

# Histopathology Image Analysis for Gastric Cancer Detection: A Hybrid Deep Learning and CatBoost Approach

Danial Khayatian<sup>a</sup>, Alireza Maleki<sup>a</sup>, Hamid Nasiri<sup>b,\*</sup> and Morteza Dorrigiv<sup>a</sup>

<sup>a</sup>Electrical and Computer Engineering Department, Semnan University, Semnan, Iran

<sup>b</sup>Department of Computer Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

## ARTICLE INFO

### Keywords:

Gastric cancer  
Binary classification  
CatBoost  
EfficientNetV2B0  
Image Classification

## ABSTRACT

**Background and Objective:** Since gastric cancer is growing fast, accurate and prompt diagnosis is essential, utilizing computer-aided diagnosis (CAD) systems is an efficient way to achieve this goal. Using methods related to computer vision enables more accurate predictions and faster diagnosis, leading to timely treatment. CAD systems can categorize photos effectively using deep learning techniques based on image analysis and classification.

**Methods:** Accurate and timely classification of histopathology images is critical for enabling immediate treatment strategies, but remains challenging. We propose a hybrid deep learning and gradient-boosting approach that achieves high accuracy in classifying gastric histopathology images. This approach examines two classifiers for six networks known as pre-trained models to extract features. Extracted features will be fed to the classifiers separately. The inputs are gastric histopathological images. The GasHisSDB dataset provides these inputs containing histopathology gastric images in three 80px, 120px, and 160px cropping sizes. According to these achievements and experiments, we proposed the final method, which combines the EfficientNetV2B0 model to extract features from the images and then classify them using the CatBoost classifier.

**Results and Conclusion:** The results based on the accuracy score are 89.7%, 93.1%, and 93.9% in 80px, 120px, and 160px cropping sizes, respectively. Additional metrics including precision, recall, and F1-scores were above 0.9, demonstrating strong performance across various evaluation criteria. In another way, to approve and see the model efficiency, the GradCAM algorithm was implemented. Visualization via Grad-CAM illustrated discriminative regions identified by the model, confirming focused learning on histologically relevant features. The consistent accuracy and reliable detections across diverse evaluation metrics substantiate the robustness of the proposed deep learning and gradient-boosting approach for gastric cancer screening from histopathology images. For this purpose, two types of outputs (The heat map and the GradCAM output) are provided. Additionally, t-SNE visualization showed a clear clustering of normal and abnormal cases after EfficientNetV2B0 feature extraction.

The cross-validation and visualizations provide further evidence of generalizability and focused learning of meaningful pathology features for gastric cancer screening from histopathology images.

## 1. Introduction

Gastric cancer is a leading cause of worldwide cancer mortality [1], with an exceptionally high rate in Asia, Eastern Europe, and Central America [1]. According to the recent Global Cancer Statistics, more than 768,000 deaths occurred out of 1 million cases in 2020, making gastric cancer the fifth most frequently diagnosed cancer and the third cause of cancer-related deaths worldwide [2]. With the rapid advancement of computer vision technology, particularly with the development of medical image classification, it is now possible to quickly and effectively examine every microscopic photo [3], [4].

Hardware advancements in Graphical Processing Units (GPUs) have also improved Convolutional Neural Network (CNN) performance. These developments were applied in various medical fields. [5],[6].

The broad utilization of Convolutional Neural Networks (CNNs) which possess the ability to autonomously extract

local features for classification purposes [7], in picture recognition came around after Krizhevsky et al. [8]. Applications of CNN are used in various fields; some of their most critical applications are in image and signal processing, natural language processing, and data analytics [9]. Since patients unaware of the nature of their illness often seek general practitioners instead of specialists. Loss of human lives can be anticipated, or the therapeutic trauma experienced in harm or an infection can be decreased through the timely diagnosis of medical anomalies. Medical anomalies include glaucoma, diabetic retinopathy, and tumors [10]. The analysis of histopathological images by pathologists is time-consuming and needs excessive expenditures. Additionally, the manual process carries a significant risk of classification errors due to resource mismatches, demanding tasks, and human errors [11]. Consequently, the timely diagnosis and early detection of cancer assume paramount significance in ensuring effective treatment and maximizing the potential for a successful cure [? ]. The ability to solve this problem appropriately and extend the survival rate is aided by early diagnosis of this condition. ML and Deep Learning (DL) provide reliable and efficient methods for developing intelligent data-driven systems in our technological era [12],[13].

\*Correspondence should be addressed to Hamid Nasiri; [h.nasiri@aut.ac.ir](mailto:h.nasiri@aut.ac.ir)

<sup>1</sup>Email address: [danial.khayatian@gmail.com](mailto:danial.khayatian@gmail.com) (Danial Khayatian), [alirezamaleki78@gmail.com](mailto:alirezamaleki78@gmail.com) (Alireza Maleki), [h.nasiri@aut.ac.ir](mailto:h.nasiri@aut.ac.ir) (Hamid Nasiri), [dorrigiv@semnan.ac.ir](mailto:dorrigiv@semnan.ac.ir) (Morteza Dorrigiv)

Computer-aided Diagnosis (CAD) provides the possibility to gain ML and DL facilities to identify border detection, extraction, and classification operations [14],[15]. By pursuing this avenue, pathologists can anticipate a reduction in work pressures, enhanced diagnostic efficiency, and improved overall reliability within their professional practice [4].

Therefore, we proposed a method that can detect Gastric cancer with high speed and accuracy and eliminate any human intervention.

In this paper, the classification problem is considered with the help of image processing. Since our dataset is the GasHisSDB, we focus on classifying two normal and abnormal types of gastric cancer, DL models are coming up to diagnose tumor type. So, in this investigation, we will provide a model using these algorithms to improve the model's performance to get higher accuracy and better classification. **The idea is to prepare the raw images to feed into the CNN-based networks, and then get some important features from the image that play an important role in diagnosing the image class. These features are hidden in the images which can not be diagnosed by human eyes and logic. According to the task which is a binary classification, we need to feed these valuable features to the classifier to predict the labels. In this phase, the vital issue is to tune the classifier parameters to perform better in classifying. In the end, we need to verify our model by implementing some experiments to ensure our findings perform well in multiple aspects.** We aim to increase the accuracy of the previous models compared with the other papers related to the GasHisSDB dataset and gastric cancer. This paper will present a pre-trained deep-learning model for the extracting feature job and gradient-boosting method. This way, multiple combinations of DL models and gradient boosting methods are examined to choose the best model based on the accuracy score.

**While prior studies have explored various computer vision and machine learning techniques for gastric cancer detection, this work puts forth a novel approach that combines the strengths of deep learning feature extraction with gradient boosting for enhanced classification performance. Specifically, we investigate various pre-trained models, the state-of-the-art CNN architecture optimized for accuracy and efficiency, for automated feature extraction from histopathology images. The extracted features are then classified using CatBoost and LightGBM algorithms, recently developed gradient-boosting algorithms shown to deliver top results across various structured data tasks. Compared to existing methods that rely on hand-crafted features or single model architectures, our hybrid deep learning and gradient boosting framework achieves significantly higher accuracy in distinguishing between normal and abnormal gastric tissue. To the best of our knowledge, this demonstrates the power of fusing efficient deep CNNs and advanced boosting techniques for this application. The proposed model's interpretability is also analyzed through visualization techniques such as ROC curves, learning curves, t-SNE projection, and Grad-CAM implementation. Overall, this work introduces**

**an innovative pipeline that advances gastric cancer detection from histology images.**

This research will answer two essential questions: 1) How does the final method detect gastric cancer? 2) How do pre-trained models extract features and prepare them for classification?

Answering previous questions is inside the body of the following topics: In Section 2, we will go deeper to explain our final method and its components. Section 3 will present our results and experiments around the proposed model. At the end of this section, a brief discussion will be done to understand better what has been done. Summarizing the paper around the results and the achievements will be organized in Section 4. **In the end, we discuss some limitations and constraints around our approach and recommend solutions to cover them in Section 5.**

## 2. Related Works

In light of the emergence and advancement of ML, artificial intelligence, and DL, numerous research institutions have undertaken diligent investigative endeavors [? ]. Also, in recent years, many researchers have been exploring various methods to diagnose gastric cancer more rapidly than human beings. According to this trend, scientists used available datasets about gastric cancer provided by health organizations and laboratories.

In the following, we aim to compare some related papers employed by multiple researchers on different datasets around gastric cancer before presenting our final proposed method. The details of these efforts are introduced as follows:

Since the research on auto-classification of gastric pathology images is valuable, Bo et al. in 2018 presented a method for automatically detecting gastric tumor images according to DL algorithms. They used an identity mapping technique to create a deep residual network with 50 layers from a dataset of gastric cancer pathology images divided into two types. The suggested approach accelerates model training while enhancing generalization performance, achieving a high score of 96% on the F-score evaluation criteria [16].

In an investigation employed in 2019, Wang et al. designed a recalibrated multi-instance DL method (RMDL) for gastric cancer diagnosis. Their proposed framework was evaluated on the gastric Whole Slide Images (WSI) dataset. The mentioned model for multiclass cases was 86.5% accurate [17].

Due to the high death rate, another study on the rapid diagnosis of stomach cancer was initiated in 2019. Horiuchi et al. improved a new technique, "magnifying endoscopy with narrow band imaging" (ME-NBI). To progress in this technique, they use the facilities of CNNs for classifying the input image. They were followed to enhance the detection of gastritis and EGC using the ME-NBI technique. They use a 22-layer CNN network on EGC, and gastritis images are split into test and train data. Based on the results of this study in 2019, the CNN system with ME-NBI achieved an accuracy

of 85.3%, demonstrating excellent sensitivity and NPV in distinguishing between gastritis and EGC [18].

