IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.006, LIIR407
DOI: https://doi.org/10.52731/liir.v006.407

# Improved Prediction of Geckler Classification in Gram Stained Smears Images for Sputum

Kei Goto [*], Masaki Kono [*], Kouichi Hirata [*]

## Abstract

In this paper, we predict Geckler classification as Geckler classes and quality classes in Gram stained smears images for sputum, by using image classification and object detection. Here, we adopt VGG, MobileNet, DenseNet, RegNet, ConvNeXt, ViT and EfficientNet as image classifiers and YOLO11, HIC-YOLO11 and SOD-YOLO11 as object detectors. Note that, whereas the image classifiers classify the classes of images directly, the object detectors first detect buccal squamous epithelial (BSE) cells and leukocytes in the image and then predict the class of the image by the number of them.

*Keywords:* Gram stained smears images, Geckler classification, image classification, object detection

## 1 Introduction

In clinical microbiology, a laboratory investigation mainly consists of a microscopic examination, a culture and an identification [3, 8, 17]. The *Gram stain* [1], which has been introduced by Hans Christian Gram (1853–1938) at 1884, is a kind of the microscopic examination for Gram stained smears images *per* $1,000\times$ *field* (*cf.*, [19]). It is a procedure by which bacteria can be classified by the ability of the cell wall to absorb a crystal violet dye, followed by a red safranin counter stain. The Gram stain colors bacteria by either purple (violet) or red (pink).

For the sample of sputum, a *Geckler classification* [7, 19, 21, 27, 29][1], which has been introduced by Geckler *et al.* [7], is known as a screening method for Gram stained smears images *per* $100\times$ *field* how the Gram stained smears image per $1,000\times$ field is quality for the microscopic examination. The quality of images is defined by the number of *buccal squamous epithelial (BSE) cells* and *leukocytes* in the Gram stained smears images per $100\times$ field. Table 1 illustrates the definition of the six classes in the Geckler classification [7, 19, 21, 27, 29], which we call *Geckler classes* in this paper. Here, #BSE (*resp.*, #leu) denotes the number of BSE cells (*resp.*, leukocytes).

[*] Kyushu Institute of Technology, Kawazu 680-4, Iizuka 820-8502, Japan

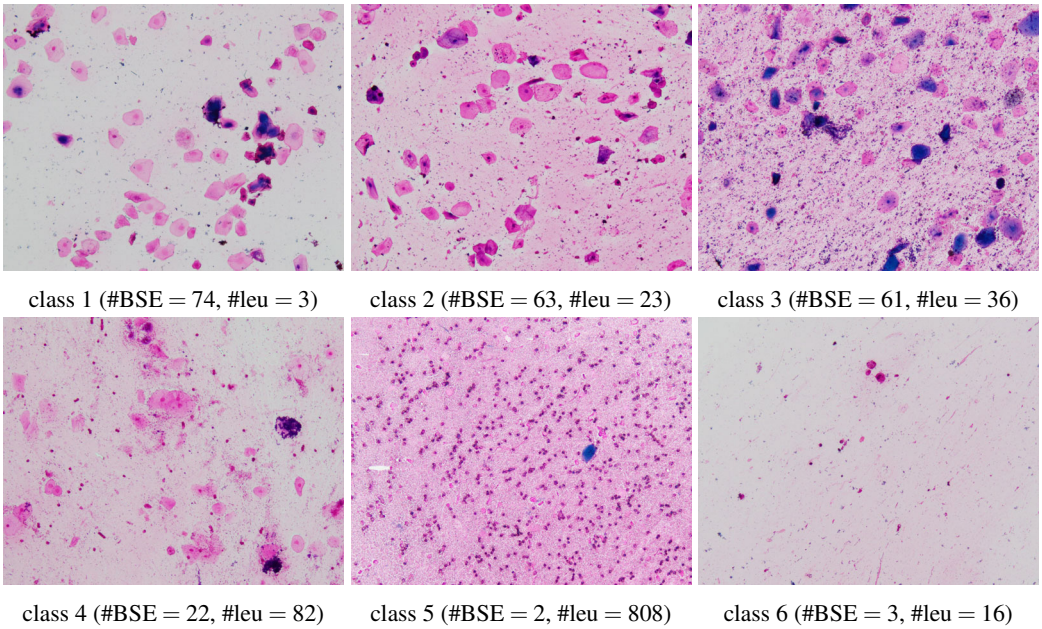
[1] The Geckler classification is also known as the criteria by Cumitech 7A [2] as presented in [19].

In the Geckler classification, the Geckler classes 4 and 5 are valuable for microscopic examination per $1,000\times$ field [7, 19, 21, 27, 29]. Then, in this paper, we denote the Geckler classes from 1 to 3 by NG (no good), those of 4 and 5 by GE (good and excellent) and that of 6 by UN (unknown), which we call *quality classes* in this paper. In other words, we say that the quality class is NG if the number of BSE cells is larger than 25, GE if the number of BSE cells is smaller than 25 and the number of leukocytes is larger than 25 and UN if both the numbers of BSE cells and leukocytes are smaller than 25.

