

Implementation of a Computational System for Substances Identification Through its pH

1nd Derlis Hernández Lara
División de Ingeniería en
Informática
Tecnológico de Estudios
Superiores de Ecatepec
Ecatepec de Morelos, México
derlis392@hotmail.com

2nd Carlos Alfonso Trejo
Villanueva
División de Ingeniería en
Informática
Tecnológico de Estudios
Superiores de Ecatepec
Ecatepec de Morelos, México
carlostrejo@tese.edu.mx

3rd Emmanuel Tonatihu Juarez
Velázquez
División de Ingeniería en
Informática
Tecnológico de Estudios
Superiores de Ecatepec
Ecatepec de Morelos, México
ejuarezv@hotmail.com

4nd Litzy Paola Barón Velázquez
División de Ingeniería en
Informática
Tecnológico de Estudios
Superiores de Ecatepec
Ecatepec de Morelos, México
paolabaron15@gmail.com

Abstract— This paper presents the process of identifying some substances through its pH, for which a prototype embedded system including a color sensor and a sampling mechanism was designed and implemented. The identification of various kinds of substances is based on the coloration of pH measuring strips, the measurement done by the sensor returns a scale divided into three levels: red color level, green color level, and blue color level, which are interpreted by the system to classify the results and try to identify the substance. The analysis of the data obtained from the sampling is carried out with various techniques such as artificial neural networks and other classification algorithms, in order to establish what is the technique that performs the optimal identification of the tested substance. The results of the analysis are presented and compared, concluding that the system complies with the identification of the substances through the coloring of pH strips, and the proposed prototype could be useful for solving several engineering problems.

Keywords— Computational system, Classification algorithms, Substances classification through pH.

I. INTRODUCTION

En la actualidad, el desarrollo de tecnología que permite emular los sentidos humanos es una tendencia. Prototipos tecnológicos que funcionan como narices electrónicas, lenguas electrónicas, oídos electrónicos, sistemas de visión artificial y otros similares tienen un gran auge y muchos trabajos de investigación se desarrollan en este campo utilizando sistemas embebidos o sistemas computacionales. La necesidad de creación de prototipos que analicen la composición de los alimentos está en creciente demanda [1].

Estos prototipos incluyen sensores de múltiples tipos y el proceso de identificación se basa principalmente en sistemas de análisis químico [2]. El vertiginoso avance de la tecnología permite la medición de parámetros como color, olor, tamaño, forma, y otros, a través del desarrollo de sensores electrónicos y virtuales con el fin de obtener valores que puedan ser interpretados para la clasificación e identificación [3]. Muchos sistemas emuladores de los sentidos humanos, se desarrollan con el objetivo de ayudar a personas con discapacidades [4]. Sin embargo, los sistemas que permiten tener una percepción de la realidad física comienzan a tener un uso más amplio en áreas de la investigación como el análisis de patrones, los sistemas de

identificación personal, los sistemas de control de calidad, los sistemas autónomos aplicados a la robótica y otros [5].

Según la información descrita en [6], el pH es la unidad de medida que describe el grado de acidez o alcalinidad de una sustancia, es medido en una escala que va de 0 a 14 tomando en cuenta que dicho valor está directamente relacionado a la proporción de las concentraciones de los iones de hidrógeno (H⁺) e hidroxilo (OH⁻). Si la concentración de H⁺ es mayor que la de OH⁻, el material es ácido; el valor del pH es menor a 7; si es el caso contrario, el material es base, donde el pH tiene un valor mayor a 7. Si las cantidades de H⁺ y de OH⁻ son las mismas, el material es neutro y su pH es 7.

En esta investigación se propone un método que contempla la identificación del nivel de pH de una sustancia a partir de la interpretación de la coloración de tiras medidoras, las cuales son analizadas con un sistema embebido compuesto de un sensor de color y cuyos valores muestreados son registrados y procesados mediante un microcontrolador, el cual tiene la tarea de identificar el tipo de sustancia a partir de un análisis previo de los datos registrados por medio de diferentes técnicas estadísticas y de inteligencia artificial para clasificación.

