VFAST Transactions on Software Engineering
http://vfast.org/journals/index.php/VTSE@ 2022, ISSN (e):2309-3978, ISSN(p): 2411-6246
Volume 10, Number 3, July-September 2022 *pp: 10-23*

Improvement for Diagnosis of Gastric Cancer from Endoscopic Images using Machine Learning

Maryam Kausar Khan[1], Muhammad Siddique[1], Naeem Aslam[1], Muntazir Hussain[2], Sara Mukhtar[3], Bushra Syed[3].

[1]Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Punjab, Pakistan
[2]Department of Computer Science, Iqra University Islamabad, Pakistan
[3]Department of Electrical Engineering, MNS UET, Multan, Punjab, Pakistan
*Corresponding author email: maryamkausar007@gmail.com

ABSTRACT

Detection of cancer disease in any part of a human body is of utmost importance as it can be cured completely. In this research work, a prognosis of early gastric cancer detection by applying modern machine learning algorithms augmented with fast and efficient classification of white light images. In earlier studies for early gastric cancer detection schemes, nominal endoscopic images demand more computational effort, which slows down process and takes more time. Moreover, in the contemporary methodologies, only basic parameters were used to detect the symptoms of gastric cancer such as accuracy. Whilst in the proposed methodology, protein structure of the cancerous part is also examined with the help of Alpha fold software. A dataset consist of white-light-images is developed from the endoscopic images of the suspected patients. By utilitarian of this dataset in the proposed scheme, results are drawn which shows greater accuracy at a lower cost as compared to contemporary techniques.

KEYWORDS
Gastric cancer, early gastric cancer detection, cancer lesions

JOURNAL INFO:
Submitted: May 23, 2022 Accepted: September 15, 2022 Published: September 22, 2022

INTRODUCTION

The stomach's outer wall develops malignant (cancerous) cells, which is known as gastric cancer. Disseminated and intestinal GC is most dangerous types. Intestinal cancer is far more prevalent in adults, whereas diffuse cancer is more frequent in women and those under the age of 50.Gastrointestinal cancer is the sixth most frequent type of cancer in the world[1], and it is the third leading cause of cancer-related death. Gastric cancer is the biggest cause of cancer-related mortality in US. Ranks 15th among all cancers diagnosed in the US. It is still a difficulty to find a cure for this disease[2]. Gastric cancer is most typically detected during upper endoscopy. There are still some GC variations that can't be identified by endoscopy, making it a challenge[3].

According to the papers, the panel of researchers collaborated on a review in 2017 constructing a convocational neural network (CNN) for H. pylori chronic infection analysis[4]. In order to diagnose and treat, CNN's accuracy was equivalent to that of endoscopy. CNN could assist endoscopists evaluate H.pylori disease in a much shorter amount of time and with less risk [5]. Then, in 2018, researchers tried to improve CNN's accuracy in detecting gastric cancer using endoscopic images. They planned to use 13,000 plus gastric cancer's endoscopic images and established a CNN-created demonstration framework based on Single Shot Multi-Box Indicator (SSD) architecture. With 6mm of diameter or sometimes even more and all invading tumors, the method provides 98.6% accuracy[6]. Then, something new was discovered about the increasing proliferation of Helicobacter pylori[7], gastric carcinogenic bacteria (HP). Gastric cancer risk is increased by prolonged gastritis, which can also promote mucosal degradation and gastrointestinal metaphase[8].Early gastric cancer growth has poor morphological characteristics, implying that intuitive discovery tactics can be difficult to use[9].

Pattern of gastritis	*Gastric histology*	*Duodenal histology*	*Acid secretion*	*Clinical condition*
 Pan-gastritis	• Chronic inflammation • Atrophy • Intestinal metaplasia	• Normal	• Reduced	• Gastric ulcer • Gastric cancer
 Antral-predominant	• Chronic inflammation • Polymorph activity	• Gastric metaplasia • Active chronic inflammation	• Increased	• Duodenal ulcer

Figure 1. Pattern of gastritis

It all comes down to distinguishing minor colour differences in the mucosa and the unpattrened of the sub mucosa beneath veins. Convolution neural network fits for perceiving explicit highlights in gastric endoscopic pictures are being transmitted. The fundamental goal of developing such a framework was to detect sickness earlier and so prevent gastric cancer

development. They attempted to turn CNN into a learning mechanism that would assist them in acquiring the HP sickness, and then assessed the accuracy of the transferring CNN, which was nearly 87.6% [10]. Automatic classification of gastrointestinal neoplasm's can be achieved through the use of endoscopic imagery and convolution neural network models. White-light images of sickeningly identified stomach bruises were utilized to diagnosis advanced gastric cancer, early gastric cancer, high-grade dysplasia, low-grade dysplasia, and non-neoplasm[11]. An extensive training dataset was used to improve the performance of three different CNN models. The Inception-Resnet-v2 model and the worst-performing endoscopes were statistically indistinguishable in their ability to distinguish between stomach cancer and normal tissue[12]. A system-based convolution neural network computer-aided detection system can be used to evaluate patients for endoscopic resection and to estimate the depth of invasion in endoscopic images (CNN-CAD)[13]. Artificial intelligence and a state-of-the-art pertained CNN architecture, ResNet50[14] are used to construct a CNN-CAD system based on CNNs.

A good prognosis is strongly linked to early identification of cancer. Physician's attention can dwindle when they see the same endoscopic images over and over again, resulting in greater medical costs and morbidity for patients.[15].During an esophagogastroduodenoscopy, DCNN detects early gastric cancer (EGC) without leaving any blind patches (EGD). In unprocessed EGD footage, a framework for the stomach is employed to identify blind regions. The DCNN correctly distinguished EGC from non-malignancy 92.5 percent of the time [16]. M-NBI stomach mucosal burrows were analyzed using a convolution neural network (CNN) (Inception-v3) based method. In diagnosing early GC, the CNN method and M-NBI demonstrated great accuracy, sensitivity, and specificity.

In addition, real-world endoscopic pictures of the proximal stomach were used to train a deep learning algorithm for detecting inflammation gastritis[11]. Deep learning (DL) was built using endoscopic photographs of the proximal stomach to detect gastritis lesions[17]. DICOM was able to recognise and export unprocessed photos of white light from the camera (DICOMs).Analysis of statistical differences between expert ratings and DL is performed using Wilcox on the signed-rank test[32]. The model achieved 93% accuracy in an independent data set of expert endoscopists (AUC: 0.98; F-score: 0.93). Data analysis competitions can be won again and over again by applying the boosting technique to machine learning in conjunction with a wide range of elements, such as input variables and outputs (biological characteristics, Helicobacter pylori infection status, endoscopic findings, and blood test results). The ability of the CNN system to distinguish between EGC and gastritis in ME-NBI patients was put to the test[18]. It is possible to treat gastric cancer more conservatively when it is found early (EGC). Even so, a magnifying endoscope and narrow-band imaging can identify EGC from gastritis (ME-NBI).The CNN process with ME-NBI can quickly discriminate among EGC and gastritis due to its excellent sensitivity and NPV[19]. Endoscopic images are being analysed using a computerised image-analysis technique to determine the likelihood of developing GC[20]. An endoscopy camera with non-magnified white light imagery was used to train the system. A multi-box detector is brought down. After that, a CNN was constructed, trained, validated, and tested using the cafe deep learning framework. The application of AI in the classification of GC risk may be useful, notwithstanding its shortcomings[21].

