Anomaly Detection in biomedical data by means of Autoencoders

Matías Cisnero[1] and Juliana Gambini[1,2]

[1] LIDEC, Universidad Nacional de Hurlingham, Pcia. de Buenos Aires, Argentina
matiasnicolas.cisnero@estudiantes.unahur.edu.ar,
juliana.gambini@unahur.edu.ar
[2] CPSI, Universidad Tecnológica Nacional, Regional Buenos Aires

Abstract. This line of research addresses the detection of anomalies in biomedical data using Autoencoders, which are artificial neural networks with a symmetrical architecture in which the input layer is the same as output layer. This method reduces the dimension of original data and then reconstructs it. The reconstruction error is the key point of this research.
The hypothesis of this work is that patients who present a large reconstruction error (greater than a threshold) can be classified as anomalous, which can help the medical professional to detect diseases. With this objective, the Autoencoder method is implemented and trained with different proportions of data from healthy and sick patients. Several architectures are used to optimize the parameters.
The threshold used to classify these data as normal or anomalies based on their reconstruction error is computed using the Youden index.
In this case, the methods are tested using the Breast Cancer Wisconsin (Diagnostic) dataset, which has samples from patients with benign and malignant patterns obtained by fine-needle aspiration (FNA).
The results are encouraging, but further research is needed to obtain more conclusive results.

Keywords: anomaly detection, biomedical data, autoencoders

Detección de Anomalías en datos biomédicos utilizando Autoencoders

Abstract. La presente línea de investigación aborda la detección de anomalías en datos biomédicos utilizando Autoencoders, los cuales son redes neuronales artificiales que poseen una arquitectura simétrica, en la que la capa de entrada es exactamente igual a la de salida. Este método reduce la dimensión de los datos originales para luego reconstruirlos. El error de reconstrucción es la pieza clave de esta investigación.

La hipótesis de este trabajo es que los pacientes que presenten un error de reconstrucción grande (mayor que un umbral), pueden ser clasificados como anómalos, lo cual puede ayudar al profesional médico a detectar enfermedades. Con este objetivo, se implementa el método Autoencoder y se entrena con distintas proporciones de datos de pacientes sanos y enfermos. Se utilizan diferentes arquitecturas para optimizar los parámetros. El umbral utilizado para clasificar estos datos como normales o anomalías según su error de reconstrucción, se calcula utilizando el índice de Youden.
En este caso, los métodos se prueban utilizando el conjunto de datos Breast Cancer Wisconsin (Diagnostic), que posee muestras de pacientes con patrones benignos y malignos obtenidas mediante punción con aguja fina (FNA).
Los resultados son alentadores pero es necesario profundizar la investigación para obtener resultados más contundentes.

Keywords: detección de anomalías, datos biomédicos, autoencoders

1 Introducción

El diagnóstico temprano de enfermedades como el cáncer de mama es una tarea prioritaria en la medicina moderna. El análisis de datos permite detectar patrones, realizar clasificaciones o regresiones que pueden colaborar con el personal de salud en el diagnóstico preciso de enfermedades.

Los Autoencoders son redes neuronales diseñados para aprender representaciones con las características más relevantes de los datos, permitiendo distinguir entre registros anómalos y normales.

Este trabajo se basa en la hipótesis de que si se entrena un Autoencoder con datos normales, los registros anómalos (malignos) tendrán un mayor error de reconstrucción en comparación con los normales (benignos). Para decidir cuál es mínimo error de reconstrucción para el cual se considera que los datos provienen de una anomalía es necesario utilizar un umbral. Este umbral de decisión se calcula con el índice de Youden (Youden, 1950).

Se han desarrollado múltiples variantes de Autoencoders, evaluadas por los autores Neloy y Turgeon (Neloy & Turgeon, 2024). Los cuales comparan la capacidad de reconstrucción, generación de muestras, visualización del espacio latente y precisión en clasificación de datos anómalos en 11 arquitecturas. Algunas de las cuales son Denosing Autoencoder, Sparse Autoencoder, Contractive Autoencoder y Variational Autoencoder junto a algunas de sus variantes (SAE, VQ-VAE, β-VAE).

