Middle Black Sea Journal of Health Science / Mid Blac Sea J Health Sci **August 2023; 9(3): 429-439**

RESEARCH ARTICLE **DOI:** 10.19127/mbsjohs.1316728

Determination of Factors Affecting Severity of *Helicobacter pylori* for Gastric Biopsy Samples by CART Decision Tree Algorithm

Türkan Mutlu Yar[1(ID)], Ülkü Karaman[1(ID)], Yeliz Kaşko Arıcı[2(ID)]

[1]Faculty of Medicine Department of Parasitology., Ordu University, Ordu, Turkey
[2]Faculty of Medicine, Department of Biostatistics and Medical Informatics, Ordu University, Ordu, Turkey

Received: 19 june 2023, Accepted: 20 July 2023, Published online: 31 August 2023

Abstract

Objective: *H. pylori* wich is one of the important gastric pathogens and is a motile, non-sporeless, encapsulated, microaerophilic, gram-negative bacterium. The aim of this study was to determine the factors affecting disease severity in patients with a positive pathologic diagnosis of Helicobacter pylori after gastric biopsy by data mining. It was aimed to utilize the more descriptive structure of data mining algorithms compared to traditional classification and regression approaches.

Methods: The study data were obtained from gastric biopsy samples of 1247 patients, 40.5% male and 59.5% female, who were sent to the pathology laboratory between 2014 and 2018. A total of 6 factors including age, gender, inflammation, metaplasia, atrophy and activation, which are thought to have an effect on gastric *H. pylori* severity, were examined. Querying the effects of factors was done with the CART (Classification and Regression Trees) decision tree algorithm, one of the data mining algorithms.

Results: The factors ranking as their effect on the severity of gastric h. pylori, as follows; activation > inflammation > metaplasia > atrophy > age > gender in a percentage of normalized importance at 100.00%, 88.6%, 51.4%, 38.1%, 12.8%, 3.3% respectively.

Conclusion: As a result, levels of activation, inflammation, and metaplasia emerged as the most important factors affecting gastric *H. pylori* severity.

Key words: Data mining, Decision Tree, CART Algorithm, *H. pylori* stomach

Suggested Citation: Suggested Citation: Yar TM, Karaman U, Kaşko Arıcı Y, Determination of Factors Affecting Severity of *Helicobacter pylori* by CART Decision Tree Algorithm. Mid Blac Sea Journal of Health Sci, 2023;9(3):429-439.

Address for correspondence/reprints:

Türkan Mutlu Yar

Telephone Number: +90 (530) 349 32 06

E-mail: mutluyarr@gmail.com

INTRODUCTION

H. pylori is one of the important gastric pathogens and is a motile, non-sporeless, encapsulated, microaerophilic, gram-negative bacterium. This bacterium settles in theantrumpart of the stomach and can generally be spiral andcurved (Dağdartan, 2011). *H. pylori* is a bacteriumthat causes important upper gastrointestinal diseases such as gastroduodenal ulcer, gastritis, adenocarcinoma, gastricmucosa-associated lymphoid tissue lymphoma. While more than 50% of the world's population is infected with H. pylori, this rate is 70-90% in developing countries, while it may be lower in developedcountries (Kanadalı et al, 2004, Kuslers et al, 2006).

Some different noninvasive and invasive methods are used in thediagnosis of *H. pylori* infection. Among these methods, non-invasive methods that do not require endoscopy, such as urea-breath test, serological methods, stool culture, antigen/nucleic acid search in stool are used. Many invasive methods that require endoscopy are gastric biopsy culture sample, histopathological examination, rapid urease test, and molecular methods (Hirschi et al, 2007, Uyanık et al, 2007). Although many invasive and noninvasive tests are used in thediagnosis of H. pylori, culture and histological examinations are accepted as the gold standard in the diagnosis of this infection. In the diagnosis of H. pylori, it is important to take, transport and store the gastric biopsy sample in appropriate ways. At least two biopsy samples should be takenfromtheantrumregionandthepatient'scimetidineintakeshould be discontinued 5-7 days before the procedure. In the diagnosis, successful results can be obtained by staining and examining the smears prepared by crushing on a slide from biopsy samples by gram stainingmethod (Erdem, 1999).

