Efficient CNN Defect Detection in Sewer Pipes with Application of Active Learning

Catarina Pires, Matthieu Protais & Lisa Warners
CS-433 Machine Learning, EPFL

***Abstract*—Despite sewer defects posing a significant problem to safety and health, monitoring of the sewage systems is limited due to inefficiency and complexity of the process. Deep learning methods on CCTV data have the potential to improve on this. In this work, the accuracy of three architectures (ResNet50, EfficientNetB1 and EfficientNetB7) are compared to discover which is most suitable for daily use in sewer monitoring, paying special notice to low computational complexity. EfficientNetB1, the optimal choice, is further fine-tuned and used to implement a basic form of active learning to ensure steady progress and possible extension of the model while it is already in use.**

I. Introduction

Despite substantial advances in sewer construction over the past decades, defects such as root intrusions, cracks, depositions and joint displacements are still commonplace [1]. This poses problems related to sewer overflow and potential contamination of soil, as well as a more substantial danger of instability and even sinkhole formation [2]. To avoid these detrimental effects, continuous data acquisition and review are necessary, as current strategies for dealing with this problem are reaction-based: municipalities take action as soon as a defect is located. However, models suggest that this approach will become impossible as age-related defects pile up [3], [4].

Monitoring of sewer defects is commonly done via CCTV videos with either direct or retrospective supervision [5]. This approach is however highly problematic for various reasons. First, the classification of defects has been shown to be unreliable, especially as the complexity of labelling increased over the past few years [6], [7]. Secondly and most importantly, the process is time-consuming [6]. This is mainly caused by the need to manually review all videos, further complicated by their duration. Due to the device documenting every region of interest extensively by turning, tilting and zooming, its velocity is low. This sluggish process, combined with limited available budget, leads to only 30-40% of sewer pipes being examined.

Considering the importance of monitoring as well as the inadequacy of current approaches, a growing body of research is devoted to automating the detection process of defects. This work serves as a proof-of-concept that a user-friendly, computationally cheap model can give good results for defect detection in sewer systems.

II. Related work

A. Computer vision and machine learning

Initial attempts to automate defect detection in sewers were based on classical computer vision techniques [8], especially for crack detection [9], [10] and improvements on image quality and distortion [11], [12]. However, these approaches are often laborious and difficult to generalise for different acquisition methods and conditions [13]. In addition, they often only consider one type of defect at the time, limiting their usefulness.

First attempts to apply machine learning techniques included multi-layer perceptron, occasionally combined with fuzzy logic [14], [15]. A one-class support vector machine (OCSVM) has been trained using only undamaged sewer pipes as input images, as to account for the fact that defects are relatively rare [16]. However, as the name suggests, an OCSVM only allows for anomaly detection and not for classifying types of defects. Unfortunately, multiple classes based on a support vector machine (SVM) resulted in low accuracy values [17]. More recently, Dang *et al.* [18] combined a cost-sensitive and an ensemble learning technique in order to deal with unbalanced data.

B. Deep learning

Deep learning techniques provide a solution to small data sets and inefficient feature extraction. As sewer data consists of images, Convolutional Neural Networks (CNNs) are most commonly used. A CNN is composed of three distinct types of layers. The fully connected part is a classical neural network, which simply classifies images with weights governing the transition between layers. These weights are continuously improved by backpropagation, i.e. computation of the gradient of the loss. The issue of simple neural networks is that the images are expressed as flattened vectors and the spatial dependency of the pixels is lost. The purpose of the convolutional and pooling layers is to retain this information. The convolutional layer separates the initial picture into distinct channels, allowing the neural network to extract patterns. The role of the pooling is to reduce the dimension of the input, by spatially averaging colour values.

As features are extracted throughout the learning process, CNNs do not require the pre-processing nor the specific design of classical computer vision. An added benefit is the more rigorous and consistent detection than possible by humans. A series of papers by Kumar, Cheng and Wang

developed application of CNNs to regional detection of multiple types of defects in sewers [4], [13], [19], [20]. These models incorporate detection and localisation in the image, finally applying multiple object tracking to incorporate sequential information. However comprehensive, it remains to be seen whether these extra computational costs are justified. Additionally, acquisition of segmented images is relatively difficult and labour-intensive. A general drawback of applying pre-trained CNN architectures is the lack of insight into feature importance. Work by Kumar et al. [21] improved upon the impretability by implementing class activation mapping (CAM), including recommendations on proper data augmentation. A final recent advance involved application of a Generative Adversarial Network (GAN) to this problem [22]. This type of unsupervised learning has been proven highly successful, however it requires a large body of high-quality data to be trained efficiently.

