Mini-Review : THE CLASSIFICATION STUDIES DONE FOR EARLY DIAGNOSIS OF THE STOMACH CANCER

Furkan ESMERAY*, İbrahim Hanifi ÖZERCAN

*Electric and Energy Department, Munzur University, Tunceli, Turkey

Pathology Department, Fırat University, Elazığ, Turkey

furkanesmeray@munzur.edu.tr ozercanih@yahoo.com

Abstract: Cancer is a heterogeneous disease. Cancer is a heterogeneous disease Cancer is composed of many different subtypes. The early diagnosis and diagnosis of gastric cancer has become imperative in cancer research because it may facilitate the subsequent clinical treatment of patients. Separation of gastric cancer patients into normal, low and high groups has become important in bioinformatics and biomedical fields. This has led to an increase in the practice of machine learning (ML) methods for early cancer diagnosis in the literature. Machine learning methods have been used to model the progression and treatment of stomach cancer. ML methods have been used to detect complex cell characteristics in cancer images. (ANN), Bayesian Networks (BNS), Random Forest (RF), Support Vector Machines (SVM), Decision Trees (DT), Linear Discriminant Analysis (LDA), Sammon mapping, Stochastic Neighbor New algorithms have been proposed using various machine learning techniques such as Embedding (SNE), Isomap, Classical multidimensional scaling (MDS), and Local Linear Embedding (LLE). Although machine learning techniques for gastric cancer have been widely applied and ultimately yielded high classification performances, an appropriate level of validation is required to take these methods into account in daily clinical treatment and practice. In this study, the methods used in algorithms for early diagnosis of stomach cancer and the classification ratios are described. In the advanced algorithms, various different features and image data are used. As a result, in this article, ML methods for gastric cancer research are increasing. For this reason, published articles have been presented to model the risk of stomach cancer.

Keywords: Stomach Cancer, Machine Learning Techniques, Classification

Introduction

Stomach cancer is a rapidly developing and spreading type of cancer (Tannapfel, 2001). It usually starts in the form of ulcer complaints. Stomach cancer can affect the organs and lymph glands (Tolbert, 1999). It can be spread by direct neighbors, lymphatic, blood and intraabdominal (Siddiqi, 2008). The stomach tumor may grow through the outer layer of the miter and extend into the surrounding organs, such as the pancreas, esophagus or intestine (Mikami, 2004). Stomach cancer cells can spread by blood metastases to the liver, lungs, and other organs (Ahmad, 2007). Cancer cells can also spread to all the lymph glands in the body through the lymphatic system (Fenoglio, 2000). Cancer first starts at the center of the cells that make up the tissues, and at the same time, these tissues bring the organs of the body to the bloom (Cappell, 2002). For a normal individual, the cells grow, develop, and the body is divided to create cells that are new in need. Cells are sufficiently digested and digested by the cell when they are old and are killed and replaced by new ones. But sometimes this process is in the wrong direction. So much so that, despite the fact that the body does not need much, new cells form and the old cells that should die do not die. These excess cells can eventually turn into tumor masses called tumors. Tumors may be benign or malignant.Stomach cancer that occurs from stomach wall and stomach tissue is a type of cancer. According to studies conducted by the Ministry of Health in our country, stomach cancer was identified as the second most common type of cancer. Endoscopy is the most important factor in the early diagnosis of this disease. Endoscopic examination of the midge and biopsy specimens are used to diagnose pathological examinations. It is observed that half of the people who have this disease are late in the diagnosis and doctors can not apply any treatment (Lambert, 2002). The most common sites of this disease in the world are far-east countries such as Japan and China. In Japan, people who get stomach cancer account for about 30% of other cancer diseases. In the United States, the number of people with stomach cancer is increasing every year (Internet, 2015), (Internet: Cancer Facts and Statictics, 2015). According to research conducted worldwide, 26% of males and 11% of females had gastric cancer. Stomach cancer is located in the third after the lungs and breast cancer in women and second place after the lung cancer in men. According to the statistics done in our country, it is estimated that the number of new gastric cancer patients is about 30 thousand per year (Fujioka, 2004). Stomach cancer is a type of cancer which spreads rapidly and also leads to death when the patient is diagnosed late (Tannapfel, 2001). Stomach cancer usually begins with ulcer and gastritis complaints. Cancer can affect lymph nodes and other peripheral organs (Tolbert, 1999). Machine learning techniques are widely used in computer assisted analysis of histopathological stomach cancer images. In the literature, machine learning techniques such as Bayesian Networks (BN), Decision