Following the previous research in 2020, Song et al. built a new CNN based on DeepLabv3 and ResNet50 to diagnose gastric cancer. Their model was evaluated on the PLAGH dataset, and the accuracy of the model is 87.3% [19].

Lizuka continued research on this type of cancer by training CNNs and recurrent neural networks (RNNs) on biopsy histopathology (WSIs) for Histopathological classification of gastric and colonic epithelial tumors. Most of their dataset originated from Hiroshima University Hospital, and the rest developed from Haradai Hospital [20].

In 2020, an investigation was done to analyze gastric cancer's histological image. Shiliang et al. introduce feature extraction approaches, review picture preprocessing techniques, and generalize segmentation and classification methods already in use. This paper studied multiple feature extraction methods, analyzed in two ways: ML approach and DL algorithms. Another analysis was performed in image preprocessing to increase the data provided in a dataset. The research shows how preprocessing helps to avoid overfitting and makes the results much more precise. Since classification plays a vital role in every image analysis, this study applied some classifiers like SVM, RM, and pre-trained models based on CNNs such as ResNet and Inception-V3. The framework structure combines DL algorithms to extract features and classifiers to do classification tasks. It is necessary to mention that postprocessing and segmentation methods are examined in this investigation [21].

In 2021, Li et al. proposed a random field based on Attention Mechanism (HCRF-AM), which consists of an Attention Mechanism (AM) module and an Image Classification (IC) module which has a CNN that is trained with the attention regions selected. A dataset containing 700 gastric histopathology images is used to build and evaluate this model. The mentioned model was 91.4% accurate on the testing images [22].

One of the numerable studies on the GasHisSDB dataset was performed by Weiming Hu et al. in 2022. They choose seven different classical ML algorithms to extract five other image virtual features to match multiple classifier algorithms. Their study was done on binary classification, divided into normal and abnormal classes. In feature extraction, methods such as color histograms, Luminance histograms, Histograms of Oriented Gradient (HOG), Local Binary Patterns (LBP), and Gray-level Co-occurrence Matrix (GLCM) were examined. The features were passed to the seven multiple classical classifications: linear regression, KNN, Random Forest, linear SVM, non-linear SVM, naïve Bayesian classifier, and ANN. The final proposed method in the ML approach combines the color histogram and Random Forest with 85.99% accuracy. It should be noted that Weiming et al. also calculated four other evaluation metrics: Precision, Recall, Specificity, and F1-Score [23].

In the latest research, Noda et al. developed a CNN-based system for diagnosing early gastric cancer, which is trained

on the training dataset comprising 1623 images. In the per-lesion analysis, the accuracy of the CNN-based findings was 86.1% [24].

A recent study by Hu et al.[23] in 2022 explored gastric histopathology image classification on the GasHisDB dataset using both traditional machine learning and deep learning techniques. They extracted handcrafted features like LBP, HOG, and GLCM to feed into linear regression, SVM, random forest, and other shallow models. Additionally, they evaluated state-of-the-art convolutional neural networks including ResNet50 and visual transformers. Their work focused only on the 160px resolution images. However, our study explores a wider range of image resolutions and proposes a novel deep CNN feature extractor paired with gradient boosting for enhanced performance across evaluation metrics. The advanced representations learned by our approach contribute uniquely to gastric histopathology analysis.

In this segment, we have sought to compile existing studies centered on the diagnosis of gastric cancer. Our emphasis in introducing and scrutinizing these selected papers lies in their shared focus on improving the diagnostic capabilities for this particular cancer by applying diverse models. Notably, a prevalent theme across these investigations involves the utilization of ML and DL methodologies. This alignment underscores a commonality between our study and prior research endeavors. Our primary objective is to develop an effective model grounded in CNNs to elevate accuracy and other pertinent metrics in the detection of gastric cancer.

### 3. Materials and Methodology

#### 3.1. Dataset

The GasHisSDB dataset [25], was employed to evaluate the proposed algorithm. This dataset includes 245,196 gastric histopathology images, consisting of 97,076 abnormal and 148,120 normal images of patients respectively. GasHisSDB contains pictures in PNG format obtained by electron microscopy. For more details of the dataset, you can see Table 1. Also, in Figure 1, we represent sample images of every cropping size for both abnormal and normal classes.

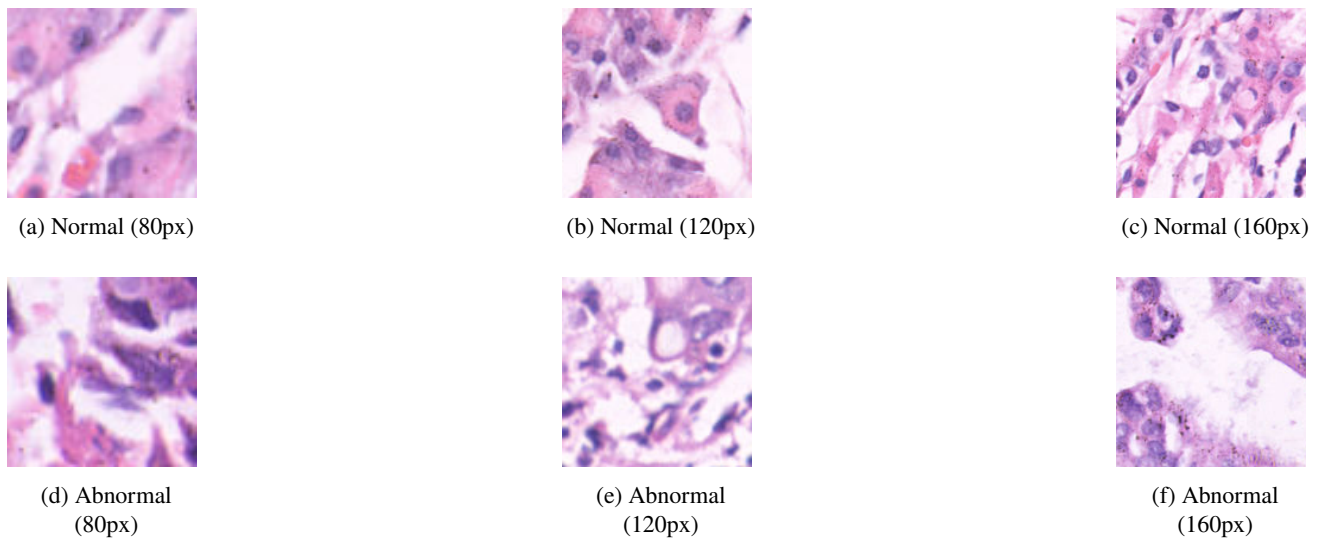
**Table 1**

Distribution of the images in the GasHisSDB dataset [25]

Cropping size	Abnormal	Normal
160px	13,124	20,160
120px	24,801	40,460
80px	59,151	87,500
<b>Total</b>	<b>97,076</b>	<b>148,120</b>

#### 3.2. Methodology

This study classifies gastric cancer into two main classes: abnormal and normal tumors. In a nutshell, the suggested method is used for the classification task mentioned earlier, as demonstrated in Figure 2 which provides a visual representation of the approach. As demonstrated by this



**Figure 1:** Samples of normal (a-c) and abnormal (d-f) tumor types in all cropping sizes. (Source: GasHisSDB dataset [25])

architecture, we used pre-trained models, which have been trained on 14 million images before, to extract features from the images. These new DL models are not feature extractors by themselves. So, a change is needed to transform them into extracting elements. This change is implemented by removing fully connected layers, which classify images based on features extracted by convolutional layers, and adding another classifier to the network. To deploy this concept, we removed the last layer of these models, known as the prediction layer. The images are fed into six various networks. Image preparation was possible using the OpenCV module and the deep learning pipelines were implemented in Python using Tensorflow and Keras. Model training and evaluation were performed on an NVIDIA Tesla T4 GPU with 16GB memory. After extracting features from each of these networks, the extracted and grouped features enter into a noble gradient boosting algorithm. CatBoost and LightGBM are two state-of-the-art algorithms that classify data into main classes. The images of the GasHisSDB dataset are divided into two sets: training (80% of the dataset) and testing (20% of the dataset). A total of 245,196 samples were used for training and testing, with 196,155 samples used for training and 49,041 samples used for testing.

### 3.3. Data Pre-processing

To prepare data for doing some operations, we understand that data pre-processing plays an essential role in data science issues. Valid data analytics are built on top of data pre-processing. Given the inherent complexity of building operations and flaws in data quality, it is a crucial stage in establishing operational data analysis [26].

