Performance of Multi Layer Perceptron and Deep Neural Networks in Skin Cancer Classification

1st Yessi Jusman
Department of Electrical Engineering
Universitas Muhammadiyah Yogyakarta
Bantul, Daerah Istimewa Yogyakarta
55183, Indonesia
yjusman@umy.ac.id

2nd Indah Monisa Firdiantika
Department of Electrical Engineering
Universitas Muhammadiyah Yogyakarta
Bantul, Daerah Istimewa Yogyakarta
55183, Indonesia
indah.monisa.2016@ft.umy.ac.id

3rd Dhimas Arief Dharmawan
Department of Electrical Engineering
Universitas Muhammadiyah Yogyakarta
Bantul, Daerah Istimewa Yogyakarta
55183, Indonesia
dhimasariefdharmawan@umy.ac.id

4th Kunnu Purwanto
Department of Electrical Engineering
Universitas Muhammadiyah Yogyakarta
Bantul, Daerah Istimewa Yogyakarta
55183, Indonesia
kunnu_p@umy.ac.id

2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech) | 978-1-6654-1875-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/LIFETECH52111.2021.9391876

***Abstract*— Skin cancer refers to a condition where there exists abnormal growth of skin cells, mostly occurs on skin exposed to the sun. There are several types of skin cancer, where the most common types include basal cell carcinoma, squamous cell carcinoma, and melanoma. Without proper treatment, skin cancer, particularly in the melanoma form, can lead to deaths. Fortunately, early detection and classification of skin cancer are highly effective in preventing serious damages from skin cancer. In this paper, we train Multi-layer Perceptron, a custom convolutional neural network, and VGG-16 for skin cancer classification on a large skin cancer dataset, HAM10000. The performance of each trained model is subsequently compared and analyzed in terms of classification accuracy and computational time. Our experimental setups reveal that the VGG-16 model can set the best classification accuracy among the compared networks while in terms of testing time, the VGG-16 and custom CNN models are being much faster than the Multi-layer Perceptron. The results of our study are beneficial in providing systematic comparison and analysis of several neural networks in skin cancer classification.**

Keywords—Deep neural networks, multi-layer perceptron, skin cancer classification, transfer learning.

I. INTRODUCTION

Skin cancer is one of the most commonly occurring cancers, caused by excessive sunlight exposure to human skin [1]. There are several types of skin cancer, where the most common types include basal cell carcinoma, squamous cell carcinoma, and melanoma. Like other types of cancer, untreatable skin cancer can cause deaths. Fortunately, early detection and classification of skin cancer can effectively increase the survival rate of people suffering from this disease [2]. Additionally, with the rapid development of machine learning algorithms, early detection of skin cancer can be made out of possible.

In the literature, several methods for skin cancer classification have been designed. The study by [3] reveals that color, texture, and shape features of melanoma are useful for skin cancer classification. Specifically, the authors of this study compare the classification results of some skin cancer classification methods built upon six different classifiers in combination with seven features. The results of this study show that the method with the HSV color features and Balanced Random Forest classifier produces the best performance on the HAM10000 dataset with 81.46% of AUC, 74.75% of accuracy, 90.09% of sensitivity and 72.84% of specificity.

In [4], several convolutional neural networks are used to predict and classify seven different types of skin lesions on the HAM10000 dataset. Additionally, the authors of this study develop a website for the implementation of the built models, providing a real-time prediction of the three most probable types of skin lesions for a given image.

More recently, Cevik and Zengin [5] build an algorithm based on the pre-trained VGG-16 model for skin disease classification on the HAM10000 dataset. The authors of this study use k-fold cross-validation to divide the images in the dataset into the training and testing sets. The experimental results of this study show that the designed model can achieve a classification accuracy of 85.62%, being better than some previous methods in the literature.

Although several machine learning-based methods for skin cancer detection and classification have been proposed, however, their performance has not been quantitatively compared with each other. Additionally, other important parameters in developing real-time implementations of skin cancer classification algorithms like the computational times of the designed methods have not been investigated and analyzed. Hence, it may be difficult to draw a proper justification regarding the performance of different machine learning models in skin cancer classification.

