IRASA International Scientific Conference
SCIENCE, EDUCATION, TECHNOLOGY AND INNOVATION
SETI VI 2024
Book of Proceedings

DOI: `https://doi.org/10.62982/seti06.duni.33

A REVIEW OF GLOBAL MODELLING WITH DEEP LEARNING OF AEROSOL OPTICAL THICKNESS FROM SATELLITE IMAGERY

Dušan Nikezić[136], *Ivana Jelić*[137], *Slavko Dimović*[138], *Milica Ćurčić*[139]

Abstract

The study provides an overview of our developed deep learning models for global spatio-temporal prediction of satellite time-series snapshots. There are several deep learning methods used for sequence-to-sequence prediction, which are presented in this paper. For global forecasting of Aerosol Optical Thickness (AOT) from satellite imagery, we have used the following models: ConvLSTM, CNN-LSTM, ConvLSTM-SA (self-attention mechanism), transfer learning technique with ResNet3D-101, and symmetric U-Net. The input data was MODAL2_E_AER_OD dataset of global AOT snapshots every 8 days from NASA Terra/MODIS (Moderate Resolution Imaging Spectroradiometer) satellite. The obtained results showed that all the developed models are capable of finding patterns in the time domain, and thus can be used for global AOT prediction.

Key words: *Deep learning, aerosol optical thickness, NASA Earth observations, spatio-temporal time-series image prediction, remote sensing satellite imagery.*

Introduction

One of the biggest uncertainties in climate modelling are aerosols with their movement through the air. In fact, interaction with solar and terrestrial radiation by aerosols disrupts the Earth's radiative budget by scattering and absorbing sunlight. Aerosol effects, like influencing cloud development or preventing clouds from forming, have effects in the water cycle (indirect effect). Since aerosol movements and dispersions are connected with weather patterns it will be useful to understand the motion of aerosols in order to predict weather patterns.

[136] Dušan Nikezić, PhD, Senior Research Associate, Vinča Institute of Nuclear Sciences, Vinča, Beograd, Republic of Serbia, dusan@vin.bg.ac.rs
[137] Ivana Jelić, PhD, Senior Research Associate, Academician of IRASA, Vinča Institute of Nuclear Sciences, Vinča, Beograd, Republic of Serbia, ivana.jelic@vin.bg.ac.rs
[138] Slavko Dimović, PhD, Full Research Professor, University of Belgrade, Vinča Institute of Nuclear Sciences, Belgrade, Republic of Serbia, sdimovic@vin.bg.ac.rs
[139] Milica Ćurčić, PhD, Research Associate, University of Belgrade, Vinča Institute of Nuclear Sciences, Belgrade, Republic of Serbia, milica.curcic@vin.bg.ac.rs

The need for atmospheric spatio-temporal forecasts has led to the development of many well-known mathematical and physical models, such as the Gaussian dispersion model, which is suitable for fast but less accurate predictions, because it does not require a lot of computational time [1]. Euler–Lagrange mathematical model is more accurate but also more demanding in terms of calculations. The Euler model is used to give spatial prediction as an explicit method for the numerical integration of ordinary differential equations. The Lagrangian method is used for forecasting a trajectory path over time with implicit time-dependence through a set of generalised coordinates [1].

The aforementioned models are based on well-known equations, but since the modelling and prediction of complex and nonlinear function approximations do not give exact solutions (ground truth), such models have discrepancies and errors. As an alternative to conventional models, Machine Learning and Artificial neural networks with Big Data offer the new possibility to capture spatial, temporal and spatio-temporal correlations more rapidly and accurately, revealing new patterns of atmospheric non-linearity [2].

