DOI 10.52209/1609-1825_2022_4_457 UDC 004.5

Principles of Machine Learning Operation

[1]**HARON Habibollah,** *PhD, Professor, habib@utm.my,*
[2]***GOLOVACHYOVA Viktoriya,** Dr. of Ped. Sci., Professor, golovacheva_vn@mail.ru,*
[2]**TOMILOVA Nadezhda,** *Cand. of Tech. Sci., Associate Professor, tomilova_kstu@mail.ru,*
[1]*University of Technology Malaysia, Malaysia, Johor, Johor Bahru, Iskandar Puteri, Skudai,*
[2]*NPJSC «Abylkas Saginov Karaganda Technical University», Kazakhstan, Karaganda, N. Nazarbayev Avenue, 56,*
**corresponding author.*

Abstract. *The article demonstrates the principles of machine learning on the example of a specific applied scientific research. The authors focus on topical issues of using machine-learning algorithms in various fields of activity, in particular, medicine. In this paper, we are talking about one of the areas of study of machine learning – the classification of plant leaf diseases. At the same time, the purpose of the article is to solve the problem using artificial intelligence with minimal human intervention. The authors consider in detail the technology of using machine learning to solve the problem of rice plant diseases. This project work used the Convolutional Neural Network (CNN) algorithm to classify four classes of different rice diseases: leaf spot, rice blast, brown spot, and tungroz. In total, the model used 5932 images, of which 4746 images were used as training data and 1186 as test data. This made it possible to experiment with a large number of data sets and thereby demonstrate the work of the machine learning method to solve a specific scientific problem.*

Keywords: *machine learning, machine learning principles, specific problems, data, artificial intelligence, models, algorithms, images, image-processing method.*

Introduction

Over the past few years, machine learning has made a breakthrough by making some of the routine tasks of humans easier. In the future, machine learning will only develop, opening up new opportunities for humanity. However, not everyone fully understands what it is and on what principle it works, although they face the result of the process every day. The work of some applications and programs in gadgets and devices is established thanks to this technology. Vivid examples are Siri and Alice. The main task of machine learning is to teach artificial intelligence, based on the information provided to it about the world around it, to independently make decisions, self-learn and constantly improve in its self-learning. It is very important to understand how machine learning works.

Research Methodology

Machine learning (ML) is a subset of artificial intelligence (AI). ML is a class of artificial intelligence methods, the characteristic feature of which is not the direct solution of a problem, but learning by applying solutions to many similar problems. For building such methods, the tools of mathematical statistics, numerical methods, mathematical analysis, optimization methods, probability theory, graph theory, various techniques for working with data in digital form are used. [1].

The principle of machine learning allows you to create computers and programs that think like a person. However, at the same time, unlike a person, they do not get tired, they can make fewer mistakes, work with any amount of data and evaluate them impartially. This opens up a huge field for the possibilities of artificial intelligence.

The goal of machine learning is to extract information from a data array, structure it to solve certain mathematical problems using ML algorithms [2].

Machine learning refers to a set of mathematical, statistical and computational methods for developing algorithms that can solve a problem not in a direct way, but based on the search for patterns in a variety of input data. With the help of machine learning, many services and programs are created, from simple ones that a person uses in everyday life to complex tools needed in industry or security.

Machine learning is a field of computer science that differs significantly from traditional digital computing approaches. Algorithms in the understanding of traditional approaches are sets of pre-programmed instructions that are used to calculate and solve problems. Machine learning algorithms are used to perform statistical analysis on input data to derive values is the solving a problem.

In the creation of such machine learning models, labeled data plays an important role, based on which the model can be trained to solve specific problems. The ML specialist passes the data and explains what

he wants to get as an output. He knows how to build a model that answers the question. The goal of machine learning is to teach a model to find a solution on its own.

Currently, machine learning has begun to gain the most popularity, as the amount of accumulated data is growing every day, as well as computing power, which makes it possible to apply complex algorithms to solve various problems, using new methods and algorithms.