C. Active learning

In general, acquisition of a large annotated data set has been an issue in this field, due to poor generalisation to different acquisition methods and required expertise for labelling [4]. A natural way to expand the data set for a specific classification and sewer system is active learning, a subset of machine learning in which the algorithm can query a user interactively to label data. These queries are usually in the form a request to a human annotator to label an unlabelled image. The algorithm prioritizes data with the highest impact on the weights, therefore being of interest when there is too much unlabelled data and priorities are required to work efficiently [23] [24].

In order to use active learning on an unlabelled data set, first a model must be initialised, providing insight on which areas of the parameter space should be labelled first to improve classification. Subsequently, the model generates predictions for the unlabelled data points. A score is chosen for each one based on the prediction, e.g. based on the entropy. This process can be repeated iteratively and a new model is trained on the new labelled data set. The unlabelled data points can be ran through this updated model to recalculate the prioritisation scores, therefore optimising the labelling strategy as the model continuously improves.

III. Models and methods

A. Data acquisition and augmentation

The presented work was done in collaboration with the city of Lausanne, Switzerland, who provided a set of sewer pipe videos to train and test the discussed models presented. Considering the acquisition speed and possible redundancy, every tenth frame of a particular video was kept, giving a 2.5 frames/s sampling rate. Two types of defects could be labelled: "fissure" (crack) and "racines" (roots), with all other images being labelled as "normal".

Data augmentation techniques were applied to the training data to improve generalisation and expand the data set. The transformations performed were the following: brightness variation, horizontal and vertical flips, rotation, contrast variations, vertical and horizontal translations, and a combination of horizontal and vertical flips with rotation, and contrast variations with rotation. The choice for these transformations follows from [21].

B. Comparison of models

1) Applied models: In 2015, the Visual Geometry Group of the University of Oxford published VGGNet [25], which follows the CNN structure explained in section II. However, VGGNet soon encountered the so-called vanishing gradient problem: an infinitely small gradient during back propagation, stopping the learning process [26].

To solve this, the concept of Residual Network (ResNet) was introduced in 2015 [27], based on shortcuts [28], [29]. Given three layers in a neural network, a shortcut connection is a path which takes the input from layer $n-2$ into the layer $n-1$ and adds it to the output of layer n, as shown in figure 1. The purpose of these shortcuts is to avoid vanishing gradient and so-called degradation. In a traditional NN, degradation appears for a large number of hidden layers. It is a reduction of training accuracy due to the over-complexity of the model with respect to the spatial dependency of the data [30].

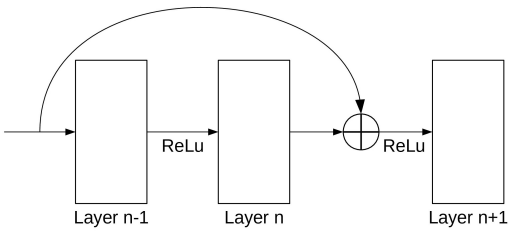

Figure 1. A shortcut connection, the building block of a residual network

ResNet can be trained far deeper than VGGNet without decrease in accuracy or training rate. As ResNet uses global average pooling instead of fully connected layers, it also requires much less parameters. In this work we therefore will be using ResNet50, i.e. the ResNet model with 50 layers, as a trade-of between accuracy and efficiency.

To further reduce the number of parameters, EfficientNet was created [31]. CNN accuracy can be improved by increasing the width or the depth of the network, or the resolution of the images. Interdependence of these dimensions results in the compound scaling method: scaling the three variables uniformly by a set of fixed constants. For example, to have 2^N times more computational resources, the network depth, width and resolution are scaled by respectively α^N, β^N and γ^N, for α, β and γ fixed scalars [31]. In 2020, this made

EfficientNet both one of the most efficient and most accurate neural networks. Gradually increasing its dimensions give us the models EfficientNetB0 to EfficientNetB7. This work compares the models EfficientNetB1 and EfficientNetB7, as well as the earlier discussed ResNet50, to determine the most suitable one for this application.