II. DISEÑO E IMPLEMENTACIÓN DEL HARDWARE

Para el desarrollo de esta investigación se construyó un prototipo de alto nivel, funcional y económico que permite la identificación de algunas sustancias a partir de la medición su pH, se usa un sensor de color TCS3200, el cual tiene un arreglo de fotodiodos con 4 diferentes filtros, 16 fotodiodos con filtro que detectan la longitud de onda de los colores rojo, verde, y azul, una imagen ampliada muestra este arreglo en la Fig. 1.



Fig. 1. Acercamiento del Chip TSC3200.

El dispositivo para el procesamiento de los datos muestreados es un Arduino Nano v. 3.0, que es una pequeña y completa placa basada en el ATmega328 que trabaja a una frecuencia de reloj de 16 MHz, y dispone de 14 pines E/S digitales y 8 analógicos. Los sistemas Arduino corresponden a sistemas *Open Source Hardware* y son ampliamente usados en proyectos tecnológicos [7]. La imagen de esta placa se muestra en la Fig. 2. El embebido Arduino es en realidad un sistema que consta de tres bloques: un conjunto de placas (*Hardware* libre), un software y un lenguaje de programación libre que nació en el año 2005 [8].



Fig. 2. Placa Arduino Nano 3.0.

Los prototipos de sistemas embebidos que usan placas con tecnología desarrollada en la plataforma Arduino han aumentado en tiempos recientes [9, 10], por ende, la elección de esta placa para este trabajo.

La automatización del proceso en la toma de las muestras requiere de un sistema mecánico que permita tomar las tiras coloradas e introducirlas al prototipo para dejarlas muy próximas al sensor de color, para lo cual se hace uso de un mini servo motor *9G tower pro MG90S* con engranajes metálicos (Fig. 3), donde se tiene adaptada una terminal para poder tomar las tiras y establecer los tiempos de recolección, toma de muestra y el tiempo para retirar la tira. Los procesos de los tiempos de la muestra son manipulados por el microcontrolador del sistema embebido, el cual los controla con precisión definiendo su posición en cada uno de los procesos de la recolección de muestras y su posterior análisis.



Fig. 3. Mini servo motor MG90S.

Además, se implementa un LCD (*Liquid Cristal Display*) de 16x2, con un adaptador de interfaz I²C para la optimización en el uso de pines de la placa, la pantalla tiene como objetivo mostrar al usuario el estado en el que se encuentra el proceso de la identificación y toma de muestra; además se utiliza un pulsador simple para controlar algunas funciones e inicio de operaciones del sistema como se muestra en la Fig. 4.



Fig. 4. LCD 16X2 con adaptador I²C y pulsador.

Para la integración del prototipo se conectan todos los dispositivos previamente mencionados a la placa Arduino Nano, la cual tiene el control sobre cada uno en los diferentes procesos de toma de muestra e identificación de las mismas. En principio se muestra la conexión del servomotor a una de las terminales digitales disponibles en el sistema embebido (Fig. 5), que es definida en el código del programa. Cabe mencionar que el servomotor solo requiere de las conexiones de alimentación y un pin especial con el que se controlará su posición.

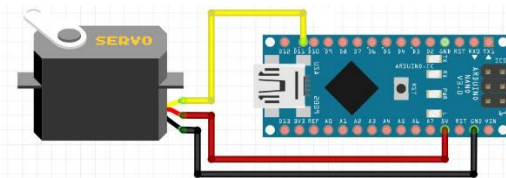


Fig. 5. Conexión del servomotor MG90S a la placa Arduino.

A continuación, se muestra la conexión del LCD a través de el adaptador I²C y el sensor TCS3200, que están conectados a dos terminales del puerto analógico y a cinco terminales del puerto digital respectivamente, como se observa en la Fig. 6.