Machine learning is used in various branches of everyday life: from automating routine tasks to creating smart systems [3]. Examples of using machine learning algorithms:

- Forecasting. ML is often used to build predictive models.
- Image recognition.
- Speech recognition.
- Medical diagnoses and others.

Machine learning is used for diagnostics, forecasting, recognition and decision making in various applied fields: from medicine to banking. Almost no research in machine learning is complete without an experiment on model or real data that confirms the practical performance of a particular method. Machine learning is based on the idea that analytical systems can learn to identify patterns and make decisions with minimal human intervention.

There are four key tasks in machine learning [4]:

- *regression* – predicting the numerical values of features, for example, predicting future sales volumes based on known sales data in the past;
- *classification* – predicting which of the known classes an object belongs to, for example, predicting whether a borrower will repay a loan, based on data on how borrowers repaid loans in the past;
- *clustering* – division of a large number of objects into clusters – classes within which objects are similar to each other, for example, market segmentation, dividing all consumers into classes so that consumers inside classes are similar to each other, and in different classes they differ;
- *dimensionality reduction* – reduction of a large number of features to a smaller one (usually 2–3) for the convenience of their subsequent visualization (for example, data compression);
- *search for anomalies* – search for rare and unusual objects that differ significantly from the bulk, for example, the search for fraudulent transactions.

This research work used more datasets to improve upon the classification accuracy. Convolutional Neural Network (CNN) technique was used to design the proposed model. CNN is preferred to other techniques for its high classification accuracy and also the ability to handle huge data. CNN works well with huge datasets. As a modern approach in image classification, CNN has the capacity to produce definite diagnosis.

Deep-CNN needs huge datasets to produce better accuracy. To obtain such data, image augmentation can be used to expand the training data using different augmentation tools such as scaling, flipping, zoom and shift.

There are three main parts that make up the entire CNN architecture, namely: Convolutional layers or convol layer, pooling layers, and fully connected layers. The convol layer and the corresponding pooling layers make up the feature extraction processes while the FC layer forms the classification process. The convol layer extracts features from the images by performing mathematical operations. The pooling layer reduces the size of the output of the convol layer called feature map. The reduction is to reduce the costs of computation. Although, the dropout layer and the activation function are also regarded as layers in the CNN framework, they are also generally considered as significant parameters. The former aid the model in dealing with overfitting while the latter helps in adding non-linearity to the neural network. For example, machine-learning algorithms can be used to recognize various diseases from available test data or MRI scans.

This work has used convolutional neural network (CNN) algorithm to classify four classes of different rice diseases of leaf blight, rice blast, brownspot and tungro. The model has used 5932 images in total, with 4746 images as training data and 1186 as testing data.

The research results

Classification of plant leaf diseases is one of the interesting areas of study in machine learning. Rice is today, one of the most consumed food worldwide. Boosting rice production across the globe has therefore become imperative to cover the high demand. Farmers need to acquire proper knowledge to classify plant leaf diseases in their farmlands. The knowledge is not readily available in rural areas [5].

Therefore, most farmers use manual method to deal with plant leaf diseases. This practice has led to the decline in rice production because of the inaccuracies involved.

Deployment of technology in agriculture has reduced many errors brought about by humans. In today's world, farmers need automated systems to tackle problem of rice plant diseases. Various techniques and methods can be used to design a model that can detect and classify diseases of plant accurately. Classification of plant leaf diseases in agricultural field today has become a notable topic of research in the aspect of disease recognition [6]. Images of the affected plants can be collected from the field for use by the model in the classification process.

However, obtaining large datasets is a big challenge that is why some of the previous research on image processing and classification worked with small datasets. This research work used more datasets to improve upon the classification accuracy, obtained in. Convolutional Neural Network (CNN) technique was used to design the proposed model. CNN is preferred to other techniques for its high classification accuracy and the ability to handle huge

data.

