

# VALIDACIÓN DE LA EFICIENCIA DE LOS ALGORITMOS DE CLASIFICACIÓN KNN Y CNN

Karla J. Guevara-Panamá, Isaura M. Jingo-Cevallos, Carla D. Rivera-Vaca

12 de diciembre 2018

## 1. Introducción

Dentro del aprendizaje automático, se encuentran técnicas de clasificación que permiten agrupar muestras de acuerdo a criterios o métodos.

En el siguiente informe se analiza los algoritmos de clasificación KNN y CNN comprobando su eficiencia de forma individual y seleccionando el de mayor validez.

KNN utiliza una base de datos en la que los datos son separados en varias clases para predecir la clasificación de un nuevo dato de muestra.

El punto de muestra es clasificado por mayoría de votos de sus vecinos, asignándole la clase más común entre

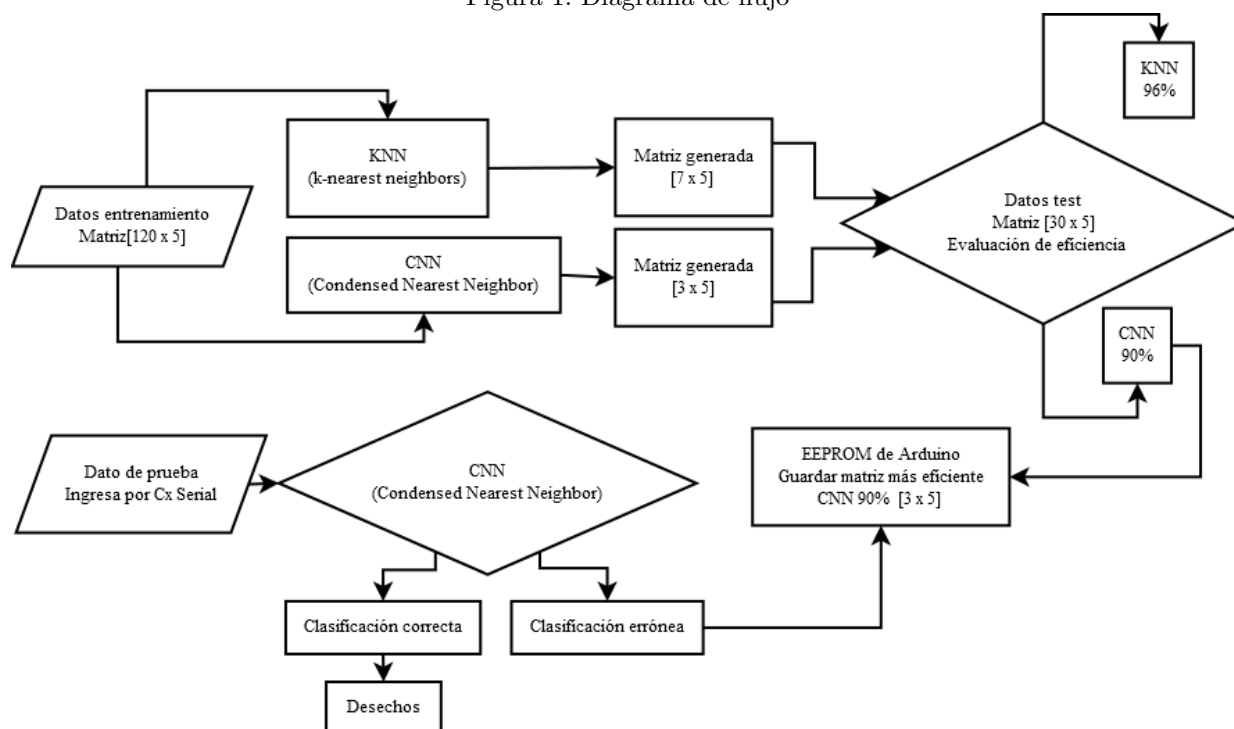
sus  $k$  vecinos más cercanos.[1]

CNN comienza con un conjunto vacío  $S$  al que se añade, de manera aleatoria, un dato correspondiente a cada una de las distintas clases. Posteriormente, cada objeto en el conjunto de entrenamiento es clasificado empleando únicamente los objetos de  $S$ ; cuando un objeto es clasificado erróneamente, este se añade a  $S$  para garantizar que en el futuro será clasificado correctamente, tras el añadido del nuevo dato se comienza nuevamente a recorrer todo el conjunto de entrenamiento. El proceso se repite hasta que no existan datos que sean clasificados erróneamente.[2]

