International Research Journal of Pharmacy and Medical Sciences
ISSN (Online): 2581-3277

# Screening of Genes That May be Associated with Gastric Cancer Using Bioinformatics and Artificial Intelligence Methods and Interpretation of Individual-Based Results

Zeynep Kucukakcali[1*], Ipek Balikci Cicek[1]
[1]Department of Biostatistics and Medical Informatics, Inonu University Faculty of Medicine, 44280, Malatya, Turkey

***Abstract*— *Aim**: There is no screening technique for gastric cancer, which is one of the most common cancer types among the causes of death from cancer, and the early diagnosis of the disease is very low. Therefore, there is a need for early detection of the disease. The aim of this study is to identify potential genes that may be associated with gastric cancer by bioinformatics methods using open access gene expression data obtained from human gastric tumor tissues and normal gastric tissues, and also to classify the data with random forest (RF), one of the machine learning models, and to evaluate the genes that may be associated with the disease on an individual basis using LIME, one of the explainable artificial intelligence (XAI) models. **Methods**: Bioinformatics analyses of the data were performed using the limma package in the R programming language. In the modelling phase, classification was performed using RF model and the classification performance was evaluated with accuracy, balanced accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F1-score metrics. LIME, one of the XAI methods, was used to explain the applied model. **Results**: According to the results of bioinformatic analysis 3283 expression was found with statistically significant differences in gene expression levels between the two groups. As a result of modelling with RF the performance metrics obtained from the model were accuracy (96.7%), balanced accuracy (96.7%), sensitivity (93.3%), specificity (100%), positive predictive value (100%), negative predictive value (93.8%), and F1-score (96.6%), respectively. According to the results of XAI model, CORO1C, CAPN13, SST, GGT6, ARSD, CYP3A5 genes were found to be effective in tumor formation. **Conclusion**: Genes that may be associated with gastric cancer were identified by bioinformatics and machine learning models. Based on the changes of the identified genes in individuals, future studies can be directed or individual trials can be conducted for the treatment of the disease.*

## I. INTRODUCTION

According to GLOBOCAN 2018 data, 1,033,701 new cases of gastric cancer (GC) are diagnosed worldwide each year, representing 5.7% of all diagnosed cancer cases. In 2018, there were 782,685 deaths due to GC. It was the fifth most commonly diagnosed cancer in 2018, accounting for 8.2% of all deaths from cancer. GC is the third most common cause of cancer-related deaths after lung and colorectal cancers (1). GC is uncommon in persons under the age of 50 in all demographics and nations. GC incidence rates rise with age, peaking between the ages of 55 and 80. Men are two to three folds as likely than women to get GC (2, 3). Eastern Asia, Central and Eastern Europe, and various Central and South American nations with the greatest incidence of GC, whereas North America, Australia, and North Africa with low incidence (4). The most common histopathological subtype of GC is gastric adenocarcinoma (GAC), accounting for approximately 90-95% of cases (5). Some causal factors that cause gastric cancer have been identified. The most important risk factors are Helicobacter pylori, nutrition, drug use, and environmental factors (6). However, the etiology and pathogenesis are still not completely determined. Recent studies examining global trends in GC incidence and mortality have confirmed an ongoing decline worldwide, given dietary regulation, awareness of drug use, and a decline in Helicobacter pylori prevalence (3, 7).

There are currently no screening techniques for GC, and because patients with early gastric cancer are frequently asymptomatic, the rate of early diagnosis of gastric cancer is low. As a result, the majority of patients (>70%) are diagnosed with advanced gastric cancer at the time of diagnosis (8). There is a need for different markers that can diagnose GC, which is usually diagnosed in late stages, is highly metastasized in these stages, and has very low survival rates in advanced stages, although early stage survival rates are quite high, and thus increase the chances of treatment.

With the development and widespread use of next-generation sequencing (NGS) technologies, genomic analyses to determine the causes of cancer development have recently revealed the relationships between various malignant tumors and genomic information and have enabled the identification of new molecular markers and intracellular pathways related to diseases. In the light of these developments, these technologies have been utilized extensively to reveal the full structure of GC genomics (9). Machine learning (ML) is a subfield of artificial intelligence that aims to make predictions about new observations by learning based on existing data, unlike traditional statistical techniques. ML, which has a wide range of applications in health, forms the basic infrastructure for the detection of genetic diseases, early diagnosis of cancer diseases, and identification of patterns in medical imaging. In the last decade, with the availability of large datasets and higher computing power, ML methods have achieved high performances in a wide range of situations (10, 11). However, a major problem with many state-of-the-art ML models is the