CNN works well with huge datasets. As a modern approach in image classification, CNN has the capacity to produce definite diagnosis [4]. Image processing methods are important in recognising diseases of plant promptly before damage is done. To increase rice production greatly, robust techniques such as CNN must be deployed.

Images of four different classes of rice diseases, namely: Bacterial leaf blight, Rice blast, Brown spot and Tungro have been used to carry out this research. Augmentation activities like zoom and rotate have been performed on the datasets to create additional data needed for the training of the model. Figure 1 (a)-(d) depict the four classes of the rice diseases that have been worked upon using Tensorflow library of Python.

Bacterial leaf blight (Rice leaf blight) is a severe rice disease that is caused by a bacterium called Xoo (*X. oryzae* pathovar oryzae). It causes yellowish ooze or wilting of leaves. Crop destruction due to this disease could be up to 75% [6].

Rice blast also called neck blast is a deadly paddy fungal illness that is caused by a fungi called Magnaporthe oryzae. It causes high economic loss of about 30% of the world annual yield [7]. Blast of rice can be recognised by the appearance of lesions on leaves, pedicles and seeds.

Brown Spot. Rice brown spot disease or paddy brown-spot is one of the vital illnesses of rice plant that is caused by a fungi called D. oryzae. It attacks seedlings and grown plants. It is one of the destructive rice disease that could lead to total loss of crop yield in the farmlands. It mostly appears in regions where water is scarce [8]. Symptoms of brown spot includes oval shaped brown spots and discolouration of stems.

Tungro. Rice tungro disease is one of the disastrous paddy viral disease in the world, most especially in South and Southeast Asia. It is caused by amalgamation of two viruses, namely: Rice tungro baciliform and rice tungro Spherical viruses [9]. These viruses can cause serious economic damage. Symptoms of rice tungro includes; Stunting of leaves, dark brown specks, yellow or orange leaves.

The dataset that has been used in this research work has been collected from Mendeley dataset. Total images of 5932 have been collected. The images comprise of 4 classes of rice leaf diseases, namely: Bacterial leaf blight, rice blast, brownspot and rice tungro. The dataset has been split into 80:20 ratios for training and validation respectively. About 4746 files have been allocated for training of the model while 1186 files for validating it. The CNN technique has been used to design the model. RGB and 64X64 pixels have been used as the size and the colour of the images respectively. The research model is shown in figure 2.

The proposed model has been implemented as follows:

Setup. The experiments have been carried out on the windows 10 with 8GB RAM, 64-bit OS. Keras and Tensorflow versions 2.6.0 have been used to run the codes on the Google colab.

Image collection. The datasets comprising four classes of rice diseases, namely; Rice blast, Bacterial blight, tungro and brownspot have been used in this project. The images have been wholly acquired from Mendeley datasets repository.

Pre-processing and augmentation. The images collected from Mendeley have been sized to 64X64 pixels and augmentation approaches like rotation, zoom and rescaling were used to expand the data for training purposes.

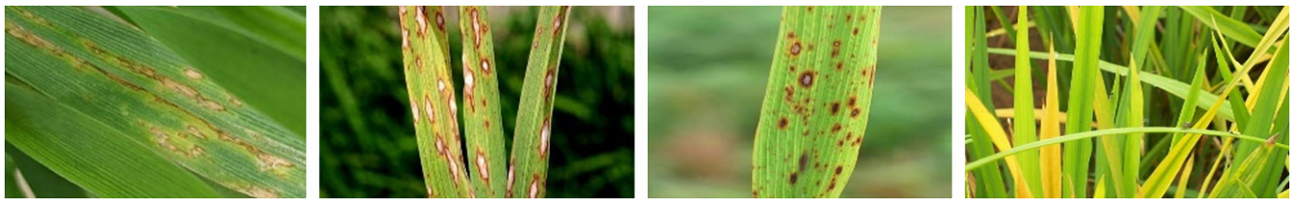
Proposed Model training. The files have been split into 80:20 ratios and then mounted for training and validation of the model. The model stops at exactly 25th epoch. This is because the error rate of the testing data was negligible at that point. To reduce losses and speed up result, Adam optimizer has been used. The developed model has strictly followed the classification process from the image collection stage up to the classification as shown in figure 1 below.