Trees (DT), Artificial Neural Networks (YSN), and Support Vector Machines (SVM) have been extensively applied in cancer surveys for the development of predictive models with effective outcomes (Kourou, 2015) An important cause of gastric cancer is H. pylori infection. There are many studies on machine learning techniques on other types of cancer besides stomach cancer in the literature (Korkmaz, 2015), (Korkmaz and Poyraz, 2014), (Korkmaz, S. A., Korkmaz, M. F., and Poyraz, 2016), (Korkmaz, S. A and Korkmaz, M. F. 2015), (Korkmaz, S. A. and Eren, H. 2013), (Korkmaz, 2016), (Korkmaz, 2015), (Korkmaz, 2018). Machine learning techniques have also been used to classify more diverse data analyzes (KORKMAZ, 2017), (Korkmaz and Poyraz 2016). But, we will examine computer-aided studies of stomach cancer in this article. In Literature, there are computer assisted studies in machine learning techniques in stomach cancer diagnosis.

Related Work

In this article, 13 studies in the literature on stomach cancer were reviewed. The first study (Korkmaz and Binol, 2017) proposed a computer-aided study to help diagnose stomach cancer early and reduce mortality rates. For this purpose, stomach cancer microscope images were taken from Pathology laboratory of Fırat University Medical Faculty. These images feature the Local Binary Patterns (LBP) and Histogram of Oriented Gradients (HOG). The dimensions of these computed features are Sammon mapping, Stochastic Neighbor Embedding (SNE), Isomap, Classical multidimensional scaling (MDS), Local Linear Embedding (LLE), Linear Discriminant Analysis (LDA), t- Distributed Stochastic Neighbor Embedding and Laplacian Eigenmaps are reduced by size reduction techniques. Artificial neural networks (ANN) and random forest (RF) classifiers were applied to these features which were reduced in size. It has been suggested that the best classification performances are found with the LBP_MDS_ANN and LBP_LLE_ANN methods.

In the second study (Korkmaz, S. A., Bínol, H., Akçiçek, A., and Korkmaz, M. F. 2017), a study for the early diagnosis of gastric cancer was proposed. For this, a total of 180 stomach image cells, normal, benign and malignant, were obtained. Of the 60 normal, 60 benign and 60 malignant stomach images, 90 were used for educational purposes and 90 were used for educational purposes. The histograms of oriented gradients (HOG) properties of these images were calculated. High-dimensional HOG features were reduced to lower dimensions by Linear Discriminant Analysis (LDA). After applying the LDA method, Artificial Neural Network (ANN) classifier was used. The images were found as 88.9% with ANN classifier. This result has been compared with some studies in the literature. It has been stated that the result obtained provides higher performance. It has also been suggested that this result in a shorter time.

In the third study (Korkmaz, S. A., Akçiçek, A., Bínol, H., and Korkmaz, M. F., 2017), histograms of oriented gradients (HOG) were used for early detection of gastric cancer. These HOG properties are found by the curve of the HOG properties. Bins and h histogram values were subtracted from the histogram plot. Bandwidth ranges were obtained by using h and bins values of normal, benign and malign images. The h values of normal and benign stomach images obtained were higher than malign gastric images. In one, it was suggested that the h values of the normal stomach image are higher than the benign stomach image.

In the fourth study (Korkmaz, 2018), Local Binary Pattern (LBP) features of images for early detection of gastric cancer were found. These properties were reduced to lower dimensions by the Locality Preserving Projections (LPP) method. The property values obtained after this process are classified by Random Forest (RF), Naive Bayes (NB), and Artificial Neural Networks (ANN) classifiers. It is stated that the highest classification performance is 96.29% with ANN classifier and the lowest classification performance is 70% with NB classifier. Moreover, it is stated that the feature size used when finding the highest performance with ANN is lower than the feature size used with other classifiers.

