

# Análisis de señales cromatográficas provenientes de muestras de orina, para el análisis de cáncer de próstata usando procesamiento de señales.

## Analysis of chromatographic signals from urine samples for prostate cancer analysis using signal processing.

Ronald de Jesús Torres-Flórez

Universidad de Pamplona

[torresfloronald@gmail.com](mailto:torresfloronald@gmail.com)

Luís Enrique Mendoza

Universidad de Pamplona

[luis.mendoza@unipamplona.edu.co](mailto:luis.mendoza@unipamplona.edu.co)

Zulmary Carolina Nieto-Sánchez

Universidad de Francisco de Paula Santander

[zulmarycarolinanisa@ufps.edu.co](mailto:zulmarycarolinanisa@ufps.edu.co)

Recibido: 09 de abril de 2019.

Aprobado: 13 de junio de 2019.

**Resumen**—La temprana detección del cáncer de próstata es de vital importancia para la puesta en marcha del tratamiento del paciente. Es por ello que este artículo demuestra que las señales cromatográficas provenientes de muestras de orina tienen relación con enfermedades relacionadas con la próstata del hombre. En este trabajo se usaron técnicas matemáticas tales como: SVM y RNA con el fin de extraer y verificar el patrón encontrado en cada señal cromatográfica. El algoritmo fue entrenado con 10 señales de cromatografía provenientes de muestras de orina, de las cuales 7 fueron obtenidas de los grupos de pacientes de control, y 7 de pacientes enfermos. La obtención de resultados positivos llevó consigo la aplicación de técnicas de pre y procesamiento sobre las señales de cromatografías, entre las que se encuentran, recorte de zona de interés, filtrado y corrección de línea base; cuyo propósito de aplicación permitió la búsqueda de los patrones característicos propios de cada grupo de pacientes, debido a la presencia o no presencia de células anormales o cancerígenas en la próstata. El sistema fue validado haciendo uso de datos ciegos y el resultado se contrastó con el médico experto en el área, permitiendo de este modo llegar a conclusiones puntuales. La especificidad del software fue del 92.86%, índice obtenido en la validación del software, cuyas señales ingresadas provienen de muestras de orinas de los pacientes analizados, pero diferentes a las utilizadas para el entrenamiento.

**Palabras clave:** Cromatografía de gases, cáncer de próstata, acondicionamiento, análisis de señales, marcadores urinarios.

**Abstract**—The early stoppage of prostate cancer is of vital importance for the start-up of the treatment of diagnosed patient, it is for them that the present project has as objective to look for an alternative tool that allows the possibility of presenting a tool of support for detecting prostate cancer, developing a method of diagnosis based on artificial intelligence, through chromatography signals of flatulence's come from urine sample for the study of the area of urology. In this project used mathematical techniques such as SVM and RNA, in order to extract and verify the pattern found. The algorithm of present project was trained 10 signals chromatography came from urine sample, which five were obtained of groups of control patients, and five of sick patients. For the development of analysis used Matlab 2014<sup>a</sup> student version. The Obtaining of positive results took with it the application of techniques of pre and processing about chromatography signals, among which are clipping of interest area, filtering and base line correction whose propose of application allowed the search of typical characteristic patterns of each group of patients, due to the presence of abnormal or cancerous cells in the prostate. The system was appreciated, doing used of blind dates and the result contrasted with the expert doctor in the area, allowing of this manner to arrive to punctual conclusions. The software specificity was 92.86%, rate obtained in the validation of software, whose signals entered come from of urine sample of analyzed patients, but different from the ones used to the treatment.

**Keywords:** Gas chromatography, prostate cancer, conditioning, signal analysis, urinal markers.

\*Autor para correspondencia.

Correo electrónico: [torresfloronald@gmail.com](mailto:torresfloronald@gmail.com) (Ronald de Jesús Torres Flórez).

La revisión por pares es responsabilidad de la Universidad de Santander.

Este es un artículo bajo la licencia CC BY-ND (<https://creativecommons.org/licenses/by-nd/4.0/>).

Forma de citar: R. J. Torres Flórez, L. E. Mendoza y Z. C. Nieto Sánchez, "Análisis de señales cromatográficas provenientes de muestras de orina, para el análisis de cáncer de próstata usando procesamiento de señales", Aibi revista de investigación, administración e ingeniería, vol. 7, no. 2, pp. 8-15, 2019.

## I. INTRODUCCIÓN

El cáncer de próstata (PC) es la neoplasia maligna masculina más común en los países desarrollados y su mortalidad solo está detrás del cáncer de pulmón y colorrectal [1]. Constituye uno de los problemas sanitarios más importantes en nuestro medio, tanto en términos de morbimortalidad, como de impacto social, económico o sobre la calidad de vida [2].

Una persona afectada por cáncer de próstata, pulmonar, de mamas y otros tipos de cáncer, presenta metabolitos volátiles específicos de cada uno de ellos, que sirven de marcadores y posibilitan el diagnóstico precoz de dichas enfermedades, el empleo de cromatografía de gases junto con otras técnicas de detección han permitido identificar estos marcadores proponiendo un método de diagnóstico para tales enfermedades, con las ventajas de ser discriminativo, sensible, rápido, seguro y no invasivo [3]. Actualmente existe suficiente evidencia que la detección precoz de cáncer de próstata disminuye la mortalidad, ya que se cuenta con tratamientos efectivos para la enfermedad diagnosticada en etapas tempranas. El desafío se centra en encontrar nuevas técnicas y marcadores que permitan identificar aquellos hombres que se encuentren en mayor riesgo, evitando así procedimientos innecesarios y en muchas ocasiones incómodos para el paciente [4]; siendo hoy en día la orina una fuente prometedora de biomarcador para una fácil detección y predicción en el pronóstico del cáncer de próstata, cuyas ventajas se ven reflejadas en la fácil disposición, rapidez y de ser no invasiva [5] [6]. Se ha demostrado que, durante el crecimiento del tumor, cambios en las proteínas de las células tumorales conducen a la peroxidación de los componentes de la membrana celular y la producción de compuestos orgánicos volátiles en la orina, que pueden ser detectados en el espacio superior de las células o en el aire espirado [10].