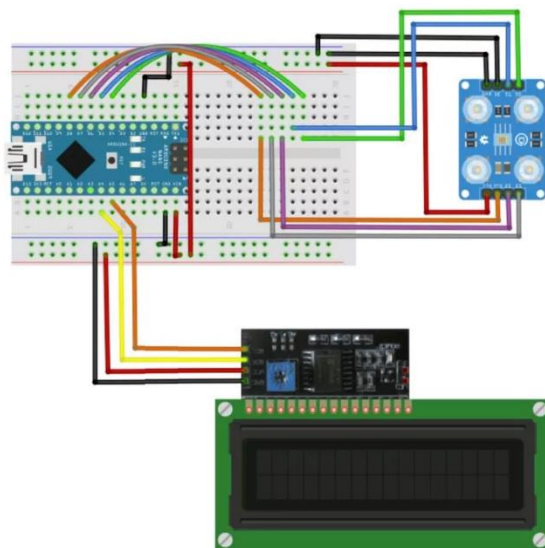


Fig. 6. Conexión del sensor de color TCS3200 y el LCD con el adaptador I²C a la placa Arduino.

La estructura final del prototipo integrado y finalizado se encapsula en un contenedor como se muestra en la Fig. 7. Se busca aislar las muestras teniendo las paredes adyacentes de color negro y solo la cara contraria al sensor de luz de color blanco, para permitir una mejor toma de muestra del color de las tiras analizadas sin interferencias por otros colores reflejados.



Fig. 7. Prototipo final implementado.

Para la medición de nivel de pH , se utilizan tiras estándar con escala de medición 1-14 que vienen acompañadas con su propia escala de identificación de nivel, las cuales se muestran en la Fig. 8.



Fig. 8. Tiras utilizadas para la medición de pH .

Así como en la presente investigación se hace uso de la medición de nivel de pH , hay muchas otras en las que sus resultados se basan en el análisis de dichos niveles para el establecimiento de sus resultados [11, 12]; no obstante, el método de obtención del nivel de pH se realiza comúnmente con sondas que analizan fluidos líquidos.

III. MÉTODO

El análisis para la identificación de sustancias a partir de la medición en tiras del nivel de pH por medio de un sensor de color se realiza como se muestra en la Fig. 9. La señal de salida entregada por el sensor permite la interpretación de la luz incidente en él, la cual es filtrada para la sensibilidad de la longitud de onda de la luz correspondiente (de los primeros tres arreglos), y un arreglo más que no tiene filtro y que sirve para la referencia de los valores obtenidos.

Los cuatro tipos de fotodiodos del sensor (que detectan la longitud de onda de los colores) están interdigitados para minimizar el efecto de la no uniformidad del incidente irradiancia. Todos los fotodiodos del mismo color están

conectados en paralelo. Los pines S2 y S3 del sensor se usan para seleccionar que grupo de fotodiodos (rojo, verde, azul, claro) están activos y los pines S0 y S1 se usan para la selección de la escala de frecuencia de salida cómo se muestra en la TABLA I y II respectivamente.

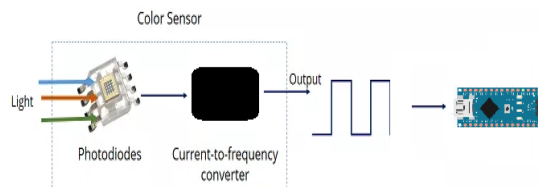


Fig. 9. Funcionamiento del sistema de detección del pH .

La identificación de los colores está definida por la respuesta espectral de los fotodiodos, cómo se muestra en la Fig. 10, la cual ha sido tomada de las hojas técnicas proporcionadas por el fabricante del TCS3200, *Texas Advanced Optoelectronic Solutions* (TAOS).