An endoscopy is essential to examine the gastrointestinal tract because it provides a close-up view for doctors. Expertise and the GI tract's intricacy limit its ability to accurately diagnose[22]. Early gastric cancer (EGC) has distinct characteristics that set it apart from other gastrointestinal diseases[3]. Un-infected (no prior history of H. pylori infection), currently infected, and eradicated are the three categories of H. pylori infection status that can be determined by using a computer-aided diagnostic system. H. pylori infection can be classified using LCI and deep learning (DL)[33]. It was decided to divide the patients into three distinct groups using endoscopic digital light microscopy (DLM) and linked colour imaging (LCI) (IEE). The validation dataset for CAD systems is based on endoscopic videos of 120 people[16]. A dataset of at least 300 patients is required for DL to be effective. LCI and white-light imaging were used to document all endoscopic procedures (WLI)[23]. A validation study based on LCI's validation data showed that LCI-diagnosis CAD's accuracy was on par with expert endoscopic diagnosis[24].

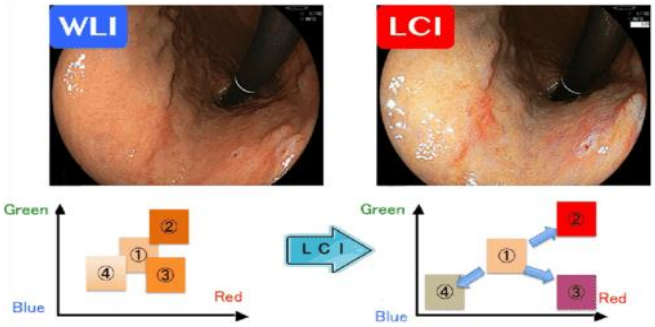


Figure 2. Difference between WLI and LCI

A predictive model for detecting stomach cancer based on non-invasive characteristics is then developed. An artificial intelligence technique called the gradient boosting decision tree (GBDT) was used to test the accuracy of their predictive model for stomach cancer diagnosis [1]. If a CNN can detect early stomach cancer better than endoscopists, then the study is successful. The CNN was trained using EGD pictures[13]. There are three methods for capturing EGD images: standard white light imaging (WLI), chromo endoscopy with the indigo carmine spray and narrow-band imaging (NBI). There are 16 or more layers in the SSD, which

stands for Single Shot Multi-Box Detector [25]. In order to speed up the diagnosis of stomach cancer, clinicians are turning to deep learning (DL) algorithms. Endoscopes can be trained quickly using DL technology, compared to normal machine learning approaches, and huge amounts of data may be evaluated[26]. Researches in this area were aimed at improving ML algorithms for endoscopic screenings and images in order to improve diagnosis[27].

Convolutional Neural Network**(CNN)**[5], Single Shot Multi-Box Detector **(SSD)**[5], **(WLI)** White Light Images [24],Gradient-boosted decision Trees **(GBDT)**[1],Extreme Gradient Boosting **(XG Boost)** [10],Magnified Endoscopy Narrow Band Imaging **(ME-NBI)**[12],Deep Convolutional Neural Network **(DCNN)**[15],Magnified Narrow Band Imaging **(M-NBI)**[12] Convolutional Neural Network-Computer Aided Detection **(CNN-CAD)**[24],Linked Colour Imaging-Computer Aided Detection **(LCI-CAD)**[24],White Light Imaging-Computer Aided Detection **(WLI-CAD)**[24], and Digital Imaging and Communication in Medicines **(DICOMs)**[25] are among the machine learning techniques which has been used earlier for detecting.

RELATED WORK

Only one machine learning model was used to predict stomach cancer in the studies we looked at.They prefer endoscopy to detect diseases the majority of the time. This strategy is also considered as a good yet efficient approach; however it is also extremely expensive. We provide a contribution to this field by merging two machine learning algorithms at the same time, resulting in accurate findings with lower the costs than previous methods. MI-NBI [11] and CAD-WLI[27] are the two techniques that make-up our proposed method. Narrow band imaging with magnified endoscopy (MI-NBI). In the initial stages of gastrointestinal detection of cancer, it is employed (EGC)[13]. The key objective of MI-NBI is to distinguish between impacted and unaffected areas of the stomach in people who have gastritis.

While CAD-WLI stands for white-light imaging, CAD is a term for computer-aided detection[24]. The use of manual drafting is no longer necessary. It makes it even easier to develop designs in 2D or 3D, making it easier to see the progress of the project. The design process can also be developed, revised, and improved using CAD[13].By combining these two methods, we hope to make early diagnosis of stomach cancer as simple and accurate as possible while keeping expenses low. We'll start by checking for H. pylori in all of our patients[4]. This is an infection that leads to chronic inflammation and the most desirable increase in the risk of gastric or duodenal ulcer cancer development. Second only to lung cancer in terms of cancer-related mortality, gastric cancer is the most common form of the disease worldwide. , and H. pylori infection is the most significant recognized risk factor [34]. If the reports are negative, we will consider our patient normal. But if the results are positive, then patients will suggest going for treatment. MI-NBI will then be used to check if we can distinguish between the impacted and unaffected stomach sections [31]. Take it for granted that the MI-NBI scans would detect any evidence of stomach cancer in the affected area.

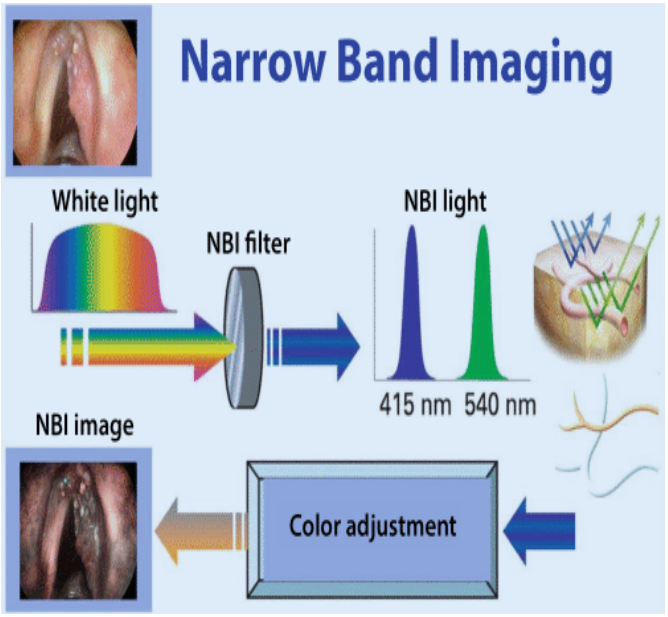


Figure 3. Conversion of WLI into MI-NBI

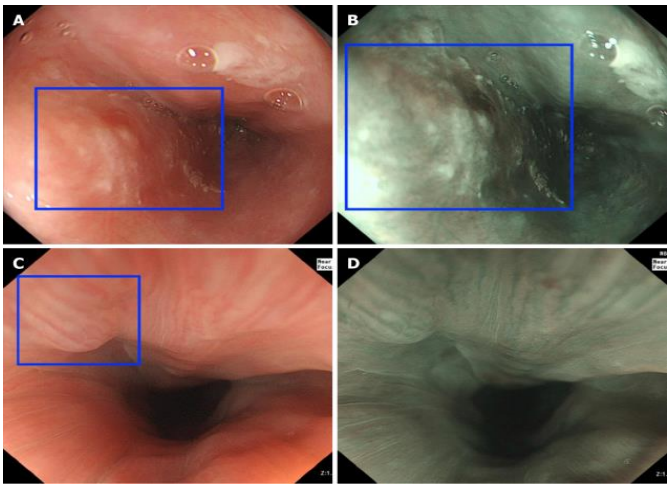


Figure 4. MI-NBI and WLI

The images will be re-run via CAD-WLI so that the computer-assisted deduction can produce more accurate and authentic results[21]. It will produce data that is both precise and detailed. This procedure will be terminated if the results are negative, at which point it will be concluded that the patient does not have gastric cancer. Using DCNN as a second set of eyes, we'll make sure nothing is missed.