Table 1: The Geckler classification.

| Geckler class | #BSE | #leu | quality | quality class |
|---|---|---|---|---|
| 1 | $> 25$ | $< 10$ | no good | NG |
| 2 | $> 25$ | $10 - 25$ | no good | |
| 3 | $> 25$ | $> 25$ | no good | |
| 4 | $10 - 25$ | $> 25$ | good | GE |
| 5 | $< 10$ | $> 25$ | excellent | |
| 6 | $< 25$ | $< 25$ | unknown | UN |

Figure 1 illustrates Gram stained smears images per $100\times$ field for sputum for every Geckler class. Here, the original size is $2,448 \times 1,920$ pixels and a larger object colored by violet or pink is a BSE cell and a small point colored by pink is a leukocyte.



class 1 (#BSE = 74, #leu = 3) class 2 (#BSE = 63, #leu = 23) class 3 (#BSE = 61, #leu = 36)

class 4 (#BSE = 22, #leu = 82) class 5 (#BSE = 2, #leu = 808) class 6 (#BSE = 3, #leu = 16)

Figure 1: A Gram stained smears image per $100\times$ field for sputum for every Geckler class.

As the related works to predicting Geckler classification, Crossman *et al.* [4] have developed the method to detect the regions of BSE cells and leukocytes by using multiple covariance approach with Kernel SVM from manually cropped images of BSE cells, leukocytes and other negative data for the Gram stained smears images per $65\times$ field. Since the

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*

shape and the size of BSE cells have wide variety in the Gram stained smear images for the sample of sputum, in order to achieve the Geckler classification by using their work, it is necessary to start to detect the regions of BSE cells carefully.

Hashimoto *et al.* [9] have developed the system to predict Geckler classes from the Gram stained smears images per 100× field for the sample of sputum directly, by using the classical image processing and the machine learning of SVM (support vector machine) and DNN (deep neural network). Today, we can improve this method as the *object detection* as counting the number of BSE cells and leukocytes.

Recently, as the direct related work, Sugimoto and Hirata [23] have predicted Geckler classes and quality classes of Gram stained smears images for sputum by using the image classification and the object detection. Here, they have adopted VGG16&19 [22], DenseNet [12] and MobileNet [10] as image classifiers and YOLOv5 [13] and YOLOv7 [28] as object detectors. However, this research contains the serious problem that some images are duplicated. Then, in this paper, first we remove the duplication of images.

Next, we predict the Geckler classes and quality classes from Gram stained smears images per 100× field for the sample of sputum by applying two methods of an *image classification* and an *object detection* based on neural networks.

For the image classification, we predict the Geckler class or the quality class for a given image by classifying the training images directly, without annotating BSE cells and leukocytes. In this paper, we adopt VGG16&19 [22], MobileNet [10], DenseNet [12], RegNet [30], ConvNeXt [18], ViT [5] and EfficientNet [25] as image classifiers.

On the other hand, for the object detection, we detect BSE cells and leukocytes from the training images annotated by them and then determine the Geckler class or the quality class of a given image by the number of detected them. In this paper, we adopt YOLO11 [14], which is the latest version of YOLO, and its improvement of HIC-YOLO11 and SOD-YOLO11. Here, both HIC-YOLOv5 [26] (Head, Involution and CBAM-YOLO) and SOD-YOLOv8 [15] (Small Object Detection-YOLO) are the improvements of YOLO to detect small objects. In this paper, we adopt HIC-YOLO11 [24] and SOD-YOLO11 [16] incorporating YOLO11 with HIC-YOLOv5 and SOD-YOLOv8.

## 2 Methods

In this section, we explain the image classifiers, the object detectors and the evaluation measures.

### 2.1 Image Classifiers

In this paper, we adopt the following image classifiers.

A VGGNet (VGG16 and VGG19) [22] is a fundamental convolutional neural network (CNN) with very small (3× 3) convolution filters. VGG16 is the VGGNet with 13 convolution layers (and 3 FC layers, which consists of total 16 layers) and VGG19 is the VGGNet with 16 convolution layers (and 3 FC layers, which consists of total 19 layers).

A MobileNet [10] is built on depthwise separable convolutions except for the first layer which is a full convolution. In this paper, we adopt MobileNetV3-S(mall) and MobileNetV3-L(arge) under MobileNetV3 [11].

A DenseNet [12] is designed to connect all layers (with matching feature-map sizes) directly with each other. To preserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers.

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*

In this paper, we adopt DenseNet-169 and DenseNet-201, whose numbers of layers are 169 and 201.

A RegNet (RNN-Regulated Residual Networks) [30] is the new architecture of the convolutional RNNs that, for every building block, a recurrent unit in the convolutional RNN takes the feature from the current building block as the input, and then encodes both the input and the serial information to generate the hidden state. In the prepared models of RegNet [20], in this paper, we adopt RegNetX-32GF and RegNetY-32GF, where 32GF means $32\times10^9$ flops to multiply-adds.

A ConvNeXt [18] is a family of pure convolution neural networks that uses the stage computer ratio, the "patchify stem" ($4\times4$ non-overlapping convolution) in the network, the ResNext design, inverted bottlenecks, $7\times7$ depthwise convolution in each block, a single GELU activation in each block, one LayerNorm as our choice of normalization in each residual block and separate downsampling layers. The ConvNeXt is prepared the variants T/S/B/L/XL. In this paper, we adopt ConvNeXt-B and ConvNeXt-L.

A ViT (Vision Transformer) [5] is the application of the Transformer in natural language processing to image processing. The ViT has three variants of "Base," "Large," and "Huge." In this paper, we adopt the variant of ViT-B/32 and ViT-L/32, where 32 means that the input patch sizes is $32\times32$.

An EfficientNet [25] is designed as a critical baseline network, called EfficientNet-B0, for the effectiveness of model scaling of depth, width and resolution. By changing the values of depth, width and resolution, the variants of EfficientNet are given as EfficientNet-B1 to B8. In this paper, we adopt EfficientNet-B7.

