

Clasificación de demencia usando métodos clásicos de Machine Learning

Georgina Concepción Waldo-Benítez¹,
Luis Carlos Padierna-García¹, Modesto Antonio Sosa-Aquino¹,
Pablo Victor Cerón-Ramírez²

¹ Universidad de Guanajuato,
Departamento de Ingeniería Física,
División de Ciencias e Ingenierías,
México

² Universidad de Quintana Roo,
México

waldobg2012@licifug.ugto.mx

Resumen. La demencia es una enfermedad que afecta la memoria, el pensamiento y las habilidades de comunicación. Más de 50 millones de personas en el mundo padecen esta enfermedad. La detección temprana de demencia permite ofrecer un mejor cuidado a los pacientes; sin embargo, el proceso de detección depende de la experiencia del médico. Las herramientas para diagnóstico asistido por computadora son de utilidad para el médico en este proceso porque le permiten tomar decisiones con mayor precisión. Objetivo clasificar demencia en adultos mayores mediante datos clínicos y métodos clásicos de Machine Learning (ML). Material y métodos: Los algoritmos de ML Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest (RF) y k-Nearest Neighbors, fueron entrenados con datos de pacientes con demencia recopilados en el dataset OASIS-2 para resolver la tarea de clasificación en tres categorías, dementes, no dementes y convertidos. Por medio de un mapa de correlación se seleccionaron las siete características más relevantes, tales como la edad y el volumen normalizado del cerebro. Después, usando el método de análisis de componentes principales (PCA) se identificó que los dos primeros componentes explicaban el 90.22% de los datos. Se realizó una partición 75/25 para entrenamiento y prueba, respectivamente. Con el método grid search se encontraron los mejores hiperparámetros de cada algoritmo. Finalmente, se realizaron 10 corridas independientes de los experimentos para evaluar su confiabilidad. Resultados: Se encontró que el algoritmo k-NN es el que obtiene mayor precisión con un $92.13\% \pm 3.48$, seguido por RF con un $92.0\% \pm 1.8$, a pesar de que obtienen valores muy cercanos, RF no clasifica la categoría de convertido. Conclusiones: El uso de datos clínicos de los pacientes junto con métodos clásicos de ML permite obtener una clasificación precisa de demencia y en consecuencia apoyar el trabajo del médico.

Palabras clave: Clasificación, demencia, machine learning, OASIS-2.

Classification of Dementia Using Methods Machine Learning Classics

Abstract. Dementia is a disease that affects memory, thinking and communication skills. More than 50 million people in the world suffer from this disease. Early detection of dementia allows us to offer better care to patients; however, the screening process depends on the doctor's experience. Computer-aided diagnosis tools are useful to the doctor in this process because they allow him to make decisions with greater precision. Objective to classify dementia in older adults using clinical data and classic Machine Learning (ML) methods. Material and methods: The ML Support Vector Machine algorithms, Naïve Bayes, Decision Tree, Random Forest (RF) and k-Nearest Neighbors, were trained with data from patients with dementia collected in the OASIS-2 dataset to solve the task of classification into three categories, insane, non-insane and converted. Using a correlation map, the seven most relevant characteristics were selected, such as age and normalized brain volume. Then, using the principal component analysis (PCA) method, it was identified that the first two components explained 90.22% of the data. A 75/25 split was performed for training and testing, respectively. With the grid search method, the best hyperparameters of each algorithm were found. Finally, 10 independent runs of the experiments were carried out to evaluate their reliability. Results: It was found that the k-NN algorithm is the one that obtains the highest precision with $92.13\% \pm 3.48$, followed by RF with $92.0\% \pm 1.8$, although they obtain very close values, RF does not classify the converted category. Conclusions: The use of patients' clinical data together with classical ML methods allows obtaining an accurate classification of dementia and consequently supporting the doctor's work.

Keywords: Classification, dementia, machine learning, OASIS-2.

1. Introducción

La demencia es una enfermedad que impacta en la memoria, el pensamiento y las habilidades de comunicación, afectando cerca de 50 millones de personas alrededor del mundo, con más de la mayoría viviendo en países de bajos y medianos ingresos. Cada año cerca de 10 millones de nuevos casos son diagnosticados. Actualmente no hay tratamiento disponible que cure la demencia o que altere su curso progresivo [1]. Sin embargo, la detección temprana de la demencia permite actuar para mejorar los cuidados del paciente e identificar el estado inicial de la enfermedad [2].