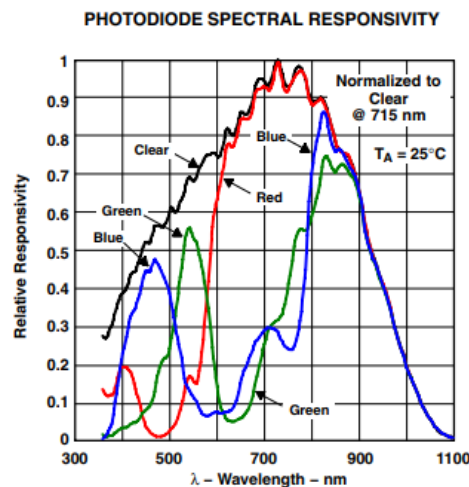


Fig. 10. Respuesta espectral de los fotodiodos del sensor de color TCS3200 para las longitudes de onda de los colores rojo, verde y azul.

TABLA I. CONFIGURACIÓN DEL CONVERTIDOR DE FRECUENCIA A CORRIENTE DEL TCS 3200

S0	S1	ESCALA DE FRECUENCIA DE SALIDA
L	L	APAGADO
L	H	2 %
H	L	20 %
H	H	100 %

TABLA II. CONFIGURACIÓN DEL CONVERTIDOR A TIPO DE FOTODIODO

S0	S1	TIPO DE FOTODIODO
L	L	ROJO
L	H	AZUL
H	L	SIN FILTRO
H	H	VERDE

El proceso de la toma de muestras se define por medio de un *software* desarrollado para el sistema Arduino, diseñado como una máquina de estados, la cual lleva la siguiente secuencia:

- **Estado 0:** Inicio del LCD y despliegue de mensaje de solicitud de inicio.
- **Estado 1:** Monitoreo de la pulsación del botón de inicio y preparación de la recepción de la tira para la obtención de la muestra a analizar con extensión del sujetador integrado al servo motor.
- **Estado 2:** Retracción del sujetador con la tira para su análisis.
- **Estado 3:** Inicio del sensor de color.
- **Estado 4:** Activación secuencial de los arreglos para la identificación de las longitudes de onda de los colores rojo, verde y azul, posteriormente toma de muestra y acumulación de 100 lecturas por cada segundo en un lapso de 6 mediciones para obtener la media del valor estabilizado y finalmente identificación de los valores de escala Rojo, Verde y Azul para la identificación de la sustancia o alimento sometido.
- **Estado 5:** Extensión del sujetador para retirar la tira medidora de pH y re inicio del sistema para la siguiente prueba.

IV. DESARROLLO DEL SOFTWARE DEL PROTOTIPO

El sistema desarrollado utiliza 3 diferentes aplicaciones de *software* las cuales son descritas a continuación:

A. Software en Arduino de muestreo para la obtención de datos de las mediciones de pH en las tiras.

Este *software* se diseña para trabajar con el prototipo de *hardware* construido. Su función principal es solicitar tiras y posteriormente realizar 100 muestreos divididos en 6 lapsos, para así, calcular la media de dichas muestras obteniendo un valor promedio del color registrado por los sensores en escala RGB, los cuales después son enviados por el puerto serial del prototipo para su almacenamiento en un archivo de datos CSV (*Comma-Separated Values*), en el que se registrarán las mediciones de cada tira analizada.

B. Interfaz de Software desarrollado en lenguaje Java para la obtención de las muestras de datos en escala RGB de las mediciones de pH en las tiras y su almacenamiento en archivos CSV.

Este *software* tiene como objetivo la obtención de los muestreos realizados por el prototipo y el *software* previamente descrito. La interfaz desarrollada permite elegir el tipo de alimento registrado (Fig. 11) y por medio de comunicación serial lee las muestras tomadas desde el prototipo, para mostrarlas en un cuadro de texto; posteriormente, cuando se tiene la cantidad de muestras deseadas por el usuario, permite almacenar el set de datos en dos archivos diferentes: uno que registra los valores RGB obtenidos por el sensor de color y el embebido del prototipo, y otro más para almacenar un arreglo

con los valores objetivo de los alimentos que se desea identificar.

