**Título:** Dataset of the paper “A multistart tabu search–based method for feature selection in medical applications". Scientific Reports, 13, 17140.

**Autoría:** Joaquín Pacheco, Olalla Saiz, Silvia Casado y Silvia Ubillos

**Palabas clave:** classification, feature selection, medical diagnosis, tabu search

**Resumen:** In the design of classification models, irrelevant or noisy features are often generated. In some cases, there may even be negative interactions among features. These weaknesses can degrade the performance of the models. Feature selection is a task that searches for a small subset of relevant features from the original set that generate the most efficient models possible. In addition to improving the efficiency of the models, feature selection confers other advantages, such as greater ease in the generation of the necessary data as well as clearer and more interpretable models. In the case of medical applications, feature selection may help to distinguish which characteristics, habits, and factors have the greatest impact on the onset of diseases. However, feature selection is a complex task due to the large number of possible solutions. In the last few years, methods based on different metaheuristic strategies, mainly evolutionary algorithms, have been proposed. The motivation of this work is to develop a method that outperforms previous methods, with the benefits that this implies especially in the medical field. More precisely, the present study proposes a simple method based on tabu search and multistart techniques. The proposed method was analyzed and compared to other methods by testing their performance on several medical databases. Specifically, eight databases belong to the well-known repository of the University of California in Irvine and one of our own design were used. In these computational tests, the proposed method outperformed other recent methods as gauged by various metrics and classifiers. The analyses were accompanied by statistical tests, the results of which showed that the superiority of our method is significant and therefore strengthened these conclusions. In short, the contribution of this work is the development of a method that, on the one hand, is based on different strategies than those used in recent methods, and on the other hand, improves the performance of these methods.

**Fecha de publicación de datos en página web (**[**https://www.ubu.es/metaheuristicos-grinubumet/ejemplos-y-datos-de-problemas**](https://www.ubu.es/metaheuristicos-grinubumet/ejemplos-y-datos-de-problemas)**):** 2023

**Financiación:** This work was partially supported by FEDER funds and the Spanish State Research Agency (Projects PID2019-104263RB-C44 and PDC2021–121021-C22); the Regional Government of “Castilla y León”, Spain (Project BU071G19); the Regional Government of “Castilla y León”; and FEDER funds (Project BU056P20).

**Publicación relacionada:** El artículo relacionado está publicado en “Scientific Reports”

**Licencia de uso de los materiales (licencia Creative Commons):** CC BY-NC

**Información necesaria para la interpretación y utilización de los datos**

La clase se guarda en la última columna: 0 = No Alzheimer; 1 = Si Alzheimer.

Las primeras variables explicativas son datos personales básicos tales como sexo, edad, altura, peso, etc.

Las siguientes indican la presencia o no de determinados síntomas, características, etc.