



# Identifying the Most Important Factors Associated with Multiple Sclerosis Using the Decision Tree Method

Sara Manochehri<sup>1\*</sup>, Zohreh Manochehri<sup>2</sup>, Tayebeh Lorestani<sup>3</sup>, Maryam Zamani<sup>1</sup>

<sup>1</sup>Department of Biostatistics, School of Public Health, Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran

<sup>2</sup>Social Development and Health Promotion Research Center, Health Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran

<sup>3</sup>Department of Ergonomics, School of Public Health, Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran

## OPEN ACCESS

### \*Corresponding Author:

Dept. of Biostatistics, School of Public Health, Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran

### Citation:

Manochehri S, Manochehri Z, Lorestani T, Zamani M. Identifying the Most Important Factors Associated with Multiple Sclerosis Using the Decision Tree Method. *Iranian biomedical journal* 2024; 28(7): 52.

## ABSTRACT

**Introduction:** Multiple sclerosis (MS) is a neurological disease that destroys the insulating covering of nerve cells. It is the third leading cause of disability after trauma and rheumatism. Studies have shown that the main cause of MS is not fully understood, and it is believed to be influenced by a combination of unknown genetic and environmental factors. Considering the ability of machine learning techniques, such as decision trees, to identify significant factors related to diseases, this study aimed to investigate various factors related to MS using the decision tree method.

**Methods and Materials:** This analytical and modeling study was conducted using the MS disease dataset. The data were obtained from the health registration system at Kermanshah University of Medical Sciences. A total of 317 individuals were studied from May 2016 to September 2017, comprising 188 individuals diagnosed with MS and 128 healthy controls. Magnetic resonance imaging was utilized for disease diagnosis. The data were processed in the R 4.0.3 software environment. The variables analyzed included gender, age, family history of MS, trauma, bowel disease, rheumatism, infectious disease, stress, depression, anxiety, migration, vitamin D deficiency, and smoking. The decision tree method and the gain ratio index were employed to assess the significance of factors influencing MS disease.

**Results:** According to the results, female gender, family history of MS, history of stress, vitamin D deficiency, and infectious disease with indices of 21.3, 18.94, 17.2, 15.78, and 15.21 were among the factors affecting MS disease.

**Conclusion and Discussion:** Three significant environmental factors associated with MS include a history of stress, a deficiency in vitamin D, and exposure to infectious diseases. Therefore, both individuals and service providers need to be aware of these factors to prevent the progression and exacerbation of its symptoms.

**Keywords:** Decision trees, Multiple sclerosis, Nervous system diseases

