

# Algoritmos de aprendizaje automático para identificación de patrones SERS de pesticidas organofosforados

Mateo Frausto-Avila, Monserrat Ochoa-Elias,  
Gonzalo Ramírez-García, Mario Alan Quiroz-Juarez

Universidad Nacional Autónoma de México,  
Centro de Física Aplicada y Tecnología Avanzada,  
México

{mfrausto, maqj}@fata.unam.mx

**Resumen.** El estudio de los pesticidas organofosforados presenta obstáculos sustanciales tanto para el medio ambiente como para la salud pública. Las peligrosas propiedades de estos insecticidas han tenido impactos adversos en las personas y otras criaturas, lo que ha generado importantes desafíos en los países en desarrollo. Las técnicas existentes son complejas y requieren mucho tiempo, y existe un gran interés dentro de la comunidad científica de crear formas más rápidas y precisas. Este estudio utiliza las capacidades de autoaprendizaje de las redes neuronales artificiales para mejorar las capacidades analíticas de los espectros SERS. Los algoritmos de aprendizaje automático se entrenan utilizando patrones SERS que se adquieren directamente de las mediciones experimentales. Los hallazgos sugieren que el aprendizaje automático ofrece una mayor precisión en el análisis de espectros, creando así oportunidades para desarrollar técnicas analíticas más rápidas y eficientes para detectar pesticidas organofosforados en diferentes matrices ambientales.

**Palabras clave:** Redes neuronales artificiales, SERS, paratión, edifenfos.

## Machine Learning Algorithms for Identification of SERS Patterns of Organophosphate Pesticides

**Abstract:** The study of organophosphate pesticides presents substantial obstacles for both the environment and public health. The dangerous properties of these insecticides have had adverse impacts on people and other creatures, creating significant challenges in developing countries. Existing techniques are complex and time-consuming, and there is a great interest within the scientific community to create faster and more accurate ways. This study uses the self-learning capabilities of artificial neural networks to improve the analytical abilities of SERS spectra. Machine learning algorithms are trained using SERS spectra, which are acquired directly from experimental measurements. The findings suggest that machine learning offers greater precision in spectral analysis, thus creating opportunities to develop faster and more efficient analytical techniques to detect organophosphate pesticides in different environmental matrices.

**Keywords:** Artificial neural networks, SERS, parathion, edifenphos.

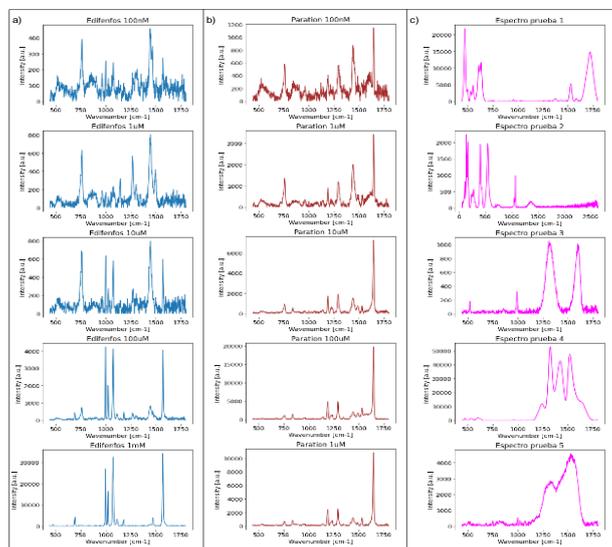
## 1. Introducción

La exposición a pesticidas organofosforados plantea desafíos significativos en términos de salud humana y cuidado del medio ambiente [1]. Estos pesticidas desempeñan un papel crucial en el aumento de la producción agrícola al prevenir y controlar una amplia variedad de plagas, pero estudios han sugerido que su presencia tiene efectos tóxicos en humanos y otros mamíferos [2]. Este hecho ha generado problemas notables en países en desarrollo, donde la falta de conocimiento generalizado sobre los riesgos asociados con estos pesticidas ha resultado en numerosos efectos nocivos entre los trabajadores agrícolas [3].