Instigated by these two facts above, we propose a study to compare and analyze the performance of some skin cancer classification algorithms. In our study, we train Multi-layer Perceptron, a custom convolutional neural network, and VGG-16 for skin cancer classification on a large skin cancer dataset, HAM10000. The performance of each trained model is subsequently compared and analyzed in terms of

978-1-6654-1875-1/21/$31.00 ©2021 IEEE

classification accuracy and computational time. Our experimental setups reveal that the VGG-16 model can set the best classification accuracy among the compared networks while in terms of testing time, the VGG-16 and custom CNN models are being much faster than the Multi-layer Perceptron.

The remaining sections of this paper are organized as follows. In Section II, we describe the detailed methodology used in our study. Section III presents the experimental results and discussions of the results. Finally, we draw a concluding remark for our study in Section IV.

II. METHODOLOGY

A. Background of Proposed Learning Model

In this study, skin cancer image classification is built using the multi-layer Perceptron and deep neural network methods. The deep neural network method uses a Convolutional Neural Network (CNN) with a model trained and transfer learning (CNN with a pre-trained model). This skin cancer detection system is designed using the python programming language and neural network library Keras. The dataset used comes from the open-source HAM10000. Tests on both methods are performed to obtain accuracy and time values. A more reliable method is seen based on the high accuracy and low execution time.

This research requires hardware to build systems in the form of computers and software to design systems in the form of IDE. The computer used in this study has the specifications in Table 1. Software needed to make a moving object detection system is found in Table 2.

TABLE I. COMPUTER SPECIFICATIONS

Processor	Intel® Core™ i3-6006U CPU
Speed	2.00GHz (4 CPUs), ~2.40GHz
System Mode	HP 240 G5 Notebook PC
RAM	8192 MB

TABLE II. SOFTWARE

Operation System	Windows 10 64 bit
Program Language	Phyton
IDE	Jupyter Notebook
Neural Network Library	Keras

B. Datasets

The research material in this final project uses images in the HAM10000 dataset ("Human Against Machine with 10000 training images") dataset. The dermatoscopy image is a good source for training the artificial neural network to diagnose skin disease pigments automatically. The HAM10000 dataset has 10015 dermatoscopic images collected for 20 years from two different places, namely The Department of Dermatology at the Medical University of Vienna, Austria, and The Skin Cancer Practice of Cliff Rosendahl in Queensland, Australia.

The HAM10000 dataset [11] consists of several pigment wounds collected from different populations. The Austrian image consists of patients heading to Europeans and specifically for early detection of melanoma in high-risk groups. This group of patients tends to nevi skin disease and someone or family who has a history of melanoma. Whereas the Australian image consists of patient wounds that have skin cancer characteristics. Australian patients mostly suffer from skin cancer caused by sunlight.

The dataset also contains images of the same lesion taken at different magnifications or angles. The lesion disease is divided into seven diagnosis categories. The seven classes of skin lesions included in the dataset are Actinic keratoses (Akiec), Basal Cell Carcinoma (Bcc), Benign Keratosis (Bkl), Dermatofibroma (Df), Melanocytic Nevi (Nevi), Vascular lesions (Vasc), and Melanoma (Mel).

The categories are divided based on malignant and benign lesions, but also to make specific diagnoses because different malignant lesions, for instance, melanoma and basal cell carcinoma, may be treated in a different way and timeframe. Except for vascular lesions, which are pigmented by hemoglobin and not by melanin.

Bkl, Df, and Nv come under benign lesion while Mel, Akiec, and Bcc is belonging to malignant lesion. This research do a binary classification of malignant and benign lesions. Besides Vascular lesion is included but Dermatofibroma from benign lesion is excluded. This research classify five different classification purpose namely, Mel with Nv, Akiec with Bkl, Bcc with Vasc, Mel with Akiec, and Bkl with Bcc.

C. Multi-Layer Perceptron Architecture

In the Multi-Layer Perceptron model before going through the hidden layer, the input must go through the input layer process. The input layer in MLP is to make the image size 1D or array in the form of 1D. The architecture in this model only has 5 hidden layer layers. The architecture of the perceptron multi-layer model can be seen in Fig. 1.