In our previous studies we developed deep learning models by ConvLSTM layers, CNN-LSTM, ConvLSTM-SA, transfer learning technique with ResNet3D-101, and symmetric U-Net. Satellite remote sensing offers imagery datasets which are sorted by date, usually sequentially (time-series). The AOT dataset from NASA Earth Observations (NEO) was used in our previous studies as input data [2-4]. Satellite Terra/MODIS measures AOT by remote sensing at a wavelength of 550 nm. A thick layer of aerosols will prevent the transmission of light by scattering or absorption through the atmosphere from the ground to the satellite's sensor, while a thin layer of aerosols allows enough light through to see the ground [5]. Measurements of AOT concentration in the snapshots are based on the colour scale in a linear range, where yellow indicates AOT < 0.1 (clear sky) and brown indicates AOT = 1 (very hazy).

Deep learning models

Several related papers on the review of deep learning models for time series prediction can be found in the literature like [6, 7]. The goal or aim of this study is to provide an overview of our developed and implemented deep learning models for global spatio-temporal prediction of satellite time-series snapshots or images that are presented in our previous studies.

CNN technique (Convolutional Neural Networks) is a way to learn what is the pattern presented in data. The data can be in any form 2D, 3D or more. Most of the explanations are explained using images because it's easy to understand how the network can be applied on data. LSTMs (Long Short-Term Memory) are predominantly used to learn, process, and classify sequential data because these networks can learn long-term dependencies between time steps of data. Common LSTM applications include sentiment analysis, language modelling, speech recognition, and video analysis. Spatio-temporal image predictions in deep learning can be done using CNNs, followed by LSTM, where the CNN (Convolution2D)

captures the spatial features and LSTM captures the correlations over time. Another method involves a new class, ConvLSTM2D in Keras, which captures spatial and temporal components at the same time.

Prediction problems with times series can be divided into three main types: one-to-sequence, sequence-to-one and sequence-to-sequence. In our first study [2] four models for global spatio-temporal image forecasting were developed for the sake of comparison: ConvLSTM sequence-to-sequence, ConvLSTM sequence-to-one, CNN-LSTM sequence-to-sequence, and CNN-LSTM sequence-to-one.

The dataset, as input data for the training of aforementioned four models, covers the period from 18 February 2000 to 14 September 2021, with a temporal resolution of 8 days, providing a total of 993 snapshots in PNG format with 3600 × 1800 resolution [8]. The evaluation metric showed that a split of 80%/20% had the best result.

A grid search for hyperparameters was attempted, but better results were achieved with manual tune testing. The optimal solution for the ConvLSTM developed model was achieved with the following settings: batch size = 1, epochs = 20, activation functions 'sigmoid' and 'relu', and optimizer 'adam' with learning_rate = 0.0005. The activation function controls the non-linearity of individual neurons and when to activate them. For further model optimization, the ConvLSTM2D layer was followed by Dropout (randomly selected neurons were ignored during training) and BatchNormalization, which standardises the inputs to a layer for each mini-batch and reduces the number of training epochs. EarlyStopping was implemented in the fit() method.

By comparing the developed models, it was determined that ConvLSTM sequence-to-one had the best prediction results according to the criteria of the smallest RMSE error and the best matching of the predicted image compared to the original image using the cosine similarity method. During the testing of the prediction speed and hardware ablation, it was determined that the CNN-LSTM models have advantages but require more space than the ConvLSTM models; the ConvLSTM models occupy less than 1 GB, making them suitable for use on weaker hardware architectures. The CNN-LSTM models showed lower inference periods and lower standard deviation error dissipation with the aforementioned metrics. The inference period was in milliseconds; therefore, all of 4 models forecast almost in real-time.

Self-attention (SA) is a type of attention used in transformers where focus is to one sequence and representation of its different parts. SA in computer vision can capture long-range spatial-temporal dependencies as explained in the study [5]. Attention consists of three matrices (query Q, key K, value V) similar to a database where data are indexed by keys, and retrieved by a query. If the query, key, value is the same, then it is a self-attention. Self-attention, by backpropagation, reweights its Q, K, and V matrices learned from the data. SA updates each component of a sequence by grouping global information from the complete input sequence.

