Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2021.020307
ISSN 2790-2692 Vol. 2, Issue 3: 53-61

Classification Algorithm for Multiple Image Formats Suorting Complex Neural Network Models

Romany Jalali*

Bangladesh University of Engineering and Technology, Bangladesh

**corresponding author*

Keywords: Complex Neural Networks, Neural Network Models, Multiple Image Formats, Classification Algorithms

Abstract: In recent years, with the continuous exploration of NNs and the upgrading of mobile phone hardware, more and more researches focus on how to design effective models to suort terminal task reasoning. IC is a classic task in many studies. It can be used in many fields such as image search, face recognition, and medical imaging. It has great practical significance in reality. The main purpose of this paper is to carry out optimization research on the classification algorithm of multiple image formats based on the suort of complex NN models (NNM). This paper proposes improved methods for problem transformation and model adaptation, and studies the problems of blurred scenes and difficult method selection in multi-format image classification (IC). In the direction of problem transformation, a basic classification model and a classification model combined with advanced strategies are designed., in the direction of model adaptation, transform multi-format output and use transfer learning for training. Finally, it is found that the problem transformation method can obtain better classification results in small data sets, but the classification time is several times as long as the model adaptation method. The model adaptation method can save a lot of time and is easier to promote.

1. Introduction

Multi-format image refers to an image that contains multiple objects of interest. The research on multi-format image is mainly multi-format IC. The current problems of multi-format IC are mainly that the classification accuracy and classification time cannot be well balanced, and the adaptability of the model is poor [1, 2]. At present, the main directions of multi-format image learning are problem conversion and model adaptation. Problem conversion is mainly to optimize ideas and strategies. Model adaptation is to adapt to multi-format by changing the structure of the model, in order to obtain stronger generalization ability. model, this paper conducts a detailed study in multi-format IC.

In related research, Ansari et al. proposed an image steganography algorithm that can process cover images in multiple formats [3]. Steganographic algorithms that can be used in a variety of cover image formats, exploit concepts such as capacity prediction, adaptive partitioning schemes, and data propagation to embed secret data with enhanced security. The proposed method is tested for robustness against steganalysis with good results. Devi et al. alied deep learning techniques to the topic of hyperspectral image exploration [4]. Unlike traditional machine vision exercises, which focus on the spatial setting, the proposed solution will use both spatial setting and phantom relations to enhance hyperspectral image grouping.

This paper is mainly based on the suort of complex NNMs, to carry out optimization research on the classification algorithm of multiple image formats. Two improved methods of problem transformation and model adaptation are proposed, and the problems of blurred scenes and difficult method selection in multi-format IC are studied. In the direction of problem transformation, a basic classification model and a classification model combined with advanced strategies are designed. In the direction of model adaptation, multi-format output is transformed and trained using transfer learning. Finally, it is found that the problem transformation method can obtain better classification results in small data sets, but the classification time is several times as long as the model adaptation method. The model adaptation method can save a lot of time and is easier to promote. Advanced strategies can reduce part of the prediction time, but there will be a certain drop in accuracy.

2. Design Research

2.1. Analysis of Research Significance

As a classic task of computer vision, IC is a technology that uses computers to analyze and understand images, thereby identifying targets and objects of different patterns, and grouping them according to relevant rules [5, 6]. It has a large number of alications in real life. In addition, face punching, medical imaging, license plate detection, etc. have also alied image recognition technology to varying degrees. Due to the excellent performance of artificial NNs in the field of artificial intelligence, the current mainstream IC models are based on NNs. Different from the combination strategy of traditional methods that use feature descriptors and machine learning models for classification, the training method of NNs is a feature maing process from input to output. With the technological innovation of equipment, the efficiency and quality of image data collection have been greatly improved, so researchers pay more attention to image-related task reasoning on the terminal [7, 8].

