# PEAR AND APPLE RECOGNITION USING DEEP LEARNING AND MOBILE

**Sergejs Kodors[1], Gunars Lacis[2], Vitaliy Zhukov[1], Toms Bartulsons[2]**
[1]Rezekne Academy of Technologies, Latvia;
[2]Latvia University of Life Sciences and Technologies, Latvia
sergejs.kodors@rta.lv, gunars.lacis@llu.lv, vz16020@edu.rta.lv, toms.bartulsons@gmail.com

**Abstract.** Apple and pear are among the most widely grown and economically important fruit species worldwide and in Latvia. In turn, scab diseases caused by ascomycetous fungi *Venturiainaequalis* and *V. pyrina*, are economically the most important diseases worldwide. Durable plant resistance has been widely regarded as the preferred disease limitation method due to environmental and food safety concerns. Whereas in cases where the use of pesticides cannot be avoided, their applications should be more precise, more targeted and reduced substantially. One way how to realize it is the smart and precision horticulture that can greatly increase the effectiveness of pesticides and use them more selectively. The smart and precision horticulture relies heavily on new technologies and digitalization, including sensing technologies, software applications, communication systems, telematics and positioning technologies, hardware and software systems, data analytics solutions, as well as knowledge linking biological information to data technologies. The aim of our project – development and implementation of mobile application with deep learning system for early identification and evaluation of apple and pear scab. The specific of project – the image processing must be completed by a mobile device without image upload into GPU cluster. This research presents the comparison of deep learning architectures developed for mobile devices (MobileNet and MobileNetV2). The classification precision and speed of neural networks are compared using open dataset"Fruits-360". The results are applicable to develop transfer learning and domain adaptation solutions. Meanwhile, decomposition into many simple subtasks can reduce required device resources to complete complex analysis using mobiles, as well as to create trustworthy AI model. The model of MobileNetV2 showed the best results: total accuracy 0.998, Cohen's Kappa 0.991 and latency 212ms/step.

**Keywords:** apple, deep learning, mobile, pear, scab.

## Introduction

Apples (*Malus* x *domestica*) and pears (*Pyruscommunis*) are widely grown fruit species worldwide and in Latvia and are among the most important fruit crops commercially [1]. The fruit production, especially in the integrated and organic production systems, is constrained by pathogenic organisms and diseases, which reduce the viability of plants, fruit development and quality. Changes in production technologies, introduction of new cultivars, as well as climatic changes can alter the pathogen populations resulting in development of new and more aggressive forms adapted to the changing environmental conditions. Scab diseases on apple and pear caused by two ascomycetous fungi *Venturiainaequalis* and *Venturiapirina*, respectively, are economically important diseases worldwide [2-5]. Severe damages by scab diseases can be caused on both apples and pears in conducive conditions in all main growing regions [5; 6].

Currently scab control heavily relies on frequent fungicide applications, the use of scab resistant cultivars and management of leaf litter [6; 7]. In some regions with climate conditions highly favourable for the apple scab, 15-20 fungicide applications may be required to control it [7]. Due to environmental and food safety concerns, and high adaptation capability of these pathogens to overcome resistance, need for changes in growing and breeding strategies have been highlighted by apple and pear scab research community [5; 8]. Durable plant genetic resistance, which could be achieved by combination of resistance from several major genes and non-specific resistance, was widely regarded as the preferred method to control these diseases in future [5]. Finding broader and diverse genetic resources can provide resources for breeding of durable host resistance.

