ORIGINAL ARTICLE

Variable selection in gender and age decision-making for traumatic spine and thoracic pathologies after various accidents with Multivariate Adaptive Regression Spline (MARS)

Gülcan Gencer[1] (ID)   **Kerem Gencer**[2] (ID)   **Ahmet Dumanlı**[3] (ID)

[1] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Afyonkarahisar Health Sciences University. Afyonkarahisar / Türkiye

[2] Common Courses Department, Afyonkarahisar Health Sciences University. Afyonkarahisar / Türkiye

[3] Department of Thoracic Surgery, Faculty of Medicine, Afyonkarahisar Health Sciences University. Afyonkarahisar / Türkiye

Abstract

Trauma is a condition that affects the body's structure and results from outside factors. After heart disease and cancer, it is the most common cause of death across all age categories. For a variety of causes, people are routinely exposed to traumatic vertebral, thoracic pathologies and rib fractures. Ribs can be harmed by simple falls, impacts, and blunt injuries as well as broken due to car accidents and falling from a height. Magnetic resonance imaging or computed tomography are used to diagnose these fractures. In this study, non-linear complex methods were used to categorize gender and age by utilizing thoracic pathologies, fractures or cracks in the body as a result of traffic accidents or falling from a height, which have the feature of being a case in forensic issues. The most important data in the classification of gender and age were determined by Multivariate Adaptive Regression Spline (MARS) method. Although autopsy should be utilized in these situations, complex regression methods is intended to have an impact on quick and accurate decision-making about events in order to speed up or direct the process in the field of forensic medicine. As a result, the effectiveness of the experts subsequent predictions will be increased by the preliminary findings produced by real-world data and artificial intelligence algorithms or complex non-linear regression problems.

Keywords: Thorax, vertebra, multivariate adaptive regression spline, variable selection, traffic collision, autopsy

Corresponding Author:
Gülcan Gencer
Email: gencergulcan@gmail.com

Introduction

For children and young adults between the ages of 5 and 29, traffic-related injuries are the main cause of death. An estimated 1.3 million people each year pass away in traffic-related incidents. Men are more likely than women to get in car accidents starting at a young age. Young men under the age of 25 account for nearly three times as many road traffic fatalities (73%) as young women [1].

As a result of 187 thousand 963 fatal and injury traffic accidents that occurred in Turkey in 2021, 2,421 people died at the accident site and 2,941 people died within 30 days due to the cause and effect of the accident after they were injured and transferred to health institutions. While 49.3% of the deaths and 72% of the injuries occurred within the residential area, 50.7% of the deaths and 28% of the injuries occurred outside the residential area [2]. Following a car collision, trauma is a condition that disturbs the body's structural order as a result of external forces [3]. 40% of all trauma cases involve thoracic trauma, the third most frequent type of injury overall [3-5]. Thoracic injuries make up one-third of all trauma emergency and hospital admissions [5-7]. A tomographic image of the vertebral and rib fractures is shown in Figure 1.

Ribs can be harmed by simple falls, impacts, and blunt injuries as well as broken due to car accidents and falling from a height. The crime scene is investigated as soon as the accident occurs on the road. Physicians are supposed to take part in gathering any biological evidence that may exist, assess if the incident was truly an accident, and decide whether an autopsy is necessary if a deceased person is present [8]. The Turkish Penal Code states in Article 280 that "if physicians, dentists, pharmacists, midwives, nurses and other healthcare providers fail to report the situation to the competent authorities or delay in doing so, despite discovering a sign that a crime has been committed while performing their duties, shall be punished with imprisonment of up to one year." Because of this, individuals who visit the emergency department or health facility; Situations including traffic accidents, falls, assaults, and work accidents should be treated as legal cases and reported in writing or verbally to the appropriate authorities [9].

However, in terms of both criminal and civil law, forensic medicine's determination of age and gender is significant. The least amount of mistakes should be used when defining anatomical features [10]. Identification will continue to be one of the most important aspects in forensic cases, according to studies done so far.

Since bone fractures pose a threat to world health, a comprehensive study of the density and burden of fractures was also carried out for the first time in 2019. This study showed the global fracture incidence rate for each anatomical region from 1990 to 2019 by age and sex [11].

