



Ciencia Latina
Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), julio-agosto 2024,
Volumen 8, Número 4.

https://doi.org/10.37811/cl_rcm.v8i4

DESARROLLO DE SOFTWARE PARA LA GESTIÓN DE DATOS DE UNA NARIZ ELECTRÓNICA CON INTERFAZ GRÁFICA

**SOFTWARE DEVELOPMENT FOR DATA MANAGEMENT OF
AN ELECTRONIC NOSE WITH GRAPHICAL INTERFACE**

Victor Jorge Cajilema Camacho

Universidad De Las Fuerzas Armadas, Ecuador

Ximena del Rocío López Chico

Universidad De Las Fuerzas Armadas, Ecuador

Milton Patricio Navas Moya

Universidad De Las Fuerzas Armadas, Ecuador

Milton Fabricio Pérez Gutiérrez

Universidad De Las Fuerzas Armadas, Ecuador

Enrique Terán

Universidad De Las Fuerzas Armadas, Ecuador

DOI: https://doi.org/10.37811/cl_rcm.v8i4.13374

Desarrollo de Software para la Gestión de Datos de una Nariz Electrónica con Interfaz Gráfica

Victor Jorge Cajilema Camacho¹

vjcajilema@espe.edu.ec

<https://orcid.org/0009-0005-0723-8622>

Universidad De Las Fuerzas Armadas ESPE
Ecuador

Ximena del Rocío López Chico

xrlopez@espe.edu.ec

<https://orcid.org/0000-0002-7519-6453>

Universidad De Las Fuerzas Armadas ESPE
Ecuador

Milton Patricio Navas Moya

mpnavas@espe.edu.ec

<https://orcid.org/0000-0002-9526-3369>

Universidad De Las Fuerzas Armadas ESPE
Ecuador

Milton Fabricio Pérez Gutiérrez

mfperez3@espe.edu.ec

<https://orcid.org/0000-0002-8639-0980>

Universidad De Las Fuerzas Armadas ESPE
Ecuador

Enrique Terán

eteran@usfq.edu.ec

<https://orcid.org/0000-0001-6979-5655>

Universidad San Francisco de Quito
Ecuador

RESUMEN

Este estudio aborda el desarrollo de un sistema enfocado en el análisis de compuestos en la exhalación humana tomados por medio de una nariz electrónica desarrollada en una fase previa, el sistema es desarrollado bajo metodología ágil scrum. En la primera etapa se realiza el levantamiento de requerimientos tomados en base a entrevistas y revisión de la fase anterior, como resultado esta etapa se obtiene el product backlog de sistema. La etapa de diseño del sistema se basa en la arquitectura C4, en esta etapa inicialmente se realiza una investigación bibliográfica para identificar las necesidades del sistema a nivel de diseño para que el producto final satisfaga las necesidades de la etapa de levantamiento de requerimientos, al finalizar esta etapa se obtuvieron los diagramas de cada nivel de la arquitectura C4 para el sistema detallando que tecnologías se utilizaran para la creación de cada parte del sistema y como se integraran. En la etapa de desarrollo se procede a implementar la solución propuesta en base al diseño, esta fase detalla cómo se crearon cada contenedor del sistema incluyendo un contenedor de análisis de datos creado con Python y una aplicación web para la presentación de resultados del sistema. Como resultado final se obtuvo un sistema capaz de registrar los datos provenientes de una nariz electrónica, identificar valores atípicos dentro de la exhalación de los pacientes y generar graficas estadísticas que permitan a los médicos analizar los datos tomados por la nariz electrónica.

Palabras clave: nariz electrónica, minería de datos, visualización de datos, procesamiento de datos

¹ Autor principal

Correspondencia: vjcajilema@espe.edu.ec

Software Development for Data Management of an Electronic Nose with Graphical Interface

ABSTRACT

This study addresses the development of a system focused on analyzing compounds in human exhalation, using an electronic nose developed in a previous phase. The system was created using the agile Scrum methodology. In the first stage, requirements were gathered based on interviews and a review of the previous phase, resulting in the system's product backlog. The system design stage is based on the C4 architecture. In this stage, a bibliographic investigation was initially carried out to identify the system's design needs to ensure that the final product meets the requirements gathered in the first stage. At the end of this stage, diagrams for each level of the C4 architecture were obtained for the system, detailing the technologies to be used in creating each part of the system and how they will be integrated. In the development stage, the proposed solution is implemented based on the design. This phase details how each container of the system was created, including a data analysis container developed with Python and a web application for presenting the system's results. As a final result, a system was obtained that can record data from an electronic nose, identify outliers within the patients' exhalation, and generate statistical graphs that allow doctors to analyze the data collected by the electronic nose.

Keywords: electronic nose, data mining, data visualization, data processing

Artículo recibido 10 julio 2024

Aceptado para publicación: 15 agosto 2024



INTRODUCCIÓN

La nariz electrónica es una herramienta diseñada para emular el sentido del olfato humano mediante sensores químicos y algoritmos de reconocimiento de patrones. Desde su introducción esta tecnología ha evolucionado significativamente, pasando de dispositivos voluminosos y costosos a modelos más compactos y eficientes. A nivel mundial se ha utilizado la nariz electrónica en diferentes campos como la agricultura para determinar el estado de diferentes alimentos, la seguridad para detectar peligro de explosivos, en el campo de la salud se ha propuesto la nariz electrónica como una alternativa de diagnóstico no invasivo de enfermedades recolectando información del paciente de manera externa sin lastimar su cuerpo (Karakaya, Ulucan, & Turkan, 2020).

La nariz electrónica, diseñada para emular el sentido del olfato humano mediante sensores químicos y algoritmos de reconocimiento de patrones, ha evolucionado de dispositivos voluminosos y costosos a modelos más compactos y eficientes. A nivel mundial, se ha aplicado en diversos campos: en la agricultura para evaluar alimentos, en la seguridad para detectar explosivos, y en la salud como una herramienta de diagnóstico no invasivo que recolecta información sin causar daño físico (Karakaya, Ulucan, & Turkan, 2020).

En la industria alimentaria, la nariz electrónica se ha empleado para detectar frescura, adulteración, malos olores y bacterias en productos como leche, pescado, vino y café. También se ha utilizado para clasificar la maduración de frutas (Cayllahue, 2020) y analizar la calidad del vino a través de su aroma (López, Arazuri, García, Mangado, & Jaren, 2014). Además, se ha investigado su uso en la predicción de acidez, solubilidad y firmeza en peras (Zhang, Wang, & Ye, 2008), así como en la detección de la adulteración del aceite de oliva con aceite de avellana (Mildner-Szkudlarz & Jeleń, 2008).

En seguridad se ha utilizado la nariz electrónica para detección de agentes químicos, explosivos peligrosos y monitorización del estado del aire en contenedores, así como edificios y medios de transporte, la inspección de contenedores, monitorización del aire en edificios de gobierno, del metro de Washington, descripción del aire ambiental en aviones comerciales; para detectar olores peligrosos o posibles amenazas de terroristas (Charumpom & Omatu, 2002).

La aplicación de la nariz electrónica en el ámbito de la contaminación se ha centrado en la detección de problemas en agua potable y ríos. Estudios han demostrado su capacidad para monitorizar la calidad



del agua potable (Gardner, Shin, Hines, & Dow, 2000), identificar emisiones de ríos altamente contaminados (Lamagna, Reich, Rodríguez, Boselli, & Cicerone, 2008), y reconocer sustancias contaminantes en el agua utilizando la nariz Kamina (Goschnick, Koroncz, Frietsch, & Kiselev, 2005). Además, la monitorización de emisiones en una planta química evidenció que la nariz electrónica puede detectar fugas de gases tóxicos, contribuyendo a la mitigación de riesgos ambientales (Wilson & Baietto, 2009).

El campo de la medicina es el menos explorado en la aplicación de la nariz electrónica, los primeros ensayos utilizando una nariz electrónica para el diagnóstico de enfermedades respiratorias se llevaron a cabo mediante el uso de aparatos construidos a partir de sensores de cromatografía y espectrometría de masas. En los últimos años las investigaciones en que se busca aplicar la nariz electrónica como un método no invasivo de diagnóstico han aumentado y contemplan temas como La detección de infecciones en el tracto urinario y vaginosis bacteriana (Persaud, Pisanelli, Evans, & Travers, 2006). La nariz electrónica ha sido utilizada también para realizar diagnóstico preliminar de diabetes, utilizando la nariz electrónica como una herramienta de diagnóstico no invasiva (Mendoza Montoya & Ávila Navarro, 2017). Aplicación de una nariz electrónica de 6 sensores para identificar y clasificar los volátiles emitidos a través del aliento exhalado obtenido de un grupo de personas sanas y pacientes con síntomas de la Enfermedad Pulmonar Obstructiva Crónica(EPOC) (Durán Acevedo, Velásquez Carvajal, & Gualdron Guerrero, 2012).

En 2020 como parte de una tesis de doctorado se plantea la aplicación de la nariz electrónica como herramienta para la detección temprana de cáncer gástrico (Jaimes Mogollón, 2020).