In cases where the use of pesticides cannot be avoided, their applications should be more precise, more targeted and reduced substantially. One way to solve it is the smart farming and precision horticulture that can greatly increase the effectiveness of pesticides and use them more selectively. This approach is supported at different political levels [9]. In practice the smart and precision horticulture typically, but not necessarily, relies heavily on new technologies and digitalization, including sensing technologies, software applications, communication systems, telematics and positioning technologies, hardware and software systems, data analytics solutions, as well as knowledge linking biological information to data technologies. Key implementation challenges for precision horticulture are following: 1) There is still a lack of accurate, scientifically-based

phenotyping techniques to develop early identification and sufficiently accurate disease prediction systems; 2) There is also a lack of linkage of existing biological information (pathogenicity of pests, plant resistance, various plant-pathogen interaction reactions) with sensory information (various types of imagining) for its automated analysis; 3) The growing amount of data and its disorder is also a challenge: research projects have produced research data covering various aspects of plant-pathogen interactions (plant field resistance, identification of plant resistance genes, pathogenicity assessment of the causative agent, etc.), but there is no internal linkage analysis, as well as implementation of other types of data (such as environmental and meteorological data, etc.).

The deep learning (DL) is the modern approach to develop solution for the smart farming. For example, Kang and Chen developed DaSNet architectures for autonomous apple harvesting with lightweight backbone based on residual network architecture experimenting with ResNet-50 and ResNet-101 [10]. Meanwhile, Bresilla et al. used YOLO-V2 to detect apples on tree [11], but Tian et al. applied YOLO-V3 network to perform real-time detection of apples to monitor and evaluate the growing of apples in orchards [12].

The aim of our project is to develop mobile application for early identification and evaluation of apple and pear scab, when analysis must becompleted by a mobile device without image upload into GPU cluster. There are different architectures developed exactly for mobile and embedded devices: MobileNet [13], EffNet [14], ShuffleNet [15] and MobileNetV2 [16]. YOLO architecture is light solution too, because it is prepared to complete detection in real time.Butthe scab detection task considers to make photo of single object to identify infected plant in the primary stageand it is complex task to visually identify the scab for experts. Therefore, a robust neural network model with high-accuracy potential and small processing latency must be selected. In the next experiments, it is possible to adapt selected model for encoding-decoding architecture to segment scab and measure its stage.

The aim of this research is to compare DL architectures developed for mobile and embedded devices considering domain of pears and apples classification. MobileNet and MobileNetV2 architectures were selected for experiment. The results of study can be applied to the smart system development, which can be used for food quality monitoring, the smart and precision horticulture domain.

## Materials and methods

A dataset "Fruits-360" (version 2019.09.21.0) [17] was applied to compare DL architectures for mobile devices. Considering the domain-specific task, the dataset was pre-processed. All classes of apples and pears were grouped into one category, other fruits – into category "Other" (see Fig. 1). Validation and training datasets have proportion 25 % and 75 %. The dataset has 120 classes of fruits and provides 6 416 images of apples and 3 640 images of pears, which are obtained photographing fruits, vertically and horizontally turning them at 360 degrees. The validation dataset contains 20 622 images, therefore the precision 0.001 is applicable considering "rule 30". The image augmentation is applied for the training dataset: horizontal flip, rotation $30^o$, shift 0.1 and zoom 0.1.
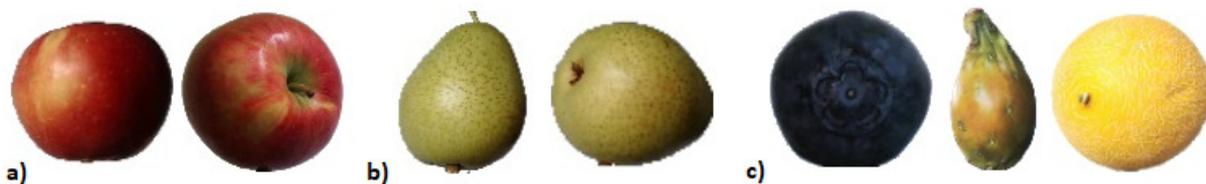


Fig. 1. **Examples of images of Fruits360 dataset:** a – apples; b – pears; c – other

Machine learning tools, Keras and TensorFlow 2.0,were applied to complete experiments using a computer equipped with NVIDIA GTX 1050 video card, which has 2GB and 640 CUDA cores. DL architectures, AlexNet [18], MobileNet and MobileNetV2, are compared in the experiment. Additionally, AlexNet model adapted for mobileis tested, too. The adapted model has all Conv2D layers replaced by SeparableConv2D layers proposed in the MobileNet architecture. All models were trained using the Adam algorithm.