2) Transfer learning and optimisation: Three models have been compared as feature extractors for sewer systems using transfer learning. The output was flattened and a dense layer including a 0.5 dropout was added to the end. The training set contained 68283 (after data augmentation was performed on a initial set of 843 images) and the validation set 700 images. To compile the models, the optimiser and loss function have been specified. For multi-class classification problems, multi-class cross entropy is commonly applied and provides an intuitive way to define the error [32]. Adam was selected as optimiser [33], which is an adaptation of first-order stochastic gradient descent. Adam has requires learning rate optimisation in order to reliably compare model architectures. Four learning rates were compared on the logarithmic scale $logspace(-4, -2, 4)$. The second hyperparameter that must be optimised is the number of epochs. As the data set used for this paper is relatively small, the danger of overfitting is imminent, as seen by a decrease in the validation error. Therefore, the models were only trained for up to 5 epochs.

The best feature extractor was chosen based on this analysis while taking into account the number of parameters, and fine-tuned further by unfreezing the last ten layers. The obtained weights were applied to construct a simple version of active learning.

IV. Results

The achieved accuracies for the compared models are displayed in table I. Interestingly, the maximum achieved validation accuracies are very close together, despite vast differences in model complexity and number of parameters. This could be partially caused by the small data set. It is in the line of expectation that EfficientNetB7 achieves the highest accuracy, though it is only marginally higher, possibly because this model only contains three classes. Considering B7's large number of parameters and the focus on computational efficiency, EfficientNetB1 was chosen to further develop for this application.

The transfer learning process was repeated, reaching 97.8% accuracy. Building on this, the last 10 layers of the model were unfrozen to fine-tune the high-level features and make it more applicable to sewer images. Early stopping was implemented to automatically monitor the validation accuracy. This allowed fine-tuning to progress for four epochs, reaching a final test accuracy of 98.2%. 11 out of 600 images were wrongly classified, all related to doubt between normal pipes and cracks (see figure 3 in appendix). Four different types of misclassifications were identified,

learning rate	epoch	EfficientNetB1	ResNet50	EfficientNetB7
0.0001	3	0.9743	0.9771	0.9771
	4	0.9714	0.9729	0.9787
	5	0.9714	0.9757	0.9787
0.00046	3	**0.9786**	**0.9786**	0.9729
	4	0.9743	0.9643	**0.9814**
	5	0.9757	0.9771	0.9787
0.0022	3	0.9729	0.9457	0.9443
	4	0.9757	0.9657	0.9771
	5	0.9743	0.9471	**0.9814**
0.01	3	0.9743	0.7900	0.8043
	4	0.9700	0.8171	0.7914
	5	0.9727	0.7686	0.7886

Table I. Validation accuracies of the models for different learning rates. Models are ordered from low to high number of parameters. Best accuracy value per model is printed in bold.

shown in figure 2. Type A will in general not be a problem, as the final goal of the model is to work on unzoomed videos. Type B can be erased by more carefully accumulating the data set. Type C and D, however, are related to finer image features. Classification of these might be improved by more deeply fine-tuning the model using sewer images.

Once the model was trained with the augmented labelled data, predictions were obtained on unlabelled data. A prioritisation score is calculated based on these predictions, which then is used to select a batch of images to be labelled. The prioritisation score chosen was entropy, defined as follows:

$$s_E = \arg\max_x \left(-\sum_i \mathbb{P}(\hat{y}_i|x) \log \mathbb{P}(\hat{y}_i|x) \right) \quad (1)$$

After the batch of images to be labelled is selected, the algorithm queries a human annotator to do so. In order to make the labelling interface intuitive and user friendly, the labelling is done using *Jupyter* widgets and the library *superintendent 0.5.3* [34]. Once labelling had been done, a new set of training data is generated from the labelled batch and used to further train the model. This model may then be used to make predictions of further unlabelled data to select a new labelling batch, therefore increasing the new training set and further train the model. Each version is saved to keep checkpoints of the progress. As the previous training data mainly featured two types of defects while the new data contained images of unknown defects, there was not an accurate label available for all. As a solution, the label "Racines" was changed to "Racines_Extrusion" to included a wider range of defects. Finally, a deletion widget was added for highly blurry or unidentifiable images.

