ISSN 2347 - 3983
Volume 8. No. 7, July 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter151872020.pdf
https://doi.org/10.30534/ijeter/2020/151872020

Image Pattern Recognition using DCT-based Feature Extraction and Genetic Search Feature Selection

Awos Kanan[1], Sreejith Munthikodu[2]
[1]Deptartment of Computer Engineering, Princess Sumaya University for Technology
Amman, Jordan, a.kanan@psut.edu.jo
[2]Department of Computer Science, University of British Columbia
Vancouver, Canada, sreejithmunthikodu@yahoo.com

ABSTRACT

Feature selection is a critical task in machine learning that is used for dimensionality reduction of data to enhance the performance of different classification and clustering algorithms. The main challenge of the feature selection problem is its large search space. Different approaches have been adopted to solve the feature selection problem. Meta-heuristic search techniques, including genetic algorithms, have recently gained attention by machine learning community and achieved enhanced performance. A genetic algorithm can generate quality solutions to optimization problems based on genetic operators such as selection, crossover, and mutation. It is an iterative process in which the best candidate solutions are combined to generate even better candidate solutions. The process is repeated until an acceptable solution is achieved or the maximum number of iterations is reached. In this paper, a genetic algorithm is used to search for the optimal subset of relevant features for image classification. Experimental results show that the adopted feature selection technique resulted in enhanced performance of neural networks and logistic regression classifiers using much less number of selected features.

Key words : Machine Learning, Image Classification, Feature Extraction, Feature Selection, Genetic Algorithms

1. INTRODUCTION

Genetic algorithms are one the most widely used techniques for feature selection in machine learning [1]. The search using genetic algorithm begins with an initial set of candidate solutions referred to as population. The quality of each candidate or individual in the population is represented by a suitable fitness function. Individuals with the best fitness values are selected to crossover and produce a new set of candidate solutions referred to as offspring. To give an exploration feature to the search technique, randomly selected offspring are subjected to random mutations. The new set of offspring produced are referred to as the second generation. The process of applying genetic operators such as selection, crossover and mutation is continued until a predefined number of generations are reached or an acceptable solution is obtained [2].

Supervised machine learning algorithms typically involve data collection, feature extraction, feature selection and classification. Feature extraction and feature selection are very critical stages in any machine learning problem. By feature extraction and feature selection, not only the machine learning becomes more accurate, but also the computational requirements are drastically reduced as less number of features are used. Since feature selection is the process of selecting the most useful features from a large number of features, the evolutionary nature of Genetic Algorithms can be a very good approach that can generate the most useful set of features. The best chromosome, on the convergence of the Genetic Algorithm, will have a set of features that can be used to train the model to achieve the best performance in terms of classification accuracy and number of features used [3].

Image classification is an important application of machine learning. The classification is primarily based on the intensity of each pixel in the image. Depending on the quality of images, the number of features can be quite large. For example, a tiny 32x32 color image has 3072 features. When a large number of training examples are required to train an image classifier, the size of training data can get huge and the training can take a lot of time and memory. Hence, dimensionality reduction using feature extraction and selection is an essential process for any image classification problem [4] [5].

In this paper, a genetic algorithm-based approach is adopted to select the optimal set of features for image classification. The dataset involves images from two different classes. Various types of genetic operators are available. The choice of which operator to use largely depends on the problem. Different genetic algorithm operators are analyzed and compared in this work, and the best operators are selected for final evaluation of the algorithm performance.

2. RELATED WORK

Several feature selection methods are used in machine learning. These methods are classified into three categories; filter methods, wrapper methods, and embedded methods [6]. Wrapper methods include sequential selection algorithms and heuristic search algorithms including GA.

The authors in [7] compared image recognition using grayscale and RGB image dataset and concluded that grayscale images resulted in better classification accuracy than color images. This observation is utilized in this work by preprocessing RGB images and convert them to grayscale images.

In [8], a DCT-based feature extraction was used for palm print recognition. The main purpose of using a 2D-DCT technique was for compressing the image and thereby generating new features with varying significance. In this work, the best subset of features will be selected from the result of 2D-DCT feature extraction technique applied to the dataset images.