## 2.2 Object Detectors

*YOLO*, which is an acronym "You only look once," is a single shot object detector integrating of the entire object detection and classification process in a single network. In this paper, we adopt YOLO11 [14] as the version 11 released in September 2024. Additionally, we also adopt HIC-YOLO [24] and SOD-YOLO [16], which are the variations of YOLO11 appropriate to detect small objects. Figures 2 and 3 illustrate the architectures of YOLO11, HIC-YOLO11 and SOD-YOLO11.

In YOLO, the several models named as n (nano), s (small), m (medium), l (large) and x (extra large) are prepared, by setting the model depth multiples and the layer width multiples. Table 2 illustrates them for YOLO11 (and HIC-YOLO11 and SOD-YOLO11). We denote the prepared models of YOLO11, HIC-YOLO11, SOD-YOLO11 as YOLO11n, HIC-YOLO11n, SOD-YOLO11n, and so on.

Table 2: The model depth multiples and the layer width multiples of prepared models for YOLO11.

| model | depth | width |
|---|---|---|
| n | 0.50 | 0.25 |
| s | 0.50 | 0.50 |
| m | 0.50 | 0.75 |
| l | 1.00 | 1.00 |
| x | 1.00 | 1.50 |

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

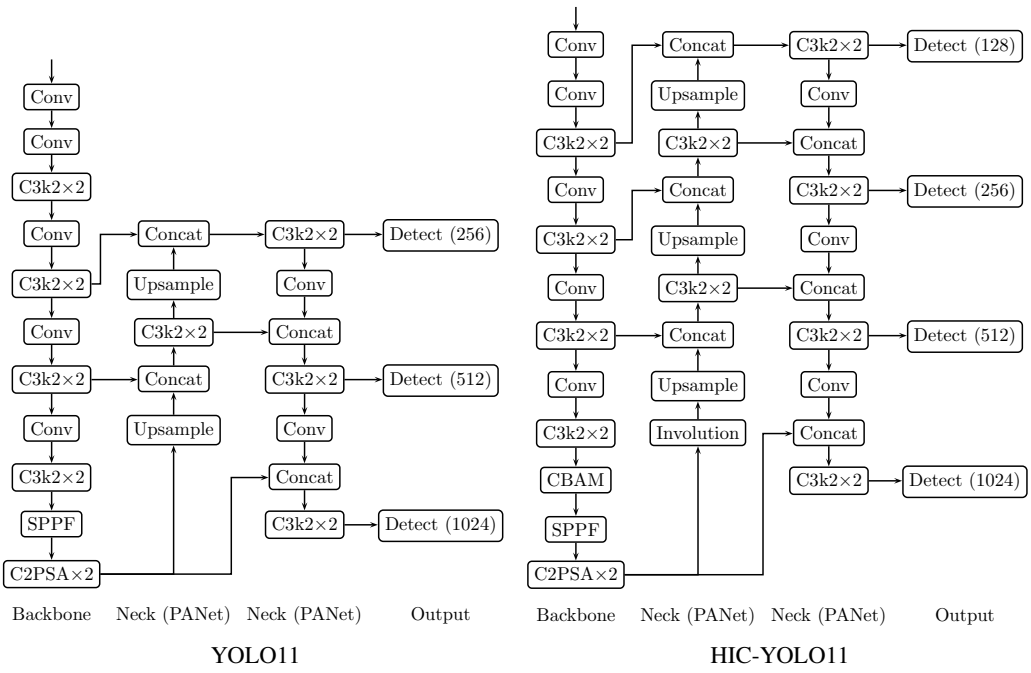



Figure 2: The architectures of YOLO11 [14] and HIC-YOLO11 [24].

## 2.3 Evaluation Measures

First, we introduce the evaluation measures for the image classification. Let $TP$ (*resp.*, $FP$) denote the number of true (*resp.*, false) images that are predicted to be true and $TN$ (*resp.*, $FN$) denote the number of false (*resp.*, true) images that are predicted to be false. We call $TP$, $FP$, $TN$ and $FN$ *true positive*, *false positive*, *true negative* and *false negative*, respectively. Then, we adopt the following standard evaluation measures of *accuracy* (*acc*), *precision* (*pre*), *recall* (*rec*) and *F1-value* (*F1*) to evaluate the results of classification.

$$acc = \frac{TP+TN}{TP+TN+FP+FN}\,,\ pre = \frac{TP}{TP+FP}\,,\ rec = \frac{TP}{TP+FN}\,,\ F1 = \frac{2\times rec\times pre}{rec+pre}.$$

