

AUTOMATED KIDNEY STONE DETECTION FROM CORONAL CT IMAGES USING DEEP VGG-19 MODEL

Abhisek Gour¹ and Vishal Suthar²

¹Department of Computer Science & Engineering, MBM University, India

²Government Polytechnic College, Mandore, India

Abstract

Kidney stone diagnosis is one of the sensitive issues in personal healthcare. Detecting kidney stones early can play a vital role in avoiding chronic kidney diseases and related surgical procedures. However, due to several associated issues, identifying a kidney stone in the early stages can be very difficult. In this research, a classification model for automated diagnosis of kidney stones utilizing coronal computed tomography (CT) images is suggested. Due to low resolution and the presence of noise, every image is passed through an image enhancement step before feeding into a VGG-19 based CNN Model. The training dataset used contains 1799 cross-sectional CT scan images from 433 individuals. Data augmentation is carried out to avoid overfitting of the deep model. The developed model can correctly identify kidney stones of even tiny size with a 97.62% precision, 98.79% recall, and 98.62% accuracy. The developed model performs better than recent similar work and is suitable for e-healthcare systems. It demonstrates that such deep-learning-based techniques can be utilized to solve other similar issues in urology.

Keywords:

Kidney Stone Detection, Computational Tomography Images, Convolutional Neural Networks, VGG19 Model

1. INTRODUCTION

Since the introduction of deep neural network algorithms, researchers have proposed several pattern recognition algorithms for accurately identifying the presence of various illnesses from imaging diagnostics. In the absence of such computer-aided diagnostic (CAD) models, the diagnosis process is majorly dependent on the availability, expertise, and cognitive ability of a radiologist. However, most visual diagnosis methods including Mammography, Computational Tomography (CT), Magnetic Resonance Imaging (MRI), X-ray, and Ultrasound (US) suffer from the presence of noise, lower resolution, process anomalies, etc.

Before the advent of deep neural network algorithms and high-performance computing infrastructure, most researchers used traditional pattern recognition algorithms and image enhancement techniques to mitigate these issues and facilitate diagnosis for radiologists. However, through the deployment of modern deep-learning systems, the classification and detection of such abnormalities can be carried out with high levels of precision without any human intervention. Furthermore, integration of such systems with ICT-based telemedicine platforms can facilitate easy delivery of clinical advice and treatment to patients in remote areas [1].

Among various kidney-related diseases, kidney stones represent a fairly frequent condition, impacting around one out of every 10 persons at a certain time in their lives. Over the last few decades, the prevalence of kidney stones seems to have grown.

Researchers have reported that this rise can be attributed, primarily, to unhealthy food, diet modifications, stressful lifestyle, and growing levels of obesity. Beyond medicinal therapies and urologic procedures, diagnosis, care plans, and follow-up procedures are crucial for patients suffering from renal stone illness. Imaging-based diagnostics play a vital role in caring for such patients [2]. Kidney stone illness (nephrolithiasis) is a frequent form of urology illness with a large incidence rate of 10% after one year, 50% after 5-10 years, and 75% after 20 years. Because of the lack of particular symptoms in the initial stages of this illness, it is hard to detect the disease without proper diagnosis. Consequently, it is discovered only after the first evidence of organ failure in most patients.

In most cases, kidney stones may lead to permanent kidney damage which can further cause persistent chronic kidney diseases. Therefore, it becomes critical to detect kidney stones early for avoiding such irreparable damage. Further, such detection is critical, not only for the treatment of renal illness but also for the management of recurring stone development.

Usually, kidney stone disease is diagnosed using conventional tests (blood tests, urine tests, and biopsies) combined with imaging tests (ultrasound, CT and MRI). Imaging tests utilizing computer tomography (CT) have become the most frequent among diagnostic tests, primarily, due to factors like time required, cost, and information received from diagnostic testing. Recently, such diagnostic images (MRI, CT and X-ray) are being used to create robust and accurate deep learning models to help in the diagnosis of critical illnesses like Covid-19, cardiac arrhythmia, prostate cancer, brain tumor, skin cancer, breast cancer, etc.

In the area of urology, several deep learning techniques have been proposed to identify ureteral stones and kidney stones. Yildirim et al. [3] proposed a deep learning approach, for the identification of kidney stones through coronal computed tomography (CT) images. They used a dataset containing CT scan images of 433 individuals and reported an accuracy of 96.82 percent in binary classification (kidney stone present/absent) of input CT images.