The data pre-processing task in this study involves resizing images to a specific dimension. Since we are supposed to feed data into the networks, the size of images should be  $224 \times 224 \times 3$  pixels. The method we utilized to do the scaling better without losing special information is linear interpolation known as bilinear interpolation. This method calculates

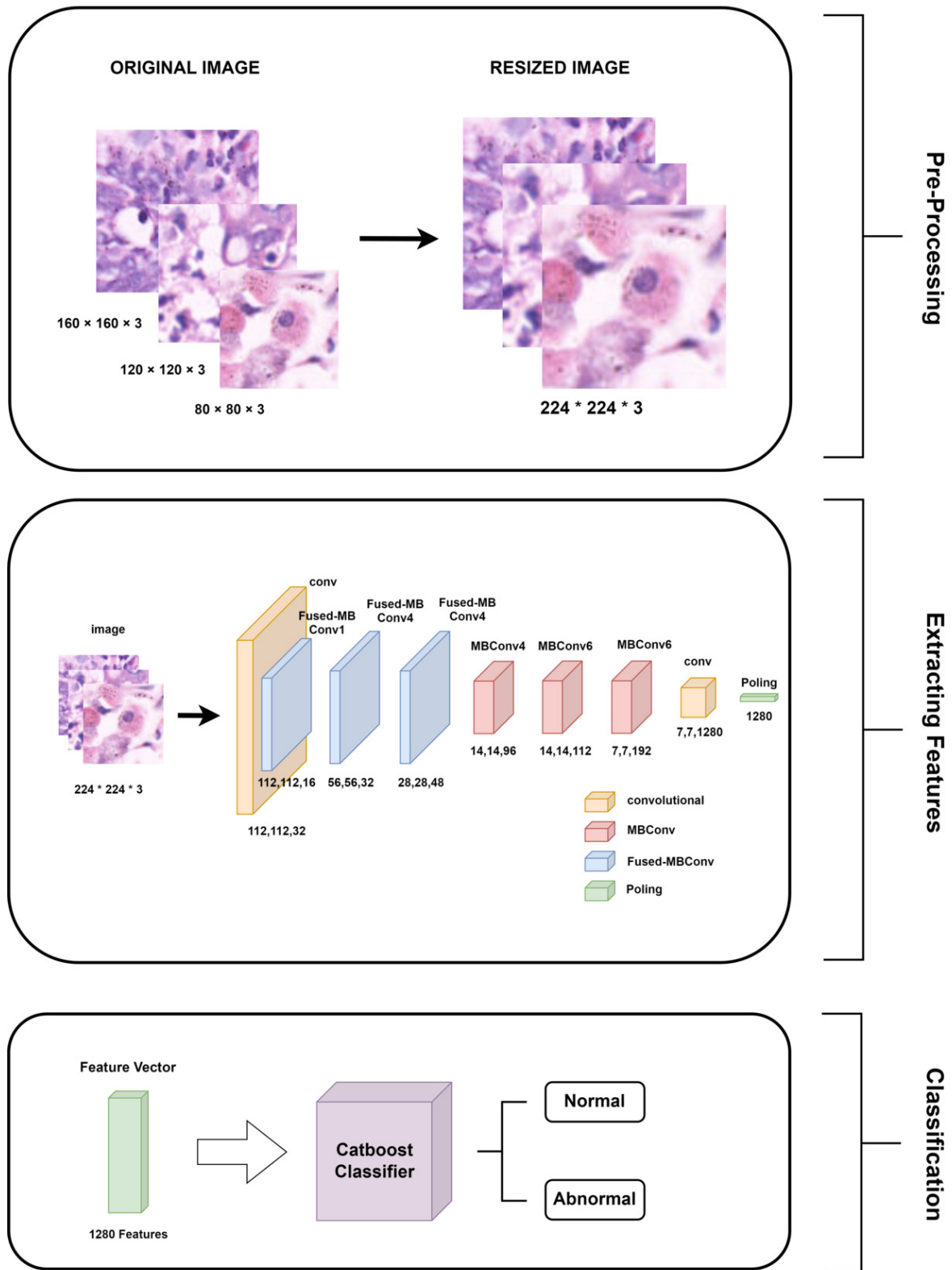
the output pixel values when transforming an input image to a new size. For each output pixel location, bilinear interpolation identifies the nearest  $2 \times 2$  neighborhood of pixels in the input image. Based on the weighted distances of the output pixel from each of these 4 neighboring pixels, their color values are blended together using linear interpolation to compute the output pixel color. Bilinear interpolation provides a good balance between resizing performance and computational complexity for most applications such as medical usage.

Although data pre-processing describes a collection of methods for improving the original data quality [26], in this study data pre-processing only involves resizing the images using the bilinear interpolation technique.

### 3.4. Feature Extractor

As mentioned before, to transform a Feature extractor, removing the last layer of the network is necessary. So it provided the possibility to use features extracted by a network with a large dataset and a different classification method in various models [27]. In the Definition of feature extraction, as taken after: "Feature extraction is generally used to mean the construction of linear combinations  $Tx$  of continuous features which have good discriminatory power between classes" [28]. Feature extraction is the method of converting the information into numerical values, which may be an exceptionally vital step for identification and visualization purposes [29]. This method performs a few changes of unique highlights to create other highlights that are more critical [30].

We used six profound learning pre-trained models for information extraction, including ResNet50, VGG16, EfficientNetV2B0, InceptionV3, Resnet101, and DenseNet201. Two fundamental parameters ought to be initialized: sort of the pooling and barring the final layer. First, we briefly explain the network that achieved higher accuracy from the dataset and our proposed model obtained from this network



**Figure 2:** The architecture of the proposed model. (The original and resized images are extracted from the GasHisSDB dataset [25])

**EfficientNetV2:** This model takes 224×224 RGB pictures as input. EfficientNetV2 is a modern family of convolutional systems that have faster training speed and superior parameter efficiency than past models [31]. Training-aware neural design search and scaling have been utilized in the improvement to optimize preparing speed and parameter efficiency for this model mutually. It employs depthwise separable convolutions which apply a single convolutional filter per input channel, reducing computations compared to standard conv layers. This is followed by a 1x1 pointwise convolution to integrate cross-channel information

In CNN, pooling layers are included primarily for down-inspecting the highlight maps by accumulating highlights from neighborhood locales. Pooling can help CNN to memorize invariant highlights and diminish computational complexity [32]. The max and the average pooling are the broadly utilized Pooling [32]. The average pooling may be more fitting for a few other scenarios, e.g., classifying unusual pictures from ordinary ones where the variation from the norm spreads all over the irregular image [32]. So, the average pooling method was chosen for all six pretrained models. The number of features obtained before and after setting the pooling method based on each network is shown in Table 2.

At that point, the final layer which was arranged to foresee dataset classes, was evacuated. To store features extracted in memory at first, we used a regular array to store the extracted features in the memory. Still, due to the memory limitation and the considerable number of dataset images, the memory capacity needed to be increased to save all the features extracted from the relevant network, so the NumPy library was used to store the extracted features in memory due to the memory limitations and the large number of dataset images. Also, since this dataset contains many images, we needed a lot of time to extract the features, so the extracted features were saved in text files every 5,000 features to manage the large number of images and the extraction time.

While this work employed fixed pre-trained parameters for the deep learning feature extractors, further fine-tuning of the networks on medical images could potentially improve results. The pre-trained models used like EfficientNetV2B0 were initially trained on general image datasets such as ImageNet which do not contain medical images, While studies such as our investigation would demonstrate how well these pre-trained networks perform on medical image datasets like GasHisSDB dataset. Fine-tuning the models by continuing the backpropagation training on gastric histopathology images may allow the networks to learn more relevant features tailored to this medical domain which could be considered in the next studies on medical image classification.

### 3.5. Classifier

**LightGBM:** LightGBM [35], a light gradient-boosting machine, is an open-source and different gradient-boosting algorithm that employs decision trees. The LightGBM method can be used in other approaches, such as the medical

**Table 2**

The number of features obtained before and after pooling

Feature Extractor	Average pooling	Without pooling
VGG16 [33]	512	7 × 7 × 512
DenseNet201 [34]	1920	7 × 7 × 1920
ResNet50 [35]	2048	7 × 7 × 2048
ResNet101 [35]	2048	7 × 7 × 2048
InceptionV3 [36]	2048	5 × 5 × 2048
EfficientNetV2B0 [31]	1280	7 × 7 × 1280

field, to assist doctors in providing an accurate and speedy medical assessment. Since the LightGBM model grows primarily horizontally and the tree depth does not significantly rise, over-learning can be avoided, so this feature tends to get a better outcome [37]. Because LightGBM generates more complex trees than the current tree-boosting algorithms, it is a more accurate decision tree algorithm [38]. Studies show that LightGBM is 20 times faster than any other gradient-boosting method in learning because of using GOSS and EFB algorithms. According to their absolute gradient values, the training data are ranked in descending order in the GOSS method. The data amplifying is more strongly influenced by the data with more extensive gradients. Since segmented instances are available, for variance evaluation gain of data point  $j$  to point  $d$  the below formula is considered.

$$v_j(d) = \frac{1}{n} \left( \frac{\left( \sum_{x_i \in A_l} g_i + \frac{1-a}{b} \sum_{x_i \in B_l} g_i \right)^2}{n_l^j(d)} + \frac{\left( \sum_{x_i \in A_r} g_i + \frac{1-a}{b} \sum_{x_i \in B_r} g_i \right)^2}{n_r^j(d)} \right) \quad (1)$$

- $x_i$  is the  $i^{th}$  sample of the training dataset.
- $g_i$  is the negative gradient of the loss function.
- $\frac{1-a}{b}$  normalize the sum of gradients.
- $A_l = \{x_i \in A : x_{i,j} \leq d\}$ ,  $A_r = \{x_i \in A : x_{i,j} > d\}$
- $B_l = \{x_i \in B : x_{i,j} \leq d\}$ ,  $B_r = \{x_i \in B : x_{i,j} > d\}$

The features' complexity is decreased through the EFB method, which also expedites model training. The EFB approach combines multiple components into a single one. It reduces the LightGBM computing difficulty from  $O(\# \text{ data} * \# \text{ feature})$  to  $O(\# \text{ data} * \# \text{ bundle})$  which  $\# \text{ bundle}$  is much less than  $\# \text{ feature}$ . So, this leads to a new high speed without losing accuracy. The result of using these algorithms to improve LightGBM can be given from the formula below.

$$F_M(x) = \sum_{m=1}^M h_m(x) \quad (2)$$

- $M$ : the maximum number of iterations.
- $h_m(x)$ : the base decision tree.