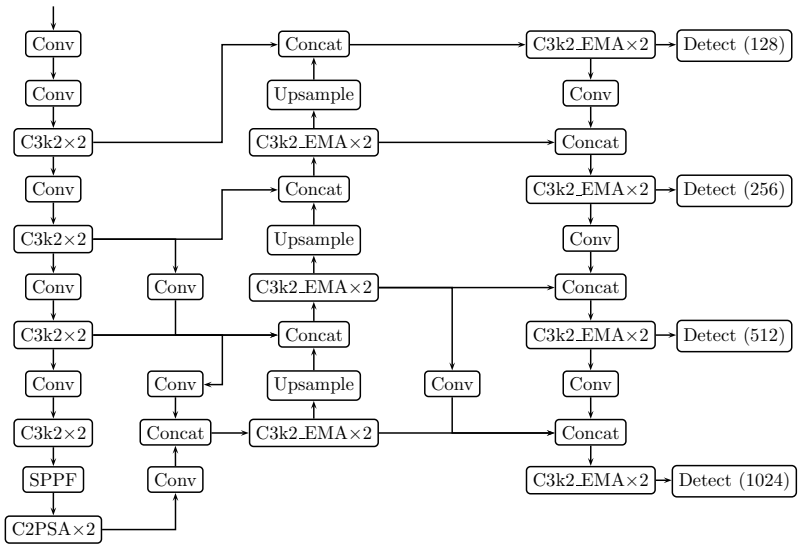
Next, in order to evaluate the object detection, we introduce the following *intersection over union* (*IoU*) [6] between the prediction box $P$ and the ground truth box $T$, where $|R|$ denotes the area of a region $R$:

$$IoU = \frac{|P\cap T|}{|P\cup T|}.$$

For a given threshold $\delta$ (%), let $TP$ be the number of the predicted boxes such that $IoU \geq \delta$, $FP$ the number of the predicted boxes such that $IoU < \delta$ and $FN$ the number of the ground truth boxes such that $IoU < \delta$.

Then, the *precision* and *recall* for the object detection are defined that $\delta$ is set to 50. Also, an *average precision for* $\delta$ ($AP\delta$) [6] is defined as the average detection precision under different recalls. Then, we adopt a (COCO) *mean AP* (*mAP*) [6] that is an average of APs when varying $\delta$ is from 50 to 95 with a step of 5.

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*




SOD-YOLO11

Figure 3: The architecture of SOD-YOLO11 [16].

### 2.4 Data and Setting

In this paper, we adopt total 808 Gram stained smears images of sputum, which have removed the duplicated images. Then, we predict Geckler classification by applying 5-fold cross validation to classification and detection as follows.

First, we divide all of the images into 5 groups uniformly. Next, after setting the first, second and third groups are set to training images, the fourth group to validating images and the fifth group to testing images, we apply the classification and the detection of training, validating and testing images by the classifiers and the detectors. Finally, we repeat the classification at 4 times with shifting the groups.

The evaluation values are the average of the testing results for 5-fold cross validation. Furthermore, in this paper, we increase the number of training images at thrice by *data augmentation* of $x$-axis and $y$-axis flippings. Table 3 illustrates the number (#) of images, the average number (#BSE) of BSE cells and the average number (#leu) of leukocytes in each image for every Geckler class into 5 group.

## 3 Experimental Results

In this section, we give experimental results of predicting Geckler and quality classes by image classifiers and object detectors. Here, the computer environment is that OS is Ubuntu 22.04.4 LTS, CPU is Intel(R) Xeon(R) CPU @ 2.20GHz, GPU is NVIDIA L4 24GB and RAM is 53GB.

### 3.1 Predicting Geckler and Quality Classes by Image Classifiers

For predicting Geckler and quality classes by image classifiers, we apply Gram stained smears images to image classifiers directly. Table 4 illustrates the evaluation values of

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*

Table 3: The number (#) of images, the average number (#BSE) of BSE cells and the average number (#leu) of leukocytes in each image for every Geckler class into 5 group.

| class | total | | | 1 | | | 2 | | | 3 | | | 4 | | | 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | # | #BSE | #leu | # | #BSE | #leu | # | #BSE | #leu | # | #BSE | #leu | # | #BSE | #leu | # | #BSE | #leu |
| 1 | 130 | 78.44 | 5.24 | 26 | 87.92 | 4.73 | 25 | 77.88 | 5.68 | 28 | 75.96 | 5.86 | 26 | 77.35 | 5.31 | 25 | 73.04 | 4.56 |
| 2 | 126 | 46.53 | 19.26 | 22 | 44.46 | 19.91 | 24 | 47.13 | 19.08 | 29 | 51.90 | 18.66 | 24 | 42.42 | 19.63 | 27 | 45.67 | 19.22 |
| 3 | 200 | 52.06 | 159.27 | 44 | 52.57 | 154.00 | 44 | 49.07 | 176.32 | 39 | 52.18 | 169.97 | 35 | 50.86 | 138.54 | 38 | 55.92 | 153.74 |
| 4 | 82 | 16.26 | 320.68 | 17 | 14.53 | 299.00 | 14 | 14.29 | 259.50 | 14 | 17.00 | 373.64 | 22 | 17.18 | 361.09 | 15 | 18.00 | 293.67 |
| 5 | 154 | 3.09 | 416.60 | 31 | 3.52 | 429.13 | 30 | 2.53 | 459.03 | 30 | 3.27 | 453.73 | 31 | 2.90 | 274.42 | 32 | 3.22 | 467.63 |
| 6 | 116 | 3.28 | 9.29 | 24 | 3.54 | 8.25 | 26 | 3.35 | 11.50 | 21 | 2.73 | 8.36 | 22 | 3.83 | 9.39 | 21 | 2.86 | 9.62 |

accuracy (acc.), precision (pre.), recall (rec.) and F1-value (F1) of predicting Geckler and quality classes by image classifiers, and F1-value ($F1_\circ$) without data augmentation. The number by bold faces denotes the largest value and the underlined number denotes the difference from the largest value is within 0.01.