Zeynep Kuukakcali and Ipek Balikci Cicek, "Screening of Genes That May be Associated with Gastric Cancer Using Bioinformatics and Artificial Intelligence Methods and Interpretation of Individual-Based Results," *International Research Journal of Pharmacy and Medical Sciences (IRJPMS)*, Volume 6, Issue 6, pp. 6-12, 2023.

lack of transparency, interpretability as well as explainability. To overcome these shortcomings, Explainable artificial intelligence (XAI) has recently gained more attention in clinical research. In this context, XAI addresses methods that aim to make ML models more understandable/interpretable by clinicians (12). One of the XAI methods, Local Interpretable Model-Agnostic Explanations (LIME), is a popular technique for explaining the predictions of black box machine learning models (13). Since LIME is designed to be model-independent, it can be applied to many different ML models. The model created by the method makes the results of the model more interpretable by determining which features in the data are more important on a patient-by-patient basis (14).

The aim of this study is to identify potential genes that may be associated with gastric cancer by bioinformatics methods using open-access gene expression data obtained from human gastric tumor tissues and normal gastric tissues. Our second aim is to classify the disease with Random Forest (RF), which is one of the ML models, using the data set to be obtained by determining the genes that show different regulation in diseased tissues compared to the normal group. Finally, using the LIME method over the RF model created, we aim to better understand the decision-making process of the classification model and make personal inferences based on the values of individuals. With these methods, it will be possible to determine at which threshold values the gene levels of individuals contribute to the disease and the disease can be evaluated based on individual-based results.

## II. MATERIAL AND METHODS

### *Dataset*

The open-access dataset used in the study was obtained from the National Center for Biotechnology Information (NCBI). The dataset consists of tissue samples from 15 pairs of gastric tumors and adjacent non-tumor (normal) tissue. When creating the dataset, care was taken to ensure that the individuals were diagnosed with gastric adenocarcinoma and had no history of other tumors. Detailed information on how gene expression profiles were obtained from tissues can be found in the related study (15).

### *Bioinformatics, gene expression*

The gathering, storage, organization, archiving, analysis, and presentation of results based on theory and practice in a subject such as biology, medical, behavioral, or health sciences is referred to as bioinformatics. Furthermore, it focuses on the research and development of computational tools and techniques to expand the use and processing of data generated by studies or the implementation of recognized processes. Obtained as a result of research or the use of well-known procedures. Bioinformatics analyses are carried out by selecting a database and a tool that permits bioinformatic analysis to be carried out in accordance with the biological question, molecule, or structure to be investigated. The results of the investigations are pooled and assessed analytically in light of previously existing information about the issue in the literatüre (16).

Changes in the physiology of an organism or cell will be accompanied by changes in the pattern of gene expression, making gene expression analysis significant in many disciplines of biological inquiry. The DNA microarray approach, which is still in development, is used to examine gene expression by hybridizing mRNA to a high-density array of immobilized target sequences, each corresponding to a different gene. Chemicals' effects on gene expression, for example, can disclose functional and toxicological properties. Expression investigations on clinical samples from both healthy and ill people may lead to the discovery of new biomarkers (17).

### *Bioinformatics analysis phase*

Gene expression analyses were done with data collected from gastric tumors and adjacent non-tumor (normal) tissue in this investigation. In the investigation, the limma (Linear Models for Microarray Analysis) package, which is accessible in the R programming language and permits expression analysis, was employed (18). Limma is a library for analyzing gene expression microarray data, with an emphasis on using linear models to analyze specific experiments and determine differential expression. The features of the packet are applicable to all gene expression methodologies, including microarrays, RNA-seq, and quantitative PCR. Thanks to the Empirical Bayes methods in the Limma package, it is also possible to obtain stable results even when the number of sequences is small. As a result of bioinformatics analysis, adjusted p-value, p-value, and Lof2FC which shows the fold change of expression differences of genes, was obtained. Genes that differed in tumor tissues compared to normal tissues were determined as up-regulated if log2FC values were greater than 1 and $p<0.05$, and down-regulated if log2FC values were less than -1 and $p<0.05$. The distribution of the data utilized in the study was shown using box-plot graphs and expression density graphs. Samples with the same qualities are represented in the graphs with the same color. The Uniform Manifold Approximation and Projection (UMAP) graph was chosen to depict the relationships between the samples in the research. Finally, the volcano plot was used to display differentially expressed genes (up and down). To rapidly identify differentially expressed genes, the volcano graph presents significance versus fold-change in log2 on the y- and x-axes. The red color in the graph represents up-regulated genes, the blue color represents down-regulated genes, and the black color represents genes that do not vary.