The reason why NNMs have achieved remarkable achievements in various fields is at the expense of their increased complexity [9, 10]. The increase in the width and depth of the NN leads to the introduction of a large number of parameters and the number of calculations, making the NNM "bloated". The amount of parameters directly brings challenges to the memory resources of the terminal. The storage space of a large model can increase to hundreds of megabytes, which is comparable to most mobile phone alications. When the model performs inference, it will take up more memory. However, the amount of calculation puts a load on the computing resources of the mobile phone processor, and a large amount of calculation will keep the mobile phone in working state, which will have a certain impact on the battery life of the mobile phone [11, 12]. Therefore, focusing on the amount of parameters and calculation can directly affect the size of the model, thereby indirectly affecting the memory, battery life, performance and other issues of the mobile phone [13, 14].

To sum up, the basis of the NNM inference on the side is the lightweight of the model, and the lightweight of the model depends on its own small amount of parameters and calculations. Therefore, the focus of this paper is to design the model around the two core tasks of reducing the complexity of the model's convolution kernel parameters and the calculation amount of participating in the NNM for image recognition, analyzing the main problems faced by the lightweight model, and designing an effective solution to solve [15, 16].

2.2. Network Structure

It is difficult for conventional NNs to learn image features with large sizes. For example, in the MNIST dataset, the size of the image is 28x28x1, and there are 28×28×1=784 weights in each neuron in the first layer of the NN. Although this order of magnitude is relatively small, there is not only one neuron in the NN, nor more than one layer of network structure, so the amount of parameters will increase rapidly with the increase of the number of network layers and neurons, which will increase the computational burden. And easily lead to model overfitting [17, 18].

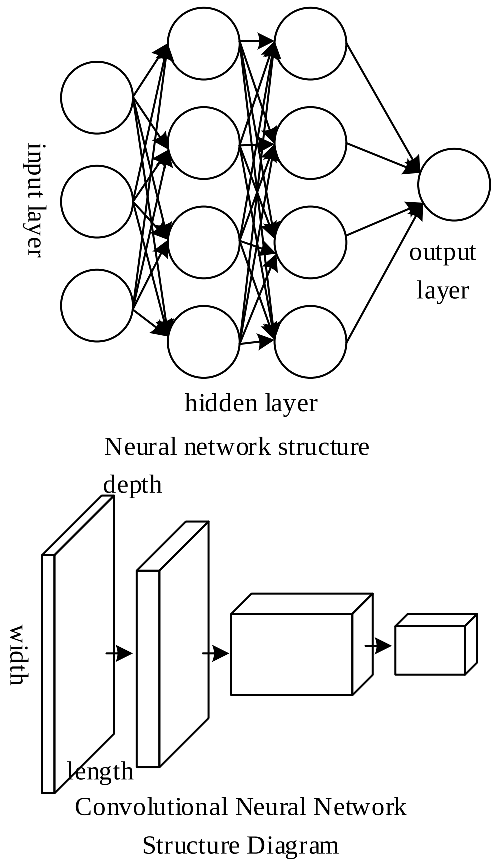

Figure 1. Network structure diagram

2.3. Model Quantization Strategy Design

The quantization in the NN training process is a quantization strategy with pseudo-quantization as the core, which includes the conversion of high-precision floating-point numbers to low-precision fixed-point numbers during quantization and the conversion of low-precision fixed-point numbers to high-precision floating-point numbers during inverse quantization. The specific strategy of model quantization is the bridge between the two fixed-point numbers and floating-point numbers. The relational transformation of the general quantization process is as follows:

$$R = (Q - Z) * T \quad (1)$$

Inverse quantization is the oosite process of the quantization process, and its relationship transformation is to make the above function identity deformation. The specific process is as follows:

$$R = (Q - Z) * T \tag{2}$$

For equations (1) and (2) R represents the real floating-point value, Q represents the quantized value, T and Z are both quantization parameters, Z represents the quantized fixed-point number corresponding to 0 floating-point number, and T represents the fixed-point quantized value. The smallest scale that can be represented. The production processes of T and Z are:

