Scientific Research Publishing

Journal of Computer and Communications, 2022, 10, 91-106
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

LGFFN-GHI: A Local-Global Feature Fuse Network for Gastric Histopathological Image Classification

Songsong Li, Wenzhong Liu

Sichuan University of Light Chemical Industry, Zigong, China
Email: 1036576676@qq.com

How to cite this paper: Li, S.S. and Liu, W.Z. (2022) LGFFN-GHI: A Local-Global Feature Fuse Network for Gastric Histopathological Image Classification. *Journal of Computer and Communications*, **10**, 91-106.
https://doi.org/10.4236/jcc.2022.1011007

Received: October 11, 2022
Accepted: November 18, 2022
Published: November 21, 2022

Abstract

Gastric cancer remains the third most common cause of cancer-related death. Histopathological examination of gastric cancer is the gold standard for the diagnosis of gastric cancer. However, manual pathology examination is time-consuming and laborious. Computer-aided diagnosis (CAD) systems can assist pathologists in diagnosing pathological images, thus improving the efficiency of disease diagnosis. In this paper, we propose a two-branch network named LGFFN-GHI, which can classify histopathological images of gastric cancer into two categories: normal and abnormal. LGFFN-GHI consists of two parallel networks, ResNet18 and Pvt-Tiny, which extract local and global features of microscopic gastric tissue images, respectively. We propose a feature blending module (FFB) that fuses local and global features at the same resolution in a cross-attention manner. This enables ResNet18 to acquire the global features extracted by Pvt-Tiny, while enabling Pvt-Tiny to acquire the local features extracted by ResNet18. We conducted experiments on a novel publicly available sub-size image database of gastric histopathology (GasHisSDB). The experimental results show that LGFFN-GHI achieves an accuracy of 96.814%, which is 2.388% and 3.918% better than the baseline methods ResNet18 and Pvt-Tiny, respectively. Our proposed network exhibits high classification performance, demonstrating its effectiveness and future potential for the gastric histopathology image classification (GHIC) task.

Keywords

Deep Learning, Gastric Cancer Screening, Histopathological Images, Transformer

1. Introduction

Gastric cancer is a global health problem, with more than 1 million people newly diagnosed with the disease each year worldwide. Despite a global decline in incidence and mortality over the past 50 years, gastric cancer remains the third leading cause of cancer-related death [1]. Gastric cancer remains a globally important cancer, According to estimates from the GLOBOCAN project of the International Agency for Research on Cancer [2], more than 1 million new cases and an estimated 769,000 deaths occur in 2020 (equivalent to 1 in 13 deaths worldwide), ranking fifth in incidence and fourth in mortality worldwide. Most gastric cancers are diagnosed at an advanced stage because they have latent and non-specific symptoms that lead to a poor prognosis. It has been reported that early and accurate detection of gastric cancer can improve the 5-year survival rate of patients by about 90% [3] [4]. However, the diagnosis of early gastric cancer is largely limited by the number of experienced imaging specialists. In addition, diagnostic accuracy depends heavily on the clinical experience of the specialist and is susceptible to a variety of factors. Qualified specialists are also unlikely to avoid all misdiagnoses and missed diagnoses [5]. The traditional method of gastric cancer diagnosis is to identify the morphological features of malignant cells by histopathological biopsy specimens, while manual pathological examination of gastric sections is time-consuming and laborious [5].

Recent advances in machine learning and image processing have enabled the development of CAD systems for faster detection and diagnosis of gastric cancer from histopathological images. CAD systems analyze histopathological images of sample tissues to identify histopathological patterns corresponding to cancerous and non-cancerous conditions and classify histopathological images as benign and malignant, respectively. The main challenge in classifying histopathological images of gastric cancer is the inherent complexity of histopathological images, such as cell overlap, subtle differences between images, and uneven color distribution. Recently developed deep learning methods using hematoxylin and eosin (H&E) stained whole section images have shown the potential to rapidly detect adenocarcinomas in gastric biopsy and resection specimens with relatively high sensitivity and specificity, which can support future diagnostic pathology workflows as well as further analysis by accurately segmenting cancer regions [6].

The aim of this study is to develop an accurate and reliable classification model for histopathological images of gastric cancer. The deep learning classification model we developed consists of three main parts, two parallel branches and an information exchange module between them. First, using ResNet18 [7] branch to extract the local information of gastric tissue image. Then, the PVT-Tiny [8] branch is used to extract the global information describing the gastric tissue image. Finally, we propose the bidirectional cross-attention module that fuses local and global information, which enables the ResNet18 branch to acquire the global information extracted by the PVT-Tiny branch, and enables the PVT-Tiny branch to acquire the local information extracted by the ResNet18 branch.

This paper is organized as follows: Section 2 presents the literature review. Section 3 introduces the classification model based on LGFFN-GHI for gastric tissue images. Section 4 presents the data and experimental setup. Section 5 compares the performance of our model with several classical deep learning models. Section 6 gives the conclusion of this paper.

2. Related Works

In the study of [9], the authors compared three aspects of feature extraction, feature dimensionality reduction, and classifier for histopathological images of gastric cancer. In the classifier, this work compares random deep forest (RF) and artificial neural network (ANN). ANN classifier outperforms RF classifier. In the study of [10], the authors proposed a method to classify histopathological images of gastric cancer by nuclear attribute relationship map. The images are pre-analyzed and the cell nuclei are first segmented, followed by selective nuclear classification. Based on the classification, different types of nuclei were constructed into cell relationship maps, and features were extracted for each cell relationship map. A total of 332 feature vectors, including mean, variance, skewness and kurtosis, were extracted based on the features of the maps. Finally, RF was used for classification. In the study of [11], the authors proposed three deep learning classification algorithms for gastric cancer histopathology. The first set of experiments was classified by convolutional neural network (CNN) method. The accuracy of the classification results was 86.4%. In the second set of experiments, features were extracted by CNN and then classified by RBF kernel support vector machine. The accuracy of the classification result is 89.2%. The third set of experiments uses K-SVD to learn the features extracted by CNN to get the complete dictionary, and then performs sparse decomposition. The classification was performed using a linear kernel support vector machine, and the classification accuracy was 95%.