Sudharson and Kokil [4] suggested utilization of deep transfer learning to classify B-mode renal ultrasound pictures using an ensemble of Deep Neural Networks. They proposed a two-layer approach. The first layer had 3 pre-trained deep neural network models (ResNet 101, ShuffleNet, and MobileNet v2) for feature extraction and the second layer was a Support Vector Machine trained for classification. In ultrasound images having relatively higher noise, they reported an accuracy of 95.58% while for a relatively better set of images they reported a 96.54% accuracy. It was proposed, that when using an ensemble of pre-trained networks and a majority voting technique, the system performance is better than any of the 3 individual models.

Wu and Yi [5] suggested a three-layer structure for detecting kidney problems using abdominal ultrasound images. In the proposed model, the first layer was designed to filter good ultrasound images, the second layer located kidney area in input images, and the third layer, a multi-feature fusion network (MF-Net) carried out the classification. Researchers reported an average true-positive fraction (TPF) of 98.0% and an average precision of 94.6%. Similarly, Ma et al. [6] and Akshaya et al. [7] also proposed back propagation neural network (BPNN) based ensemble systems for detecting kidney stone illness. Sudharson and Kokil [8] also proposed a deep residual learning network architecture to reduce speckle noises in ultrasound images before classifying them with the help of an SVM classifier.

Apart from form classification, several models have also been proposed to perform segmentation on kidney images. Thein et al. [9][10] proposed a model to remove unwanted areas in kidney CT images using 3 thresholding techniques – soft-organ removal, bony skeleton removal, and bed-mat removal. Akkasaligar et al. proposed a level set segmentation method to detect kidney stones. Akshaya et al. [11] proposed a Fuzzy C-Mean based Clustering technique to carry out segmentation of kidney stones in input images. However, it can be observed that the performance of segmentation methods is usually not as reliable as classification methods.

2. PROPOSED WORK

In this research, a VGG-19 based deep convolutional neural network is proposed to classify input Coronal Computational Tomography (CT) images gathered from more than 400 subjects into one of the two categories namely, Kidney Stone or Normal. The training dataset contains 1799 coronal CT images collected from Yildirim et al. [3]. The dataset was initially divided into two parts – each part representing one of the two output classes of the classification problem. Further, since the VGG-19 is a deep CNN Model, training with such a small number of training images has a higher chance of network overfitting. It has been noted in literature that using preprocessing methods, the classification accuracy can be improved significantly [12]. Therefore, some pre-processing steps including shifting, flipping, scaling, and resizing were carried out to augment the training set of images. Finally, this augmented dataset was used to train the proposed model for carrying out the classification. Some sample images from the dataset after pre-processing are shown in Fig.1.

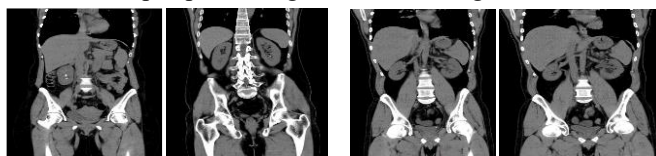


Fig.1. Sample images from input dataset (a) With kidney stone
(b) Normal

2.1 VGG-19 NETWORK MODEL

Simonyan and Zisserman [13] developed a Convolutional Neural Network framework called VGG Net (Visual Geometry Group Network) at Oxford University's Computer Laboratory. 1.3 million images from the ImageNet dataset were used to train VGG-Net. Later, a revised version titled VGG-19 was proposed

which had 19 highly linked layers and outperformed most other similar models at that time. Maxpooling layers were used instead of average pooling before categorization to improve feature extraction. VGG-19 Network takes $224 \times 224 \times 3$ resolution images as input and uses 16 convolution layers (with 3×3 kernels), 5 maxpooling layers, 3 fully connected layers, and one classification layer to achieve the categorization of input images. VGG-19 Network Model uses Rectified Linear Unit (ReLU) activation function at the convolution and fully connected layers and Softmax activation function at the output layer. [14] Fig.2 depicts a pictorial representation of the VGG-19 model's network architecture derived from Xiao et al. [15]. The major purpose of down-sampling layers is to enhance the network's anti-distortion capabilities while preserving primary characteristics and decreasing the dimensionality of the input samples [16] [17] [18].