AlexNet is architecture, which significantly outperformed all prior competitors in ImageNet 2012 and strongly influenced into computer vision development. Therefore, AlexNetis used to draw a baseline for comparison.

The primary models of AlexNet and MobileNet architectures have input tensors (227x227x3) and (224x224x3), but Fruits-360 images have size equal to 100x100px. Therefore, each architecture had been experimentally adapted for Fruits-360 dataset, and then each developed model wasassessed 10 times to calculate mean and median accuracy to consider randomness arisen in training time. The accuracy of trained model is calculated using the total accuracy (1) and the Cohen's kappa coefficient [19] (2) based on confusion matrix of validation dataset.

$$A = P / N,$$  (1)

where    $A$ – total accuracy;
           $P$ – positive rate;
           $N$ – total population.

$$K = \frac{N \sum_{i=1}^{m} C_{ii} - \sum_{i=1}^{m} \left( C_i \cdot C_i' \right)}{N^2 - \sum_{i=1}^{m} \left( C_i \cdot C_i' \right)},$$  (2)

where    $K$ – Cohen's kappa coefficient;
           $N$ – number of samples;
           $C_{ii}$ – number of correctly predicted samples of class $i$ (diagonal of confusion matrix);
           $C_i$ – number of samples, which belong to class $i$;
           $C_i'$ – number of samples classified as class $i$.

**Results and discussion**

The adapted models of DL architectures for image classification and their accuracies are provided in Table 1 and in Fig. 2.

A box-plot diagram (see Fig. 2) shows strong difference between AlexNet and the set of models {AlexNet(SConv2D), MobileNetV1 and MobileNetV2}. The difference among AlexNet (SConv2D), MobileNetV1 and MobileNetV2 are assessed using statistical analysis. Mann-Whitney-Wilcoxon test is applied, because populations are unpaired and the number of measurements is smaller than 30. The results of the test are shown in Table 2. The statistical analysis proves that precision of AlexNet model is the smallest among all models. However, the significant difference among other models is not identified ($\alpha = 0.05$).

Keras library provides prebuilt models of MobileNetV1 and MobileNetV2. These models with option "alpha = 0.5" showed next results:

- MobileNetV1: total accuracy ≈1.000, Cohen's Kappa = 0.998, 109 ms·step$^{-1}$, size 3 210 051;
- MobileNetV2: total accuracy = 0.999, Cohen's Kappa = 0.998, 171 ms·step$^{-1}$, size 2 227 715.

Therefore, the domain-adapted models are not strongly worse with regard to the accuracy, but their sizes are significantly smaller. Comparing AlexNet and AlexNet (SConv2D) models, the separated convolution layer strongly influences into accuracy of neural network simultaneously reducing its size. A block of bottleneck with ReLU6 proposed by MobileNetV2 constructs a good solution to decrease the size and save accuracy.

The authors of Fruits-360 dataset experimented with AlexNet architecture, too [17]. Their model had similar number of filters, but it applied kernel sizes 5x5 and strides (2, 2). The obtained accuracy was 0.944 for RGB colour model that is close to our result 0.974. Meanwhile, other authors trained AlexNet model with only 2 blocks of convolution and pooling layers to recognize images with accuracy 0.998 [20]. Their model contained 64 kernels per convolution layer and 500 neurons per fully-connected layer.