$$T = \frac{R_{\max} - R_{\min}}{Q_{\max} - QR_{\min}} \tag{3}$$

$$T = Q_{\max} - \frac{R_{\max}}{T} \tag{4}$$

For equations (3) and (4) Rmax and Rmin respectively represent the maximum floating point value and the minimum floating point value during a certain operation, and Qmax and Qmin respectively represent the maximum fixed point value and the minimum fixed point value in this operation. Through the above formal conversion, the mutual conversion of fixed-point numbers and floating-point numbers can be conveniently realized, thereby suorting the quantization process of model weights and activation values.

3. Experimental Study

3.1. Dataset Preprocessing

The process of generating digital images is often affected by various factors, such as shooting tools, external environment and so on. The main purpose of image preprocessing is to eliminate irrelevant information in the image, extract valuable real information, normalize the data to adapt to the network structure, simplify the data, etc., so as to improve the accuracy and reliability of the model. For traditional CNN, the work of image preprocessing is more complicated. The process generally includes: grayscale, geometric transformation, image denoising, region extraction and other steps. The processed image features are more prominent, and there is basically no irrelevant information, so it can be very It can well adapt to the training needs of traditional CNN, which is helpful to improve the performance of the model.

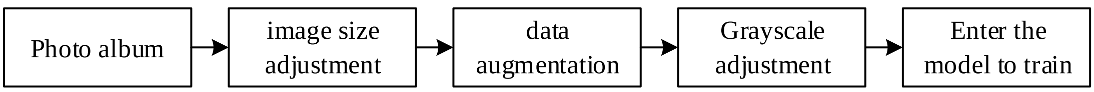

Figure 2. Image processing flow

The data set used in this paper has a single shooting background and some noise. Due to the problem of shooting conditions, the brightness of the pictures is not uniform, and the collected leaves of the same species have certain differences, the size of the leaves is not uniform, and some leaves are defective, as shown in Figure 3 -4 shown. This data set generally needs to be preprocessed when used as a training data set of traditional CNN, but the CapsNet network can establish spatial connections between features. The network requires normalization and grayscale processing.

Traditional CNN requires a large amount of data to train the model, and the number of pictures is generally tens of thousands, while CapsNet does not require a huge data set like CNN. However, the number of each 300 images is relatively small for training the CapsNet classification model, so it needs to be enhanced to a certain extent. The image enhancement method used in this paper is as follows:

(1) By translating the original image by a distance in a random direction;
(2) Rotate the image by randomly selecting an angle;
(3) Slightly blur the image;
(4) Add random noise to the image.

3.2. Experimental Verification

Based on the specific design method of the custom NN basic module and the basic principles of different model compression rate variant settings, the main purpose here is to verify which basic module variant can be more effectively used for feature extraction and classification. The data set used for verification is Public dataset CIFAR-10. Table 1 details the NN architecture stacked by the base modules ($\alpha = 0.5$).

The operation represents the current layer, which can be the convolution layer (Conv) the basic module, the global pooling (AvgPool) and the dense connection layer (Dense) using Softmax for the activation function; the input feature map represents the current layer. The input feature map of the three Dimension; output channel refers to the number of feature maps; convolution kernel size/step size represents the size and step size of the convolution kernel respectively. The basic module has been described in detail above, and it contains many types of convolution kernels, so it will not be described.

To verify the effectiveness of the model, the following training strategies were adopted. First, the training set is divided into a training set and a validation set by 4:1, in order to effectively avoid the overfitting phenomenon that the model performs well on the training set and performs poorly on the test set during the training process of the NN. For 200 epochs, the iteration of each epoch is 1250 times, and the corresponding batch size is 32; monitor the accuracy on the validation set, and perform an Early Stop operation if there is no accuracy improvement in 10 epochs.