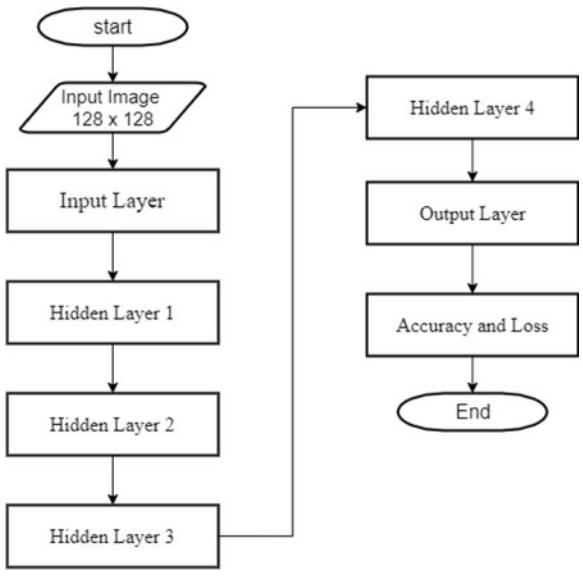

Fig. 1 Multi-Layer Perceptron

D. Convolutional Neural Network

1. Model Trained

The CNN model in this study consists of 4 layers Convolutional Layer, 4 Max pooling layers, and 2 dense layers at the end of the architecture. The architecture of the convolutional neural network model can be seen in Fig. 2.

978-1-6654-1875-1/21/$31.00 ©2021 IEEE

2. Pre-trained Model (Transfer Learning)

In transfer learning using a pre-trained VGG-16 model. In the transfer learning process, it is the same as the Convolutional Neural Network because the architecture also consists of feature extraction and classification. In this study using the VGG-16 model that has been trained on ImageNet. The transfer learning method uses the VGG16 model which has 16 layers. The input shape of VGG16 is 224 x 224 because VGG16 is trained using ImageNet data which has an image of 224 x 224. However, this study was changed to 128 x 128 due to the very long training time for processing 224 x 224 images. Flowchart transfer learning can be seen in Fig. 3.

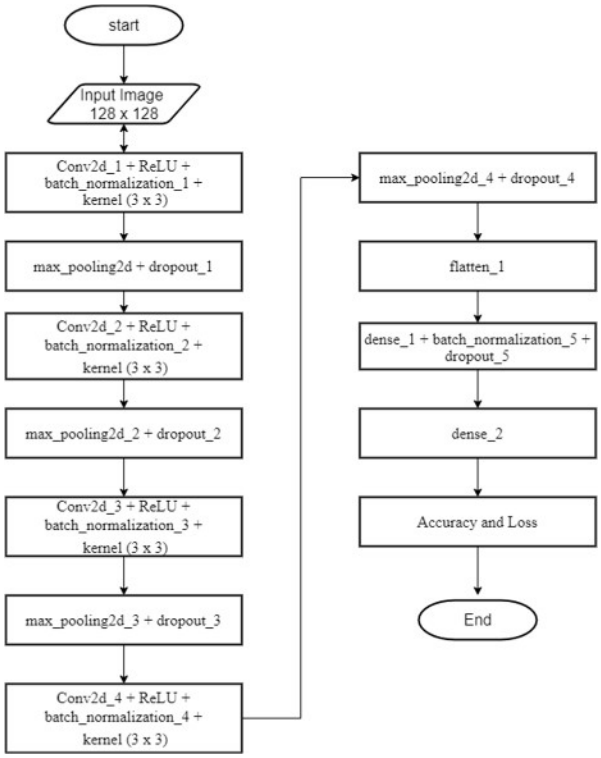

Fig. 2 CNN with Model Trained

III. Experimental Results and Discussions

The performance of deep neural networks and multi-layer Perceptron in the training and testing phases are evaluated in terms of the classification accuracy. This metric indicates the correctness of predictions made by the neural networks. Additionally, we also report and analyse the time required for the training and testing of the networks. The best network is subsequently defined as it produces the highest classification accuracy and, at the same time, does not render a high processing time.

A. Classification accuracy of the networks

The classification accuracies of each network on the defined classes are presented in Table 3. The best classification accuracy for each class is listed in bold. As can be observed in Table 3, both the custom CNN and pre-trained models can learn the representative features for skin cancer better compared to the Multi-Layer Perceptron. Based on Table 3, the pre-trained model can achieve good validation accuracy in the classification of all classes of the imagery. The higher value is in the classification of Basal Cell Carcinoma and Vascular classes. Ultimately, it is worth stating that the CNN models can outperform the MLP by a noticeable margin. This fact is better illustrated by the graphs depicted in Figures 4-6.