With the intent of testing this implementation, frames obtained from four unlabelled videos were used and the algorithm was run 10 times, selecting 100 images to be labelled in each iteration. Due to the small number of images in the newly generated labelled training set, the tuning of the model is initially done using 3 epochs. As the data set grew and a decrease in accuracy was observed, after the sixth run

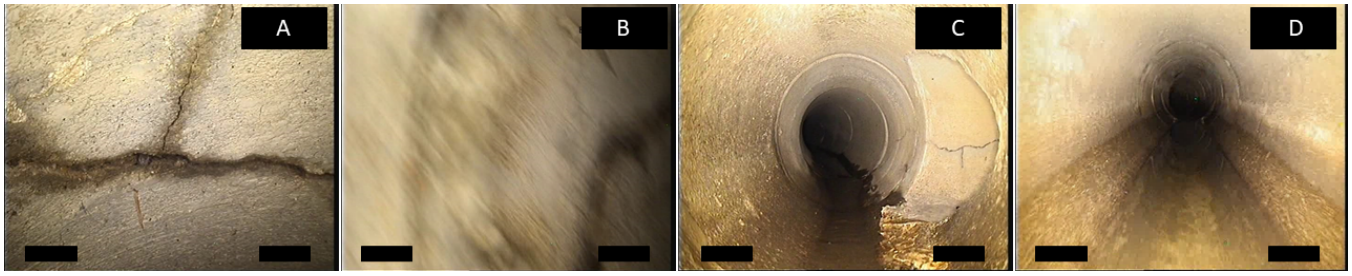

Figure 2. Four representing images of the misclassified instances. Cracks misclassified as normal are zoomed in (A), blurry (B), and branching (C) (occurring 3, 2, and 2 times resp.). Normal misclassified as cracks is likely due to marks of water damage (D) (occurring 4 times).

iteration	epoch	accuracy	validation accuracy
1	3	0.9500	0.9371
2	3	0.9394	0.9200
3	3	0.9486	0.8943
4	3	0.9177	0.8986
5	3	0.9465	0.8243
6	3	0.9034	0.8000
7	5	0.9589	0.7843
8	5	0.9456	0.7371
9	5	0.9617	0.6043
10	5	0.9601	0.5957

Table II. Accuracies and validation accuracies for each iteration of the Active Learning algorithm

5 epochs were used in an effort to achieve better accuracy. The results obtained are shown in table II.

Finally, the model obtained at the end of the tenth iteration was used to make predictions to the test set previously mentioned. The results are shown on a confusion matrix in the appendix.

V. Discussion and conclusion

The data set used in this report to compare ResNet and the more recent but cheaper EfficientNet came solely from two labelled videos and was expanded using data augmentation. The models were compared on transfer learning. Preliminary fine tuning was done building further on EfficientNetB1, resulting in a set of weights. This initial data set was fairly small. The use of a larger data set would improve the model, potentially using a publicly available data set such as [35]. Alternatively, extensive use of active learning might be slower, but results in a very specific data set for the given sewer system. Another important improvement is the extension to a larger number of classes. Application of the discussed model on a CCTV video reveals that many deposits are now classified as roots, likely due to the model recognising them as foreign intrusions. However, the practical treatment of this type of defect is very different. It will therefore be important to expand the model in future development.

The problem of an unbalanced data set as described in section II was encountered in this model. The training set contained a larger number of defects than an arbitrary video would. This resulted in relatively many defects being detected in a single video. Application of methods such as described in [18] could improve this model characteristic. Additionally, the active learning will, as discussed before, also allow for better generalisation if used properly.

Regarding active learning, there is a decrease in accuracy up to the point when the number of epochs was changed from 3 to 5. It subsequently increased back to the original value, surpassing it at a final value of 96.01% and maximum value of 96.17% in iteration 9. The validation accuracy, however, decreases in every iteration, with a final value of 59.57% and a maximum value in the first iteration of 93.71%. This suggest that the model generalises badly through iterations. This may be due to the fact that the initial labelled validation, test and training data (though augmented) were not representative of all the defects that might be found in sewers systems. In addition, the newly labelled data was stored in such a way that the newly labelled batches were added to the previously labelled batches. Therefore, the batches from previous iterations were continuously used to train the model, resulting in overfitting. Solely using the newly labelled data and labelling bigger batches at a time might prove more beneficial. From the confusion matrix in the appendix it is clear that the model continues to show difficulty distinguishing between ”normal” and ”fissure”. The broadening of the label ”racines” also proved to make the model more uncertain, which was expected.

In conclusion, EfficientNetB1 was selected and fine-tuned for defect detection in sewage systems, showing good accuracy values. To improve on its generalisation capabilities and allow for a natural expansion of the data set, a basic form of active learning was implemented. This model provides a baseline for future automation of defect detection and as such, a safer and more efficient sewer maintenance.

Acknowledgements

We would like to Mr. Antonio da Silva, chef du Pôle Développement et Intégration of the city of Lausanne, for his input and the entire team for providing the data (and keeping the ground under our feet stable). Additionally, we want to thank Dr. Martin Jaggi for his supervision.