Table 4: The evaluation values of predicting Geckler and quality classes by image classifiers.

| | Geckler class | | | | | quality class | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| classifiers | acc. | pre. | rec. | F1 | $F1_\circ$ | acc. | pre. | rec. | F1 | $F1_\circ$ |
| VGG16 | 0.891 | 0.884 | 0.889 | 0.884 | $>$ 0.861 | 0.943 | **0.946** | 0.933 | **0.939** | $>$ 0.919 |
| VGG19 | 0.872 | 0.862 | 0.867 | 0.861 | $>$ 0.842 | 0.929 | 0.927 | 0.914 | 0.919 | $>$ 0.915 |
| MobileNetV3-S | 0.876 | 0.870 | 0.877 | 0.868 | $>$ 0.858 | 0.932 | 0.926 | 0.926 | 0.924 | $>$ 0.918 |
| MobileNetV3-L | 0.884 | 0.877 | 0.889 | 0.878 | $>$ 0.870 | 0.931 | 0.921 | 0.933 | 0.924 | $>$ 0.919 |
| DenseNet-169 | 0.885 | 0.875 | 0.887 | 0.878 | $>$ 0.873 | 0.937 | 0.935 | 0.931 | 0.932 | $>$ 0.920 |
| DenseNet-201 | 0.885 | 0.877 | 0.890 | 0.879 | $>$ 0.874 | 0.937 | 0.933 | 0.929 | 0.930 | $>$ 0.921 |
| RegNetX-32GF | 0.885 | 0.876 | 0.885 | 0.876 | $<$ 0.877 | 0.933 | 0.928 | 0.930 | 0.927 | $<$ **0.937** |
| RegNetY-32GF | **0.896** | **0.886** | **0.900** | **0.889** | $>$ 0.882 | 0.934 | 0.929 | 0.930 | 0.929 | $<$ 0.934 |
| ConvNeXt-B | 0.884 | **0.886** | 0.885 | 0.880 | $>$ 0.879 | **0.946** | 0.945 | **0.934** | **0.939** | $>$ 0.930 |
| ConvNeXt-L | 0.884 | 0.877 | 0.885 | 0.877 | $<$ **0.884** | 0.933 | 0.926 | 0.924 | 0.923 | $<$ 0.931 |
| ViT-B/32 | 0.859 | 0.851 | 0.858 | 0.850 | $<$ 0.877 | 0.919 | 0.915 | 0.907 | 0.909 | $<$ 0.925 |
| ViT-L/32 | 0.879 | 0.875 | 0.877 | 0.871 | $>$ 0.861 | 0.929 | 0.928 | 0.910 | 0.916 | $<$ 0.919 |
| EfficientNet-B7 | 0.876 | 0.870 | 0.882 | 0.871 | $>$ 0.855 | 0.931 | 0.918 | 0.927 | 0.921 | $>$ 0.914 |

Table 4 shows that, for predicting Geckler classes, all of the F1-values are larger than 0.850. In particular, RegNetY-32GF has the largest F1-value of 0.889, VGG16 has the second largest F1-value of 0.884 and ConvNeXt-B has the third largest F1-value of 0.880. Also, RegNetY-32GF has the largest values of other evaluation measures, VGG16 has the next largest values of accuracy and precision and ConvNeXt-B has the largest value of precision. Hence, RegNetY-32GF is most appropriate classifier to predict Geckler classes. Additionally, RegNetY-32GF, VGG16 and ConvNeXt-B are more appropriate classifiers than others.

On the other hand, Table 4 also shows that, for predicting quality classes, all of the F1-values are larger than 0.909. In particular, VGG16 and ConvNeXt-B have the largest

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*

F1-value of 0.939. Also, VGG16 has the largest value of precision and the second largest values of accuracy and recall, and ConvNeXt-B has the largest values of accuracy and recall and the second largest value of precision. Hence, VGG16 and ConvNeXt-B are more appropriate classifiers to predict quality classes than others.

Finally, by comparing F1 with $F1_\circ$, the data augmentation for ViT-B/32, RegNetX-32GF and ConvNeXt-L is not effective for predicting Geckler and quality classes.

### 3.2 Predicting Geckler and Quality Classes by Object Detectors

For predicting Geckler and quality classes by object detectors, first we detect BSE cells and leukocytes in Gram stained smears images. Next, we predict Geckler classes and quality classes by using the number of detected BSE cells and leukocytes. Table 5 illustrates the evaluation values of accuracy (acc.), precision (pre.), recall (rec.) and F1-value (F1) of predicting Geckler and quality classes by object detectors, and F1-value ($F1_\circ$) without data augmentation.

Table 5: The evaluation values of predicting Geckler and quality classes by object detectors.