2.2 PROPOSED NETWORK MODEL

In this research, two network architectures based on the original VGG-19 Network Model were prepared and evaluated for their performance. In both the proposed networks, the input layer and the convolutional layers of the VGG-19 network were used. The network was initialized with weights computed after pre-training the network on imagenet dataset (from the TensorFlow library). Schematic diagrams of both network architectures are presented in Fig.3. As can be observed from the figure, pretrained CNN layers are used to extract relevant features from input images before classification. In the first architecture (shown in Fig.3 (a)), one flattening layer, one fully connected layer, and one softmax layer were added after VGG-19 layers; while in the second architecture (shown in Fig.3(b)), two fully connected layers surrounded by a flattening and a softmax layer were added after the VGG-19 layers.

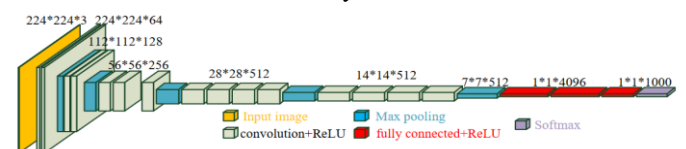


Fig.2. VGG-19 network model (derived from [14])

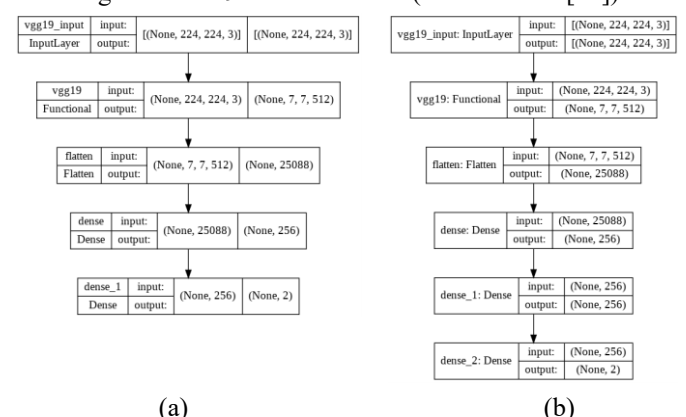


Fig.3. Proposed network model architectures with (a) single dense layer and (b) two dense layers in the classifier block

2.3 NETWORK TRAINING

For training the proposed network, the Binary Cross Entropy was used as the loss function since the network classifies between two classes (Kidney Stone, Normal). The computed loss is an

average of cross-entropy losses on both the output categories as presented in Eq.(1).

$$L = -\frac{1}{n} \sum_{i=1}^n [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)] \quad (1)$$

where, $\hat{y}_i$ is the i^{th} predicted output, y_i are the equivalent target value and n denotes the number of output classes.

This loss function denotes the average result of applying a categorical cross-entropy loss function to a large no. of independent classification samples, where each sample has only two possible classes (with target probabilities y_i and $(1-y_i)$).

While training the proposed network, Adam optimizer (python TensorFlow implementation) was used with a learning rate of $2e^{-5}$ to find the suitable weights for the network.

The input dataset was divided into two parts – 80% of the available images were used in the training process and the rest (346 images) were kept aside to test the efficiency of the trained model. Out of the 1453 CT images used for training, an 80-20 split

was used for training and validation respectively. For each network architecture, results were collected after training the networks for 20 and 30 epochs respectively. Since the pre-trained VGG-19 network weights were used, a small number of epochs was sufficient for obtaining significant results.

3. EXPERIMENTAL RESULTS

The proposed networks were implemented with the help of Keras and Tensorflow libraries in the Python programming language. Performance Results were collected from three different configuration:

1. Configuration (1): Network with a single Dense Layer (shown in Fig.3(a)), Training for 20 epochs
2. Configuration (2): Network with a single Dense Layer (shown in Fig.3(a)), Training for 30 epochs
3. Configuration (3): Network with two Dense Layers (shown in Fig.3(b)), Training for 30 epochs