### *XAI, ML*

Machine learning is a collection of techniques that can predict future events or classify data by extracting patterns from previous data. In a wide variety of industries, including medical sciences, machine learning approaches have had significant success in the analysis of datasets, with many predictive models (19). Most ML models focus on accuracy in the estimation process and rarely tend to explain this result. This is the black box feature of ML methods (20). Understanding, explaining, and interpreting the results of ML approaches is very important. ML models that can offer ideas



Zeynep Kuukakcali and Ipek Balikci Cicek, “Screening of Genes That May be Associated with Gastric Cancer Using Bioinformatics and Artificial Intelligence Methods and Interpretation of Individual-Based Results,” *International Research Journal of Pharmacy and Medical Sciences (IRJPMS)*, Volume 6, Issue 6, pp. 6-12, 2023.

about why and why certain predictions are produced are called explicable artificial intelligence models. Interpretation of model predictions in the use of ML models, explaining why and how results are obtained will be a priority for clinical practitioners when it comes to application and use (21). For this reason, the importance of applying XAI to recent studies has increased and in addition to modeling, it has gained importance to explain the determinant factors in the explanation of the results. The LIME method, one of the XAI methods included in this study, is within the scope of the local interpretability model, and local interpretability is the case of trying to understand why the model makes its decision by examining a single prediction of a model locally.

*Random Forest*

The Random Forest technique is a classification and regression approach that includes voting. It is made up of multiple decision trees that are combined, and the individual trees are voted on to decide the winner class. The decision trees in the forest are independent of one another and are generated using the bootstrap process using samples selected from the data set (22). The random forest approach employs a large number of classification trees. Each piece of incoming data is processed by all of these classification trees. Classification trees are used to classify each piece of the incoming data. After each input data set is entered into all classification trees and voted on, data is assigned to the class with the most votes from the tree structures (23).

*Modelling phase*

Random Forest one of the machine-learning approaches was employed in the modeling. Before modeling, the data set to be used in the modeling was created by using the Lasso variable selection method, which is one of the variable selection methods from the genes determined as up and down, and based on the sum of the absolute values of the model parameters being less than a fixed value (upper limit). The n-fold cross-validation approach was used in the analyses. The data is separated into n parts in the n-fold cross-validation procedure, and the model is applied to n parts. One of the n-parts is utilized for testing, while the remaining n-1 parts are used to train the model. For the cross-validation procedure, the mean of the acquired values is assessed. The modeling approach in this study was carried out using 5-fold cross-validation. As performance assessment criteria, accuracy, balanced accuracy, sensitivity, selectivity, positive predictive value, negative predictive value, and F1-score were employed. The LIME method, which is one of the XAI approaches, was used to interpret the model results.

## III. RESULTS

Distribution graphs for 15 pairs of gastric tumors and adjacent non-tumor (normal) tissue used in the study are given in Figure 1 and Figure 2.

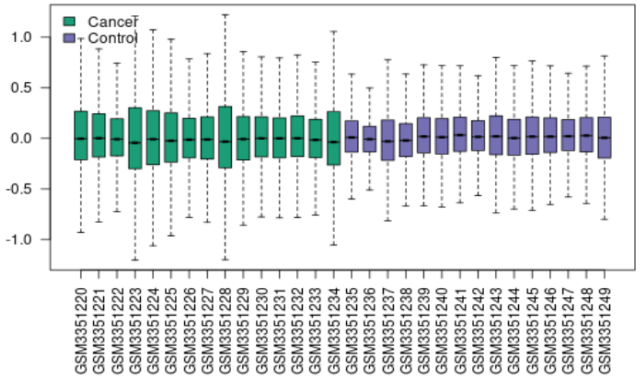



Figure 1: The distribution of the values of the selected samples.

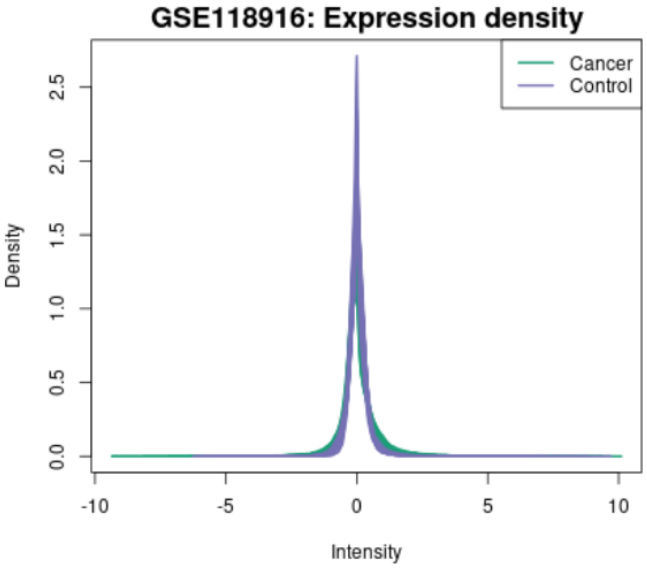



Figure 2: The expression density graph of the selected samples.