In LightGBM, two methods (leaf-wise and level-wise) are utilized for growing decision trees, making this algorithm different from others [35]. According to [37], LightGBM brings more accurate predictions for medical classification.

**CatBoost:** CatBoost is a new gradient-boosting technique for the problem of prediction proposed by Prokhorenkova et al. (2018) and Dorogush et al. (2018) [39]. CatBoost performs differently from other gradient-boosting algorithms. This algorithm begins with ordered boosting, an adequate adjustment of gradient boosting algorithms, to overcome the issue of target leakage [39] CatBoost machine is considered one of the foremost effective machines [40]. During the last few years, it has been used for various issues within arranged frameworks such as pharmaceuticals, science, and natural chemistry [40]. In addition, CatBoost uses an ordered boosting method Instead of the traditional gradient boosting algorithm for gradient estimation **which pre-sorts training data to avoid repeatedly resorting during tree construction. Categorical features are processed using a novel permutation-driven approach to handle categoricity. These enhancements accelerate the gradient boosting process compared to XGBoost [41],[39].**

For each of the six feature extraction networks per every sub-size, we found optimal parameters for better performance of CatBoost. The values tested to obtain the best parameters are as follows: For the number of iterations, the dimension is tuned from 100 to 1000 every 100 steps; moreover, for the number of splits for the numerical features parameter, values of 5, 10, 20, 30, 50, 100, and 200 were tested, also, we tried the values of 1,3,5,10, and 100 for the regularizing to find the best possible parameter. For setting the depth parameter, the dimension is tuned from 1 to 11. For setting the minimum number of training samples in a leaf parameter, the measurement is adjusted from 1 to 100.

## 4. Results and Discussion

### 4.1. Evaluation Criteria

In the following section, we calculated five metrics to evaluate our proposed method. Mainly, we focus on accuracy, but precision, recall, specificity, and  $F_1 - Score$  are used to get a better view of the proposed method's performance. In the classification problem, if an image is predicted to its category, it is known as a positive classification, otherwise known as negative. Hence, we have TP (True-Positive), TN (True-Negative), FP (False-Positive), and FN (False-Negative). Despite having these, we are easily able to calculate the mentioned metrics which are defined as follows:

$$Accuracy = \frac{TP + TN}{TN + TP + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$F_1 - Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (7)$$

### 4.2. Results

We used the different combinations of DL models and gradient boosting methods to get better classification based on accuracy score. In this paper, we deploy six pre-trained models to extract features from images. Then, two gradient-boosting algorithms are used to do the classification duty. ResNet50, VGG16, EfficientNetV2B0, InceptionV3, Resnet101, and DenseNet201 pre-trained models are experimented with as feature extractors. All pre-trained models are based on CNN models, which have been tested before. In Table 3 below, some hyperparameter values are set after fine-tuning the LightGBM algorithm to get higher accuracy.

**Table 3**  
LightGBM hyperparameters setting

sub-size	Max number of leaves	Minimal number of data	Max number of bins	Limit the max depth	Minimal sum hessian	Accuracy (%)
160px	53	250	12	8	20	92.84
120px	82	500	6	14	25	92.26
80px	35	500	15	8	48	89.06

The values of the parameters that led to the best accuracy obtained by the Catboost algorithm are given in Table 4. In Figure 3, we illustrated the impact of varying CatBoost parameter values on the accuracy metric. The figure demonstrates a systematic progression of the parameter effects, indicating that the addition of each parameter, along with its corresponding value, results in an incremental improvement in the accuracy score.

**Table 4**  
CatBoost hyperparameters setting

sub-size	Number of iterations	L2 regularization	Depth	Border count	Number of training samples	Accuracy (%)
160px	700	3	6	5	1	93.99
120px	900	5	6	20	1	93.18
80px	900	10	8	30	1	89.72

The extracted features are classified by LightGBM and CatBoost classifiers. Table 5 presents results achieved from every combination of feature extractors with classifiers. The results are reached and compared in three cropping factors. According to the following table, the proposed method known as EfficientNetV2B0-CatBoost achieved accuracies

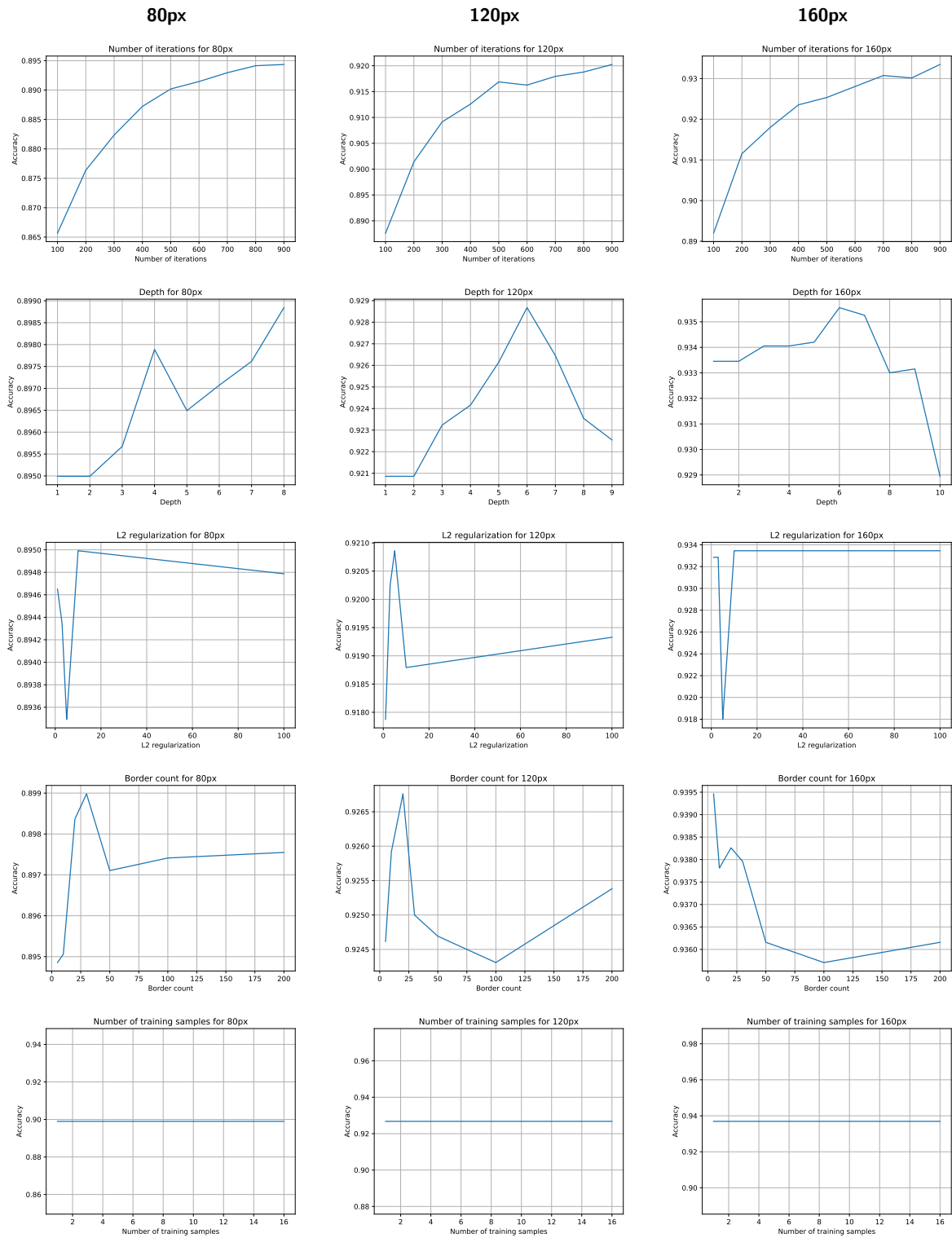


Figure 3: Parameter tuning effect on the accuracy score illustrated for each parameter according to all cropping sizes.

of 93.99%, 93.1%, and 89.7% in 160px, 120px, and 80px cropping sizes respectively.

As illustrated in Table 5, 18 experiments were conducted for this research. The EfficientNetV2B0-CatBoost achieved better results than the 11 models in the table through cropping sizes. For the 80px cropping size, the

proposed model outperformed almost all the other models by a considerable discrepancy, the only close ones being ResNet50-CatBoost with 89.4% and EfficientNetV2B0-LightGBM with 89.0%. In the 120px cropping size, the proposed model achieved the appropriate result with a 93.1% accuracy score, with EfficientNetV2B0-LightGBM as the



**Table 5**  
Binary classification accuracy score in the GasHisSDB dataset

Feature Extractor	Classifier	Results (based on accuracy (%))		
		80px	120px	160px
ResNet50	CatBoost	89.4	90.4	91.8
	LightGBM	88.8	90.1	91.4
VGG16	CatBoost	87.9	89.8	90.2
	LightGBM	87.4	89.1	89.4
EfficientNetV2B0	CatBoost	<b>89.7</b>	<b>93.1</b>	<b>93.9</b>
	LightGBM	89.0	92.2	92.8
InceptionV3	CatBoost	72.0	73.5	71.8
	LightGBM	71.9	73.6	72.6
ResNet101	CatBoost	88.9	91.5	92.2
	LightGBM	88.0	90.5	91.4
DenseNet201	CatBoost	88.0	89.9	90.6
	LightGBM	88.0	89.6	90.4

second-best model; The best individual result was obtained for the 160px cropping size by the proposed model. For this size, the second-best accuracy is related to the same network by feeding the extracted features to the LightGBM Classifier, which achieved an accuracy of 92.8%. InceptionV3 network achieved the worst results of all our experiments after feeding into both Classifiers. The average accuracies of this network differ by approximately 16% from the others.