Table 1. NN architecture parameters table

operate	Input feature map	output channel	Kernel size/step size
Input image	32×32×3	-	-
Conv1	32×32×3	32	3*3/1
base2	32×32×32	16/1	
base3	32×32×16	32/1	
base4	16×16×32	32/2	
base5	16×16×32	64	/1
base6	16×16×64	64/2	
base7	8×8×64	128	/1
Conv8	8×8×128	1000	1*1/1
AvgPool	8×8×1000	1000	-
Dense10	1×1×1000	10	-

4. Experiment Analysis

4.1. Parameter Performance Comparison

After experiments, it has been verified that the NN constructed by the basic module can effectively extract image features and perform IC. Next, for model variants with different compression rates, the design model compares the amount of parameters (in millions) and the amount of calculation (MFLOPs: million floating-point multiplication-adds, million floating-point addition operations) between them.) and the changing law of accuracy, so as to determine a better solution to reasonably design a lightweight NN (the units of parameters and calculations in subsequent chapters are the same as those in this chapter, and will not be explained again). Table 2 shows the relevant parameters of different model variants, and the rest of the experimental settings are the same for each variant except for the architecture.

Table 2. Parameters and performance comparison of model variants

Model hyperparameter α	Parameter quantity (M)	Compute volume (MFLOPs)	Storage Space (Mb)	precision
α=0.125	1.1	74	4.2	74.9%
α=0.25	1.2	97	4.6	75.2%
α=0.5	1.5	142	5.5	76.4%
α=0.75	1.7	185	6.3	75.8%