La revisión bibliográfica muestra la aplicación de la nariz electrónica en diferentes campos de estudio siendo la medicina una de las áreas de menor concurrencia, sin embargo, en los últimos años ha aumentado el estudio de esta área, impulsada por los avances tecnológicos y científicos en campos como inteligencia artificial y análisis de datos (Li, Wei, Zhou, Wang, & You, 2023).

En Latinoamérica, la aplicación de la nariz electrónica está en una fase emergente, especialmente en la industria alimentaria, agrícola y ambiental. Países del continente han comenzado a explorar su uso para asegurar la calidad de productos alimenticios, como el café, el vino y las frutas tropicales.

A pesar de que las investigaciones en la región con respecto a esta tecnología son claramente menores que en otras regiones del mundo, en los últimos años se ha mostrado un mayor interés por parte de universidades e institutos de educación en realizar investigaciones en esta área. Por ejemplo, el trabajo realizado por investigadores de la Universidad Tecnológica de Panamá que incluye estudios sobre la utilización de narices electrónicas, abarcando desde la detección de contaminación ambiental hasta la seguridad industrial. Estos dispositivos se utilizan para identificar y cuantificar compuestos volátiles en diferentes entornos, lo que permite una evaluación más precisa de la calidad del aire y la detección de contaminantes (Moreno et al., 2009). En Colombia por otro lado se ha realizado investigaciones en el área alimentaria aplicadas al control de calidad de materias primas y productos elaborados; monitoreo de procesos, frescura y madurez, investigaciones de estimación del tiempo de vida útil (Quicazán, Díaz & Zuluaga, 2011).

La Universidad de las Fuerzas Armadas ha iniciado el desarrollo de una nariz electrónica con el objetivo de contribuir al diagnóstico médico con una herramienta no invasiva para analizar el estado de salud de los pacientes a partir de los compuestos en su exhalación. En la primera fase, se construyó un módulo electrónico con 6 sensores que capturan la exhalación y almacenan los datos localmente, permitiendo generar gráficas estadísticas manualmente con Excel (Cueva & Erazo, 2022).

Esta fase del estudio se enfoca en desarrollar un sistema integral que utiliza una nariz electrónica para capturar compuestos en la exhalación.

Una vez tomada subida la información, el sistema permite identificar datos atípicos y, si es necesario, excluirlos del análisis, para luego generar gráficas estadísticas que faciliten la comparación con los resultados de la primera etapa.

Este informe se organiza en varias secciones: primero, se presenta una investigación en la sección de estado del arte para definir los términos clave; luego, en la sección de métodos se detalla la elicitación, el diseño del sistema utilizando el modelo C4 y el desarrollo que describe el módulo de captura de información, que permite cargar los datos obtenidos por la nariz electrónica junto con el estado del paciente, una vez subida la información se identifican datos atípicos y se generan gráficas estadísticas.

Estado del Arte

Nariz electrónica

Es una herramienta que combina sensores químicos y un módulo de reconocimiento de patrones para emular el olfato humano. Los sensores capturan datos de la exhalación, mientras que el módulo analiza estos datos (Moreno et al., 2009).

Las técnicas de reconocimientos de patrones utilizadas por la nariz electrónica pueden ser paramétricas y no paramétricas.

- Las técnicas paramétricas asumen que los datos se ajustan a un tipo de distribución específica como la distribución normal por ejemplo y requieren la estimación de parámetros como la media y la varianza (Moreno et al., 2009).
- Las técnicas no paramétricas a diferencia del grupo anterior no asumen ninguna distribución específica en los datos, son menos potentes pero más fáciles de utilizar. Se dividen en supervisadas y no supervisadas; las técnicas supervisadas requieren relaciones previamente identificadas, mientras que las no supervisadas operan sin necesidad de conocimiento previo (Moreno et al., 2009).

2.2 Modelo C4

El modelado C4 organiza la arquitectura de un sistema en cuatro niveles de abstracción, facilitando su comprensión en distintos grados de detalle (Rabehi, Helal, Zappa, & Comini, 2024).

- El nivel de contexto es el nivel más alto de abstracción que establece al sistema de manera general identificando los actores y sistemas externos con los que interactúa (Rabehi, Helal, Zappa, & Comini, 2024).
- El nivel de contenedores presenta la composición del sistema en base a unidades ejecutables que internamente tienen varios componentes como aplicaciones y bases de datos mostrando cómo se distribuyen las responsabilidades entre cada uno y cómo se comunican (Rodas, Ríos, & Solarte, 2016).
- Los componentes muestran a mayor detalle cómo están formados los contenedores permitiendo comprender cuáles son las funcionalidades asignadas a cada contenedor (Rodas, Ríos, & Solarte, 2016).

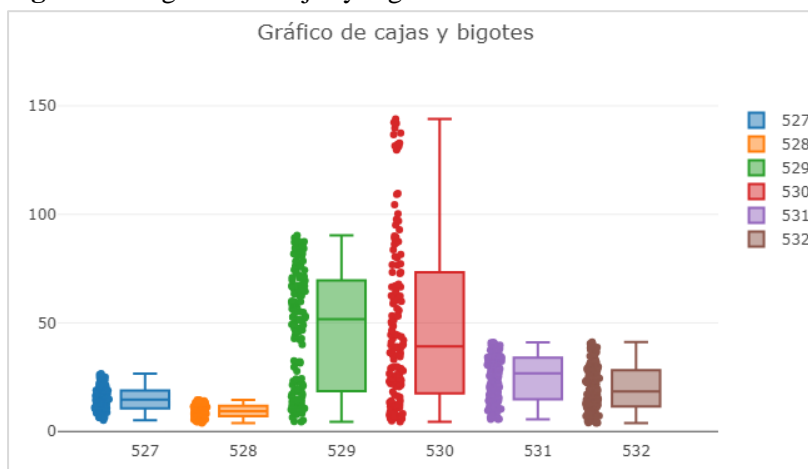


- El diagrama de clases es un nivel opcional en el cual se muestra como cada componente será implementado en términos de clases y sus relaciones (Rodas, Ríos, & Solarte, 2016).

Diagrama de cajas y bigotes

Es una distribución gráfica en la que intervienen la mediana, primer cuartil, tercer cuartil, valor máximo y valor mínimo (Miller, Heeren, & Hornsby, 2005). Estas gráficas son especialmente útiles para comparar distribuciones entre varios grupos, facilitando la identificación de diferencias y datos atípicos (Minnaard, Condesse, Minnaard, & Rabino, 2005).

Figura 1 Diagrama de cajas y bigotes

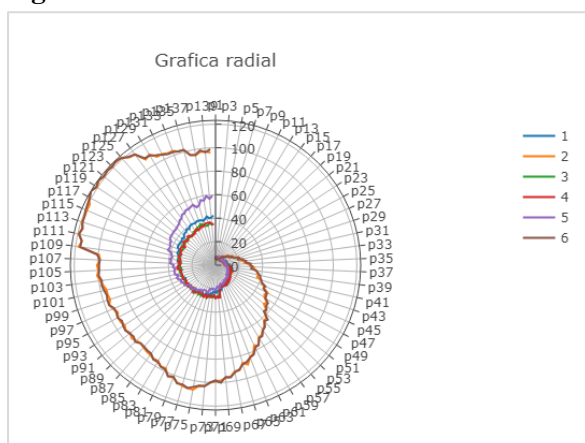


Gráficos radiales.

Los gráficos radiales muestran múltiples variables a la vez, con ejes que se extienden desde un punto central para reflejar la cantidad de variables en el análisis (Benedito, 2013).

En este estudio, permiten representar varios compuestos de la nariz electrónica de manera compacta, destacando la relación y proporción entre ellos.

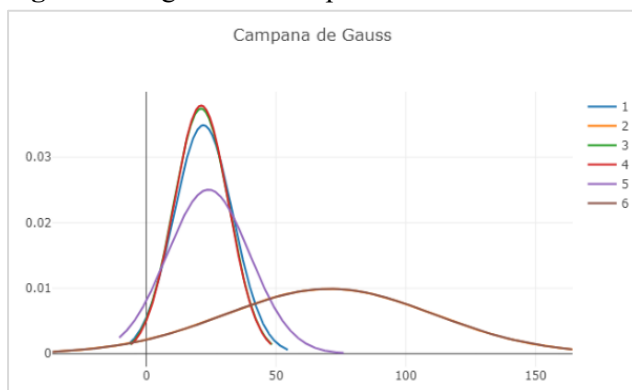
Figura 2 Gráfico radial



Campanas de Gauss.

El gráfico de campanas de Gauss es importante en aspectos como la identificación de la normalidad de los datos y la identificación de desviaciones de lo que sería una distribución ideal normal (Minnaard, Condesse, Minnaard, & Rabino, 2005).

Figura 3 Diagrama de campanas de Gauss.



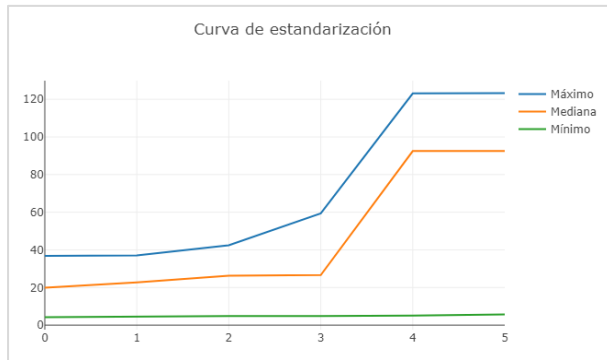
Estandarización de las muestras.