In the fifth study (Korkmaz, S. A. and Esmeray, F. 2018), Maximally Stable Extremal Regions (MSER) properties of stomach images for early diagnosis of gastric cancer were calculated. Discrete Fourier Transform (DFT), Local Tangent Space Alignment (LTSA) and Neighborhood Preserving Embedding methods have been applied to reduce the size to the calculated high dimensional MSER properties. The Random Forest (RF) classifier was applied to these low dimensional feature values. It has been stated that a higher classification result is obtained compared to other classifiers in the literature.

In the Sixth Study (Sasaki, 2010), S. Yoshihiro and colleagues studied a computerized system to predict risk factors for stomach cancer. Digital endoscopy images of patients with H. pylori bacteremia were studied in the system. Three parameters have been used to classify the gastric mucosa.The data obtained from this class is processed by

Bayes theorem and output is obtained. This study sheds light on the identification of patients at high risk for endoscopies.

In the seventh study (Ahmadzadeh, 2013), D. Ahmadzadeh and colleagues developed a stomach cancer diagnosis system using the local pattern algorithm and the DSM (Decision Supporting Machine) method. In the advanced system, a system is obtained that corrects early cancer after noise reduction, feature extraction, feature identification, classification steps.55 volunteer patients were randomly selected. The diagnostic accuracy rate was 91.8%. The authors suggested that it is a system that helps experts save time and money.

In the eighth study (Garcia, 2017), Garcia and colleagues used Deep Convolutional Neural Networks to suggest an approach for automated tumor-infiltrating lymphocytes on immunohistochemical images of gastric cancer. The accuracy rate of this study was 96.88%.

In the ninth study (Zhang, 2017), authors used a concurrent model called the Gastric Precancerous Disease Network (GPDNet) to distinguish between erosion, ulcer and polyp classes using convolutional neural networks (CNN). The classification rate with GPDNet was 88.90%.

The tenth article (Shichijo, 2017), established the convulsion nerve network (CNN) and assessed its ability to diagnose Helicobacter pylori gastritis infection. The first deep CNN has 22 layers. A total of 32,208 images, positive or negative, were used for the education of H. Pylori. Another CNN was trained using images classified according to 8 anatomic regions. For the test, 11,481 histopathological stomach images were used. Sensitivity, specificity, accuracy and diagnostic time for the first CNN were 81.9%, 83.4%, 83.1% and 198 s respectively. For secondary CNN, 88.9%, 87.4%, 87.7% and 194 s respectively are obtained. A higher accuracy result was obtained with the secondary CNN. For 23 endoscopists, these values were 79.0%, 83.2%, 82.4% and 230 ± 65 minutes respectively.According to endoscopists, H. pylori gastritis can be diagnosed more precisely and in a shorter time using CNN.

In the eleventh study (Bollschweiler, 2004), artificial neural networks (ANN) were used to estimate 4302 patients with lymph node metastases with gastric cancer. And, a marine computer program (MCP) was developed. ANN was compared with MCP. The authors suggest that the predictive value for lymph node metastasis varies in each of the lymph node metastases. This variation is between 42% and 70% for MCP and between 64% and 64% for ANN. The classification rate of YSA was found to be 93%.

In the twelfth study (Cosatto, 2013), haematoxylin and eosin stained gastric tissue samples were examined. And in these cases, a machine-learning based computer system that can detect cancer is recommended. In the computer system, high-level medically relevant features such as nuclei and glands are used. The Multilevel Learning Framework (MIL) classifier is used to classify the data set. The system has been trained and tested with a large-scale dataset of 26K textures. 1'196 positive tissue and 8'558 patients from 16'692 negative tissue. A test data consisting of 4,168 patients was obtained from 582 positive tissues and 8125 negative tissues. The accuracy performance of the system was 96% with AUC.

In the thirteenth study Akbari and colleagues (Akbari, 2011) developed a stomach cancer diagnosis system using infrared ultra-spectral imaging. This study was conducted by selecting patients with gastric cancer. The spectral features were extracted from cancerous and normal tissues, and the comparison was made and the detection of the cancerous regions was performed by using the KDM method with the spectral diagram.