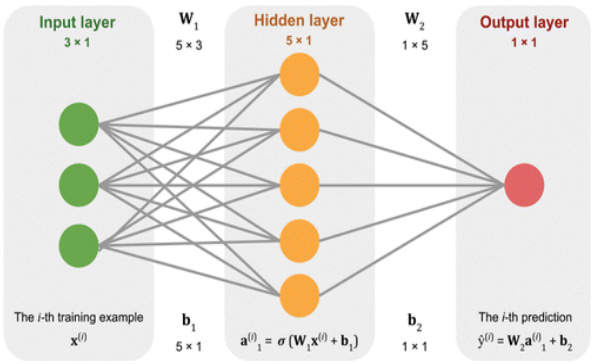


Figure 5.CNN working

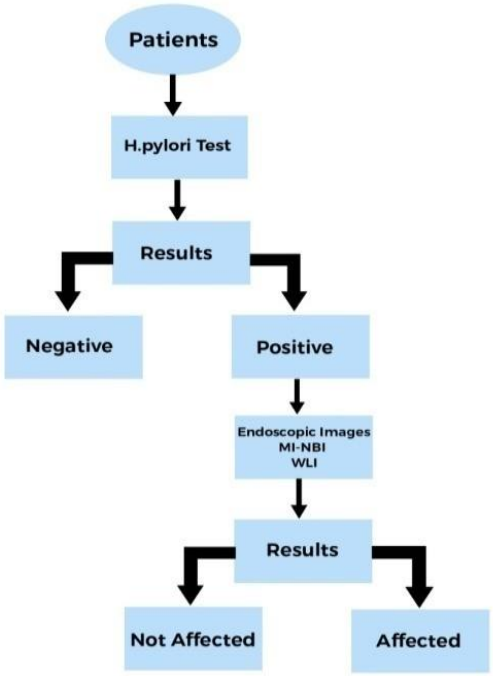
A deep conventional neural network, or DCNN, is a form of deep neural network. DCNN is made up of multiple neural network layers, including convolution and pooling layers. During esophagogastroduod endoscopy, DCNN is also utilized to identify early gastric cancer (EGC) without blind spots (EGD). If we receive positive outcomes after running our images via DCNN, we'll need some more criteria's to confirm the presence of gastric cancer. We'll also look at the stages in this phase, and we'll use the TNM staging method to figure out where they are. This approach is most commonly used to determine the stage of stomach cancer. The specialist assigns a number to the tumour size (T1–4) and whether or not lymph nodes are involved (N0–N3). And whether or not the cancer has spread (metastasized) (M0 or M1). The lower the number, the earlier the cancer has progressed. This assurance is based on factors such as gender, CBC, and LFT data, which allow us to learn about patient's physical state and how they feel. Is their weight is in balance with their other physical issues, or not? However, if we do not receive good results, we will issue a negative report to the patient, and no more testing will be required.All of the publications we have studied use these techniques separately. Convolutional Neural Network(CNN) [5], Single Shot Multi-Box Detector (SSD)[5], (WLI) White Light Images [24],Gradient-boosted decision Trees (GBDT) [1],Extreme Gradient Boosting (XG Boost) [10],Magnified Endoscopy Narrow Band Imaging (ME-NBI) [12],Deep Convolutional Neural Network (DCNN) [15],Magnified Narrow Band Imaging (M-NBI) [12] Convolutional Neural Network-Computer Aided Detection (CNN-CAD)[24],Linked Colour Imaging-Computer Aided Detection (LCI-CAD) [24],White Light Imaging-Computer Aided Detection (WLI-CAD)[24], and Digital Imaging and Communication in Medicines (DICOMs) [25].Our model provides accuracy of 98 percent in just 8 seconds of computation time, compared to these technique's accuracy of close to 80 percent and 28 seconds of computing time overall. Because of this, our suggested solution is far superior to those now in use.

DATASET DESCRIPTION

The dataset we have used is taken from Kaggle https://www.kaggle.com/code/kerneler/starter-the-nerthus-dataset-16ee48df-9/data?select=nerthus-dataset-frames.We took our dataset and optimized it using our code to get better and more accurate results. We took total of 5460 images. These images are a collection of gastric cancer WLI and Endoscopy images. Our WLI light images are first transformed into MI-NBI images.

METHODOLOGY

Our purpose is to extract (endoscopic/magnified narrow-band imaging) of affected parts in the stomach and develop training datasets from it, as well as to detect early gastric cancer by considering the damage of affected parts through those images, to enhance the DCNN system, and to proclaim the stages of gastric cancer by keeping the condition of lesions in mind (affected parts).



Flowchart 1: Process for detection of cancer patient

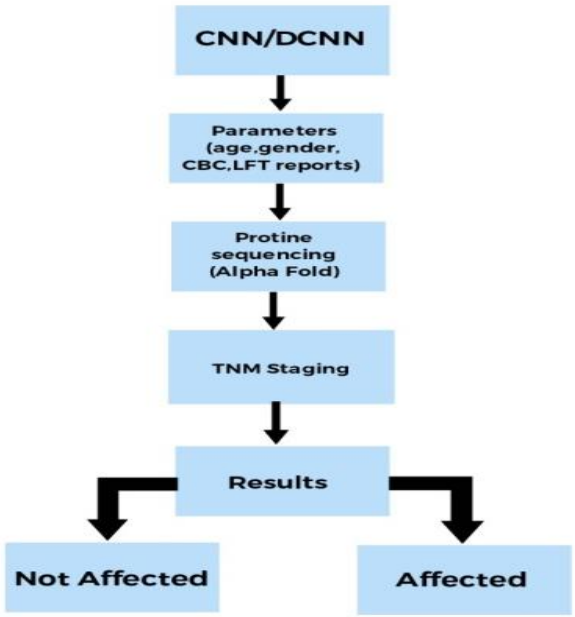
A. IMAGE ACQUISITION AND DATASET PROCESSING

Our work must be in the form of images taken from some endoscopic video while we are working on the images. The endoscopic videos are usually shot in white light. As a result, we can refer to these images as WLI (white light images). The images we used in our suggested work are 24-bit full-colour images with a size of (1000x870) pixels, which we have collected from the frames of that particular video. These photographs provide details such as the location and size of the lesions, the colour of the affected section, the swollen area, and, most significantly, how the affected part differs from the normal part. These info graphic visuals assist us in spotting

gastric cancer in its early stages. A total of 1007 images were gathered, with 448 of them including multiple lesions. We didn't consider the remaining photographs because they were deemed normal. As a result, we turned our photos into binary numbers in order to extract accurately the information from them. Because early stomach tumours have certain morphological characteristics[3], the detailed texture of these images will provide us with important information that will aid our training. We cropped some little images with a size of 224 x224 pixels to acquire that information from these images. Each of the cropped images comprises over 80% of the malignant zone, as per ground truth. We were able to obtain 1007 images of cancer as part of our training data as a result. On the other hand, 9,800 typical images with a size of 224 x224 pixels do not contain any malignant regions, thus they were clipped out of the images at random and are not included in the datasets. However, the new images produced 5,453 cancer images and 5,397 standard images for our test datasets.We know that while training CNN or DCNN systems, we need a large number of images because the larger the training dataset, the more accurate the findings will be. We require some well-explained training datasets; however they are relatively costly, primarily derived from malignant images. However, we can expand the number of training datasets by employing the Data Augmentation approach, which uses a geometric or appearance alteration to enhance the number of training datasets. Data enrichment was accomplished through the use of a Python-based open-source neural network library. All of these enhancements are performed to the same object twice: rotation and shift (rotation), shear (shift), zoom and flip More than 172,000 cancer photos and 176,388 conventional photographs in 224x224-pixel format have been collected. Consequently, it could be helpful in certain cases.