La estandarización es un proceso meticuloso que dependiendo de la técnica analítica utilizada, la cantidad de parámetros a estandarizar y la logística involucrada, puede requerir un tiempo variable para su implementación. Este proceso incluye los parámetros esenciales que permiten que un método analítico comience a generar datos con un grado adecuado y verificable de confianza (Cueva & Erazo, 2022).

Modelo discriminante de gases.

Esta técnica estadística multivariante se utiliza para evaluar si existen diferencias significativas entre los grupos de gases detectados en función de un conjunto de variables medidas, tiene como objetivo establecer procedimientos para la clasificación sistemática de los grupos analizados. En este estudio, se analizaron los puntos máximos y mínimos de las muestras recolectadas, excluyendo aquellas que mostraban discrepancias significativas con la mayoría de los valores. Esto permitió la estandarización del nivel de cada compuesto (Cueva & Erazo, 2022).

Figura 4 Diagrama de curvas de estandarización.



Histogramas

Los histogramas representan la distribución de datos mediante barras verticales, donde el ancho de cada barra indica el rango del intervalo y la altura indica el número de puntos en el intervalo. Los datos se dividen en sub-rangos llamados intervalos, mostrando cuántos puntos pertenecen a cada uno (Gilat, 2005).

Diagramas Q-Q Plot.

Los diagramas Q-Q Plot (quantil to quantil plot) permiten comparar la distribución de datos con una distribución teórica, es útil en casos que se necesite comparar la variable de estudio con la distribución normal, en estos casos se traza una línea diagonal junto con la distribución de datos y se define visualmente si los datos siguen la misma dirección de la línea en cuyo caso presentan una tendencia normal (Di Rienzo et al., 2008).

Pruebas de Shapiro-Wilk.

Las pruebas de Shapiro-Wilk permiten definir si un conjunto de datos presenta una distribución normal. El método inicia ordenando el conjunto de datos de menor a mayor, calcular la media y la varianza muestral (Flores Tapia & Flores Cevallos, 2021).

Pruebas de Lilliford.

Las pruebas de Lilliford están basadas en Kolmogorov-Smirnov que se utiliza para evaluar una hipótesis cuando los parámetros de media y desviación estándar (Rial & Varela, 2008).

Minería de datos.

La minería de datos es un proceso que permite analizar datos para descubrir nuevas relaciones, patrones, tendencias utilizando técnicas avanzadas como el muestreo y el análisis de datos.

La minería de datos es principalmente útil cuando se dispone de grandes cantidades de información que no podrían ser analizadas manualmente por una persona (Rodríguez, 2010).

DB-Scan.

DB-SCAN (Density-Based Spatial Clustering of Applications with Noise) es un método de clustering basado en densidad. Fue introducido por Martin Ester en 1996. Este algoritmo agrupa puntos que están densamente conectados en un espacio de datos y puede identificar clústeres de formas arbitrarias en presencia de ruido (Khan, Rehman, Aziz, Fong, & Sarasvady, 2014).

METODOLOGÍA

Se propone desarrollar una nariz electrónica capaz de identificar el perfil de salud de los pacientes mediante la recolección y análisis de seis compuestos en su exhalación (Cueva & Erazo, 2022). Esta etapa se enfoca en crear un software, aplicando un ciclo de desarrollo que comienza con la elicitación de requisitos para identificar necesidades, roles y procesos necesarios para continuar con el diseño y desarrollo del sistema.

Análisis

Para la elicitación de requerimientos, se realizaron entrevistas y se analizó la documentación previa del desarrollo de la nariz electrónica. Esto permitió generar las historias de usuario y el product backlog según la metodología Scrum. La figura 5 muestra el product backlog, destacando historias clave como la captura de datos, la visualización estadística y la detección de valores atípicos.

Figura 5 Product Backlog del sistema.

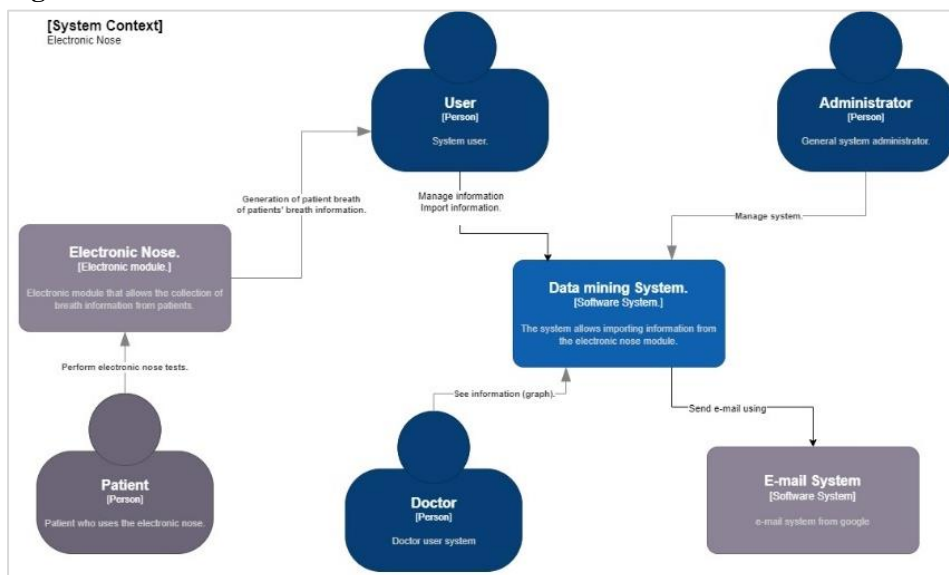
Identificador (ID) de la Historia	Enunciado de la Historia	Alias	Estado	Dimensión / Esfuerzo	Iteración (Sprint)	Prioridad
1	Como un Usuario, necesito iniciar sesión, con la finalidad de acceder a las funcionalidades del sistema	Login	cumplido	8	1	1
2	Como un Usuario, necesito crear pacientes, con la finalidad de disponer de pacientes para posteriores registros	Registro paciente	cumplido	8	1	1
3	Como un Usuario, necesito registrar los datos de la nariz electrónica y estado del paciente, con la finalidad de guardar la información para su análisis	Toma de datos	cumplido	13	2	1
4	Como un Doctor, necesito visualizar las gráficas estadísticas de los datos tomados, con la finalidad de realizar el análisis de los datos	Visualización estadística	cumplido	21	3	1
5	Como un Doctor, necesito visualizar los datos que el sistema considera atípicos, con la finalidad de validar los datos que se mostraron en el análisis	Visualización de atípicos	cumplido	22	4	1
6	Como un Administrador, necesito registrar usuarios en el sistema, con la finalidad de darles acceso a Doctores y Usuarios generales al sistema	Registro de usuario	cumplido	8	5	1
7	Como un Usuario, necesito recuperar mi contraseña, con la finalidad de acceder al sistema	Reiniciar contraseña	cumplido	8	5	2
8	Como un Administrador, necesito registrar compuestos, con la finalidad de habilitarlos en el sistema	Registrar compuestos	cumplido	8	6	2

Diseño

Para el diseño del sistema se empleó el modelo C4, ideal para visualizar y comunicar la estructura compleja de sistemas de software mediante cuatro niveles de abstracción. El nivel de contexto, ilustrado en la Figura 6, identifica a los diferentes tipos de usuarios: un usuario general, un administrador y un doctor, quienes interactúan con el sistema para gestionar información y visualizar resultados, y un paciente, que interactúa principalmente con el módulo de la nariz electrónica para la recolección de datos.

Se incluyen módulos externos clave, como la nariz electrónica usada para captar datos del paciente, y el sistema de correo electrónico para verificar direcciones de los usuarios registrados. La integración de estos módulos es esencial para garantizar precisión y seguridad en el manejo de la información.

Figura 6 Nivel de contexto del modelo C4.



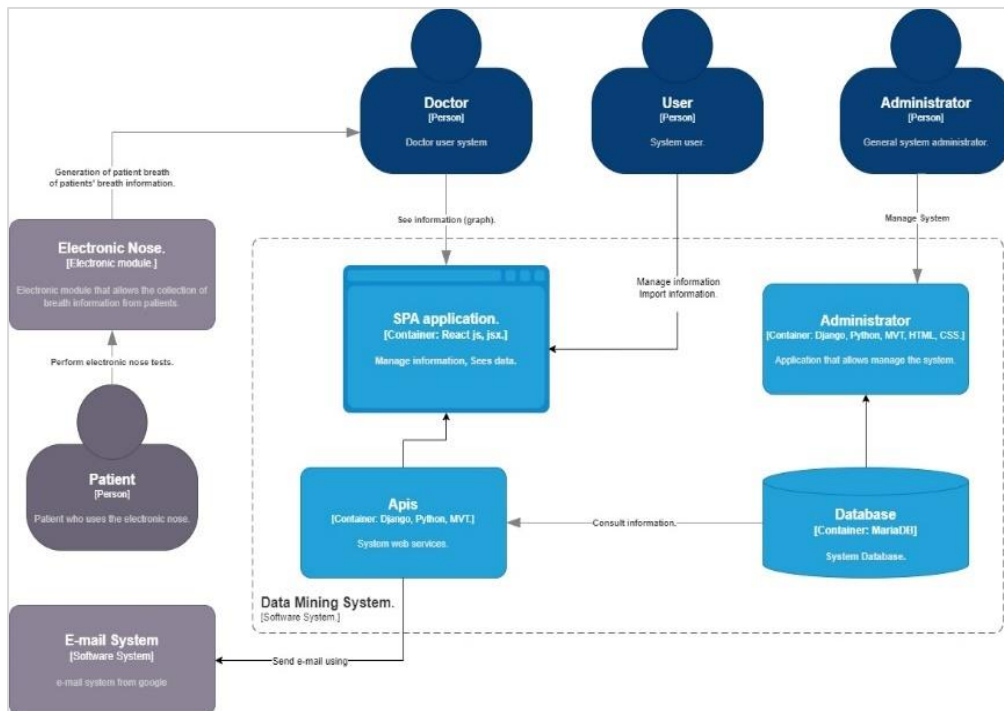
La segunda fase del modelado C4, ilustrada en la Figura 7, muestra la interacción de los usuarios con cada contenedor y la tecnología utilizada para su desarrollo.