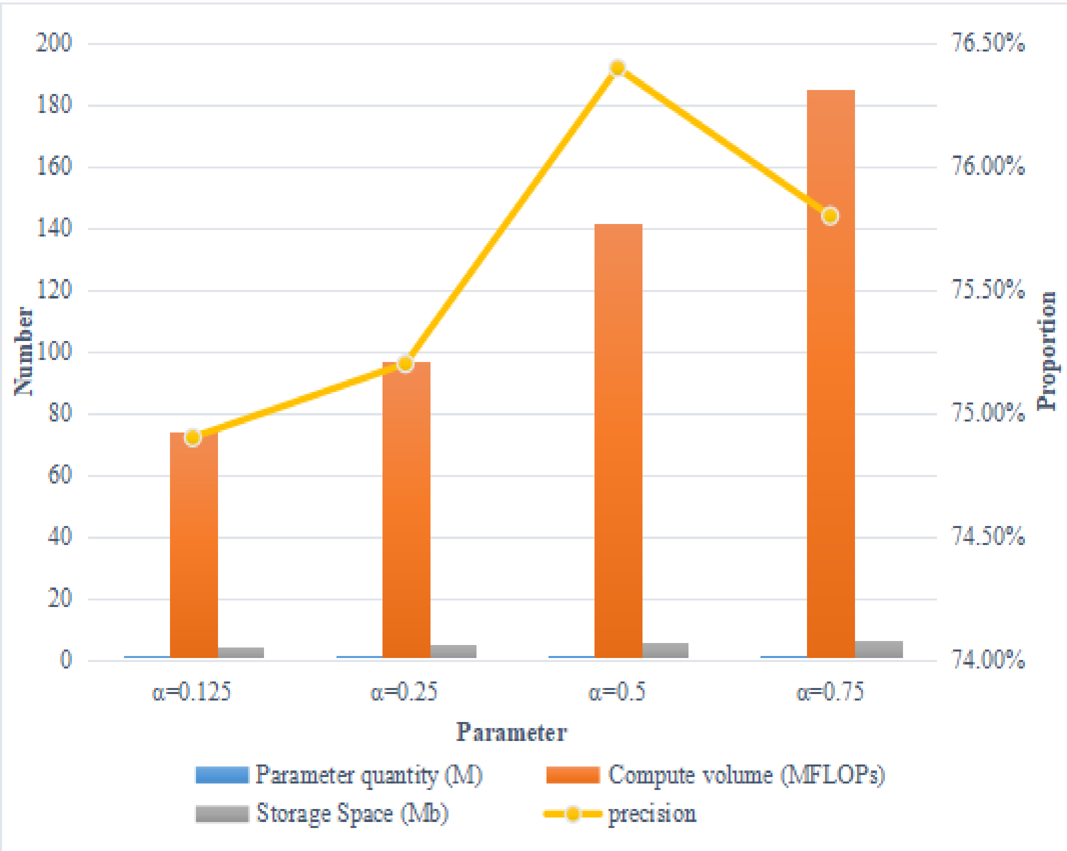

Figure 3. Parameter and performance comparison analysis plot of model variants

In Figure 3, it can be seen that the best model accuracy is achieved when the hyperparameter is 0.5. This is because as the value of continues to rise, the proportion of standard convolution in shunt feature extraction increases, which can maintain relative Fewer parameters ensure that more features are extracted, so the performance is improved. At this time, the parameter amount of the model is only 1.5M, which is nearly 100 times smaller than the parameter amount of the widely used VGG16 model of about 138M. The memory size occupied by the calculation model is 5.5Mb, which fully satisfies the reasoning of the image recognition task of mobile phones that lack memory resources. However, when it exceeds 0.5, the performance of the model decreases, which indicates that the model complexity and accuracy are not completely positively correlated.

4.2. Validation Verification

Taking the basic module of the NN (without the attention algorithm) as the baseline, to explore the influence of the channel domain attention and the spatial domain attention, design and adopt one of the model variants, and determine the performance of the model variants and the baseline. The effect of two attention algorithms. The overall comparison is shown in Table 3.

Table 3. Channel domain and spatial domain model variant performance

Model	Parameters (M)	Compute volume (MFLOPs)	precision
Model(baseline)	1.5M	142	76.4%
Model(channel)	1.7M	143	80.5%
Model(space)	1.7M	145	78.9%