Conclusions

Stomach cancer is a rapidly developing and spreading type of cancer (Tannapfel, 2001). It usually starts in the form of ulcer complaints. Stomach cancer can affect the organs and lymph glands (Tolbert, 1999). It can be spread by direct neighbors, lymphatic, blood and intraabdominal (Siddiqi, 2008). The stomach tumor may grow through the outer layer of the miter and extend into the surrounding organs, such as the pancreas, esophagus or intestine (Mikami, 2004).
Stomach cancer cells can spread by blood metastases to the liver, lungs, and other organs (Ahmad, 2007). Cancer cells can also spread to all the lymph glands in the body through the lymphatic system (Fenoglio, 2000). Cancer first starts at the center of the cells that make up the tissues, and at the same time, these tissues bring the organs of the body to the bloom (Cappell, 2002). For a normal individual, the cells grow, develop, and the body is divided to create cells that are new in need. Cells are sufficiently digested and digested by the cell when they are old and

are killed and replaced by new ones. But sometimes this process is in the wrong direction. So much so that, despite the fact that the body does not need much, new cells form and the old cells that should die do not die. These excess cells can eventually turn into tumor masses called tumors. Tumors may be benign or malignant. For this reason, the early diagnosis and diagnosis of gastric cancer has become imperative in cancer research because it may facilitate the subsequent clinical treatment of patients. Separation of gastric cancer patients into normal, low and high groups has become important in bioinformatics and biomedical fields. This has led to an increase in the practice of machine learning (ML) methods for early cancer diagnosis in the literature. Machine learning methods have been used to model the progression and treatment of stomach cancer. ML methods have been used to detect complex cell characteristics in cancer images. Artificial Neural Networks (ANN) ,Bayesian Networks (BNS), Random Forest (RF), Support Vector Machines (SVM), Decision Trees (DT), Linear Discriminant Analysis (LDA), Sammon mapping, Stochastic Neighbor new algorithms have been proposed using various machine learning techniques such as Embedding (SNE), Isomap, Classical multidimensional scaling (MDS), and Local Linear Embedding (LLE), which is used to diagnose stomach cancer early. Although machine learning techniques for gastric cancer have been widely applied and ultimately yielded high classification performances, an appropriate level of validation is required to take these methods into account in daily clinical treatment and practice. In this study, the methods used in algorithms for early diagnosis of stomach cancer and the classification ratios are described. In the advanced algorithms, various different features and image data are used. As a result, in this article, ML methods for gastric cancer research are increasing. For this reason, published articles have been presented to model the risk of stomach cancer.

References

Tannapfel, A., Schmelzer, S., Benicke, M. et al. (2001). Expression of the p53 homologues p63 and p73 in multiple simultaneous gastric cancer. The Journal of Pathology, 195, 163–170. 18.

Tolbert, D., Fenoglio-Preiser, C., Noffsinger, A. et al. (1999). The relation of p53 gene mutations to gastric cancer subsite and phenotype. Cancer Causes and Control, 10, 227–231. 19.

Siddiqi, A. M., Li, H., Faruque, F. et al. (2008). Use of hyperspectral imaging to distinguish normal, precancerous, and cancerous cells. Cancer Cytopathology, 114(1), 13–21. 20.

Mikami, T., Sasaki, Y., Fukuda, S., Hada, R., Munakata, A. (2004). Computer-aided diagnosis of superficial type early oesophageal cancer by image processing on ordinary endoscopic pictures. Gastrointestinal Endoscopy, 59, 258. 21.

Ahmad, Z., Idrees, R., Sahar, A. N., Ahmed, R., Ahsan, A., Asghar, N. (2007). Gastric carcinoma: typing, staging, lymph node and resection margin status on gastrectomy specimens. Journal of the College of the Physicians and Surgeons Pakistan, 17(9), 539. 22.

Fenoglio-Preiser, C., Carneiro, F., Correa, P. (2000). Tumours of the stomach, Pathology and Genetics of Tumors of the Digestive System. International Agency for Research on Cancer, 37–52. 23.