El proceso de tamizaje representa un trabajo extenuante para los radiólogos y recae en su experiencia. Recientemente las herramientas basadas en algoritmos de Machine Learning (ML) han probado ser útiles para reducir la carga de trabajo por medio de una correcta clasificación donde solo la información de la imagen MRI del cerebro, es utilizada sin ningún otro tipo de parámetros, reduciendo el tiempo y mejorando la precisión [3, 4, 5, 6, 7, 8].

En [3] los datos clínicos del OASIS-2 son usados para entrenar una máquina de soporte vectorial (SVM) con un kernel R.i.e.BF que clasifica tres tipos de pacientes (dementes, no dementes y convertidos). Utilizan una técnica de preprocesamiento de

datos para completar los datos faltantes y seleccionar un conjunto de características específicas para predecir el resultado, con métricas de rendimiento como la precisión.

En [4], XGBoost fue comparado con otros algoritmos como Random Forest (RF), SVM, Decision Tree (DT), k-Nearest Neighbors (k-NN) para clasificación de los datos del OASIS-2. Una selección de características fue aplicada para mejorar los algoritmos. En resumen, los trabajos previos han resuelto la clasificación de demencia basados en el OASIS-2 con niveles de precisión en el rango entre 68% y 98%.

Los algoritmos clásicos de ML requieren entrenamiento con datos y características seleccionadas por el especialista. Este trabajo propone clasificar la demencia con los métodos existentes de ML buscando el más simple y con la mejor precisión que sea comparable reportado por la literatura.

2. Materiales y métodos

El dataset OASIS-2 provee acceso a un base de datos de neuroimágenes procesadas en un amplio rango demográfico, cognitivo y genético para el uso en investigación de neuroimagen, clínico y cognitivo en el envejecimiento normal y el declive cognitivo [9]. Existen cuatro dataset principales, OASIS-1, OASIS-2, OASIS-3 y OASIS-4. En este trabajo solo nos enfocamos en el OASIS-2, considerando los datos clínicos para el entrenamiento de los métodos clásicos de ML. El OASIS-2 está constituido por una colección de 150 pacientes (72 dementes, 14 convertidos, y 64 no dementes) en un rango de edad de 60 a 96 años. Cada sujeto fue escaneado dos o más veces en visitas separadas por al menos un año.

2.1. Métodos clásicos de Machine Learning

Para encontrar el mejor clasificador de demencia, se entrenaron diversos métodos de Machine Learning que son frecuentemente utilizados en la literatura. A continuación, se da una breve descripción de cada uno de ellos.

Máquinas de soporte vectorial (SVM, por sus siglas en inglés): Es aplicado en la práctica cuando los datos no son linealmente separables, aprende predictores lineales con características espaciales de alta dimensionalidad. Esta alta dimensionalidad aumenta la complejidad de la muestra y de los requerimientos computacionales [10]. Árboles de decisión (DT): pueden ser usados en clasificación multiclase, regresión o clustering. DT son descritos como débiles aprendices con un boost, lo cual los convierte en algoritmos de aprendizaje efectivo.

Son rápidos de entrenar y relativamente fáciles de interpretar. Predice la etiqueta asociada con la instancia x viajando de un nodo a una hoja del árbol [11].

Random Forest (RF): es un clasificador que consiste en una colección de árboles de decisión, donde cada árbol es construido aplicando un algoritmo A en el set de entrenamiento S y un vector adicional aleatorio. La predicción del RF es obtenida por la mayoría de los votos sobre la predicción de los árboles individuales [10].

Naïve Bayes: es un clasificador que utiliza suposiciones generativas y parámetros de estimación simplificando el proceso de aprendizaje. Considera el problema de predecir una etiqueta basado en un vector de características donde se asume que le corresponde una etiqueta para cada valor de x [10].