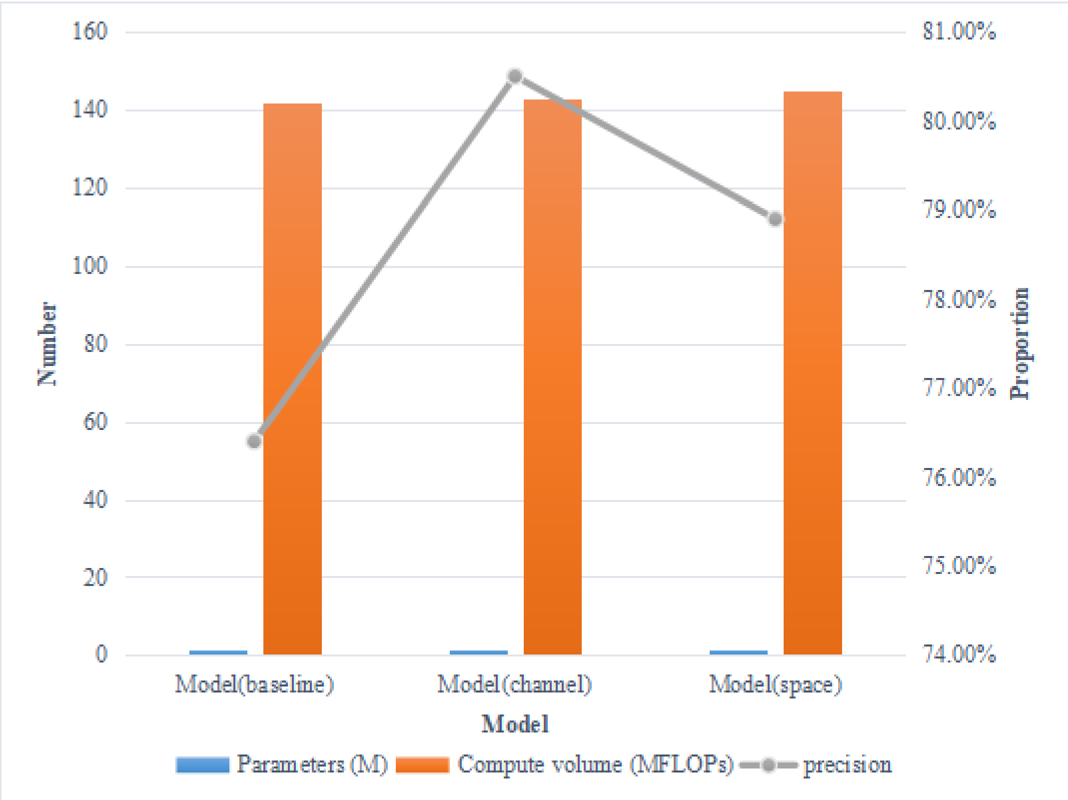

Figure 4. Performance analysis graph of channel domain and spatial domain model variants

As can be seen from Figure 4, both channel domain attention and spatial domain attention improve the performance of the baseline to varying degrees. The former only introduces about 200,000 parameters and 1,000,000 floating-point calculations to improve the accuracy by 4.1%, while the latter Introduces roughly the same amount of parameters and floating-point operations to improve accuracy by 2.5%. This shows that the design of the soft attention algorithm has great practicability in the light-weight NN construction idea. On the other hand, the performance of the channel domain attention is better than the spatial domain attention algorithm. This paper believes that the channel domain attention describes all the information of a feature map. If this feature map is extracted separately, it can be aroximated. When the feature map is rich in information, it can get a good effect. The spatial domain feature map is the feature information of the same index position of the channel dimension. Although all the spatial information can be aroximately combined into one picture, the feature information in this picture is "split and spliced" and cannot really describe a complete picture. Meaningful pictures. Therefore, the attention algorithm in the channel domain is slightly better than the feature map algorithm in the spatial domain.

After exploring the NN modules in the channel domain and the space domain respectively, the effectiveness of the two is verified, but another question is how to effectively combine the two to enhance the feature representation, lay a good foundation for high-quality model inference, and further Effectively improve the overall performance of the model.

5. Conclusion

This paper proposes a new lightweight NNM for terminal IC. The model is divided into three parts to realize image recognition on the terminal, which effectively reduces the "volume" of the model while ensuring the accuracy of recognition. Improve the feasibility of the model in practical alications. This paper finds that the existing lightweight model design is mainly divided into automatic network search, model compression, and tight network design through preliminary investigation. The automatic NN search uses a strategy to search for an effective model in a predefined space, but this process is expensive and not universal. The tight network design mainly reduces the connection between the feature map channels through special convolution to reduce the amount of model parameters and calculation, but it also leads to incomplete fusion of feature information, which reduces the accuracy to a certain extent. By analyzing the defects of existing work, this paper proposes an attention-embedded NN design method, and uses the model quantization during training as a sulementary optimization model.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Puthongkham P, Wirojsaengthong S, Sueangam A. Machine Learning and Chemometrics for Electrochemical Sensors: Moving Forward To the Future of Analytical Chemistry. The Analyst. (2021) 146(21):6351-6364. https://doi.org/10.1039/D1AN01148K

[2] Hanif A M, Beqiri S, Keane P A, et al. Alications of Interpretability in Deep Learning Models for Ophthalmology. Current Opinion in Ophthalmology. (2021) 32(5):452-458. https://doi.org/10.1097/ICU.0000000000000780

[3] Ansari A S, Mohammadi M S, Parvez M T. A Multiple-Format Steganography Algorithm for Color Images. IEEE Access. (2020) (99):1-1. https://doi.org/10.1109/ACCESS.2020.2991130

[4] Devi T, Darwin P. Hyper Spectral Fruit IC for Deep Learning Aroaches and NN Techniques. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. (2021) 30(03):357-383. https://doi.org/10.1142/S0218488522400116

[5] Suguna G, Lavanya R. Performance Assessment of EyeNet Model in Glaucoma Diagnosis. Pattern Recognition and Image Analysis. (2021) 31(2):334-344. https://doi.org/10.1134/S1054661821020164

[6] Oliveira J, Silva E G, Neto P. A Hybrid System Based on Dynamic Selection for Time Series Forecasting. IEEE Transactions on NNs and Learning Systems. (2021) (99):1-13.

