínimo	1	2	3	4	Máximo
PAS(mmHg)	90	110	130	150	170	190
PAD (mmHg)	50	62	74	86	98	110
CA (cm)	68,0	80,4	92,8	105,2	117,6	130,0
Glucosa Basal (mg/dl)	49,10	83,28	117,47	151,66	185,85	220,04
Triglicéridos (mg/dl)	54,60	158,80	263,00	367,20	471,40	575,60
HDL (mg/dl)	10,90	34,68	58,46	82,24	106,02	129,80

Tabla 2. Normalización por Intervalo.

Valor entrada	(Min-1)	(1-2)	(2-3)	(3-4)	(4-Max)
Valor normalizado	0,2	0,4	0,6	0,8	1

2.4 Experimentos

Se realizaron tres experimentos, considerando en cada caso los dos tipos de pre-procesamiento y las seis variables del SM, en cada experiencia se definieron tres clústeres.

- Toda la población: 270 AM para aprendizaje y 117 para validación
- Solo población femenina: 170 para aprendizaje y 75 para validación.
- Solo población masculina: 100 para aprendizaje y 42 para validación.

3. RESULTADOS

Resultados para toda la población

La figura 2 muestra la distribución de los clústeres según el tipo de pre-procesamiento. En las tablas 3 y 4 se presentan las características de los tres clústeres obtenidos para la normalización binaria y por rangos respectivamente para el caso del experimento 1 (N=117).

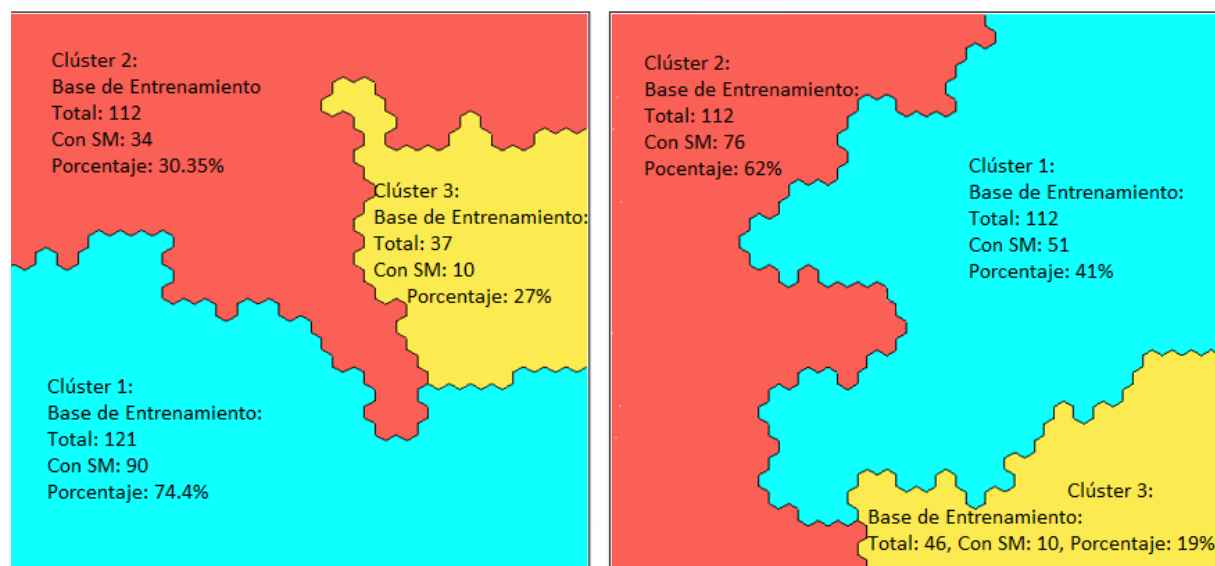


Figura 2. A) SOM usando pre-procesamiento binario. B) SOM usando pre-procesamiento por rangos

Tabla 3. Características de los clústeres para toda la población usando normalización binaria (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAS: Presión Arterial diastólica, HDL: Lipoproteínas de alta densidad).