B: ECG IDENTIFICATION UTILIZING ENDOSCOPIC/NI-MBI

Early identification of gastric cancer in hospitals and medical labs was employed for screening or preoperative procedures. GIF-H290Z and GIF-H290 endoscopes (GIF-Q260J), GIF-XP290N and EVIS LUCERO CV-260/CLV-260 endoscopic video systems (EVIS LUCERO CV-290/CLV-290SL) were used to capture the images that we have cropped. The photos were taken using conventional endoscopes and standard endoscopic video systems. When making the decision of which photographs to preserve and which to discard, we've devised a set of standards. Standard white light SWL, chromo endoscopy, indigo carmine spraying, and narrow-band imaging are all acceptable methods for obtaining the necessary images (NBI). There must be at least one lesion in each of these images to qualify as a photograph. As a result of their poor quality, photos taken under dim lighting or in an isolated area are also discarded. We used 5,453 photos of cancer and 5,397 images of healthy tissue for our test and training datasets. Detecting EGC in photos is the next step after the generation of datasets. For this purpose, we submitted CAD-WLI photos that will help us identify any areas of damage. It determines if a GC or another sort of lesion is to blame for the harm. Observing the shape and colour of lesions teaches us about cancer stages.



Flowchart 2: Process for validation of cancer patient

C: CNN & DCNN SYSTEM TRAINING

We utilize the test datasets to estimate the performance of the CNN once it has been built with the training datasets. Our CNN and DCNN systems can detect the lesions. In these images, the technology is able to swiftly and precisely tell us the cancer stage based on the one lesion detected. Our system is able to identify the known area of this image thanks to certain brightly coloured frames. The highlighting was necessary because the same location appeared in multiple images. We utilise a distinct colour for that region in order to bring attention to it. As a result, we consider our findings correct if we see the same lesions in multiple images. Additionally, the CNN is capable of calculating the sensitivity, as well as the PPV and NPP for gastric cancer. As it is defines that the ratio of correctly diagnosed gastric cancer lesions to the total number of cancer lesions. True gastric cancer lesions identified are divided by the total number of lesions diagnosed by CNN as gastric cancer. PPV results are total of 714 of the 1007 test image sets (51 percent) showed signs of gastric cancer. Atrophy of the stomach mucosa was found in 58 of the patients (84.1 percent).It was shown that in the study population, there were 42 cases of early gastric cancer (T1) and 25 cases of advanced gastric cancer T2–T4 (32.5 percent).Tumors had an average diameter of 24 millimetres (range 3 to 170 mm). There were a total of 55 lesions, the majority of which were superficial (types 0, 1, 2, 3, and 4), 71%. All 2296 of the test images were processed by the CNN in 47 seconds. When it comes to find

gastric cancer lesions, the CNN has an overall success rate of 30.6% and 92.22% respectively. On the other hand, only 161 of the 232 malignant tumors were proven to be cancerous. For any study to be useful, it must have a high degree of accuracy, precision, sensitivity, and specificity.

Accuracy: Accuracy demonstrates the proposed overall system performance.

Below is an equation to calculate accuracy.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: The equation to compute the correctness of a system is given below.

$$Precesion = \frac{TP}{TP + FP}$$

Sensitivity: It verifies that the targeted person has the disease if the diagnostic test to predict the gastric cancer is confirmed positive. An equation to compute the sensitivity is given.

$$Sensectivity \frac{TP}{TP + FN}$$

Specificity: Specificity confirms that the targeted person is healthy if the diagnostic test to predict the gastric cancer is a true negative.

$$\text{Specificity} \frac{\text{TN}}{\text{TN} + \text{FP}}$$

ALPHA FOLD

Since there are so many protein-structured databases, it's not surprising. Alpha fold is one of the latest protein structure databases that have recently been produced. Alpha fold is a Deep Mind artificial intelligence system that excels at predicting the three-dimensional structure of proteins based on the amino acid sequence. Every day, Alpha fold maintains its precision and expertise. It is also regarded as the most rapid method available for predicting proteins at this time.

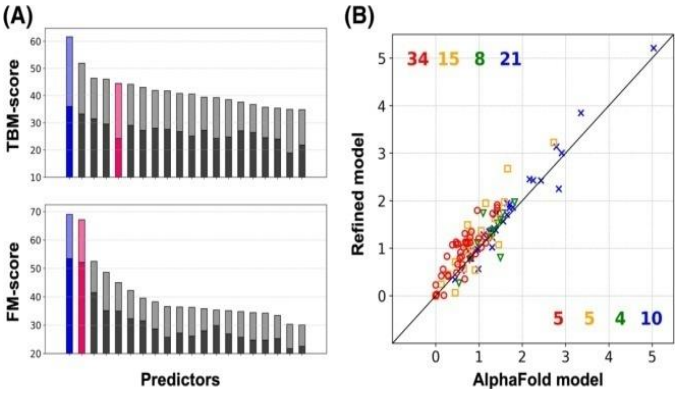


Figure 6. Alpha fold statics

There are two significant works of Alpha Fold that have made this database famous.

1) A database of protein structures

2) Predicting the 3D structure of a given protein

A gene known as CASP14 is encoded by the enzyme CASP14, which we know is found in humans. As one of the best methods for predicting protein structure, the alpha fold provides results that are 90% to 95% accurate. Using the Alpha Fold, buildings may be predicted quickly and reliably. Alpha Fold can readily anticipate any structure that is rare and has not yet been found in another database.1 lac plus sequence structures related to various organisms can be predicted by it. We can now predict the structure of proteins by utilizing a somewhat reduced version of alpha fold version 2 in Alpha fold.

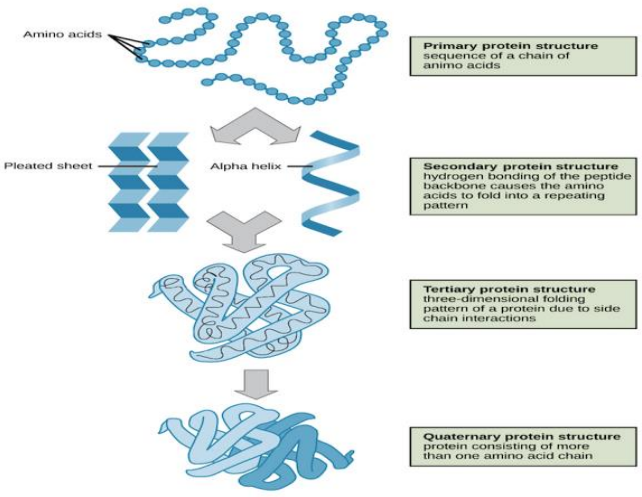


Figure7. Alpha fold step by step working

TNM STAGING SYSTEM FOR GASTRIC CANCER

The TNM approach is one of the ways that practitioners characterize the stage of the disease. Using the results of diagnostic tests and scans, doctors have come to the following conclusions. In the human body, a tumor (T) is a form of cancer. Are the tumor's walls penetrated at a significant level?