In the study of [12], a nine-layer deep convolutional neural network (DCNN) is proposed, which consists of three convolutional layers, three maximum pooling layers, two fully connected layers, and one output layer. This work yielded an accuracy of 96.88%. In the study of [13], the authors designed a supervised feed-forward CNN model. The network classified tumor and necrotic regions with 69.9% and 81.4% accuracy, respectively. In addition, several comparative experiments were conducted in this work. AlexNet was used for deep learning, color and texture features were used for machine learning, and RF was used for classification. In the study of [14], the authors propose a new deep learning network-based model for classification of histopathological images of gastric cancer. To extract deep features, the proposed deep learning network has different structures, a shallow multiscale module and a deep network module. Several comparison experiments were performed, such as AlexNet, VGG-16, ResNet-50, ResNet-101, Inception-V4, and DenseNet-121. After comparison, the proposed network achieved good results. For the patch level, the classification accuracy of the model is 97.93%. For the slice level, the classification accuracy of the model is 100%. In the study of [15], the authors constructed a 50-layer residual network model.

In this model, multi-size convolutional kernels are used to extract features, and after extensive training, the network achieves an output F-score of 95.5%. In the study of [16], a feature balancing module (FBM) is proposed that can distinguish subtle differences in images. The balancing module has two types of channels: the first is a channel attention (CA) module and the second is a spatial attention (SA) module. In the study of [17], the authors propose a ten-layer convolutional neural network, in which three convolutional layers extract features, four pooling layers reduce the image size, and three fully connected layers output feature values. In the study of [18], the authors propose a recalibration-based multi-instance deep learning method for histopathological image classification of gastric cancer. In this method, two convolutional layers and one pooling layer are added to transform the ResNet-v2 network into a full network model. The pooling layer is averaged pooling with two convolutional layers: one for feature extraction and the other for classification. In the study of [19], the authors fused the two networks, DeepLab-V3 and ResNet-50, and introduced the structure of the ResNet-50 network into DeepLab-V3 and built a new convolutional neural network on top of DeepLab-V3. In this work, 2,166 complete slices were selected as the training set and 300 slices as the test set. After extensive training, the final accuracy of the model is 87.3%, sensitivity is 99.6%, and specificity is 84.3%. In the study of [20], the authors propose a multi-scale deep learning network, in which images with different magnifications are selected from the whole WSI images, patches of the same size are extracted from the images with different magnifications, and then these patches are put into the deep HIPO. Then, this multi-scale deep learning network can learn images of multiple scales. In the study of [21], the authors used the standard Inception-V3 network framework. By changing the depth multiplier, the parameters can be reduced. To increase the robustness of the images, data enhancement methods such as mirroring and rotation are used. The Adam optimization algorithm is used to optimize the network. After extensive training, the network model with the lowest validation error was selected. In the study of [22], the authors propose three classical convolutional neural networks for image classification: AlexNet, ResNet-50 and Inception-V3. In the data selection, ten-fold cross validation is used to test the performance of the classification. The data is divided into ten parts, train:validation:test = 8:1:1. For each combination, the classification results of the three classical networks are obtained. Finally, the ten results are a series of outputs that are used to calculate accuracy, sensitivity and specificity. In the study of [23], an intelligent attention mechanism (HCRF-AM) model based on hierarchical conditional random fields is proposed in. The HCRF-AM model consists of an attention mechanism (AM) module and an image classification (IC) module. In the AM module, the HCRF model is built to extract attention regions. In the IC module, a convolutional neural network model is trained using the selected attention regions, and then an integrated learning algorithm for classification probability is used to obtain image-level results from the patch-level output of the CNN. In addition, the AM module and migration learning techniques enable the network to generalize well to other types of image data in addition to histopathological images. In the study of [24], a novel multi-instance classification framework based on graph convolutional network (GCN) is proposed for gastric microscopy image classification. First, patch embeddings are generated by feature extraction. Then, graph structure is introduced to model the spatial topological structure relationship between instances. In addition, a graph classification model with hierarchical pooling is constructed to implement this multi-instance classification task.

Previous studies have used convolutional neural network to recognize pathological images of gastric cancer. CNN model is the main type of deep learning, which can be applied to many machine vision tasks. CNN model also has some shortcomings, one of which is that CNN model can not handle global information well. In contrast, the vision transformer (ViT) model used in the field of computer vision can extract more abundant global information. In medicine, the composition of histopathological images is complex. The proportion of abnormal areas in some abnormal images is large, while the proportion of abnormal areas in some abnormal slices is small. Therefore, the model used for the classification task of tissue pathological images must have a strong ability to comprehensively extract global and local information. Considering the actual situation of CNN and ViT models, we heuristic proposed a hybrid model for GHIC tasks, namely LGFFN-GHI, which integrates local information and global information into an organic whole (**Figure 1**).

3. Methodology

3.1. CNN Branch: ResNet18

The branch of the LGFFN-GHI model used to extract local features from gastric tissue images is ResNet18 [7]. ResNet18 consists of a pyramid model with a backbone block stacked with four stages of residual blocks, and the size of the feature map output at each stage is halved while the number of channels is doubled.

3.2. Transformer Branch: PVT-Tiny

Since Vision Transformer (ViT) [25] was introduced to image classification tasks in 2021 and has been successful. There has been a lot of work devoted to various variants of ViT [8] [26]-[33] for computer vision tasks such as image classification, target detection, image segmentation, etc. PVT (Pyramid Vision Transformer) [8] is one of the ViT variants for intensive prediction tasks such as image detection and image segmentation. PVT introduces a pyramid structure similar to CNN compared to ViT, making PVT apply as a backbone like CNN for intensive prediction tasks, such as segmentation and detection. Similar to the CNN backbone [7] similarly, PVT has four stages to generate feature maps at different scales. All stages share a similar architecture, which consists of a patch embedding layer and L_i Transformer encoder layers.