B. Time Comparison Results

In deep learning, computational times are also important aspects. Therefore, we also present the training and testing times for each model in Table 4 and Figures 7 to 8. As can be observed in the table, the CNN models require much longer times than the MLP to complete the training procedures. The higher computational complexity of CNN models is mainly due to the number of layer used by the models. The larger number of the layer generally results in the higher computational complexity.

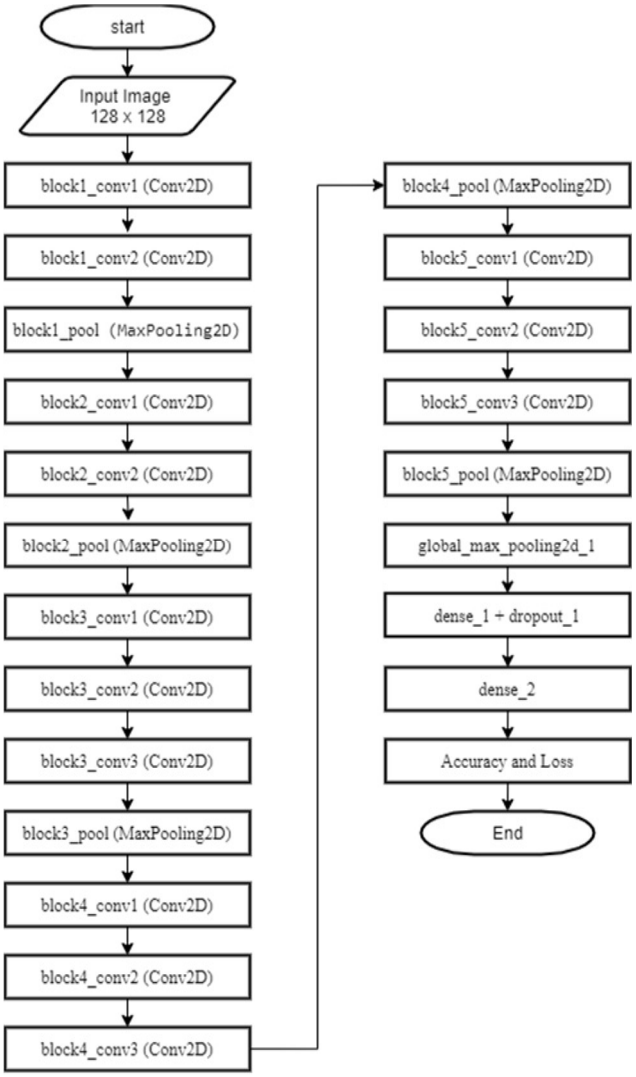

Fig. 3 CNN with Pre-trained Model

It is interesting to note that much longer training times required by the CNN models do not always make their real-time implementations impractical. Our argument is based on the fact that in practical applications, the CNN models do not longer require any training phase. In other words, the CNN models can be trained offline, such that the long training time may no longer be an important issue. Rather, the testing time is the key parameter that should be considered when one builds a CAD system. From Table 4, we may observe that CNN models are much faster than the MLP in making

978-1-6654-1875-1/21/$31.00 ©2021 IEEE

predictions from the given testing images. Hence, they would be more useful than the MLP in practical applications.

From the produced classification accuracies and testing times, we may suggest the use of CNN models, particularly the pre-trained VGG-16 model for skin cancer classification. Our suggestion is supported by the fact that the VGG-16 model can achieve the best accuracy without rendering a long testing time. Finally, we may highlight that CNN models are useful for skin cancer classification, and skin cancer classification using other CNN models should be aimed for future works.

