

# A review of a hybrid feature selection method for DNA microarray data

Laura Rodríguez-Navas<sup>[0000–0003–4929–1219]</sup>

Universidad Internacional Menéndez Pelayo (UIMP), Madrid, Spain  
rodrigueznava@posgrado.uimp.es

**Abstract.** The emergence of DNA Microarray technology has enabled researchers to analyse the expression level of thousands of genes simultaneously. The Microarray data analysis is the process of finding the most informative genes as well as remove redundant and irrelevant genes. One of the most important applications of Microarray data analysis is cancer classification. However, the curse of dimensionality and the curse of sparsity make classifying gene expression profiles a challenging task. One of the most effective methods to overcome these challenges is feature (gene) selection. Feature (gene) selection can be used to successfully extract those genes that directly influence classification accuracy and to eliminate genes that have no influence on it. This significantly improves calculation performance and classification accuracy. In this paper, I aim to review the correlation-based feature selection (CFS) and the Taguchi-genetic algorithm (TGA) merged into a new hybrid method. The classification accuracy obtained by the proposed method is higher when it is compared to other classification methods from the literature.

**Keywords:** Feature selection · Taguchi-genetic algorithm · K-nearest neighbour · Leave-one-out cross-validation.

## 1 First Section

### 1.1 A Subsection Sample

Please note that the first paragraph of a section or subsection is not indented. The first paragraph that follows a table, figure, equation etc. does not need an indent, either.

Subsequent paragraphs, however, are indented.

**Sample Heading (Third Level)** Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

*Sample Heading (Fourth Level)* The contribution should contain no more than four levels of headings. Table 1 gives a summary of all heading levels.

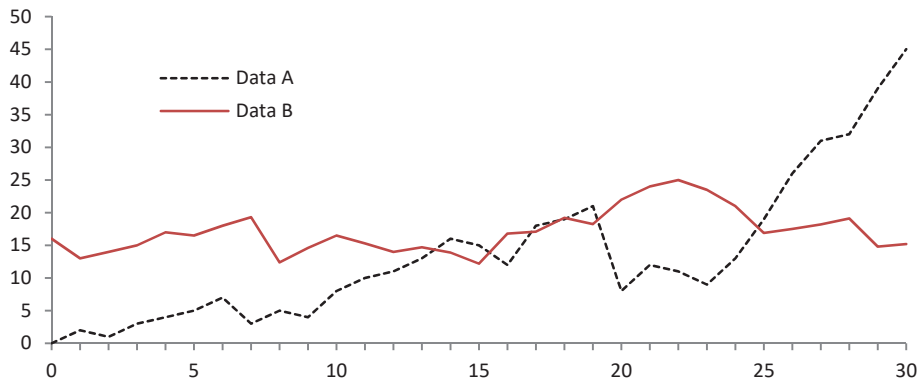
**Table 1.** Table captions should be placed above the tables.

Heading level	Example	Font size and style
Title (centered)	<b>Lecture Notes</b>	14 point, bold
1st-level heading	<b>1 Introduction</b>	12 point, bold
2nd-level heading	<b>2.1 Printing Area</b>	10 point, bold
3rd-level heading	<b>Run-in Heading in Bold.</b> Text follows	10 point, bold
4th-level heading	<i>Lowest Level Heading.</i> Text follows	10 point, italic

Displayed equations are centered and set on a separate line.

$$x + y = z \tag{1}$$

Please try to avoid rasterized images for line-art diagrams and schemas. Whenever possible, use vector graphics instead (see Fig. 1).



**Fig. 1.** A figure caption is always placed below the illustration. Please note that short captions are centered, while long ones are justified by the macro package automatically.

**Theorem 1.** *This is a sample theorem. The run-in heading is set in bold, while the following text appears in italics. Definitions, lemmas, propositions, and corollaries are styled the same way.*

*Proof.* Proofs, examples, and remarks have the initial word in italics, while the following text appears in normal font.

For citations of references, we prefer the use of square brackets and consecutive numbers. Citations using labels or the author/year convention are also acceptable. The following bibliography provides a sample reference list with entries for journal articles [1].

## References

1. Chuang, L.Y., Yang, C.H., Wu, K.C., Yang, C.H.: A hybrid feature selection method for dna microarray data. *Computers in biology and medicine* **41**(4), 228–237 (2011)