The UMAP graph, where we can see the relationships of the samples with each other, is given in figure 3. With this graph, it is seen that the samples with the same characteristics are clustered together. In the graph, green dots show pancreatic ductal adenocarcinoma samples, while purple dots show normal tissues.

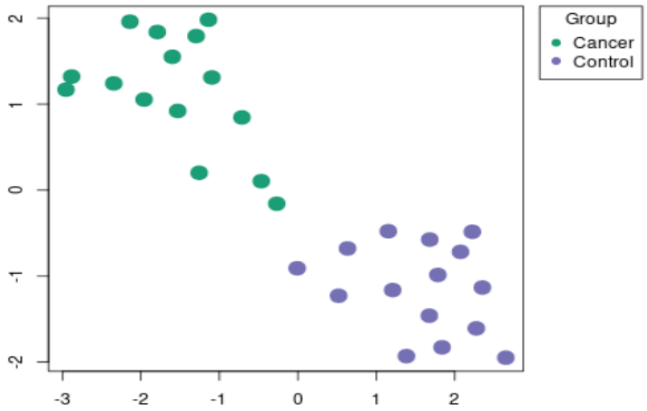



Figure 3: UMAP plot of the samples



Zeynep Kuukakcali and Ipek Balikci Cicek, "Screening of Genes That May be Associated with Gastric Cancer Using Bioinformatics and Artificial Intelligence Methods and Interpretation of Individual-Based Results," *International Research Journal of Pharmacy and Medical Sciences (IRJPMS)*, Volume 6, Issue 6, pp. 6-12, 2023.

The data set in the study contained 49396 transcripts and according to the gene expression analysis, 3283 expressions were found with statistically significant differences in gene expression levels between the two groups (|log2FC| > 1.0, p<0.05). Information on the top 10 genes showing up- and down-regulation in expression between the two groups is given in Table 1 and Table 2.

TABLE 1: Genes whose expression level is up-regulated in gastrict adenocarcinoma samples relative to normal tissue

| ID | Adj.P Val | P Value | t | Log2FC | Gene Symbol |
|---|---|---|---|---|---|
| 11746506_a_at | 4,72E-05 | 3,54E-06 | 5,65 | 3,54964758 | SPP1 |
| 11742711_at | 4,23E-05 | 3,05E-06 | 5,7 | 3,40092154 | THBS2 |
| 11731475_a_at | 6,04E-05 | 4,88E-06 | 5,54 | 3,37497786 | SFRP4 |
| 11723174_a_at | 1,37E-04 | 1,42E-05 | 5,16 | 3,33741112 | FNDC1 |
| 11726339_s_at | 2,68E-03 | 6,11E-04 | 3,82 | 3,23578135 | MAGEA3 /// MAGEA6 |
| 11721212_a_at | 3,09E-04 | 4,03E-05 | 4,79 | 3,23195156 | THBS4 |
| 11757941_s_at | 3,98E-05 | 2,80E-06 | 5,73 | 3,09180619 | THBS2 |
| 11749919_a_at | 8,63E-07 | 1,80E-08 | 7,56 | 3,02620436 | GUCY1A3 |
| 11755955_a_at | 1,13E-04 | 1,11E-05 | 5,25 | 2,79062237 | FAP |
| 11740290_a_at | 1,30E-06 | 3,18E-08 | 7,35 | 2,76385628 | HOXC6 |

TABLE 2: Genes whose expression level is down-regulated in gastrict adenocarcinoma samples relative to normal tissue

| ID | Adj.P Val | P Value | t | Log2FC | Gene Symbol |
|---|---|---|---|---|---|
| 11748336_a_at | 1,17E-06 | 2,75E-08 | -7,4 | -5,79202286 | GIF |
| 11733660_a_at | 3,01E-06 | 9,52E-08 | -6,95 | -5,64793899 | CHIA |
| 11729079_s_at | 4,30E-09 | 1,04E-11 | -1,06E+01 | -5,53280808 | ESRRG |
| 11728308_at | 2,77E-10 | 2,07E-13 | -1,24E+01 | -5,47904038 | KRT20 |
| 11732742_at | 1,30E-05 | 6,63E-07 | -6,24 | -5,44095774 | ATP4B |
| 11756545_a_at | 8,45E-07 | 1,75E-08 | -7,57 | -5,2803817 | GKN2 |
| 11723302_a_at | 6,09E-09 | 1,69E-11 | -1,04E+01 | -5,26753987 | CHGA |
| 11734596_a_at | 4,56E-06 | 1,66E-07 | -6,74 | -5,16872816 | ATP4A |
| 11744246_at | 2,64E-09 | 5,13E-12 | -1,09E+01 | -5,07184373 | KCNE2 |
| 11715481_a_at | 2,89E-09 | 6,02E-12 | -1,08E+01 | -5,02258642 | SST |

Figure 4 depicts the volcano plot used to visualize the differentially expressed genes between groups.