TABLE III. CLASSIFICATION ACCURACIES OF THE MULTI-LAYER PERCEPTRON AND DEEP NEURAL NETWORKS

No.	Classification	Method	Training Accuracy	Validation Accuracy	Testing Accuracy
1.	Melanoma and Melanocytic Nevi	MLP	84,88%	84,69%	71,9%
		CNN	77,7%	73,7%	74,82%
		TL	**87,58%**	**86,39%**	**84,89%**
2.	Actinic Keratosis and Benign Keratosis	MLP	77,34%	76,85%	74,25%
		CNN	**78,15%**	77,88%	**87,12%**
		TL	78,05%	**77,94%**	77,22%
3.	Basal Cell Carcinoma and Vascular	MLP	79,31%	75,66%	73,17%
		CNN	83,55%	79,54%	95,12%
		TL	**88,67%**	**89,91%**	**97,56%**
4.	Melanoma and Actinic Keratosis	MLP	77,71%	79,67%	60%
		CNN	**82,54%**	73,9%	**80%**
		TL	82,47%	**86,50%**	78,46%
5.	Benign Keratosis and Basal Cell Carcinoma	MLP	67,71%	68,18%	71,42%
		CNN	75,81%	70,32%	**86,67%**
		TL	**76,5%**	**77,8%**	80%

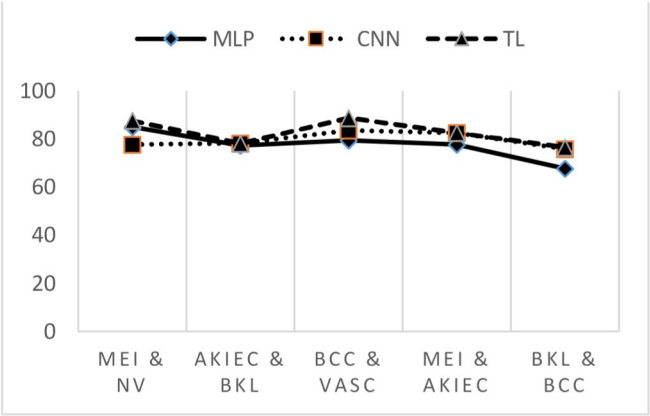

Figure 4. Training Accuracy

TABLE IV. TRAINING AND TESTING TIMES OF MLP AND DEEP NEURAL NETWORKS

No.	Classifications	Method	Training Time (s)	Testing Time(s)
1.	Melanoma and Melanocytic Nevi	CNN	9312,83	35,77
		MLP	7628,3	199,76
		TL	26318,78	39,51
2.	Actinic Keratosis and Benign Keratosis	CNN	2901,54	27,13
		MLP	1618,92	152,30
		TL	7118,39	146,58
3.	Basal Cell Carcinoma and Vascular	CNN	2530	37,29
		MLP	699,19	79,23
		TL	2571	31,82
4.	Melanoma and Actinic Keratosis	CNN	4969,97	57,36
		MLP	1747,59	88,23
		TL	7685,92	62,84
5.	Benign Keratosis and Basal Cell Carcinoma	CNN	4198,3	52,42
		MLP	2114,89	149,33
		TL	5664	31,7