The disadvantage of experiments with neural network models trained on Fruits-360 dataset are laboratory conditions – images with single object and white background. Therefore, the trained models cannot classify the natural images, but it is not important to the experiment, because the unified dataset provides the similar semantic features, which must be recognized, that provides possibility to compare

the effectiveness of each architectures. Our project considers to create own dataset with apples and pears filmed in natural conditions. However, development of own dataset is a time consuming process, therefore, approach with synthetic data generation can be applied mixing synthetic images with natural images. For example, this approach was applied by authors, who trained YOLO models to detect apples [11]. It is worth noting that the authors of YOLO model for apple recognition conclude about limitation of their platform, which is the compute power needed for the system to run. Therefore, it is a perspective assumption to continue research combining two ideas related with MobileNet2V and synthetic data application.

Table 1

## Models of DL architectures

| AlexNet | AlexNet (SConv2D) | MobileNetV1 | MobileNetV2 |
|---|---|---|---|
| conv2d, 16, 5x5, s2, relu<br>batch norm<br>max pool, 2x2<br>conv2d, 32, 3x3, relu<br>batch norm<br>max pool, 2x2<br>conv2d, 64, 3x3, relu<br>conv2d, 128, 3x3, relu<br>max pool, 2x2<br>fcl, 512, relu<br>fcl, 512, relu<br>fcl, 3, softmax | s_conv2d, 16, 5x5, s2, relu<br>batch norm<br>max pool, 2x2<br>s_conv2d, 32, 3x3, relu<br>batch norm<br>max pool, 2x2<br>s_conv2d, 64, 3x3, relu<br>s_conv2d, 128, 3x3, relu<br>max pool, 2x2<br>fcl, 512, relu<br>fcl, 512, relu<br>fcl, 3, softmax | conv2d, 16, 3x3, s2, relu<br>batch norm<br>fcl, 16, tanh<br>s_conv2d, 32, 3x3, s1, relu<br>batch norm<br>s_conv2d, 64, 3x3, s1, relu<br>batch norm<br>*block A( k ):*<br>s_conv2d, $k$, 3x3, s1, relu<br>batch norm<br>s_conv2d, $k$, 3x3, s2, relu<br>batch norm<br>*block A*(128)<br>*block A*(256)<br>*block A*(512)<br>glob avg pool<br>fcl, 512, relu<br>fcl, 3, softmax | c0: conv2d, 16, 3x3, s2, relu6<br>c1: batch norm<br>*block A( k, z ):*<br>s_conv2d(z), $k$, 3x3, s1, relu6<br>batch norm<br>*block B( k, z ):*<br>s_conv2d(z), $k$, 3x3, s2, relu6<br>batch norm<br>c2: conv2d, 8, 1x1, s1, relu6<br>c3: batch norm<br>c4: *block A( 8, c3 )*<br>c5: *block B( 16, {c3, c4} )*<br>c6: *block B( 24, c5 )*<br>c7: *block B( 32, c6 )*<br>c8: conv2d, 64, 1x1, s1, relu6<br>c9: batch norm<br>c10: *block A( 64, c9 )*<br>c11: *block B( 256, {c9, c10} )*<br>c12: glob avg pool<br>c13: fcl, 256, relu6<br>c14: fcl, 3, softmax |
| *Trainable params* | | | |
| 952 835 | 866 750 | 801 795 | 98 379 |
| *Latency* | | | |
| *35ms/step* | *137s/step* | *287ms/step* | *212ms/step* |
| *Total accuracy* | | | |
| *min* = 0.967<br>*mean* = 0.970<br>*med* = 0.971<br>*max* = 0.974 | *min* = 0.978<br>*mean* = 0.995<br>*med* = 0.998<br>*max* = 0.999 | *min* = 0.976<br>*mean* = 0.991<br>*med* = 0.992<br>*max* = 0.999 | *min* = 0.993<br>*mean* = 0.997<br>*med* = 0.998<br>*max* = 0.999 |
| *Cohen's Kappa* | | | |
| *min* = 0.883<br>*mean* = 0.893<br>*med* = 0.890<br>*max* = 0.907 | *min* = 0.949<br>*mean* = 0.983<br>*med* = 0.900<br>*max* = 0.995 | *min* = 0.953<br>*mean* = 0.977<br>*med* = 0.980<br>*max* = 0.993 | *min* = 0.959<br>*mean* = 0.987<br>*med* = 0.991<br>*max* = 0.994 |
| *Training time (minutes)* | | | |
| *min* = 61<br>*mean* = 105<br>*med* = 100<br>*max* = 156 | *min* = 66<br>*mean* = 108<br>*med* = 103<br>*max* = 166 | *min* = 78<br>*mean* = 158<br>*med* = 160<br>*max* = 229 | *min* = 47<br>*mean* = 65<br>*med* = 67<br>*max* = 79 |
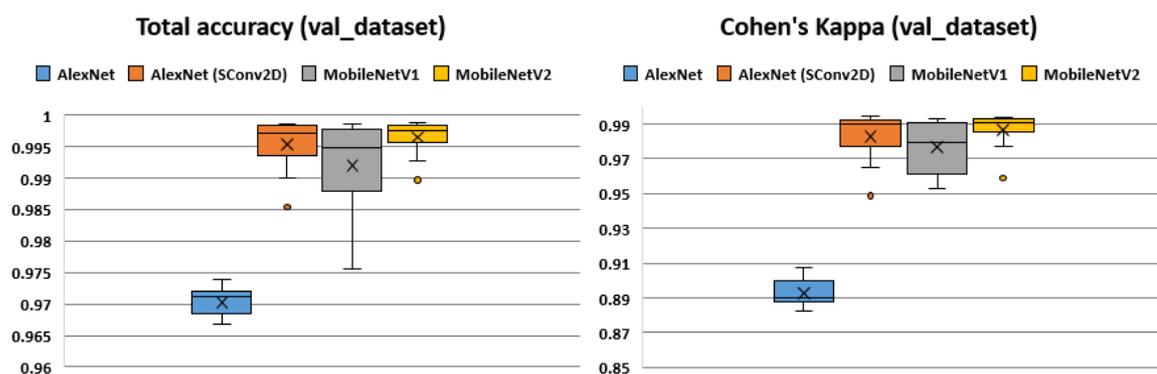
Fig. 2. **Accuracies of trained models**

