ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

A Survey on Automatic Detection and Classification of Gastric Cancer using Deep Learning

Prof. Akshita Chanchlani[1], Anjali Bhatt[2], Prajakta Awaghade[3], Sumit Chauhan[4], Sujay Pasalkar[5]

U.G. Student, Department of Computer Engineering, SKNSITS Engineering College, Maharashtra, India

Professor, Department of Computer Engineering, SKNSITS Engineering College, Maharashtra, India

ABSTRACT: From all the worst cancers in the world, Gastric Cancer is a malign of all with the highest ratio of deaths and increasing mortality rate every year in the globe. As gastric cancer is so harmful to health, the main stress should be given on detection of the cancer first and then treating it. Currently, cancer detection is done by pathologists by examining the biological tissues of a patient and then perform various tests on it and come to conclusion. But this process in turn consumes more time and manpower. In this paper, a framework for automatic detection and classification of Gastric Cancer has been proposed based on Deep Learning. A deep neural network has been built for detecting and classifying abnormal and normal images (CT images) from obtained datasets. This research in automatic detection and classification of gastric cancer has a greater value as it will help doctors and medical sciences field.

KEYWORDS: Gastric cancer, Image recognition, Convolutional neural network, Classification

I. INTRODUCTION

Gastric cancer is the fourth most common cause of cancer-related death in the world. Nearly one million new cases occur each year. Gastric cancer or stomach cancer, is a type of cancer that begins in the mucus-producing cells on the inside lining of the stomach. The most common type of stomach cancer is adenocarcinoma. The most important measure to diagnose gastric cancer is the detection and treatment of diseases early. Using Computer-Aided Diagnosis (CAD) to classify the pathology images can improve the diagnostic efficiency and provide doctors with more objective and accurate diagnosis results, which has significant clinical value. [1]

Despite a decrease in its incidence in some regions of the world, gastric cancer still poses a major clinical challenge because most cases are diagnosed in an advanced stage, with a poor prognosis and limited treatment options. The most common causes are *Helicobacter pylori* infection (proven), Epstein-Barr virus infection (suspected), and familial. Major predisposing factors include high salt intake, smoking, and a familial genetic component. Primary prevention (i.e., *H. pylori* eradication) is increasingly recommended. The current and traditional methods for detection of cancer are labour-intensive and also the result may vary from pathologists to pathologists. Due to which this method becomes error-prone. Hence, to overcome these obstacles we need to implement a trustworthy way of detecting the cancer at an early stage and then treating it. [8]

We have used python languagetensor flow library with Anaconda spyder application for basic algorithm generation and then have done the further heavy processing with the help of other online processing sources.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

II. RELATED WORK

Nowadays, deep learning has been a very successful technique in the field of machine learning and computer vision. We achieved the best performances in many applications like image classification and recognition. It has shown the capability to train a very complex model from practical dataset and extract very high-level features. Although computer-assisted image analysis systems have been developed to aid the detection of cancer in pathological images such as breast and cervical cancer; At the first stage, handcraft features such as geometric and pixel features were extracted at low resolution, support vector machine was employed to yield candidate cancer areas. At the second stage, convolutional neural network (CNN) was employed to analyse the candidate areas at high resolution and make the final diagnosis. The image was separated into small patches, CNN were employed to identify the epithelial and stromal tissues. CNN was employed for patch classification. In this paper, we proposed a deep learning framework for gastric cancer identification. We proposed a network using different structures for shallow and deep layers, and applied it for gastric cancer identification. Experimental results illustrate the excellent classification performance of proposed deep learning network, which outperforms several well-known deep learning frameworks. [1] Refer Table 1 for detailed Literature Survey result.