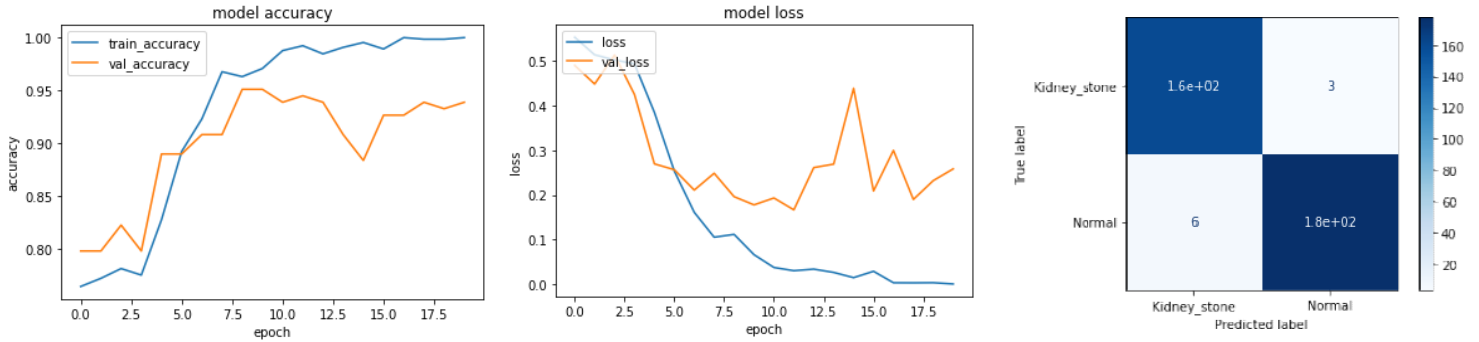


Fig.4. A plot of (a) Training Accuracy, (b) Training Loss and (c) Confusion Matrix corresponding to the configuration 1

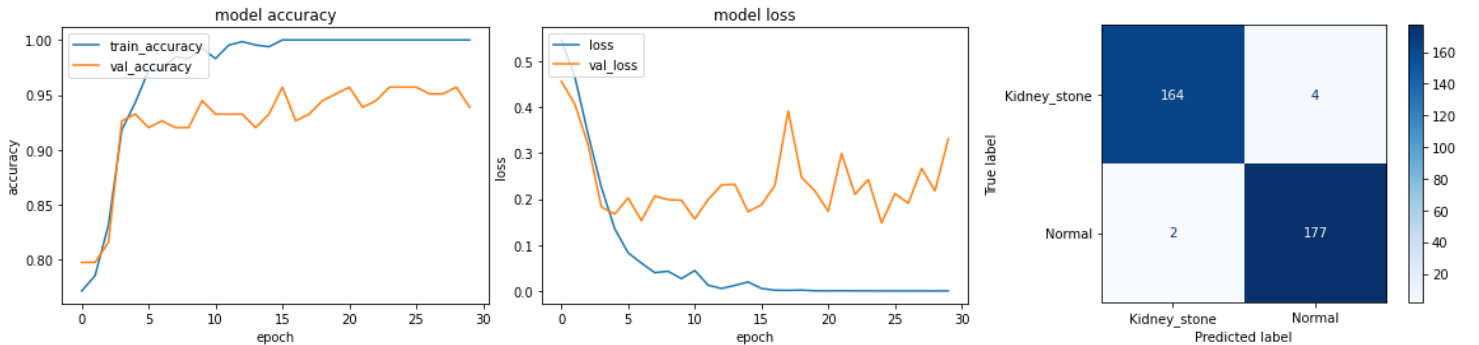


Fig.5. A plot of (a) Training Accuracy, (b) Training Loss and (c) Confusion Matrix corresponding to the configuration 2