3. PROBLEM FORMULATION

A genetic algorithm is used in this paper to select the most useful features from a given 3072 features of an RGB image dataset with two classes. The image dataset consists of 20,000 images for training and 2,000 images for testing. The dataset is selected from 60,000 32x32 color images, commonly known as the CIFAR-10 dataset, which was collected by the authors of [9]. Randomly selected 100 sample images from two classes (Car and Bird) are shown in Figure 1.

Figure 1: Randomly selected 100 sample images from the training dataset. (Class 1: Car, Class 2: Bird) [9]

The feature selection problem can be formulated as a search problem. The goal state of this search problem is a subset with a minimum number of features from the 3072 features of the image that maximizes classification performance. There are 3072 features and each feature has two possibilities, either to be included in the selected subset or not. Hence the search space of this problem consists of 2^{3072} states which makes brute-force search to solve this problem infeasible. As a pre-processing step, the color images are converted into gray-scale images. This results in reducing the number of features from 3072 to 1024. Hence, with the gray-scale image dataset, the search space is reduced to 2^{1024} which is still very huge number in the order of 10^{307}.

Feature selection is a classic problem suitable for binary encoding to map the search problem (phenotype) to a genetic algorithm (genotype). Every feature has only two possible values (selected, or not selected). If a gene has an allele of 1, then the feature that it maps to is selected to be used in classification. If it is 0, the corresponding feature is not selected. The initial population is randomly selected with a predetermined number of individuals. Figure 2 shows initial population of 30 individuals.. During each iteration, all individuals are replaced by the offspring.

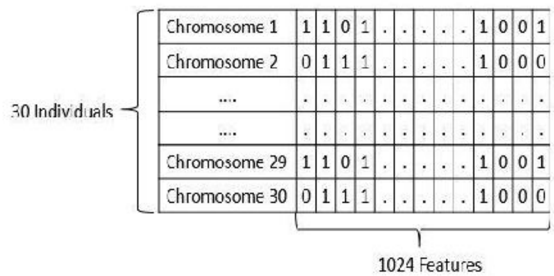

Figure 2: Binary encoding of population with 30 individuals.

The objective of feature selection is to find a subset of features that maximizes the classification accuracy while minimizing the number of features used for classification. Hence, the fitness function should be based on a combination of the number of misclassifications and the number of features used for classification. The proposed fitness function is given by:

$$f = No.\ of\ misclassifications + 0.5n \qquad (1)$$

Where *n* is the number of selected features in the individual.

The genetic algorithm objective is to minimize the fitness function. A regularization parameter of 0.5 is multiplied by the ‘number of features’ component in the fitness function to give higher priority to classification accuracy over the number of features used. The value 0.5 was experimentally calculated by running the algorithm using several other values for the regularization parameter. For finding the number of misclassifications, two classification algorithms are used; logistic regression [10] [11] and multilayer neural networks [5] [12]. The performance of the proposed genetic algorithm is evaluated using these two classifiers.

4. METHODOLOGY

The proposed genetic algorithm-based approach is designed to find the optimal minimum number of features that result in maximizing the classification accuracy. The methodology adopted in this paper is shown in Figure 3.

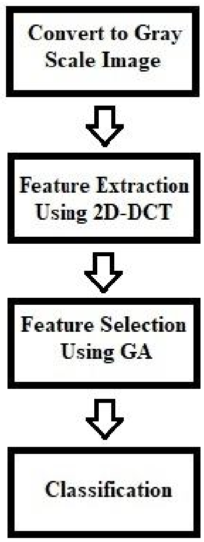

Figure 3: The proposed Methodology.

As preprocessing of the dataset, the images were converted into grayscale, as shown in Figure 4.

Figure 4: Preprocessing (Grayscale Images)

For the feature selection to work well, it is important that the features have a varying degree of significance. Rather than applying feature selection to the raw pixel data of the original images, a suitable feature extraction technique is applied to generate new features with varying significance. A feature extraction technique called 2 Dimensional-Discrete Cosine Transformation (2D-DCT) was applied onto the grayscale images which resulted in an entirely new set of features some of which are most relevant for classification and some are completely irrelevant.