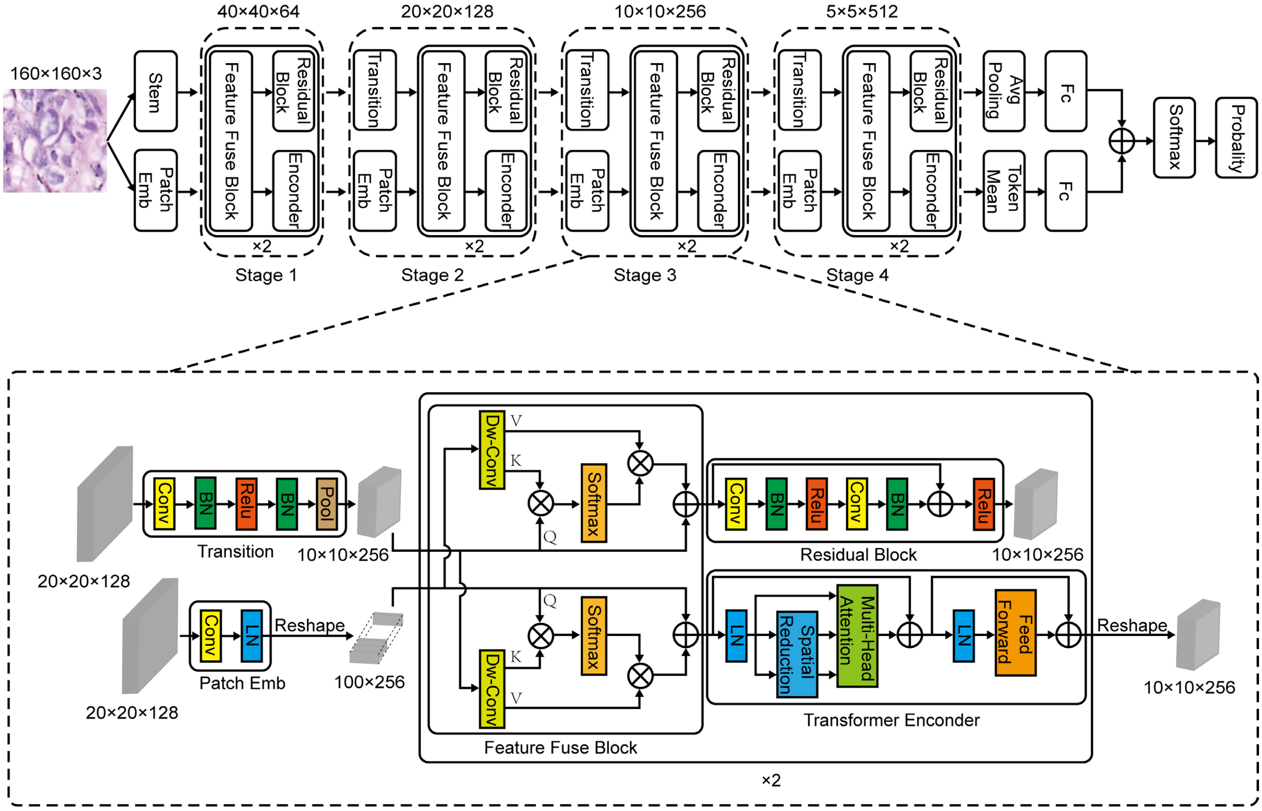

Figure 1. Local-global feature fuse network (LGFFN-GHI). The upper row shows the overall structure of LGFFN-GHI. The bottom row shows the specific implementation of the third phase.

In the first stage, given an input image of size $H \times W \times 3$, it is first divided into $\frac{HW}{4^2}$ patches, each of size is $4 \times 4 \times 3$. Then, the flattened patches are input into the linear projection to obtain patches of size $\frac{HW}{4^2} \times C_1$ of the embedded patches. Then, the embedded patches are passed together with the positional embedding through a Transformer encoder with L_1 layers, and the output is reconstructed as a feature map of size $\frac{H}{4} \times \frac{W}{4} \times C_1$ feature map of size F_1. Similarly, using the feature map from the previous stage as input, the following feature map is obtained: F_2, F_3, F_4, which are in steps of 8, 16, 32 pixels with respect to the input image, respectively.

The Transformer encoder for stage i has L_i encoder layers, each consisting of an attention layer and a feedforward layer [34]. Since PVT needs to handle high-resolution (e.g., 4-stride) feature mapping, a space-reduction attention (SRA) layer is proposed to replace the traditional multi-head attention (MHA) layer in the encoder [34]. The SRA is similar to the MHA. Similar to MHA, SRA receives a query Q, a key K and a value V as input and outputs a fine-grained feature. The difference is that SRA reduces the spatial scale of K and V before the attention operation, which greatly reduces the computational or memory overhead. The details of the SRA in phase i can be expressed as follows:

$$SRA(Q,K,V) = Concat\left(head_0,\cdots,head_{N_i}\right)W^O \tag{1}$$

$$head_j = Attention\left(QW_j^Q, SR(K)W_j^K, SR(V)W_j^V\right) \tag{2}$$

which $Concat(\cdot)$ is the connection operation in [34]. $W_j^Q \in \mathbb{R}^{C_i \times d_{head}}$, $W_j^K \in \mathbb{R}^{C_i \times d_{head}}$, $W_j^V \in \mathbb{R}^{C_i \times d_{head}}$ and $W^O \in \mathbb{R}^{C_i \times C_i}$ are the linear projection parameters. N_i is the number of heads of the attention layer in stage *i*. Therefore, the size of each head (*i.e.* d_{head}) is equal to $\frac{C_i}{N_i}$. $SR(\cdot)$ is the operation to reduce the spatial dimensionality of the input sequence (*i.e.*, *K* or *V*), is denoted as follows:

$$SR(X) = Norm\left(Reshape(X, R_i)W^S\right) \tag{3}$$

where $X \in \mathbb{R}^{H_i W_i \times C_i}$ denotes an input sequence, and R_i is the reduction rate of the attention layer in stage *i*. $Reshape(X, R_i)$ is the operation that reshapes the input sequence *x* into a sequence of size $\frac{H_i W_i}{R_i^2} \times \left(R_i^2 C_i\right)$. $W_s \in \mathbb{R}^{R_i^2 C_i \times C_i}$ is reduces the dimensionality of the input sequence to a linear projection of C_i. $Norm(\cdot)$ representation layer normalization [35]. Same as the attention operation in the original Transformer [34], the formula for the calculation of $Attention(\cdot)$ as follows:

$$Attention(q,k,v) = Softmax\left(\frac{qk^T}{\sqrt{d_{head}}}\right)v \tag{4}$$

3.3. LGFFN-GHI

Figure 1 depicts the structure of LGFFN-GHI. The CNN branch uses the ResNet18 network structure. It consists of a backbone network, four stages of residual blocks and classifiers connected. Two residual blocks are included in each stage. The first residual block in each stage halves the size of the input feature map and doubles the number of channels. In contrast to the way the feature maps of the original network are reduced in resolution, we introduce the Transition module [36]. The Transition module is a Conv-Batch Normalization-Relu-Batch Normalization-Avg Pooling sequence. Transformer branch uses the PVT-Tiny network. It is similar to ResNet18 and contains four stages and a classifier. Each stage consists of a patch embedding layer and two Transformer encoders. To be consistent with ResNet18 feature dimension, we change the output feature dimension of the third stage from 320 to 256. The linear mapping of the traditional Transformer is changed to a convolutional operation in PVT, and the zero-fill operation of convolution can imply position information [37], so we remove the position encoding in PVT-Tiny. For classification, PVT uses the traditional ViT approach to add a category tag to the last layer of Transformer encoder. We use PVTv2 on the PVT-Tiny branch [38] approach, *i.e.*, the output of the last layer of Transformer encoder is averaged as the category marker.

3.4. Feature Fuse Block (FFB)

In order to blend the local features extracted by the ResNet18 branch and the global features extracted by the PVT-Tiny branch, we propose a local-global feature blending module. The module consists of two branches using bidirectional concordance to mix local and global features respectively. In each stage, it is placed in front of the residual block and the Transformer encoder layer. The specific formula for the PVT-Tiny-> ResNet18 branch is follows:

$$X = Attention\left(X, SR\left(ZW_Z^K\right), SR\left(ZW_Z^V\right)\right) + X \tag{5}$$

where $X \in \mathbb{R}^{H_i \times W_i \times C_i}$ is the input feature map of ResNet18 in stage *i*, and $Z \in \mathbb{R}^{H_i W_i \times C_i}$ is the input sequenc of PVT-Tiny in stage *i*. $W_Z^K \in \mathbb{R}^{C_i \times C_i}$ and $W_Z^V \in \mathbb{R}^{C_i \times C_i}$ are the linear projection parameters. The formulas $SR(\cdot)$ according to (3), in order to reduce the computational, the settings of R_i from the first stage to the fourth stage are 8, 4, 2 and 1, respectively. The $Attention(\cdot)$ calculation formula differs from equation (4) by changing d_{head} is changed to C_i. Similarly, the specific formula for the ResNet18 -> PVT-Tiny branch is follows:

$$Z = Attention\left(Z, SR\left(XW_X^K\right), SR\left(XW_X^V\right)\right) + Z \tag{6}$$

where $W_X^K \in \mathbb{R}^{C_i \times C_i}$ and $W_X^V \in \mathbb{R}^{C_i \times C_i}$ are linear projection parameters.

4. Data and Experimental Parameter Settings

4.1. GasHisSDB Dataset

Four pathologists from Longhua Hospital of Shanghai University of Traditional Chinese Medicine provided 600 pathological images of gastric cancer in size of 2048 × 2048. Three sizes of pathology images (160 × 160, 120 × 120, 80 × 80 pixels) were directly cropped to obtain the database GasHisSDB [39]. Among them, 13, 124 abnormal images and 20, 160 normal images were sub-database of 160 × 160 pixel size. Normal: no cancer cells were present in the images. Abnormal: cancer cells were present in the images. Each normal image did not contain cancerous areas. Each cell had no or very small anisotropy. In the abnormal images, the cancer cells are often in irregularly arranged multilayers with nuclei of variable size and division [40]. **Figure 2** shows some samples of normal and abnormal gastric cancer pathology images. The training set, validation set and test set were randomly divided in the ratio of 6:2:2 on a 160 × 160 sub-data set.

4.2. Experimentation Settings

We used Pytorch to write our code. The computing platform used was an Nvidia Tesla M40 24GB graphics processing unit. In the training phase, AdamW was used as an optimizer with a learning rate set to 1e−5 and trained for 100 epochs. During training, no data enhancement strategy was used. After training, the model with the lowest loss in the validation set was selected for testing.

4.3. Evaluation Metrics

In this study, Accuracy, Precision, Recall, Specificity, and F1 Score were used as

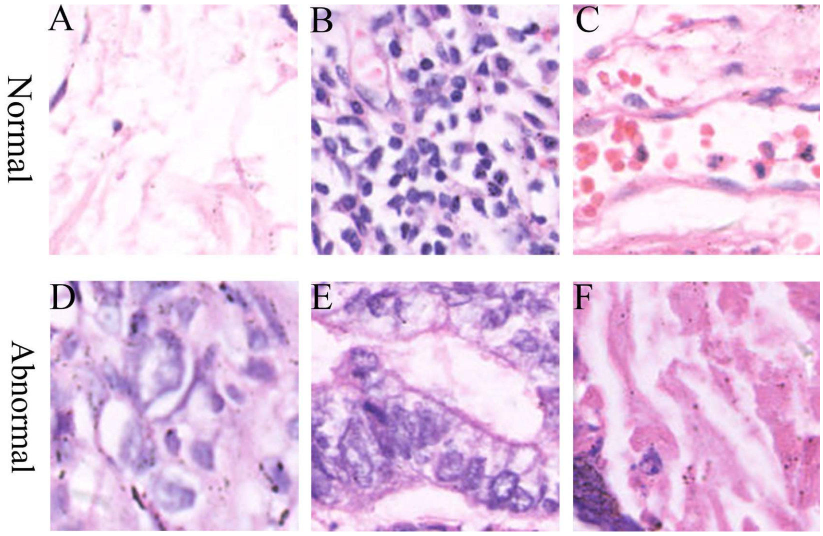

Figure 2. Some samples of 160 × 160 pixel size in GasHisSDB. A, B, and C are images of normal tissues, and D, E, and F are images of tissues containing cancer cells.