A lymph node (N) is a site where the neoplasm has spread. If this is the case, where are they located and how many of them are there?

Is there evidence that the cancer has metastasized to other parts of your body?

Each patient's cancer stage is determined by combining the test results. Stage 0 (zero) is followed by stages I through IV (1 through 4). In order to coordinate on the best possible treatments, the stage is a standard way to define cancer.

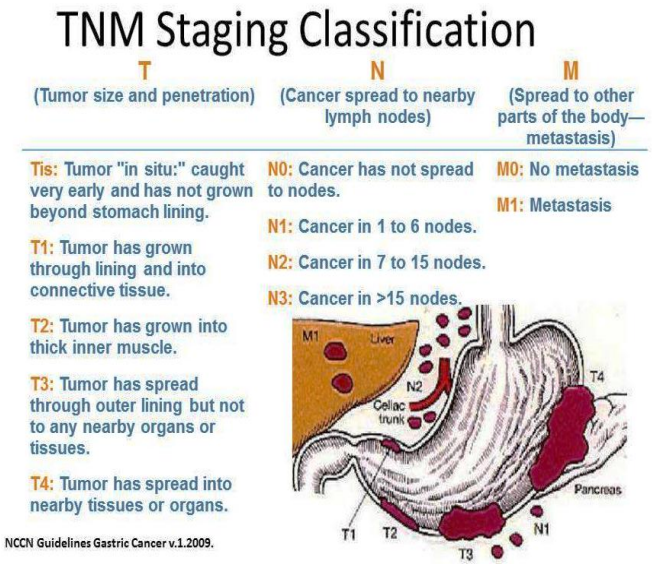


Figure 8.TNM Staging for stage detection

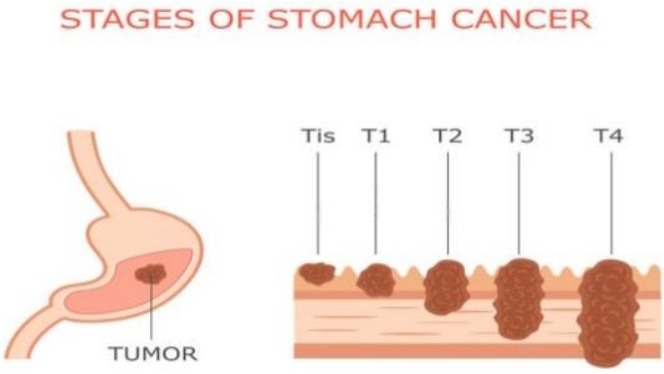


Figure 8. Stages of Gastric cancer

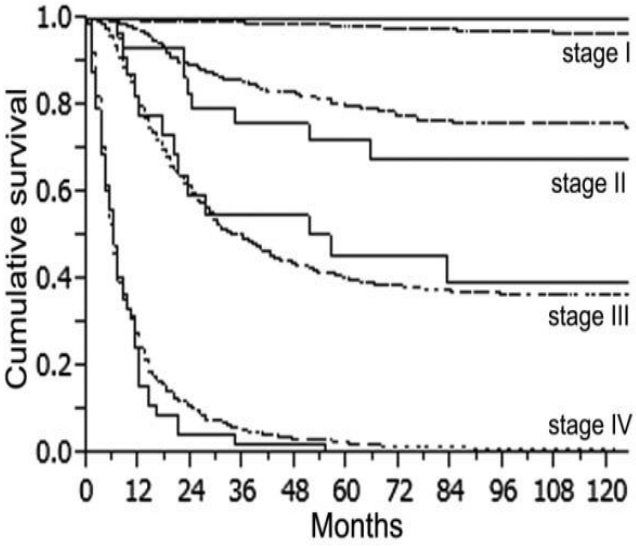


Figure 9. Survival rate of youth

RESULTS AND DISCUSSION

WHY IT IS VITAL TO CONVERT WLI TO MI-NBI AND HOW DO WE DO IT?

Using no dyes or treatments, this is the only light endoscopic technology in the world that can successfully target tumor biopsies that aren't visible under white light. The NBI to WLI to MI-NBI transition is vital and useful in distinguishing between cancerous and non-cancerous portions of cancerous tumors. MI-NBI technology makes it possible to detect tumor borders more accurately. As a diagnostic technique, NBI should not be used in place of histological sampling.Gastric cancer can still be difficult to detect using white-light imaging (WLI), despite WLI's extensive history. Researchers have found that WLI's sensitivity and specificity can range from 40 percent to 60 percent. WLI alone cannot discriminate between malignant and non-cancerous tumors. Using narrow-band imaging to magnify endoscopy has changed this situation (ME-NBI). Blue (415 nm) and green (540 nm) light is emitted selectively from the scope's tip by a narrow-band filter. Haemoglobin absorbs both blue and green light. Capillary networks can be seen more clearly in green light since it has a lower reflectivity than red light. This is not the case with blue light, which can be employed for subsurface imaging of the vascular system. In the future, it will be possible to see the surface mucosa and the blood vessels. Even while ME-NBI has been claimed to be more accurate than WLI endoscopy in diagnosing early stomach cancer, the sensitivity and specificity of ME-NBI have been inconsistent from research to study. This is the primary reason: Variations in the characteristics and dimensions of stomach lesions exist. Only one study comparing WLI and ME-NBI diagnostic efficiency has already done, however both investigations focused on the diagnostic efficiency of WLI and ME-NBI.

Some investigations employed multiple diagnostic criteria, such as ABC, VS, and Type A-E. Consequently, we conducted a review of previous studies to see if ME-NBI can accurately distinguish between gastric cancer and non-cancerous lesions.

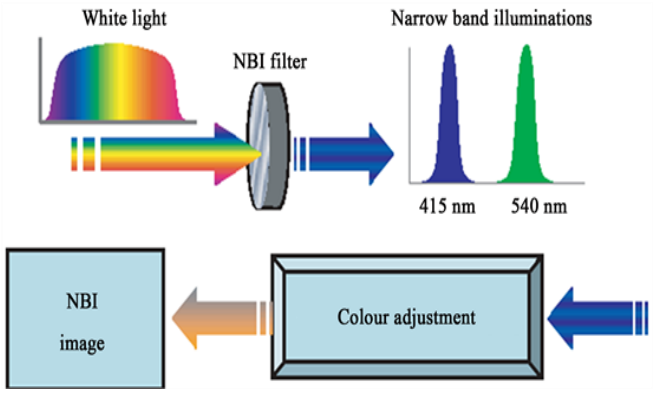


Figure 10. NBI images conversion

OUR APPROCH

TRANING DATASET

As per our training dataset, we take 3800 images from 5460 images and train them with our relevant code. We trained them to notice the change in the gastric cancer images. By noticing the size of gastric lesions, it detects the stages of gastric cancer. We train our system on detecting both cancerous and non-cancerous images. After concluding the results, we got 60% cancerous images and 40% non-cancerous images, respectively. In other words, we get 2280 cancerous images and 1520 non-cancerous images from 3800 images.