Table 2

**Results of Mann-Whitney-Wilcoxon test**

|  | AlexNet | AlexNet(SConv2D) | MobileNetV1 | MobileNetV2 |
|---|---|---|---|---|
|  |  |  |  | Cohen's Kappa |
| **AlexNet** | $p = 1$ | $p < 0.01$ | $p < 0.01$ | $p < 0.01$ |
| **AlexNet (SConv2D)** | $p < 0.01$ | $p = 1$ | $p > 0.05$ | $p > 0.05$ |
| **MobileNetV1** | $p < 0.01$ | $p > 0.05$ | $p = 1$ | $p > 0.05$ |
| **MobileNetV2** | $p < 0.01$ | $p > 0.05$ | $p > 0.05$ | $p = 1$ |
|  | Total Accuracy |  |  |  |

## Conclusions

1. The MobileNetV2 architecture showed the best results considering the trade-off between the size of neural network and its accuracy. The domain-adapted model of MobileNetV2 showed the total accuracy 0.998 and the Cohen's Kappa coefficient 0.991 providing 10 times smaller neural network than AlexNet architecture. Additionally, the MobileNetV2 models provided better accuracy than AlexNet population with a confidence level 0.95.
2. There are other DL architectures, like ShuffleNet and EffNet developed for the mobile device, which can be adapted and compared for apple and pear classification in future experiments.
3. The perspective solution is to continue research with MobileNetV2 architecture and synthetic data mixed with natural images.