El uso de perros adecuadamente entrenados con la capacidad de detectar CaP con un alto grado de éxito [11], sugiere que a partir de la orina de los pacientes se liberan al aire compuestos orgánicos volátiles (COV), la mayoría desconocidos en la actualidad, capaces de ser identificados por el olfato, años posteriores en el proyecto realizado en [12], se estableció con precisión en términos de sensibilidad y especificidad con la cual un sistema olfativo canino rigurosamente entrenado podría reconocer compuestos orgánicos volátiles específicos del cáncer de próstata en muestras de orina. En [13] se evalúa la capacidad de una nariz electrónica para discriminar el cáncer de próstata de la hiperplasia prostática benigna. Se ha desarrollado un nuevo sistema basado en electrodos metálicos que a pesar de recoger información “poco precisa y selectiva”, el procesamiento del conjunto de datos obtenidos es capaz de determinar si una persona tiene cáncer de próstata y vejiga, ofreciendo hasta el momento mejores resultados que la prueba de PSA (antígeno prostático específico) en sangre [6], en [14] se estudió los aminoácidos Ácido aspártico, Serina, Arginina, Leucina e Isoleucina en muestras de orina de diferentes pacientes, concluyendo que es posible encontrar diferencia para los pacientes con cáncer con respecto a los grupos de sujetos sanos. En [15] emplearon Cromatografía de Gases-Espectrometría de Masas (CG/EM) y Cromatografía Líquida de Alta Eficiencia-Espectrometría de Masas (HPLC-EM) para estudiar muestras biológicas como tejido, sangre y orina de individuos con cáncer de próstata localizado características propias de este tipo de cáncer. El trabajo realizado en [16], desarrollaron y validaron un método para la determinación de metabolitos, como biomarcadores del cáncer de próstata en muestras de orina, entre los metabolitos de estudio se encontraban la sarcosina, alanina, leucina y prolina.

## II. SEÑAL CROMATOGRÁFICA

Se utilizaron señales de cromatografía proveniente de muestras de orinas, cuyas longitudes constan de un total de 49994 puntos. La base de datos cuenta con siete señales de pacientes sujetos a control; así como un total de siete señales provenientes de pacientes enfermos. La

figura 1 y la 2 muestran un registro cromatográfico real de un paciente sano y un paciente enfermo respectivamente.

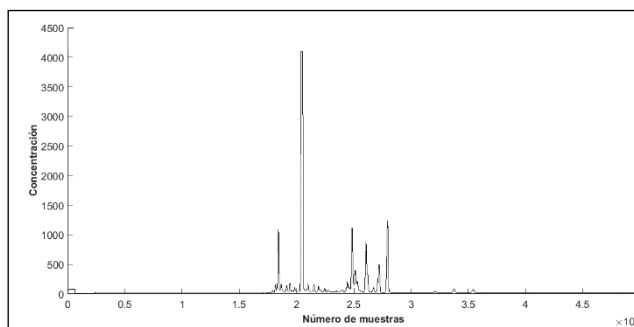


Figura: 1 Señal de cromatografía de paciente de control.  
Fuente: Elaboración propia.

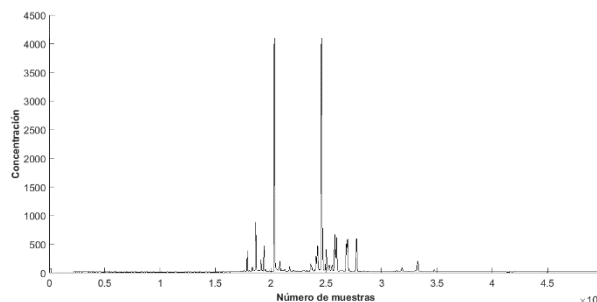


Figura: 2 Señal de cromatografía de paciente enfermo.  
Fuente: Elaboración propia.

## III. ACONDICIONAMIENTO DE LAS SEÑALES

El diagrama de la 3 muestra las etapas del procedimiento del acondicionamiento para la señal de cromatografía. A continuación, se describe cada una de estas etapas.



Figura: 3 Procedimiento para el acondicionamiento de la señal de cromatografía.

Fuente: Elaboración propia.

### a. Obtención de la zona de interés

La morfología de las señales de cromatografías obtenidas, tienen la particularidad de presentar pocos picos (compuestos) en los primeros y en los últimos minutos del examen, cumpliéndose así para ambos tipos de pacientes (ver **s;Error! No se encuentra el origen de la referencia.**), es por ello que no se tendrá en cuenta el rango de datos obtenido en estos intervalos de tiempo. La selección exacta del rango de tiempo donde se tomará la señal (zona de interés), se realiza de una forma dependiente de la ubicación del compuesto más representativo en ambos grupos de pacientes, cumpliendo siempre con las características de ser el de mayor concentración en las muestras y el primero (de los más altos) en ser detectado. Se ha tomado un número de 5000 puntos antes y 25000 puntos después del pico representativo.