| | Geckler class | | | | | quality class | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| detectors | acc. | pre. | rec. | F1 | $F1_\circ$ | acc. | pre. | rec. | F1 | $F1_\circ$ |
| YOLO11n | 0.840 | 0.850 | 0.835 | 0.838 | $>0.722$ | 0.964 | 0.950 | 0.957 | 0.953 | $>0.939$ |
| YOLO11s | 0.835 | 0.850 | 0.826 | 0.832 | $>0.772$ | 0.966 | 0.953 | 0.961 | 0.956 | $<$ **0.959** |
| YOLO11m | **0.863** | **0.867** | **0.855** | **0.857** | $>0.767$ | 0.965 | 0.951 | 0.962 | 0.955 | $>0.952$ |
| YOLO11l | 0.841 | 0.849 | 0.832 | 0.835 | $>0.762$ | 0.966 | 0.955 | 0.962 | 0.958 | $>0.949$ |
| YOLO11x | 0.834 | 0.846 | 0.822 | 0.828 | $>0.762$ | 0.963 | 0.948 | 0.957 | 0.952 | $<0.953$ |
| HIC-YOLO11n | 0.802 | 0.827 | 0.812 | 0.810 | $>0.707$ | 0.965 | 0.951 | 0.959 | 0.955 | $>0.943$ |
| HIC-YOLO11s | 0.843 | 0.855 | 0.835 | 0.839 | $>0.784$ | **0.973** | 0.960 | **0.968** | **0.963** | $>0.944$ |
| HIC-YOLO11m | 0.826 | 0.845 | 0.816 | 0.822 | $>0.767$ | 0.969 | **0.962** | 0.959 | 0.960 | $>0.943$ |
| HIC-YOLO11l | 0.839 | 0.852 | 0.828 | 0.834 | $>0.777$ | 0.968 | 0.959 | 0.960 | 0.959 | $>0.949$ |
| HIC-YOLO11x | 0.831 | 0.844 | 0.823 | 0.827 | $>0.765$ | 0.963 | 0.946 | 0.958 | 0.951 | $>0.945$ |
| SOD-YOLO11n | 0.789 | 0.810 | 0.800 | 0.798 | $>0.720$ | 0.968 | 0.949 | 0.964 | 0.956 | $>0.955$ |
| SOD-YOLO11s | 0.835 | 0.845 | 0.827 | 0.831 | $>0.778$ | 0.966 | 0.953 | 0.959 | 0.956 | $>0.951$ |
| SOD-YOLO11m | 0.814 | 0.825 | 0.812 | 0.812 | $>$ **0.807** | 0.969 | 0.952 | 0.966 | 0.958 | $>0.953$ |
| SOD-YOLO11l | 0.829 | 0.844 | 0.820 | 0.824 | $>0.783$ | 0.961 | 0.948 | 0.953 | 0.950 | $<0.953$ |
| SOD-YOLO11x | 0.846 | 0.858 | 0.840 | 0.842 | $>0.778$ | 0.968 | 0.956 | 0.961 | 0.958 | $>0.947$ |

Table 5 shows that, for predicting Geckler classes, YOLO11m has the largest values of all the evaluation measures, which are at least 0.1 larger than the values for other detectors. Hence, YOLO11m is the most appropriate detector to predict Geckler classes.

On the other hand, Table 5 also shows that, for predicting quality classes, HIC-YOLO11s has the largest F1-value and the largest values of accuracy and recall. Then, HIC-YOLO11s is the most appropriate detector to predict quality classes. In contrast to predicting Geckler classes, for predicting quality classes, almost evaluation values for detectors are within 0.1 whose difference from the largest value.

In order to investigate the results in Table 5, Table 6 illustrates the evaluation values of AP50, AP75 and mAP for detecting BSE cells and leukocytes and the ratio (%) of images that satisfy the criteria of the number of BSE cells and the number of leukocytes for Geckler classification of Table 1 for all the 808 images through testing.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Table 6: The evaluation values of AP50, AP75 and mAP for detecting BSE cells and leukocytes and the ratio (%) of images that satisfy the criteria for Geckler classification of Table 1.

| | BSE cells | | | | leukocytes | | | |
|---|---|---|---|---|---|---|---|---|
| detector | AP50 | AP75 | mAP | % | AP50 | AP75 | mAP | % |
| YOLO11n | 0.917 | 0.816 | 0.691 | 0.901 | 0.477 | 0.166 | 0.218 | 0.870 |
| YOLO11s | 0.923 | 0.835 | 0.709 | 0.890 | 0.497 | 0.190 | 0.235 | 0.870 |
| YOLO11m | 0.923 | 0.839 | 0.714 | 0.885 | 0.507 | 0.208 | 0.248 | 0.896 |
| YOLO11l | **0.924** | 0.839 | 0.715 | 0.899 | 0.509 | 0.210 | 0.249 | 0.879 |
| YOLO11x | 0.923 | 0.836 | 0.713 | 0.903 | 0.510 | 0.213 | 0.252 | 0.861 |
| HIC-YOLO11n | 0.907 | 0.818 | 0.692 | 0.900 | 0.554 | 0.210 | 0.262 | 0.837 |
| HIC-YOLO11s | 0.914 | 0.834 | 0.707 | 0.892 | 0.571 | 0.232 | 0.279 | 0.866 |
| HIC-YOLO11m | 0.917 | 0.837 | 0.712 | 0.899 | 0.583 | 0.249 | 0.292 | 0.859 |
| HIC-YOLO11l | 0.916 | 0.836 | 0.710 | 0.901 | 0.584 | 0.245 | 0.291 | 0.869 |
| HIC-YOLO11x | 0.917 | **0.842** | **0.716** | 0.902 | 0.586 | 0.251 | 0.293 | 0.855 |
| SOD-YOLO11n | 0.906 | 0.817 | 0.689 | 0.901 | 0.551 | 0.202 | 0.259 | 0.818 |
| SOD-YOLO11s | 0.914 | 0.830 | 0.704 | 0.900 | 0.573 | 0.237 | 0.283 | 0.860 |
| SOD-YOLO11m | 0.914 | 0.836 | 0.710 | 0.890 | 0.581 | 0.242 | 0.288 | 0.845 |
| SOD-YOLO11l | 0.914 | 0.834 | 0.708 | 0.905 | 0.581 | 0.241 | 0.288 | 0.860 |
| SOD-YOLO11x | 0.919 | **0.842** | **0.716** | 0.915 | **0.592** | **0.253** | **0.297** | 0.879 |

Table 6 shows that the detection of BSE cells succeeds whereas the detection of leukocytes fails. In particular, SOD-YOLO11x is the most appropriate to detect BSE cells and leukocytes, whereas the values of mAP to detect BSE cells and leukocytes are 0.716 and 0.297, respectively.