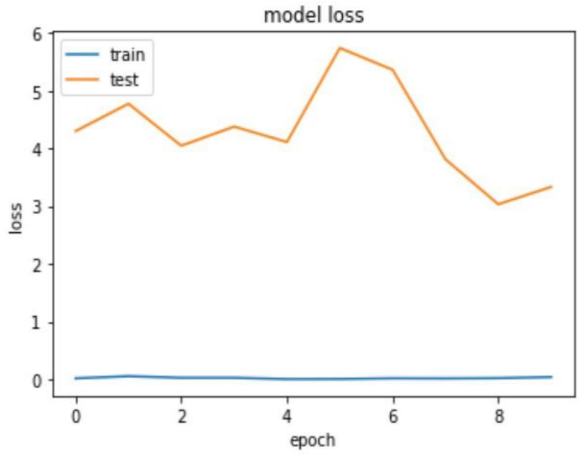


Figure 10. NBI images conversion

Here we test, our tested and trained dataset that how efficiently they worked.

TESTING DATASET

We used 1660 images in total for this study. Validation accuracy, as well as the accuracy of genuine positives and negatives (and their inverses), were all obtained after the tests were completed. By measuring the size of the lesion, we may look for all four stages of stomach cancer in our dataset.

DETERMINATION OF GASTRIC CANCER STAGES

ZERO STAGE (Tis) OF GASTRIC CANCER

Determining the "Zero" (Tis) stage of Gastric cancer. As we know that Tis is known as the very initial stage of gastric cancer we can consider it as a very first sign of cancer occurrence/development, It is a stage of GC in which tumor are in a situation where they got caught very early and has not able to grown beyond stomach lining. So that's why it is named as "zero' stage of cancer. Including this there are **5 stages**: stage 0 (zero) and stages I through IV (1 through 4).Our trained model is efficient enough to detect and then separate out all the images that belongs to "zero" stage of gastric cancer. This detection is based on the size of the lesions present on every single image.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 98\%$$

$$\text{True positive} = \frac{3800\text{x}59}{100} = 2242$$

$$\text{True negative} = 3800\text{x}\frac{39}{100} = 1482$$

$$\text{False positive} = 3800\text{x}\frac{1}{100} = 38$$

$$\text{False negative} = 38$$

As per results the images that are detected and separated out from our dataset with the help of our trained model contains 98% of accurate images which belongs exact to zero stage, and the true positive images are 2242 whereas false positive and false negative are 38 respectively.

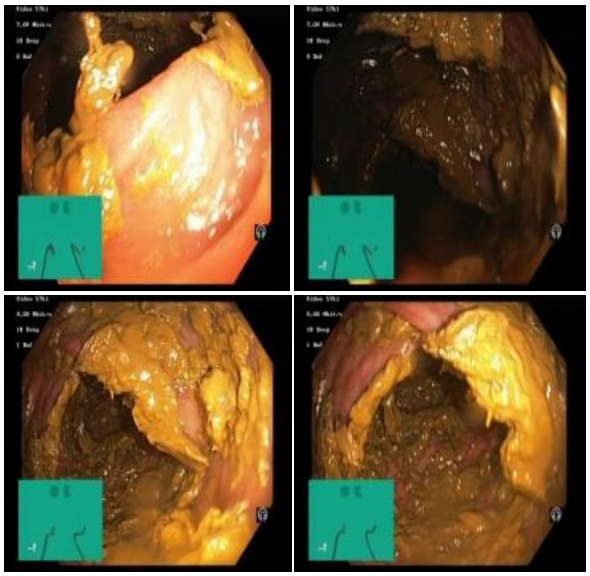


Figure 12.Zero stage of Gastric Cancer

T1 SECOND STAGE OF GASTRIC CANCER

Moving forward with the first stage of gastric cancer which is termed as T1, in this particular stage of GC the tumor has grown through the lining and then into the connective tissues. So here we underwent 3300 total images from our dataset, and our trained model detects 98% accurate images, which perfectly belongs to the first stage of gastric cancer.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 98\%$$

$$\text{True positive} = 3300\text{x}\frac{59}{100} = 1947$$

$$\text{True negative} = 3300\text{x}\frac{39}{100} = 1287$$

$$\text{False positive} = 3300\text{x}\frac{1}{100} = 33$$

$$\text{False negative} = 33$$

As per results our trained model shows 1947 images as a true positive, 1287 as a true negative and 33, 33 as a false positive and false negative

T2 THIRD STAGE OF GASTRIC CANCER

T2 is known as the third stage of the gastric cancer. In this stage of GC tumor has grown into thick inner muscle. For getting the results of T2 stage of gastric cancer we gave our model a command of 3000 images from our dataset to detect all the images which belongs to second stage of gastric cancer. As we have already explained that these stages are defined on the size of the lesions, larger the size of lesion larger the risk of being on a serious stage of GC.

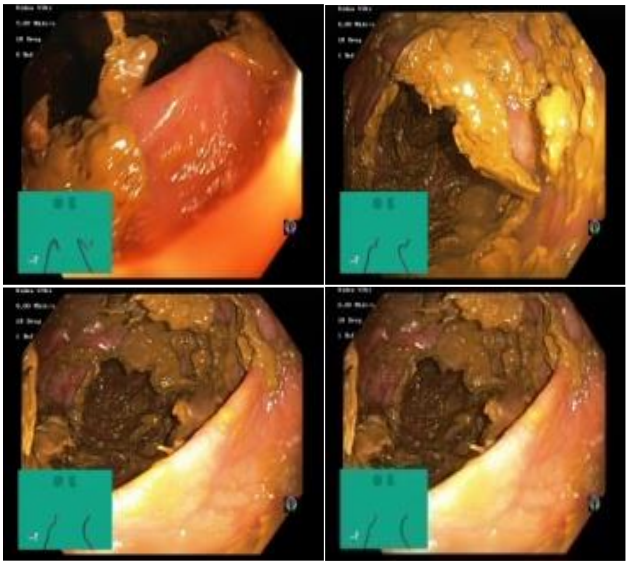


Figure 12.T1 stage of Gastric Cancer

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 98\%$$

$$\text{True positive} = 3000\text{x}\frac{59}{100} = 1770$$

$$\text{True negative} = 3000\text{x}\frac{39}{100} = 1170$$

$$\text{False positive} = 3000\text{x}\frac{1}{100} = 30$$

$$\text{False negative} = 30$$

For the results of T2 stage detection,our trained model shows the correctly detected images with the accuracy of 98% and the images that are truly predicted are 1770 in numbers. Whereas true negative images are 1170 in numbers, and lastly false positive and false negative shares the same number which is 30.

T3 FOURTH STAGE OF GASTRIC CANCER

T3 which is known as the fourth stage of gastric cancer is a stage in which tumor got spread through outer lining but not affect any tissue or the organ. We give command to our model to detect such cancer images which belongs to that stage. The larger the size of the lesion the more sensitive the stage is of gastric cancer

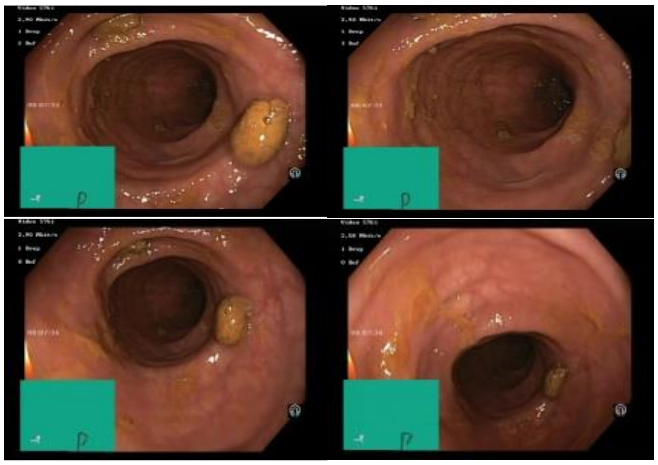


Figure 13.T2 stage of Gastric Cancer

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = 98\%$$

$$\text{True positive} = 2800\text{x}\frac{59}{100} = 1652$$

$$\text{True negative} = 2800\text{x}\frac{39}{100} = 1092$$

$$\text{False positive} = 2800\text{x}\frac{1}{100} = 28$$

$$\text{False negative} = 28$$

The below mentioned images are the images which belongs to the T3 stage of GC. For detecting the images of this particular stage our trained model again got 98% of accurate results with the 1652 images that belongs to true positive and 1092 images that belongs to true negative. And false positive plus false negative got 28 images only.

