# Summary

Multiple myeloma is a malignant tumor of the bone marrow that develops as a result of plasma cell mutations with complex pathogenesis. It is still an incurable disease, and the main goal of treatment is to extend survival time. An important task, therefore, becomes its early diagnosis, which is particularly important for detecting stage II and III disease. Multiple myeloma is a difficult disease to diagnose, so attempts are being made to use developments in computational technology in the form of artificial intelligence tools for data processing and analysis to develop algorithms that can select relevant early signs of the disease. The main method used for this purpose will be machine learning.

The dissertation presents methods for analyzing incomplete data, along with methods to recover it. This is the pre-processing step of data analysis, essential for properly executed studies. Data preprocessing is crucial for proper analysis and interpretation of results, especially when studying complex diseases such as multiple myeloma. Since the data is characterized by high dimensionality, the thesis therefore discusses feature sleight-of-hand methods used in machine learning. Metrics such as entropy, conditional entropy, mutual information, and Kulback-Leibler divergence are used to evaluate the relationship between features and the target variable. The analysis showed a high dependence of these methods on the information characteristics of the analyzed data.

In the dissertation, a study of an experimental data set taken among patients of a clinical hospital in Lublin was conducted. Pre-processing of the data was carried out, which included filling in missing results. At the next stage of data processing, feature selection was carried out, resulting in a reduction to 20 features from a set of 283 features. Twelve classifiers were used to assess the quality of the selected subsets.

A total of 1,534 computational experiments were conducted, testing different combinations, nine selection methods (between filtering, convolution, and embedding methods) as well as 12 different classifiers. The quality of each model was assessed using measures such as accuracy, sensitivity, specificity, F-score, and AUC parameters. The quality of each model was evaluated using measures such as accuracy, sensitivity, specificity, F-score, and AUC (area under the curve) parameters.

Based on the calculations, the two models JMI-B-20 and NJMIM-NB-20 were selected for multiple myeloma feature classification (diagnosis), as they achieved the highest accuracy and operated with fewer features. In the final part of the thesis, a proposal for further research on multiple myeloma diagnosis is presented.

**Keywords**: Multiple myeloma, data processing, classification methods and algorithms, diagnostic features