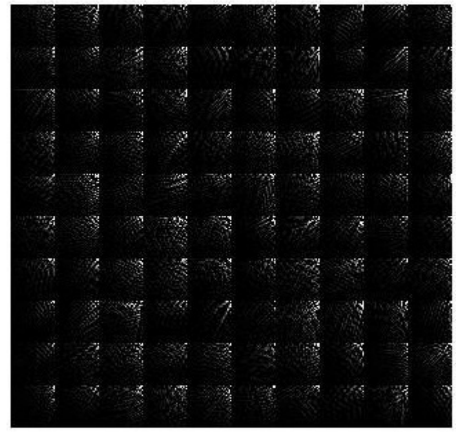

Figure 5: Feature Extraction (2D DCT Compression)

The dataset with new features after applying the 2D-DCT, as shown in Figure 5, is used as the input to the genetic algorithm for feature selection. Two different classifiers were used to evaluate the fitness function. A multilayer neural network with 1 hidden layer and 20 neurons in the hidden layer was used as the first classifier. A logistic regression classifier was used as the second classifier. The initial population of 60 individuals was randomly chosen. The search using genetic algorithm was performed using various genetic algorithm operators to compare the performance of each operator to the feature selection problem. The following subsections provide brief description of the genetic operators used in this paper.

4.1 Selection

- **Roulette Wheel Proportionate Selection:** Chromosomes/individuals are selected based on probability, which is proportional to their fitness value. A candidate with higher fitness value has a better chance of getting selected for the next step (crossover).

- **Rank Selection:** Chromosomes are ranked based on their fitness value. The probability of selection is proportional to the rank of a chromosome rather than their fitness value. Rank selection is preferred when the fitness values of the chromosomes are very close to each other. If Roulette wheel selection is applied in such as case, each chromosome will have an almost equal probability of being selected.

- **Best half Rest Random Selection:** In this selection method, the best 30 chromosomes are selected by default. Remaining 30 chromosomes are selected with a uniform probability of being selected from 60 chromosomes.

4.2 Crossover

- **Single Point Crossover:** a point is selected at random. Values from one parent are copied to fill values from the beginning of the current individual to the selected point. The remaining values are copied from the other parent.

- **Two Point Crossover:** Two points are selected randomly at random. Values from one parent are copied to fill values from the beginning of the current individual to the first point. The values from the first point to the second point are copied from the other parent. The remaining values are copied from the first parent.

- **Uniform Crossover:** Each bit has a 50% chance of inheriting values from either of the parents.

4.3 Mutation

- **Time-Independent Flip Inversion:** A chromosome is selected with a probability of 0.2 for mutation and a bit is selected at random and the value of the bit is flipped.

- **Time-Dependent Flip Inversion:** It is similar to time-independent flip inversion. However, the probability of selecting a chromosome for mutation starts at 0.3 and decreases as the generation progresses.

- **Reversing Mutation:** A chromosome is selected with a probability of 0.3 in the beginning for mutation and a bit is selected at random. All bits after the selected bit are reversed. The probability of selecting a chromosome for mutation decreases as the generation progresses.

- **Interchanging Mutation:** A chromosome is selected with a probability of 0.2 for mutation, two random bits are selected, and their values are swapped.

5. EXPERIMENTAL RESULTS

Several experiments were carried out using MATLAB to find the optimal GA parameters for the feature selection problem. Initially, the focus was on finding the proper fitness function, as discussed earlier in the paper. Then the focus was on finding the initial population size. After experimenting with a population size of 10, 20, 30 and 60 it was decided to set it to 60 as higher population size gave better results at the cost of higher computations. Different genetic algorithm operators were analyzed and compared in the following subsections, and the best operators are selected for final evaluation of the algorithm performance.

5.1 Comparison of Selection Methods

The performance of the selection methods presented in subsection 4.1 are shown in Figure 6 and Figure 7 for neural networks and logistic regression classifiers, respectively. Rank selection proved to be the best selection method for the feature selection problem as it gave the best improvement in fitness value with generations using both classifiers. As the generations progress, the fitness values of the Best Half Rest Random method gets closer to rank selection. However, Rank selection still gives better probability of selection to candidates who have higher fitness values than others. Other selection methods tend to give equal probability of selection to each individual as generation advances. Hence, rank selection always tends to pick the best individuals and performs better. Roulette wheel proportionate selection was performing poorly with both classifiers.