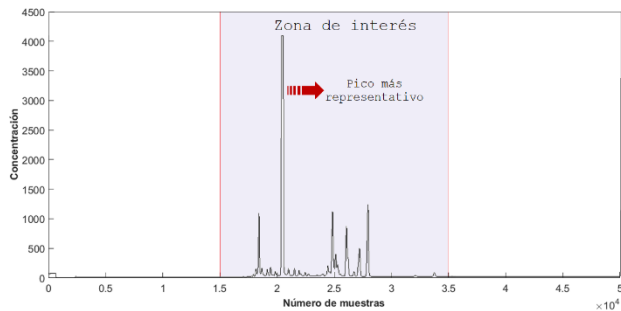


Figura: 4 Visualización de la zona de interés.  
Fuente: Elaboración propia.

### b. Suavizado de la señal

El suavizado de la señal tiene un único objetivo, y es el de eliminar pequeños rizados en la señal de cromatografía, rizados que pueden provocar conflictos en la corrección de la línea base. Este suavizado no involucra la acción de algún tipo de filtrado digital, si no la aplicación de condicionales. En la 5; **Error! No se encuentra el origen de la referencia.** se muestra una porción de la señal de cromatografía antes (recuadro rojo) y después (recuadro azul) de ser suavizada. Nótese como la señal suavizada no presenta cambios rápidos en amplitud lo cual permite tener una señal más adaptable para realizar la corrección de línea base. Esta técnica se realizó teniendo en cuenta los picos más significativos de la señal, y así mismo calculando el promedio una vez se presenta un cambio muy rápido en el tiempo y una amplitud muy pequeña.

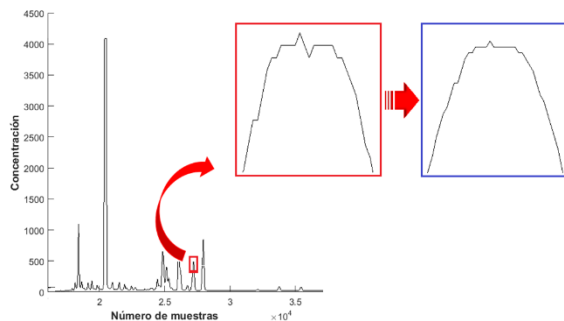


Figura: 5 Suavizado de la señal de cromatografía.  
Fuente: Elaboración propia.

### c. Corrección de línea base

Un registro cromatográfico puede ser considerado como una serie temporal conformada por la superposición de tres tipos de señales. La primera señal que representa los componentes químicos, la segunda señal se manifiesta como componentes de ruido aleatorio y finalmente la última componente se manifiesta como una señal de baja frecuencia denominada línea base (LB), que tiende a introducir error en la medición de los picos, en especial cuando hay picos de baja altura en las adyacencias de picos elevados [38]. El algoritmo propuesto produce una corrección de la línea base cumpliendo el diagrama de la **Error! No se encuentra el origen de la referencia.** **Error! No se encuentra el origen de la referencia.**

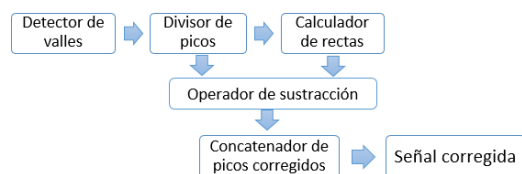


Figura: 6 Pasos para la corrección de Línea Base.  
Fuente: Elaboración propia.

La **Error! No se encuentra el origen de la referencia.** muestra una señal de cromatografía de gases de una muestra de orina,

proveniente de un paciente sano con las técnicas de recortado de zona de interés y suavizado aplicadas (gráfica azul). Como se observa en el diagrama de la **Error! No se encuentra el origen de la referencia.**, la señal es pasada a través de un detector de valles, que consta de un sistema de derivación de primer orden y de un sistema diferenciador entre un máximo (cresta o pico) y un mínimo (valle).

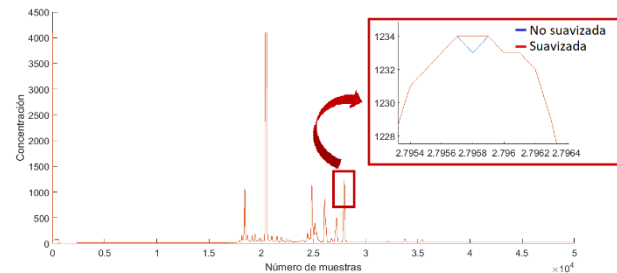


Figura: 7 Señal de Cromatografía recortada y suavizada.  
Fuente: Elaboración propia.

La aplicación de la derivada de primer orden permite encontrar la localización en el dominio del tiempo de los valles o crestas presentes sobre la señal. La derivada de una señal digital se representa matemáticamente como se muestra en la ecuación (1).

$$y[n] = x[n + 1] - x[n] \quad (1)$$

Donde:

$x[n]$ , Señal recortada y suavizada.  
 $y[n]$ , Señal derivada.

Determinar que la derivada represente un valle, depende de la transición de negativo a positivo de la señal derivada  $y[n]$ . Conociendo los valles de la señal  $x[n]$ , se procede a dividir a  $x[n]$  en sub-señales o picos  $x_i$ , cuyas longitudes comprenden de valle a valle. En la **Error! No se encuentra el origen de la referencia.** se aprecia la localización de los valles sobre  $x[n]$ , así como la representación de una sub-señal  $x_i$ .