In many traffic collisions, especially those
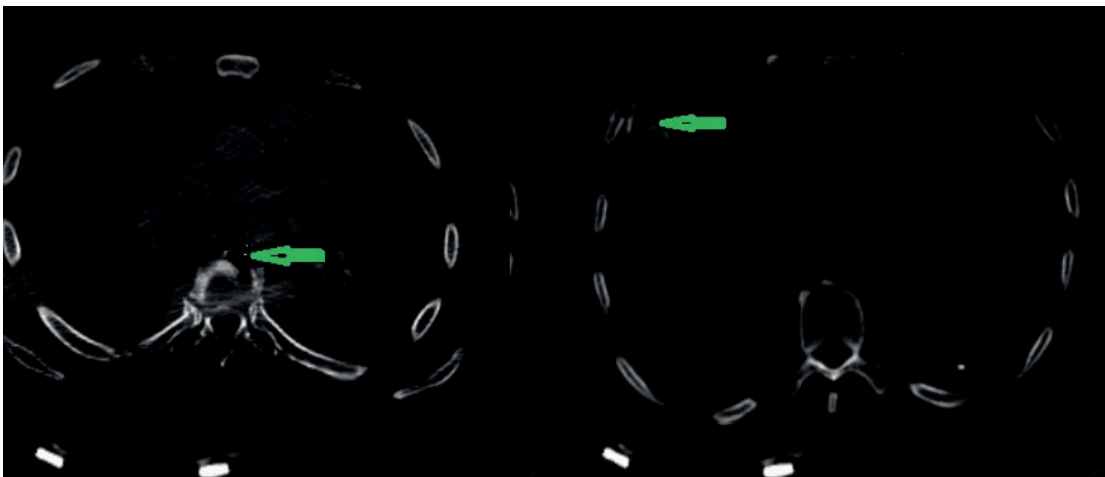
Figure 1. Tomography images of rib and vertebral fractures.

involving large cars, there are unrecognizably dead bodies [12,13].

To date, fractures have been found using techniques like magnetic resonance (MR) and computed tomography (CT) for a variety of causes. However, in this study, it has been proposed to estimate gender and age by looking for fractures or cracks in the body by combining deep learning and machine learning techniques. First, the variable selection was done, and then 150 features were chosen for both gender and age requirements. Making the data noiseless brought the feature reduction step before classification to a successful conclusion. For feature selection, a data collection with 251 samples overall was employed.

Materials and Methods

Thorax computed tomography scans were examined to determine which patients had spinal, rib fractures and thoracic pathologies. In forensic medicine, cases such as traffic accidents, falls, assaults, and work accidents are accepted as forensic cases. Therefore, the study included patients with rib fractures or thoracic vertebral fractures after a car collision and fall from a height. On a chest computed tomography, the degree of rib fractures, displacement, or non-displacement, were identified. Patients with thoracic vertebral fractures had their fracture levels, kinds (corpus fracture, transverse, and spinous process fractures) and thoracic pathologies identified. For statistical calculations, the R 4.0.2 environment was employed. Patients provided both written and verbal informed consents for the study, which were prepared in accordance with the Helsinki Declaration and approved by the local clinical research ethics committee of the Afyonkarahisar Health Sciences University (2022/16).

Feature Selection

A technique for identifying and removing unimportant participants from a data collection is feature selection. As a result of this procedure, the study moves forward using the information that best explains the topic under investigation. As a result, multidimensionality's complexity is eliminated. The literature describes three feature selection techniques: filtering, wrapping, and embedding. One of the wrapping approaches, MARS, was employed in this work to pick features [14].

Multivariate Adaptive Regression Spline

MARS is one of the new non-parametric regression methods. It was created in the early 1990s by Jerome H. Friedman [15], a statistician at Stanford University. MARS is a hybrid of Recursive Partitioning Regression (RPR) and Projection Pursuit Regression (PPR). This regression procedure is designed to accommodate both binary and continuous output variables. MARS is a versatile, precise, and quick technique. In contrast to linear methods, it considers subsets of variables [16]. This method, which is used for classification with a categorical output variable, has a wide range of applications because it is very flexible, accurate, and fast [17].

For $(x \in R)$, the following is the basis function:

$$(x - t)_+ = \{x - t, \text{ if } x > t \quad 0 \text{ , } otherwise \text{ and} $$
$$(t - x)_+ = \{t - x, \text{ if } x < t \quad 0 \text{ , } otherwise \quad (1)$$

In the expression above, t represents the point at which every function becomes piecewise linear at the value t. A key part of MARS models are hinge functions taking the form. Figure 2 on the right shows a mirrored pair of hinge functions with a knot at 2. The goal of this inflection is to obtain the projected pair x_j via nodes with the value x_{ij} [18].

$$C = \left\{ \left(x_J - t \right)_+, \left(t - x_J \right)_+ \right\} \quad (2)$$
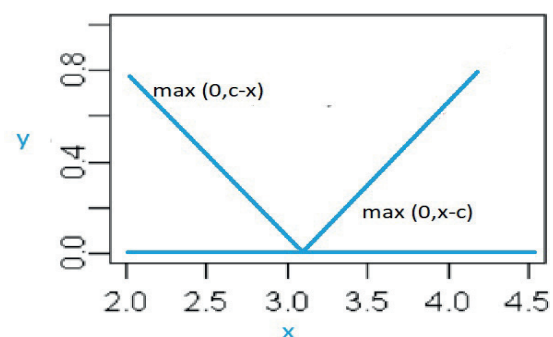
Figure 2. A mirrored pair of hinge functions with a knot at x=c.