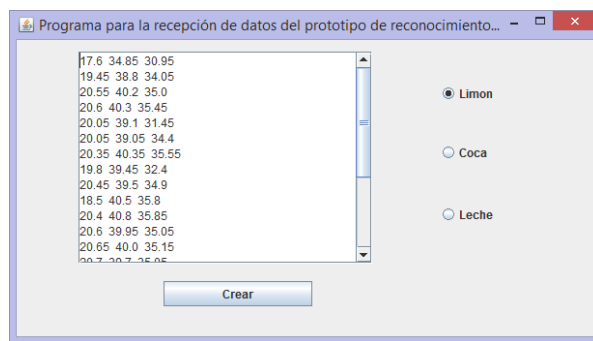


Fig. 11. Interfaz desarrollada en Java para la obtención de las muestras de datos de pH y su almacenamiento en archivos CSV.

C. Software para el análisis e identificación de alimentos y sustancias.

El *software* final realiza la toma de la muestra de algún alimento a través de los archivos generados por el *software* B, con los cuales procesa la información obtenida de los muestreos basándose en el análisis por medio de algunos métodos para su posterior identificación. El algoritmo y método elegido para la identificación puede ser cambiado para mejorar la precisión del reconocimiento de las sustancias sometidas. Los métodos de análisis se describen en la sección de resultados.

V. RESULTADOS Y DISCUSIÓN

Con la descripción de las secciones anteriores se describe la creación de dos archivos de datos obtenidos de los muestreos de las tiras indicadoras de pH que se asocian con valores en escala RGB correspondientes a diferentes sustancias sometidas. Una vez obtenidas las muestras de los valores RGB del prototipo y ya almacenadas en archivos CSV, estas son utilizadas para realizar pruebas de clasificación para la identificación de las sustancias con el uso de dos herramientas y diferentes métodos de análisis. Las herramientas utilizadas son: *NprTool de Matlab* y *SPSS de IBM*, mientras que los métodos de análisis usados en estas herramientas fueron: *Redes Neuronales Artificiales (RNA)*, *Clúster Bietápico*, *Análisis por Discriminante*, y *Análisis por KNN*.

A. Análisis con NprTool de Matlab

Para comenzar el análisis se utiliza la herramienta *NprTool (Neural network pattern recognition tool)* de Matlab (Fig. 12), que sirve para resolver de problemas de clasificación y reconocimiento de patrones utilizando una red *PatternNet* configurada para trabajar con una alimentación de diferente número de capas para sus pruebas y una función de activación sigmoide. Las redes neuronales artificiales son muy utilizadas en los procesos de clasificación por la certeza alcanzada en sus resultados [13, 14]. La red neuronal artificial se diseñó haciendo pruebas con 8, 10, 12 y 15 capas de neuronas para realizar comparativas de la efectividad en el reconocimiento y certeza en la clasificación.

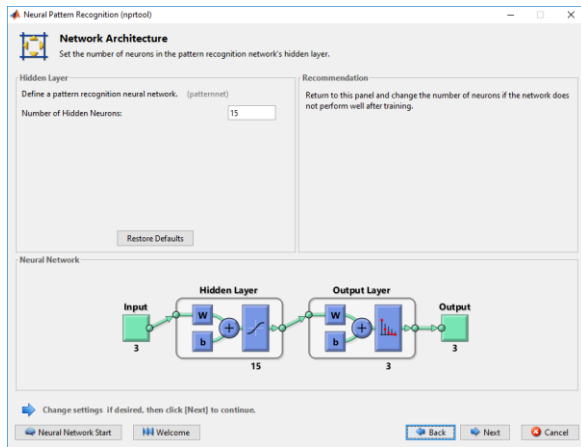


Fig. 12. Herramienta NprTool para la clasificación basada en el reconocimiento de patrones por una red neuronal artificial.