A deep residual networks model to extract the features of the images and automatically classify the pathological images of gastric cancer. At the same time, the method of data enhancement and migration learning are used to improve the recognition accuracy to meet the high medical standards. Deep convolutional neural networks are usually composed of one or more volume layer and full connection layer, and also include associated weight and pool layer. Generally, the gradient descent method and the chain rule are used to update the network parameters. ResNet has very deep network architecture. It has shown good characteristics in many previous pattern recognition and image classification. We proposed a new residual unit, which makes the training simpler, but also improves the generalization ability of network. [2] The convolution layer is used for feature learning. In this paper, we used the convolution kernels size as 7*7 and 3*3, etc. and ReLU,etc as activation function. Each neuron input from a fixed area in the feature maps layer of neurons, a layer of map in the area size is determined by the size of the convolution kernel, a convolution kernels can be learning from the previous layer several map as the convolution, the corresponding element accumulation after add a bias, through nonlinear activation function, such as ReLU or Sigmoid, get a feature map, which implements a feature extraction. [4]. Below is the figure [1] which shows CNN architecture.

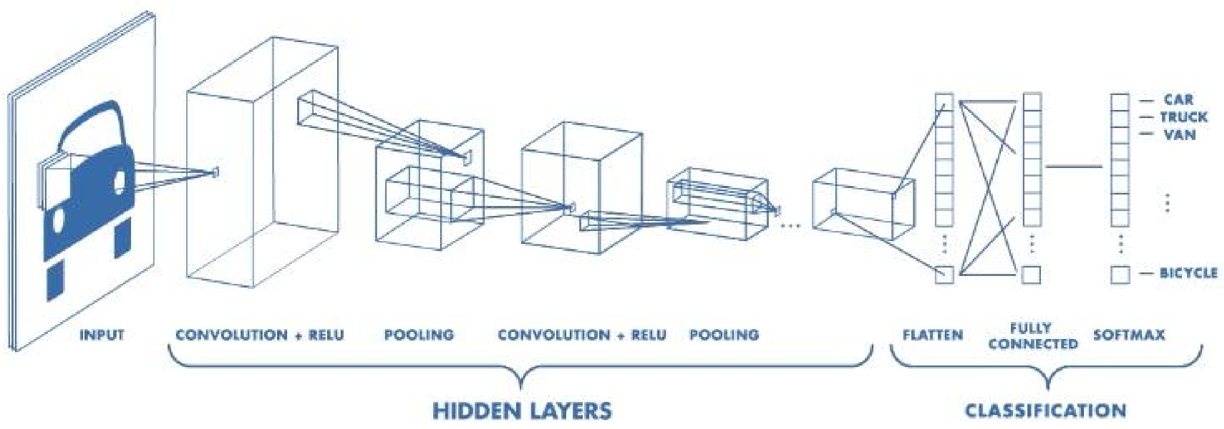

Figure.1: CNN Architecture

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Table I: Algorithms Implemented which were obtained from Related Work

Sr.No.	Title	Work/Methodologies Used	Accuracy obtained
1	Deep Learning Based Gastric Cancer Identification [1]	Gastric Net Algorithm	Classification accuracy of 100% for sliced-based classification.
2	Gastric Pathology Image Recognition Based on Deep Residual Networks [2]	Deep convolutional neural networks [1]	The final score of the test set reached 96%.
3	Gastric Lymph Node Cancer Detection of Multiple Features Classifier for Pathology Diagnosis Support System [3]	HLAC feature, wavelet feature, Delaunay feature.	In the best performance, sensitivity and specificity were 94.6% and 84.9% respectively.
4	A Novel Gastric Ulcer Differentiation System Using Convolutional Neural Networks [4]	Deep Convolutional Neural Network (CNN)[1][2]	Five different models were tested from which ERFCN achieves best accuracy of 86%
5	Gastric Lymph nodes detection Based on Visual Saliency and Dictionary Learning [5]	Visual Saliency Algorithm	The accuracy reached was 79.5%
6	Automatic Lymphocyte Detection on Gastric Cancer IHC Images using Deep Learning [6]	Model based on Deep Convolutional Neural Networks for Automatic lymphocyte detection on IHC images of gastric cancer.[1][2]	Detection on IHC images of gastric cancer using Deep Convolutional Neural Networks produced an acceptable 96.88% accuracy,
7	Detection of Gastric Cancer Risk from X-Ray Images via Patch-Based Convolutional Neural Network [7]	Patch-based Convolutional Neural Network (CNN).	The Proposed system obtained the accuracy rate of 89%.
8	Helicobacter pylori Infection Detection from multiple X-Ray images based on Combination use of Support Vector Machine and Multiple Kernel Learning [8]	Support Vector Machine (SVM) and Multiple Kernel Learning (MKL).[1]	Support Vector Machine (SVM) and Multiple Kernel Learning (MKL) together reached 94% accuracy rate.
9	Partial Labelled Gastric Tumour Segmentation via patch-based Learning [9]	Fully Convolutional Networks (FCN)-based model [1][2]	The mean accuracy of 91% was achieved.
10	Development of a Gastric Cancer Diagnostic Support System with a Pattern Recognition Method Using Hyperspectral Camera [10]	Optimal wavelength selection technique	An average sensitivity of 72.2% and average specificity of 98.8% was obtained.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

