

Identificación automática de pulsares a partir de su caracterización estadística

S. Ruppel¹, A. Venere¹, J. Cogo¹, J. Areta^{1,2}, N. Maffione^{1,2}, M. Orellana^{1,2}, A. Granada^{1,2}, G.M.Gancio³

¹ Universidad Nacional de Río Negro. Sede Andina, Argentina

² Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

³ Instituto Argentino de Radioastronomía, CONICET-CICPBA-UNLP, Argentina

INTRODUCCIÓN

Dada la problemática en la identificación de radio-pulsares, se explora la implementación, a través de técnicas de software, de un sistema autónomo capaz de descartar falsos positivos como señales periódicas en una base de datos pública formada por descriptores estadísticos de observaciones centradas en 1352Mhz.

Se enmarca este trabajo como proyecto final obligatorio para la finalización de la carrera de Ingeniería Electrónica, en curso en la Universidad Nacional de Río Negro. Se propone una metodología basada en Inteligencia Artificial, la cual toma como punto de partida la base de datos High Time Resolution Universe Survey (HTRU2). Esta base de datos contiene parámetros estadísticos de unas 17898 señales de las cuales 13007 no son pulsares y 1639 si lo son (Lyon et al. 2016). Los conceptos de inteligencia artificial que se usaron puede dividirse en dos subgrupos, Machine Learning y Deep learning.

ENFOQUE DE LA PROPUESTA

Muy brevemente podemos decir que Machine learning consiste en alimentar un algoritmo con una gran cantidad de datos que permita que una tarea se lleve a cabo de una mejor manera a partir de cierto aprendizaje. En cambio Deep learning se centra en utilizar como algoritmo una red neuronal de algún tipo, que pueda auto entrenarse para realizar una tarea de forma más eficiente (Campesato 2020). Aquí la propuesta consistió en estudiar la performance de las técnicas de machine y de deep learning y evaluar su desempeño para la identificación de candidatos a pulsares.

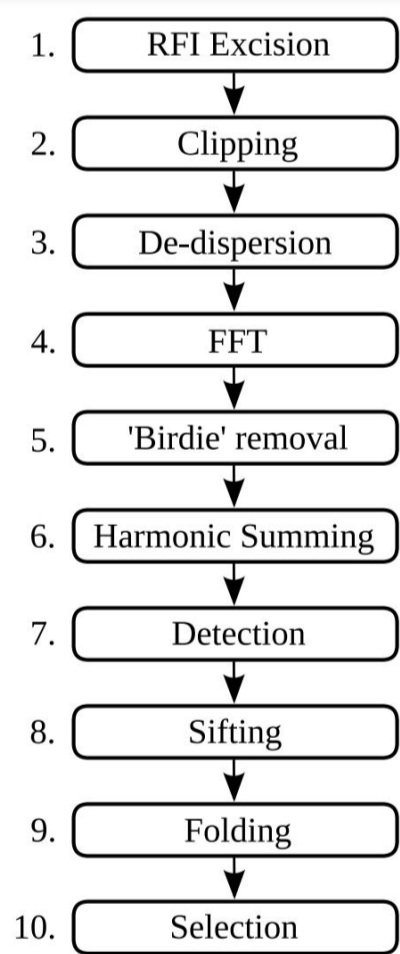
BASE DE DATOS

Procesar una señal para determinar si en la misma existe o no un pulsar requiere de una serie de pasos de post-procesado de la señal obtenida. Debido a la naturaleza de los mismos, las técnicas son CPU o GPU intensivas y dependiendo de la duración de las mediciones pueden ser RAM dependientes también. El proceso es largo e iterativo probando a través de varios valores de DM (Dispersion Measure) hasta concluir con una muestra representativa de la señal que se cree puede contener un pulsar. Luego un astrónomo evalúa si efectivamente lo es.