Los mejores resultados se obtienen con las RNA de 10 y 15 capas ocultas de neuronas, lo cual se puede confirmar con la matriz de confusión y la gráfica de performance de mejor validación mostradas en las Fig. 13 y 14 respectivamente, en donde se indica el número de épocas utilizado para alcanzar el resultado esperado en el entrenamiento para el experimento de 10 capas.

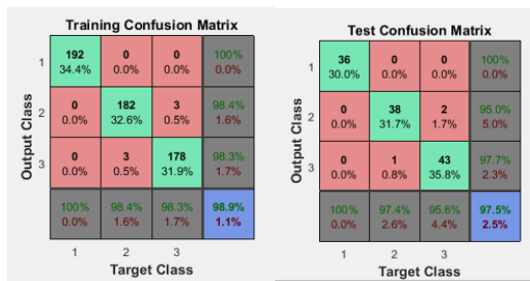


Fig. 13. Matriz de confusión de entrenamiento para una RNA con 10 capas.

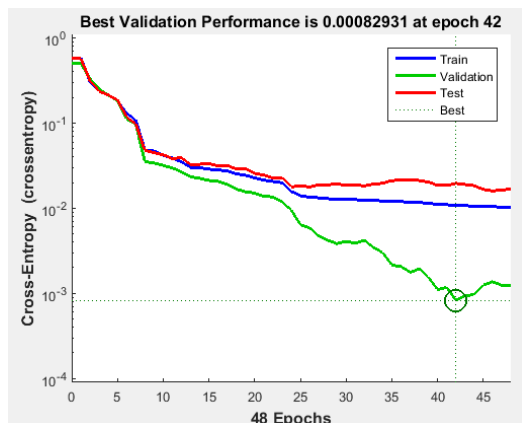


Fig. 14. Gráfica de performance para una RNA con 10 capas.

En ambas pruebas de redes de 10 y 15 capas se alcanzó un nivel de confiabilidad que va del 96.7% al 97.5% respectivamente, las cuales fueron obtenidas con los datos RGB

del prototipo y los tres elementos que se deseaba identificar en el experimento, limón, refresco y leche y que denotan la confiabilidad de la clasificación basada en el método de análisis por RNA.

B. Análisis con SPSS de IBM

SPSS es una plataforma gráfica de *data science* y analítica predictiva, que permite analizar conjuntos de datos estructurados y no estructurados utilizando algoritmos de *machine learning* y capacidades de manipulación de datos con opciones automatizadas.

El primer análisis realizado con SPSS es el del método de clasificación por clúster bietápico, que es una herramienta de exploración diseñada para revelar agrupaciones naturales dentro de un conjunto de datos que de otro modo no serían evidentes. Cabe mencionar que, las pruebas que implicaron 3 diferentes sustancias para ser reconocidos contaron con un número similar de muestras, lo cual se puede observar en el reconocimiento de los grupos al tener prácticamente el mismo porcentaje de distribución en su reconocimiento. En la primera sección de la TABLA III entregada por SPSS y la herramienta de distribución de clústeres se observa el total de las muestras analizadas y clasificadas en cada uno de los grupos por separado, además se observa el porcentaje parcial de cada uno de los clústeres identificados sobre el total de la muestra sometida. También se muestra el estado de cada clúster por separado indicando el porcentaje de coincidencia y diferencia entre cada uno de los clústeres y el clúster seleccionado en cuestión. Los resultados de estas tablas muestran la eficacia de la clasificación a partir de los resultados en los que se puede verificar a partir de los porcentajes presentados en la sección identificada como “*.00*”, que identifica a los elementos excluidos del grupo que corresponden con los elementos de los otros clústeres y los elementos presentados en la sección identificada como “*1.00*”, los cuales corresponden a los elementos clasificados dentro del clúster que corresponden con los elementos de uno de los elementos en particular.