The ConvLSTM model was integrated with the SA mechanism in order to develop ConvLSTM-SA model [5]. ConvLSTM-SA model was built using Keras and since there is no SA layer in Keras we used an Attention layer and fed the same tensor twice as SA is of multiplicative kind. The optimal parameters for ConvLSTM2D layers were

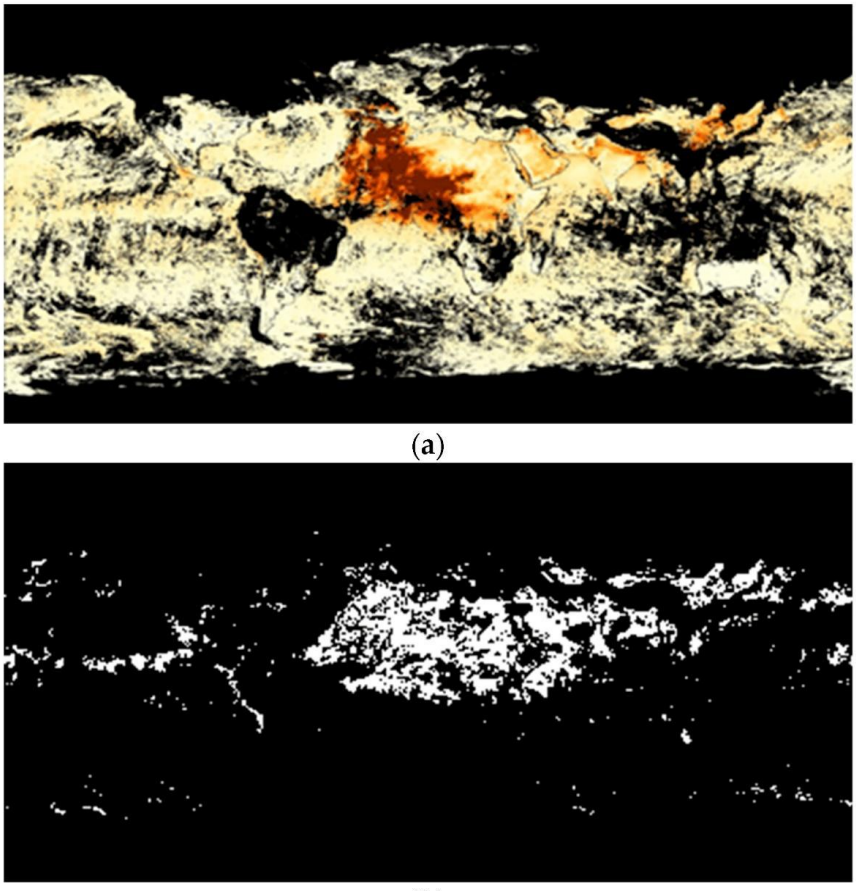
found by grid search technique. Dropout and BatchNormalization were both added after ConvLSTM2D layers to avoid overfitting, and both dropout values were calculated with learning rate by PSO algorithm. The last layer is Conv2D in order to forecast a single image.

Hyperparameter tuning was done in 3 stages. The first and second stage were based on grid search with two parameter types. The first grid search was done in order to choose the best activation function for ConvLSTM2D layers (one was for output layer activation) and the second grid search was done for recurrent_activation. During testing, we noticed that with "linear" and "ReLu" activation functions in any combination, it is impossible to fit the model because during training, all metrics got NAN values, but in combination with other functions it was possible. The obtained results show that the best combination for model is "linear" function for output activation and "hard_sigmoid" for recurrent_activation with CS = 0.9975, EUCD (Euclidean Distance) = 22.8787, and RMSE = 0.0662 for test dataset. The second grid search was done in order to choose an adequate number of filters with filter size for both ConvLSTM2D layers. The last stage in hyperparameters tuning was PSO, which as a swarm intelligence algorithm was used to find the global minimum of the function with *pyswarms* Python library.