Cappell, M. S., Friedel, D. (2002). The role of esophagogastroduodenoscopy in the diagnosis and management of upper gastrointestinal disorders. Medical Clinics of North America, 86, 1165–216.

Lambert, R., Guilloux, A., Oshima, A., Pompe-Kirn, V., Bray, F., Parkin, M., Ajiki, W., Tsukuma, H. (2002). Incidence and mortality from stomach cancer in Japan, Slovenia and the USA. International Journal Of Cancer, 97, 811–818.

İnternet: Cancer Facts and Figures. (2015). American Cancer Society. URL: http://www.webcitation.org/query?url=http%3A%2F%2Fwhyquit.+com%2Fstudies%2F2003_acs_cancer_facts.+pdf&date=2016-02-11. Son Erişim Tarihi : 11.02.2016.

İnternet: Cancer Facts and Statictics. (2015). American Cancer Society. URL:http://www.webcitation.org/query?url=http%3A%2F%2Fwww.+cancer.+org%2Fresearch%2Fcancerfactsstatistics%2Fcancerfactsfigures2015%2F&date=2016-02-11. Son Erişim Tarihi : 11.02.2016 .

Fujioka, N., Morimoto, Y., Arai, T., Kikuchi, M. (2004). Discrimination between normal and malignant human gastric tissues by Fourier transform infrared spectroscopy. Cancer Detection and Prevention, 28(1), 32–36.

Tannapfel, A., Schmelzer, S., Benicke, M. et al. (2001). Expression of the p53 homologues p63 and p73 in multiple simultaneous gastric cancer. The Journal of Pathology, 195, 163–170.

Tolbert, D., Fenoglio-Preiser, C., Noffsinger, A. et al. (1999). The relation of p53 gene mutations to gastric cancer subsite and phenotype. Cancer Causes and Control, 10, 227–231.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal, 13, 8-17.

Korkmaz, S. A., & Poyraz, M. (2015). Least square support vector machine and minumum redundacy maximum relavance for diagnosis of breast cancer from breast microscopic images. Procedia-Social and Behavioral Sciences, 174, 4026-4031.

Korkmaz, S. A., & Poyraz, M. (2014). A New Method Based for Diagnosis of Breast Cancer Cells from Microscopic Images: DWEE—JHT. Journal of medical systems, 38(9), 92.

Korkmaz, S. A., Korkmaz, M. F., & Poyraz, M. (2016). Diagnosis of breast cancer in light microscopic and mammographic images textures using relative entropy via kernel estimation. Medical & biological engineering & computing, 54(4), 561-573.

Korkmaz, S. A., & Korkmaz, M. F. (2015). A new method based cancer detection in mammogram textures by finding feature weights and using Kullback–Leibler measure with kernel estimation. Optik-International Journal for Light and Electron Optics, 126(20), 2576-2583.

Korkmaz, S. A., & Eren, H. (2013, December). Cancer detection in mammograms estimating feature weights via Kullback-Leibler measure. In Image and Signal Processing (CISP), 2013 6th International Congress on (Vol. 2, pp. 1035-1040). IEEE.

Korkmaz, S. A., Korkmaz, M. F., Poyraz, M., & Yakuphanoglu, F. (2016). Diagnosis of breast cancer nano-biomechanics images taken from atomic force microscope. *Journal of Nanoelectronics and Optoelectronics*, *11*(4), 551-559.

Korkmaz, S. A., Poyraz, M., Bal, A., Binol, H., Özercan, I. H., Korkmaz, M. F., & Aydin, A. M. (2015). New methods based on mRMR LSSVM and mRMR KNN for diagnosis of breast cancer from microscopic and mammography images of some patients. *International Journal of Biomedical Engineering and Technology*, *19*(2), 105-117.

Korkmaz, S. A. (2018). DETECTING CELLS USING IMAGE SEGMENTATION OF THE CERVICAL CANCER IMAGES TAKEN FROM SCANNING ELECTRON MICROSCOPE. The Online Journal of Science and Technology-January, 8(1).

KORKMAZ, S. A. (2017). VARIATION OF DRIVING CONCENTRATION WITH DRIVER PERCEPTION THROUGH IN-CAR VIEW ROAD SCENE AS VISUAL STIMULANT. The Online Journal of Science and Technology-July, 7(3).

