Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 7, July 2021: 1065 - 10475

Research Article

Classification of Image Blood Cancer by Using Multi-Training RNN

Baker Khalid Baker
College of education, Akre
Duhok University,
Duhok, IRAQ
email: baker.khalid@uod.ac

Rakan Mohammed Rashid
College of Basic Education,
Duhok University,
Duhok, IRAQ
rakan.rashid@uod.ac

Nashat Salih Abdulkarim Alsandi
Health Management information System
Department Duhok, Private Technical Institute,
Duhok, IRAQ nashat.alsandi@dhk-pti.com

Omar Farook Mohammad
Computer Technology Engineering
Al-Hadba University College
Mosul, IRAQ
email: ofmalobaidy@hcu.edu.iq
https://orcid.org/0000-0001-7347-5283

Abstract: new method presented in this research to classify bone marrow based on features classification that extracted by human body. Towards this end, new features derived from image based on blood taken by microscope used in proposed descriptor: also human pose Human pose plays important role in extracted features then using these features as the blood cancer input with classifier. In this paper we focused on using retrieval image processing techniques for divided into several steps that include image acquisition, features extraction, and classification. The Retrievable Neural Network (RNN) was used to classify the segmented cells into either normal or abnormal classes based on the features selected by the genetic algorithm (GA). As a result, the classification of cells achieved an accuracy of 98.4%. Subsequently, after the manual review of blood smears, the model will act as a second reader, and it would increase the diagnostic accuracy.

Keywords: *Retrievable Neural Network (RNN), A Multi-Trainable, Classification, Feature extraction.*

I. INTRODUCTION

Nowadays the components of the blood in the human body are considered very important as the cells are an indicator of disturbance in one of the organs, which is difficult to access for diagnosis. Recently, Blood, since it helps transfer minerals and oxygen to all areas of the body, is one of the most essential parts of the human body. Blood consists of three basic components: white blood cells (WBC), red blood cells (RBC) and platelets [1]. Blood may be infected with different diseases that lead to malfunction of the body and thus influence human life and health. Some of these disorders are not very serious and can be treated, such as anaemia, while others, such as blood cancer, are very serious and not possible to treat. Cancer of the blood that affects the bone marrow, characterized by an increase in the number of immature WBCs called (blast cells) as shown in Figure 1, illustrates the blood components presented [2].

The precise identification and classification of acute blood cancer that use image-processing techniques remains difficult despite comprehensive study. The reasons for different forms of Acute Blood Cancer (ABC) are due to the difference in scale, shapes, locations, and image intensity. Blood cancer detection using image processing and computer vision techniques consists of five steps, including image acquisition, image pre-processing, segmentation of WBCs, extraction of features, and cancer blood cell classification [3].

The main goal of this paper is to design, develop, implement and evaluate the performance of a computerized model that analyzes the microscopic blood images to segment and diagnose the ABC cells. This model is designed by creating a multi trainable machine learning technique that is able to perform segmentation and detection of the cells of interest (COI) or (blast cells) on the microscopic blood images and this multi-trainable model can distinguish between the ABC and Non-ABC cells.

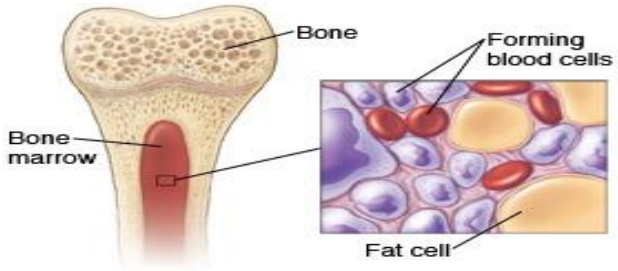

Fig.1. The blood components [2]

In this paper, the multi-trainable machine-learning approach for classification of acute blood cancer provides two significant factors that include solve the blast cell issues and getting an accurate diagnosis. These two factors can be achieved using the proposed machine learning techniques using Back-Propagation with Neural Network with back-Retrievable Neural Network (RNN). Hence, this technique allows for the screening of a large number of complex and variation blood images; as well as eliminating human error such as the error caused by lack of experience, fatigue, and repetition [4].

II. RELATED WORK

In term of health blood cancer has a record of current symptoms such as fatigue, weakness, weight loss, bone pain, swelling of the lymph nodes (such as the neck and armpits), enlarged liver or spleen, red spots

on the skin, etc., and a patient's past problems. In addition, the medical history of the family of a patient also helps in diagnosing cancer. [5].

According to image processing includes the following steps: firstly, importing the digital image via the acquisition tools. Secondly, manipulating and analyzing an image to extract useful information. Finally, the output result can be a report or image related to an analysis of the original image. Image analysts have used various fundamentals of interpretation of visual techniques (image processing techniques). Image processing techniques use computers to help in the manipulation and analysis of digital images. There are three general phases using the digital technique that include pre-processing by enhancing an image, information extraction and classification [6].

Trainable segmentation has used to isolate the ROI from the rest of an image background followed by feature extraction to identify the abnormality of cases. This approach starts by train the model with some classes (i.e. labeling of cases). By selecting some of the regions from the image and then extraction and calculating a number of features, thus, pass the results of the features extraction technique to a classifier for identification process [7].

According to feature selection is crucial and important to improve the classification results. In acute lymphoblastic leukemia, some of the features selection techniques have introduced for optimization, such as Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Dragonfly Algorithm (DA), and Cuckoo Search (CS). These techniques dedicated to dimension reduction for acute lymphoblastic leukemia classification with more accurate results [8,20].