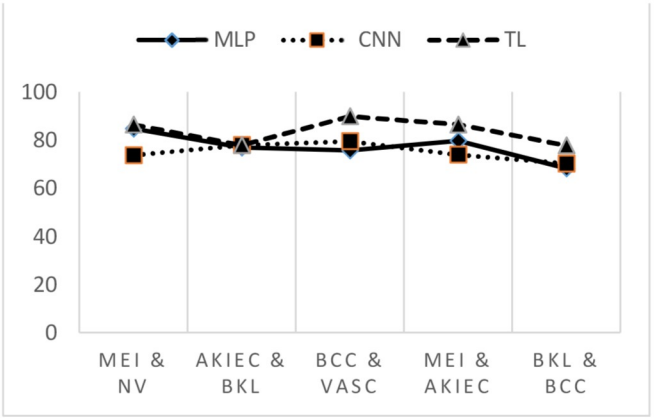

Figure 5. Validation Accuracy

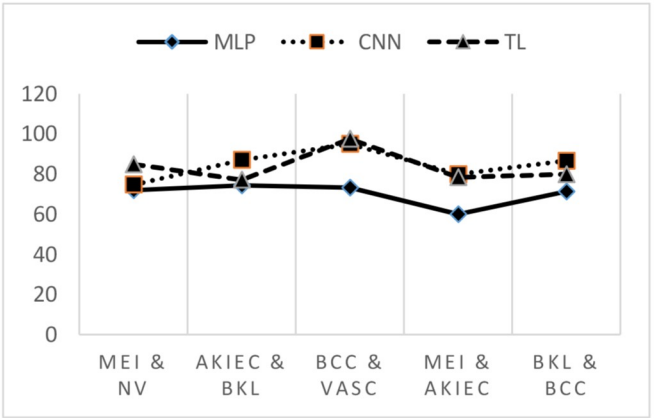

Figure 6. Testing Accuracy

978-1-6654-1875-1/21/$31.00 ©2021 IEEE

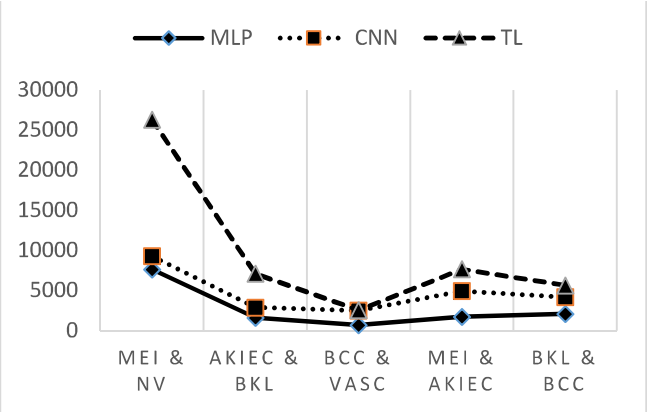

Figure 7. Training Time

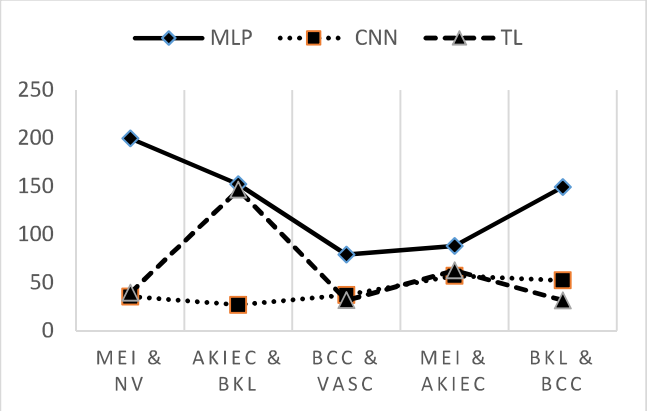

Figure 8. Testing Time

IV. CONCLUSIONS

In this paper, we have trained Multi-layer Perceptron, a custom convolutional neural network, and VGG-16 for skin cancer classification on a large skin cancer dataset, HAM10000. The performance of each trained model has subsequently been compared and analyzed in terms of classification accuracy and computational time. Our experimental setups have revealed that the VGG-16 model can set the best classification accuracy among the compared networks while in terms of testing time, the VGG-16 and custom CNN models are being much faster than the Multi-layer Perceptron.

ACKNOWLEDGMENT

The authors thank to Universitas Muhammadiyah Yogyakarta, Indonesia, for supporting this research.

REFERENCES

[1] R. Segal, K. Miller and A. Jemal, "Cancer statistics 2018", CA Cancer. J Clin., vol. 68, pp. 7-30, 2018.

[2] A. C. Geller, S. M. Swetter, K. Brooks, M.-F. Demierre and A. L. Yaroch, "Screening early detection and trends for melanoma: current status (2000–2006) and future directions", Journal of the American Academy of Dermatology, vol. 57, no. 4, pp. 555-572, 2007.

[3] T. C. Pham, G. S. Tran, T. P. Nghiem, A. Doucet, C. M. Luong and V. Hoang, "A Comparative Study for Classification of Skin Cancer," 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 2019, pp. 267-272, doi: 10.1109/ICSSE.2019.8823124.

[4] K. Pai and A. Giridharan, "Convolutional Neural Networks for classifying skin lesions," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), Kochi, India, 2019, pp. 1794-1796, doi: 10.1109/TENCON.2019.8929461.

[5] E. Çevİk and K. ZengİN, "Classification of Skin Lesions in Dermatoscopic Images with Deep Convolution Network Derin Konvolüsyon Ağı ile Dermatoskopik Görüntülerde Deri Lezyonlarının Sınıflandırılması," Eur. J. Sci. Technol., pp. 309–318, 2019, doi: 10.31590/ejosat.638247.

978-1-6654-1875-1/21/$31.00 ©2021 IEEE