A decision tree is a model that shows classification outcomes and decision rules in a data structure that resembles a tree. As an inductive learning method, its goal is to turn the ostensibly disorganized and disorganized known cases into a tree model that, by technological means, can predict unknown instances (Greff et al., 2016).

In this study, the decision tree model was used to take advantage of the more descriptive nature of this method compared to the traditional Classification and Regression approaches. For this purpose, CART (Classification and Regression Trees) Decision Tree Algorithm was preferred.

Our aim in the study is to determine the factors affecting the severity of the disease in patients who applied to Ordu University Training and Research Hospital with dyspeptic complaints in May-Grünwald & Giemsa and Hematoxylin & eosin and *H. pylori* retrospectively found positive gastric antrum biopsy samples by data mining. It is aimed to benefit from the more descriptive structure of data mining algorithms compared to

traditional classification and regression approaches.

METHODS

Data of the study

The study data were obtained retrospectivelyfrom gastric biopsy samples of 1247 patients, who were sent to the pathology laboratory and diagnosed as *H. pylori* positive between 2014 and 2018 in Ordu University Research and Training Hospital.The patients were included into the study by considering the Sydney criteria in accordance with *H. pylori* severity degrees in 3 groups; mild (n=538), moderate (n=445) and severe (n=264).

Statistical Analysis

In this study, the decision tree model was used to take advantage of the more descriptive structure of this method compared to traditional Classification and Regression approaches. For this purpose, CART (Classification and Regression Trees) Decision Tree Algorithm was preferred.

The classification appears to be one of the most crucial approaches and strategies utilized in machine learning or data mining among the many others (Rutkowski et al., 2014).

With a maximum tree depth of 5, all minimum cases in the parent node were 100, and the minimum cases in the child node were 50, the CART method was used. The Gini index was used to calculate the homogeneity of the nodes. IBM SPSS v28 (IBM, Armonk, NY, USA) statistical software was used to calculatethe algorithmand to calculate descriptive statistics of the dataset.

The frequencies and percentages were given as descriptive statistics for categorical variables, whereas the mean+standard deviation (SD) and minimum-maximum values were given for continuous variables.

CART Algorithm

To handle categorization and prediction issues, data miners utilize decision trees, a visual technique that is simple to grasp and interpret. Breiman et al. (1984) initially created Classification and Regression Trees (CART), one of the decision tree methods, which uses both categorical and continuous data to solve classification and regression issues. Regression Trees-RT are used when the dependent variable is continuous, while technique classification trees (CT) are used when the dependent variable is categorical (Chang and Wang, 2006).

CART may be used serially and is based on Hunt's algorithm. When choosing the dividing attribute, it makes use of the Gini index splitting measure. Because it uses regression analysis with the aid of regression trees, CART differs from previous Hunt's-based algorithms (Kumar and Vijayalakshmi, 2011; Priyama et al. 2013).

RESULTS

The sample of the study consisted of a total of 1247 gastric biopsy samples, 40.5% male and 59.5% female, sent to the pathology laboratory and diagnosed as *H. pylori* positive between 2014 and 2018. The samples' frequency

distributions of gender,levels of the severityof *H. pylori*, inflammation, metaplasia, atrophy, and activation (increased neutrophil count) were given in Table 1.Descriptive statistics of the samples' according to the severity of *H. pylori*were given in Table 2.The mean age of mild, moderate and severe groups, which are *H. pylori* severity levels, were 50.17±15.10(17-84),47.67±15.15(17-83) and 47.58±14.31(20-86) years, respectively.

Table 1: Frequency distributions of gender,levels of the severityof H. pylori, inflammation, metaplasia, atrophy, and activation

Variables		**n**	**%**
Gender	Female	742	59.5
	Male	505	40.5
H. pylori severity	Mild	538	43.1
	Moderate	445	35.7
	Severe	264	21.2
Inflammation	Negative	19	1.5
	Mild	395	31.7
	Moderate	449	36.0
	Severe	384	30.8
Metaplasia	Negative	813	65.2
	Type 1	294	23.6
	Type 2	112	9.0
	Type 3	28	2.2
Atrophy	Negative	612	49.1
	Mild	500	40.1
	Moderate	115	9.2
	Severe	20	1.6
Activation	Negative	209	16.8
	Mild	459	36.8
	Moderate	417	33.4
	Severe	162	13.0