T4 FIFTH STAGE OF GASTRIC CANCER

T4 which is also termed as the last stage of gastric cancer is a stage where tumor got spread into tissues as well in organs. It's the most dangerous stage of all.

TOTAL TIME TAKEN IN EXECUTION

As per our results, MI-NBI takes a minimum of 4s and a maximum of 8s to execute on our trained system and below mentioned machine along with the help of Jupiter Notebook because the code is written in the python language, so that's why we use Jupiter notebook and give us 98% accurate results.

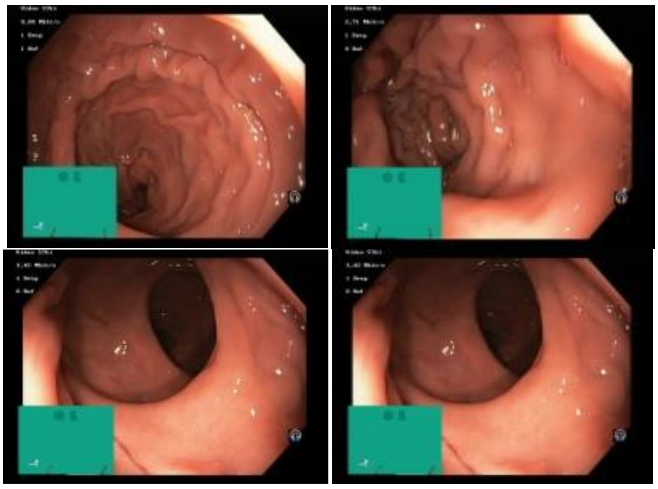


Figure 13. T3 stage Gastric Cancer

Whereas the dataset containing standard endoscopy/WLI takes more time to execute more than 10-15s and gives only 92.0% results, we can also say that our proposed method also helps in getting the more accurate results in less time.

MACHINE USED

The system, or dataset, over which we train, test, and execute, has the following characteristics: Processor 3rd Generation Intel® Core TM i5-3230M (2.6 GHz, 3 MB L3 cache, two cores) the machine on which we train, test &run, or dataset has properties mentioned below.

Processor	Intel Turbo Boost up to 3.20 GHz
	Enhance Technology
Graphics	Intel HG Intel HD Graphics 4000 is a graphic
	15.6" diagonal LED-backlit
Display Size	HD display
	anti-glare(1366 x 768 pixels)

FUTURE WORK

Using a modest number of datasets, our CNN-based approach was able to detect early gastric cancer with high accuracy. More pictures and videos with more surface features similar to gastritis can easily be added to the Training datasets in the future, we believe.

Technique	**Paper Name**	**Sensitivity**	**Specificity**	**Accuracy**	**Validation**	**PPV**(positive predictive value)	**NPV**(negative predictive value)	**Processing power**	**Age**	**Results**
CNN(SSD)	Application of artificial intelligence using a convolution neural network for detecting gastric cancer in endoscopic images	92.2%		92.2%	-	30.6%	-	47 s to analyse more than 2000 test images	(30-85)	Half of the false-positive lesions were gastritis with changes in color tone or an irregular mucosal surface.
CNN	Automated classification of gastric neoplasm in endoscopic images using a convolution neural network	84.0%	87.3 %		-	-	-	-	(50-79)	The evaluated deep-learning models have the potential for clinical application in classifying gastric cancer or neoplasm on endoscopic white-light images.
DCNN	A deep neural network improves endoscopic detection of early gastric cancer without blind spots	94.0 %	91.0 %	92.5 %	-	91.3 %	93.8 %	-	-	DCNN achieved automated performance for detecting EGC and monitoring blind spots.
CNN-CAD	Application of convolution neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy	76.47%	95.56 %	89.16 %	-	89.66%	88.97%	-	-	This system distinguished early gastric cancer from deeper sub mucosal invasion and minimized overestimation of invasion depth, which could reduce unnecessary mastectomy.
LCI (CAD)	Endoscopic three categorical diagnosis of Helicobacter pylori infection	Unaffected (92.5%) Unaffected	Unaffected (81.2 %) Currently	84.2%	120 datasets	Unaffected (69.8%) Currently affected(80.6)	-	-	(57-59)	The results of this study suggest the feasibility of an innovative gastric cancer screening

	using linked color imaging and deep learning: a single center prospective study (with video)	(92.5%) Currently affected (62.5%) Post-era distinction (60.0%)	affected (92.5 %) Post-era distinction (90.0 %)			Post-era distinction (70.3)				program to determine cancer risk in individual subjects based on LCI-CAD.
CAD-WLI	Automatic detection of early gastric cancer in endoscopic images using a transferring convolution neural network	80.0%	94.8%	87.6%	-	93.4%	-	4 ms per image & 10 hr over all	-	The detection accuracy was 82.8%. This means that our proposed scheme may offer substantial assistance to endoscopists in decision making.
ME-NBI	Convolution Neural Network for Differentiating Gastric Cancer from Gastritis Using Magnified Endoscopy with Narrow Band Imaging	95.4%	71.0%	85.3%	-	82.3% 87% 83%	91.7%	test speed was 51.83 images/s (0.02 s/image)	-	The accuracy of the CNN system with ME-NBI images was 85.3%, with 220 of the 258 images being correctly diagnosed
GBDT	Application of machine learning in the diagnosis of gastric cancer based on non-invasive characteristics	Training dataset (88.0%) Test dataset (83.0%)	83.4% 84.1%	85.9% 83.0%	-	87.0% 83.0%	84.7% 87.8%	Formula (M iterations)	(45-72)	We construct a predictive model to diagnose gastric cancer with high sensitivity and specificity. The model is non-invasive and may reduce the medical cost.
CNN (XG Boot)	Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive-medical checkup data:	0.933%	0.768 %	0.777 %					(46-56)	

CREDIT AUTHOR STATEMENT

Maryam Kausar Khan: Computational Neural Network, Disease Staging system, Methodology and Writing- Original draft preparation. **Muhammad Siddique:** Supervision: **Naeem Aslam**: Visualization, Investigation: **Muntazir Hussain:** Data curtain: **Sara Mukhtar** System Validation: **Bushra Syed:** Writing- Reviewing and Editing

COMPLIANCE WITH ETHICAL STANDARDS

It is declared that all authors don't have any conflict of interest. Furthermore, informed consent was obtained from all individual participants included in the study

REFERENCES

[1] P.-H. Niu, L.-L. Zhao, H.-L. Wu, D.-B. Zhao, and Y.-T. Chen, "Artificial intelligence in gastric cancer: Application and future perspectives," World Journal of Gastroenterology, vol. 26, no. 36, pp. 5408–5419, Sep. 2020.

[2] P. Jin, X. Ji, W. Kang, Y. Li, H. Liu, F. Ma, S. Ma, H. Hu, W. Li, and Y. Tian, "Artificial Intelligence in gastric cancer: A systematic review," Journal of Cancer Research and Clinical Oncology, vol. 146, no. 9, pp. 2339–2350, 2020.