III. PROPOSED ALGORITHM

Firstly, standard data augmentation techniques are applied on the available gastric cancer dataset and thousands of images of size 512x512 are generated. Different CNN architectures are empirically studied to observe the behaviour of variation in model characteristics (network depth, layer properties, training parameters, etc.) by training them from scratch on a representative subset of whole data for cancer classification. The self-designed CNN architecture with best classification rates is selected. Later, the proposed CNN is also applied for further detection. [6]

Based on deep learning, we formulated automatic classification according to the existence of cancerous tissue in the pathological section of gastric cancer. A deep residual neural network model is proposed which has deeper and more complex structures. [1] The segmentation is carried out with the help of SVM (Support Vector Machine) algorithm. [8]

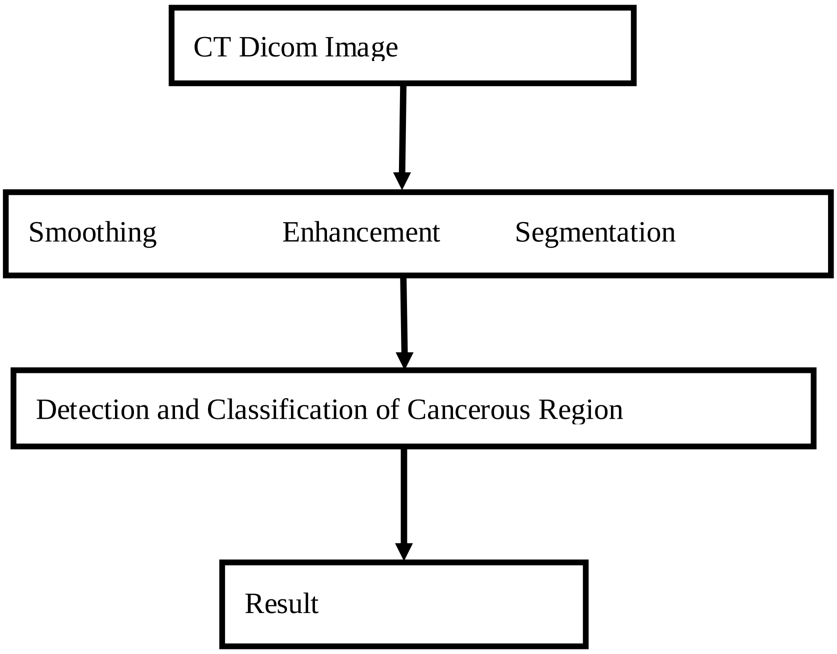

Figure 2: Proposed Architecture

IV. PSEUDO CODE

Step 1: Read the CT dicom images with help of pydicom library.

Step 2: Numpy array will contain the data.

NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Step 3: Preprocessing begins with CNN

1. Convolution

A convolution is a combined integration of two functions that shows you how one function modifies the other. There are three important items to mention in this process: the input image, the feature detector, and the feature map. The input image is the image being detected. The feature detector is a matrix, usually 3x3 (it could also be 7x7). A feature detector is also referred to as a kernel or a filter. Intuitively, the matrix representation of the input image is multiplied element-wise with the feature detector to produce a feature map also known as a convolved feature or an activation map. The aim of this step is to reduce the size of the image and make processing faster and easier. Some of the features of the image are lost in this step. However, the main features of the image that are important in image detection are retained. These features are the ones that are unique to identifying that specific object. For example, each animal has unique features that enable us to identify it. The way we prevent loss of image information is by having many feature maps. Each feature map detects the location of certain features in the image.

2. Apply the ReLu (Rectified Linear Unit)

In this step we apply the rectifier function to increase non-linearity in the CNN. Images are made of different objects that are not linear to each other. Without applying this function, the image classification will be treated as a linear problem while it is actually a non-linear one.

3. Pooling

Spatial invariance is a concept where the location of an object in an image doesn't affect the ability of the neural network to detect its specific features. Pooling enables the CNN to detect features in various images irrespective of the difference in lighting in the pictures and different angles of the images. There are different types of pooling, for example, max pooling and min pooling. Max pooling works by placing a matrix of 2x2 on the feature map and picking the largest value in that box. The 2x2 matrix is moved from left to right through the entire feature map picking the largest value in each pass. These values then form a new matrix called a pooled feature map. Max pooling works to preserve the main features while also reducing the size of the image. This helps reduce overfitting, which would occur if the CNN is given too much information, especially if that information is not relevant in classifying the image.

4. Flattening

Once the pooled featured map is obtained, the next step is to flatten it. Flattening involves transforming the entire pooled feature map matrix into a single column which is then fed to the neural network for processing.

5. Full connection

After flattening, the flattened feature map is passed through a neural network. This step is made up of the input layer, the fully connected layer, and the output layer. The fully connected layer is similar to the hidden layer in ANNs but in this case it's fully connected. The output layer is where we get the predicted classes. The information is passed through the network and the error of prediction is calculated. The error is then back-propagated through the system to improve the prediction.

The final figures produced by the neural network don't usually add up to one. However, it is important that these figures are brought down to numbers between zero and one, which represent the probability of each class. This is the role of the Softmax function.

Step 4: Random Forest algorithm for Detection of Tumor.

Random Forest is a supervised learning algorithm. Like you can already see from its name, it creates a forest and makes it somehow random. The forest it builds, is an ensemble of Decision Trees, most of the time trained with the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

"bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result.

Step 5: SVM (Support Vector Machine) for staging of cancer.

The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N—the number of features) that distinctly classifies the data points.

Step 6: go to step 4.

Step 7: End.

V. CONCLUSION

A self-designed CNN architecture is proposed for imageanalysis (cancer classification and detection based CNN) on datasets of gastric cancer. The datasets were obtained fromwww.cancerimagingarchive.net. Thus, a deep residual networks model to extract the features of the images andautomatically classify the pathological images of gastric cancer is implemented

REFERENCES

1.Yuexiang Li, Xuechen Li, Xinpeng Xie and Linlin Shen: DEEP LEARNING BASED GASTRIC CANCERIDENTIFICATION ,IEEE 15th InternationalSymposium on Biomedical Imaging (ISBI 2018) April 4-7, 2018, Washington, D.C., USA

2. Bo Liu, Kelu Yao, Mengmeng Huang, Jiahui Zhang,Yong Li, Rong Li : Gastric Pathology ImageRecognition Based on Deep Residual Networks 0730-3157/18/$31.00 ©2018 IEEE DOI10.1109/COMPSAC.2018.10267

3. Takumi ISHIKAWA, Junko TAKAHASHI, HiroshiTAKEMURA, Hiroshi MIZOGUCHI : GastricLymph Node Cancer Detection of Multiple FeaturesClassifier for Pathology Diagnosis SupportSystem 978-1-4799-0652-9/13 $31.00 © 2013 IEEEDOI 10.1109/SMC.2013.446