Justification. CNN model was chosen for its ability to handle huge data and also to offer excellent result.

The proposed model used 5932 files, of which 4746 were used for training and 1186 for verification. The model gave an improved classification accuracy of 99.12%. The batch size, image size, and kernel used were 16, 64x64, and 3x3 pixels, respectively.

The proposed model has classified the testing data as presented in figure 3 below.

In the creation of such machine learning models, labeled data plays an important role, on the basis of which the model can be trained to solve specific problems. By itself, artificial intelligence is not able to evaluate or predict something. In order for the



a b c d

a – Bacterial blight; b – Rice blast; c – brownspot; d – Tungro

Figure 1 – Four classes of the rice diseases that have been worked upon using Tensorflow library of Python

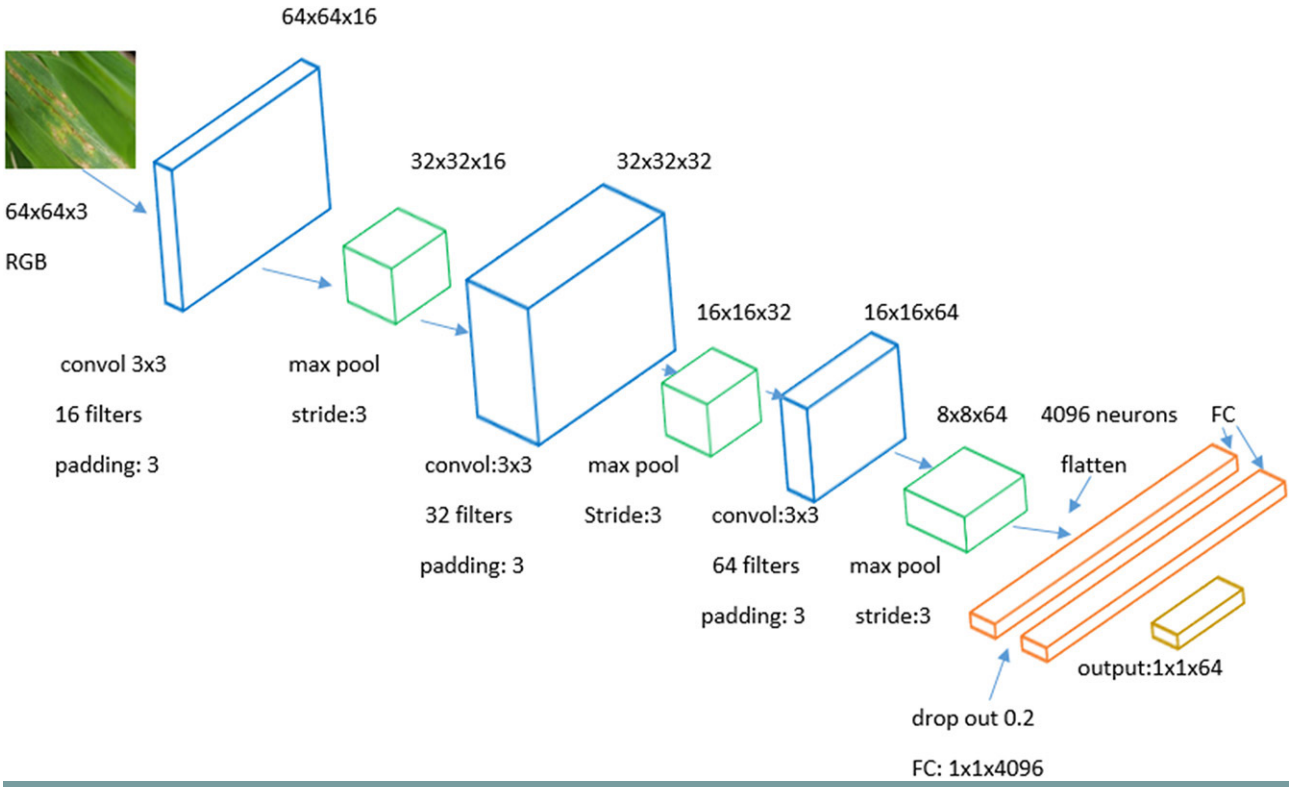


Figure 2 – Research framework

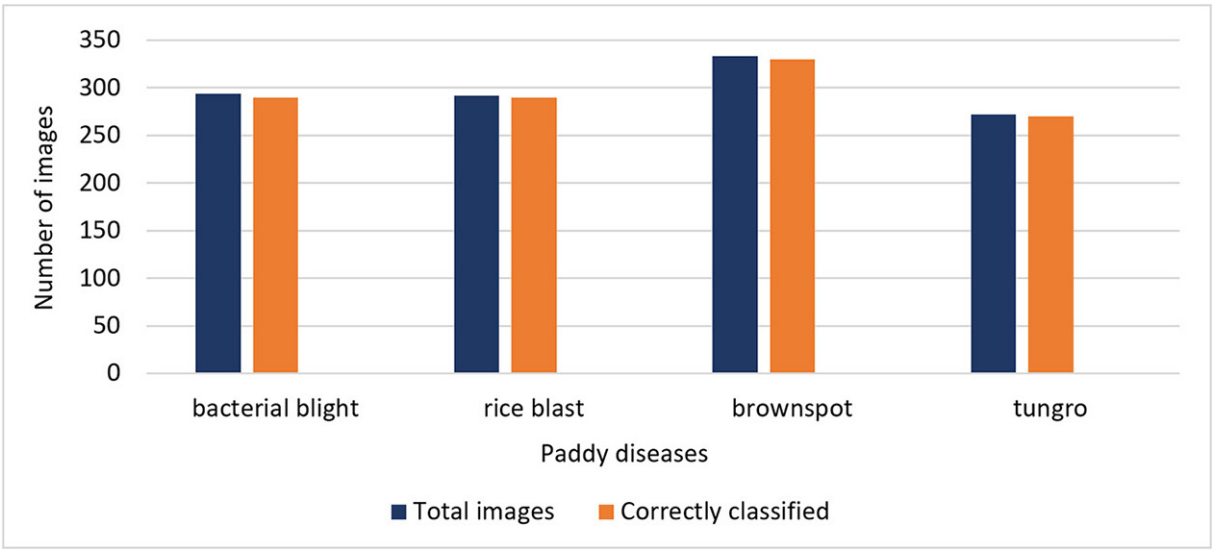


Figure 3 – Classified data chart

model to understand, it needs to be trained to work with data. Various techniques and methods can be used to design a model that can detect and classify diseases of plant accurately. Classification of plant leaf diseases in agricultural field today, has become a notable topic of research in the aspect of disease recognition. Images of the affected plants can be collected from the field for use by the model in the classification process.

Conclusion

In general, this study is aimed at an in-depth understanding of the principles of machine – learning. This work, designed to reveal the principles of operation of machine learning methods, is focused on a specific scientific study. The basics of machine learning look like this: a model is like a black box: it takes input data – the condition of the problem – and produces a certain answer, but the box has many additional parameters. They affect how the data will be processed. A machine-learning specialist selects suitable algorithms for solving a problem, trains several promising models.