[3]J. Taninaga et al., "Prediction of future gastric cancer risk using a machine learning algorithm and comprehensive medical check-up data: A case-control study," Scientific Reports, vol. 9, no. 1, Aug. 2019.

[4] W. Pan, X. Li, W. Wang, L. Zhou, J. Wu, T. Ren, C. Liu, M. Lv, S. Su, and Y. Tang, "Identification of Barrett's esophagus in endoscopic images using Deep Learning," 2021.

[5] T. Hirasawa et al., "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," Gastric Cancer, vol. 21, no. 4, pp. 653–660, Jan. 2018.

[6] L. Wu, W. Zhou, X. Wan, J. Zhang, L. Shen, S. Hu, Q. Ding, G. Mu, A. Yin, X. Huang, J. Liu, X. Jiang, Z. Wang, Y. Deng, M. Liu, R. Lin, T. Ling, P. Li, Q. Wu, P. Jin, J. Chen, and H. Yu, "A deep neural network improves endoscopic detection of early gastric cancer without blind spots," Endoscopy, vol. 51, no. 06, pp. 522–531, 2019.

[7] V. E. Strong, "Progress in gastric cancer," Updates in Surgery, vol. 70, no. 2, pp. 157–159, Jun. 2018.

[8] T. Itoh, H. Kawahira, H. Nakashima, and N. Yata, "Deep learning analyzes Helicobacter pylori infection by upper gastrointestinal endoscopy images," Endoscopy International Open, vol. 06, no. 02, pp. E139–E144, Feb. 2018.

[9] S.-L. Zhu, J. Dong, C. Zhang, Y.-B. Huang, and W. Pan, "Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics," PLOS ONE, vol. 15, no. 12, 2020.

[10] T. Hirasawa et al., "Current status and future perspective of artificial intelligence applications in endoscopic diagnosis and management of gastric cancer," Digestive Endoscopy, vol. 33, no. 2, pp. 263–272, Dec. 2020.

[11] Y. Ikenoyama et al., "Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists," Digestive Endoscopy, vol. 33, no. 1, pp. 141–150, Jun. 2020.

[12] Y. Xue, N. Li, X. Wei, R. A. Wan, and C. Wang, "Deep learning-based earlier detection of esophageal cancer using improved empirical wavelet transform from Endoscopic Image," IEEE Access, vol. 8, pp. 123765–123772, 2020.

[13] B.-J. Cho, C. S. Bang, S. W. Park, Y. J. Yang, S. I. Seo, H. Lim, W. G. Shin, J. T. Hong, Y. T. Yoo, S. H. Hong, J. H. Choi, J. J. Lee, and G. H. Baik, "Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network," Endoscopy, vol. 51, no. 12, pp. 1121–1129, 2019.

[14] K. Kubota, J. Kuroda, M. Yoshida, K. Ohta, and M. Kitajima, "Medical Image Analysis: Computer-aided diagnosis of gastric cancer invasion on Endoscopic Images," Surgical Endoscopy, vol. 26, no. 5, pp. 1485–1489, 2011.

[15] W. Qiu, J. Xie, Y. Shen, J. Xu, and J. Liang, "Endoscopic Image Recognition Method of gastric cancer based on Deep Learning Model," Expert Systems, vol. 39, no. 3, 2021.

[16] Y. Zhu, Q.-C. Wang, M.-D. Xu, Z. Zhang, J. Cheng, Y.-S. Zhong, Y.-Q. Zhang, W.-F. Chen, L.-Q. Yao, P.-H. Zhou, and Q.-L. Li, "Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy," Gastrointestinal Endoscopy, vol. 89, no. 4, 2019.

[17] J. K. Min, M. S. Kwak, and J. M. Cha, "Overview of deep learning in gastrointestinal endoscopy," Gut and Liver, vol. 13, no. 4, pp. 388–393, 2019.

[18] Y. Sakai, S. Takemoto, K. Hori, M. Nishimura, H. Ikematsu, T. Yano, and H. Yokota, "Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018.

[19] M. K. Cheung, A. Y. Wong, L. Chen, E. W. Chan, I. C. Wong, and W. K. Leung, "Long-term use of proton pump inhibitors and risk of gastric cancer development after treatment for H. pylori : A population-based study," Gastroenterology, vol. 152, no. 5, 2017.

[20] Y. Sakai, S. Takemoto, K. Hori, M. Nishimura, H. Ikematsu, T. Yano, and H. Yokota, "Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018.

[21] K. Zhang, "Acupuncture for quality of life in gastric cancer patients: Methodological issues," Journal of Pain and Symptom Management, vol. 63, no. 4, 2022.

[22] A. Jibawi, M. Baguneid, and A. Bhowmick, "Gastric cancer," Current Surgical Guidelines, pp. 279–294, 2018.

[23] T. de Lange, P. Halvorsen, and M. Riegler, "Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy," World Journal of Gastroenterology, vol. 24, no. 45, pp. 5057–5062, 2018.

[24] H. Nakashima, H. Kawahira, H. Kawachi, and N. Sakaki, "Endoscopic three-categorical diagnosis of helicobacter pylori infection using linked color imaging and Deep Learning: A single-center prospective study (with video)," Gastric Cancer, vol. 23, no. 6, pp. 1033–1040, 2020.

[25] S. Shichijo, S. Nomura, K. Aoyama, Y. Nishikawa, M. Miura, T. Shinagawa, H. Takiyama, T. Tanimoto, S. Ishihara, K. Matsuo, and T. Tada, "Application of convolutional neural networks in the diagnosis of helicobacter pylori infection based on endoscopic images," EBioMedicine, vol. 25, pp. 106–111, 2017.

[26] H. J. Yoon and J.-H. Kim, “Lesion-based convolutional neural network in diagnosis of early gastric cancer,” Clinical Endoscopy, vol. 53, no. 2, pp. 127–131, 2020.

[27] Y. Horiuchi, K. Aoyama, Y. Tokai, T. Hirasawa, S. Yoshimizu, A. Ishiyama, T. Yoshio, T. Tsuchida, J. Fujisaki, and T. Tada, “Convolutional Neural Network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging,” Digestive Diseases and Sciences, vol. 65, no. 5, pp. 1355–1363, 2019.

[28] H. Ueyama, Y. Kato, Y. Akazawa, N. Yatagai, H. Komori, T. Takeda, K. Matsumoto, K. Ueda, K. Matsumoto, M. Hojo, T. Yao, A. Nagahara, and T. Tada, “Application of artificial intelligence using a convolutional neural network for diagnosis of early gastric cancer based on magnifying endoscopy with narrow-band Imaging,” Journal of Gastroenterology and Hepatology, vol. 36, no. 2, pp. 482–489, 2020.

[29] J. H. Lee, Y. J. Kim, Y. W. Kim, S. Park, Y.-i Choi, Y. J. Kim, D. K. Park, K. G. Kim, and J.-W. Chung, “Spotting malignancies from gastric endoscopic images using Deep Learning,” Surgical Endoscopy, vol. 33, no. 11, pp. 3790–3797, 2019.

[30] J. Taninaga, Y. Nishiyama, K. Fujibayashi, T. Gunji, N. Sasabe, K. Iijima, and T. Naito, “Prediction of future gastric cancer risk using a machine learning algorithm and Comprehensive Medical Check-up data: A case-control study,” Scientific Reports, vol. 9, no. 1, 2019.