The maximum model obtained initially is cropped with the second step, the backtracking algorithm and the best model is discovered by

eliminating the least effective variables one by one in the MARS model construction process. The submodel with the best approximation is determined by comparing the submodels obtained in this process using the Generalized Cross-Validation (GCV) criterion. The GCV method is described below.

$$GCV = \frac{\sum_{i=1}^{N} \left(y_i - \hat{f}_\lambda(x_i)\right)^2}{(1 - M(\lambda)/N)^2} \tag{3}$$

$M(\lambda)$ and denote the number of effective parameters and the number of observations, respectively. $M(\lambda)$ is found in this expression as $M(\lambda) = r + cK$ where r and K denote the number of independent basis functions and the number of nodes chosen in the incremental part, respectively.

Results

251 patients who experienced various trauma cases traffic accidents and fall injuries were included in the study. In the model constituting the data set, 63.35% occurred as a result of falling from a height, and 36.65% as a result of traffic accidents. These patients had thoracic vertebral fractures as well as rib fractures or rupture, injury, bleeding, and other conditions in the lung. According to fracture or fracture status, it was shown that the degrees and types of fractures (corpus fracture, transverse, and spinous process fractures) in patients with spinal and/or rib fractures and in patients with thoracic vertebral fractures varied. 150 different variables were taken into account. Variable status was coded with a 1 rather than a 0.

For two variables in this investigation, feature reduction was done. Age and gender are these factors. A total of 150 features were reduced to 9 features using the MARS feature selection method for the age variable. These are Lung Parenchymal laceration, Lung left lower lobe atelectasis, right 7th rib displaced fracture, right 10th rib displaced fracture, left 6th rib displaced difference, left 8th rib non-displaced fracture, left 9th rib non-displaced fracture, C7 vertebra transverse process fracture, T10 vertebra transverse fracture.

A total of 150 features were reduced to 5 features using the MARS feature selection method for the gender variable. These were identified as right 3rd rib displaced fracture, right 7th rib displaced fracture, left 5th rib non-displaced fracture, left 9th rib non-displaced fracture, T2 vertebra corpus fracture.

Table 1 shows demographic data. Table 1 shows demographic data. Of the respondents, 148 were men and 103 were women. In the data

Table 1. Features of the population.

Variables		Frequency	Mean
Sex	Male	148	45,5
	Female	103	55,2
Age	16-44	97	
	45-70	114	49,25
	71 and above	40	
	Total	**251**	

Table 2. The treatment method according to the disease etiopathology.

Treatment method		Traffic accidents	Fall from height	Total
Surgical	n	54	113	169
	%	58.7%	71.1%	66.8%
Medical	n	38	46	84
	%	41.3%	28.9%	33.2%
	n	92	159	251

collection, there are 40 patients over the age of 71, 114 patients between the ages of 45-70 and 97 patients between the ages of 16-44.

Table 2 shows the treatment method applied according to the etiopathology of the diseases. As a result of traffic accidents, surgical methods were used in 58.7% and medical methods were used in 41.3%. As a result of falling from a height, 71.1% of the patients were treated with surgical methods, while 28.9% were treated with medical methods.

In the MARS feature selection, 150 features were reduced to 5 features for sex and 9 features for age, as shown in Table 3. Lung parenchymal laceration, Lung left lower lobe atelectasis, right

7th rib displaced fracture, right 10th rib displaced fracture, left 6th rib displaced difference, left 8th rib non-displaced fracture, left 9th rib non-displaced fracture, C7 vertebra transverse process fracture, T10 vertebra transverse fracture for age.

The features selected by the algorithms are given in Figure 3. Commonly selected are left 9th rib non-displaced and right 7th rib displaced fracture.

It is suggested in this study to increase the use of complex non-linear regression methods in the field of health management and services. In our study, we found that the physician's decisions are consistent with the outcomes of MARS. This

Table 3. Selected features performances for gender and age.