El contenedor de APIs representa el backend del sistema, responsable del procesamiento de datos. Para la implementación de este contenedor se eligió Python y el framework Django por su eficiencia en el análisis de datos.

El contenedor de la Aplicación SPA (Single Page Application) constituye el frontend, que gestiona la interfaz visual del sistema. Esta aplicación web dinámica, desarrollada con React, permite subir y visualizar información de los pacientes, además de gestionar los compuestos y gráficos estadísticos.

El contenedor Administrador, desarrollado con Django y tecnologías web como HTML y CSS, facilita la gestión de las configuraciones generales del sistema, ofreciendo una interfaz administrativa robusta. El contenedor de Base de Datos utiliza MariaDB por su robustez, eficiencia y licencia abierta, ideal para la investigación. Este contenedor organiza y almacena todos los datos del sistema utilizando una estructura relacional definida por SQL.

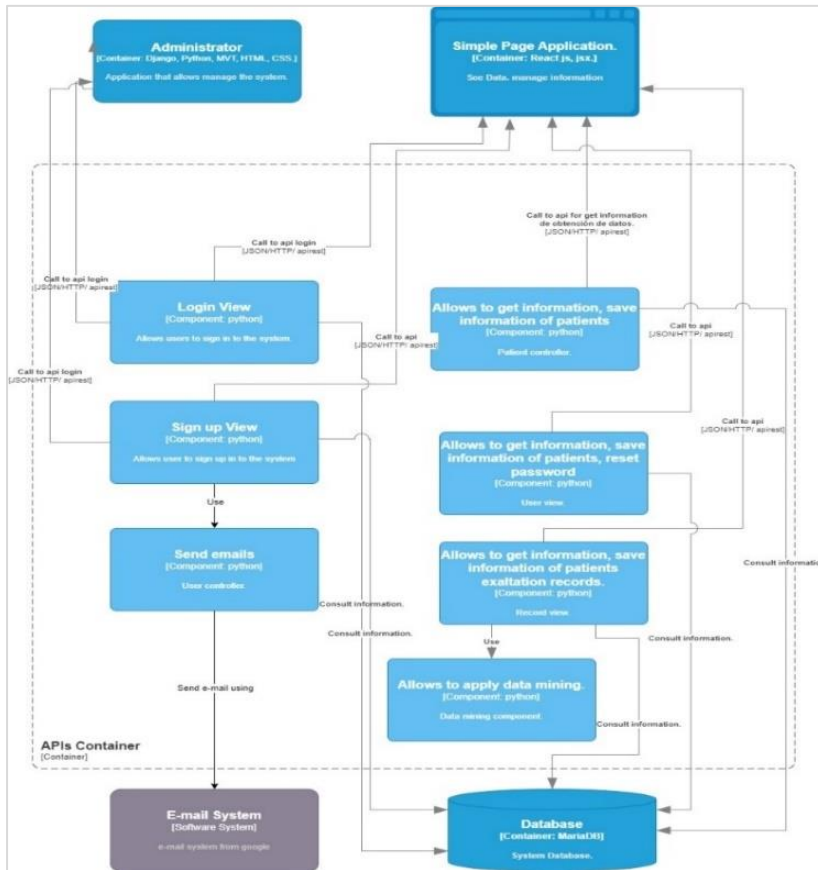
Figura 7 Nivel de contenedores del modelo C4.



La Figura 8 muestra la tercera fase del modelo C4, centrada en los componentes. Este nivel detalla cómo cada contenedor se forma por componentes que colaboran para cumplir las funciones del sistema. El contenedor de APIs gestiona la lógica del backend, manejando operaciones CRUD, la interacción con servicios externos y bases de datos.

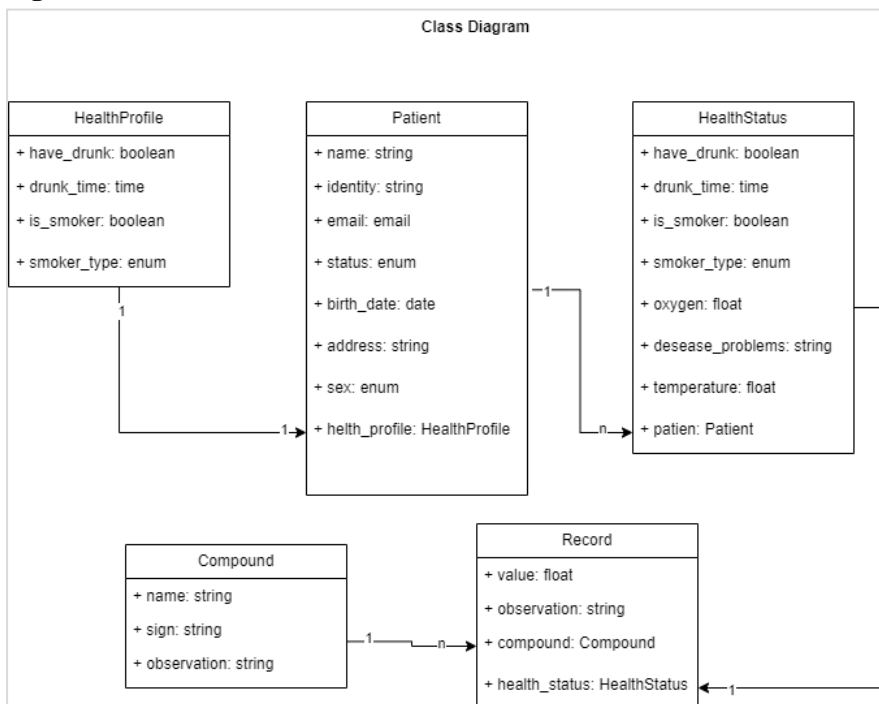
Dentro de este contenedor, se destacan componentes como el de subida de información, que envía datos de los pacientes a la base de datos; el de minería de datos, que detecta valores atípicos; y el de obtención de información, que envía los datos a la aplicación SPA para su visualización.

Figura 8 Nivel de componentes del modelo C4.



En la Figura 9 se puede apreciar el cuarto nivel C4 perteneciente al diagrama de clases que intervienen directamente en la fase de análisis de datos.

Figura 9 Nivel de clases del modelo C4.



Desarrollo

La aplicación SPA desarrollada facilita la gestión de perfiles de pacientes y el acceso a visualizaciones gráficas y resultados obtenidos mediante técnicas de minería de datos. Además, permite registrar mediciones basadas en el estado de salud del paciente, complementadas con datos de la nariz electrónica.

Como se muestra en la Figura 10, la interfaz del sistema captura y carga estos datos en la base de datos para su análisis. Recoge información detallada, como diagnóstico de enfermedades, consumo de alcohol, hábitos de fumar (incluyendo tipo de fumador), temperatura corporal, y oxigenación en sangre. También permite adjuntar un archivo generado externamente por la nariz electrónica, que contiene los datos detallados, para su integración y almacenamiento en el sistema.

Figura 10 Formulario de toma de datos.

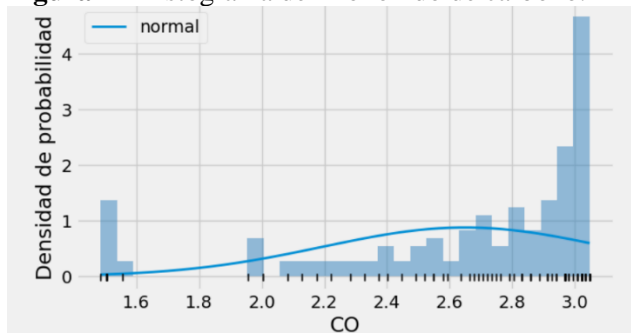
Pruebas de normalidad

Una vez se ha realizado la toma de datos se procede a determinar si los datos tienen distribución normal para lo cual se eligen 2 pruebas gráficas y 2 pruebas cuantitativas.

Entre las pruebas gráficas se consideran a los histogramas y a las gráficas QQplot.