La determinación de residuos de pesticidas organofosforados en alimentos es de suma importancia debido a su impacto potencial en la salud humana. La mayoría de los métodos disponibles para esta determinación son complicados y laboriosos. Por esta razón, el desarrollo de métodos rápidos y precisos para la determinación de estos pesticidas es de gran interés para la comunidad científica. Existen algunos métodos de determinación rápida, los cuales se dividen en dos categorías principales: métodos electroquímicos [4-7] y espectrales [8-11]. Si bien, los métodos electroquímicos presentan ventajas en términos de tiempos de determinación cortos, alta sensibilidad y bajo costo, la necesidad de sintetizar reactivos intrincados los hace inadecuados para una amplia variedad de aplicaciones. Por otro lado, los métodos espectrales, como la espectrometría de fluorescencia, infrarroja y Raman, son rápidos y emplean procedimientos simples [12].

Entre las técnicas espectrales, la tecnología de dispersión Raman mejorada en superficie (Surface Enhanced Raman Spectroscopy, SERS por sus siglas en inglés) ha surgido como una herramienta poderosa para la determinación rápida de pesticidas [13]. Los sustratos SERS se basan principalmente en nanopartículas de metales nobles (Ag, Au), y son capaces de intensificar las señales Raman con un factor de mejora (EF) de hasta 107 veces, permitiendo identificar los pesticidas con alta sensibilidad y precisión [14,15]. Sin embargo, el análisis de los espectros es complejo y usualmente tardado, impidiendo que esta estrategia analítica se aplique de manera masiva. A este respecto, es necesario desarrollar estrategias que permitan simplificar y automatizar el análisis de datos. Actualmente, el aprendizaje automático (machine learning, ML por sus siglas en inglés) se ha convertido en una potencia innovadora, demostrando su capacidad para detectar patrones complejos y ofrecer nuevas perspectivas en numerosos ámbitos científicos [16,17]. El campo de la investigación ha experimentado una profunda revolución como resultado de la capacidad de los algoritmos de aprendizaje automático para revelar conexiones intrincadas y ocultas dentro de vastos conjuntos de datos [18].

En este trabajo, explotamos las características de autoaprendizaje de redes neuronales artificiales (Neural Networks, NN por sus siglas en inglés) para potenciar la capacidad analítica de los espectros SERS. Estos algoritmos de ML son entrenados con patrones SERS directamente obtenidos del espectrómetro Raman. En este punto, resaltamos que nuestro algoritmo no demanda preprocesamiento de datos además de que permanece compacto, al contar únicamente con una capa oculta de 5 neuronas. Se analizaron espectros SERS con diferentes concentraciones de paratión y edifenfos, dos



**Fig. 1.** Ejemplos de espectros SERS en concentraciones variables (100 nM, 1  $\mu$ M, 10  $\mu$ M, 100  $\mu$ M y 1mM) para: a) Edifenfos, b) Paratión y c) compuestos desconocidos.

pesticidas usados como insecticida y fungicida respectivamente. Mediante un modelo de redes neuronales de clasificación se pudieron identificar exitosamente los espectros SERS de ambas moléculas, así como una tercera molécula de prueba desconocida que sirve como tercera clase en la clasificación. Los resultados indican que ML proporciona una mayor precisión en el análisis de los espectros, lo que abre nuevas posibilidades para avanzar hacia métodos analíticos más rápidos y eficientes que permitan la detección de pesticidas organofosforados en diversas matrices ambientales y alimentos.

## 2. Metodología

### 2.1. Preparación de sustratos SERS

Las nanoestrellas de oro se sintetizaron conforme a la metodología reportada por Panikar et al [19]. Brevemente, una disolución 0.25 de  $\text{HAuCl}_4$ , volúmenes fijos de semillas de oro previamente sintetizadas mediante el método de Turkevich [20], y ácido clorhídrico (HCl) 1 M fueron mezclados con agitador magnético durante 2 minutos. Después, se añadió nitrato de plata ( $\text{AgNO}_3$ ) 3 mM y de ácido ascórbico 0.1M ( $\text{C}_6\text{H}_8\text{O}_6$ ). La reacción se continuó durante 3 minutos antes de agregar 2 mL de solución de bromuro de hexadeciltrimetilamonio (CTAB). Finalmente, se dejó en agitación durante 30 minutos. Por último, se centrifugaron y enjuagaron 2 veces a 600 revoluciones por minuto (rpm) por 10 minutos.