metrics to evaluate the model for classifying gastric pathology images. For the problem of class imbalance in gastric pathology images, Matthew correlation coefficient (MCC) [41] [42] was also used for evaluation. It is one of the balance metrics for unbalanced datasets, considering true and false predictions for both negative and positive categories. The diagnostic significance of the classifier at different thresholds is evaluated by plotting the accuracy-recall (PR) and receiver operating characteristic (ROC) curves. They are mathematically formulated as follows:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \tag{7}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{9}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN}+\text{FP}} \tag{10}$$

$$\text{F1} = 2\cdot\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{11}$$

$$\text{MCC} = \frac{(\text{TP}\cdot\text{TN})-(\text{FP}\cdot\text{FN})}{\sqrt{(\text{TP}+\text{FP})(\text{TP}+\text{FN})(\text{TN}+\text{FP})(\text{TN}+\text{FN})}} \tag{12}$$

where TP refers to correctly predicted abnormal images, TN refers to correctly predicted normal images, FP refers to incorrectly predicted abnormal images, and FN refers to incorrectly predicted normal images.

5. Results and Discussion

To validate the performance of the proposed model for classifying pathological images of gastric cancer, we compared it with several classical deep learning models, including AlexNet [43], VGG16 [44], Inception-V3 [45] and ResNet50 [7]. To further validate this benefit by combining the two networks, we also trained ReNet18 and PVT-Tiny independently for comparison with our proposed approach.

The training process of the proposed deep learning model is shown in Figure 3 and Figure 4, where the model converges after 100 epochs of training. The loss

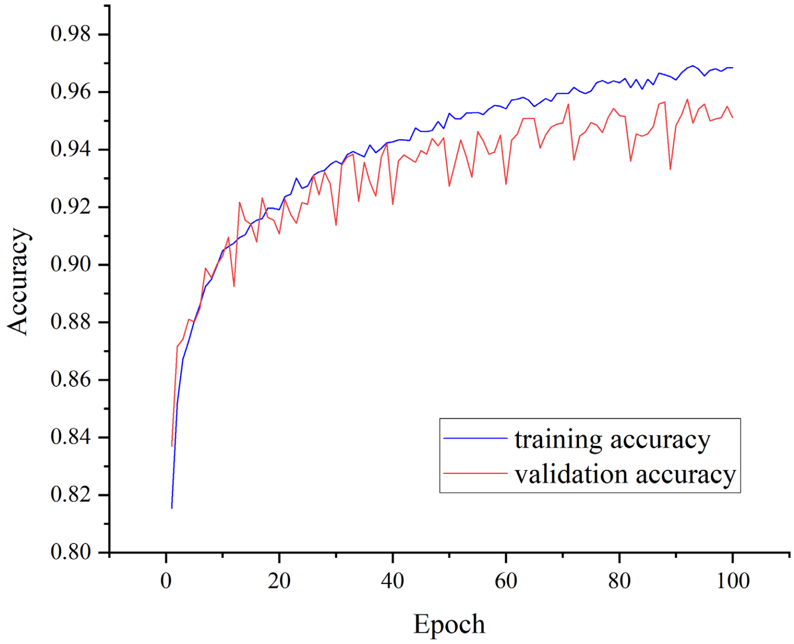

Figure 3. Accuracy curves of LGFFN-GHI on the training and validation sets.

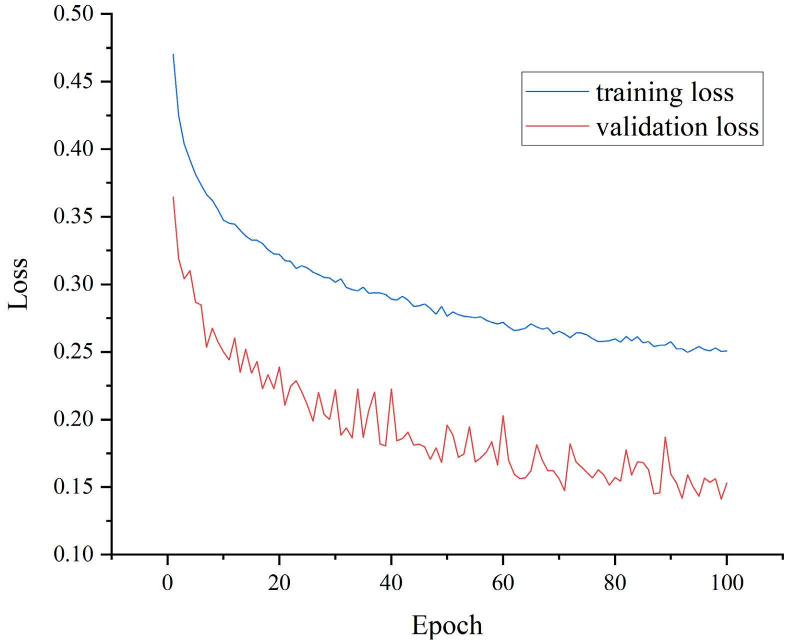

Figure 4. Loss curves of LGFFN-GHI on the training and validation sets.

of the validation set reaches a minimum of 0.1417 at the 91st epoch (**Figure 4**), at which time the accuracy of the training and validation sets are 96.83 and 95.75, respectively (**Figure 3**). **Figure 5** shows the LGFFN-GHI classification confusion matrix on the test set. Out of a total of 6656 test images, 4032 images were in the normal category and 2624 images were in the abnormal category. Among them, 3960 images of the normal category were classified correctly and 72 images were classified incorrectly. The images in the abnormal category were classified correctly in 2484 images and incorrectly in 140 images. As can be seen from **Figure 5**, our model is generally accurate for the secondary classification of gastric cancer, which fully demonstrates that our proposed method is effective and can be applied to the binary classification problem of gastric cancer.