Model ConvLSTM-SA and model without SA layer was verified with two evaluation criteria which refers to models that are able to generate image of the global AOT concentration and it could find patterns in the time domain. ConvLSTM-SA model had better performances than reference model ConvLSTM sequence-to-one from [2]. RMSE and EUCD show improvements of −13.86% and −7.01%, respectively.

In the next study [9], the new deep learning model was developed in order to better predict the high aerosol concentrations associated with air pollution and climate change using transfer learning and the segmentation process of global satellite images. Transfer learning does not change the weights in layers to be re-initialized, so layers do not change during training. Newly added layers must be trained on the input dataset in order for the developed model to make predictions. Optionally, the performance of the new model can be improved by fine tuning it at a very low learning rate. This will increase the model's performance on the new data and prevent overfitting.

Transfer learning is usually used with convolutional neural networks for classification. Low-level features, i.e., lines, are extracted from convolutional layers closer to the input, while middle layers catch complex abstract features from the lower-level features extracted. Classification is performed with layers closer to the output. The use of transfer learning for recurrent neural networks such as LSTM is not common. Therefore, in the study [9] was made a model with a prediction of the peak AOT index for early warning. In most cases, the AOT is smaller than 0.4; an AOT > 0.6 occurs only 2% of the time and an AOT > 0.4 occurs only 6% of the time. The threshold value for high AOT concentrations used in this study was chosen to be above 0.8, and through that the developed model forecasts an AOT in the range of (0.8, 1] for the eighth day.



(a)

(b)

Figure 1: (a) Original snapshot of AOT; (b) segmented image with an AOT threshold of 0.8

The residual network (ResNet) is a deep learning model intended for computer vision tasks, and it is characterised by the fact that it enables the use of a huge number of convolutional layers using the skip connections technique. Skipping some layers results in residual blocks that are partly connected to the nearest layers and partly connected to distant layers. This technique solves the vanishing and exploding gradient problem caused by the many layers connected in a deep neural network. The ResNet3D-101 transfer learning model, which was pre-trained with 2D images was used and the input dataset of AOT was MODAL2_E_AER_OD with sequences of 10 Terra MODIS images in RGB format resized to 400 × 200 pixels. The input data format was 3D since 10 images in one sequence make up the third temporal component, and the output was one 2D segmented image with the same resolution, 200 × 400. The original input dataset was converted into 2 values in images, black/white (0/1), for

binary image segmentation. The model was trained for classifying each pixel in output segmented images. Values represented by 0 are unread data and AOT values less than 0.8 (black pixels [0, 0.8]), while values represented by 1 are AOT values above 0.8 (white pixels (0.8, 1]), fig. 1.

In this study we have defined our own two evaluation criteria as follows [10]:

1. The distance domain criterion - the machine learning model is capable of generating adequate data if metrics for the predicted data, in comparison with the original data, are equal or better than an average difference between randomly selected output data from the database.
2. The time domain criterion - the ML model is capable of finding patterns in the time domain if metrics for predicted data, in comparison with the original data, are equal or better than an average difference of two adjacent output data from the database.

It can be concluded, based on the mentioned criteria, that the model can generate adequate data, and recognize patterns in the time domain.

We have done a comparison of our 3 deep learning models from studies [2, 5, 9] and the obtained results are shown in Table 1, where *ddc* is distance domain criterion, *dtc* is time domain criterion, *dm* is the model prediction metric for the test set, *ddr* is the metric relative coefficient for *ddc*, and *dtr* is the metric relative coefficient for *dtc*.