REFERENCES

1. Lisa Tagliaferri, 2017. An Introduction to Machine Learning. Available via link: https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning
2. Material from Wikipedia. Alice (voice assistant). Available via link: https://ru.wikipedia.org/wiki/Алиса_(голосовой_помощник)
3. What is machine learning? Available via link: https://practicum.yandex.ru/blog/chto-takoe-mashinnoe-obuchenie/.
4. Machine learning. Introduction. Available via link: https: //infourok.ru/statya-na-temu-mashinnoe-obuchenie-vvedenie-5169527.html
5. R. Swathika, S. Srinidhi., N. Radha and K. Sowmya., «Disease Identification in paddy leaves using CNN based Deep Learning», 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1004-1008, doi: 10.1109/ICICV50876.2021.9388557.
6. S. Pratapagiri, R. Gangula, R. G, B. Srinivasulu, B. Sowjanya and L. Thirupathi, «Early Detection of Plant Leaf Disease Using Convolutional Neural Networks», 2021 3rd International Conference on Electronics Representation and Algorithm (ICERA), 2021, pp. 77-82, doi: 10.1109/ICERA53111.2021.9538659.
7. S. Norhalina & A. Muhammad & I. Rosziati & S., N. & Wan, W.H.N. (2020). An Efficient Convolutional Neural Network for Paddy Leaf Disease and Pest Classification. International Journal of Advanced Computer Science and Applications. doi: 11.10.14569/JACSA.2020.0110716.
8. Islam, Md & Shuvo, Md & Shamsojjaman, Muhammad & H. Shazid & Hossain, Md & K. Tania. (2021). An Automated Convolutional Neural Network Based Approach for Paddy Leaf Disease Detection. International Journal of Advanced Computer Science and Applications. 12. 10.14569/IJACSA.2021.0120134.
9. S.S. Hari, M. Sivakumar, P. Renuga, S. Karthikeyan and S. Suriya. «Detection of Plant Disease by Leaf Image Using Convolutional Neural Network» International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), pp.1-5,2019, Doi:10.1109/ViTECoN.2019.8899748.

Машиналық оқыту қалай жұмыс істейді

*[1]**ХАРОН Хабиболла,** PhD, профессор, habib@utm.my,*
*[2]***ГОЛОВАЧЁВА Виктория Николаевна,** п.ғ.д., профессор, golovacheva_vn@mail.ru,*
*[2]**ТОМИЛОВА Надежда Ивановна,** т.ғ.к., доцент, tomilova_kstu@mail.ru,*
[1]Малайзия технологиялық университеті, Малайзия, Джохор, Джохор-Бару, Искандар Путери, Скудай,
[2]«Әбілқас Сағынов атындағы Қарағанды техникалық университеті» КеАҚ, Қазақстан, Қарағанды, Н. Назарбаев даңғылы, 56,
**автор-корреспондент.*

***Аңдатпа.** Мақалада нақты қолданбалы ғылыми зерттеу мысалында машиналық оқыту принциптері көрсетілген. Авторлар қызметтің әртүрлі салаларында, атап айтқанда, медицинада машиналық оқыту алгоритмдерін қолданудың өзекті мәселелеріне назар аударады. Бұл жұмыста біз машиналық оқытудың зерттеу бағыттарының бірі – өсімдік жапырақ ауруларының классификациясы туралы айтып отырмыз. Сонымен қатар, мақаланың мақсаты – адамның минималды араласуымен жасанды интеллект арқылы мәселені шешу. Авторлар күріш өсімдіктерінің аурулары мәселесін шешу үшін машиналық оқытуды қолдану технологиясын егжей-тегжейлі қарастырады. Бұл жоба жұмысында әртүрлі күріш ауруларының төрт класын жіктеу үшін конволюционды нейрондық желі (CNN) алгоритмі қолданылды: жапырақ дақтары, күріш жарылуы, қоңыр дақ және тунгроз. Модельде барлығы 5932 сурет пайдаланылды, оның ішінде 4746 сурет оқу деректері және 1186 сынақ деректері ретінде пайдаланылды. Бұл көптеген деректер жиынтығымен тәжірибе жасауға және сол арқылы белгілі бір ғылыми мәселені шешу үшін машиналық оқыту әдісінің жұмысын көрсетуге мүмкіндік берді.*

***Кілт сөздер:** машиналық оқыту, машиналық оқыту принциптері, деректер, жасанды интеллект, модельдер, алгоритмдер, кескіндер, кескінді өңдеу әдісі.*