Volúmenes fijos de las nanoestrellas fueron depositados sobre sustratos de aluminio, y se permitió que se secaran a temperatura ambiente. Finalmente, se añadieron concentraciones variables (100 nM, 1  $\mu$ M, 10  $\mu$ M, 100  $\mu$ M y 1mM) de los pesticidas sobre la muestra anterior, y se permitió nuevamente que se secaran a temperatura

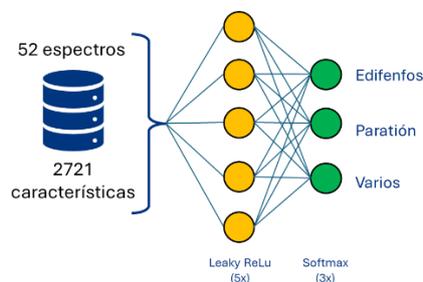


Fig. 2. Modelo de la red neuronal empleada.

ambiente. Los espectros se obtuvieron con el espectrómetro Raman Senterra de Bruker, con un láser de 785 nm. Espectros representativos de las concentraciones analizadas pueden verse en la Fig. 1.

## 2.2. Obtención de los espectros SERS

Los espectros Raman de los pesticidas paratión y edifenfos se obtuvieron mediante la técnica SERS con nanoestrellas de oro, las cuales generan un incremento del campo electromagnético debido a su plasmón de resonancia. En la molécula edifenfos las bandas 1000, 1020 y 1080  $\text{cm}^{-1}$  corresponden a los enlaces C=C, C-H y C-S, respectivamente. En el paratión las bandas 760, 1347 y 1590  $\text{cm}^{-1}$  representan los enlaces de P=S, N-O y C-H.

## 2.3. Preparación de la base de datos

Para la creación de la base de datos, se emplearon 52 espectros SERS; 18 espectros de edifenfos, 19 de paratión y 15 de compuestos desconocidos, estos últimos fueron usados con el fin de propiciar variabilidad a la clasificación de los espectros. Cada espectro está conformado de 2721 filas y 2 columnas (número de onda y amplitud respectivamente). La base de datos fue construida únicamente utilizando la columna de amplitud de cada espectro ya que el vector de número de onda se conserva en cada registro. Este nuevo vector con los datos de amplitud de la medición pasa a ser nuestro vector de características, y con cada registro se construye un dataframe de 52 filas y 2721 columnas.

## 2.4. Modelo de aprendizaje automático

Nuestro modelo de ML fue programado en Python y consiste en una red neuronal de 2 capas (ver Fig. 2) donde la capa oculta está compuesta de 5 neuronas tipo Leaky-ReLU. Finalmente, la capa de salida consta de 3 neuronas tipo Softmax para clasificar si el espectro corresponde a edifenfos, paratión o una molécula diferente. Los pesos sinápticos fueron optimizados a través del método Adam [21], en la dirección que minimiza la función de costo, específicamente la entropía cruzada. En esencia, el método de Adam es una modificación del método del descenso del gradiente

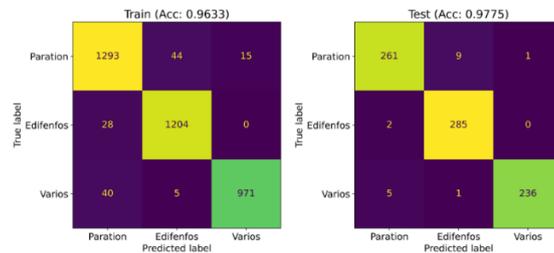


Fig. 3. Matrices de confusión para los datos de entrenamiento y prueba.

estocástico, que se basa en la estimación del primer y segundo momento de la derivada direccional.

### 3. Resultados

Durante el entrenamiento se usó el 70% de los datos. 15% de los datos se reservaron para el segmento de prueba, y finalmente, el 15% restante de los datos se destinaron para validación. Es importante mencionar que se llevaron a cabo 100 realizaciones del entrenamiento para fines estadísticos. Los resultados obtenidos indican que la red neuronal exhibe una precisión del 96.33%, 97.75% y 97.12% en las etapas de entrenamiento, prueba y validación, respectivamente. Las matrices de confusión promedio para el conjunto de datos de entrenamiento y prueba se muestran en la Fig. 3. Con el fin de evitar sobre entrenamiento del modelo propuesto, incorporamos regularización L2 con un factor de penalización del orden de  $10^{-3}$ .