Validación N=117	Clúster 1 (N=33)	Clúster 2 (N=63)	Clúster 3 (N=21)			
Con SM	28(86,36%)	27(42,37 %)	10(47,62%)			
Características	Media \pm std			p12	p13	p23
PAS (mmHg)	137,48 \pm 14,34	123,46 \pm 8,73	112,57 \pm 12,04	0,001	0,001	0,001
PAD (mmHg)	84,73 \pm 8,54	77,67 \pm 6,69	69,29 \pm 7,79	0,001	0,001	0,001
CA (cm)	95,71 \pm 11,41	94,57 \pm 11,66	94,67 \pm 9,87	0,841	0,972	0,832
Glucosa (mg/dl)	92,26 \pm 23,41	93,47 \pm 27,76	93,45 \pm 27,34	0,951	0,716	0,844
Triglicéridos (mg/dl)	147,65 \pm 79,66	176,90 \pm 109,78	155,17 \pm 71,28	0,247	0,546	0,642
HDL (mg/dl)	38,04 \pm 6,67	43,71 \pm 18,39	43,91 \pm 8,34	0,157	0,005	0,173

Tabla 4. Características de los clústeres para toda la población usando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAD: Presión Arterial diastólica, HDL: Lipoproteínas de alta densidad).

Validación N=117	Clúster 1 (n=85)	Clúster 2 (n=31)	
Con SM	41 (48,23%)	22 (70,97 %)	
Características	Media ± std		p12
PAS (mmHg)	120,14±11,01	140,35±10,51	0,000
PAD (mmHg)	75,65±8,05	84,65±8,15	0,000
CA (cm)	93,54±10,32	98,82±12,68	0,076
Glucosa (mg/dl)	90,47±25,22	100,52±28,61	0,040
Triglicéridos (mg/dl)	168,55±95,88	154,02±99,82	0,295
HDL (mg/dl)	39,77±8,67	46,23±18,51	0,080

En el caso del pre-procesamiento binario (tabla 3) se observa que las poblaciones son diferentes entre ellas en relación a la presión arterial. Se observan tres clústeres con valores de presión arterial diastólica y sistólica bien diferenciados. Para el resto de las características del SM no se presentan diferencias entre los clústeres.

En la tabla 4, se observa que la mayor parte de la población se divide entre los clústeres 1 y 2. El clúster 3 solo tiene un AM. El clúster 1 presenta valores promedios normales para todas las variables a excepción de los triglicéridos. La mayoría de valores promedios del clúster 2 se encuentran alterados o en el límite del valor corte de diagnóstico. Se observa que los sujetos agrupados en este grupo presentan características más relacionadas con el SM y por ende un porcentaje mayor de AM con SM. Estas mismas tendencias fueron observadas para la población femenina entre los clústeres 1 y 2 (tabla 5), el clúster 3 agrupo 10 AM sin SM. Igualmente los clústeres 1 y 2 de población de AM masculina siguen la misma tendencia observada en las experiencias anteriores, en este caso el clúster 3 agrupa 1 solo sujeto con valores de glicemia y triglicéridos claramente muy elevados.

Tabla 5. Características de los clústeres para la población femenina usando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAD: Presión Arterial diastólica, HDL: Lipoproteínas de alta densidad).

Validación N=75	Clúster 1 (n=45)	Clúster 2 (n=20)	Clúster 3 (n=10)			
Con SM	35 (77.78%)	9 (45 %)	0 (0%)			
Características	Media ± std			p12	p13	p23
PAS (mmHg)	124,89±13,02	136,00±10,96	111,00±8,76	0,002	0,003	0,001
PAD (mmHg)	75,29±6,86	88,00±6,81	68,00±4,22	0,001	0,002	0,001
CA (cm)	99,51±9,40	86,78±7,60	89,30±9,66	0,001	0,007	0,300
Glicemia (mg/dl)	102,77±33,26	84,60±12,02	81,72±8,47	0,071	0,038	0,416
Triglicéridos (mg/dl)	187,98±111,32	125,39±63,03	133,86±42,38	0,016	0,321	0,416
HDL (mg/dl)	38,86±7,78	46,58±21,70	45,59±11,10	0,172	0,046	0,495

Tabla 6. Características de los clústeres para la población masculina usando normalización por rangos (CA: Circunferencia Abdominal, PAS: Presión Arterial Sistólica, PAD: Presión Arterial diastólica, HDL: Lipoproteínas de alta densidad).

Validación N=42	Clase 1 (n=23)	Clase 2 (n=18)	Clase 3 (n=1)			
Con SM	3 (13.04%)	12 (66.67 %)	1 (100%)			
Características	Media \pm std			p12	p13	p23
PAS (mmHg)	115,87 \pm 6,36	135,17 \pm 13,85	135,00 \pm 0,00	0,001	0,092	0,632
PAD (mmHg)	75,61 \pm 7,38	84,00 \pm 7,00	65,00 \pm 0,00	0,001	0,133	0,105
CA (cm)	89,30 \pm 12,13	102,72 \pm 7,81	87,00 \pm 0,00	0,001	0,828	0,105
Glicemia (mg/dl)	78,58 \pm 7,37	96,67 \pm 16,17	197,80 \pm 0,00	0,001	0,112	0,105
Triglicéridos (mg/dl)	164,15 \pm 80,53	160,05 \pm 113,48	278,20 \pm 0,00	0,554	0,248	0,211
HDL (mg/dl)	45,10 \pm 20,22	40,57 \pm 9,67	35,90 \pm 0,00	0,590	0,563	0,632