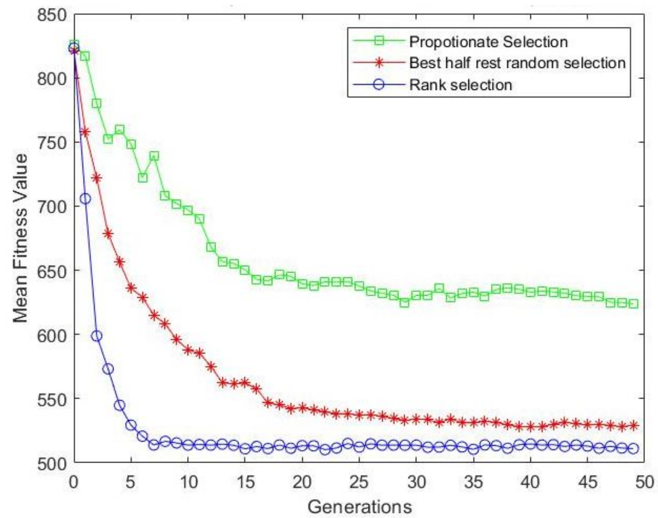

Figure 6: Performance of Different Selection Methods (NNs)

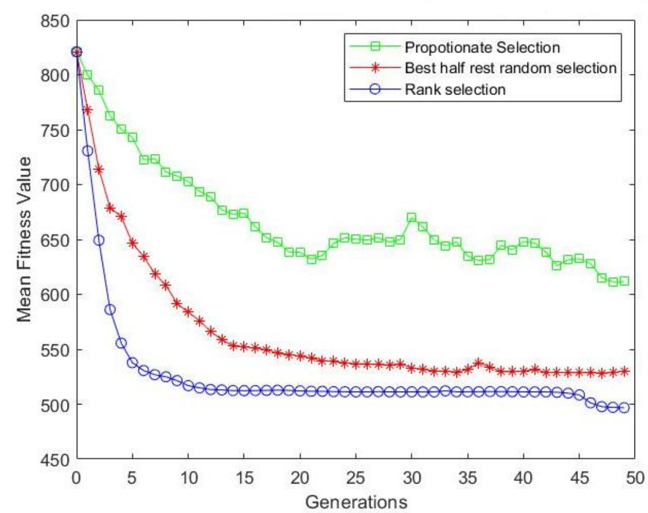

Figure 7: Performance of Different Selection Methods (LR)

5.2 Comparison of Crossover Methods

Figure 8 and Figure 9 show the performance of the three crossover methods presented in subsection 4.2. Although all three crossover methods performed well and showed steady convergence with neural networks, single point crossover is found to be the best method for the feature selection problem with logistic regression classifier. Hence, it will be used for final evaluation of the proposed approach.

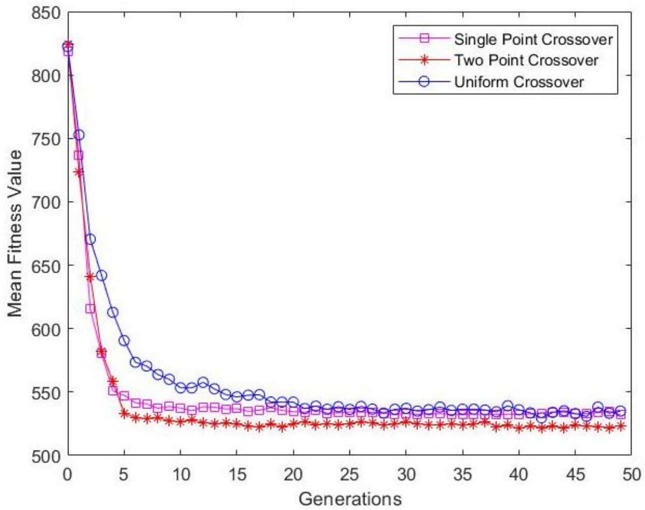

Figure 8: Performance of Different Crossover Methods (NNs)