Table2: Descriptive statistics of the samples according to the severity of *H. pylori*

H. pylori	Female		Male		*Total*		Age	
Severity	n	%	n	%	*n*	*%*	Mean±SD	Min.-Max.
Mild	342	63.6	196	36.4	*538*	*43.1*	50.17±15.10	17-84
Moderate	260	58.4	185	41.6	*445*	*35.7*	47.67±15.15	17-83
Severe	140	53.0	124	47.0	*264*	*21.2*	47.58±14.31	20-86

The tree was designed in the root (node 0), branch (nodes: 1, 2, 4, 5, 8, 9 and 12) and leaf (nodes: 3, 6, 7, 10, 11, 13, 14, 15 and 16) nodes (Figure 1).

In the study, theseverity of *H. pylori* was found to be 43.1%, 35.7% and 21.2%, respectively, in the mild, moderate and severe groups (Node 0). Potential factors that caused this variation in the severity of *H. pylori* were ranked according to their estimation importance with the CART Decision Tree Algorithm.With the CART, the most effective factor on the severity of H. pyloriwas estimated as activation level (p=0.048). Therefore, the root node was first divided into two branches negative-mild (53.6%) and moderate-severe (46.4%) according to activation levelwhile the rate of H. pylorimildwashigh (59.7%) in those with negative-mild activation (Node 1), the rate of those with moderate *H. pylori* severity (43.4%) washigher in thosewithmoderate-severe activation (Node 2). Those with negative-mild activation were dividedinto two childnodes as

moderate-severe (Node3) and negative-mild (Node4) according to the level of inflammation (p=0.010). In both groups, the rate of those with mild *H. pylori* severity was high (47.9 and 69.6%, respectively), but the rate of those with severe *H. pylori* severity was approximately 2 times higher in the moderate-severe group (14.9% versus 8.2%).

Those with negative-mild inflammation level were dividedinto two childnodes as ≤ 39.5 (Node7) and >39.5 (Node8) according to age (p=0.002). The rate of thosewith severe *H. pylori* severity was higher in those whose age was ≤39.5 (15.2% versus 5.9%). In those with age>39.5, two childnodeswereseparated as negative (Node 11) and type1-type-2 (Node 12) according to metaplasia level (p=0.001). Compared to type1-type 2 in terms of metaplasia, in negativeones; The rate of those with mild *H. pylori* severity was lower, whereas the rate of those with moderate and severe levels was higher (67.2% versus 77.3%; 26.1% versus 17.5%; 6.7% versus 5.2%). While there was no other factor affecting theseverity of H. pyloriin patient saged>39.5 years with metaplasia negative, those with metaplasia type1-type2 were divided into two age groups as ≤60.5 (Node 15) and >60.5 (Node 16) (p=0.001). Compared to those aged >60.5, those aged ≤60.5; The rate of those with mild *H. pylori* severity was lower, where as the rate of those with moderate and severe levels was higher (70.0% versus 87.5%; 22.2% versus 10.9%; 7.8% versus 1.6%).

Those with moderate-severe activation levels were divided into two childnodes, which see the metaplasial level in themselves, as negative-type 1 (Node5) and type2-type3 (Node4) (p=0.003). In the metaplasia negative-type 1 group, the proportion of those with moderate *H. pylori* severity was high (44.0%) in themetaplasia type2-type3 group. The rate of those with severe pylori severity was high (45.0%) *H. pylori* severity was not affected by any other factor.

However, Metaplasia negative-type 1 group was divided into two childnodes, severe (Node 9) and negative-mild-moderate (Node 10) in terms of inflammation level (p=0.001). Those with severe inflammation had a higher rate of *H. pylori* (34.3% versus 24.3%). Those with severe inflammation showed branching according to the level of activation (p=0.001). When the two groups showing moderate level activation (Node 13) and severe level activation (Node 14) were compared in terms of *H. pylori* severity; The rate of mild was moderate in the moderate activation group (28.8% versus 16.1%), and the rate of severe was higher in the severe activation group (43.7% versus 40.4%; 40.2% versus 30.8%). The rate of thosewith severe *H. pylori* inflammation was approximately 2 times higher (14.9% versus 8.2%) in the moderate-severe group. Those with negative-mild inflammation level were