4. CONCLUSIONES

Los resultados usando un pre-procesamiento por rangos permiten una mejor clasificación de la población con SM. La selección del método de pre-procesamiento de los datos es de suma importancia para la utilización de SOM. Este estudio ha permitido escoger el tipo de pre-procesamiento para el diagnóstico de SM en la población de AM de la ciudad de Cuenca. Los trabajos futuros están orientados a experimentar con diferentes números de clases para la normalización por rangos y de clústeres para el diagnóstico de SM en AM usando SOM.

AGRADECIMIENTOS

Este trabajo fue realizado gracias al financiamiento de la Dirección de Investigación de la Universidad de Cuenca (DIUC).

REFERENCIAS

- Chimbo JC, Chuchuca, AJ, Encalada LE, Wong S. Nivel de actividad física medida a través del Cuestionario Internacional de Actividad Física, en Adultos Mayores de las parroquias urbanas de Cuenca-Ecuador, 2015. Revista de la Facultad de Ciencias Médicas, Universidad de Cuenca. 2016; 34(2), 51-56.
<http://dx.doi.org/10.4067/S0034-98872010001100004>
- Grundy, S. M., Cleeman, J. I., Daniels, S. R., Donato, K. A., Eckel, R. H., Franklin, B. A., ... & Spertus, J. A. Diagnosis and management of the metabolic syndrome. Circulation 2005; 112(17), 2735-2752
<https://doi.org/10.1161/CIRCULATIONAHA.105.169404>
- Isasi Viñuela, P., & Galván León, I. M. (2004). Redes de neuronas artificiales. Un Enfoque Práctico, Editorial Pearson Educación SA Madrid España.
- Klein BE, Klein R, Lee KE. Components of the metabolic syndrome and risk of cardiovascular disease and diabetes in Beaver Dam. Diabetes care 2002; 25(10), 1790-1794. PMID:12351479

<https://doi.org/10.2337/diacare.25.10.1790>

Kohonen, T., & Honkela, T. (2007). Kohonen network. Scholarpedia, 2(1), 1568.

<https://doi.org/10.4249/scholarpedia.1568>

Montaño Moreno, J. J. (2002). Redes neuronales artificiales aplicadas al análisis de datos.

Serrano, A. J., Soria, E., & Martin, J. D. (2009). Redes neuronales artificiales. Universidad de Valencia (Escuela Técnica Superior Ingeniería, Departamento de Ingeniería Electrónica): Valencia, España.





UNIVERSIDAD DE CUENCA
desde 1867

Bibliografía

- [1] M. de Salud Pública, *Guía de Práctica Clínica (GPC) de Diabetes mellitus tipo 2*. Dirección Nacional de Normatización, 2017, num. 1.
- [2] A. Makroglou, J. Li, y Y. Kuang, “Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview,” *Applied numerical mathematics*, vol. 56, num. 3-4, pp. 559–573, 2006.
- [3] A. J. Serrano, E. Soria, y J. D. Martín, “Redes neuronales artificiales,” *Universidad de Valencia (Escuela Técnica Superior Ingeniería, Departamento de Ingeniería Electrónica): Valencia, España*, 2009.
- [4] E. Severeyn, S. Wong, G. Passariello, J. L. Cevallos, y D. Almeida, “Methodology for the study of metabolic syndrome by heart rate variability and insulin sensitivity,” *Revista Brasileira de Engenharia Biomédica*, vol. 28, num. 3, pp. 272–277, 2012.
- [5] M. L. Gimeno, C. B. Martínez, I. P. Calleja, y J. A. C. Lenguas, “Síndrome metabólico. concepto y fisiopatología,” *Revista Española de Cardiología Suplementos*, vol. 5, num. 4, pp. 3D–10D, 2005.
- [6] F. Parra, D. Andrade, J. Cruz, L. Solano-Quinde, K. Palacio-Baus, L. Encalada, y S. Wong, “Plataforma basada en ecgml para el estudio de las complicaciones cardiovasculares en el adulto mayor con síndrome metabólico,” *Maskana*, vol. 6, num. Supl., pp. 157–164, 2015.
- [7] J. Velásquez, E. Severeyn, H. Herrera, L. Encalada, y S. Wong, “Anthropometric index for insulin sensitivity assessment in older adults from ecuadorian highlands,” in *12th International Symposium on Medical Information Processing and Analysis*. International Society for Optics and Photonics, 2017, pp. 101 600S–101 600S.
- [8] G. Crepaldi y S. Maggi, “El síndrome metabólico: contexto histórico,” *Diabetes Voice.*, vol. 51, num. especial, 2006.