Nevertheless, all the detectors have larger F1-values in Table 5. The reason is that the thresholds for the criteria in Table 1 are small as 10 and 25. In other words, whereas the detectors fail to predict the exact number of BSE cells and leukocytes well, they can predict the sufficient number of BSE cells and leukocytes to determine whether or not the number is larger than 10 or 25 in the criteria, as shown by the column of "%" in Table 6. Note that the successful detectors to predict Geckler and quality classes in Table 5 are YOLO11m and HIC-YOLO11s, respectively, whereas the successful detector to detect BSE cells and leukocytes in Table 6 is SOD-YOLO11x.

Finally, by comparing F1 with $F1_\circ$, the data augmentation for all of the detectors is effective for predicting Geckler classes, whereas the data augmentation for YOLO11s, YOLO11x and SOD-YOLO11x is not effective for predicting quality classes. In particular, the difference between F1 and $F1_\circ$ for predicting Geckler classes is larger than 0.1 for YOLO11n and HIC-YOLO11n.

### 3.3 Discussion

Table 7 illustrates the appropriate classifiers and detectors whose F1-value is the top three largest values of predicting Geckler and quality classes in Tables 4 and 5.

Table 7 shows that, for predicting Geckler classes, the F1-values of image classifiers are larger than those of object detectors. By comparing the largest F1-value, the former is 0.032 larger than the latter. In particular, the F1-values for the object detectors except YOLO11m are at least 0.047 smaller than the largest F1-value of 0.889 of RegNetY-32GF.

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*

Table 7: Appropriate classifiers and detectors of predicting Geckler and quality classes.

| Geckler classes | acc. | pre. | rec. | F1 | quality classes | acc. | pre. | rec. | F1 |
|---|---|---|---|---|---|---|---|---|---|
| RegNetY-32GF | **0.896** | **0.886** | **0.900** | **0.889** | ConvNeXt-B | **0.946** | 0.945 | **0.934** | **0.939** |
| VGG16 | 0.891 | 0.884 | 0.889 | 0.884 | VGG16 | 0.943 | **0.946** | 0.933 | **0.939** |
| ConvNeXt-B | 0.884 | **0.886** | 0.885 | 0.880 | DenseNet-169 | 0.937 | 0.935 | 0.931 | 0.932 |
| YOLO11m | **0.863** | **0.867** | **0.855** | **0.857** | HIC-YOLO11s | **0.973** | 0.960 | **0.968** | **0.963** |
| SOD-YOLO11x | 0.846 | 0.858 | 0.840 | 0.842 | HIC-YOLO11m | 0.969 | **0.962** | 0.959 | 0.960 |
| HIC-YOLO11s | 0.843 | 0.855 | 0.835 | 0.839 | HIC-YOLO11l | 0.968 | 0.959 | 0.960 | 0.959 |

On the other hand, Table 7 also shows that, for predicting quality classes, the F1-values of image classifiers are smaller than those of object detectors. By comparing the largest F1-value, the former is 0.024 smaller than the latter.

# 4 Conclusion

In this paper, we have predicted Geckler classification of Geckler classes and quality classes, by using image classifiers and object detectors. Then, we have confirmed that the image classifiers predict Geckler classes with larger F1-value than the object detectors, whereas the object detectors predict quality classes with larger F1-value than the image classifiers.

Note that, whereas we can predict Geckler classification by applying Gram stained smears images directly (without prepossessing) to image classifiers, it is necessary to annotate BSE cells and leukocytes in Gram stained smears images to predict Geckler classification by object detectors. Hence, the image classifiers are more appropriate to implement automated Geckler classification than the object detectors.

Whereas we have also applied the *transfer learning* [31] to the image classification, we have confirmed that the transfer learning is not effective to predict Geckler classification well. The largest F1-value of predicting Geckler classes is 0.890 for RegNetY-32GF with transfer learning, whereas it is 0.889 for RegNetY-32GF without transfer learning. Also, the largest F1-value of predicting quality classes is 0.936 for RegNetX-32GF with transfer learning, whereas it is 0.939 for VGG16 and ConvNeXt-B without transfer learning.

Then, it is a future work to improve the image classifiers to predict Geckler classification well by applying another methods except transfer learning. As shown in Section 3.2, the object detectors fails to detect leukocytes, so it is a future work to improve the detection of leukocytes. Finally, it is a future work to investigate appropriate methods of data augmentation of predicting Geckler classification.

# References

[1] J. W. Bartholomew, T. Mittwer: *The Gram stain*, Bacteriol. Rev. **16**, 1–29 (1952). https://doi.org/10.1128/br.16.1.1-29.1952.

[2] J.G. Bartlett, N. S. Brewer, K. J. Ryan: *Cumitech 7: Laboratory diagnosis of lower respiratory tract infections*, American Society for Microbiology (1978).

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*

[3] E. M. Chen, S. S. Kasturi: *Microbiology and immunology* (2nd edition), McGraw Hill (2010).

[4] M. Crossman, A. Wiliem, J. Jennings, B. C. Lovell: *A multiple covariance approach for cell detection of Gram-stained smears images*, Proc. ICASSP'15, 932–936 (2015). https://doi.org/10.1109/ICASSP.2015.7178106.