dividedinto two childnodes as ≤ 39.5 (Node7) and >39.5 (Node8) according to age (p=0.002). The rate of those with severe *H. pylori* severity was higher in those whose age was ≤39.5 (15.2% versus 5.9%). In those with age>39.5, two childnodeswereseparated as Negative (Node 11) and type1-type-2 (Node 12) according to metaplasia level (p=0.001). Compared to type1-type 2 in terms of metaplasia, in negative ones; The rate of those with mild *H. pylori* severity was lower, whereas the rate of those with moderate and severe levels was higher (67.2% versus 77.3%; 26.1% versus 17.5%; 6.7% versus 5.2%).

While there was no other factor affecting the severity of *H. pylori* in patients aged > 39.5 years with metaplasia negative, those with metaplasia type1-type2 were divided into two age groups as ≤60.5 (Node 15) and >60.5 (Node 16) (p=0.001). Compared to those aged >60.5, those aged ≤ 60.5; The rate of those with mild H. pyloriseveritywaslower, whereas the rate of those with moderate and severe levels was higher (70.0% versus 87.5%; 22.2% versus 10.9%; 7.8% versus 1.6%).

DISCUSSION

H. pylori is an aerophilic bacteria species that can be the cause of many diseases such as chronic gastritis, gastritis malt lymphoma, aseptic carriage colonized in the gastric mucosa, usually in the antrumand corpus, and gastric cancer in humans (Mirza, 2011). It has been reported that *H. pylori* can generally settle in the antrum and cause chronic active gastritis, atrophic gastritis, intestinal metaplasia, dysplasia, and gastric cancer (Topal et al, 2004). Most of the samples examined in the study were takenfromtheantrumregionandthe condition of the patients was evaluated in terms of intestinal metaplasia, atrophy, inflammationand activation.

In the study, inflammation was found in 98% of *H. pylori* positive patients, intestinal metaplasia in 34.6% and atrophy in 50.7%. Activation was positive in 82.9%. In studies on intestinal metaplasia; Kesici (2018) found a significant relationship between *H. pylori* positivity and intestinal metaplasia. They also found a significant relationship between intestinal metaplasma positivity and *H. pylori* density. Kurtulus et al. (2017) could not detect a significant relationship between *H. pylori* positivity and intestinal metaplasia. Sasa et al. (2002) found an increase of 50.7% in intestinal metaplasia and atrophy frequency and in *H. pylori* positivity.

Güner andTuncer (2019) found that H. pylori positivity was effective on duodenal ulcer, atrophic gastritis, intestinal metaplasia. A significant correlation was found between the colonization status of the bacteria and inflammation in patients with H. pylori positivity (p<0.01) Sipponen et al. (1997) reported a positive correlation between H. pylori inflammation and colonization.