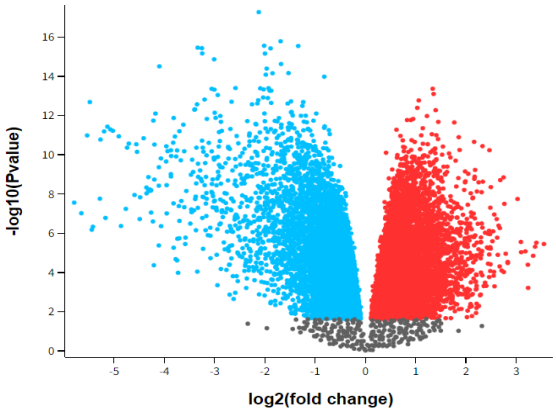



Figure 4: Volcano plot of genes with up-regulated (red dots), down-regulated (blue dots) gene expression, and not regulated (black dots) among the group of gastric adenocarcinoma samples and normal tissue

The findings of the performance metrics from the RF model are given in Table 3.

TABLE 3: Performance metrics of the RF model

| Metric | Value (%) |
|---|---|
| Accuracy | 96.7 |
| Balanced Accuracy | 96.7 |
| Sensitivity | 93.3 |
| Specificity | 100 |
| Positive predictive value | 100 |
| Negative predictive value | 93.8 |
| F1 score | 96.6 |

*LIME results*

Considering the LIME results, with 100% probability for patient 2 gastric tumors were presumed. The "11721940_a_at > 0.96, 11715481_a_at > 1.22, 11736909_at > 0.76, 11724692_a_at > 0.6, -0.31 < 11731212_x_at <= 0.70, 1 1748972_a_at <= -0.44, 11716288_s_at <= -0.22, -0.08 < 11721421_s_at <= 1.06, 11717057_x_at <= -0.43, 11715222_at > 0.77" values positively affected the probability of predicting this patient as a gastric tumor with a 100% probability. In addition, with 98% probability for patient 6 non-tumor were presumed. The "11715222_at <= -0.84, 11717370_at <= -0.58, 11715373_a_at > 0.46, 11721941_x_at <= -1.86, 0.07 < 11719964_a_at <= 0.56, 11724692_a_at <= -0.58, 11762251_at <= -0.36, 0.17 < 11717057_x_at <= 0.77, 0.16 < 11715580_a_at <= 0.53, -0.74 < 11731212_x_at <= -0.31" values negatively affected the probability of predicting this patient as non-tumor with a 98% probability. Results for other patients will be interpreted similarly.

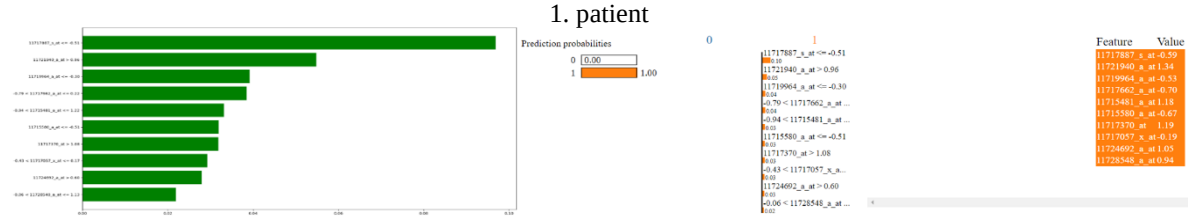
## IV. Discussion

Gastric cancers, a common global disease, is the fifth most common cancer after lung, breast, colorectal and prostate cancers and constitutes a significant health burden worldwide (24). More than 1 million people worldwide develop stomach cancer each year, and despite a decline in incidence and mortality worldwide over the last 50 years, stomach cancer remains the third leading cause of cancer-related deaths (25). GAC is the most common histological type with an incidence of 90-95% and despite decades of advances in diagnostic and therapeutic modalities, the mortality rate of GAC remains

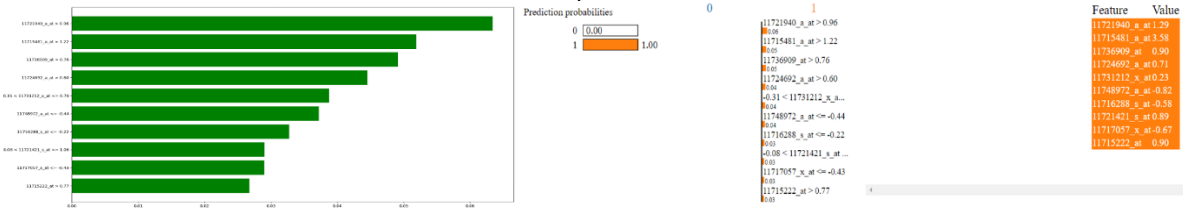
International Research Journal of Pharmacy and Medical Sciences
ISSN (Online): 2581-3277

high. In addition, the global 5-year survival rate of patients is not satisfactory (15). It is well known that cancers are often caused by abnormal cell cycle activity. These activities are usually caused by genetic lesions in genes encoding proteins involved in the cell cycle or by mutations in upstream or downstream signalling pathways (26).