Table 1: Relative coefficients of RMSE metric

Metric RMSE	***ddc***	***dtc***	***dm***	***ddr***	***dtr***
ConvLSTM [2]	0.4613	0.4219	0.3199	0.6935	0.7582
ConvLSTM-SA [5]	0.0931	0.0663	0.0613	0.6584	0.9246
ResNet3D-101 [9]	0.2916	0.2590	0.2446	0.8388	0.9444

As can be seen from Table 1, the proposed relative coefficients of the RMSE metric have high sensitivity, which is why they are suitable for comparing different deep learning models. The model that most reliably generates data from the above is ConvLSTM-SA, which is based on ConvLSTM layers with an added self-attention layer with a *ddr* coefficient for the RMSE metric of 0.6584, while the model that best follows the changes in the time domain is the ConvLSTM model with an achieved *dtr* coefficient for the RMSE metric of 0.7582.

The advantage of the ConvLSTM model compared with the other models is reflected in the better finding of patterns in the time domain, and disadvantage is the prediction of satellite AOT images that contain some undefined pixels. The advantage of the ConvLSTM-SA model is better image reproduction than that of the other two models, and its disadvantage is the weaker detection of changes in the time domain than that of the model without the self-attention layer. The advantage of the ResNet3D-101 model compared with the other two models is that its predictions are much easier to interpret because the pixels are segmented and focus only on high-concentration aerosols, while the disadvantage is weaker characteristics according to the relative *ddr* and *dtr* coefficients.

In our most recent study [11], we implemented a fully symmetric U-Net deep learning model for forecasting a segmented image of high global aerosol concentrations. The

input dataset consisted of snapshots from 18 February 2000 to 17 November 2023 on every 8 days of global satellite images with a total of 1094 images. The images used were selected from NASA's Terra MODIS satellite with an original resolution of 3600 × 1800 pixels. This resolution provided us with clean information without the effect of averaging neighbouring pixels as a side effect of the downscaling method. Unread pixels in images are shown in black, and AOT intensity ranges are shown from white to dark brown. The useful reference range is scaled with decimal values from 0 to 1. The input data were 3D sequences of 8 Terra MODIS images in RGB format resized to 400 × 200 pixels with the third temporal component. The output was one 2D-segmented image with the same resolution of 200 × 400. The original input dataset was converted by masking to, black/white (0/1) values, Binary image segmentation model was trained for classifying each pixel. The unread data and AOT values less than 0.8 (black pixels [0, 0.8]) were masked by 0, while AOT values above 0.8 (white pixels (0.8, 1]) were masked by 1. An image of probabilities ranging from 0 to 1, as segmentation result, was encoded as bytes ranging from 0 to 255.