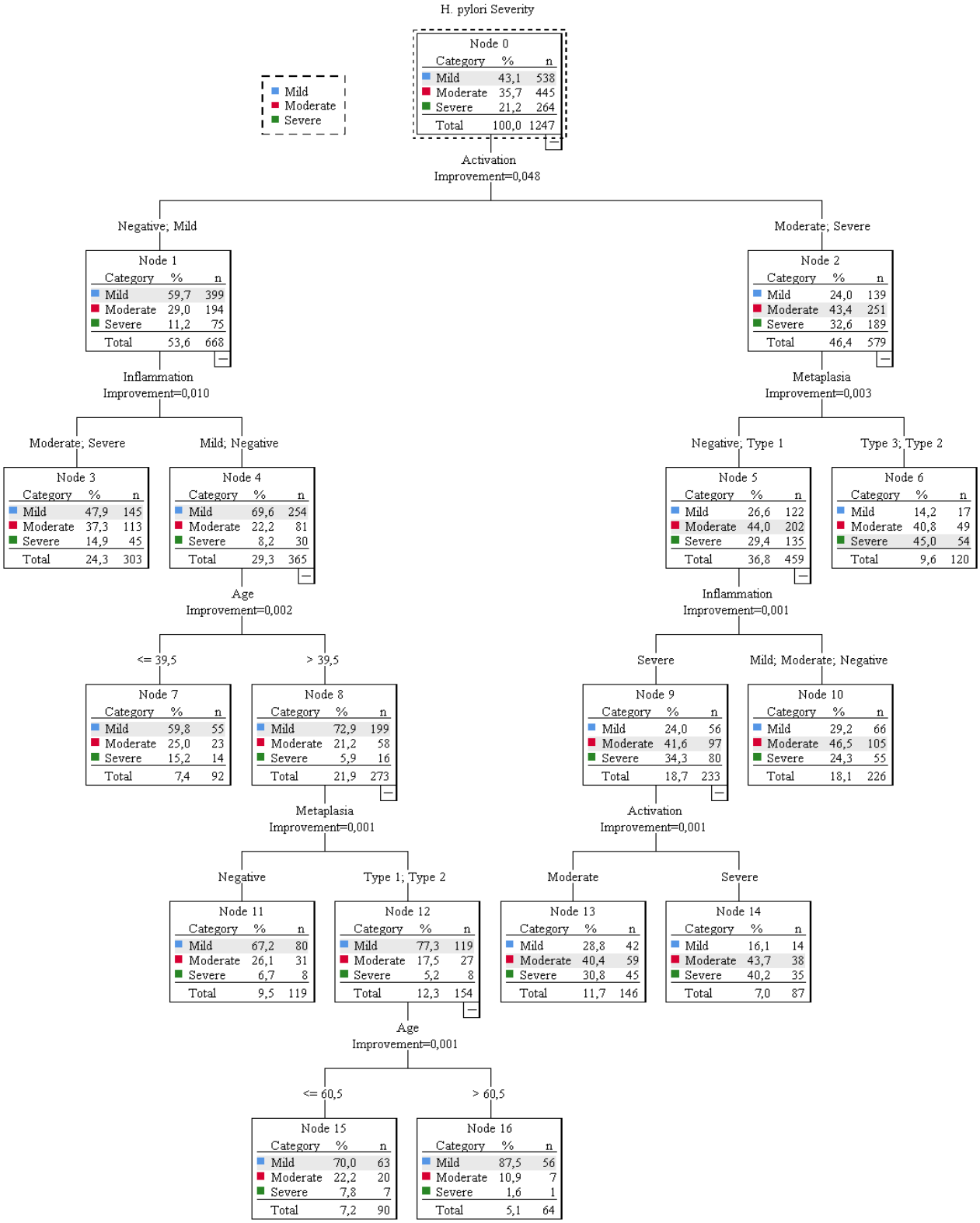

Figure 1. Decision treediagram obtained by CART algorithm for the severity of gastric *H. pylori* The importance levels of the factors (independent variables) based on CART predictive algorithm for the severity of gastric *H. pylori* was presented in Figure 2.