- [9] P. Zimmet, M. Alberti, K. George, y M. Serrano Ríos, “Una nueva definición mundial del síndrome metabólico propuesta por la federación internacional de diabetes: fundamento y resultados,” *Revista española de cardiología*, vol. 58, num. 12, pp. 1371–1376, 2005.
- [10] D. Alejandro, “Síndrome metabólico en personal de salud de una unidad de medicina familiar,” *Rev Med Inst Mex Seguro Soc*, vol. 48, num. 3, pp. 297–302, 2010.
- [11] J. Rosas Guzmán, A. González Chávez, P. Aschner, y R. Bastarrachea, “Consenso latinoamericano de la asociación latinoamericana de diabetes (alad): Epidemiología, diagnóstico, control, prevención y tratamiento del síndrome metabólico en adultos,” *Rev ALAD*, vol. 18, num. 1, pp. 25–42, 2010.
- [12] P. Zimmet, G. Alberti, y J. Shaw, “Nueva definición mundial de la fid del síndrome metabólico: argumentos y resultados,” *Diabetes Voice*, vol. 50, num. 3, pp. 31–33, 2005.
- [13] J. E. Hall, *Guyton y Hall. Tratado de fisiología médica*. Elsevier Health Sciences, 2011.
- [14] M. Schnell, Z. A. Dominguez, y C. Carrera, “Aspectos genéticos, clínicos y fisiopatológicos del síndrome metabólico.” in *Anales venezolanos de nutrición*, vol. 20, num. 2, 2007.
- [15] M. I. Harris, W. C. Hadden, W. C. Knowler, y P. H. Bennett, “Prevalence of diabetes and impaired glucose tolerance and plasma glucose levels in us population aged 20–74 yr,” *Diabetes*, vol. 36, num. 4, pp. 523–534, 1987.
- [16] C. R. Kahn, “Insulin resistance, insulin insensitivity, and insulin unresponsiveness: a necessary distinction,” *Metabolism*, vol. 27, num. 12, pp. 1893–1902, 1978.
- [17] R. J. Brown y J. A. Yanovski, “Estimation of insulin sensitivity in children: methods, measures and controversies,” *Pediatric diabetes*, vol. 15, num. 3, pp. 151–161, 2014.
- [18] N. Friedrich, B. Thuesen, T. Jørgensen, A. Juul, C. Spielhagen, H. Wallaschofski, y A. Linneberg, “The association between igf-i and insulin resistance,” *Diabetes care*, vol. 35, num. 4, pp. 768–773, 2012.
- [19] N. Motamed, S. J. H. Miresmail, B. Rabiee, H. Keyvani, B. Farahani, M. Maadi, y F. Zamani, “Optimal cutoff points for homa-ir and quicki in the diagnosis of metabolic syndrome and non-alcoholic fatty liver disease: A population based study,” *Journal of diabetes and its complications*, vol. 30, num. 2, pp. 269–274, 2016.
- [20] R. Kahn, “Follow-up report on the diagnosis of diabetes mellitus: the expert committee on the diagnosis and classifications of diabetes mellitus,” *Diabetes care*, vol. 26, num. 11, p. 3160, 2003.
- [21] I. G. Goikoetxea, “Aportaciones a la clasificación no supervisada ya su validación. aplicación a la seguridad informática,” Ph.D. dissertation, Universidad del País Vasco-Euskal Herriko Unibertsitatea, 2010.



- [22] A. K. Jain, M. N. Murty, y P. J. Flynn, “Data clustering: a review,” *ACM computing surveys (CSUR)*, vol. 31, num. 3, pp. 264–323, 1999.
- [23] A. K. Jain, “Data clustering: 50 years beyond k-means,” *Pattern recognition letters*, vol. 31, num. 8, pp. 651–666, 2010.
- [24] P. Gil, C. Mateo Agulló, J. Pomares, G. J. García Gómez, y F. Torres, “Reconocimiento de objetos 3d con descriptores de superficie,” 2016.
- [25] T. Kawada, T. Andou, y M. Fukumitsu, “Waist circumference, visceral abdominal fat thickness and three components of metabolic syndrome,” *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 10, num. 1, pp. 4–6, 2016.
- [26] J. J. Montaña Moreno y otros, “Redes neuronales artificiales aplicadas al análisis de datos,” 2002.
- [27] T. Masters, *Practical neural network recipes in C++*. Morgan Kaufmann, 1993.