El segundo tipo de análisis con SPSS fue el del método de clasificación por discriminante, que es una técnica estadística multivariante para analizar si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas sobre los mismos. Los resultados de la clasificación muestran que los porcentajes de identificación de cada una de las sustancias por separado va desde un 93.3% hasta un 99.9% de certeza.

El siguiente procedimiento fue el método del vecino más cercano, el cual sirve para clasificar casos en función de su similitud con otros. Los resultados de este análisis muestran una identificación muy cercana al 33%.

En la Fig. 15 se muestra el espacio de predictores basado en el reconocimiento de los valores RGB para uno de las sustancias a identificar, haciendo la observación de que se eligió la búsqueda de 3 vecinos para la clasificación de los 3 elementos.

TABLA III. DISTRIBUCIÓN DE GRUPOS PARA LAS SUSTANCIAS DE ESTUDIO CON SPSS.

Distribución del clúster				
Clúster	N	% de combinado	% del total	
1	273	34.5	34.5	
2	267	33.7	33.7	
3	252	31.8	31.8	
Combinado	792	100.0	100.0	
Total	792		100.0	

0.00		Limón			1.00	
Clúster	Frecuencia	Porcentaje (%)	Frecuencia	Porcentaje (%)		
1	0	0.0	273	100.0		
2	267	51.4	0	0.0		
3	252	48.6	0	0.0		
Combinado	519	100	273	100		

0.00		Coca			1.00	
Clúster	Frecuencia	Porcentaje (%)	Frecuencia	Porcentaje (%)		
1	273	50.6	0	0.0		
2	267	49.4	0	0.0		
3	0	0.0	252	100.0		
Combinado	540	100.0	252	100.0		

0.00		Leche			1.00	
Clúster	Frecuencia	Porcentaje (%)	Frecuencia	Porcentaje (%)		
1	273	52.0	0	0.0		
2	0	0.0	267	100.0		
3	252	48.0	0	0.0		
Combinado	525	100.0	267	100.0		

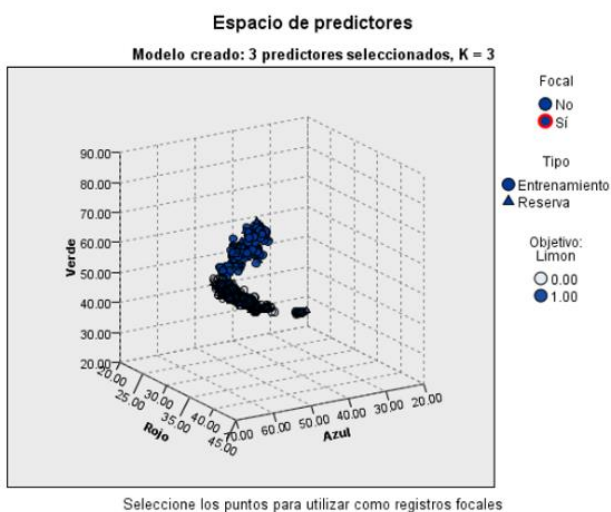


Fig. 15. Espacio de predictores para uno de los elementos en la clasificación por análisis del vecino más cercano de SPSS.

VI. CONCLUSIONES

Se ha logrado diseñar e implementar un sistema computacional mediante la construcción de un prototipo de alto nivel y el uso de software y algoritmos de clasificación, con el objetivo de poder identificar sustancias a partir de la toma de muestras con tiras medidoras de pH, analizando la coloración obtenida para después revisar las muestras y con el uso de técnicas de análisis como redes neuronales artificiales, clúster bietápico, método del discriminante y el método del vecino más cercano, lograr la clasificación o identificación de las sustancias obteniendo resultados precisos y satisfactorios.