The proposed hybrid framework leverages a convolutional neural network (EfficientNetV2B0) for automated feature extraction followed by an ensemble method (CatBoost) for classification. This pipeline was selected to balance accuracy and computational efficiency. EfficientNetV2B0 employs optimizations like depthwise separable convolutions to reduce the number of parameters and Floating-point operations per second (FLOPS) compared to standard convolutional layers [42]. Meanwhile, CatBoost uses algorithmic improvements like ordered boosting to accelerate decision tree training. By combining an efficient deep CNN with a high-speed boosting algorithm, the resulting model aims to deliver state-of-the-art accuracy without excessive complexity during training or inference.

As a result, considering that the best accuracies in all three sizes were related to the EfficientNetV2B0-CatBoost, additional evaluation metrics were calculated alongside accuracy.

Based on the findings in Table 6, the proposed method in 160px size has better evaluation metrics than other cropping sizes. This finding provides a compelling explanation for the higher accuracy achieved with this size. As indicated by the data in Table 6, the proposed method has high precision besides high accuracy, indicating consistent and accurate predictions across repeated measurements. The high  $F_1$  –  $Score$  achieved by the proposed model demonstrates the balanced performance of recall and precision, particularly in handling the imbalanced dataset.

To gain further insights, we generated a Grad-CAM heatmap using the proposed model. In Grad-CAM, we would like to preserve the spatial position information of objects lost in a fully connected layer. The last convolutional layer is used because its neurons identify the class type we are not interested in. Using images generated by Grad-CAM,

**Table 6**  
Evaluation criteria for the proposed method

Cropping size	Accuracy	Precision	Recall	Specificity	F1-Score
80px	89.72	0.87	0.86	0.91	0.87
120px	93.18	0.92	0.89	<b>0.95</b>	0.91
160px	<b>93.99</b>	<b>0.93</b>	<b>0.91</b>	<b>0.95</b>	<b>0.92</b>
Average	92.29	0.90	0.88	0.93	0.90

we obtained a pre-trained model and dataset to see what the model learned about the abnormal and normal classes for all three crop sizes. This analysis reveals the regions of the image that the model pays more attention to and the regions that receive less emphasis. Additional detailed outputs from the GradCAM algorithm can be found in Table 10.

In Table 7 and Table 8, we gathered the times taken to complete each network which is combined with the CatBoost and LightGBM classifier. Also, the size of the file containing the extracted features is presented. From the provided timing data, it's evident that the training and parameter tuning times for each network architecture and image size vary significantly between LightGBM and CatBoost models. For LightGBM, it's noticeable that the training and extraction times tend to increase as the image size and model complexity grow. This is particularly evident with ResNet101 and DenseNet201 models, as their larger file sizes and deeper architectures result in longer training and feature extraction times. Additionally, the parameter exploration times differ considerably across the various image sizes for each network, suggesting that the impact of image size on model performance is non-trivial with LightGBM.

As for CatBoost, the training and parameter tuning times display a similar trend to LightGBM, where larger image sizes and more complex models lead to longer training and feature extraction durations. Notably, the impact of image size appears to be more pronounced for CatBoost compared to LightGBM, with substantial disparities in execution times between the different image sizes. This emphasizes the significance of image size in influencing the computational workload and time efficiency of the training and tuning processes for CatBoost models. Overall, the data underscores the importance of taking image size and model complexity into account when evaluating and optimizing the performance of LightGBM and CatBoost models.

### 4.3. Discussion

Throughout our experiments, we obtain a novel method to predict 148,120 images. The proposed method, As evident from Figure 1, combines EfficientNetV2B0 as a feature extractor and CatBoost as a classifier. Figure 2 illustrates the architecture of the final model. Table 11 presents a comparison of the accuracy achieved by our proposed method with other relevant papers.

A key advantage of the proposed approach is the balance between predictive performance and computational complexity. The use of an optimized deep CNN architecture for feature extraction avoids the high overhead of larger models like VGG16 while still capturing meaningful representations

**Table 7**

Time spent and the file size of each network combined with CatBoost (specified for each parameter).

Network	Image Size	File Size	Extracting Feature	Iteration	L2 Leaf Reg	Depth	Border Count	Min data in one leaf	Total
ResNet50	160px	533.4 MB	11193s	925s	900s	4269s	1235s	13809s	32331s
ResNet50	120px	1.38 GB	18515s	1569s	1245s	2659s	2044s	2541s	28573s
ResNet50	80px	1.4 GB	33579s	3081s	2502s	1386s	3620s	5812s	49980s
EfficientNetV2B0	160px	476.9 MB	10950s	530s	458s	7047s	1142s	6235s	26362s
EfficientNetV2B0	120px	1.54 GB	33682s	1010s	841s	1155s	794s	2475s	39957s
EfficientNetV2B0	80px	2.29 GB	45235s	1940s	1676s	2315s	6096s	8962s	66224s
ResNet101	160px	889.3 MB	23145s	777s	938s	6113s	306s	19868s	51147s
ResNet101	120px	1.38 GB	29345s	1610s	1348s	3125s	2544s	13256s	51228s
ResNet101	80px	3.01 GB	63900s	3143s	2653s	4683s	4695s	22526s	101600s
DenseNet201	160px	598.2 MB	12741s	905s	769s	11849s	1454s	1480s	29198s
DenseNet201	120px	1.75 GB	37270s	1106s	260s	5862s	599s	2653s	47750s
DenseNet201	80px	3.26 GB	69438s	2786s	2315s	2568s	6823s	5876s	89806s
inceptionV3	160px	598.2 MB	12737s	684s	355s	4623s	897s	6960s	26256s
inceptionV3	120px	2.13 GB	45368s	113s	581s	650s	1291s	2315s	50318s
inceptionV3	80px	4.73 GB	104777s	2941s	2426s	1236s	3127s	1941s	116448s
VGG16	160px	140.6 MB	2982s	274s	180s	1398s	458s	8659s	13951s
VGG16	120px	434.4 MB	8943s	436s	360s	3972s	759s	10423s	24893s
VGG16	80px	903 MB	19248s	653s	546s	4181s	1028s	6853s	32509s

**Table 8**

Time spent and the file size of each network combined with LightGBM (specified for each parameter).

Network	Image Size	File Size	Extracting Feature	Number of Leaves	Min Child Samples	Max Bin	Max Depth	Min child Weight	Total
ResNet50	160px	533.4 MB	11352s	8462s	320s	67s	2532s	1026s	23759s
ResNet50	120px	1.38 GB	29436s	352s	556s	537s	1066s	2317s	34264s
ResNet50	80px	1.4 GB	29820s	25880s	1920s	1528s	3014s	7223s	69385s
EfficientNetV2B0	160px	476.9 MB	10351s	8753s	243s	183s	387s	806s	20723s
EfficientNetV2B0	120px	1.54 GB	32802s	13292s	711s	486s	1014s	2336s	50641s
EfficientNetV2B0	80px	2.29 GB	48777s	7477s	807s	801s	1833s	3989s	63684s
ResNet101	160px	889.3 MB	18935s	4858s	409s	186s	450s	1172s	26010s
ResNet101	120px	1.38 GB	28968s	6892s	1256s	304s	1490s	2379s	41289s
ResNet101	80px	3.01 GB	63921s	13560s	1316s	1560s	2701s	6256s	89314s
DenseNet201	160px	598.2 MB	12737s	8617s	455s	307s	639s	1236s	23991s
DenseNet201	120px	1.75 GB	37275s	1486s	499s	641s	885s	2486s	43272s
DenseNet201	80px	3.26 GB	69438s	27840s	1727s	861s	1258s	5876s	107000s
inceptionV3	160px	598.2 MB	12737s	12673s	297s	241s	417s	1001s	27366s
inceptionV3	120px	2.13 GB	45369s	1514s	1282s	4502s	648s	2658s	55973s
inceptionV3	80px	4.73 GB	2994s	4546s	1158s	1208s	2539s	4727s	17172s
VGG16	160px	140.6 MB	2994s	3496s	104s	80s	174s	373s	7221s
VGG16	120px	434.4 MB	9252s	7017s	207s	161s	341s	878s	17856s
VGG16	80px	903 MB	19248s	8141s	350s	321s	606s	1464s	30130s

of the histology images. Meanwhile, CatBoost incrementally builds an ensemble classifier through boosted decision trees in a faster, more efficient manner compared to contemporary gradient boosting techniques like XGBoost.

In Figure 4 an experiment was done to evaluate our proposed method. In this test, we utilized the Receiver Operating Characteristic (ROC) curve on the proposed method on all cropping sizes. ROC curves are appropriate for evaluating the model's performance. ROC evaluates classification performance at various threshold values such as Area Under the ROC Curve (AUC) which shows a summary of the ROC curve that measures a binary classifier's capacity to

discriminate between classes. In a nutshell, the higher the ROC curve, the better our model will perform.

The experiment visualized in Figure 5 demonstrates that the training curve stabilizes after approximately 600 iterations and remains stable from 700 to 800 iterations. It is worth noting that the validation curve exhibits a lesser decrease compared to the training curve due to the larger amount of data included in these groups. The more data embedded into a set, the more it learns based on errors. For better understanding, this experiment was done on all cropping sizes in Figure 5.