Table 1 shows the model evaluation results of LGFFN-GHI with other models on the test set. In terms of classification accuracy, the lowest is 84.960% for Inception_v3, while our proposed method achieves the highest 96.814%. Among them, ResNet18 and Pvt-Tiny reached 94.426% and 92.833%, respectively, and LGFFN-GHI improved by 2.388% and 3.918%, respectively. The proposed classification method achieves the highest precision, recall, specificity, F1 score and Matthew correlation coefficient all with 97.183%, 94.664%, 98.214%, 95.907% and 93.322%, respectively. Among them, the precision improved by 4.754% and 7.262%, the recall improved by 1.143% and 2.515%, the specificity improved by 3.2% and 4.936%, the F1 score improved by 2.935% and 4.885%, and the Matthew correlation coefficient improved by 4.964% and 8.245%. It can be seen that the two branches in LGFFN-GHI, ResNet18 and Pvt-Tiny, can benefit from each other's extracted information through FFB as a bridge of information exchange. LGFFN-GHI outperforms the base model in accuracy, precision, recall, specificity, F1 score and Matthew correlation coefficient. The CNN branch of LGFFN-GHI is ResNet18, which is good at capturing local information in images, such as the structural information of cell clusters composed of several cells in gastric pathology tissue images. The Transformer branch of LGFFN-GHI is Pvt-Tiny, which can directly extract the global information in the image, such as the relationship between cell clusters in the gastric pathology tissue image and the image texture information. FFB in LGFFN-GHI serves as a bridge for information exchange, enabling ResNet18 and Pvt-Tiny to benefit from each other's extracted information.

Figure 6 shows the ROC curves of the three classifiers LGFFN-GHI, ResNet18 and Pvt-Tiny on the test set, and their AUC values are 99.3%, 98.9% and 98.1%, respectively, and our proposed method improves 0.4% and 1.2% over ResNet18 and Pvt-Tiny, respectively. This indicates that LGFFN-GHI can effectively extract feature information, and has high robustness, so as to effectively improve the accuracy of gastric cancer classification. Relatively speaking, the AUC value of Pvt-Tiny is significantly lower, indicating that the extracted features are not sufficient for gastric cancer classification, and there is a certain gap between PVT-TINY and LGFFN-GHI.

Table 1. Results of LGFFN-GHI compared with different models on the test set (%).

Model	Accuracy	Precision	Recall	Specificity	F1Score	MCC
ResNet18	94.426	92.429	93.521	95.014	92.972	88.358
Pvt-Tiny	92.833	89.921	92.149	93.278	91.022	85.077
AlexNet	90.985	88.953	88.071	92.881	88.510	81.096
Vgg16	95.582	96.120	92.53	97.569	94.291	90.735
Inception_v3	84.960	78.089	85.975	84.300	81.842	69.291
ResNet50	91.195	87.463	90.663	91.542	89.034	81.720
LGFFN-GHI	**96.814**	**97.183**	**94.664**	**98.214**	**95.907**	**93.322**

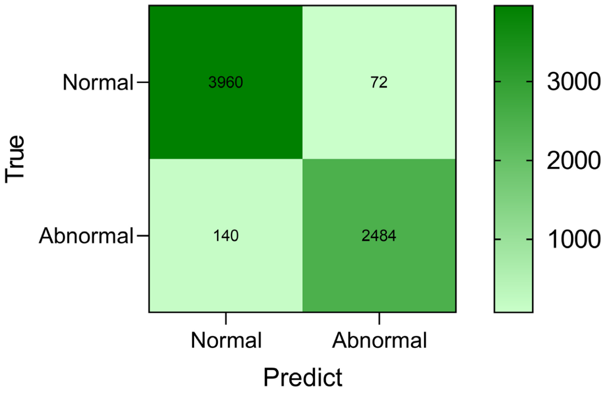

Figure 5. Confusion matrix of LGFFN-GHI in the test set.

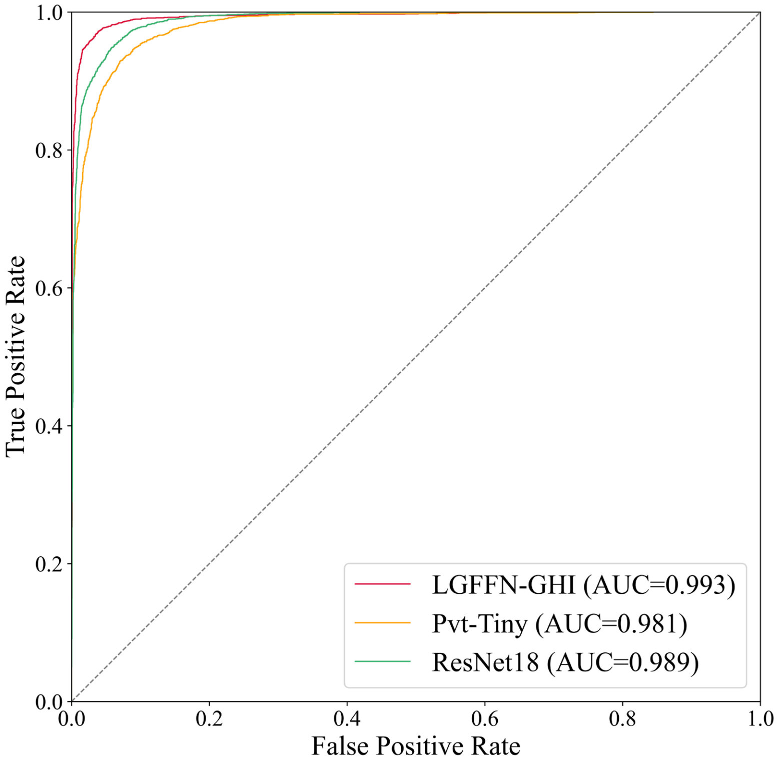

Figure 6. ROC curves of LGFFN-GHI, ResNet18 and Pvt-Tiny on the test set.