Dicho proceso ya fue realizado por Lyon et al. (2016) en el caso de la base de datos que analizamos (HTRU2). Los autores detallan los siguientes pasos:

Como resultado obtienen parámetros estadísticos para cada fuente puntual que se consideró candidato a pulsar.

Fig. 1: Para la generación de la base de datos HTRU2 se siguieron estos pasos, el resultado era un pulso candidato del cual se obtuvieron una serie de parámetros estadísticos.



	mean_int_pf	std_pf	ex_kurt_pf	skew_pf	mean_dm	std_dm	kurt_dm	skew_dm
0	140.562500	55.683782	-0.234571	-0.699648	3.199833	19.110426	7.975532	74.242225
1	102.507812	58.882430	0.465318	-0.515088	1.677258	14.860146	10.576487	127.393580
2	103.015625	39.341649	0.323328	1.051164	3.121237	21.744669	7.735822	63.171909
3	136.750000	57.178449	-0.068415	-0.636238	3.642977	20.959280	6.896499	53.593661
4	88.726562	40.672225	0.600866	1.123492	1.178930	11.468720	14.269573	252.567306
...
17893	136.429688	59.847421	-0.187846	-0.738123	1.296823	12.166062	15.450260	285.931022
17894	122.554688	49.485605	0.127978	0.323061	16.409699	44.626893	2.945244	8.297092
17895	119.335938	59.935939	0.159363	-0.743025	21.430602	58.872000	2.499517	4.595173
17896	114.507812	53.902400	0.201161	-0.024789	1.946488	13.381731	10.007967	134.238910
17897	57.062500	85.797340	1.406391	0.089520	188.306020	64.712562	-1.597527	1.429475

METODOLOGÍA

A partir de estrategias de machine y deep learning buscamos derivar dicha tarea a algoritmos que operen de manera autónoma. Entonces, trabajamos con estas técnicas para finalmente realizar una prueba sobre la base de datos HTRU2, observando si los algoritmos detectan claramente las señales de pulsares de las que no lo son. A partir de esa visualización y utilizando las herramientas que ofrece Scikit-learn, se determinó el atributo más relevante que permite diferenciar una señal que contenía un pulsar de otra que no. La visualización de la base de datos se realizó mediante histogramas que se muestran en los paneles de la Figura 2.