REFERENCES

[1] Z. Yazdanfar and A. Sharma, "Urban drainage system planning and design – challenges with climate change and urbanization: a review," *Water Sci Technol*, vol. 72, no. 2, p. 165–179, 2015.

[2] United States Environmental Protection Agency. Sanitary Sewer Overflows (SSOs). [Online]. Available: https://www.epa.gov/npdes/sanitary-sewer-overflows-ssos Accessed 06-12-2021.

[3] N. Caradot, H. Sonnenberg, I. Kropp, A. Ringe, S. Denhez, A. Hartmann, and P. Rouault, "The relevance of sewer deterioration modelling to support asset management strategies," *Urban Water J*, vol. 14, no. 10, pp. 1007–1015, 2017.

[4] S. S. Kumar, D. M. Abraham, M. R. Jahanshahi, T. Iseley, and J. Starr, "Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks," *Autom Constr*, vol. 91, pp. 273–283, 2018.

[5] M. R. Halfawy and J. Hengmeechai, "Efficient algorithm for crack detection in sewer images from closed-circuit television inspections," *J Infrastruct Syst*, vol. 20, no. 2, p. 04013014, 2014.

[6] R. R. Harvey and E. A. McBean, "Predicting the structural condition of individual sanitary sewer pipes with random forests," *Can J Civ Eng*, vol. 41, no. 4, pp. 294–303, 2014.

[7] A. J. van der Steen, J. Dirksen, and F. H. Clemens, "Visual sewer inspection: detail of coding system versus data quality?" *Struct Infrastruct Eng*, vol. 10, no. 11, pp. 1385–1393, 2014.

[8] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Adv Eng Inf*, vol. 29, no. 2, pp. 196–210, 2015.

[9] R. G. Lins and S. N. Givigi, "Automatic crack detection and measurement based on image analysis," *IEEE Trans Instrum Meas*, vol. 65, no. 3, pp. 583–590, 2016.

[10] S.-N. Yu, J.-H. Jang, and C.-S. Han, "Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel," *Autom Constr*, vol. 16, no. 3, pp. 255–261, 2007.

[11] P. Prasanna, K. J. Dana, N. Gucunski, B. B. Basily, H. M. La, R. S. Lim, and H. Parvardeh, "Automated crack detection on concrete bridges," *IEEE Trans Autom Sci Eng*, vol. 13, no. 2, pp. 591–599, 2016.

[12] Z. Chen and T. C. Hutchinson, "Image-based framework for concrete surface crack monitoring and quantification," *Adv Civ Eng*, vol. 2010, 2010.

[13] J. C. P. Cheng and M. Wang, "Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques," *Autom Constr*, vol. 95, pp. 155–171, 2018.

[14] M. J. Chae and D. M. Abraham, "Neuro-fuzzy approaches for sanitary sewer pipeline condition assessment," *Journal of Computing in Civil Engineering*, vol. 15, no. 1, pp. 4–14, 2001.

[15] T. Shehab and O. Moselhi, "Automated detection and classification of infiltration in sewer pipes," *Journal of Infrastructure Systems*, vol. 11, no. 3, pp. 165–171, 2005.

[16] J. Myrans, Z. Kapelan, and R. Everson, "Using automatic anomaly detection to identify faults in sewers," 2018, 1st International WDSA / CCWI 2018 Joint Conference.

[17] M.-D. Yang and T.-C. Su, "Automated diagnosis of sewer pipe defects based on machine learning approaches," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1327–1337, 2008.

[18] L. M. Dang, S. Kyeong, Y. Li, H. Wang, T. N. Nguyen, and H. Moon, "Deep learning-based sewer defect classification for highly imbalanced dataset," *Computers & Industrial Engineering*, vol. 161, p. 107630, 2021.

[19] S. S. Kumar, M. Wang, D. M. Abraham, M. R. Jahanshahi, T. Iseley, and J. C. P. Cheng, "Deep learning-based automated detection of sewer defects in cctv videos," *J Comp Civ Eng*, vol. 34, no. 1, p. 04019047, 2020.

[20] M. Wang, S. S. Kumar, and J. C. P. Cheng, "Automated sewer pipe defect tracking in cctv videos based on defect detection and metric learning," *Autom Constr*, vol. 121, p. 103438, 2021.