Fig. 1 Mapa de correlación para el OASIS-2 utilizado para la selección de los mejores datos de entrada.

k-Nearest Neighbors: es un método de aprendizaje supervisado que clasifica los datos en clústeres formados por la respectiva etiqueta de clasificación. Puede ser vista como un esquema de votación entre los k vecinos más cercanos, siendo k impar para evitar empates. El parámetro k controla la estabilidad del estimador k-NN, cuando k es pequeña el algoritmo es sensible a los datos y cuando incrementa el estimador se convierte más estable [12].

Para la visualización de las características se utilizó un método que nos permite trabajar con aquellas que son más relevantes y así poder describir gráficamente el trabajo de nuestro clasificador.

Análisis de componentes principales (PCA, por sus siglas en inglés): reduce el número de características transformando los datos en características que presenten alta varianza, independencia y ortogonalidad uno con otro. Asume que los datos tienen señales que se mezclan de diferentes maneras. PCA intenta ‘desmezclar’ estas señales para recuperar las originales.

PCA transforma un conjunto de observaciones de variables correlacionadas en un set de variables no correlacionadas llamadas componentes principales. La idea base del PCA es transformar la dimensión de las señales en número más pequeños de componentes principales [13].

2.2. Estrategias de evaluación y métricas de rendimiento

Para encontrar los mejores parámetros de los algoritmos es necesario emplear técnicas como grid search o bootstrapping, que nos permiten aumentar el rendimiento y precisión del método de ML empleado.

Grid search es el proceso de ajustar hiperparámetros para determinar los valores óptimos dados para un modelo seleccionado. Usa un bucle a través de hiperparámetros predefinidos y ajusta el modelo en el set de entrenamiento, seleccionando los parámetros de una lista dada [14].

La validación cruzada de k-pliegues involucra dividir aleatoriamente un set de observaciones en k grupos o pliegues, de un tamaño aproximadamente igual. El primer pliegue es tratado como un conjunto de validación, y el método es ajustado en los k-1 pliegues restantes [15].

Bootstrapping es cualquier prueba o métrica que usa muestreo aleatorio con remplazo que cae bajo un tipo más amplio de métodos de remuestreo. Asigna mediciones de precisión a los estimadores de muestra. Esta técnica permite la estimación de una distribución de muestras de casi cualquier estadístico usando un método de muestreo aleatorio [16].

Precisión es la proporción de predicciones correctas (ambas verdaderas positivas y verdaderas negativas) dentro del número total de casos examinados, siguiendo la ecuación (1) [17]:

$$\text{Precisión} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

donde TP es verdadero positivo, FP es falso positivo, TN es verdadero negativo y FN es falso negativo.

2.3. Preprocesamiento de los datos clínicos de OASIS-2

Selección de características. Un mapa de correlación (Fig. 1) fue utilizado para seleccionar las mejores características de entrada para el modelo de ML, siendo estas siete: sexo, edad, educación, minimental state examination (MMSE), clinical dementia rating (CDR) normalized whole brain volume (nWBV) y Atlas Scaling Factor (ASF). Estas características son usadas en trabajos previamente publicados [4], con la diferencia de ASF fue usado en lugar del Estimated Total Intracranial Volume (eTIV). Retraso, estatus socioeconómico (SES) y visita fueron eliminados por no ser características relevantes para el modelo. eTIV fue omitido debido a que está fuertemente correlacionado con ASF.

Visualización. De las siete características restantes, PCA fue utilizado para reducir el número de dimensiones del OASIS-2 dataset y poder visualizar la frontera de decisión. Los dos primeros componentes principales explican el 90.22% de los datos. La Fig. 2 muestra la frontera de decisión obtenida por el clasificador k-NN, donde los grupos 0, 1 y 2 para convertido, no demencia y demencia, respectivamente.

2.4. Diseño experimental

Después de seleccionar las características más relevantes del OASIS-2, los registros clínicos resultantes fueron separados en 75% para el entrenamiento y 25% para la prueba. Métodos de ML clásicos (SVM, RF, DT, NB, y k-NN) y sus configuraciones experimentales reportados en previos trabajos son presentados en la Tabla 1.