**Table 9**

Model Performance: Robust cross-validation scores at 160px and 120px cropping sizes, with  $k=5$  and  $k=9$ , show strong predictive abilities and effective generalization potential.

	k = 5			k = 9		
	Mean CV Score	Max CV Score	Avg CV Score	Mean CV Score	Max CV Score	Avg CV Score
<b>80px</b>	89.54	89.77	89.54	89.63	89.82	89.63
<b>120px</b>	92.82	93.21	92.82	92.83	93.25	92.83
<b>160px</b>	93.61	94.15	93.56	93.56	93.81	93.61

In order to evaluate our proposed framework, we implement K-Fold cross-validation to be confident about our model performance. As obvious in Table 9, the model demonstrates consistently strong cross-validation performance at 80px, 120px, and 160px cropping sizes. For 160px, with 5 and 9 splits, scores range from 93.56% to 94.15%, and the mean score of 93.61% highlights the model's robust predictive abilities and proficiency in generalization. At 120px, the model achieves slightly lower scores (92.82%) but maintains strong consistency and reliability for 5 and 9 splits, and finally in the case of an 80px cropping size, employing 5 and 9 splits reveals scores within the range of 89.54% to 89.84%. The computed mean score of 89.77% emphasizes the model's resilient predictive capabilities and aptitude for effective generalization. Overall, All cropping sizes exhibit strong cross-validation scores, affirming the model's capacity for accurate predictions and effective generalization to diverse data subsets.

Also, We employed the t-distributed stochastic neighbor embedding (t-SNE) technique to visualize the high-dimensional data in a 2-dimensional space. Similar objects are represented by neighboring points, while dissimilar ones are depicted as distant points. Figure 6 depicts the application of the T-SNE algorithm on a sample of 300 data points, consisting of two main classes: normal and abnormal. We utilized this technique both before and after feature extraction using the EfficientNetV2 network. The results show how significantly objects are clustered and moved towards the ones with the same class. The primary objective of the T-SNE algorithm is to visually represent data while preserving the distances between objects based on various metrics. In this study, we employed the Manhattan metric to illustrate the distances between points. Figure 7 presents the Confusion Matrix for each cropping size of the two classes.

In Table 5, we conclude that our proposed method with a high achieved accuracy of 93.9% with EfficientNetV2B0 + CatBoost framework. this conclusion could be followed by some ambiguous questions about whether this framework works well on other datasets or even how this model prevents overfitting and provides hand-crafted results.

While this work focused on the GasHisSDB dataset for model development and evaluation, testing the proposed EfficientNetV2B0 + CatBoost framework on additional gastric histopathology image datasets could further validate its generalizability. As a highly optimized convolutional neural network architecture, EfficientNetV2B0 may transfer well to other datasets given its strong representation learning

capabilities from a diversity of training data. However, factors like differences in image resolution, staining techniques, and capture equipment could impact results. Applying our approach to multiple external datasets and comparing performance would provide useful insights into the broader applicability of this model. The proposed combination of EfficientNetV2B0 for feature extraction and CatBoost for classification outperformed other model architectures in our experiments, which may be attributed to two key factors. First, EfficientNetV2B0 leverages a cutting-edge CNN design tuned specifically for high accuracy and efficiency through automated architecture search and scaling. Second, CatBoost is an advanced gradient-boosting algorithm that effectively handles categorical features and prevents overfitting which is clear in 5 that no early-stopping is needed. Together, the deep representation learning of EfficientNetV2B0 combined with the robust boosted decision trees of CatBoost offer complementary modeling strengths. The high dimensionality of the 1280 features extracted by EfficientNetV2B0 provides a rich input feature space that CatBoost can then optimize as part of its boosted tree construction process.

## 5. Limitations and Future Recommendations

In our endeavor to enhance the efficiency and reliability of gastric cancer detection for clinicians, we have chosen to address acknowledged inherent limitations and put forth future directions that resonate with the broader discourse on Explainable Artificial Intelligence (XAI). While our hybrid approach significantly enhances diagnostic accuracy, the interpretability of the intricate features extracted by deep learning models remains a challenge. In future studies, integrating XAI methodologies could facilitate a more transparent understanding of model decisions, aiding clinicians in comprehending and trusting the diagnostic outcomes.

Another challenge in this study domain is the availability and diversity of histopathological images which strongly can impact model generalization because of the variations in imaging protocols and staining techniques. Future research should explore methods to enhance the model's compatibility with diverse datasets, ensuring robust performance across different clinical settings.

Finally, the most likely problem that can obstacle our approach is constraints associated with computational demands in DL models. Investigating strategies to optimize model architectures without compromising performance is

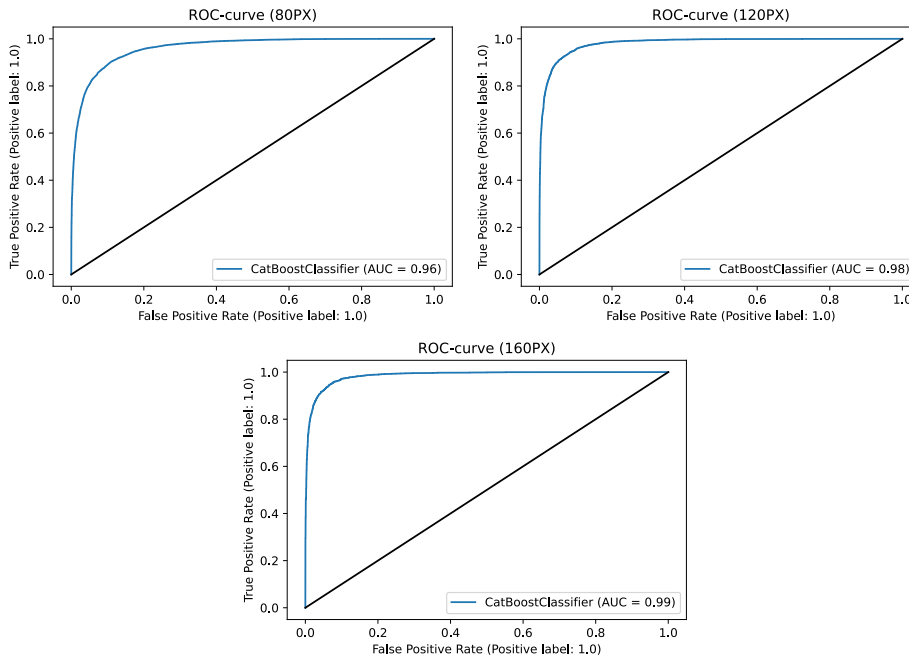


Figure 4: ROC curves of the proposed method for different cropping sizes

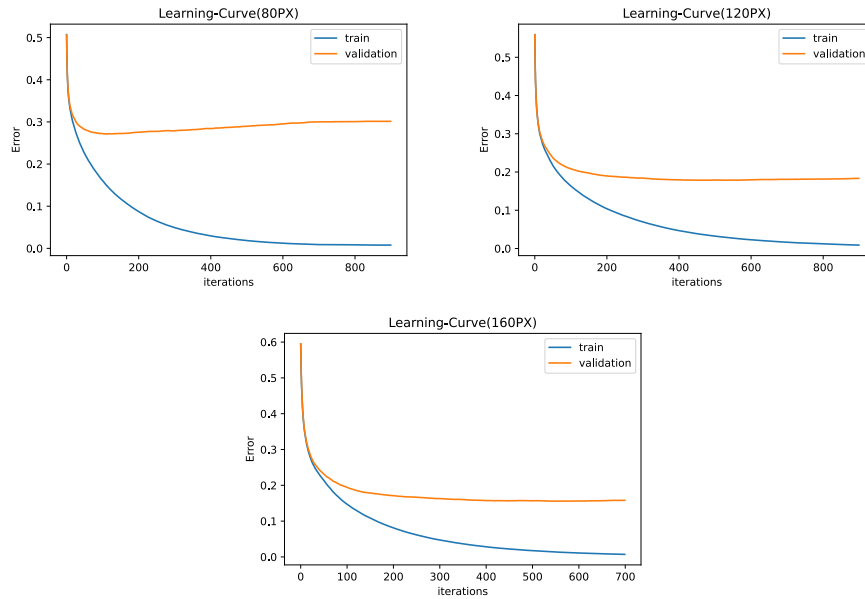


Figure 5: Learning curves of the proposed method for 80px, 120px, and 160px cropping sizes

crucial for real-world applicability. Also, collaborating with healthcare practitioners and institutions will be essential for the seamless integration of our model into clinical workflows.

## 6. Conclusion

This paper suggests a novel method to classify gastric cancer into two main classes: normal and abnormal tumors. The presented model was chosen based on accuracy.