[21] S. S. Kumar, D. Abraham, and M. Rosenthal, "Leveraging visualization techniques to develop improved deep neural network architecture for sewer defect identification," construction Research Congress 2020. [Online]. Available: 10.1061/9780784482858.089

[22] Z. Situ, S. Teng, H. Liu, J. Luo, and Q. Zhou, "Automated sewer defects detection using style-based generative adversarial networks and fine-tuned well-known cnn classifier," *IEEE Access*, vol. 9, pp. 59 498–59 507, 2021.

[23] A. Solaguren-Beascoa, "Active Learning in Machine Learning," https://towardsdatascience.com/active-learning-in-machine-learning-525e61be16e5, 2020, [Online].

[24] S. Mhosein, "Active Learning: Curious AI Algorithms," https://www.datacamp.com/community/tutorials/active-learning?fbclid=IwAR34SCj3p5YbwpRMddDNY_rM6OUxLs703cf54Grqn8JrLPp2Lb8xRGPwBL8, 2018, [Online].

[25] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015.

[26] T.-V. Pricope, "An analysis on very deep convolutional neural networks: Problems and solutions," *Studia Universitatis Babeș-Bolyai Informatica*, vol. 66, p. 5, 2021.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," 2015.

[28] S. Targ, D. Almeida, and K. Lyman, "Resnet in Resnet: Generalizing Residual Architectures," 2016.

[29] S. Li, J. Jiao, Y. Han, and T. Weissman, "Demystifying ResNet," 2017.

[30] C. S. Wickramasinghe, D. L. Marino, and M. Manic, "ResNet Autoencoders for Unsupervised Feature Learning From High-Dimensional Data: Deep Models Resistant to Performance Degradation," *IEEE Access*, vol. 9, pp. 40 511–40 520, 2021.

[31] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv:1905.11946*, 2020.

[32] V. Martinek. (2020, 05) Cross-entropy for classification. [Online]. Available: https://towardsdatascience.com/cross-entropy-for-classification-d98e7f974451 Accessed 14-12-2021.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization v9," 2017, 3rd International Conference for Learning Representations. [Online]. Available: arXiv:1412.6980

[34] J. Freyberg, "Superintendent documentation," https://superintendent.readthedocs.io/en/latest/, 2018, [Online].

[35] J. B. Haurum and T. B. Moeslund, "Sewer-ml: A multi-label sewer defect classification dataset and benchmark," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 13 456–13 467.

APPENDIX

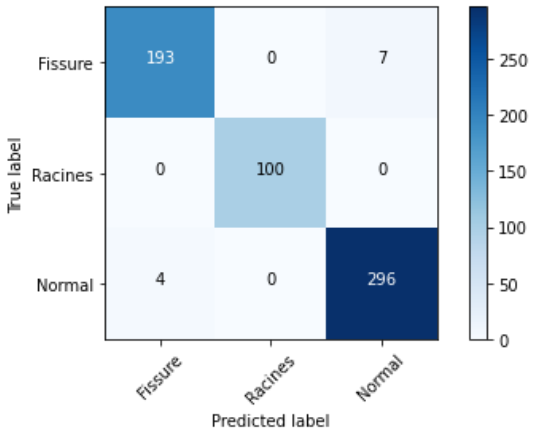

Figure 3. Confusion matrix of the predicted and the true labels. Produced using the test set of 300 normal images, 100 racines, and 200 fissures.

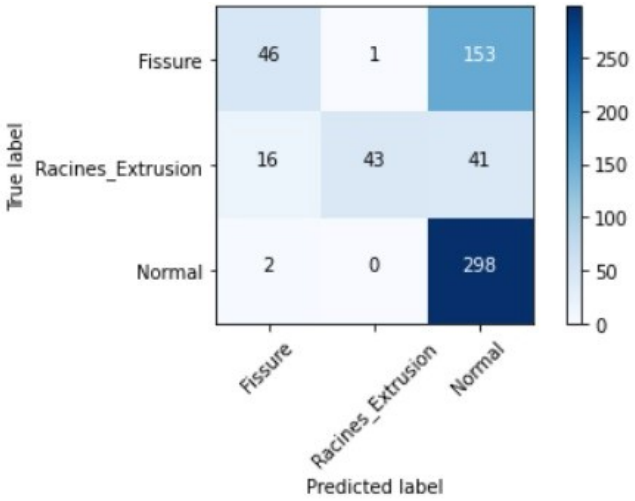

Figure 4. Confusion matrix of the predicted and the true labels. Produced using the test set of 300 normal images, 100 racines, and 200 fissures.