7.1. Preprocessing of data: filtering, summarization and analysis of differential gene expression. Feature selection: subset of relevant genes for metastatic sarcoma using FOCI method. Association rule mining: discovering interesting relationships between genes, leading to the identification of potential biomarkers related to metastatic sarcoma and low survival rates.

2. Methodology

Methodology phases:
1. Preprocessing of data: filtering, summarization and analysis of differential gene expression.
2. Feature selection: subset of relevant genes for metastatic sarcoma using FOCI method.
3. Association rule mining: discovering interesting relationships between genes, leading to the identification of potential biomarkers related to metastatic sarcoma and low survival time.

Input data:
- Gene expression dataset GSE21050 – French Sarcoma Group (FSG) database available in Gene Expression Omnibus (GEO).
- 309 patients:
  - Time Survivor
  - Metastasis status: Yes (121) + No (188).
  - 54,613 probes (Affymetrix Human Genome U133 Plus 2.0).

Association rules mining results:
- Example of associations between genes with low (0) and high (1) expression levels in patients with <5-year survival (t_survival = 0) and metastasis.

3. Results

Preprocessing results:
- Filtering and summarization process: 11,901 genes.
- Differential expression analysis: 1,516 genes.

Feature selection results:
- Number of genes selected using FOCI that best explains the response variable (metastasis): 120.
- Hierarchical clustering and Volcano Plot to analyze the selected genes expression levels.
- Coexpression network.

Association rules mining results:
- Identification of relevant genes to metastatic sarcoma.
- Discovery of significant relationships between metastasis-associated genes and < 5-year patient survival.

- Sarcomas are rare and heterogeneous mesodermal tumors that primarily affect children and adolescents (over 20% of all pediatric tumors).
- The relative 5-year survival rate for patients with metastatic sarcoma is usually low.
- Investigating the genetic basis of these tumors through genome-wide analysis is crucial due to their rarity and late diagnosis.

4. Conclusions

- Relevant group of genes associated with metastasis and low survival rates in sarcoma*.
- Potential new treatments for metastatic sarcoma patients integrating drug repurposing strategies, association rules, and gene-drug associations**.
- Future works: heterogeneous omic data and other diseases, advanced algorithms for identifying relevant association rules.

Acknowledgements: The authors would like to thank the Spanish Ministry of Science and Innovation for the support under PID2020-117954RB-C22 project.