4. Jee-Young Sun, Sang-Won Lee, Mun-Cheon Kang,Seung-Wook Kim, Seung-Young Kim, Sung-Jea Ko :A Novel Gastric Ulcer Differentiation System UsingConvolutional Neural Networks 2372-9198/18/$31.00©2018 IEEE DOI 10.1109/CBMS.2018.00068

5. Nuo Tong, Shuiping Gou, Yao Yao, Chenjiao Wang,ling Bai : Gastric Lymph nodes detection Based onVisual Saliency and Dictionary Learning 978-1-5090-2597-8/16/$31.00 ©2016 IEEE

6. Emilio Garcia, Renato Hermoza, Cesar Beltran Castanon: Automatic Lymphocyte Detection Of Gastric Cancer IHC Images using Deep Learning 1063-7125/17 $31.00© 2017 IEEE DOI 10.1109/CBMS.2017.94

7. Kenta Ishihara, Takahiro Ogawa and Miki Haseyama :DETECTION OF GASTRIC CANCER RISK FROM XRAY IMAGES VIA PATCH-BASEDCONVOLUTIONAL NEURAL NETWORK IEEE 2017ICIP2017

8. Kenta Ishihara, Takahiro Ogawa and Miki Haseyama :HELICOBACTER PYLORI INFECTION DETECTION FROM MULTIPLE X-RAY IMAGES BASED ONCOMBINATION USE OF SUPPORT VECTORMACHINE AND MULTIPLE KERNEL LEARNING978-1-4799-8751-1/15/$31.00©2015 IEEE

9. Yang Nan , Gianmarc Coppola Dan Zhang , SeniorMember, IEEE, Yaonan Wang , Guanzhen Yu : Partial 4Labeled Gastric Tumor Segmentation via patch-basedCollege of Electrical and Information Engineering,Hunan University, Changsha 410082, China

10. Hiroyuki Ogihara, Yoshihiko Hamamot , Yusuke Fujita ,Atsushi Goto ,Jun Nishikawa, and Isao Sakaida : Development of a Gastric Cancer Diagnostic SupportSystem with a Pattern Recognition Method Using Hyperspectral Camera Journal of Sensors · January 2011DOI: 10.1155/2016/1803501

11. Jun Tang A Color Image Segmentation algorithm Basedon Region Growing School of Electronic Engineering Xi'an Shiyou University Xi'an China 978-1-4244-6349-7/10/$26.00 c 2010 IEEE

12. M. Coimbra, Member, IEEE, F. Riaz, Student Member,IEEE, M. Areia, F. Baldaque Silva, and M. Dinis-Ribeiro 32nd Annual International Conference of the IEEEEMBS Buenos Aires, Argentina, August 31 – September4, 2010 978-1-4244-4124-2/10/$25.00 ©2010 IEEE

13. CHEN Xian-lai, YANG Lu-ming, YANG Rong, XIAOXiao-dan Research on Application of BP Neural Network to Recognizing Gastric Cancer Cell 1-4244-1120-3/07/$25.00 ©2007 IEEE

14. Robert Caprara, Student Member, IEEE, Keith L.Obstein, Gabriel Scozzarro,Christian Di Natali, Student Member, IEEE, Marco Beccani,Student Member, IEEE, Douglas R. Morgan and PietroValdastri*, Senior Member, A Platformfor Gastric CancerScreening in Low-and Middle-Income Countries DOI10.1109/TBME.2014.2386309, IEEE

15. Kenta Ishihara, Takahiro Ogawa and Miki HaseyamaHELICOBACTER PYLORI INFECTION DETECTION FROMMULTIPLE X-RAY IMAGES BASED ON COMBINATIONUSE OF SUPPORT VECTOR MACHINE AND MULTIPLEKERNEL LEARNING 978-1-4799-8339-1/15/$31.00 ©2015IEEE

16. Kenta Ishihara, Takahiro Ogawa and Miki HaseyamaGraduate School of Information Science and Technology Hokkaido University Helicobacter Pylori Infection Detection FromGastric X-ray Images Using KLFDA-based Decision Fusion 978-1-4799-8751-1/15/$31.00 ©2015 IEEE