El entrenamiento de los modelos se realizó usando una validación cruzada de 10-pliegues y sus hiperparámetros fueron ajustados usando una estrategia grid search. SVM con un gamma de 0.4 y con el kernel radial basis function (RBF); RF con una profundidad máxima de 10; DT con una profundidad máxima de 2, y un separador aleatorio; un NB gaussiano; y k-NN, fue entrenado calculando la distancia Manhattan

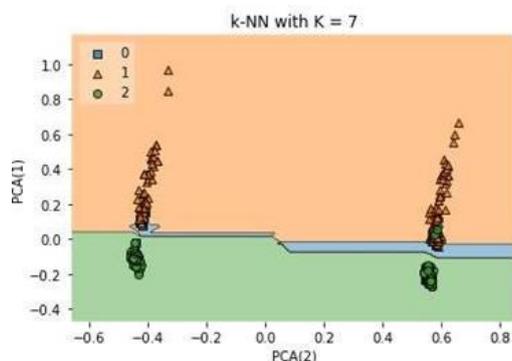


Fig. 2. Clasificación de demencia con k-NN usando los dos componentes principales, que muestran la barrera de decisión y la no linealidad de los datos.

Tabla 1. Precisión de los diferentes algoritmos. Promedio \pm desviación estándar, a partir de la validación cruzada de 10-pliegues.

Algoritmos Clásicos de ML	Precisión (promedio \pm desviación estándar) Nuestro experimento	Mejores reportados [8, 6, 4, 18]
k-NN	92.13 % \pm 3.48	90.74%
RF	92.0% \pm 1.8	96.66%(i), 93.56%(ii)
DT	90.0% \pm 4.3	99.28% (iii)
SVM	89.9% \pm 3.2	92.57%
NB	85.3% \pm 3.5	87.29%

con cada punto y asignándolos al clúster con la mínima distancia desde los centros del clúster, el número de vecinos cercanos fue de siete de acuerdo con grid search.

A pesar de que los datos no son linealmente separables en OASIS-2, como se observa en la Fig. 2, los métodos clásicos de ML pueden ser implementados para clasificar la demencia con precisión.

3. Resultados y discusión

Los algoritmos clásicos de ML fueron entrenados y probados con los datos clínicos del OASIS-2. Los mejores resultados obtenidos en nuestros experimentos, y una comparación con trabajos previos son reportados en la Tabla 1. La precisión promedio más alta fue de 92.13% \pm 3.48 obtenida con el algoritmo k-NN.

(i) La precisión del algoritmo RF reportada por Shanmuga et al. [4] del 96.66%, es el valor máximo obtenido en sus experimentos. Ningún intervalo de confianza es dado, por lo que es un resultado que indica sesgo en el experimento.

(ii) Antes de realizar la clasificación con el algoritmo RF; se realizó una selección de características con un algoritmo de optimización (Particle Swarm Optimization) [18].

(iii) La precisión reportada para DT por Bansal et al. [8] del 99.28% es dudosa ya que indica un sobreajuste del modelo. Esto es común en los algoritmos basados en árboles de decisión. Después de una revisión, se encontró que los autores no explican

el número de instancias usadas en el cálculo de la precisión, y la falta de un esquema de validación como k-pliegues o bootstrapping.

En [3] los autores obtuvieron una precisión de a 68.75% usando un SVM con un kernel RBF en los datos clínicos de OASIS-2. Este bajo rendimiento puede deberse al esquema de ajuste de hiperparámetros usados, en los que sólo valores altos y bajos de C y gamma fueron considerados. En nuestros experimentos, un esquema de grid search obtuvo una precisión de $89.9\% \pm 3.2$, lo cual concuerda con los resultados encontrados en [4] de 92.57%.

4. Conclusiones

Se clasificó la demencia en adultos mayores en tres grupos, dementes, no dementes y convertidos, mediante información clínica como la edad y características del cerebro obtenidas por los médicos mediante imágenes MRI, identificando como mejor algoritmo de clasificación, después de probar con distintos algoritmos de ML, un clasificador k-NN, el cual alcanza una precisión del $92.13\% \pm 3.48$. A comparación del RF, este clasificador pudo distinguir las tres categorías correctamente.