Feature Selection Techniques	Number of Features Selected	Selected Features (MARS)
Sex	5	Right 3rd rib displaced fracture, right 7th rib displaced fracture, left 5th rib non-displaced fracture, left 9th rib non-displaced fracture, T2 vertebra corpus fracture
Age	9	Lung Parenchymal laceration, Lung left lower lobe atelectasis, right 7th rib displaced fracture, right 10th rib displaced fracture, left 6th rib displaced difference, left 8th rib non-displaced fracture, left 9th rib non-displaced fracture, C7 vertebra transverse process fracture, T10 vertebra transverse fracture.

Common Selected Features (MARS)
Left 9th rib non displaced
Right 7th rib displaced fracture
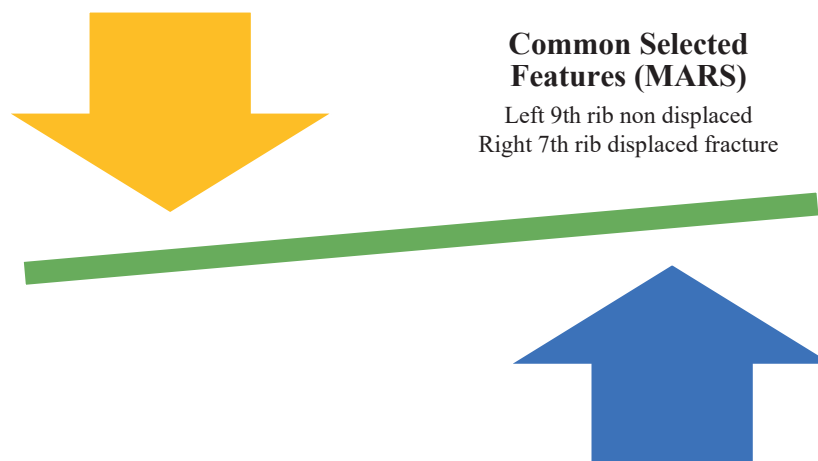
Figure 3. Commonly selected features.

demonstrates that complex non-linear regression can be used to reduce human-induced errors in diagnosis and treatment processes, as well as to aid in medical decision-making.

Discussion

It is very important to determine the age and gender of the corpses that are not recognized in the discipline of forensic medicine. Tülüce and Altuntaş [19] reported that traumatic pneumothorax is a very important pathology that requires early intervention due to rib fractures and contusion, Ergün and Topuz [20] reported that mortality is high in fracture patients accompanied by a chest injury, Dumanli et al [7] emphasized that the coexistence of vertebral and costal fractures in trauma patients should be kept in mind, and diagnosis and treatment planning should be made accordingly. In this study, in light of the literature, criteria such as fractures or cracks, bruises and injuries in the body were reviewed while determining the first criteria to be considered in the estimation of gender and age. Those who experienced trauma after a traffic accident and fell from a height were included in the study. These patients had rib fractures as well as possible thoracic vertebral fractures and thoracic pathologies. The factors seen in these patients were first reduced and then the most important variables were found. The basis of the data used in the study was traffic accidents and falls from height, which could be the subject of forensic medicine. In Table 2, as a result of the treatment method applied according to the etiopathology of the diseases, 58.7% of the traffic accidents were treated with surgical methods, 41.3% with medical methods, and as a result of falling from a height, 71.1% were treated with surgical methods and 28.9% with medical methods.

In this study, characteristics that should be considered in estimating age and gender were estimated using the regression curve technique. It has been suggested that deep learning, artificial intelligence and machine learning algorithms or complex nonlinear regression problems can be used in forensic applications.

Conclusion

A person may need recognition and separation from others for any reason. Both the government and the person's relatives attach great importance to this scenario. In these cases, forensic medicine institutions provide assistance. A person's identity consists of all the characteristics that allow identification and distinguish him from other people. A person may need recognition and separation from others for any reason. Using the complex nonlinear regression method on photographs of patients with spine and / or rib fractures and patients with thoracic vertebral fractures, this study identified the first features to consider when determining the age and sex of patients. X-rays were taken first. Patients with spine and/or rib fractures and patients with thoracic vertebral fractures were classified in a database by looking at these photographs. As a result, analysis of fractures or cracks seems to offer valuable information as a starting point for research. However, it is clear that the purpose of such a classification is to arrive at a general conclusion rather than a definitive answer. Accuracy rates are expected to serve as a guideline for the study to have appropriate evidential value. This research seeks to guide future research. As a result, machine learning algorithms or complex nonlinear regression problems used in forensics could help open the door to more experimental work in other disciplines. It is estimated that these people will be a new guide to the authorities when it is determined that the people who suddenly disappeared are buried in a place years later.