6. Conclusions and Future Work

In this paper, an LGFFN-GHI model is proposed to classify histopathological images of gastric cancer into normal and abnormal. In the experiment, LGFFN-GHI was tested on the publicly available gastric cancer histopathology dataset (GasHisSDB) to achieve 96.814% accuracy. Our model outperforms classical neural network models, including AlexNet, VGG16, Inception-V3, ResNet50, ReNet18 and PVT-Tiny, showing its potential for gastric cancer histopathology image classification tasks. The model not only considers the advantages of classical CNN models in describing local information, but also uses the latest Transformer model for global information description, considering both global and local associations of images in spatial context.

Although LGFFN-GHI is very effective for gastric cancer diagnosis, this method needs to be validated on a larger dataset before it can be used in the clinic. We intend to include this work as part of our future work. In addition, we would like to explore LGFFN-GHI for breast, colon and prostate cancer diagnosis. The number of parameters of LGFFN-GHI is larger than the general model, and we will further optimize it to reduce the number of parameters in the future.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Thrift, A.P. and El-Serag, H.B. (2020) Burden of Gastric Cancer. *Clinical Gastroenterology and Hepatology*, **18**, 534-542. https://doi.org/10.1016/j.cgh.2019.07.045

[2] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I. and Jemal, A. (2021) Global Cancer Statistics 2020: Globocan Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA*: *A Cancer Journal for Clinicians*, **71**, 209-249. https://doi.org/10.3322/caac.21660

[3] Amin, M.B., Greene, F.L., Edge, S.B., Compton, C.C., Gershenwald, J.E. and Brookland, R.K. (2017) The Eighth Edition AJCC Cancer Staging Manual: Continuing to Build a Bridge from a Population-Based to a More "Personalized" Approach to Cancer Staging. *CA*: *A Cancer Journal for Clinicians*, **67**, 93-99. https://doi.org/10.3322/caac.21388

[4] Sano, T., Coit, D.G., Kim, H.H., Roviello, F., Kassab, P. and Wittekind, C. (2017) Proposal of a New Stage Grouping of Gastric Cancer for TNM Classification: International Gastric Cancer Association Staging Project. *Gastric Cancer*, **20**, 217-225. https://doi.org/10.1007/s10120-016-0601-9

[5] Niu, P., Zhao, L., Wu, H., Zhao, D. and Chen, Y. (2020) Artificial Intelligence in Gastric Cancer: Application and Future Perspectives. *World Journal of Gastroenterology*, **26**, Article No. 5408. https://doi.org/10.3748/wjg.v26.i36.5408

[6] Smyth, E.C., Nilsson, M., Grabsch, H.I., van Grieken, N.C. and Lordick, F. (2020) Gastric Cancer. *The Lancet*, **396**, 635-648. https://doi.org/10.1016/S0140-6736(20)31288-5

[7] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image

Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[8] Wang, W., Xie, E., Li, X., Fan, D., Song, K. and Liang, D. (2021) Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions. 2021 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 568-578. https://doi.org/10.1109/ICCV48922.2021.00061

[9] Korkmaz, S.A. and Binol, H. (2018) Classification of Molecular Structure Images by Using Ann, Rf, Lbp, Hog, and Size Reduction Methods for Early Stomach Cancer Detection. *Journal of Molecular Structure*, **1156**, 255-263. https://doi.org/10.1016/j.molstruc.2017.11.093

[10] Sharma, H., Zerbe, N., Böger, C., Wienert, S., Hellwich, O. and Hufnagl, P. (2017) A Comparative Study of Cell Nuclei Attributed Relational Graphs for Knowledge Description and Categorization in Histopathological Gastric Cancer Whole Slide Images. 2017 *IEEE* 30*th International Symposium on Computer-Based Medical Systems* (*CBMS*), Thessaloniki, 22-24 June 2017, 61-66. https://doi.org/10.1109/CBMS.2017.25

[11] Liu, B., Zhang, M., Guo, T. and Cheng, Y. (2018) Classification of Gastric Slices Based on Deep Learning and Sparse Representation. 2018 *Chinese Control and Decision Conference* (*CCDC*), Shenyang, 9-11 June 2018, 1825-1829. https://doi.org/10.1109/CCDC.2018.8407423

[12] Garcia, E., *et al*. (2017) Automatic Lymphocyte Detection on Gastric Cancer Ihc Images Using Deep Learning. 2017 *IEEE* 30*th International Symposium on Computer-Based Medical Systems* (*CBMS*), Thessaloniki, 22-24 June 2017, 200-204. https://doi.org/10.1109/CBMS.2017.94

[13] Sharma, H., Zerbe, N., Klempert, I., Hellwich, O. and Hufnagl, P. (2017) Deep Convolutional Neural Networks for Automatic Classification of Gastric Carcinoma Using Whole Slide Images in Digital Histopathology. *Computerized Medical Imaging and Graphics*, **61**, 2-13. https://doi.org/10.1016/j.compmedimag.2017.06.001

[14] Li, Y., Li, X., Xie, X. and Shen, L. (2018) Deep Learning Based Gastric Cancer Identification. 2018 *IEEE* 15*th International Symposium on Biomedical Imaging* (*ISBI* 2018), Washington DC, 4-7 April 2018, 182-185. https://doi.org/10.1109/ISBI.2018.8363550

[15] Liu, B., Yao, K., Huang, M., Zhang, J., Li, Y. and Li, R. (2018) Gastric Pathology Image Recognition Based On Deep Residual Networks. 2018 *IEEE* 42*nd Annual Computer Software and Applications Conference* (*COMPSAC*), Vol. 2, 408-412. https://doi.org/10.1109/COMPSAC.2018.10267

[16] Zhu, Z., Ding, X., Zhang, D. and Wang, L. (2020) Weakly-Supervised Balanced Attention Network for Gastric Pathology Image Localization and Classification. 2020 *IEEE* 17*th International Symposium on Biomedical Imaging* (*ISBI*), Iowa City, 3-7 April 2020, 1-4. https://doi.org/10.1109/ISBI45749.2020.9098567

[17] Kloeckner, J., Sansonowicz, T.K., Rodrigues, Á.L. and Nunes, T.W. (2020) Multi-Categorical Classification Using Deep Learning Applied to the Diagnosis of Gastric Cancer. *Jornal Brasileiro de Patologia e Medicina Laboratorial*, **56**, 1-8. https://doi.org/10.5935/1676-2444.20200013