Con el fin de evaluar el rendimiento de nuestra red neuronal, comparamos nuestro modelo con dos algoritmos diferentes de aprendizaje automático: Random Forest (RF) y Maquinas de Soporte Vectorial (SVM por sus siglas en ingles). Estos algoritmos fueron sometidos a 100 realizaciones como en el caso de la red neuronal. El algoritmo RF se compone de un conjunto de árboles predictores, donde las predicciones de cada árbol están influenciadas por los valores de un vector aleatorio. Este vector aleatorio se muestrea de forma independiente y sigue la misma distribución para todos los árboles del bosque. El error de generalización para los bosques converge hacia un límite a medida que aumenta el número de árboles en el bosque. Por su parte, el algoritmo SVM es un método de aprendizaje automático que tiene como objetivo crear hiperplanos en un espacio N-dimensional para distinguir entre conjuntos de datos que pertenecen a clases distintas. Los hiperplanos sirven como límites de decisión que establecen una distancia de margen, lo que permite la clasificación. La Tabla 1 muestra la precisión obtenida en NN, RF y SVM.

### 4. Conclusiones

Los algoritmos de aprendizaje automático permiten analizar los espectros SERS de manera rápida, automatizada, y con gran precisión, abriendo nuevas posibilidades para

**Tabla 1.** Precisiones obtenidas de la red neuronal (NN), random forest (RF) y máquina de soporte vectorial (SVM).

Algoritmo	Precision (%)
NN	97.75%
RF	94.30%
SVM	94.15%

la detección de pesticidas organofosforados en diversas matrices ambientales y de alimentos. El estudio realizado utilizando espectros SERS de paratión y edifentos demostró la efectividad de esta técnica en la identificación de pesticidas. Nuestra red neuronal presenta una precisión de 97.75%, mostrando su capacidad para clasificar con precisión los espectros correspondientes a diferentes pesticidas y compuestos desconocidos.

Para establecer un punto de comparación se entrenó una máquina de soporte vectorial y un algoritmo de random forest. Estos modelos exhiben precisiones cercanas al 94% en ambos casos.

**Agradecimientos.** MAQJ agradece el apoyo de la Dirección General de Asuntos del Personal Académico (DGAPA)-Universidad Nacional Autónoma de México a través del proyecto UNAM-PAPIIT TA101023 y al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT) de México bajo el proyecto CF-2023-I-1496.

## Referencias

1. Phopin, K., Tantimongcolwat, T.: Pesticide Aptasensors—State of the Art and Perspectives. *Sensors*, vol. 20, no. 23, pp. 6809 (2020). DOI: 10.3390/s20236809.
2. Yu, J.J.: Removal of Organophosphate Pesticides from Wastewater by Supercritical Carbon dioxide extraction. *Water Research*, vol. 36, no. 4, pp. 1095–1101 (2002). DOI: 10.1016/S0043-1354(01)00293-7.
3. Rani, L., Thapa, K., Kanojia, N., Sharma, N., Singh, S., Grewal, A.S., Srivastav, A.L., Kaushal, J.: An Extensive Review on the Consequences of Chemical Pesticides on Human Health and Environment. *Journal of Cleaner Production*, vol. 283, pp. 124657 (2021). DOI: 10.1016/j.jclepro.2020.124657.
4. Sidhu, G.K., Singh, S., Kumar, V., Dhanjal, D.S., Datta, S., Singh, J.: Toxicity, Monitoring and Biodegradation of Organophosphate Pesticides: A Review. *Critical Reviews in Environmental Science and Technology*, vol. 49, no. 13, pp. 1135–1187 (2019). DOI: 10.1080/10643389.2019.1565554.
5. Kim, Y.J., Cho, Y.A., Lee, H.S., Lee, Y.T., Gee, S.J., Hammock, B.D.: Synthesis of Haptens for Immunoassay of Organophosphorus Pesticides and Effect of Heterology in Hapten Spacer Arm Length on Immunoassay Sensitivity. *Analytica Chimica Acta*, vol. 475, no. 1-2, pp. 85–96 (2003). DOI: 10.1016/S0003-2670(02)01037-1.
6. Liang, Y., Liu, X.J., Liu, Y., Yu, X.Y., Fan, M.T.: Synthesis of Three Haptens for the Class-Specific Immunoassay of O, O-dimethyl Organophosphorus Pesticides and Effect of Hapten Heterology on Immunoassay Sensitivity. *Analytica Chimica Acta*, vol. 615, no. 2, pp. 174–183 (2008). DOI: 10.1016/j.aca.2008.03.050.