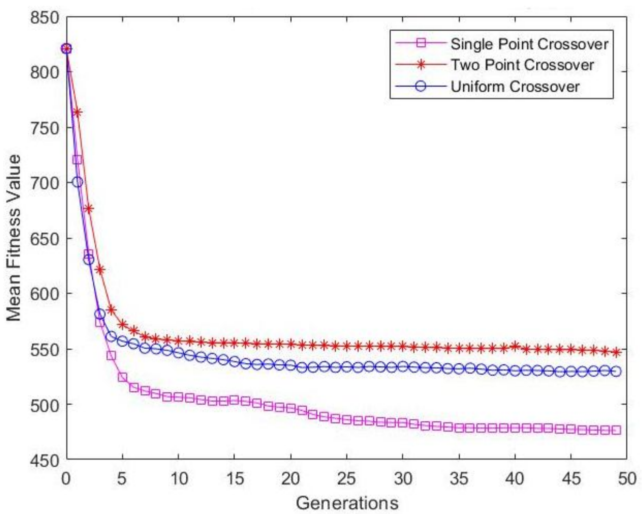

Figure 9: Performance of Different Crossover Methods (LR)

5.3 Comparison of Mutation Methods

The performance of mutation methods presented in subsection 4.3 was also compared based on the mean fitness values achieved using neural networks and logistic regression classifiers as shown in Figure 10 and Figure 11, respectively. All the mutation methods performed well except reversing mutation where all the bits after a randomly selected bit are flipped. It is a more aggressive mutation method that results in higher exploration that does not help keeping good candidate individuals for next generations. The probability of selecting a chromosome for reversing mutation is kept as a decreasing function of the number of generations. The other three mutation methods performed almost in a similar way with time independent flip inversion achieved slightly better performance and hence chosen for the final genetic algorithm evaluation.

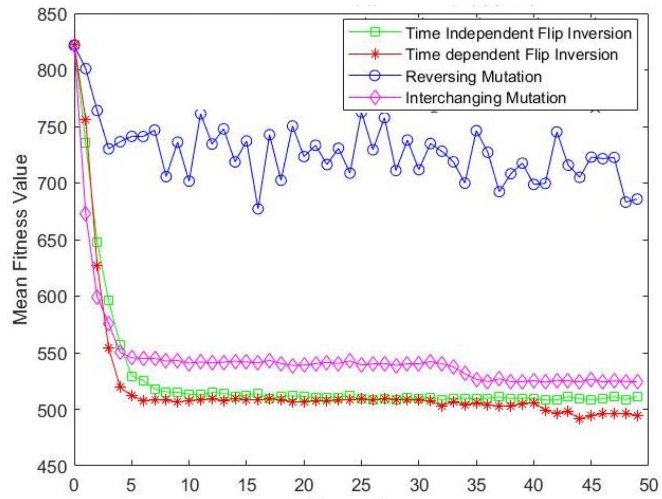

Figure 10: Performance of Different Mutation Methods (NNs)

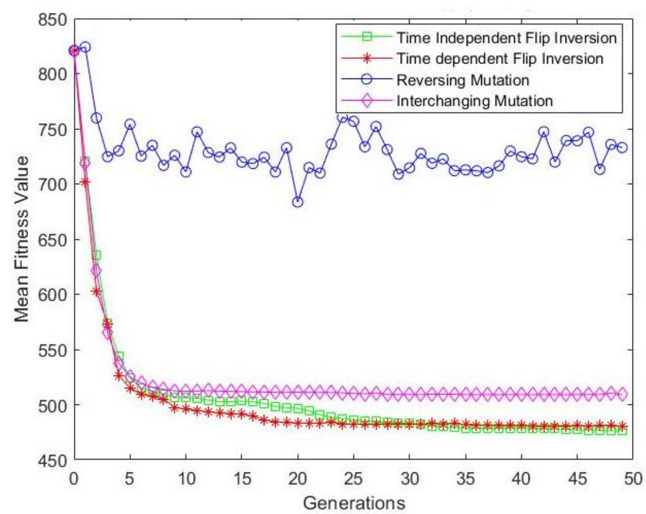

Figure 11: Performance of Different Mutation Methods (LR)

5.4 Evaluation Using Optimal Genetic Operators

The final evaluation of the proposed genetic algorithm-based feature selection for image classification was performed using GA parameters shown in Table 1.

Table 1: GA settings used for final evaluation

Parameter	Selected value\method
Population Size	60 Individuals
Selection Method	Rank Selection
Crossover Method	Single Point Crossover
Mutation Method	Time-Dependent Flip Inversion

Two set of experiments were carried out to find the optimal set of features for the image classification problem. Firstly with logistic regression classifier and then neural networks classifier. From initial experiments, logistic regression classifier resulted in lower accuracy compared to neural network classifier for the same data and selected features.