Conflict of interest

The authors declared no conflicts of interest.

References

1. World Health Organization. Road traffic injuries 2022. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

2. Türkiye İstatistik Kurumu. Karayolu Trafik Kaza İstatistikleri 2021.

3. Haberal M, Köksal E, Civan M, Tülüce K, Karadağ H, Köksal Z. Traumatic hemothorax: Analysis of 108 cases. J Exp Clin Med. 2013;30(1):31. doi:10.5835/jecm.omu.30.01.008.

4. Apilioğulları B, Esme H, Ceran S, Düzgün N. Retrospektive analysis of 48 cases with thoracic trauma. Anatol J Med Sci. 2015;1(1):14-8. doi: 1015197/sabad.3.1.04.

5. Ceran S, Sunam GS, Aribas OK, Gormus N, Solak H. Chest trauma in children. Eur J Cardio-Thorac Surg. 2002;21(1):57-9. doi: 1016/S1010-7940(01)01056-9.

6. Büyükkarabacak Y, Şengül AT, Gürz S, Pirzirenli MG, Başoğlu A. Associated traumas in thoracic trauma patients: Their effects on mortality and morbidity. BSJ Health Sci. 2019;2(3):78-84.

7. Dumanli A, Aydin S, Gencer G. Evaluation of traumatic vertebra and rib fractures. J Med Res Surg. 2022:3(5), 86-95.

8. Çetin G, Özaslan A. Trafik kazasına bağlı yaralanmalar. Soysal Z, Çakalır C. Editors: Adli Tıp İstanbul: İstanbul Üniversitesi; 1999.

9. Şahinoğlu S, Büken NÖ. Türk Ceza Kanunu Madde 280'nin Tıp Etiği açısından İncelenmesi, Uluslararası Birleşik Biyoetik Kongresi. Kongre Kitabı, Şanlıurfa, 2005; 168-169.

10. Aktaş E. Kostaların sternal uç kemik morfolojisinde yaşa ilişkin progressif değişikliklerin kişinin öldüğü zamanki yaşının saptanmasında kullanılabilirliği [in Turkish]. 1997. [Dissertation-Ege University]

11. Wu A-M, Bisignano C, James SL, Abady GG, Abedi A, Abu-Gharbieh E, et al. Global, regional, and national burden of bone fractures in 204 countries and territories, 1990–2019: A systematic analysis from the Global Burden of Disease Study 2019. Lancet Healthy Longev. 2021;2(9):e580-92. doi: 10.1016/S2666-7568(21)00172-0.

12. İHA. Smoke covered the sky! Oil tanker collided with truck in China. 2022. https://www.cnnturk.com/dunya/dumanlar-gokyuzunu-kapladi-cinde-petrol-tankeri-kamyonla-carpisti.

13. Cihan. 'U-turn' on highway destroyed a family 2010. https://www.haber7.com/guncel/haber/632235-otoyolda-u-donusu-bir-aileyi-yok-

14. Chandrashekar G, Sahin F. Engineering E. A survey on feature selection methods. Computers & Electrical Engineering. 2014;40(1):16-28. doi: 10.1016/j.compeleceng.2013.11.024.

15. Friedman J. Multivariate adaptive regression splines. Ann Stat. 1991;19(1):1-67. doi: 10.1214/aos/1176347963.16.

16. Mukhopadhyay A, Iqbal A. Prediction of mechanical property of steel strips using multivariate adaptive regression splines. J Appl Stat. 2009;36(1):1-9. doi: 10.1080/02664760802193252.

17. Put R, Xu QS, Massart DL, Vander Heyden Y. Multivariate adaptive regression splines (MARS) in chromatographic quantitative structure–retention relationship studies. J Chromatogr A. 2004;1055(1-2):11-9. doi: 10.1016/j.chroma.2004.07.112.

18. Kartal Koc E, Bozdogan H. Model selection in multivariate adaptive regression splines (MARS) using information complexity as the fitness function. Mach Learn. 2015;101:35-58. doi: 10.1007/s10994-014-5440-5.

19. Tülüce K, Altuntaş G. Evaluation of 127 patient with traumatic pneumothorax: Single center experience [in Turkish]. Sakarya Med J. 2020;10(4):655-60. doi: 10.31832/smj.793475.

20. Ergün T, Obuz Topuz Ç. The effect of traumatic lung injury on mortality in fracture patients; retrospective examination of polytrauma patients [in Turkish]. Med J Mugla Sitki Kocman University. 2022;9(3):291-4. doi: 10.47572/muskutd.998252.