However, other metrics such as precision, recall (sensitivity), specificity, and  $F_1 - Score$  are calculated for a better view of different angles of our model. Also, the application of cross-validation results further underscores the robust performance of our proposed algorithm for gastric cancer classification. The final method combines the EfficientNetV2B0 pre-trained model and CatBoost gradient boosting. We changed EfficientNetV2B0 as a feature extractor and set CatBoost as the appropriate classifier by choosing proper hyperparameters. We experimented with 12 combinations of

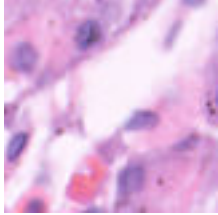
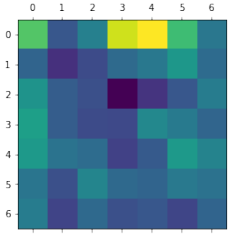

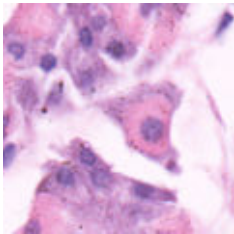
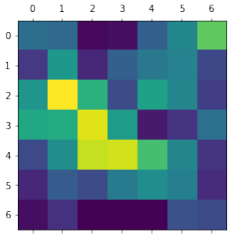
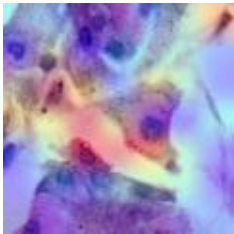
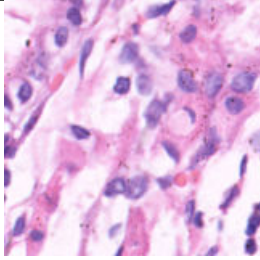
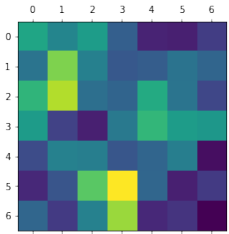
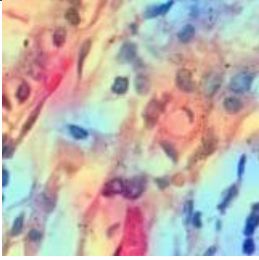
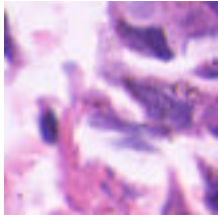
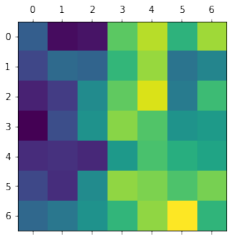

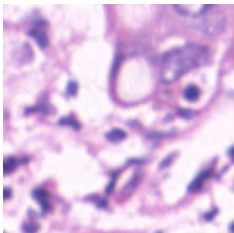
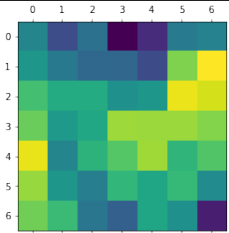
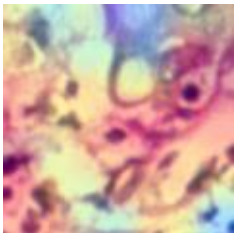
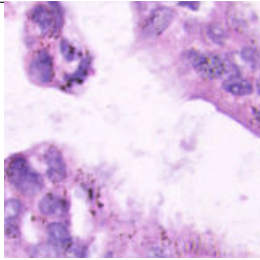
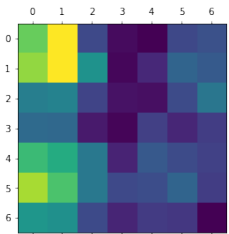
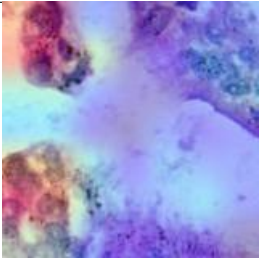
six pre-trained models and two gradient boosting algorithms in all three cropping sizes: 80px, 120px, and 160px. Through all these models, EfficientNetV2B0-CatBoost is the chosen method with the highest accuracy: 89.7%, 93.1%, and 93.9% for 80px, 120px, and 160px cropping sizes respectively. According to the results and empirical evaluations around the proposed method, our model can thoroughly diagnose gastric cancer images provided in the GasHisSDB dataset. In future works, new techniques such as augmentation will be considered **to cover the dataset limitations and availability**, and multiple classifications will be the main aim to get better accuracy. Moreover, **as discussed the limitations, we will consider integrating XAI approaches to make the model more interpretable for commercial purposes.**

## 7. Declaration of Conflicts of Interest

The authors express that they have no competing interests affecting the reported work in this article.

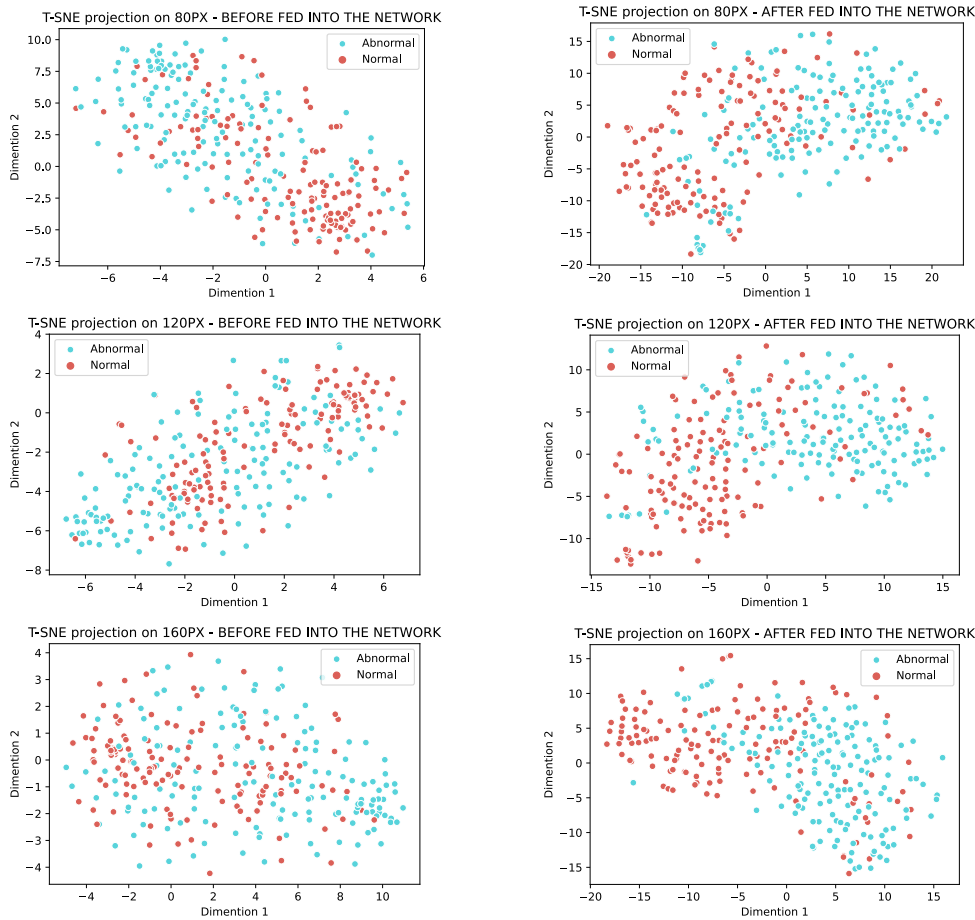
**Table 10**

Grad-CAM outputs consist of a heat map for normal and abnormal tumors in gastric cancer. (Source of original images are extracted from GasHisSDB dataset [25].)

	Size	Original image	Heat-map	Grad-CAM output
Normal	80px			
	120px			
	160px			
Abnormal	80px			
	120px			
	160px			

**Table 11**  
Performance comparison with state-of-the-art counterparts

Paper	Method	Dataset	Accuracy (%)
Li et al. (2022) [22]	HCRF + IC	700 gastric histopathology images	91.4
Wang et al. (2019) [17]	RMDL	Gastric WSI	86.5
Song et al. (2020) [19]	DeepLabv3 + ResNet50	PLAGH	87.3
Weiming et al. (2022) [23]	Color Histogram + Random Forest	GasHisSDB	85.99
Yusuke et al. (2020) [18]	CNN + ME-NBI	1492 EGC and 1078 gastritis images	85.3
Noda et al. (2022) [24]	CNN	1623 gastric cancer images	86.1
<b>Proposed Method</b>	<b>EfficientNetV2B0 + CatBoost</b>	<b>GasHisSDB</b>	<b>93.99</b>



**Figure 6:** t-SNE projection on 300 images in GasHisSDB dataset before and after extracting features by the EfficientNetV2B0 model

## 8. Data and Code Availability

A publicly available dataset, i.e., BreakHis, was used in this study, which is available at GasHisSDB<sup>1</sup>. In addition, The source code of the proposed method required to reproduce the predictions and results is available at Github<sup>2</sup>.

<sup>1</sup><https://gitee.com/neuhwm/GasHisSDB.git>

<sup>2</sup><https://github.com/danialkh/Gastric-cancerdetection>

## 9. Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## References

- [1] Arash Etemadi, Saeid Safiri, Sadaf G Sepanlou, Kevin Ikuta, Catherine Bisignano, Ramin Shakeri, Mohammad Amani, Christina Fitzmaurice, Molly Nixon, Nooshin Abbasi, et al. The global, regional, and national burden of stomach cancer in 195 countries, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The lancet Gastroenterology & hepatology*, 5(1):42–54, 2020.