[18] Wang, S., Zhu, Y., Yu, L., Chen, H., Lin, H. and Wan, X. (2019) Rmdl: Recalibrated Multi-Instance Deep Learning for Whole Slide Gastric Image Classification. *Medical Image Analysis*, **58**, Article ID: 101549. https://doi.org/10.1016/j.media.2019.101549

[19] Song, Z., Zou, S., Zhou, W., Huang, Y., Shao, L. and Yuan, J. (2020) Clinically Applicable Histopathological Diagnosis System for Gastric Cancer Detection Using Deep Learning. *Nature Communications*, **11**, Article No. 4294. https://doi.org/10.1038/s41467-020-18147-8

[20] Kosaraju, S.C., Hao, J., Koh, H.M. and Kang, M. (2020) Deep-Hipo: Multi-Scale Receptive Field Deep Learning for Histopathological Image Analysis. *Methods*, **179**, 3-13. https://doi.org/10.1016/j.ymeth.2020.05.012

[21] Iizuka, O., Kanavati, F., Kato, K., Rambeau, M., Arihiro, K. and Tsuneki, M. (2020) Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Scientific Reports*, **10**, Article No. 1504. https://doi.org/10.1038/s41598-020-58467-9

[22] Cho, K., Lee, S.H. and Jang, H. (2020) Feasibility of Fully Automated Classification of Whole Slide Images Based on Deep Learning. *The Korean Journal of Physiology & Pharmacology*, **24**, 89-99. https://doi.org/10.4196/kjpp.2020.24.1.89

[23] Li, Y., Wu, X., Li, C., Li, X., Chen, H. and Sun, C. (2022) A Hierarchical Conditional Random Field-Based Attention Mechanism Approach for Gastric Histopathology Image Classification. *Applied Intelligence*, **52**, 1-22. https://doi.org/10.1007/s10489-021-02377-4

[24] Xiang, X. and Wu, X. (2021) Multiple Instance Classification for Gastric Cancer Pathological Images Based on Implicit Spatial Topological Structure Representation. *Applied Sciences*, **11**, Article ID: 10368. https://doi.org/10.3390/app112110368

[25] Kolesnikov, A., Dosovitskiy, A., Weissenborn, D., Heigold, G., Uszkoreit, J. and Beyer, L. (2021) An Image Is Worth 16X16 Words: Transformers for Image Recognition at Scale.

[26] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y. and Zhang, Z. (2021) Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 10012-10022. https://doi.org/10.1109/ICCV48922.2021.00986

[27] Li, W., Wang, X., Xia, X., Wu, J., Xiao, X. and Zheng, M. (2022) Sepvit: Separable Vision Transformer.

[28] Dong, X., Bao, J., Chen, D., Zhang, W., Yu, N. and Yuan, L. (2021) Cswin Transformer: A General Vision Transformer Backbone with Cross-Shaped Windows. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 12114-12124. https://doi.org/10.1109/CVPR52688.2022.01181

[29] Li, Y., Yao, T., Pan, Y. and Mei, T. (2022) Contextual Transformer Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2022.3164083

[30] Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y. and Jiang, Z. (2021) Tokens-to-Token Vit: Training Vision Transformers from Scratch on ImageNet. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 58-567. https://doi.org/10.1109/ICCV48922.2021.00060

[31] Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X. and Yuan, L. (2021) Cvt: Introducing Convolutions to Vision Transformers. 2021 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 2-31. https://doi.org/10.1109/ICCV48922.2021.00009

[32] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E. and Wang, Y. (2021) Transunet: Transformers Make Strong Encoders for Medical Image Segmentation.

[33] Han, K., Xiao, A., Wu, E., Guo, J., Xu, C. and Wang, Y. (2021) Transformer in Transformer. *Advances in Neural Information Processing Systems* 34: *Annual Conference on Neural Information Processing Systems*, December 2021, 15908-15919.

[34] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L. and Gomez, A.N. (2017) Attention Is All You Need. 31*st Conference on Neural Information Processing Systems* (*NIPS* 2017), Long Beach, 4-9 December 2017, 5998-6008.

[35] Ba, J.L., Kiros, J.R. and Hinton, G.E. (2016) Layer Normalization.

[36] Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 4700-4708. https://doi.org/10.1109/CVPR.2017.243

[37] Chu, X., Zhang, B., Tian, Z., Wei, X. and Xia, H. (2021) Do We Really Need Explicit Position Encodings for Vision Transformers?

[38] Wang, W., Xie, E., Li, X., Fan, D., Song, K. and Liang, D. (2022) Pvtv2: Improved Baselines with Pyramid Vision Transformer. *Computational Visual Media*, **8**, 415-424. https://doi.org/10.1007/s41095-022-0274-8

[39] Hu, W., Li, C., Li, X., Rahaman, M.M., Ma, J. and Zhang, Y. (2022) Gashissdb: A New Gastric Histopathology Image Dataset for Computer Aided Diagnosis of Gastric Cancer. *Computers in Biology and Medicine*, **142**, Article ID: 105207. https://doi.org/10.1016/j.compbiomed.2021.105207

[40] Wang, F., Shen, L., Li, J., Zhou, Z., Liang, H. and Zhang, X. (2019) The Chinese Society of Clinical Oncology (CSCO): Clinical Guidelines for the Diagnosis and Treatment of Gastric Cancer. *Cancer Communications*, **39**, 1-31. https://doi.org/10.1186/s40880-019-0349-9

[41] Boughorbel, S., Jarray, F. and El-Anbari, M. (2017) Optimal Classifier for Imbalanced Data Using Matthews Correlation Coefficient Metric. *PLOS ONE*, **12**, e177678. https://doi.org/10.1371/journal.pone.0177678

[42] Chicco, D. and Jurman, G. (2020) The Advantages of the Matthews Correlation Coefficient (Mcc) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics*, **21**, Article No. 6. https://doi.org/10.1186/s12864-019-6413-7

[43] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, **60**, 84-90.

[44] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition.

[45] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z. (2016) Rethinking the Inception Architecture for Computer Vision. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 2818-2826. https://doi.org/10.1109/CVPR.2016.308