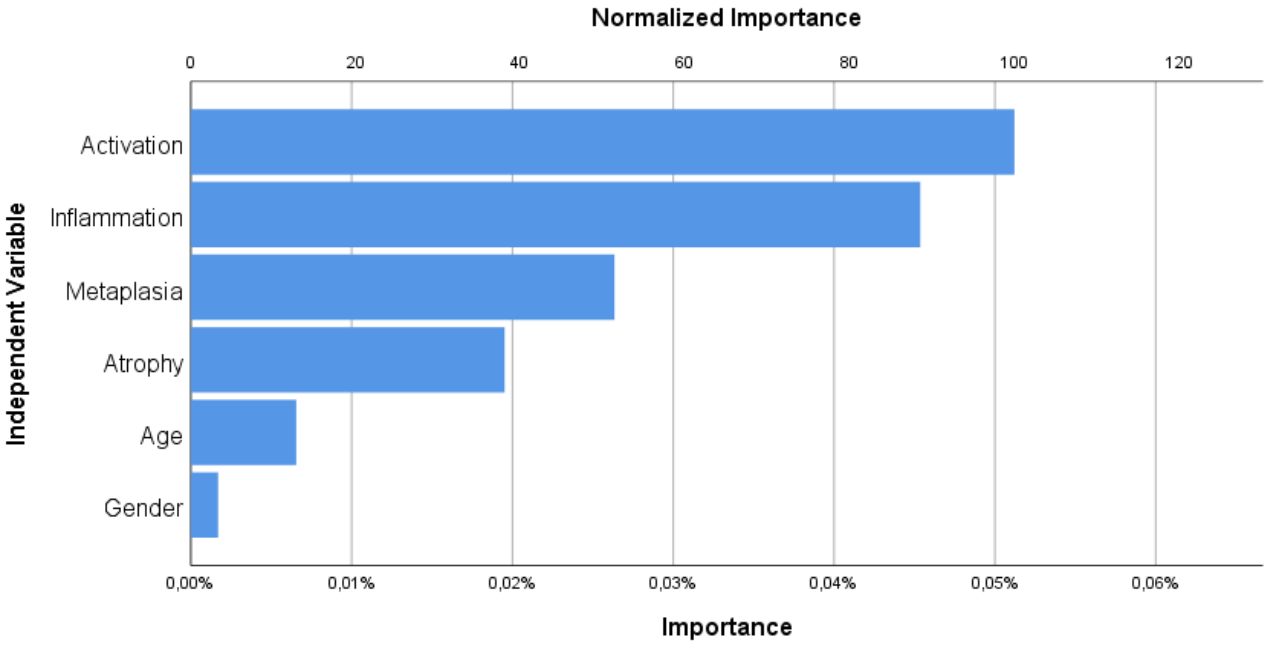

Figure 2. Normalized importance levels of the factors,sorted by decreasing importance

They reported that there is a positive correlation between H. pylori severity and activation. A significant correlation was found between *H. pylori* severity and activation positivity ($p<0.001$). The increase in the severity of *H. pylori* increased the rate of activation positivity. The rate of patients with mild and moderate *H. pylori* positivity was found to be higher than those without atrophy, and the rate of those with atrophy (56.4%) in severe H. pylori-positive patients was higher than the rate (43.6%) in patientswith severe *H. pylori* positivity. Asaka et al. (2001) also reported that atrophic gastritis increased in *H. pylori* positivity. Kurtulus et al. (2017) found a significant relationship between atrophy and *H. pylori* positivity.

In the study of *Karaman*, inflammation was found in 98% of *H. pylori* positive patients, intestinal metaplasia in 34.6% and atrophy in 50.7%. Activation was positive in 82.9 of the patients. It has been determined that there is a significant relationship between the colonization status of the bacteria and inflammation in patients with *H. pylori* positive. The relationship between positivity of intestinal metaplasia and *H. pylori* density was found to be statistically significant. Intestinal metaplasia and an increase of 50.7% were detected in bacterial positivity. The increase in *H. pylori* severity increased the incidence of activation positivity. While the rate of patients with mild and moderate *H. pylori* positivity was higher than those without atrophy, the rate of patients with severe *H. pylori* positivity

(56.4%) was higher than the rate of patients without atrophy (43.6%) (Karaman, 2020). Similar results were obtained with data mining in the study. *H. pylori* activation > inflammation > metaplasia > atrophy > age > sex gives normalized significance percentages of 100%, 88.6%, 51.4%, 38.1%, 12.8%, and 3.3%, respectively. Activation and inflammation with a given normalized significance greater than 50% indicated that themarkers of metaplasiaweresignificant.

CONCLUSION

In this study activation showed the highest percentage of normalized importance due to the severity of gastric *H. pylori* because therefore, the presence of activation is important in the follow-up of the treatment of the bacteria. Inflammation showed 88.6% of importance where it conveys high effect although it was not the highest because Inflammation was also found to increase in the presence of bacteria.In this the study as a result, activation, inflammation and metaplasia were determined as the most important factors affecting the severity of gastric *H. pylori* .

Ethical Approval: Ethical approval for this study was not required as data mining was used.

Peer-review: Externally peer-reviewed.

Author Contributions:

Concept: TMY, ÜK, YKA, Design: TMY, ÜK, YKA, Supervision: TMY, ÜK, YKA, Data Collection and/or Processing: TMY, ÜK, YKA, Analysis and/or Interpretation: TMY, ÜK, YKA, Writing: HS, EU, MCU, EA, CA, SD