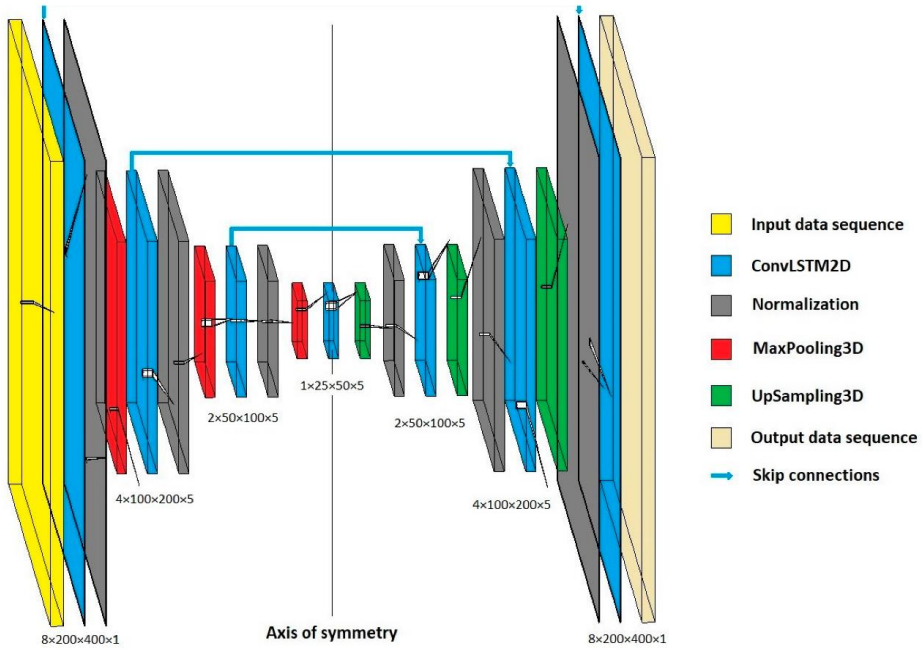


Figure 2: Implemented structure of the symmetric U-Net architecture

During the manual search of effective and useful layers, we found that Conv3D, Attention, Dropout, and AvgPooling3D layers do not have a positive impact on the model results. Also in the mentioned process, we found that the ConvLSTM2D, Normalisation, and MaxPooling3D layers have a positive impact on the model. The model development process was divided into two phases: the first was manual, and the second phase was automated metaheuristic screening. The reason for this is that as the required variables increase in metaheuristic search, hardware resources and

time increase significantly, a problem better known as the 'curse of dimensionality'. The best combination of layers in a symmetric U-Net model that we found is shown in Figure 2. The encoding path is on the left half (side), while the decoding path is on the right half (side). Every block in the encoding path of the U-Net contains a ConvLSTM2D, Normalisation, and MaxPooling3D layer. The max pooling reduces the dimensionality of each feature map. Each block in the decoding path contains symmetrically to the encoding path the up-sampling and dimension recovery, the fusion operation, as well as Normalisation and ConvLSTM2D layers. Each blue box corresponds to a multi-channel feature map.

Symmetric U-Net model parameters (Activation, Recurrent activation, Threshold, Recurrent dropout, Number of filters in hidden layers, and Learning rate) have been searched by the FOX metaheuristics. Threshold value represents the limit when classifying pixel values in model output data. The required criterion was finding the maximum value of the AUC-PR (Area Under Curve for Precision vs. Recall) metric on the test dataset.

Comparing the best-achieved result of the AUC-PR metric of 0.572 with the *ddc* and *dtc* criteria of 0.368 and 0.496, respectively, it can be concluded that the symmetric U-Net model is capable of generating adequate data and finding patterns in the time domain. The calculated relative coefficients *ddr* and *dtr* have values of 1.554 and 1.153, respectively. The previously developed ResNet3D-101 model with transfer learning [9] gave an AUC-PR metric for global aerosol prediction of 0.553, achieving values of *ddr* and *dtr* relative coefficients of 1.502 and 1.115, respectively. Based on the comparison of the values of relative coefficients, it can be can concluded that the symmetric U-Net model is capable of generating better data than the developed ResNet3D-101 model from previous study, i.e., better-achieved relative coefficient *ddr* and more capable of finding patterns in the time domain based on a better relative coefficient *dtr*.

Discussion and conclusions

This study presents our several deep learning models for global modelling AOT from satellite imagery. The stochastic nature of deep learning models introduces a certain level of uncertainty, due to the random setting of the initial values of weights and biases, as well as due to the parallelization process of training where the results of individual epochs were not obtained simultaneously. The stochastic nature of machine learning algorithms is most commonly seen on complex and nonlinear methods used for classification and regression predictive modelling problems, furthermore the obtained results may vary.

Deep learning models for classification are easier to compare because they use the accuracy metric. Deep learning models for time series prediction are not easy to evaluate precisely, and therefore are hard to make comparison. It is especially difficult to compare models for classification and prediction with each other, which is the case in this study. Although this study provides a comparison of deep learning models

with the results obtained, the presented models should be viewed as different types of global AOT modelling, i.e. image-based methods for time-series forecasting.
The presented deep learning models in this study (ConvLSTM, CNN-LSTM, ConvLSTM-SA, transfer learning technique with ResNet3D-101, and symmetric U-Net), they have shown satisfactory results of global AOT prediction since there is no exact solution, not even with complicated mathematical-physical models. Furthermore, the hyperparameters optimizations were done with some of state-of-the-art methods, as well as the evaluation metrics like our two evaluation criteria.