Korkmaz, S. A., & Poyraz, M. (2016, October). Path planning for rescue vehicles via segmented satellite disaster images and GPS road map. In Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), International Congress on (pp. 145-150). IEEE.

Korkmaz, S. A., & Binol, H. (2017). Analysis of Molecular Structure Images by using ANN, RF, LBP, HOG, and Size Reduction Methods for early Stomach Cancer Detection. Journal of Molecular Structure.

Korkmaz, S. A., Bínol, H., Akçiçek, A., & Korkmaz, M. F. (2017, September). A expert system for stomach cancer images with artificial neural network by using HOG features and linear discriminant analysis: HOG_LDA_ANN. In Intelligent Systems and Informatics (SISY), 2017 IEEE 15th International Symposium on (pp. 000327-000332). IEEE.

Korkmaz, S. A., Akcicek, A., Bínol, H., & Korkmaz, M. F. (2017, September). Recognition of the stomach cancer images with probabilistic HOG feature vector histograms by using HOG features. In *Intelligent Systems and Informatics (SISY), 2017 IEEE 15th International Symposium on* (pp. 000339-000342). IEEE.

Korkmaz, S. A. (2018). LBP Özelliklerine Dayanan Lokasyon Koruyan Projeksiyon (LPP) Boyut Azaltma Metodunun Farklı Sınıflandırıcılar Üzerindeki Performanslarının Karşılaştırılması. Sakarya University Journal of Science, 22(4), 1-1.

Korkmaz, S. A., Esmeray, F. (2018). "Classification with Random Forest Based on Local Tangent Space Alignment and Neighborhood Preserving Embedding for MSER features: MSER_DFT_LTSA-NPE_RF." *International Journal of Modern Research in Engineering and Technology*, 3.2:31-37.

Sasaki, Y., Hada, R., Yoshimura, T., Hanabata, N., Mikami, T., & Fukuda, S. (2010, September). Computer-aided estimation for the risk of development of gastric cancer by image processing. In IFIP International Conference on Artificial Intelligence in Theory and Practice (pp. 197-204). Springer, Berlin, Heidelberg.

Ahmadzadeh, D., Fiuzy, M., & Haddadnia, J. (2013). Stomach Cancer Diagnosis by Using a Combination of Image Processing Algorithms, Local Binary Pattern Algorithm and Support Vector Machine. Journal of Basic and Applied Scientific Research www. textroad. com.

Garcia, E., Hermoza, R., Castanon, C. B., Cano, L., Castillo, M., & Castanñeda, C. (2017, June). Automatic Lymphocyte Detection on Gastric Cancer IHC Images Using Deep Learning. In Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on (pp. 200-204). IEEE.

Zhang, X., Hu, W., Chen, F., Liu, J., Yang, Y., Wang, L., ... & Si, J. (2017). Gastric precancerous diseases classification using CNN with a concise model. PloS one, 12(9), e0185508.

Shichijo, S., Nomura, S., Aoyama, K., Nishikawa, Y., Miura, M., Shinagawa, T., ... & Tada, T. (2017). Application of Convolutional Neural Networks in the Diagnosis of Helicobacter pylori Infection Based on Endoscopic Images. EBioMedicine, 25, 106-111.

Bollschweiler EH, Monig SP, Hensler K, et al. (2004). Artificial neural network for prediction of lymph node metastases in gastric cancer: a phase II diagnostic study. Ann Surg Oncol, 11:506-11.

Cosatto, E., Laquerre, P. F., Malon, C., Graf, H. P., Saito, A., Kiyuna, T., ... & Kamijo, K. I. (2013, March). Automated gastric cancer diagnosis on h&e-stained sections; ltraining a classifier on a large scale with multiple instance machine learning. In Medical Imaging 2013: Digital Pathology (Vol. 8676, p. 867605). International Society for Optics and Photonics.

Akbari, H., Uto, K., Kosugi, Y., Kojima, K., Tanaka, N. (2011). Cancer detection using infrared hyperspectral imaging. The Official Journal of The Japanese Cancer Association, 102(4), 852-857.