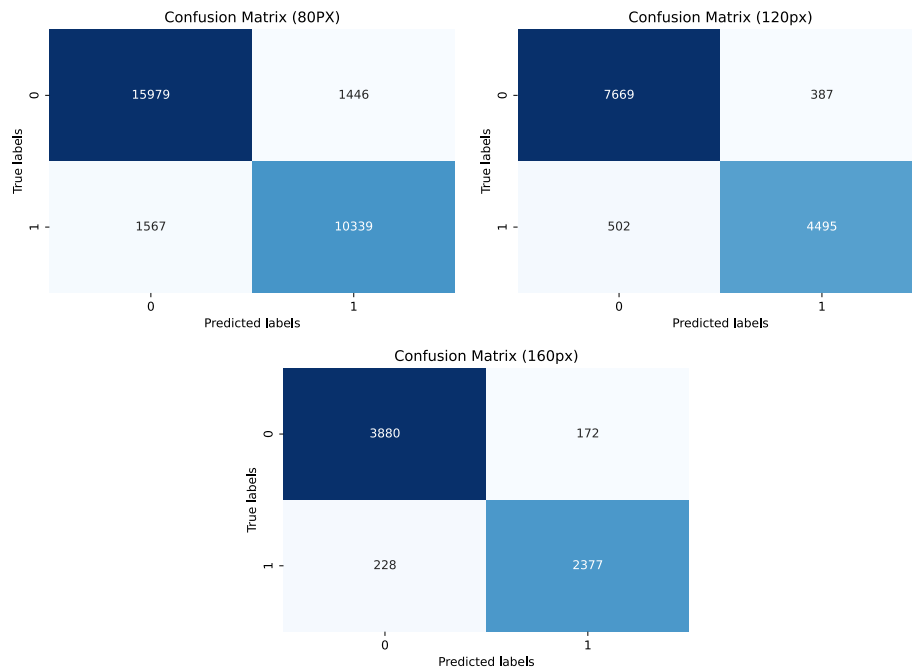


Figure 7: Confusion matrix of the proposed method in all three cropping sizes

- [2] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L Siegel, Lindsey A Torre, and Ahmedin Jemal. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6):394–424, 2018.
- [3] Vinay Kumar, Abul K Abbas, and Jon C Aster. *Robbins basic pathology e-book*. Elsevier Health Sciences, 2017.
- [4] Jing Wang and Xiuping Liu. Medical image recognition and segmentation of pathological slices of gastric cancer based on deeplab v3+ neural network. *Computer Methods and Programs in Biomedicine*, 207:106210, 2021.
- [5] Hamid Nasiri and Sharif Hasani. Automated detection of covid-19 cases from chest x-ray images using deep neural network and xgboost. *Radiography*, 28(3):732–738, 2022.
- [6] Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep learning applications in medical image analysis. *Ieee Access*, 6:9375–9389, 2017.
- [7] Xufeng Huang, Qiang Lei, Tingli Xie, Yahui Zhang, Zhen Hu, and Qi Zhou. Deep transfer convolutional neural network and extreme learning machine for lung nodule diagnosis on ct images. *Knowledge-Based Systems*, 204:106230, 2020.
- [8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.
- [9] DR Sarvamangala and Raghavendra V Kulkarni. Convolutional neural networks in medical image understanding: a survey. *Evolutionary intelligence*, 15(1):1–22, 2022.
- [10] Zhicheng Jiao, Xinbo Gao, Ying Wang, and Jie Li. A deep feature based framework for breast masses classification. *Neurocomputing*, 197:221–231, 2016.
- [11] Vicky Mudeng and Se-woon Choe. Deep neural network incorporating domain and resolution transformations model for histopathological image classification. *Computers and Electrical Engineering*, 104:108468, 2022.
- [12] Mobina Ezzoddin, Hamid Nasiri, and Morteza Dorrigiv. Diagnosis of covid-19 cases from chest x-ray images using deep neural network and lightgbm. In *2022 International Conference on Machine Vision and Image Processing (MVIP)*, pages 1–7. IEEE, 2022.
- [13] Monika Sachdeva, Alok Kumar Singh Kushwaha, et al. The power of deep learning for intelligent tumor classification systems: A review. *Computers and Electrical Engineering*, 106:108586, 2023.
- [14] Hamid Nasiri, Seyed Ali Alavi, et al. A novel framework based on deep learning and anova feature selection method for diagnosis of covid-19 cases from chest x-ray images. *Computational intelligence and neuroscience*, 2022, 2022.
- [15] Areej A Malibari, Jaber S Alzahrani, Majdy M Eltahir, Vinita Malik, Marwa Obayya, Mesfer Al Duhayyim, Aloísio V Lira Neto, and Victor Hugo C de Albuquerque. Optimal deep neural network-driven computer aided diagnosis model for skin cancer. *Computers and Electrical Engineering*, 103:108318, 2022.
- [16] Bo Liu, Kelu Yao, Mengmeng Huang, Jiahui Zhang, Yong Li, and Rong Li. Gastric pathology image recognition based on deep residual networks. In *2018 IEEE 42nd annual computer software and applications conference (COMPSAC)*, volume 2, pages 408–412. IEEE, 2018.
- [17] Shujun Wang, Yaxi Zhu, Lequan Yu, Hao Chen, Huangjing Lin, Xiangbo Wan, Xinjuan Fan, and Pheng-Ann Heng. Rmdl: Recalibrated multi-instance deep learning for whole slide gastric image classification. *Medical image analysis*, 58:101549, 2019.
- [18] Yusuke Horiuchi, Kazuharu Aoyama, Yoshitaka Tokai, Toshiaki Hirasawa, Shoichi Yoshimizu, Akiyoshi Ishiyama, Toshiyuki Yoshio, Tomohiro Tsuchida, Junko Fujisaki, and Tomohiro Tada. Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging. *Digestive diseases and sciences*, 65:1355–1363, 2020.
- [19] Zhigang Song, Shuangmei Zou, Weixun Zhou, Yong Huang, Liwei Shao, Jing Yuan, Xiangnan Gou, Wei Jin, Zhanbo Wang, Xin Chen, et al. Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning. *Nature communications*, 11(1):4294, 2020.
- [20] Osamu Iizuka, Fahdi Kanavati, Kei Kato, Michael Rambeau, Koji Arihiro, and Masayuki Tsuneki. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. *Scientific reports*, 10(1):1504, 2020.
- [21] Shiliang Ai, Chen Li, Xiaoyan Li, Tao Jiang, Marcin Grzegorzec, Changhao Sun, Md Mamunur Rahaman, Jinghua Zhang, Yudong Yao,



- and Hong Li. A state-of-the-art review for gastric histopathology image analysis approaches and future development. *BioMed Research International*, 2021, 2021.
- [22] Yixin Li, Xinran Wu, Chen Li, Xiaoyan Li, Haoyuan Chen, Changhao Sun, Md Mamunur Rahaman, Yudong Yao, Yong Zhang, and Tao Jiang. A hierarchical conditional random field-based attention mechanism approach for gastric histopathology image classification. *Applied Intelligence*, pages 1–22, 2022.
- [23] Weiming Hu, Haoyuan Chen, Wanli Liu, Xiaoyan Li, Hongzan Sun, Xinyu Huang, Marcin Grzegorzec, and Chen Li. A comparative study of gastric histopathology sub-size image classification: From linear regression to visual transformer. *arXiv preprint arXiv:2205.12843*, 2022.
- [24] Hiroto Noda, Mitsuru Kaise, Kazutoshi Higuchi, Eriko Koizumi, Keiichiro Yoshikata, Tsugumi Habu, Kumiko Kirita, Takeshi Onda, Jun Omori, Tepei Akimoto, et al. Convolutional neural network-based system for endocytoscopic diagnosis of early gastric cancer. *BMC gastroenterology*, 22(1):1–10, 2022.
- [25] Weiming Hu, Chen Li, Xiaoyan Li, Md Mamunur Rahaman, Jiquan Ma, Yong Zhang, Haoyuan Chen, Wanli Liu, Changhao Sun, Yudong Yao, et al. Gashissdb: A new gastric histopathology image dataset for computer aided diagnosis of gastric cancer. *Computers in biology and medicine*, 142:105207, 2022.
- [26] Cheng Fan, Meiling Chen, Xinghua Wang, Jiayuan Wang, and Bufu Huang. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Frontiers in Energy Research*, 9:652801, 2021.
- [27] Mohammad Reza Abbasniya, Sayed Ali Sheikholeslamzadeh, Hamid Nasiri, and Samaneh Emami. Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods. *Computers and Electrical Engineering*, 103:108382, 2022.
- [28] Samina Khalid, Tehmina Khalil, and Shamila Nasreen. A survey of feature selection and feature extraction techniques in machine learning. In *2014 science and information conference*, pages 372–378. IEEE, 2014.
- [29] Mesut Toğaçar, Burhan Ergen, Zafer Cömert, and F Özyurt. A deep feature learning model for pneumonia detection applying a combination of mrmr feature selection and machine learning models. *Irbm*, 41(4):212–222, 2020.
- [30] Yuanjie Zheng, Brian Vanderbeek, Ebenezer Daniel, Dwight Stambolian, Maureen Maguire, David Brainard, and James Gee. An automated drusen detection system for classifying age-related macular degeneration with color fundus photographs. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1448–1451. IEEE, 2013.
- [31] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.
- [32] Rajendran Nirthika, Siyamalan Manivannan, Amirthalingam Ramanan, and Ruixuan Wang. Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study. *Neural Computing and Applications*, 34(7):5321–5347, 2022.
- [33] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [34] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [35] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.
- [36] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [37] Sami Akbulut, Ipek Balıkcı Cicek, and Cemil Colak. Classification of breast cancer on the strength of potential risk factors with boosting models: A public health informatics application. *Medical Bulletin of Haseki/Haseki Tip Bulteni*, 60(3), 2022.
- [38] Hamid Nasiri, Ghazal Kheyroddin, Morteza Dorrigiv, Mona Esmaeili, Amir Raesi Nafchi, Mohsen Haji Ghorbani, and Payman Zarkesh-Ha. Classification of covid-19 in chest x-ray images using fusion of deep features and lightgbm. In *2022 IEEE World AI IoT Congress (AIoT)*, pages 201–206. IEEE, 2022.
- [39] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.
- [40] John T Hancock and Taghi M Khoshgoftaar. Catboost for big data: an interdisciplinary review. *Journal of big data*, 7(1):1–45, 2020.
- [41] S Chehreh Chelgani, H Nasiri, A Tohry, and HR Heidari. Modeling industrial hydrocyclone operational variables by shap-catboost-a “conscious lab” approach. *Powder Technology*, 420:118416, 2023.
- [42] Yunho Jeon and Junmo Kim. Constructing fast network through deconstruction of convolution. *Advances in neural information processing systems*, 31, 2018.