7. Amine, A., Mohammadi, H., Bourais, I., Palleschi, G.: Enzyme Inhibition-based Biosensors for Food Safety and Environmental Monitoring. *Biosensors and Bioelectronics*, vol. 21, no. 8, pp. 1405–1423 (2006). DOI: 10.1016/j.bios.2005.07.012.
8. Rainina, E.I., Efremenco, E.N., Varfolomeyev, S.D., Simonian, A.L., Wild, J.R.: The Development of a New Biosensor based on Recombinant E. Coli for the Direct Detection of Organophosphorus Neurotoxins. *Biosensors and Bioelectronics*, vol. 11, no. 10, pp. 991–1000 (1996). DOI: 10.1016/0956-5663(96)87658-5.
9. Viveros, L., Paliwal, S., McCrae, D., Wild, J., Simonian, A.: A Fluorescence-based Biosensor for the Detection of Organophosphate Pesticides and Chemical Warfare Agents. *Sensors and Actuators B. Chemical*, vol. 115, no. 1, pp. 150–157 (2006). DOI: 10.1016/j.snb.2005.08.032.
10. Simonian, A.L., Good, T.A., Wang, S.S., Wild, J.R.: Nanoparticle-based Optical Biosensors for the Direct Detection of Organophosphate Chemical Warfare Agents and Pesticides. *Analytica Chimica Acta*, vol. 534, no. 1, pp. 69–77 (2005). DOI: 10.1016/j.aca.2004.06.056.
11. Janotta, M., Karlowatz, M., Vogt, F., Mizaikoff, B.: Sol–gel based Mid-Infrared Evanescent Wave Sensors for Detection of Organophosphate Pesticides in Aqueous Solution. *Analytica Chimica Acta*, vol. 496, no. 1-2, pp. 339–348 (2003). DOI: 10.1016/S0003-2670(03)01011-0.
12. Tanner, P.A., Leung, K.H.: Spectral Interpretation and Qualitative Analysis of Organophosphorus Pesticides using FT-Raman and FT-infrared Spectroscopy. *Applied Spectroscopy*, vol. 50, no. 5, pp. 565–571 (1996)
13. Weng, S., Li, M., Chen, C., Gao, X., Zheng, S., Zeng, X.: Fast and Accurate Determination of Organophosphate Pesticides using Surface-Enhanced Raman Scattering and Chemometrics. *Analytical Methods*, vol. 7, no. 6, pp. 2563–2567 (2015). DOI: 10.1039/C4AY03067B.
14. Li, B., Zhang, W., Chen, L., and Lin, B.: A Fast and Low-Cost Spray Method for Prototyping and Depositing Surface-Enhanced Raman Scattering Arrays on Microfluidic Paper based Device. *Electrophoresis*, vol. 34, no. 15, pp. 2162–2168 (2013). DOI: 10.1002/elps.201300138.
15. Liron, Z., Zifman, A., Heleg-Shabtai, V.: Surface-Enhanced Raman Scattering Detection of Cholinesterase Inhibitors. *Analytica Chimica Acta*, vol. 703, no. 2, pp. 234–238 (2011). DOI: 10.1016/j.aca.2011.07.033.
16. Mjolsness, E., DeCoste, D.: Machine Learning for Science: State of the Art and Future Prospects. *Science*, vol. 293, no. 5537, pp. 2051–2055 (2001). DOI: 10.1126/science.293.5537.2051.
17. Device: *Electrophoresis*, vol. 34, no. 15, pp. 2162–2168 (2020)
18. Sarker, I.H.: Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, vol. 2, no. 3, pp. 160 (2021). DOI: 10.1007/s42979-021-00592-x.
19. Panikar, S.S., Ramírez-García, G., Sidhik, S., Lopez-Luke, T., Rodriguez-Gonzalez, C., Ciapara, I.H., De la Rosa, E.: Ultrasensitive SERS Substrate for Label-Free Therapeutic-Drug Monitoring of Paclitaxel and Cyclophosphamide in Blood Serum. *Analytical Chemistry*, vol. 91, no. 3, pp. 2100–2111 (2018). DOI: 10.1021/acs.analchem.8b04523.
20. Ramírez-García, G., Honorato-Colin, M.A., De la Rosa, E., López-Luke, T., Panikar, S.S., de Jesús Ibarra-Sánchez, J., Piazza, V.: Theranostic Nanocomplex of Gold-Decorated up Conversion Nanoparticles for Optical Imaging and Temperature-Controlled Photothermal Therapy. *Journal of Photochemistry and Photobiology A: Chemistry*, vol. 384, pp. 112053 (2019)

*Mateo Frausto-Avila, Monserrat Ochoa-Elias, et al.*

21. Kingma, D.P., Lei-Ba, J.: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980 (2014)