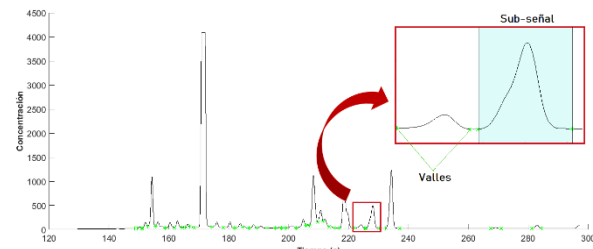


Figura: 8 Representación de la Sub-señal y ubicación de los valles en la señal de cromatografía.  
Fuente: Elaboración propia.

El cálculo de la línea base  $r_i$  para cada sub-señal  $x_i$  (ver **Error! No se encuentra el origen de la referencia.**), hace uso de la ecuación (2).

$$r_i = \left( \frac{x[v_{i+1}] - x[v_i]}{v_{i+1} - v_i} \right) (n - v_i) + x[v_i] \quad (2)$$

Donde  $v_i$  es la posición de los valles  $i$ .

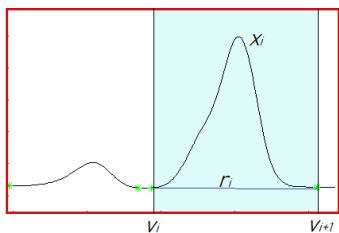


Figura: 9 Parámetros de la sub-señal de la recta de línea base  $r_i$ .  
Fuente: Elaboración propia.

Para corregir la línea base de la señal de cromatografía  $x[n]$ , se realiza la operación matemática descrita en la ecuación (3) para cada pico o sub-señal.

$$X[n] = x_i - r_i \quad \because \quad v_i \leq n \leq v_{i+1} \quad (3)$$

Donde  $X[n]$ , representa el pico con corrección de la línea base.

Si concatenamos todas las sub-señales  $X[n]$ , se obtiene de forma final la señal completa con corrección de línea base. La **¡Error! No se encuentra el origen de la referencia.0** muestra la corrección de línea base de la señal  $x[n]$  luego de ser concatenada (señal roja), y sin corrección de línea base (señal azul). Nótese como los diferentes picos se encuentran sobre un nivel cero y sin modificaciones en los tiempos de retención de cada uno de ellos.

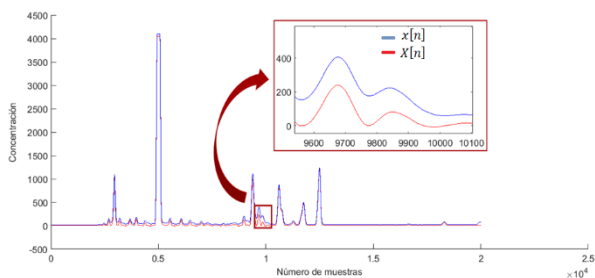


Figura: 10 Señal de cromatografía con y sin corrección de línea base.  
Fuente: Elaboración propia.

En este tipo de señales se presenta un caso particular mostrado en la 1a. Cuando la recta  $r_i$  intercepta en más de 2 puntos (incluyendo el punto inicial y final) con la sub-señal  $x_i$ , la sub-señal con corrección de línea base no presenta una distribución positiva como debería ocurrir (**¡Error! No se encuentra el origen de la referencia.1b**), para ello se realiza un ajuste en su forma, conservando su energía y tiempo de retención. Este ajuste puede observarse en la **¡Error! No se encuentra el origen de la referencia.2**.

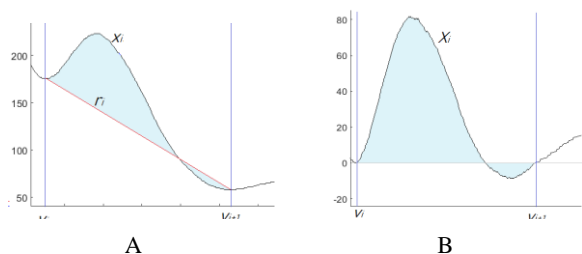


Figura: 11 Caso especial en la corrección de la línea base.  
Fuente: Elaboración propia.

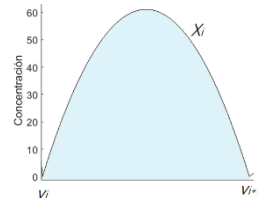


Figura: 12 Sub-señal  $X_i$  con corrección de línea base y redistribución de energía.  
Fuente: Elaboración propia.

Posteriormente se procede a eliminar los picos con un valor menor al 0.5% del valor máximo son llevados a cero. El comportamiento de este resultado final se muestra en la **¡Error! No se encuentra el origen de la referencia.3**.

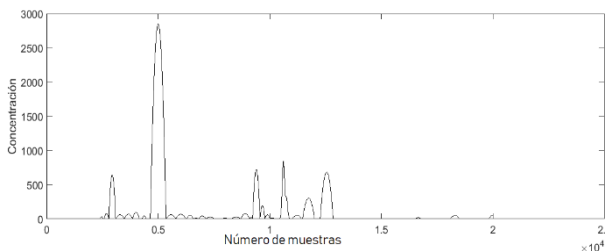


Figura: 13 Señal concatenada con energía redistribuida.  
Fuente: Elaboración propia.

#### IV. EXTRACCIÓN DE PATRONES

Para la extracción de patrones se utilizaron los diferentes procedimientos mencionados a continuación.

- 1 Encontrar los picos diferentes entre cada uno de los grupos de pacientes.
- 2 Crear un cromatógrama que contenga los picos representativos de los pacientes con cáncer de próstata.
- 3 Extraer en la señal de ambos grupos los respectivos patrones.