Conflict of Interest: No conflict of interests

Financial Disclosure: No financial support

REFERENCES

1. Asaka M, Sugiyama T, Nobuta A, Kato M, Takeda H, Graham DY. Atrophic gastritis and intestinal metaplasia in Japan: Results of a largemulticenterstudy. Helicobacter, 2001;6, 294-9.
2. Breiman L, Friedman JH, Olshen R, & Stone, ACG. (1984). Classification and Regression Trees. Wadsworth International Group, Belmont, California, USA.
3. Chang LY & Wang HW. Analysis of Traffic Injury Severity: An Application of NonParametric Classification Tree Techniques. Accident Analysis & Prevention, 2006;38, 1019-1027.
4. Dağdartan U. (2011). Isolation of Helicobacter pylori from Gastric and Duodenum Biopsy Samples and Investigation of Antimicrobial Resistance. Department of Medical Microbiology, Specialization Thesis in Medicine. Myrtle
5. Erdem B: Campylobacterand Helicobacter. Basic and Clinical Microbiology. UstacelebiSh. Güneş Bookstore, Ankara, 1999; pp. 531–40
6. Greff K, Srivastava RK, Koutník J, Steunebrink BR. and SchmidhuberJ. LSTM: a search space odyssey. IEEE Transactions

Neural Network Learning System, 2016;28(10): 2222–2232

7. Güner A, Telli N *Helicobacter pylori*: A New Food Pathogen? Erciyes University Journal of Vet Fak 2012;9(1) 51-63.
8. Hirschl AM, Makristathis A. MethodstodetectHelicobacterpylorifromculturetomolecularbiology. Helicobacter 2007; 12(Suppl 2): 6-11.
9. Kadanalı A, Özkurt Z. *Helicobacter pylori* infection: Epidemiology, pathogenesis and related diseases. KlimikJournal 2004; 17(3): 146-50.
10. Cutter U, Comparison of intestinal metaplasiaandHelicobacterpylori scores in patients undergoing upper gastrointestinal endoscopy. CukurovaMed J, 2018;43(3): 574-580.
11. Karaman Ü, (2020). Evaluation of the diagnosis of Helicobacter pylori from gastric biopsy samples by staining methods. Giresun University Medical Microbiology Department, Master's thesis. Giresun.
12. Kumar SA & Vijayalakshmi MN. (2011). Efficiency of decisiontrees in predictingstudent'sacademicperformance. First International Conference on Computer Science, Engineering and Applications, India.
13. Kurtuluş A, Akın M, Buldukoğlu OÇ, Yalçınkaya T, Yıldırım B, Gelen MT. The Frequency of Helicobacter pylori and the Demographic, EndoscopicandHistopathologicalCharacteristics of the Patients Performed Endoscopy in the Tertiary Health Institution in Antalya RegionAkd Tip D, 2017;2:101-106.
14. Kusters JG, vanVliet AHM, Kuipers EJ. Pathogenesis of Helicobacterpylori infection. Clin Microbiol Rev 2006; 19(3): 449-90.
15. Mirza E. (2001). The prevalence of Helicobacter pylori in patients with metabolic syndrome. Gazi UniversityFaculty of Medicine, Department of Internal Medicine. Master thesis. Ankara.
16. Rutkowski, L., Jaworski, M., Pietruczuk, L., &Duda P. The CART decision tree for mining data streams. Information Sciences, 2014;266,1-15. https://doi.org/10.1016/j.ins.2013.12.060
17. Priyama A, Abhijeeta RG, Ratheeb A, Srivastavab S. Comparative analysis of decision tree classification algorithms. International Journal of Current Engineering and Technology. 2013;3(2):334-7.
18. Sasa G, Milosav S, Vuka K. (2002). The relationship between the density of Helicobacter pylori colonisation and the degree of gastritis severity. Gastro enterol hepatol, 21, 3- 4.
19. Sipponen P, Stolte M. (1997). Clinical impact of routine biopsies of the gastric ant rumand body. Endoscopy, 29, 671-8.
20. Topal D, Göral V, Yılmaz F. Association of *Helicobacter pylori* with Intestinal

Metaplasia, Gastric Atrophy and Bcl-2. Turkey Clinics J Gastroentero hepatol, 2004;15, 65-73.

21. Tuncel F, BozkurtF, Gülseren A, Usta Y. The relationship between the incidence of celiac disease and *Helicobacter pylori* gastritis in childhood. Endoscopy. 2019;27(1):16-19.
22. Uyanık MH, Aktaş O. Microbiological diagnosis of Helicobacter pylori. EAJM 2007;39(3): 205-9.