## 2. Diseño del Sistema

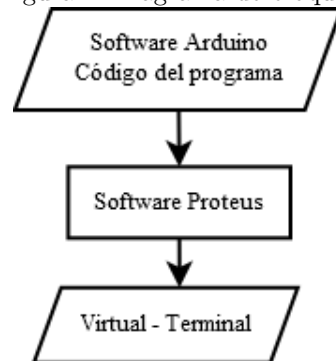
### 2.1. Diagrama de Flujo

Figura 1: Diagrama de flujo



## 2.2. Diagrama de bloques

Figura 2: Diagrama de bloques



## 3. Desarrollo

### 3.1. Simulación

Figura 3: Matriz reducida CNN

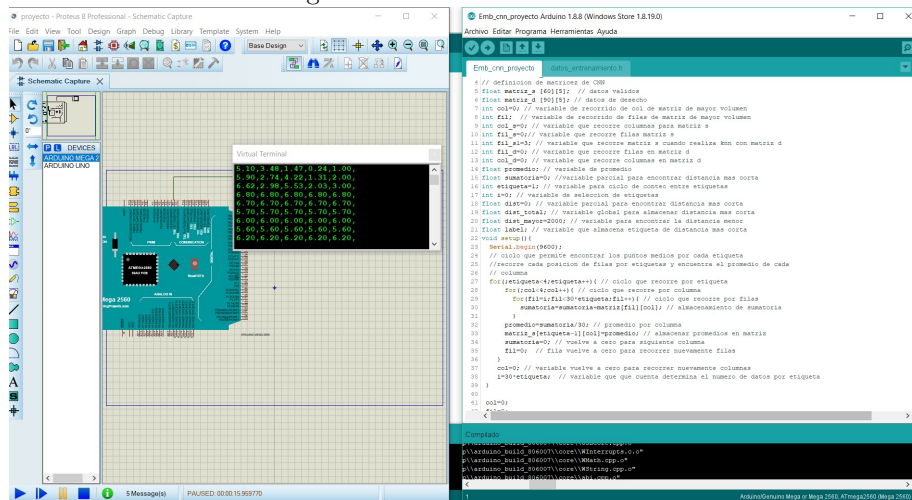


Figura 4: Datos guardados en EEPROM

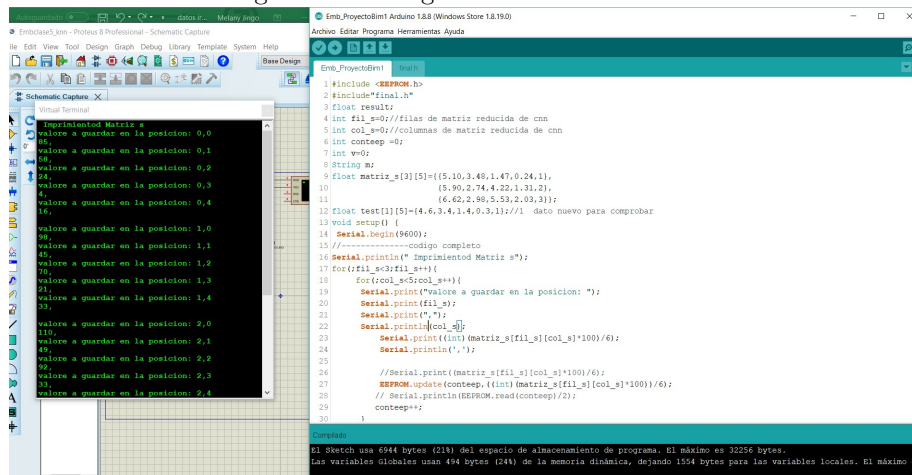


Figura 5: Decisión aciertos y desiertos

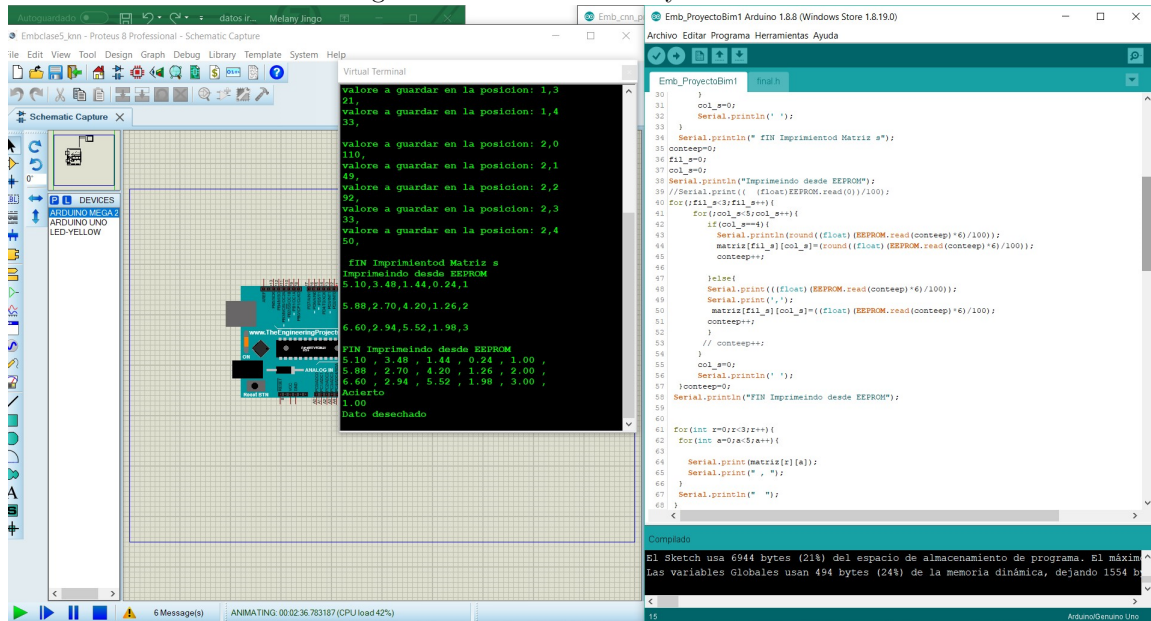
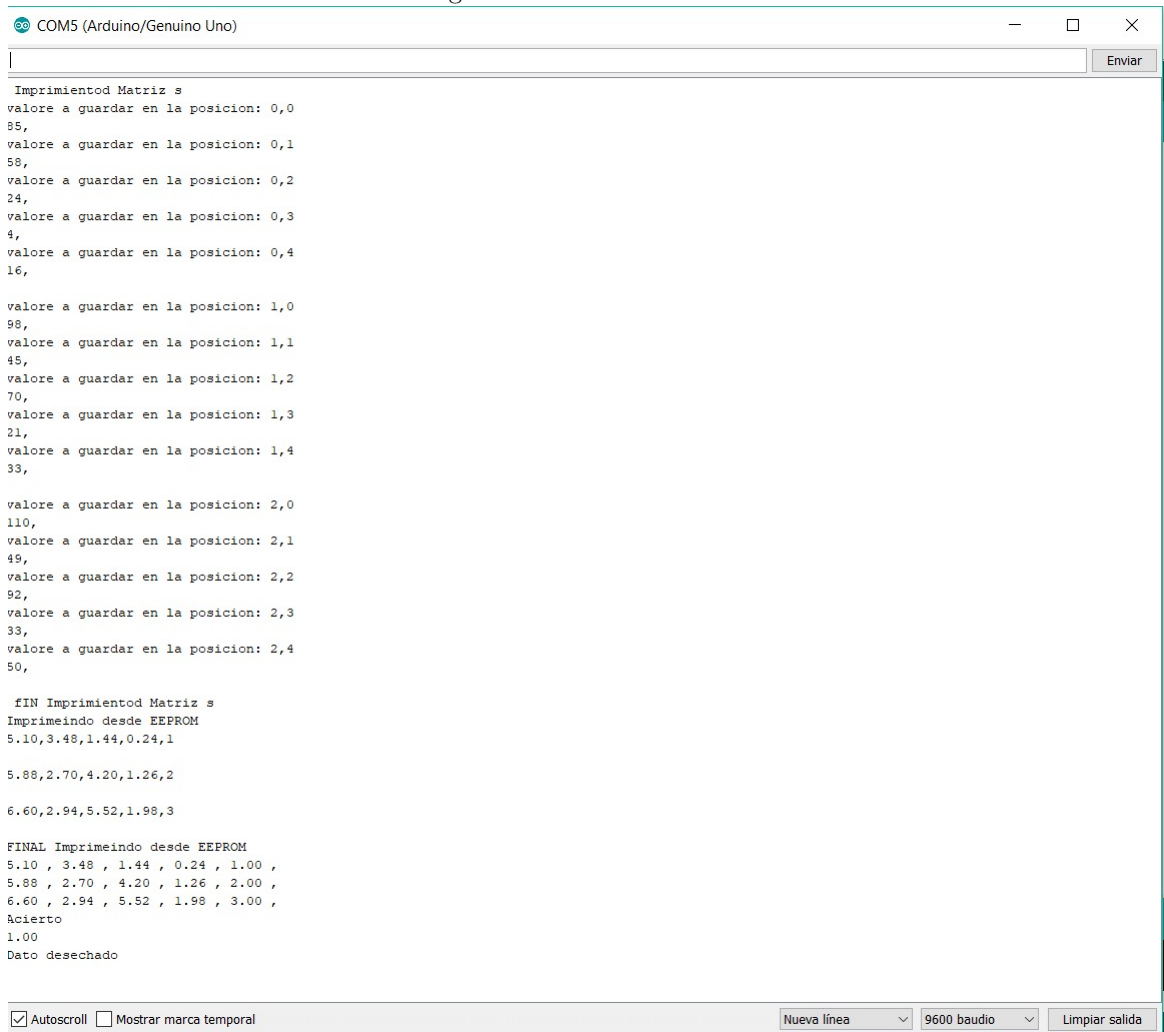