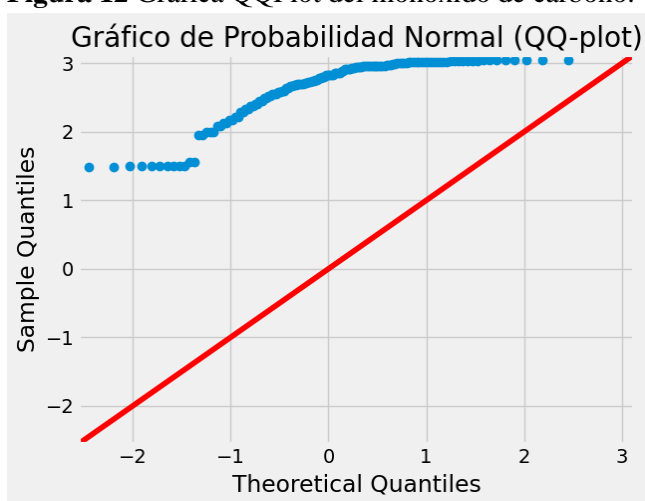
En la figura 11 se aplica la prueba gráfica de histogramas en la cual se puede apreciar que los datos no siguen una distribución normal ya que son demasiado irregulares y no pueden tener simetría gráfica.

Figura 11 Histograma del monóxido de carbono.



En la figura 12 se puede apreciar la gráfica QQplot del monóxido de carbono la cual muestra que en ningún punto los datos siguen una línea diagonal recta, por lo tanto, no tienen distribución normal.

Figura 12 Gráfica QQPlot del monóxido de carbono.



Una vez se ha realizado la toma de datos se procede a determinar si los datos tienen distribución normal para lo cual se eligen 2 pruebas gráficas y 2 pruebas cuantitativas.

Entre las pruebas graficas se consideran a los histogramas y a las gráficas QQplot.

Como método cuantitativo se elige el método de Lilliford y Shapiro-Wilk para definir si los datos de la investigación tienen distribución normal, para lo cual se siguieron los pasos descritos a continuación:

1. Definir la hipótesis nula (H_0) como que los datos tienen una distribución normal y la hipótesis alternativa (H_1) como que los datos no siguen una distribución normal.
2. Realizar los cálculos del valor p y el estadístico en cada uno de los métodos.

Tabla 1 Cálculos de distribución normal.

Método	Estadístico	Valor p
Lilliford	0.19101916	0.0009999999999998
Shapiro-Wilk	0.79406166	9.4815897754702e-13

- Finalmente se analiza los valores obtenidos y presentes en la tabla 2 para determinar si los datos tienen una distribución normal para lo cual se define que el nivel de significancia será de 0,05. En ambos casos tanto para el método de Lilliford como para el método de Shapiro-Wilk el valor p calculado es mucho más pequeño que el nivel de significancia rechazando de esta manera la hipótesis nula.

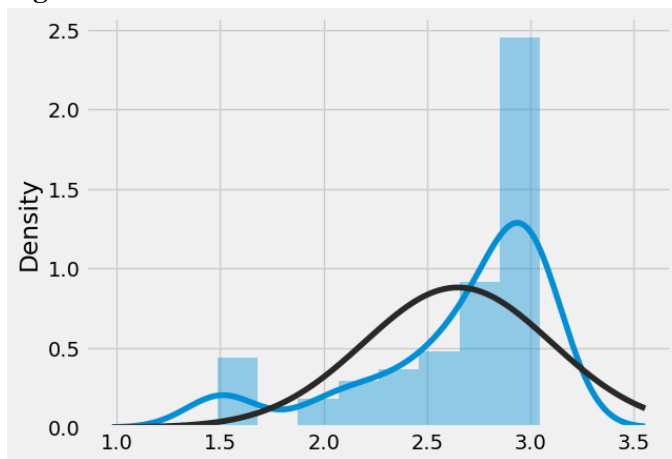
En conclusión, después de aplicar 4 diferentes pruebas a los datos de la investigación se pueden definir que los datos no tienen distribución normal.

Transformación de normalidad

Al no disponer de una distribución normal es necesario aplicar transformaciones de normalidad para poder graficar las campanas de Gauss se aplicaron las transformaciones de normalidad por medio de la librería scipy de Python.

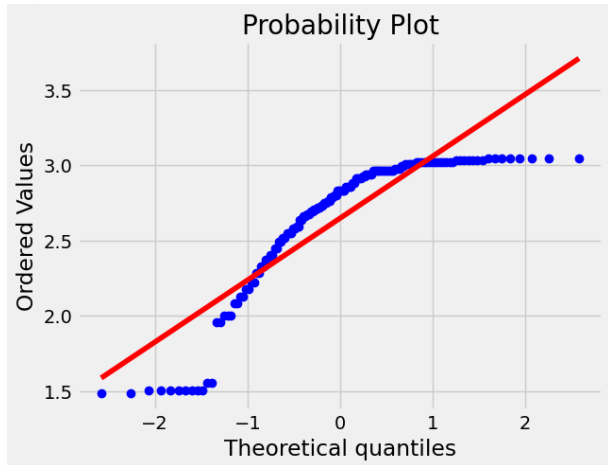
En la figura 13 se puede apreciar de fondo los histogramas junto con las gráficas de las curvas creadas por los datos disponibles, la curva de color negro representa los datos que han sido sometidos a la transformación de normalidad.

Figura 13 Curvas con datos transformados.



En la figura 14 se puede apreciar el gráfico QQplot con los datos transformados, los cuales muestran que se dispone de una distribución repartida a lo largo de la línea diagonal en comparación con la figura 6.

Figura 14 Gráfica QQPlot con datos transformados

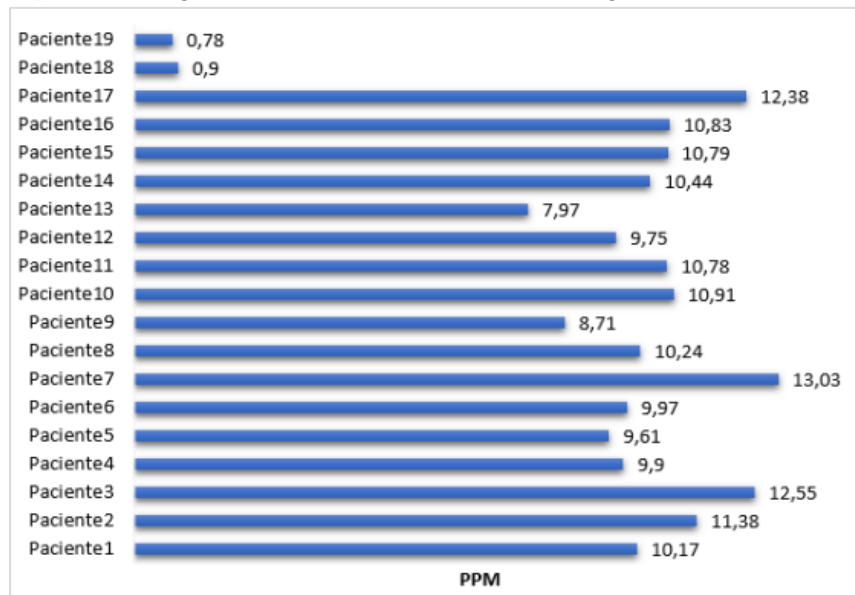


RESULTADOS Y DISCUSIÓN

Selección de la muestra

Se consideraron 19 personas para las pruebas del sistema. En las primeras pruebas, se observó que 2 pacientes mostraron mediciones bajas de los elementos. Por ejemplo, en la Figura 15, los pacientes 18 y 19 presentan bajas mediciones de dihidrógeno, con 0.9 PPM y 0.78 PPM respectivamente, evidenciando una anomalía en la gráfica. Este patrón se repite en todos los compuestos, lo que hace necesario aplicar un criterio de exclusión técnico para no afectar la muestra (Cueva & Erazo, 2022).

Figura 15 Diagrama de barras laterales del dihidrógeno.



Nota: Figura tomada de la primera fase de la investigación. Fuente: C. A. Cueva 2022

Dados estos resultados, se seleccionaron 17 personas como participantes en el proyecto. Durante la toma de datos, se identificaron pacientes con mediciones anormales en comparación con el resto, lo que llevó a la exclusión de estos datos atípicos en las pruebas iniciales de la primera etapa, según criterio médico (Cueva & Erazo, 2022). La Tabla 2 muestra que tanto el criterio médico como el técnico coinciden en excluir a ciertas personas del análisis por presentar datos atípicos.

Tabla 2 Análisis de toma de datos iniciales.

Compuesto	Análisis por criterio técnico	Análisis por criterio medico
H2	Pacientes 18 y 19	Pacientes 3, 10, 15 y 17
CH4	Pacientes 18 y 19	Paciente 10
NH3	Pacientes 18 y 19	Pacientes 3, 7 y 15
H2S	Pacientes 18 y 19	Pacientes 3 y 7
CO	Pacientes 18 y 19	Pacientes 3, 7 y 15
CO2	Pacientes 18 y 19	Pacientes 3, 7 y 15
Total	2	3

Nota: Tabla tomada de la primera fase de la investigación. Fuente: C. A. Cueva 2022

La Figura 16 muestra una tabla con los valores atípicos identificados en el monóxido de carbono, resaltando la eficacia de la minería de datos aplicada. Esta detección automática valida visualmente los resultados del proceso de exclusión. En este caso, se utilizó DB-Scan, un algoritmo de agrupamiento que identifica clústeres basados en la densidad de los datos (Khan, Rehman, Aziz, Fong, & Sarasvady, 2014).