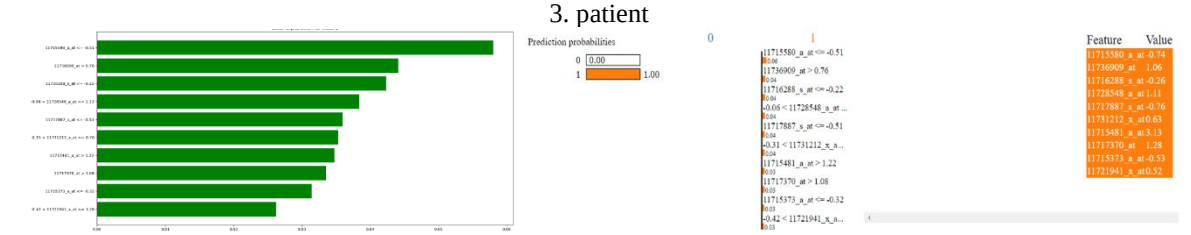
1. patient



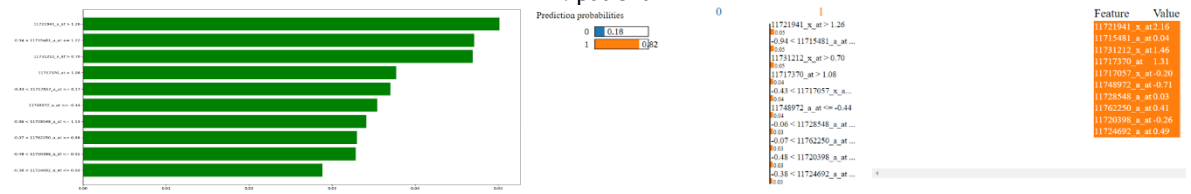

2. patient



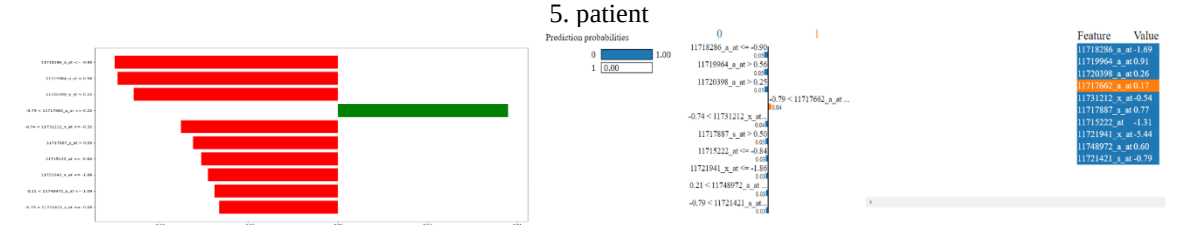

3. patient



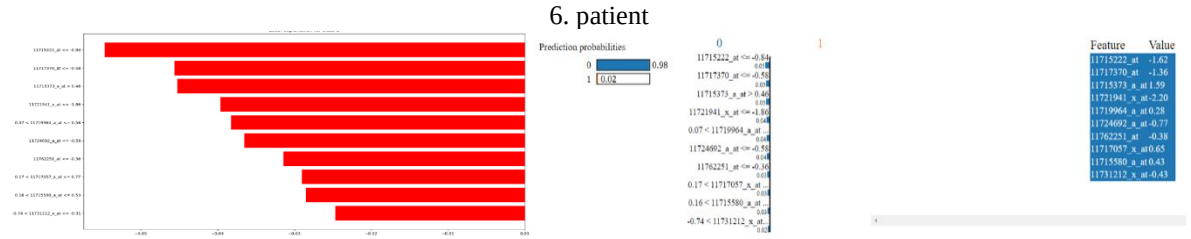

4. patient




5. patient




6. patient




Identifying and correlating important genes and key pathways involved in the initiation and progression of GAC, one of the most lethal malignancies in the world, is vital for the management of the disease and the formulation of treatment strategies. In order to elucidate the molecular structure of GAC and to identify therapeutic targets for the disease, we used an open-access dataset consisting of samples from 15 pairs of gastric tumors and adjacent non-tumor (normal) tissue. Firstly, bioinformatic analyses were performed with the gene expression data sets of the samples. As a result of the bioinformatic analyses, the results of modelling at the individual level using LIME, one of the XAI methods, were discussed by modelling with genes that meet the conditions of up and down regulation.