[7] Himanshu Aneja Deepanshu Jindal and Ajay Tiwari Utkarsh Pandey. Comparative study of IC Algorithms. International Journal for Modern Trends in Science and Technology. (2021) 7(1):88-92. https://doi.org/10.46501/IJMTST070120

[8] Kobayashi M. Synthesis of Complex- and Hyperbolic-Valued Hopfield NNs. Neurocomputing. (2021) 423(4):80-88. https://doi.org/10.1016/j.neucom.2020.10.002

[9] Yosefvand F, Rajabi A, Shabanlou S. Predicting Non-deposition Sediment Transport in Clean Pipes: Assessing Heuristic Models. Iranian Journal of Science and Technology, Transactions of Civil Engineering. (2021) 46(1):615-629. https://doi.org/10.1007/s40996-021-00766-w

[10] Yasin Q, M Majdański, Sohail G M, et al. Fault and Fracture Network Characterization Using Seismic Data: a Study Based on NNMS Assessment. Geomechanics and Geophysics for Geo-Energy and Geo-Resources. (2021) 8(2):1-26. https://doi.org/10.1007/s40948-022-00352-y

[11] Poloziuk K, Yaremenko V. NNs and Monte-Carlo Method Usage in Multi-Agent Systems for Sudoku Problem Solving. Technology Audit and Production Reserves. (2020) 6(2(56)):38-41. https://doi.org/10.15587/2706-5448.2020.218427

[12] Jassim M S, Coskuner G, Zontul M. Comparative Performance Analysis of Suort Vector Regression and Artificial NN for Prediction of Municipal Solid Waste Generation:. Waste Management & Research. (2021) 40(2):195-204. https://doi.org/10.1177/0734242X211008526

[13] Chatterjee O, Chakrabartty S. Resonant Machine Learning Based on Complex Growth Transform Dynamical Systems. IEEE Transactions on NNs and Learning Systems. (2020) (99):1-15.

[14] Hewage P, Trovati M, Pereira E, et al. Deep Learning-Based Effective Fine-Grained Weather Forecasting Model. Pattern Analysis and Alications. (2021) 24(1):343-366. https://doi.org/10.1007/s10044-020-00898-1

[15] Huynh P H, Nguyen V H, Do T N. Improvements in the Large p, Small n Classification Issue. SN Computer Science. (2020) 1(4):1-19. https://doi.org/10.1007/s42979-020-00210-2

[16] Raziani S, Ahmadian S, Jalali S, et al. An Efficient Hybrid Model Based on Modified Whale Optimization Algorithm and Multilayer Perceptron NN for Medical Classification Problems. Journal of Bionic Engineering. (2021) 19(5):1504-1521. https://doi.org/10.1007/s42235-022-00216-x

[17] Ekici B, Kazanasmaz Z T, Turrin M, et al. Multi-zone Optimisation of High-Rise Buildings Using Artificial Intelligence for Sustainable Metropolises. Part 1: Background, Methodology, Setup, and Machine Learning Results. Solar Energy. (2021) 224(2):373-389. https://doi.org/10.1016/j.solener.2021.05.083

[18] Vadiati M, Yami Z R, Eskandari E, et al. Alication of Artificial Intelligence Models for Prediction of Groundwater Level Fluctuations: Case Study (Tehran-Karaj alluvial aquifer). Environmental Monitoring and Assessment. (2021) 194(9):1-21. https://doi.org/10.1007/s10661-022-10277-4