Figura 16 Tabla de valores atípicos del monóxido de carbono detectados por el sistema en los datos de la segunda fase.

#	Usuario	Compuesto	Valores atipicos encontrados	Ocultar del análisis
1	Paciente 3	Monóxido de carbono	SI	ATIPICO
2	Paciente 5	Monóxido de carbono	SI	ATIPICO
3	Paciente 7	Monóxido de carbono	SI	ATIPICO

Análisis de las gráficas al tomar datos de la nariz electrónica

En la primera etapa del proyecto, se recopilaron datos con la nariz electrónica y se unificaron en un archivo Excel para generar gráficas. En la segunda fase, las gráficas se generaron mediante un nuevo sistema que utiliza Python para el procesamiento de datos y JavaScript para su visualización. Este sistema robusto almacena datos de sensores, identifica valores atípicos y produce gráficos estadísticos,

empleando técnicas como gráficos boxplot (cajas y bigotes), representaciones radiales, curvas de estandarización y la campana de Gauss para el análisis de datos.

Las Figuras 17 y 18 comparan los gráficos de cajas y bigotes del monóxido de carbono en ambas fases del estudio. La Figura 18 muestra la capacidad del nuevo sistema para generar visualizaciones detalladas automáticamente, mejorando la interpretación de los datos. Las nuevas gráficas permiten acercamientos y deslizamientos, facilitando el análisis, especialmente en gráficos de caja pequeños como se muestra en la Figura 19. Además, la distribución de los datos en cada cuartil se visualiza claramente, lo que ayuda a identificar patrones, densidades y acumulaciones en el gráfico.

Figura 17. Gráficas de cajas y bigotes del monóxido de carbono en la primera fase.

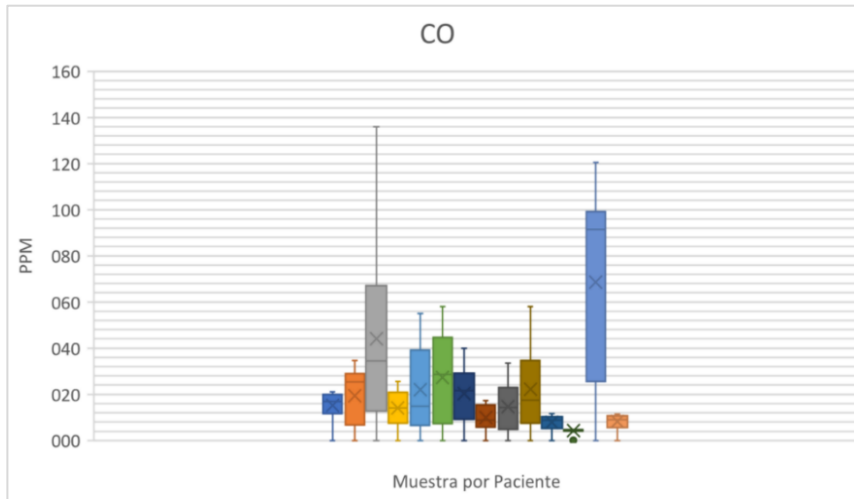


Figura 18 Gráficas de cajas y bigotes del monóxido de carbono en la segunda fase.

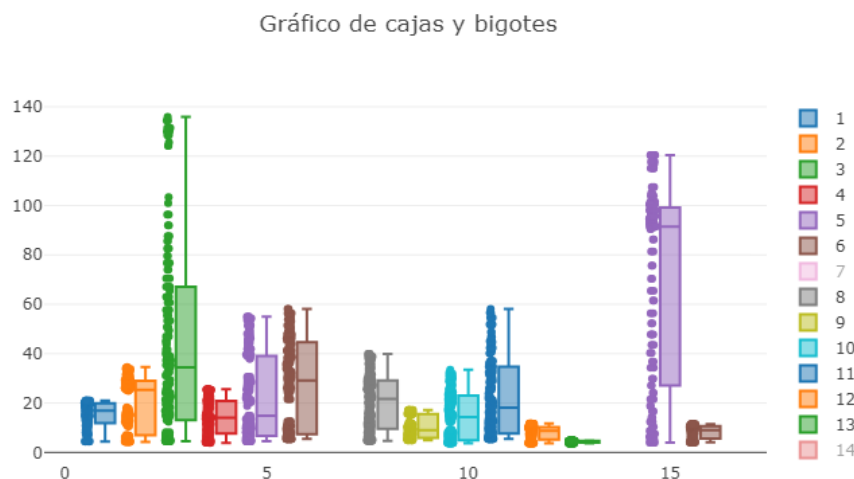
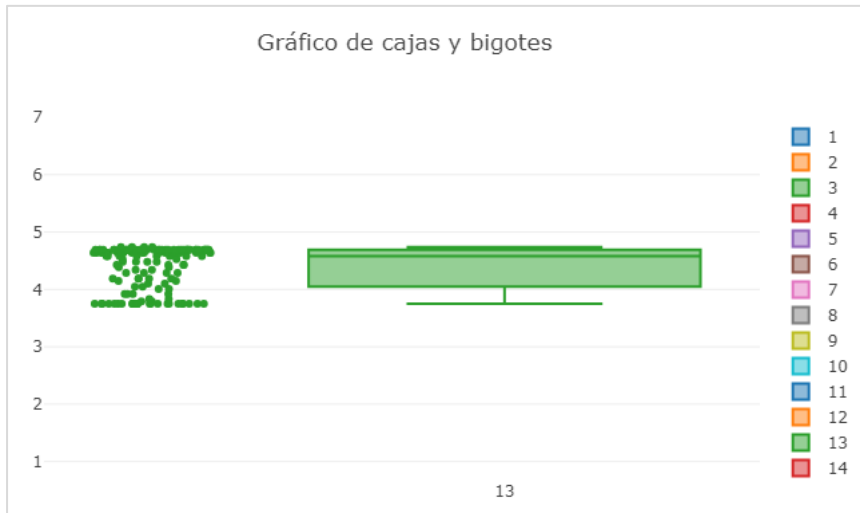


Figura 19 Gráficas de cajas y bigotes con acercamiento para mejor apreciación de los datos



Las Figuras 20 y 21 muestran la distribución radial del dióxido de carbono en ambas fases del estudio, demostrando que el sistema generó correctamente la gráfica radial. La Figura 21 presenta esta gráfica, la cual permite manejar la presentación de las curvas según la necesidad, ocultando visualmente información sin eliminarla de la base de datos. Esta funcionalidad es especialmente útil cuando las curvas se cruzan, ya que facilita el análisis enfocado al permitir que solo se muestren las curvas relevantes en un momento dado, ayudando a centrar la atención en los pacientes que se están analizando.

Figura 20 Gráfica radial del dióxido de carbono en la primera fase.

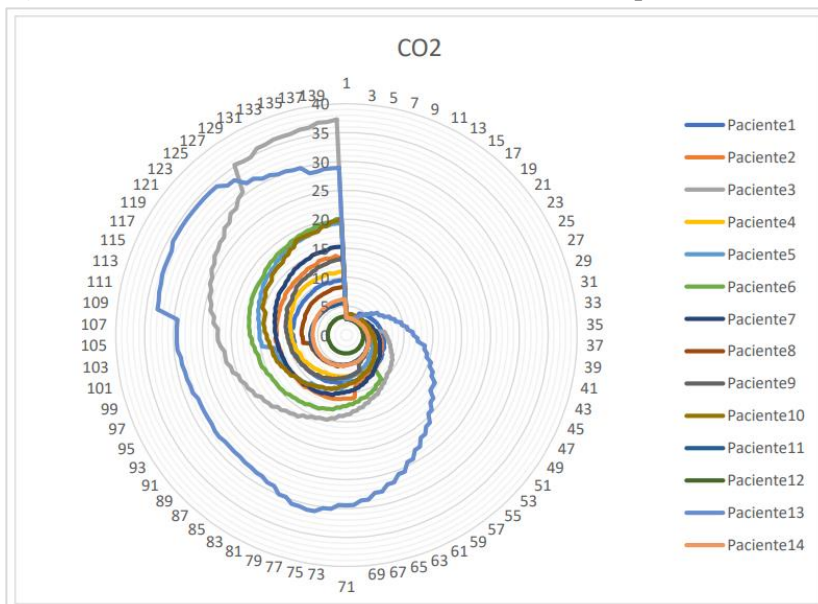
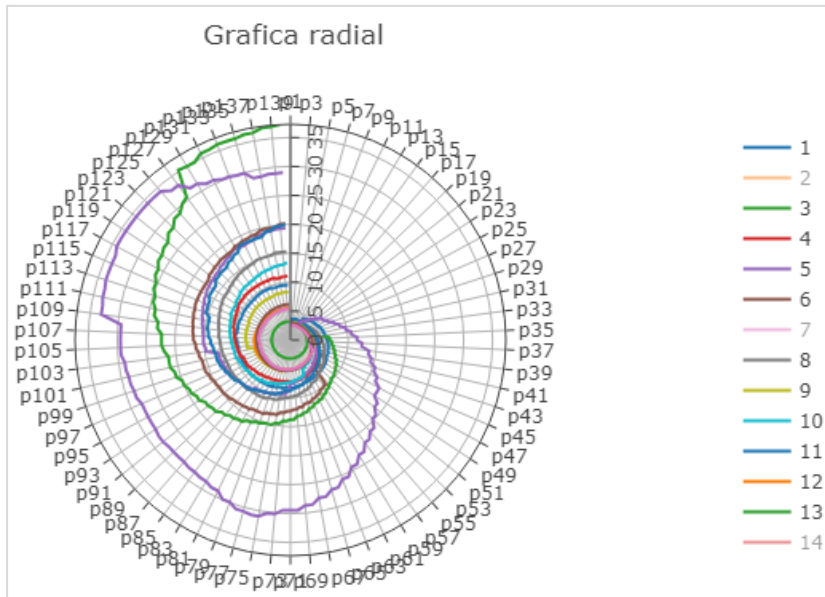


Figura 21 Gráfica radial del monóxido de carbono en la segunda fase.