Otras propuestas como la de Nawaz et al. (Nawaz et al., 2024) incluyen a los *Ensemble of Autoencoders* como una alternativa muy robusta porque combina varios modelos.

Sin embargo, los autores Bouman y Heskes (Bouman & Heskes, 2025), cuestionan la capacidad de los autocodficadores para detectar anomalías mediante el error de reconstrucción. Los autores afirman que las anomalías lejanas a los datos normales pueden ser reconstruidas perfectamente, denotando una falta de fiabilidad en los modelos.

Finalmente, las propuestas de Tekgöz y Topuz (Tekgöz & Topuz, 2025) y de Reshan et al. (Reshan et al., 2023) muestran trabajos recientes utilizando el mismo conjunto de datos que se utiliza en este trabajo. En el primero, se utiliza el Autoencoder para reducir la dimensionalidad de los datos. En el segundo, se utilizan técnicas de *ensemble machine learning* (EML), logrando excelentes resultados, con un accuracy de 0.9989 y un AUC de 1 en el mejor modelo.

En este trabajo analizamos la capacidad de tres Autoencoders diferentes para detectar anomalías en un conjunto de datos de pacientes, comparando sus resultados al entrenarlos con distintas proporciones de anomalías. Algunos registros corresponden a personas enfermas y otros no. Si bien es necesario seguir trabajando, podemos observar que los resultados son alentadores.

2 Materiales y Métodos

Para la realización de este trabajo, se utiliza el conjunto de datos ("Breast Cancer Wisconsin (Diagnostic) Data Set", 2016). El cual está compuesto por 569 registros de muestras citológicas con 357 de éstas etiquetadas como benignas (B) y 212 como malignas (M), los atributos corresponden a características numéricas derivadas del análisis citológico mediante FNA.

El método empleado en este trabajo es el Autoencoder, un tipo de perceptrón multicapa caracterizado por su estructura simétrica respecto a la capa latente y por poseer la misma dimensión tanto en la entrada como en la salida. La arquitectura de un Autoencoder se compone de tres partes principales: el *encoder*, que transforma los datos de entrada en una representación de menor dimensionalidad, la capa del medio, denominada ¨capa latente¨ y el *decoder*, que intenta reconstruir la entrada original a partir de la representación compacta. La Figura 1 muestra la estructura de esta red neuronal.

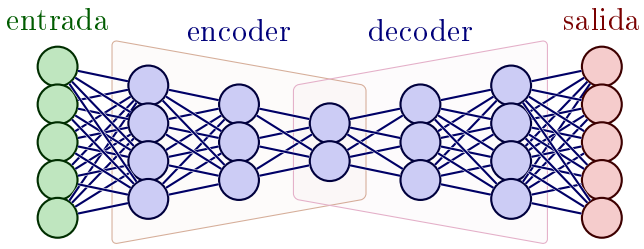

Fig. 1. Arquitectura de un Autoencoder.

También se utilizan dos variantes del Autoencoder, el *Sparse Autoencoder (SAE)* y *Contractive Autoencoder (CAE)*, los cuales imponen restricciones adicionales sobre la capa latente.

Para evaluar el desempeño del autocodificador en la tarea de detección de anomalías, se analizan métricas derivadas del error de reconstrucción $\|\hat{t} - t\|$, donde t es un registro tomado del conjunto de datos y $\hat{t}$ es el estimado. Este valor se utiliza como criterio de decisión para distinguir entre casos normales y anómalos mediante un umbral ϵ, calculado con el índice de Youden. El registro se considera anomalía si el error de reconstrucción supera el umbral.

Para analizar la robustez del Autoencoder frente a distintos niveles de contaminación en los datos, se realiza una división del conjunto en entrenamiento (60%), validación (20%) y prueba (20%).

A partir del subconjunto de entrenamiento, se generan tres nuevos subconjuntos de igual tamaño con distintas proporciones de anomalías: A con 0%, B con 10% y C con 25%.