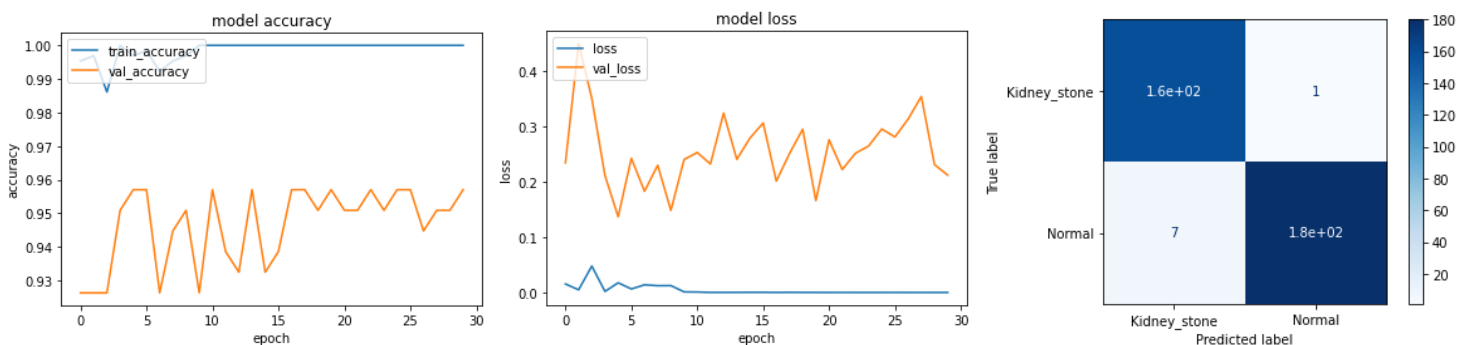


Fig.6. A plot of (a) Training Accuracy, (b) Training Loss and (c) Confusion Matrix corresponding to the configuration 3

3.1 CONFIGURATION 1 TRAINING

Curves representing model accuracy and loss corresponding to the training of the first configuration are presented in Fig.4 (a) and (b) respectively. It can be observed that while the training accuracy is very close to maximum, the difference between network and validation accuracy is very large. Similarly, validation loss is significantly higher than training loss. Fig.4(c) depicts the confusion matrix obtained from testing the network on test samples kept aside during training. In this configuration, out of the 346 test CT images, the model produced 6 false positive and 3 false negative predictions while classifying all remaining test samples accurately.

3.2 CONFIGURATION 2 TRAINING

In this configuration, the number of epochs was increased to check if the model is underfitting during training. Corresponding accuracy and loss curves are presented in Fig.5 (a) and (b) respectively. It can be observed that validation accuracy improves when compared with the first configuration. Similarly, validation loss has also reduced significantly with a few spikes. The testing confusion matrix for this configuration is presented in Fig.5 (c). In this configuration, the model produced 2 false-positive and 4 false-negative predictions while classifying all remaining test samples accurately.

3.3 CONFIGURATION 3 TRAINING

In this configuration, the number of dense layers was incremented to further improve the network performance while keeping the training epochs at 30. Accuracy and Loss Curves for this configuration are presented in Fig.6 (a) and Fig.6 (b) respectively. It can be observed that the training-validation gap increases (for accuracy as well as a loss) when compared to the second configuration. It could be due to the overfitting of the model on the training dataset caused by an additional fully connected dense layer. The impact of the same can be seen in the confusion matrix presented in Fig.6 (c). In this configuration, the network produced 7 false positives and 1 false-negative prediction.

4. PERFORMANCE COMPARISON

A comparison of the performance of all three configurations is presented in Table.1.

Table.1. Performance of different experiment configurations

Network Configuration	Class	Precision	Recall	F1-score
Configuration 1	Kidney Stone	0.96	0.98	0.97
	Normal	0.98	0.97	0.98
Configuration 2	Kidney Stone	0.99	0.98	0.98
	Normal	0.98	0.99	0.98
Configuration 3	Kidney Stone	0.96	0.99	0.98
	Normal	0.99	0.96	0.98

It can be observed that the configuration 2 (single dense layer, 30 epochs) offers the best average accuracy, recall, and F1 score.

In comparison with similar work carried out by Yildirim et al. (2021) on the same dataset using XResNet-50 based network architecture, the proposed network model based on VGG-19 achieves better average accuracy across both classes. See Table.2 for comparison.

Table.2. Performance comparison of the proposed model

Model	Precision (TP/TP+FP)	Recall (TP/TP+FN)	Accuracy (TP+TN/N)
XResnet-50 Architecture [4]	0.9753	0.9576	0.9682
Proposed Network	0.9762	0.9879	0.9827

5. CONCLUSION

In this study, a deep convolutional neural network model based on VGG-19 architecture is proposed for the classification of Coronal Computational Tomography (CT) images for detecting the presence of kidney stones. The proposed model provides the best average accuracy of 98.27%, precision of 97.62%, and recall of 98.79% when trained and tested on a dataset containing 1799 CT images collected from 433 individuals. When compared with similar work, the proposed model is relatively smaller in size (resulting in faster prediction) while providing better performance metrics on the same dataset, which makes it more suitable for deployment in a healthcare system. Researchers interested in extending this study can tune the network further to improve the performance or design a segmentation framework for identifying the location of kidney stones in a given input image.