Acknowledgements

The organizer gratefully acknowledges the work done by Scientific Committee of the SETI V 2023 International Scientific Conference for efforts done for the success of this event.

References

[1] Nikezić P. Dušan, Zoran J. Gršić, Dramlić M. Dragan, Dramlić D. Stefan, Lončar B. Boris, Dimović D. Slavko, Modeling air concentration of fly ash in Belgrade, emitted from thermal power plants TNTA and TNTB, *Process Saf. Environ. Prot., 106*, (2017), February, pp. 274–283, https://doi.org/10.1016/j.psep.2016.06.009
[2] Nikezić P. Dušan, Ramadani R. Uzahir, Radivojević S. Dušan, Lazović M. Ivan, Mirkov S. Nikola, Deep learning model for global spatio-temporal image prediction, *Mathematics, 10*, (2022), 18: 3392,
https://doi.org/10.3390/math10183392
[3] Ramadani R. Uzahir, Nikezić P. Dušan, Radivojević S. Dušan, Mirkov S. Nikola, Lazović M. Ivan, Primena ConvLSTM modela za predikciju optičke debljine aerosola, *Proceedings*, LXVI Conference ETRAN, Novi Pazar, Serbia, 2022, https://www.etran.rs/2022/zbornik/ETRAN-22_radovi/152-VI1.2.pdf
[4] Mirkov S. Nikola, Radivojević S. Dušan, Lazović M. Ivan, Ramadani R. Uzahir, Nikezić P. Dušan, Satelitsko osmatranje i duboko učenje za predviđanje aerosola, *Vojnotehnički glasnik, 71*, (2023), 1, pp. 66-83,
http://dx.doi.org/10.5937/vojtehg71-40391
[5] Radivojević S. Dušan, Lazović M. Ivan, Mirkov S. Nikola, Ramadani R. Uzahir, Dušan P. Nikezić, A comparative evaluation of self-attention mechanism with convLSTM model for global aerosol time series forecasting, *Mathematics*, *11*, (2023), 7: 1744, https://doi.org/10.3390/math11071744
[6] Han Z., Zhao J., Leung H., Ma K. F., Wang W., A review of deep learning models for time series prediction, *IEEE Sensors Journal, 21*, (2021), 6, pp. 7833-7848, DOI: 10.1109/JSEN.2019.2923982
[7] Emmert-Streib F., Yang Z., Feng H., Tripathi S., Dehmer M., An introductory review of deep learning for prediction models with big data, *Front. Artif. Intell., 3*, (2020), 4, DOI: 10.3389/frai.2020.00004
[8] Available online:

https://neo.gsfc.nasa.gov/archive/rgb/MODAL2_E_AER_OD/
(accessed on 19 July 2024)
[9] Nikezić P. Dušan, Radivojević S. Dušan, Lazović M. Ivan, Mirkov S. Nikola, Marković J. Zoran, Transfer learning with ResNet3D-101 for global prediction of high aerosol concentrations, *Mathematics*, *12*, (2024), 6: 826,
https://doi.org/10.3390/math12060826
[10] Radivojević S. Dušan, Introducing two evaluation criteria for the next data prediction using machine learning models, *Proceedings*, DSC (Data Science Conference) Europe, Belgrade, Serbia, 2023, DOI:10.13140/RG.2.2.34745.13928
[11] Nikezić P. Dušan, Radivojević S. Dušan, Mirkov S. Nikola, Lazović M. Ivan, Miljojčić A. Tatjana, Symmetric U-Net model tuned by FOX metaheuristic algorithm for global prediction of high aerosol concentrations, *Symmetry*, *16*, (2024), 5: 525, https://doi.org/10.3390/sym16050525