[5] A. Dosoviskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby: *An image is worth 16×16 words: Transformers for image recognition at scale*, Proc. ICLR'21 (2021). https://openreview.net/forum?id=YicbFdNTTy.

[6] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, A. Zisserman: *The PASCAL visual object classes (VOC) challenge*, Internat. J. Comput. Vision **88**, 303–338 (2010). https://doi.org/10.1007/s11263-009-0275-4.

[7] R. W. Geckler, D. H. Gremillon, C. K. McAllister, C. Ellenbogen: *Microscopic and bacteriological comparison of paired sputa and transtracheal aspirates*, J. Clin. Microbio. **6**, 396–399 (1977). https://doi.org/10.1128/jcm.6.4.396-399.1977.

[8] S. Gillespie, K. Bamford: *Medical microbiology and infection at a glance* (3rd edition), Blackwell Publishing (2007).

[9] K. Hashimoto, R. Iida, K. Hirata, K. Matsuoka, S. Yokoyama: *Detecting Geckler classification from Gram stained smears images for sputum*, Proc. ICPRAM'20, 469–476 (2020). https://doi.org/10.5220/0008962304690476.

[10] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adams: *MobileNets: Efficient convolution neural networks for mobile vision applications*, arXiv 1704.04861 (2017). https://doi.org/10.48500/arXiv.1704.04861.

[11] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, H. Adam: *Searching for MobileNetV3*, Proc. ICCV'19, 1314–1324 (2019). https://doi.org/10.1109/ICCV.2019.00140

[12] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger: *Densely connected convolution networks*, Proc. CVPR'17, 2261–2269 (2017). https://doi.org/10.1109/CVPR.2017.243.

[13] G. Jocher: *YOLOv5*, https://github.com/ultralytics/yolov5 (2020).

[14] G. Jocher, J. Qui: *YOLO11 by Ultralytics*, https://github.com/ultralytics/ultralytics (2024).

[15] B. Khalili, A. W. Smyth: *SOD-YOLO: Enhancing YOLOv8 for small object detection in aerial imagery and traffic scenes*, Sensor **24**, 6209 (2024). https://doi.org/10.3309/s24196209

[16] H. Kobata, H. Kashino, K. Hirata: *Detecting bacteria in Gram stained smears images by YOLO and SOD-YOLO*, Proc. ESKM'25 (2025) (to appear).

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*

[17] M. Laposate, P. Mocaffrey: *Clinical laboratory methods: Atlas of commonly performed tests*, McGraw-Hill (2022).

[18] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, S. Xie: *A ConvNet for the 2020s*, Proc. CVPR'22, 11966–11976 (2022). https://doi.org/10.1109/CVPR52688.2022.01167.

[19] L. M. Marler, J. A. Siders, S. D. Allen: *Direct smear atlas: A monograph of Gram-stained preparations of clinical specimens*, Lippincott Williams & Wilkins (2001).

[20] I. Radosavoic, R. P. Kosaraju, R. Girshick, K. He, P. Dollár: *Designing network design spaces*, Proc. CVPR'20, 10425–10433 (2020). https://doi.org/10.1019/CVPR42600.2020.01044.

[21] A. Robinson: *Assessing microbiology test orders and specimen quality*, Clinical Microbiology Newsletter **27**, 78–82 (2005). https://doi.org/10.1016/S0196-4399(05)80030-3.

[22] K. Simonyan, A. Zisserman: *Very deep convolution networks for large-scale image recognition*, Proc. ICLR'15 (2015).

[23] H. Sugimoto, K. Hirata: *Predicting Geckler classification from Gram stained smears images for sputum: Image classification versus object detection*, Proc. ESKM'23, 122–127 (2023). https://doi.org/10.1109/IIAI-AAI59060.2023.00033

[24] H. Tabakotani, H. Kashino, K. Hirata: *Detecting bacteria in Gram stained smears images by HIC-YOLO and Its Variants*, Proc. ESKM'25 (2025) (to appear).

[25] M. Tan, Q. V. Le: *EfficientNet: Rethinking model scaling for convolutional neural networks*, Proc. ICML'19, 6105–6114 (2019).

[26] S. Tang, S. Zhang, Y. Fang: *HIC-YOLOv5: Improved YOLOv5 for small object detection*, arXiv.2309.16393 (2023). https://doi.org/10.48500/arXiv.2309.16393.

[27] The Japanese Respiratory Society: *Diagnosis of hospital-acquired pneumonia and methods of testing for pathogens*, Respirology **14**, 510–522 (2009). https://doi.org/10.1111/j.1440-1843.2009.01572.x.

[28] C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao: *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*, arXiv.2207.02696 (2022). https://doi.org/10.48500/arXiv.2207.02696.

[29] D. F. Welch, M. T. Kelly: *Sputum screening by Nomarski interference contrast microscopy*, J. Clin. Microbio. **9**, 520–524 (1979). https://doi.org/10.1128/jcm.9.4.520-524.1979.

[30] J. Xu, Y. Pan, X. Pan, S. Hoi, Z. Yi, Z. Xu: *RegNet: Self-regulated network for image classification*, IEEE Trans. Neural Netw. Learn. Sys. **34**, 9562–9567 (2023). https://doi.org/10.1109/TNNLS.2022.3158966.

[31] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He: *A comprehensive survey on transfer learning*, Proc. IEEE **109**, 43–76 (2021). https://doi.org/10.1109/JPROC.2020.3004555

*Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.*