Basándose en la teoría de que cada compuesto está representado por un tiempo específico de retención (aunque se presenten pequeñas variaciones en el tiempo), encontrar los picos diferentes entre los pacientes de control y los pacientes enfermos, implica conocer cuales picos están más cerca en el tiempo entre una señal y otra, para realizar una correcta asignación. De la figura 14 se puede decir:

- El pico X es la misma sustancia que el pico A, si  $d1 < d2$ , por lo tanto, el pico B es una sustancia no común entre las dos señales, siendo propia de la señal con cáncer de próstata.
- El pico X es la misma sustancia que el pico B, si  $d1 > d2$ , por lo tanto, el pico A es una sustancia no común entre las dos señales, siendo propia de la señal con cáncer de próstata.

Donde,

$$d1 = t_x - t_{a2} \quad (4)$$

$$d2 = t_{b1} - t_x \quad (5)$$

Tabla 1: Tabla de variables de tiempo

Variable	Significado
$t_{a1}$	Tiempo de inicio del pico A
$t_{a2}$	Tiempo final del pico A

$t_{b1}$	Tiempo de inicio del pico B
$t_{b2}$	Tiempo final del pico B
$t_{x1}$	Tiempo de inicio del pico X
$t_{x2}$	Tiempo final del pico X
$t_x$	Tiempo de ocurrencia del punto máximo del pico X

Fuente: Elaboración propia.

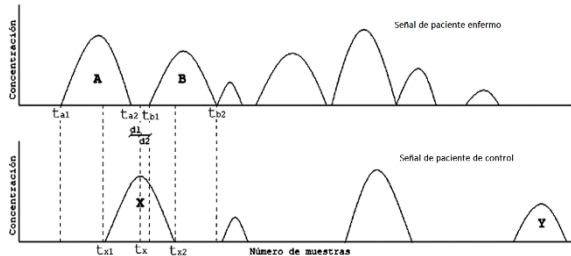


Figura: 14 Determinación de picos iguales.  
Fuente: Elaboración propia.

Como resultado de este procedimiento se obtiene la señal picos representativos  $P_{r1}$  (denominada así por ser la primera señal obtenida) mostrada en la **Figura 5** (para este ejemplo), la cual contiene únicamente los picos que están presentes en la señal de paciente enfermo y no en la señal cromatográfica del paciente de control. Nótese que el pico Y en la señal patrón de paciente sano no es tomada en cuenta como un pico representativo de la señal de cáncer de próstata, ya que pertenece al paciente sano y puede ser común entre los dos grupos.

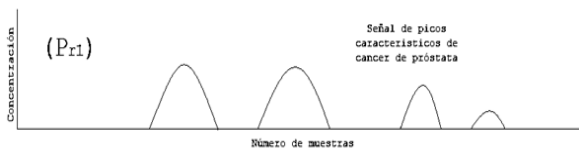


Figura: 15 Señal de picos representativos del cáncer de próstata ( $P_{r1}$ ).  
Fuente: Elaboración propia.

Este procedimiento es aplicado nuevamente utilizando la misma señal patrón de paciente de control, pero cambiando la señal de paciente de enfermo, obteniendo el supuesto resultado mostrado en la **Figura 6**.

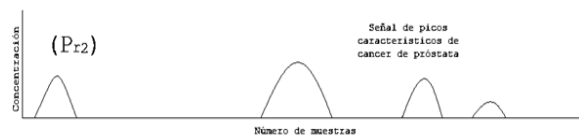


Figura: 16 Señal de picos representativos del cáncer de próstata ( $P_{r2}$ ).  
Fuente: Elaboración propia.

Este procedimiento es aplicado una y otra vez utilizando la misma señal de paciente control, pero cambiando la señal de paciente enfermo (3 señales). Luego se realiza empleando una señal de cromatografía de un segundo paciente de control con todas las señales de pacientes enfermo, y así sucesivamente hasta obtener un total de 12 señales de picos representativos  $P_{ri}$  (3 señales de pacientes enfermos con las 4 señales de control).

Obtenidas las 12 señales de picos representativos del cáncer de próstata  $P_{ri}$ , se procede a encontrar un único cromatograma característico de los picos representativos ( $P_{rCap}$ ), para ello se toman todos los picos que están en las diferentes señales  $P_{r1}, P_{r2}, P_{r3} \dots P_{r12}$ . Por ejemplo, de las señales  $P_{r1}$  y  $P_{r2}$  de la **Figura 5** y la **Figura 6**.

7 respectivamente, se obtiene ( $P_{rCap}$ ), tal como se muestra en la **Figura 7**.



Figura: 17 Cromatograma de picos representativos de cáncer de próstata ( $P_{rCap}$ ).  
Fuente: Elaboración propia.

Donde C1, C2... y C5, representan los compuestos diferentes.

Obtenido el cromatograma característico de los picos representativos  $P_{rCap}$ , se procede a encontrar cada uno de las señales patrones  $S_p$ , para ello se toma cada señal de cromatografía con corrección de línea base  $X[n]$ , donde cada pico ha sido normalizado, y se le extrae únicamente los picos presentes en el cromatograma  $P_{rCap}$ . El propósito de normalizar cada pico es de evitar que la concentración entre una misma muestra, en señales diferentes, tenga influencia significativa en la clasificación. Por ejemplo, si decimos que la señal de la **Figura 5** es  $X[n]$ , la representación de los picos normalizados es la señal de la **Figura 18**, representada como  $X_{norm}[n]$  mediante la ecuación (4). La señal patrón  $S_p$  para este ejemplo se muestra en la **Figura 19**.

$$X_{norm}[n] = \frac{X[n]}{\max(X[n])} \quad \therefore \quad v_1 \leq n \leq v_2 \quad (4)$$



Figura: 18 Señal de "paciente de control".  
Fuente: Elaboración propia.