A partir de esta regla se construye la matriz de confusión, y se calculan las métricas de exactitud, sensibilidad y especificidad, entre otras. Asímismo, se emplean la curva ROC y el área bajo la curva (AUC) para caracterizar el rendimiento del modelo.

3 Resultados

Las configuraciones de cada modelo surgen de un proceso de *grid search*. Obteniendo los hiperparámetros de la Tabla 1.

Table 1. Hiperparámetros de entrenamiento para cada modelo.

Modelo	**Dimensiones de capas**	***lr***	***Batch size***	**Épocas**
AE	[30, 32, 16, 8, 4, 2]	0.001	16	200
SAE	[30, 64]	0.001	16	200
CAE	[30, 32, 16]	0.001	16	200

Los resultados se visualizan en la Tabla 2. Donde se observa que el mayor valor de *Recall* lo posee el modelo **CAE (A)** y el modelo **SAE (A)** cuenta con el mayor valor de $F1_{score}$, *Precision* y *Specificity*. También se visualiza el decrecimiento de las métricas al aumentar las anomalías en el entrenamiento de los modelos, especialmente en el modelo SAE. Por ejemplo, en el modelo **SAE (A)** más del 90% de las muestras resultaron bien clasificadas.

4 Conclusiones y trabajos futuros

Los resultados experimentales muestran una buena capacidad discriminativa de los modelos, destacando por sus métricas generales al SAE y por su *Recall* junto a su robustez ante la presencia de anomalías al CAE.

Table 2. Métricas para cada modelo entrenado con los subconjuntos A, B y C.

Modelo	*Accuracy*	*Precision*	*Recall*	*Specificity*	*Balanced Accuracy*	$F1_{score}$
AE (A)	0.860	0.771	0.881	0.847	0.864	0.822
AE (B)	0.746	0.618	0.810	0.708	0.759	0.701
AE (C)	0.632	0.500	0.619	0.639	0.629	0.553
SAE (A)	0.939	0.973	0.857	0.986	0.922	0.911
SAE (B)	0.754	0.652	0.714	0.778	0.746	0.682
SAE (C)	0.640	0.538	0.167	0.917	0.542	0.255
CAE (A)	0.930	0.886	0.929	0.931	0.930	0.907
CAE (B)	0.798	0.757	0.667	0.875	0.771	0.709
CAE (C)	0.693	0.566	0.714	0.681	0.697	0.632

Como trabajo futuro planeamos realizar una comparación exhaustiva entre el algoritmo utilizado en este trabajo y otras metodologías ya existentes aplicadas a este mismo conjunto de datos.

Referencias

Bouman, R., & Heskes, T. (2025). Autoencoders for anomaly detection are unreliable. *CoRR*, *abs/2501.13864*. https://doi.org/10.48550/arXiv.2501.13864

Breast cancer wisconsin (diagnostic) data set [Last accessed 2025/11/27]. (2016). https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data/data

Nawaz, A., Khan, S., & Ahmad, A. (2024). Ensemble of autoencoders for anomaly detection in biomedical data: A narrative review. *IEEE Access*, *12*, 17273–17289.

Neloy, A., & Turgeon, M. (2024). A comprehensive study of auto-encoders for anomaly detection: Eficiency and trade-offs. *Machine Learning with Applications*, *17*, 100572.

Reshan, M. S. A., Amin, S., Zeb, M. A., Sulaiman, A., Alshahrani, H., Azar, A. T., & Shaikh, A. (2023). Enhancing breast cancer detection and classification using advanced multi-model features and ensemble machine learning techniques. *Life*, *13*(10). https://doi.org/10.3390/life13102093

Tekgöz, S., & Topuz, D. (2025). Autoencoders for clinical data analysis: Application of neural network-based dimensionality reduction on fine-needle aspiration breast data. *Asian Journal of Research in Computer Science*, *18*(12), 165–183. https://doi.org/10.9734/ajrcos/2025/v18i12797

Youden, W. (1950). Index for rating diagnostic tests. *Cancer*, *3*, 32–35. https://doi.org/10.1002/1097-0142