## Acknowledgements

## References

[1] Skrivele M. etc. Overview of fruit and berry growing in Latvia. Proc. Intl. Sci. Conf. "Sustainable Fruit-Growing: From Fruit to Product, Jurmala-Dobele, May 28-31, 2008, pp.5-14.
[2] Postman J.D. etc. Scab Resistance in *Pyrus* Germplasm. Acta Hort., 671, 2005, pp. 601- 608.
[3] Lespinasse Y. etc. Pear Breeding for Scab and Psylla Resistance. Acta Hort., 800, 2008, pp. 475-482.
[4] Bouvier L. etc. A new pear scab resistance gene*Rvp*1 from the European pear cultivar 'Navara' maps in a genomic region syntenic to an apple scab resistance gene cluster on linkage group 2. TGG, 2012,pp. 53-60.

[5]  Bus V.G.M. etc. Revision of the Nomenclature of the Differential Host-Pathogen Interactions of *Venturiainaequalis*and *Malus*. Annu. Rev. Phytopathol., 49, 2011, pp. 391-413.

[6]  Stehmann C. etc. Molecular identification of a sexual interloper: the pear pathogen, *Venturiapirina*, has sex on apple. Phytopathology, 91,2001, pp. 633-641.

[7]  Broggini G.A.L. etc. Genetic mapping of 14 avirulence genes in an EU-B04 x 1639 progeny of *Venturiainaequalis*. Fungal Genetics and Biology, 48, 2011, pp. 166-176.

[8]  Chevalier M. etc.Behaviour studies of new strains of *Venturiapirina* isolated from 'Conference' cultivar on a range of pear cultivars. Acta Hort., 800, 2008, pp. 817-824.

[9]  Precision agriculture and the future of farming in Europe Scientific Foresight Study IP/G/STOA/FWC/2013-1/Lot 7/SC5, 2016.

[10] Kang H., Chen C. Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards.Sensors, 2019, 19, pp. 4599. DOI: 10.3390/s19204599

[11] Bresilla K.,Perulli G.D.,Boini A. etc.Single-Shot Convolution Neural Networks for Real-Time Fruit Detection Within the Tree. Frontiers in Plant Science, vol.10, 2019, pp. 611. DOI: 10.3389/fpls.2019.00611

[12] Tian Y., Yang G., Wang Z.etc. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Computers and Electronics in Agriculture, vol. 157, 2019, pp. 417-426.

[13] Howard, A.G., Zhu, M., Chen etc. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017. [online] [20.03.2020].Available at:ArXiv, abs/1704.04861.

[14] Freeman I., Roese-Koerner L., Kummert A. Effnet: An Efficient Structure for Convolutional Neural Networks. Proceedings of 25th IEEE International Conference on Image Processing (ICIP), 7-10 Oct.,2018, Athens, Greece, pp. 6-10.

[15] Zhang X., Zhou X., Lin M., Sun J. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices.Proceeding of IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018, 18-23 June, 2018, Salt Lake City, USA, pp. 6848-6856.

[16] Sandler M., Howard A., Zhu M. etc. MobileNetV2: Inverted Residuals and Linear Bottlenecks. Proceeding of IEEE/CVF Conference on Computer Vision and Pattern Recognition 2018, 18-23 June, 2018, Salt Lake City, USA, pp. 4510-4520.

[17] Mureşan, H., Oltean, M. Fruit recognition from images using deep learning. Acta Universitatis Sapientiae, Informatica, 10(1), 2018, pp. 26-42. DOI: 10.2478/ausi-2018-0002

[18] Krizhevsky A., Sutskever I., Geoffrey E.H. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems, 25, 2012, pp. 1097-1105.

[19] Tallón-Ballesteros A.J., Riquelme J. Data Mining Methods Applied to a Digital Forensics Task for Supervised Machine Learning. Studies in Computational Intelligence, 2014, pp. 413-428. doi: 10.1007/978-3-319-05885-6-17.

[20] Sakib S., Ashrafi Z., Siddique M.A. Implementation of Fruits Recognition Classifier using Convolutional Neural Network Algorithm for Observation of Accuracies for Various Hidden Layers, 2019. [online] [20.03.2020]. Available at: ArXiv, abs/1904.00783.