Figura: 19 Señal de "paciente de control" con picos normalizados.  
Fuente: Elaboración propia.



Figura: 20 Señal Patrón de "paciente de control".  
Fuente: Elaboración propia.

Realizando el mismo procedimiento para diferentes señales  $X[n]$ , se obtiene un total de 14 señales patrón (7 de pacientes de control y 7 de pacientes enfermos), sin embargo, la localización de un patrón característico en este espacio es difícil de encontrar, por lo que es importante el uso de técnicas de inteligencia artificial supervisadas, tales como Máquina de soporte vectorial y redes neuronales.

Para obtener un mejor patrón característico entre ambos grupos, se realiza la eliminación de los picos que no aporten significativamente un buen resultado de clasificación. La eliminación de estos picos es la siguiente. Se elimina un pico P de todas las señales, tanto las de entrenamiento como las de simulación, y se calcula el porcentaje de clasificación. Si la clasificación aumenta con la eliminación de ese

pico, indica que el pico P no es importante, ya que representa una similitud entre ambos grupos, o no es común entre un mismo grupo; de lo contrario si la clasificación disminuye, indica que el pico P si es de gran importancia entre los dos grupos de clasificación. Este proceso es realizado hasta recorrer todos los picos de las señales, y la clasificación sea la de mayor grado, permitiendo de este modo encontrar el mejor espacio de clasificación.

### V. CLASIFICACIÓN

La clasificación implica el uso de técnicas de inteligencia artificial, como las redes neuronales artificiales (red neuronal probabilística) y las máquinas de soporte vectorial. El resultado de cada técnica implica un porcentaje total que influye en la clasificación final.

### VI. RESULTADOS

#### a. Cromatógrama de picos característicos del Cáncer de Próstata

La **¡Error! No se encuentra el origen de la referencia.** muestra el cromatógrama de todos los picos que presentan diferencias entre los pacientes de control y los pacientes enfermos de la base de datos que se obtuvo. Nótese que existe un vasto número de picos, por lo que fue indispensable eliminar picos que no representan un marcador del cáncer de próstata.

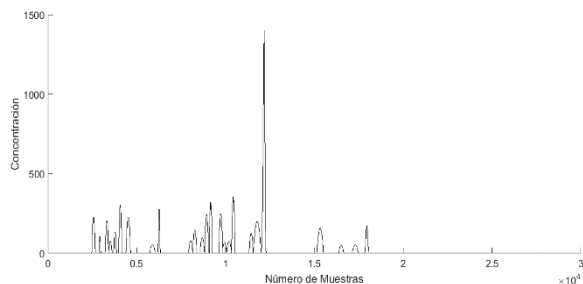


Figura: 21 Cromatógrama de picos representativo en señales de pacientes enfermos.  
Fuente: Elaboración propia.

#### b. Señales patrón de pacientes sujetos a control y pacientes enfermos

La figura 22 muestra la matriz patrón obtenido luego de realizar la eliminación de los picos que no contribuyen a una buena clasificación. Los niveles de color amarillo indican un valor elevado, mientras se hace menor cuando se torna de color azul. Visiblemente es difícil encontrar un patrón para cada uno de los grupos, es por ello que el uso de inteligencia artificial busca dar solución.

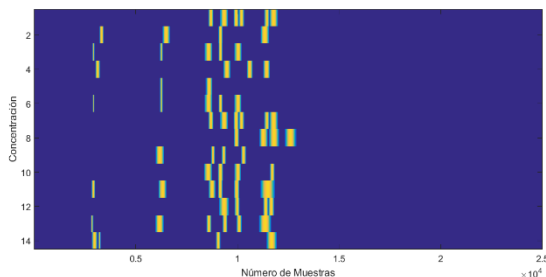


Figura: 22 Patrones característicos.  
Fuente: Elaboración propia.

Donde,

Tabla 2: Relación señal-patrón-paciente.

Señal Patrón	Paciente
N° 1	Enfermo N° 1
⋮	⋮
N° 7	Enfermo N° 7
N° 8	Control N° 1
⋮	⋮
N° 14	Control N° 7

Fuente: Elaboración propia.

#### a. Clasificación

La simulación de cada una de las técnicas de inteligencia artificial se realizó con las señales de entrenamiento y las otras cuatro señales las cuales el sistema no fue entrenado. A continuación, se muestra el porcentaje de clasificación de cada técnica de clasificación para cada señal, las cuales forman parte del grupo de señales patrón con las que el sistema no fue entrenado.

Tabla 3: Paciente de control N° 6 (Ver Figura 1, Anexo).

Clasificación del Paciente N° 6.				
Diagnóstico Medico	Diagnóstico del Software			
	RNA		MSV	
Paciente con Cáncer de próstata	*CaP.	%	*CaP.	%
	NO	0%	SI	50%
%Total	50%			
	Paciente con Cáncer de próstata			

Fuente: Elaboración propia.  
\*CaP: Cáncer de próstata.

Tabla 4: Paciente de Control N° 7 (Ver Figura 2 en Anexo).

Clasificación del Paciente N° 4.				
Diagnóstico Medico	Diagnóstico del Software			
	RNA		MSV	
Paciente con Cáncer de próstata	*CaP.	%	*CaP.	%
	NO	0%	NO	50%
%Total	50%			
	Paciente con Cáncer de próstata			

Fuente: Elaboración propia.  
\*CaP: Cáncer de próstata.