REFERENCES

- [1] L. Prisilla and I. Laurance Aroqiaraj, "Kidney Stone Detection using Contrast Limited Adaptive Histogram Equalization (CLAHE) on CT Scan Images", *International Journal on Computational Intelligence and Information*, Vol. 7, No. 4, pp. 199-206, 2018.
- [2] C.J. McCarthy, V. Baliyan, H. Kordbacheh, Z. Sajjad, D. Sahani and A. Kambadakone, "Radiology of Renal Stone Disease", *International Journal of Surgery*, Vol. 36, pp. 638-646, 2016.
- [3] K. Yildirim, P.G. Bozdog, M. Talo, O. Yildirim, M. Karabatak, and U.R. Acharya, "Deep Learning Model for Automated Kidney Stone Detection using Coronal CT Images", *Computers in Biology and Medicine*, Vol. 135, pp. 104569-104578, 2021.
- [4] S. Sudharson and P. Kokil, "An Ensemble of Deep Neural Networks for Kidney Ultrasound Image Classification", *Computer Methods and Programs in Biomedicine*, Vol. 197, pp. 105709-105718, 2020.
- [5] Y. Wu and Z. Yi, "Automated Detection of Kidney Abnormalities using Multi-Feature Fusion Convolutional Neural Networks", *Knowledge-based Systems*, Vol. 200, pp. 105873-105885, 2020.
- [6] F. Ma, T. Sun, L. Liu and H. Jing, "Detection and Diagnosis of Chronic Kidney Disease using Deep Learning-Based Heterogeneous Modified Artificial Neural Network",

- Future Generation Computer Systems*, Vol. 111, pp. 17-26, 2020.
- [7] M. Akshaya, R. Nithushaa, N.S.M. Raja and S. Padmapriya, "Kidney Stone Detection using Neural Networks", *Proceedings of International Conference on System, Computation, Automation and Networking*, pp. 1-4, 2020.
- [8] S. Sudharsan and P. Kokil, "Computer-Aided Diagnosis System for the Classification of Multi-Class Kidney Abnormalities in the Noisy Ultrasound Images", *Computer Methods and Programs in Biomedicine*, Vol. 205, pp. 106071-106079, 2021.
- [9] N. Thein, H.A. Nugroho, T.B. Adjil and K. Hamamoto, "An Image Preprocessing Method for Kidney Stone Segmentation in CT Scan Images", *Proceedings of International Conference on Computer Engineering, Network and Intelligent Multimedia*, pp. 147-150, 2018.
- [10] N. Thein, K. Hamamoto, H. A. Nugroho and T. B. Adjil, "A Comparison of Three Preprocessing Techniques for Kidney Stone Segmentation in CT Scan Images", *Proceedings of International Conference on Biomedical Engineering*, pp. 1-5, 2018.
- [11] P. T. Akkasaligar, S. Biradar and V. Kumbar, "Kidney Stone Detection in Computed Tomography Images", *Proceedings of International Conference on Smart Technologies for Smart Nation*, pp. 353-356, 2017.
- [12] A. Gour, "A Hybrid Feature Extraction Approach for Finding Local Discriminative Coordinates for Face Recognition", *Proceedings of International Conference on Smart and Innovative Trends in Next Generation Computing Technologies*, pp. 1-5, 2018.
- [13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", *Proceedings of International Conference on System, Computation and Intelligence*, pp. 1-7, 2014.
- [14] A. Victor Ikechukwu, S. Murali, R. Deepu and R.C. Shivamurthy, "ResNet-50 vs VGG-19 vs Training from Scratch: A Comparative Analysis of the Segmentation and Classification of Pneumonia from Chest X-ray Images", *Global Transitions*, Vol. 2, No. 2, pp. 375-381, 2021.
- [15] J. Xiao, J. Wang, S. Cao and B. Li, "Application of a Novel and Improved VGG-19 Network in the Detection of Workers Wearing Masks", *Journal of Physics: Conference Series*, Vol. 1518, No. 1, pp. 1-12, 2020.
- [16] H. Luo, Y. Yang, B. Tong, F. Wu and B. Fan, "Traffic Sign Recognition using a Multi-Task Convolutional Neural Network", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 19, No. 4, pp. 1100-1111, 2018.
- [17] T. Liu and Z. Wang, "HiCNN: A very Deep Convolutional Neural Network to better enhance the Resolution of Hi-C Data", *Bioinformatics*, Vol. 35, No. 21, pp. 4222-4228, 2019.
- [18] L. Alzubaidi, "Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions", *Journal on Big Data*, Vol. 8, No. 1, pp. 53-73, 2021.