El prototipo se desarrolla con dispositivos electrónicos simples, económicos y que son tendencia tecnológica, mientras que las aplicaciones desarrolladas son en su totalidad creadas con *software* libre, lo que reduce costos y hace que la propuesta tecnológica sea viable. Los métodos de análisis utilizados entregaron resultados que rondan entre el 90% y el 99% de certeza en la identificación de sustancias, por lo que la propuesta del dispositivo es alentadora y cumple con el propósito para el cual fue concebido, además de ser simple y económica.

VII. BIBLIOGRAFÍA

- Á. Arrieta Almario y O. Fuentes Amín, «Lengua electrónica portátil para el análisis de leche cruda basada en tecnología PSoc (Programmable System on Chip) y Android.» *Ingeniare: Revista Chilena de Ingeniería*, vol. 24, n° 3, pp. 445-453, 2016.
- Á. Arrieta Almario, O. Fuentes Amín y J. Díaz, *Revista de la Sociedad Química del Perú*, vol. 82, n° 2, pp. 152-161, 2016.
- P. Rosado Rodrigo, E. Figueras Ferrer, M. Planas Rosselló y F. Reverter Comes, «La visión artificial, un nuevo aliado para el análisis de imágenes artísticas.» *Arte, Individuo y Sociedad*, vol. 28, n° 2, pp. 239-354, 2016.
- I. Aguirre Gil y J. Gaviria, «Sistema de movimiento del cursor del computador para personas con discapacidad utilizando técnicas de visión artificial.» *Multiciencias*, vol. 15, n° 3, pp. 328-338, 2015.
- F. Martínez Santa, S. Orjuela Rivera y M. Arbulú Saavedra, «Global navigation approach for assistant robot.» *Tecnura*, vol. 21, n° 51, pp. 105-117, 2017.
- O. a. s. company, «Omega.» México, 8 enero 2020. [En línea]. Available: <https://mx.omega.com/prodinfo/medidores-de-ph.html>.
- C. X. Rosero Chandi y C. F. Vaca Orellana, «Metodología para evaluación de usabilidad del entorno de desarrollo integrado de Arduino.» *SATHIRI*, vol. 13, n° 1, pp. 214-226, 2018.
- M. Ázua Barrón, M. Vázquez Peña, R. Arteaga Ramírez y R. Hernández Saucedo, «Sistema de adquisición de datos de bajo costo con la plataforma arduino.» *Revista Mexicana de Ciencias Agrícolas*, vol. 8, n° 1, pp. 1-12, 2017.
- P. Pérez Díaz, «Sistema embebido basado en el monitoreo de consumo de energía eléctrica de electrodomésticos para el control del gasto.» *Revista Tecnología Digital*, vol. 8, n° 1, pp. 69-80, 2018.
- L. J. Silva Díaz y Y. Morejón Mesa, «Secadores solares artesanales para especias y control de temperatura.» *Revista Ingeniería Agrícola*, vol. 9, n° 3, pp. 65-68, 2019.
- J. Connor, M. B. Yatvin y L. Huang, «pH-sensitive liposomes: Acid-induced liposome fusion.» *Proceedings of the National Academy of Sciences*, vol. 81, n° 6, pp. 1715-1718, 1984.
- J. Liu, Y. Huang, A. Kumar, A. Tan, S. Jin, A. Mozhi y X.-J. Liang, «pH-sensitive nano-systems for drug delivery in cancer therapy.» *Biotechnology advances*, vol. 32, n° 4, pp. 693-710, 2014.
- M. Culebro de la Piedra, «Modelo neurodifuso para el control de humedad del suelo en cultivo hidropónico para la planta de tomate.» *Revista Tecnología Digital*, vol. 6, n° 1, pp. 43-56, 2016.
- C. M. Pinilla Gomez, S. M. Rosenstiehl, E. Gomez Flor, F. Gonzalez, O. Perdomo y F. J. Rodríguez Alvira, Red neuronal artificial para la detección de edema macular diabético en imágenes de tomografía de coherencia óptica: estudio transversal de prueba diagnóstica., Colombia: Universidad del Rosario, 2018.