Tabla 5: Paciente Sano N° 6 (Ver Figura 3 en Anexo).

Clasificación del Paciente N° 5.				
Diagnóstico Medico	Diagnóstico del Software			
	RNA		MSV	
Paciente sano	*CaP.	%	*CaP.	%
	NO	50%	NO	50%
%Total	100%			
	Paciente sano			

Fuente: Elaboración propia.  
\*CaP: Cáncer de próstata.

Tabla 6: Paciente Sano N° 7 (Ver Figura 4 en Anexo).

Clasificación del Paciente N° 7.				
Diagnóstico Medico	Diagnóstico del Software			
	RNA		MSV	
Paciente sano	*CaP.	%	*CaP.	%
	NO	50%	NO	50%
%Total	100%			
	Paciente sano			

Fuente: Elaboración propia.  
\*CaP: Cáncer de próstata.

El porcentaje de clasificación del sistema fue calculada haciendo uso de la ecuación (5) Téngase en cuenta que en este valor se toman en cuenta los porcentajes de clasificación con los datos de entrenamiento, los cuales fueron de un 100%.

$$\%P = \frac{\%C(\text{Pac. N}^\circ 1) + \dots + \%C(\text{Pac. N}^\circ 14)}{n} \quad (5)$$

$$\%P = \frac{12(100\%) + 2(75\%)}{14}$$

$$\%P = 96.43 \%$$

La sensibilidad del sistema la cual nos indica la capacidad de nuestro estimador para dar como casos positivos los casos realmente enfermos, es calculada con la ecuación (6).

$$\text{Sensibilidad} = \frac{VP}{VP + FN} \quad (6)$$

Donde *VP* es verdaderos positivos y *FN* falsos negativos.

La especificidad del sistema la cual nos indica la capacidad de nuestro estimador para dar como casos negativos los casos realmente sanos, es calculada con la ecuación (7).

$$\text{Especificidad} = \frac{VN}{VN + FP} \quad (7)$$

Donde *VN*, serían los verdaderos negativos y *FP* los falsos positivos.

Para las técnicas de clasificación de Redes Neuronales la sensibilidad y la especificidad para los valores de *VP*, *FP*, *VN* y *FN* mostrados en la son:

Tabla 7: Resultados de clasificación de redes neuronales.

Verdaderos Positivos	<i>VP</i>	7
Falsos Positivos	<i>FP</i>	2
Verdaderos Negativos	<i>VN</i>	5
Falsos negativos	<i>FN</i>	0

Fuente: Elaboración propia.

$$\%Sensibilidad = \left(\frac{7}{7+0}\right) * 100\% = 100\%$$

$$\%Especificidad = \left(\frac{5}{5+2}\right) * 100\% = 71.43\%$$

Para las técnicas de clasificación de Maquinas de soporte vectorial la sensibilidad y la especificidad para los valores de *VP*, *FP*, *VN* y *FN* mostrados en la son:

Tabla 8: Resultados de clasificación de Máquinas de soporte vectorial.

Verdaderos Positivos	<i>VP</i>	7
Falsos Positivos	<i>FP</i>	0
Verdaderos Negativos	<i>VN</i>	7
Falsos negativos	<i>FN</i>	0

Fuente: Elaboración propia.

$$\%Sensibilidad = \left(\frac{7}{7+0}\right) * 100\% = 100\%$$

$$\%Especificidad = \left(\frac{7}{7+0}\right) * 100\% = 100\%$$

El porcentaje de sensibilidad y especificidad del sistema lo hemos definido como el promedio de especificidad entre las dos técnicas de clasificación.

$$\%Sensibilidad = \frac{100\% + 100\%}{2} = 100\%$$

$$\%Especificidad = \frac{71.43\% + 100\%}{2} = 85.71\%$$

## VII. CONCLUSIONES

A pesar de estar definidos dos grupos de señales de cromatografía de gases provenientes de muestras de orina de pacientes sanos y pacientes de control, no existe homogeneidad en los patrones obtenidos, por lo que se hace complejo determinar con otras técnicas (correlación, energías particulares) cuales pacientes presentan las posibilidades de cáncer de próstata y cuáles no, sin embargo validamos la clasificación haciendo uso de inteligencia artificial. La corrección de línea base sin duda alguna mejora sustancialmente la morfología de las señales de cromatografía, permitiendo determinar con gran estimación los tiempos de retención de cada uno de los picos que representan a una sustancia en particular, de tal forma que permite encontrar con la aplicación de otras técnicas de pre y procesamiento, una mejor representación para la clasificación en uno de los dos grupos manejados (paciente sano o paciente sujeto a control). Así mismo, el porcentaje de clasificación alcanzado del 88% permite validar la clasificación y decir que los patrones característico en cada señal de cromatografía se obtuvo de manera adecuada, esto conlleva a decir que si existe una diferencia entre los dos grupos de pacientes (sanos y de control), sin embargo, este porcentaje de clasificación no es suficiente para realizar un análisis correcto (Redes Neuronales), consecuencia de usar pocas señales para el entrenamiento, es por ello que se aconseja aumentar de forma significativa el número de datos de entrenamiento. En la clasificación las máquinas de soporte vectorial jugaron un papel importante, destacándose sobre las redes neuronales artificiales. Los porcentajes de sensibilidad y especificidad del software puede variar si se realiza la validación con un mayor número de pacientes, ya que los resultados en pacientes con *Verdaderos* positivos, Falsos positivos, Verdaderos negativos o Falsos negativos pueden variar significativamente, por lo tanto, estas cifras no son completamente de confiar.