Las Figuras 22 y 23 muestran las curvas de estandarización para el monóxido de carbono en cada fase del estudio, destacando valores representativos como el mínimo, la media y el máximo de cada paciente.

Figura 22 Curva de estandarización del Monóxido de carbono primera fase.

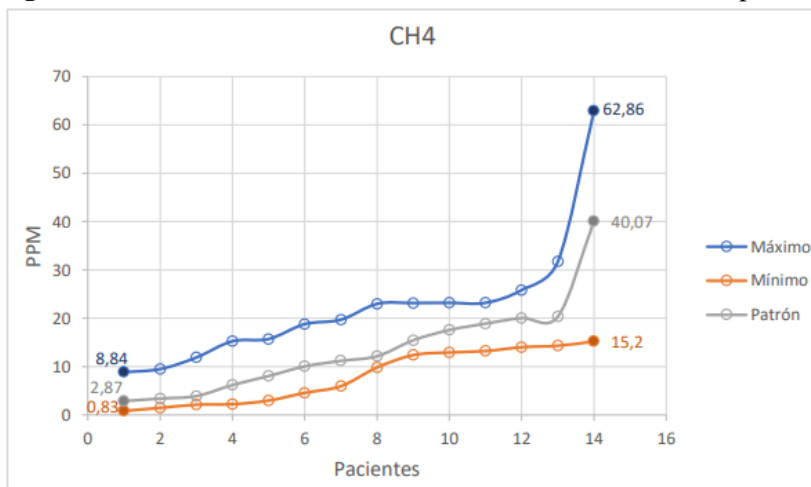
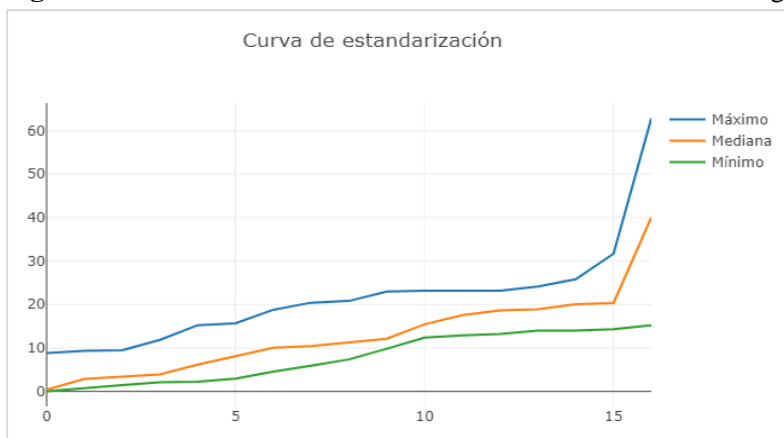


Figura 23 Curva de estandarización del Monóxido de carbono segunda fase..



Las distribuciones de la campana de Gauss para el ácido sulfhídrico presentadas en las Figuras 24 y 25, muestran consistencia entre ambas fases. La Figura 25 ilustra las curvas generadas por el sistema en la segunda etapa, donde, tras cargar la información en la base de datos, se aplican pruebas de normalidad, esenciales para generar la campana de Gauss. La generación de estas curvas requiere datos con tendencia normal, determinada por pruebas de normalidad. En caso de que no se detecte normalidad, se procede a transformar los datos para ajustarlos antes de su visualización en el frontend, que ofrece la opción de acercar, desplazar y ocultar no requeridas.

Este análisis resalta cómo los ajustes de normalidad mejoran la interpretación estadística, confirmando la capacidad del sistema para ajustarse a distribuciones normales post-transformación y proporcionar una herramienta robusta para la visualización de datos de sensores.

Figura 24 Campana de gauss del ácido sulfhídrico primera fase.

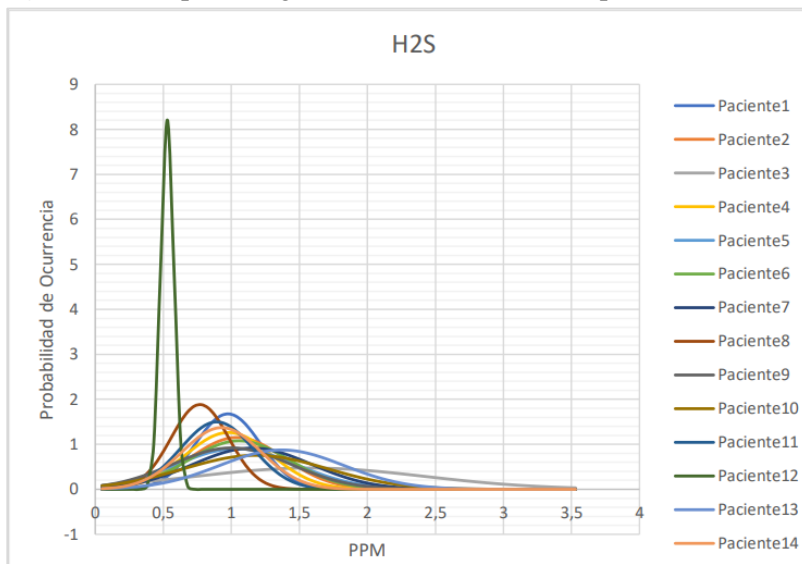
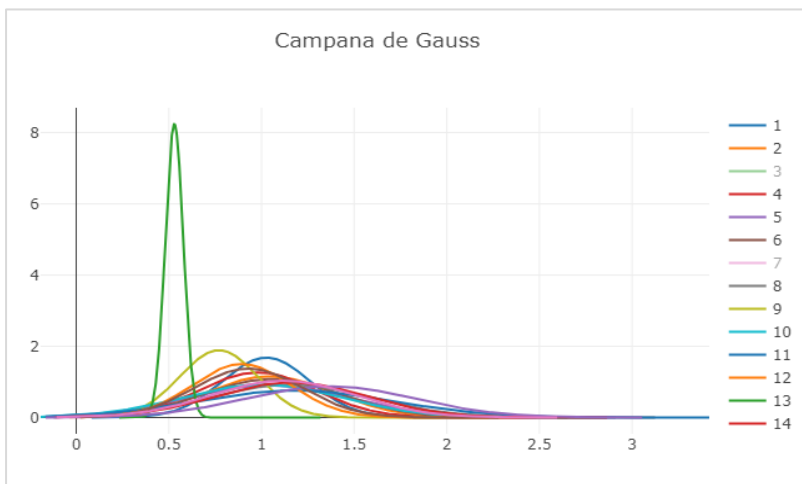


Figura 25 Campana de gauss del ácido sulfhídrico, segunda fase



DISCUSIÓN

Este estudio se centró en desarrollar un sistema que captura y almacena datos de los sensores y del estado de salud de los pacientes en una base de datos estructurada. Esto permite centralizar la información, facilitando su análisis y acceso desde cualquier ubicación, gracias al alojamiento en un servidor. Además, se priorizó un diseño responsive para garantizar su uso en diversos dispositivos.

Las funcionalidades clave del sistema incluyen la minería de datos y herramientas de análisis estadístico que asisten a los profesionales médicos en la elaboración de diagnósticos. Esta investigación se enfoca en proveer gráficos estadísticos que faciliten la interpretación de los datos por parte de profesionales de la salud, recordando que estas herramientas están diseñadas como un apoyo para el análisis médico.

Una etapa crítica en nuestro análisis fue la transformación de normalidad aplicada a los conjuntos de datos, lo que permitió el uso de curvas de Gauss. Este procedimiento es esencial cuando los datos no siguen una distribución normal, como se observó en las muestras. La transformación garantiza que las técnicas estadísticas, que asumen normalidad, puedan aplicarse de manera efectiva, facilitando una interpretación más precisa. Las curvas de Gauss resultantes destacan características y tendencias que serían menos evidentes en datos no transformados.

Al utilizar herramientas de software libre el sistema desarrollado permite dar un control total del procesamiento de la información sin necesidad del pago de una licencia por las herramientas y crear un sistema robusto que automatice partes del procesamiento de data de la nariz electrónica.

El uso de herramientas especializadas en cada aspecto del desarrollo del sistema permite responder adecuadamente a las necesidades. Python ofrece capacidades para la transformación de normalidad y análisis estadístico, mientras que React facilita la creación de aplicaciones visuales con diseño responsive. Además, el uso de Plotly, una herramienta especializada en la generación de gráficos, permite funcionalidades avanzadas como acercamientos y la visualización dinámica de cada línea de graficación dentro del sistema.