El uso de algoritmos más potentes como XGBoost podría ser otra alternativa para continuar utilizando estos métodos clásicos, y mejorar la precisión. Otro trabajo a futuro consiste en realizar la clasificación de demencia con métodos de aprendizaje profundo utilizando solamente imágenes de MRI y no datos clínicos. Debido al nivel de precisión alcanzado por k-NN, este puede ser considerado para su validación en la práctica clínica.

Referencias

1. World Health Organization: Demencia (2023) www.who.int/news-room/fact-sheets/detail/dementia
2. Custodio, N., Duque, L., Montesinos, R., Alva-Díaz, C., Mellado, M., Slachevsky, A.: Systematic review of the diagnostic validity of brief cognitive screenings for early dementia detection in spanish-speaking adults in Latin America. *Frontiers in Aging Neuroscience*, vol. 12, pp. 270 (2020) doi: 10.3389/fnagi.2020.00270
3. Battineni, G., Chintalapudi, N., Amenta, F.: Machine learning in medicine: performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, vol. 16 (2019) doi: 10.1016/j.imu.2019.100200
4. Vinayak, S. S., Shahina, E. A., Nayeemulla-Khan, A.: Dementia prediction on OASIS dataset using supervised and ensemble learning techniques. *International Journal of Engineering and Advanced Technology*, vol. 10, no. 1, pp. 244–254 (2020)
5. Dua, M., Makhija, D., Manasa, P. Y. L., Mishra, P. A.: CNN–RNN–LSTM based amalgamation for Alzheimer’s disease detection. *Journal of Medical and Biological Engineering*, vol. 40, pp. 688–706 (2020) doi: 10.1007/s40846-020-00556-1
6. Naidu, C., Kumar, D., Maheswari, N., Sivagami, M., Li, G.: Prediction of Alzheimer’s disease using Oasis dataset. *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 7, no. 6S3, pp. 36–42 (2019)
7. Yagis, E., Citi, L., Diciotti, S., Marzi, C., Atnafu, S. W., De-Herrera, A. G. S.: 3D convolutional neural networks for diagnosis of dementia. In: 33rd International Symposium on Computer-Based Medical Systems (CBMS), IEEE, pp. 65–70 (2020) doi: 10.1109/CBMS49503.2020.00020

8. Bansal, D., Chhikara, R., Khanna, K., Gupta, P.: Comparative analysis of various machine learning algorithms for detecting dementia. *Procedia Computer Science*, vol. 132, pp. 1497–1502 (2018) doi: 10.1016/j.procs.2018.05.102
9. OASIS-BRAINS: Open access series of imaging studies (2023) <https://www.oasis-brains.org/>
10. Shalev-Shwartz, S., Ben-David, S.: *Understanding machine learning: from theory to algorithms*. Cambridge: Cambridge University Press (2014)
11. Mohri, M., Rostamizadeh, A., Talwalkar, A.: *Foundations of machine learning*. MIT Press (2018)
12. Rosasco, L.: *Introductory machine learning notes*. Genova: MIT (2017)
13. Winn, J., Bishop, C. M., Diethe, T., Guiver, J., Zaykov, Y.: *Model-based machine learning*. Taylor & Francis (2014)
14. Mujtaba, H.: *Great Learning* <https://www.mygreatlearning.com/blog/gridsearchcv/>
15. James, G., Witten, D., Hastie, T., Tibshirani, R.: *An introduction to statistical learning*. Springer (2017) doi: 10.1007/978-1-0716-1418-1
16. Varian, H. R.: Bootstrap tutorial. *Mathematica Journal*, vol. 9, no. 4, pp. 768–775 (2005)
17. Metz, C.: Basic principles of ROC analysis. *Seminar of Nuclear Medicine*, vol. 8, no. 4, pp. 283–298 (1978) doi: 10.1016/S0001-2998(78)80014-2
18. Saputra, R. A., Agustina, C., Puspitasari, D., Ramanda, R., Pribadi, D., Indriani, K.: Detecting Alzheimer’s disease by the decision tree methods based on particle swarm optimization. *Journal of Physics: Conference Series*, IOP Publishing, vol. 1641, no. 1, p. 012025 (2020) doi: 10.1088/1742-6596/1641/1/012025