## VIII. REFERENCIAS

- [1] J. Morote, X. Maldonado y R. Morales, «Prostate cancer,» *Elsevier*, vol. I, n° 1, p. 7, 2016.
- [2] A. Jiménez Pacheco, J. M. Peinado Herreros y M. C. Iribar Ibabe, «Evaluacion del en el diagnóstico actual del cáncer de prostata. Desarrollo de nuevos Biomarcadores Urinarios,» vol. 96, p. 10, 2011.
- [3] R. Marbot Ramada y I. Martín Hernández, «Análisis del aliento: un método promisorio para el diagnóstico del cáncer y factores desencadenantes,» *Revista CENIC Ciencias Biológicas*, vol. XXXIX, n° 1, p. 7, 2008.
- [4] C. Ramos G. y J. Fullá O., *Detección Precoz De Cáncer De Próstata*, Departamento de Urología. Clínica las Condes, 2013.
- [5] M. Truong, B. Yang y D. Jarrard, «Towards the Detection of Prostate Cancer in Urine: A Critical analysis,» *The Journal of Urology*, vol. I, n° 1, p. 21, 2012.
- [6] S. Digital, «Lenguas electrónicas, sistema no invasivo que detecta cáncer de próstata y de vejiga,» Consalud, Madrid, 2017.
- [7] W. Filipiak , A. Sponring , A. Filipiak, C. Ager, J. Schubert y W. Miekisch , «TD-GC-MS analysis of volatile metabolites of

human lung cancer and normal cells in vitro. *Cancer Epidemiol Biomarkers*,» 2010.

- [8] J. Cornu, G. Cancel-Tassin, V. Ondet, C. Girardet y O. Cussenot , «Olfactory Detection of Prostate Cancer by Dogs Sniffing Urine: A Step Forward in Early Diagnosis,» *Eur Urol*, 2011.
- [9] G. Taverna , L. Tidu, F. Grizzi , V. Torri, A. Mandressi, P. Sardella , G. La Torre, G. Cocciolone , M. Seveso, G. Giusti, R. Hurle , A. Santoro y P. Graziotti , «Highly-Trained Dogs' Olfactory System Detects Prostate Cancer in Urine Samples,» *The Journal of Urology*, vol. I, n° 1, p. 28, 2014.
- [10] A. Roine, . E. Veskimäe, A. Tuokko, P. Kumpulainen , . J. Koskimäki , A. Keinänen Tuom, R. Häkkinen Merja , J. Vepsäläinen , T. Paavonen , . J. Lekkala , . T. Lehtimäki , . L. Tammela Teuvo y K. J. Oksala Niku , «Detection of Prostate Cancer by an Electronic Nose: Proof of principle study,» *Science Direct*, 2015.
- [11] F. C. Romero Segura, Estudio preliminar de posibles Biomarcadores del Cáncer de Próstata, Caracas: Universidad Central de Venezuela. Facultad de Ciencias, 2014.
- [12] A. Sreekumar , I. Poisson , T. Rajendiran, A. Khan , Q. Cao , J. Yu , B. Laxman, R. Mehra , R. Lonigro , Y. Li , M. Nyati, A. Ahsan , . S. Kalyana-Sundaram, B. Han, X. Cao , J. Byun, J. Wei , . S. Varambally, C. Beecher y A. Chinnaiyan, «Metabolomic profiles delineate potential role sarcosine in prostate cancer progression,» *Nature*, 2009.
- [13] . M. Shamsipur, M. Taghi y M. Babri, «Quantification of candidate prostate cancer metabolite biomarkers in urine using dispersive derivatization liquid–liquid microextraction followed by gas and liquid chromatography–mass spectrometry,» *Journal of Pharmaceutical and Biomedical Analysis*, pp. 81-81:65-75, 2013.
- [14] E. Gómez Sotomayor y B. Serrano Ortega, Urología Básica para Estudiantes de Medicina, Primera ed., Loja: Unidad de Comunicación e Imagen Institucional, 2016, p. 236.
- [15] L. H. Alba, M. Alba, D. Ortiz S., M. Otálora E. y D. Roselli, Análisis De Los Registros Individuales De Prestación De Servicios De Salud (Rips), Bogotá: Medicina: Cancer en Colombia, 2016.
- [16] F. Lara Rosano, Fundamento de redes neuronales artificiales, Unam.
- [17] R. Torres Cabeza , O. Llanes Santiago, E. Barrero Viciedo y V. Moreno Vega, «Faults Diagnostic using Hopfield Artificial Neural Network in front of Incomplete Data,» *Journal of Engineering and Technology for Industrial Applications*, vol. IV, n° 13, pp. 77-82, 2018.
- [18] E. J. Carmona Suárez, Tutorial sobre Máquinas de Vectores Soporte (SVM)., Madrid: Dpto. de Inteligencia Artificial, Universidad Nacional de Educación a Distancia., 2014.
- [19] J. L. Semmlow, Biosignal and Medical Image Processing, Segunda ed., vol. I, New York: Taylor & Grancis Group, 2008, p. 450.
- [20] O. M. d. l. Salud, «Informe de la 58a Asamblea Mundial de la Salud A58/16.,» Organización Mundial de la Salud, 2005.
- [21] P. C. Walsh y J. F. Worthington, The prostate, a guide for men and the women who love them, New York: Warner Books, 1997.
- [22] S. Dijkstra, I. L. Birker, F. P. Smit, G. H. J. M. Leyten, T. M. de Reijke, F. A. M. Peter y I. M. van Oort, «Prostate Cancer Biomarker Profiles in Urinary,» *Investigative Urology*, vol. I, n° 1, p. 7, 2013.