CONCLUSIONES

Al concluir esta etapa de investigación, se puede afirmar que el sistema desarrollado ha demostrado una notable eficiencia en la captura de compuestos en la exhalación de los pacientes mediante una nariz electrónica. Este sistema optimiza el procesamiento y la visualización de datos, facilitando a los



médicos la interpretación de las gráficas generadas. Una de sus principales ventajas es la sistematización del análisis, que centraliza el procesamiento, incluyendo pruebas y transformaciones de normalidad, minimizando errores y eliminando la necesidad de cálculos externos por parte de los usuarios.

Al comparar los resultados de la etapa 1 con los actuales, se concluye que la implementación del sistema ha mejorado significativamente la recogida, procesamiento y visualización de datos. La funcionalidad de generación de gráficas se ha completado, ofreciendo a los médicos gráficos estadísticos que facilitan una visualización y comparación más eficientes. Esta mejora refleja un avance considerable en la facilidad de uso y accesibilidad de la información.

Al concluir esta etapa de investigación, se ha desarrollado un sistema que procesa datos de la nariz electrónica utilizando la metodología Scrum y el modelado C4. Este enfoque ágil ha permitido un ciclo de desarrollo iterativo e incremental. Python fue elegido para el procesamiento de datos por sus extensas librerías de código abierto, ideales para análisis y estadística, mientras que JavaScript se seleccionó para la presentación de la información por su compatibilidad con herramientas de graficación. En resumen, la combinación de estas metodologías y tecnologías ha resultado en un sistema escalable, que sienta las bases para futuras integraciones y desarrollos.

El sistema ha demostrado ser capaz de identificar datos atípicos a partir de la información de la nariz electrónica. Sin embargo, el método DBSCAN no detectó todos los datos según los criterios médicos. Se sugiere explorar otros métodos, como K-means y análisis de cuartiles, y aumentar la cantidad de datos poblacionales en futuras investigaciones para mejorar la precisión y robustez del sistema.

REFERENCIAS BIBLIOGRAFICAS

Benedito, M. Á. P. (2013). Aplicación de los gráficos radiales en la evaluación del servicio público de transportes. *Auditoría Pública*, 60, 73-96.

Cayllahue Quille, R. (2020). Estudio de la nariz electrónica en la industria alimentaria. *SUNEDU*.

Charumpom, B., & Omatu, S. (2002, August). Classifying smokes using an electronic nose and neural networks. *In Proceedings of the 41st SICE Annual Conference (Vol. 5, pp. 2661-2665)*. IEEE.
<https://doi.org/10.1109/SICE.2002.1196116>

Cueva, C. A., & Erazo, K. G. (2022). *Diseño e implementación de una nariz electrónica para detección de compuestos orgánicos volátiles presentes en el aliento de seres humanos, en colaboración con*



- la Escuela de Medicina de la Universidad San Francisco de Quito (Tesis de ingeniería).*
Universidad de las Fuerzas Armadas ESPE, Ecuador.
- Di Rienzo, J. A., Casanoves, F., González, L. A., Tablada, E. M., Díaz, M. P., Robledo, C. W., & Balzarini, M. G. (2008). *Estadística para las ciencias agropecuarias (7ª ed.)*. Universidad Nacional de Córdoba.
- Durán Acevedo, C. M., Velásquez Carvajal, A. E., & Gualdron Guerrero, O. E. (2012). Implementación de una nariz electrónica para detectar pacientes con EPOC desde el aliento exhalado. *Ingeniería y Desarrollo*, 30(2), 143-159.
- Flores Tapia, C. E., & Flores Cevallos, K. L. (2021). Pruebas para comprobar la normalidad de los datos en procesos productivos: Anderson-Darling, Ryan-Joiner, Shapiro-Wilk y Kolmogórov-Smirnov. *Societas: Revista de Ciencias Sociales y Humanidades*, 23(2), 83-106.
- Gardner, J. W., Shin, H. W., Hines, E. L., & Dow, C. S. (2000). An electronic nose system for monitoring the quality of potable water. *Sensors and Actuators B: Chemical*, 69(1), 336-341.
[https://doi.org/10.1016/S0925-4005\(00\)00482-2](https://doi.org/10.1016/S0925-4005(00)00482-2)
- Gilat, A. (2005). *MATLAB: An introduction with applications (2ª ed.)*. John Wiley & Sons.
- Goschnick, J., Koronczai, I., Frietsch, M., & Kiselev, I. (2005). Water pollution recognition with the electronic nose KAMINA. *Sensors and Actuators B: Chemical*, 106(1), 182-186.
<https://doi.org/10.1016/j.snb.2004.05.055>
- Irigoin Cabrera, U. O., Goicochea Espino, J. O., Cabanillas Oliva, E. G., Sánchez Cáceres, V., Heredia Quezada, G. R., & Saldaña Ferreyra, K. (2023). *Métodos estadísticos para el control de la calidad y la productividad*. Editorial Mar Caribe.
- Jaimes Mogollón, A. L. (2020). *Desarrollo de un sistema de percepción sensorial artificial para la detección del cáncer gástrico en el aliento exhalado (Tesis doctoral)*. Universidad Complutense de Madrid, Madrid, España.
- Karakaya, D., Ulucan, O., & Turkan, M. (2020). Electronic nose and its applications: A survey. *International Journal of Automation and Computing*, 17(2), 179-209.
<https://doi.org/10.1007/s11633-019-1212-9>



- Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014, February). DBSCAN: Past, present and future. In **The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)** (pp. 232-238). *IEEE*.
- Lamagna, A., Reich, S., Rodríguez, D., Boselli, A., & Cicerone, D. (2008). The use of an electronic nose to characterize emissions from a highly polluted river. *Sensors and Actuators B: Chemical*, *131*(1), 121-124. <https://doi.org/10.1016/j.snb.2007.12.026>
- Li, Y., Wei, X., Zhou, Y., Wang, J., & You, R. (2023). Research progress of electronic nose technology in exhaled breath disease analysis. *Microsystems & Nanoengineering*, *9*(1), 129. <https://doi.org/10.1038/s41378-023-00594-0>
- López, A., Arazuri, S., García, I., Mangado, J., & Jaren, C. (2014). A review of the application of near-infrared spectroscopy for the analysis of fruits and vegetables. *Journal of Food Engineering*, *143*, 99-114. <https://doi.org/10.1016/j.jfoodeng.2014.07.019>
- Mendoza Montoya, J. J., & Ávila Navarro, E. (2017). Desarrollo de un sistema olfativo electrónico para el diagnóstico preliminar no invasivo de la diabetes. *Revista Científica Andina: Science & Humanities*, *1*(2), 157-176.
- Miller, C. D., Heeren, V. E., & Hornsby, J. (2005). *Matemática: Razonamiento y aplicaciones* (Vol. 10). Pearson Educación.
- Mildner-Szkudlarz, S., & Jeleń, H. H. (2008). The potential of different techniques for volatile compounds analysis coupled with PCA for the detection of the adulteration of olive oil with hazelnut oil. *Food Chemistry*, *110*(3), 751-761. <https://doi.org/10.1016/j.foodchem.2008.02.053>
- Minnaard, C., Condesse, V., Minnaard, V., & Rabino, C. (2005, May 10). Los gráficos de caja: Un recurso innovador. *Revista Iberoamericana de Educación*, *35*(8).
- Moreno, I., Caballero, R., Galán, R., Matía, F., & Jiménez, A. (2009). La nariz electrónica: Estado del arte. *Revista Iberoamericana de Automática e Informática Industrial RIAI*, *6*(3), 76-91
- Persaud, K. C., Pisanelli, A. M., Evans, P., & Travers, P. J. (2006). Monitoring urinary tract infections and bacterial vaginosis. *Sensors and Actuators B: Chemical*, *116*(1), 116-120. <https://doi.org/10.1016/j.snb.2005.12.062>

- Quicazán, M. C., Díaz, A. C., & Zuluaga, C. M. (2011). La nariz electrónica, una novedosa herramienta para el control de procesos y calidad en la industria agroalimentaria. *Vitae*, *18*(2), 209-217.
- Rabehi, A., Helal, H., Zappa, D., & Comini, E. (2024). Advancements and prospects of electronic nose in various applications: A comprehensive review. *Applied Sciences*, *14*(4506). <https://doi.org/10.3390/app14114506>
- Rial, A., & Varela, J. (2008). *Estadística práctica para la investigación en ciencias de la salud*. Netbiblo.
- Rodas, A., Ríos, J. I., & Solarte, G. R. (2016). Creación
- Rodríguez, J. E. (2010). *Fundamentos de minería de datos*. Universidad Francisco José de Caldas.
- Wilson, A. D., & Baietto, M. (2009). Applications and advances in electronic-nose technologies. *Sensors*, *9*(7), 5099-5148. <https://doi.org/10.3390/s90705099>
- Zhang, H., Wang, J., & Ye, S. (2008). Predictions of acidity, soluble solids and firmness of pear using electronic nose technique. *Journal of Food Engineering*, *86*(3), 370-378. <https://doi.org/10.1016/j.jfoodeng.2007.08.026>