Zeynep Kuukakcali and Ipek Balikci Cicek, “Screening of Genes That May be Associated with Gastric Cancer Using Bioinformatics and Artificial Intelligence Methods and Interpretation of Individual-Based Results,” *International Research Journal of Pharmacy and Medical Sciences (IRJPMS)*, Volume 6, Issue 6, pp. 6-12, 2023.

When the results of bioinformatic analyses were examined, it was determined that 3283 genes showed different regulation (up or down) in gastric tumors compared to normal tissues. SPP1 gene showed 11.63 fold up-regulation in gastric tumor samples compared to normal tissue samples. Likewise, THBS2, SFRP4, FNDC1, MAGEA3 /// MAGEA6, THBS4, THBS2, GUCY1A3, FAP, and HOXC6 genes had up-regulated gene expression of 10.55, 10.33, 10.05, 9.44, 9.38, 8.51, 8.11, 6.91, and 6.77 fold, respectively. GIF gene showed 55.33 fold down-regulation in gastric tumor samples compared to normal tissue samples. Likewise, CHIA, ESRRG, KRT20, ATP4B, GKN2, CHGA, ATP4A, KCNE2, and SST genes had down-regulated gene expression of 49.86, 46.20, 44.32, 43.41, 38.85, 38.31, 35.75, 33.59, and 32.44 fold, respectively.

The RF model was performed with 30 genes selected by Lasso variable selection method from the genes determined as up and down. The performance metrics obtained from the model were accuracy (96.7%), balanced accuracy (96.7%), sensitivity (93.3%), specificity (100%), positive predictive value (100%), negative predictive value (93.8%), and F1-score (96.6%), respectively.

When the LIME findings are examined, the genes that are positively or negatively associated with gastric tumor are especially 11717057_x_at (CORO1C), 11724692_a_at (CAPN13), 11715481_a_at (SST), 11724692_a_at (CAPN13), 11731212_x_at (GGT6), 11717370_at (ARSD), 11721941_x_at (CYP3A5). In one study, it was shown that the CORO1C gene was expressed at higher levels in gastric cancer tissues and this was associated with poorer survival (27). In another study, CORO1C gene showed high expression in gastric cancer tissues (28). In a study, genetic variability in CYP3A4 and CYP3A5 in primary liver, gastric and colorectal cancer patients was analysed and disease-associated variants were identified (29). In addition, CYP3A5 gene and its variants have been analysed for many diseases and cancer types in many studies (30-32).

According to the results of bioinformatic analyses and modelling, genes that may be associated with the disease were identified. Bioinformatic analyses identified genes showing up or down expression in tumor status. These genes were considered as priority biomarkers for the disease and modelling was performed. With the modelling, tumor and non-tumor tissues were separated with a very high accuracy value and the disease was classified. The LIME method was applied to make the machine learning model more understandable and to determine the critical values in gene expression levels in individuals with and without disease and to reveal individual-based results. The genes obtained by examining the results obtained with LIME on an individual basis are the genes that are present in all individuals and contribute to whether they are sick or not.

In addition, LIME, which is the XAI method applied, allows to analyse the disease on an individual basis by revealing the genes associated with each patient's condition and the critical values at the gene levels of the genes. The results obtained may also be effective for the "personalised medicine" movement, which is rapidly developing today and based on the idea of treating patients on an individual basis. Because the results of the method are based on the changes in individuals. With the comprehensive research and studies to be carried out in the future regarding these genes, perhaps targeted therapies can be developed and new treatment strategies can be added to the treatment strategies of the disease.

## REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA: a cancer journal for clinicians. 2019;69(1):7-34.
3. Luo G, Zhang Y, Guo P, Wang L, Huang Y, Li K. Global patterns and trends in stomach cancer incidence: age, period and birth cohort analysis. International journal of cancer. 2017;141(7):1333-44.
4. Eusebi LH, Telese A, Marasco G, Bazzoli F, Zagari RM. Gastric cancer prevention strategies: A global perspective. Journal of gastroenterology and hepatology. 2020;35(9):1495-502.
5. Ajani JA, Lee J, Sano T, Janjigian YY, Fan D, Song S. Gastric adenocarcinoma. Nature reviews Disease primers. 2017;3:17036.
6. Fuccio L, Eusebi LH, Bazzoli F. Gastric cancer, Helicobacter pylori infection and other risk factors. World journal of gastrointestinal oncology. 2010;2(9):342-7.
7. Arnold M, Karim-Kos HE, Coebergh JW, Byrnes G, Antilla A, Ferlay J, et al. Recent trends in incidence of five common cancers in 26 European countries since 1988: Analysis of the European Cancer Observatory. European journal of cancer. 2015;51(9):1164-87.
8. Digklia A, Wagner AD. Advanced gastric cancer: Current treatment landscape and future perspectives. World journal of gastroenterology. 2016;22(8):2403-14.
9. Onoyama T, Ishikawa S, Isomoto H. Gastric cancer and genomics: review of literature. Journal of Gastroenterology. 2022;57(8):505-16.
10. Polikar R. Ensemble learning. Ensemble machine learning: Springer, Boston, MA; 2012. p. 1-34.
11. Akman M, Genç Y, Ankarali H. [Random Forests Methods and an Application in Health Science]. Turkiye Klinikleri J Biostat. 2011;3(1):36.
12. Došilović FK, Brčić M, Hlupić N, editors. Explainable artificial intelligence: A survey. 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO); 2018: IEEE.
13. Garreau D, Luxburg U, editors. Explaining the explainer: A first theoretical analysis of LIME. International Conference on Artificial Intelligence and Statistics; 2020: PMLR.
14. Hu L, Chen J, Nair VN, Sudjianto A. Locally interpretable models and effects based on supervised partitioning (LIME-SUP). arXiv preprint arXiv:180600663. 2018.
15. Li L, Zhu Z, Zhao Y, Zhang Q, Wu X, Miao B, et al. FN1, SPARC, and SERPINE1 are highly expressed and significantly related to a poor prognosis of gastric adenocarcinoma revealed by microarray and bioinformatics. Scientific reports. 2019;9(1):7827.
16. Akalin PK. Introduction to bioinformatics. Molecular nutrition & food research. 2006;50(7):610-9.
17. van Hal NL, Vorst O, van Houwelingen AM, Kok EJ, Peijnenburg A, Aharoni A, et al. The application of DNA microarrays in gene expression analysis. Journal of biotechnology. 2000;78(3):271-80.
18. Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and Bioconductor: Springer; 2005. p. 397-420.
19. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. XAI—Explainable Artificial Intelligence. Sci Robot. 2019;4(37):eaay7120.
20. Samek W, Müller K-R. Towards explainable artificial intelligence. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. 11700: Springer; 2019. p. 5-22.
21. Tjoa E, Guan C. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. IEEE Trans Neural Netw Learn Syst. 2020;32(11):4793-813.

22. Breiman L. Random forests. Machine Learning. 2001;45(1):5-32.
23. Cutler A, Cutler DR, Stevens JR. Random Forests. In: Cha Z, Yunqian M, editors. Ensemble Machine Learning. 1st ed: Springer New York; 2012. p. 157-75.
24. Tarver T. American cancer society. cancer facts and figures 2014. J Consumer Health Internet. 2012;16:366-7.
25. Thrift AP, El-Serag HB. Burden of Gastric Cancer. Clinical Gastroenterology and Hepatology. 2020;18(3):534-42.
26. Sherr CJ, Bartek J. Cell cycle–targeted cancer therapies. Annual Review of Cancer Biology. 2017;1:41-57.
27. Cheng X, Wang X, Wu Z, Tan S, Zhu T, Ding K. CORO1C expression is associated with poor survival rates in gastric cancer and promotes metastasis in vitro. FEBS Open Bio. 2019;9(6):1097-108.
28. Vidal AF, Ribeiro-dos-Santos AM, Vinasco-Sandoval T, Magalhães L, Pinto P, Anaissi AK, et al. The comprehensive expression analysis of circular RNAs in gastric cancer and its association with field cancerization. Scientific reports. 2017;7(1):14551.
29. Gervasini G, García-Martín E, Ladero JM, Pizarro R, Sastre J, Martínez C, et al. Genetic variability in CYP3A4 and CYP3A5in primary liver, gastric and colorectal cancer patients. BMC Cancer. 2007;7(1):118.
30. Hustert E, Haberl M, Burk O, Wolbold R, He Y-Q, Klein K, et al. The genetic determinants of the CYP3A5 polymorphism. Pharmacogenetics and Genomics. 2001;11(9):773-9.
31. Wang B-S, Liu Z, Xu W-X, Sun S-L. CYP3A5* 3 polymorphism and cancer risk: a meta-analysis and meta-regression. Tumor Biology. 2013;34:2357-66.
32. Meulendijks D, Rozeman E, Cats A, Sikorska K, Joerger M, Deenen M, et al. Pharmacogenetic variants associated with outcome in patients with advanced gastric cancer treated with fluoropyrimidine and platinum-based triplet combinations: a pooled analysis of three prospective studies. The pharmacogenomics journal. 2017;17(5):441-51.