However, it performed much faster. 20 experiments were performed with different random initial population. Best candidate from each trial was saved to compare the results. A generation size of 50 was sufficient enough as in all cases, the algorithm converged within 50 iterations.

The convergence of the algorithm was found to be dependent on the initial population selected. In some cases, the algorithm converged in as low as 15 iterations and resulted in very low fitness value. On the other hand, some populations took about 50 iterations to converge. Table 2 shows accuracy results for different number of features selected by the genetic algorithm using logistic regression and neural networks classifiers.

Table 2: Classification accuracy results

No. of Selected Features	Accuracy (%)	
	LR	NN
43	76.85	78.85
40	76.9	78.75
56	77.9	84.05
66	77.9	83.9
70	78.1	81

As shown in Table 2, the best accuracy for both classifiers was achieved using only 56 selected features out of the original 1024 features. It is clear from the results that neural network classifier outperforms logistic regression in all cases.

6. CONCLUSION

A genetic algorithm was successfully applied for feature selection in an image classification task. The optimal genetic methods and parameters were selected based on a detailed analysis and comparison of different available options in MATLAB. Final evaluation of the proposed algorithm using the best genetic methods showed that neural network classifier achieved better classification accuracy compared to logistic regression using only 56 features out of the original 1024 features of the image dataset. The proposed GA-based feature selection approach can be further improved by optimizing the fitness function and using a heuristic initialization of the population..

REFERENCES

1. Li J., Cheng K., Wang S., Morstatter F., Trevino R. P., Tang J., Liu H. **Feature selection: a data perspective**, *ACM Computing Surveys*, Vol. 50, no. 6, p. 94, 2017. https://doi.org/10.1145/3136625
2. Bäck, Thomas, David B. Fogel, and Zbigniew Michalewicz, **Evolutionary computation 1: Basic algorithms and operators**, *CRC press*, 2018.
3. Mafarja M.M., Mirjalili S., **Whale optimization approaches for wrapper feature selection**, *Applied Soft Computing*, Vol. 62, pp. 441-453, 2017.
4. Cai, Jie, et al. **Feature selection in machine learning: A new perspective.** *Neurocomputing*, Vol. 300, pp. 70-79, 2018.
5. R. Margarita, A. Gamarra, G. Christian, M. Quintero, **Using Genetic Algorithm feature selection in neural classification systems for image pattern recognition,** *Ing. Inv*., Vol. 33, no. 1, pp. 52-58, 2013.
6. Girish Chandrashekar, Ferat Sahin, **A survey on feature selection methods,** *Computers and Electrical Engineering*, Vol. 40, pp. 16–28, 2014. https://doi.org/10.1016/j.compeleceng.2013.11.024
7. H. M. Bui, M. Lech, E. Cheng, K. Neville and I. S. Burnett, **Using grayscale images for object recognition with the convolutional-recursive neural network**, *IEEE Sixth International Conference on Communications and Electronics (ICCE)*, Ha Long, 2016, pp. 321-325.
8. H. Imtiaz and Shaikh Anowarul Fattah, **A DCT-based feature extraction algorithm for palm-print recognition**, *International Conference On Communication Control And Computing Technologies*, Ramanathapuram, 2010, pp. 657-660.
9. A. Krizhevsky and G. Hinton, **Learning Multiple Layers of Features from Tiny Images**, technical report, Univ. of Toronto, 2009.
10. Peng CY, Lee KL, Ingersoll GM., **An introduction to logistic regression analysis and reporting**. *The journal of educational research*, Vol. 96, no. 1, pp. 3-14, 2002. https://doi.org/10.1080/00220670209598786
11. Tran, Dung. **Hospital Bed Support System Based on Machine Learning.** *International Journal of Emerging Trends in Engineering Research*. Vol. 8, no. 5, pp. 1987-1989, 2020. https://doi.org/10.30534/ijeter/2020/84852020
12. Andriyanto, Artha. **Sectoral Stock Prediction Using Convolutional Neural Networks with Candlestick Patterns as input Images**. *International Journal of Emerging Trends in Engineering Research*. Vol. 8, no. 6, pp. 2249-2252, 2020. https://doi.org/10.30534/ijeter/2020/07862020