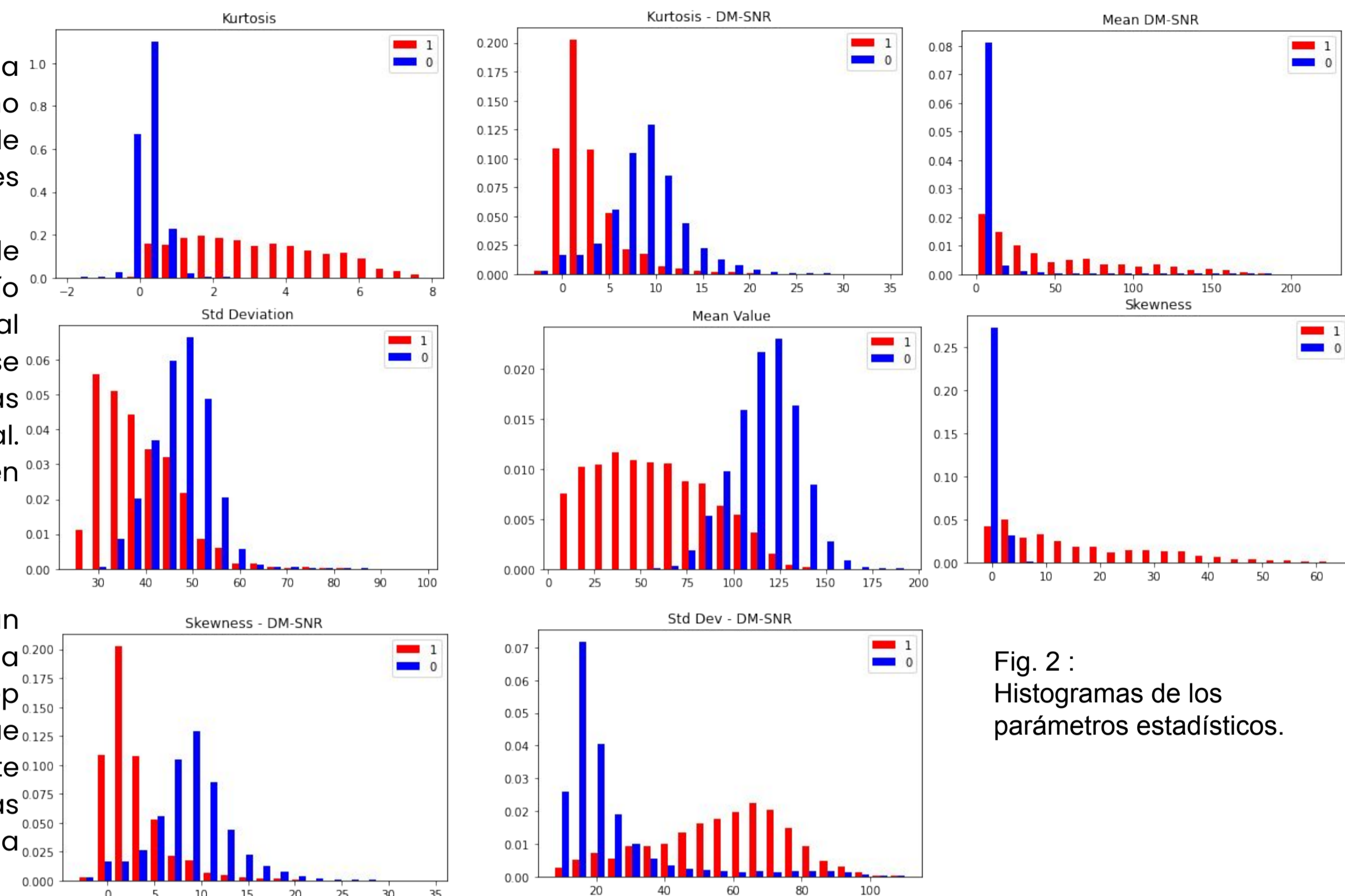


Fig. 2 : Histogramas de los parámetros estadísticos.