Принципы работы машинного обучения

*[1]**ХАРОН Хабиболла,** PhD, профессор, habib@utm.my,*
*[2]***ГОЛОВАЧЁВА Виктория Николаевна,** д.п.н., профессор, golovacheva_vn@mail.ru,*
*[2]**ТОМИЛОВА Надежда Ивановна,** к.т.н., доцент, tomilova_kstu@mail.ru,*
[1]Университет технологий Малайзии, Малайзия, Джохор, Джохор-Бару, Искандар Путери, Скудай,
[2]НАО «Карагандинский технический университет имени Абылкаса Сагинова», Казахстан, Караганда, пр. Н. Назарбаева, 56,
**автор-корреспондент.*

***Аннотация.** В статье демонстрируются принципы машинного обучения на примере конкретного прикладного научного исследования. В центре внимания авторов находятся актуальные вопросы использования алго-*

ритмов машинного обучения в различных областях деятельности, в частности, медицине. В данной работе речь идет об одном из направлений изучения машинного обучения – классификации болезней листьев растений. При этом целью статьи является решение задачи с помощью искусственного интеллекта с минимальным вмешательством человека. Авторы подробно рассматривают технологию применения машинного обучения для решения проблемы болезней растений риса. В этой проектной работе использовался алгоритм сверточной нейронной сети (CNN) для классификации четырех классов различных болезней риса: пятнистости листьев, пирикуляриоза риса, бурой пятнистости и тунгроза. В общей сложности модель использовала 5932 изображения, из которых 4746 использовались в качестве обучающих данных, а 1186 – в качестве тестовых данных. Это дало возможность поэкспериментировать с большим количеством наборов данных и тем самым продемонстрировать работу метода машинного обучения к решению конкретной научной задачи.

Ключевые слова: *машинное обучение, принципы машинного обучения, данные, искусственный интеллект, модели, алгоритмы, изображения, метод обработки изображений.*

REFERENCES

1. Lisa Tagliaferri, 2017. An Introduction to Machine Learning. Available via link: https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning
2. Material from Wikipedia. Alice (voice assistant). Available via link: https://ru.wikipedia.org/wiki/Алиса_(голосовой_помощник)
3. What is machine learning? Available via link: https://practicum.yandex.ru/blog/chto-takoe-mashinnoe-obuchenie/.
4. Machine learning. Introduction. Available via link: https: //infourok.ru/statya-na-temu-mashinnoe-obuchenie-vvedenie-5169527.html
5. R. Swathika, S. Srinidhi., N. Radha and K. Sowmya., «Disease Identification in paddy leaves using CNN based Deep Learning», 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1004-1008, doi: 10.1109/ICICV50876.2021.9388557.
6. S. Pratapagiri, R. Gangula, R. G, B. Srinivasulu, B. Sowjanya and L. Thirupathi, «Early Detection of Plant Leaf Disease Using Convolutional Neural Networks», 2021 3rd International Conference on Electronics Representation and Algorithm (ICERA), 2021, pp. 77-82, doi: 10.1109/ICERA53111.2021.9538659.
7. S. Norhalina & A. Muhammad & I. Rosziati & S., N. & Wan, W.H.N. (2020). An Efficient Convolutional Neural Network for Paddy Leaf Disease and Pest Classification. International Journal of Advanced Computer Science and Applications. doi: 11.10.14569/JACSA.2020.0110716.
8. Islam, Md & Shuvo, Md & Shamsojjaman, Muhammad & H. Shazid & Hossain, Md & K. Tania. (2021). An Automated Convolutional Neural Network Based Approach for Paddy Leaf Disease Detection. International Journal of Advanced Computer Science and Applications. 12. 10.14569/IJACSA.2021.0120134.
9. S.S. Hari, M. Sivakumar, P. Renuga, S. Karthikeyan and S. Suriya. «Detection of Plant Disease by Leaf Image Using Convolutional Neural Network» International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), pp.1-5,2019, Doi:10.1109/ViTECoN.2019.8899748.