Figura 6: Simulación Física



## 4. Análisis de Resultados

Cuadro 1: Eficiencia Algoritmo de radios y CNN

	Algoritmo de radios	CNN
Matriz original	120 x 5	90 x 5
Matriz reducida	7 x 5	3 x 5
Eficiencia	96 %	90 %

## 5. Conclusiones y Recomendaciones

### 5.1. Conclusiones. -

La principal desventaja de los algoritmos de clasificación analizados. Es la capacidad de memoria que se requiere para almacenar toda la muestra en el microprocesador. Cuando la muestra es demasiado grande, el tiempo de respuesta del computador también es grande. La alternativa a este problema es deshacerse de patrones en el conjunto de entrenamiento que pueden ser muy similares y no agrega información extra o relevante. De esta forma se declaró un subconjunto de la muestra que pequeño y preciso para ambos algoritmos

Posterior al cálculo de la eficiencia de cada algoritmo se decidió que el de mayor validez es CNN, debido a que este mantiene un conjunto de entrenamiento basado en los puntos mal clasificados a partir de la muestra inicial. Se conjetura que estos puntos mal clasificados están próximos a las fronteras de decisión. Consiguiendo un subconjunto de muestras bastante reducido con respecto a la muestra inicial

En la memoria EEPROM no se pueden almacenar datos con decimales pero lo que se procedió a una transformación a entero y luego a escala 6, pero para recuperar los datos de la memoria EEPROM ocurre un error de un 2.04 aproximadamente en ciertos valores de la matriz reducida, lo cual puede afectar a la efectividad , en este caso no afectó y se mantuvo en 90 con CNN.

### 5.2. Recomendaciones. -

Es importante tomar en cuenta un radio equilibrado al momento de depurar los datos, para evitar escasez o exceso de datos.

Verificar de no saturar el microprocesador incluyendo demasiadas líneas de código, esto puede dificultar el rendimiento del programa.

## 6. Referencias

[1] A. Bronshtein, "Medium," 11 Abril 2017. [Online]. Available: <https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7>. [Accessed 11 Diciembre 2018].

[2] J. A. Olvera López , J. A. Carrasco Ochoa and J. F. Martínez Trinidad, "INAOEP," 7 Febrero 2005. [Online]. Available: <http://ccc.inaoep.mx/Reportes/CCC-05-002.pdf>. [Accessed 11 Diciembre 2018].