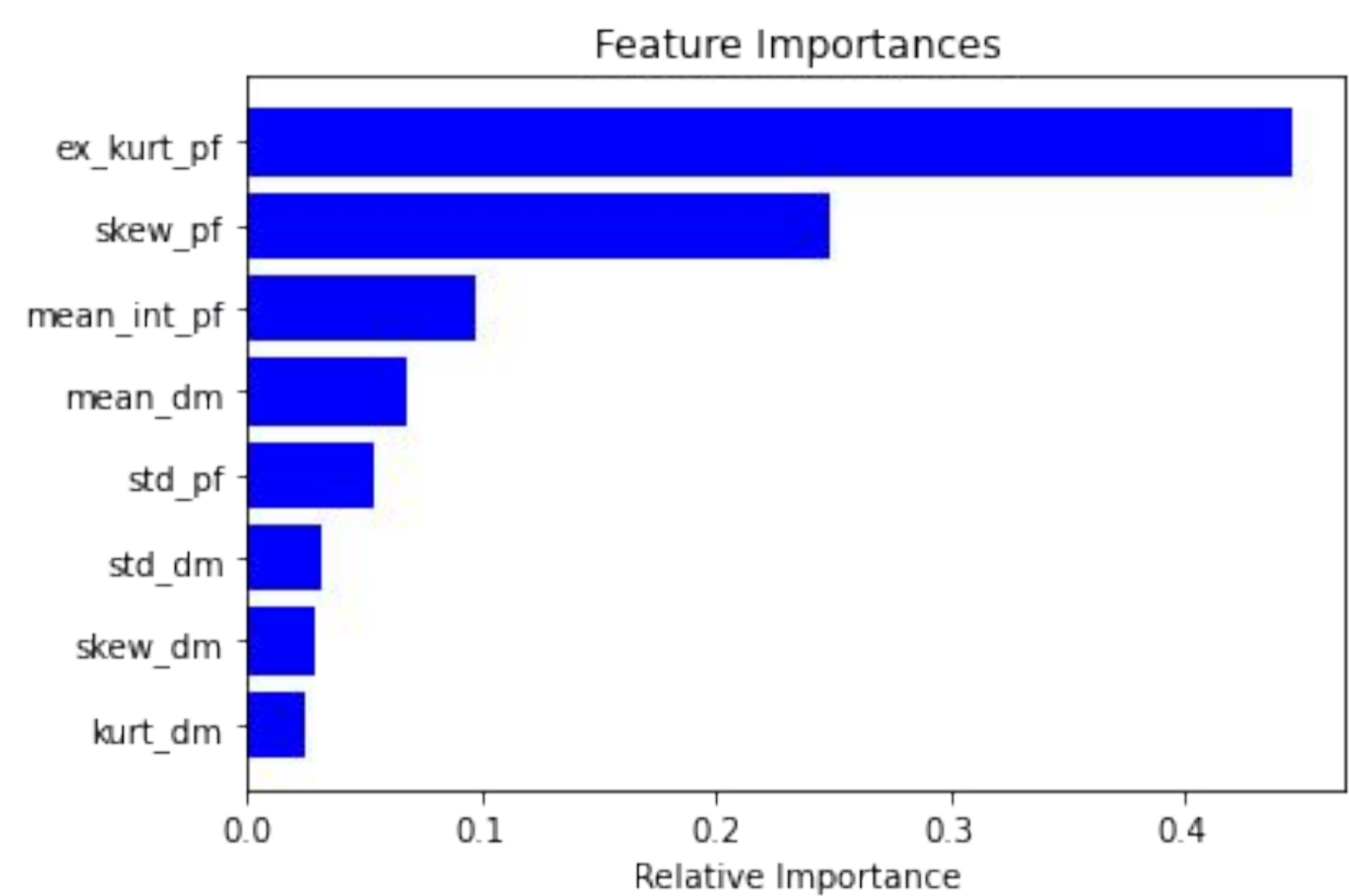


Fig. 3: Importancia relativa de los parámetros

RESULTADOS

Detectado el atributo más relevante para la validación del candidato, que resulta ser la Kurtosis de la señal, se propone entonces dos algoritmos para Machine Learning y una red neuronal para Deep Learning y se evalúa el desempeño teniendo en cuenta que las redes neuronales requieren de un tiempo de entrenamiento previo y luego un tiempo para la clasificación. Se evalúa la tasa de éxito, el tiempo de operación del algoritmo, los falsos positivos y negativos, así como el conjunto de datos obtenidos y se obtuvieron los siguientes resultados resumidos en la siguiente tabla.

	Random Forest	Support Vector Machine	Deep learning (*)
Presicion	97,88%	97.23%	97.9551%
Runtime	0:00:03.725348s	0:05:41.493335	0:02:04.506093 (KerasClassifier) +0:00:15.746782 (Fitting)
Falsos negativos	52	25	43
Positivos en el dataset	330	332	332
Positivos predcidos	302	381	357

A futuro, el trabajo se puede replicar realizando pruebas con otro hardware.

REFERENCIAS

- Campesato, O. (2020). Artificial intelligence, machine learning, and deep learning. Mercury Learning and Information.
 Thornton, D. (2013). The high time resolution radio sky. The University of Manchester (United Kingdom).
 Lyon, R. J. (2016). Why are pulsars hard to find?. The University of Manchester (United Kingdom)
 Lyon, R.J. HTRU2, DOI: 10.6084/m9.figshare.3080389.v1.

CONTACTO:

ruppel33@gmail.com
(UNRN)

Agradecemos financiamiento parcial del proyecto PI 40B885 de la UNRN.