In addition, the classification technique is one of the processes used to classify the image input patterns (e.g. blast cells) into one of the sets previously identified on the basis of characteristics, such as form, color, and texture. There are different methods to construct an object classifier, including the two most common classifiers known as: first, Support Vector Machine, and secondly, Artificial Neural Network (ANN) [9].

Multiple stages recommended by investigator for a new diagnosis of the technique of acute lymphoblastic leukaemia. This approach includes the acquisition of blood images using the public dataset (ALL-IDB1), the segmentation of blast cells using marker-based segmentation (MBS), and then the extraction of features determined using the grey level co-occurrence matrix (GLCM) from segmented blast cells with reduced and selected specific features using probabilistic main component analysis In addition, at the classification level, the Random Forest (RF) was proposed to classify the segmented cell into a normal or abnormal group [10].

Moreover, introduced a new approach for diagnosis of the acute image, which includes the following stages. Firstly, it converted the blood image from RGB to gray-scale and then implemented the histogram equalization to enhance the images. Secondly, it implemented the segmentation stage based on the WBC edges detection and then used HSV conversion with erosion to segment the blast cells from the rest of blood image background. Thirdly, features extraction stage based on geometrical and statistical features, color and texture which was calculated from the segmented blast cells. Finally, the classification stage was implemented using Fuzzy C-means clustering technique to diagnose image classification [11].

The author presented tried to simulate the learning of human brain strategies for decision making purposes. These techniques are often used for medical image classification, such as feed-forward neural networks (FFNN), Retrievable Neural Network (RNN), Kohonen self-organizing maps (SOMs), Hopfield neural networks (HNNs), and others [12].

III. TRAINING BPNN IN MACHINE LEARNING

The RNN was trained using the features selected by the GA in the previous step for the positive class, negative class, and background (three classes that were cropped in truncate regions step). Thus, RNN optimization was used to identify the best parameters namely; the number of hidden layers and the number of nodes in the hidden layers are needed for the neural network.

In order to implement the RNN classifiers, the blood images needed to be selected for the training segmentation model. The total number of blood image samples equal to 352 images, out of which 179 were class A image (abnormal) samples and the remaining 189 were class B (normal) samples. It is worth mentioning that the classification was performed based on three main classes that include positive class (blast cells), negative class (RBC), and background.

In order to select the datasets for training processes, this work has used 70 percent of the blood image samples (281 blood images), two-thirds (55 percent) of the blood image samples (238 blood images), and 50 percent of the blood image samples (200 blood images), out of 352 blood images to train the classification model. Table I, illustrates the specification of samples for training and testing the trainable classification model.

TABLE I. TRAINING AND TESTING MODEL SAMPLES FOR

Number of samples	Training set	Testing set
352	281	62
352	238	101
352	200	133

These parameters include learning rate, number of neurons, number of hidden layers, number of epochs and the activation function, the number of RNN classifier input nodes would be equal to the number of selected features; and the output node was set to three outputs (blast cells, RBC and background). The number of RNN hidden layers was chosen as one at this stage of the proposed model, as expanding the number of hidden layers which result in over-fitting the data. This explains why one secret layer is normally appropriate [13].

The learning rate was tested as 0.001, 0.01 and 0.1, the maximum number of epochs (training cycle) would be 1000 and the number of neurons in the hidden layer was 2, 4, 6, 8, 10, and 12. From these processes, the validation accuracies were taken every 100 training cycles and the performance of learning rates was tested, which revealed that the best of them was at 0.01 with 6 nodes in the hidden layer, as shown in Figure 2.

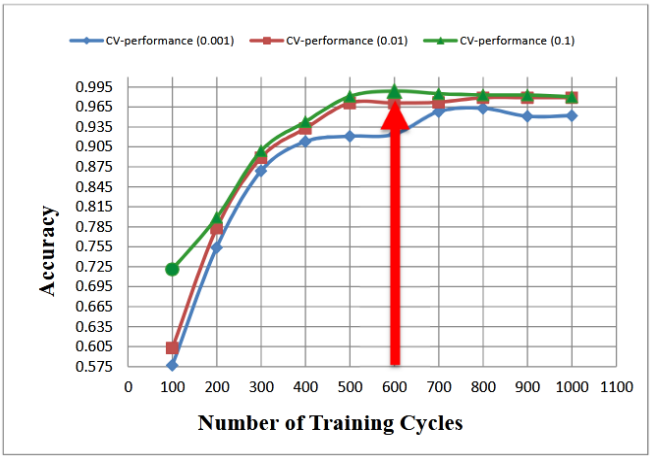

Fig.2. Training cycles and the performance of hidden layers of accuracy

IV. METHOD AND DISCUSSION

The proposed multi-trainable model contains several steps for segmentation and diagnosis of acute blood cancer cells using region features information for instance, proposed trainable segmentation (i.e. one of the segmentation methods that depends on training the model to isolate the region of interest from the rest of an image) to isolate the COI of image background.

The main aim of trainable segmented use to successfully isolate blast cells from the rest of blood images background based on local region information. Due to the complex setting of the microscope with the camera, the capturing of microscopic blood images is a complicated task for people. There are two public image datasets used in this paper, namely, ABC-IDB1 and ABC-IDB2. These datasets include images of acute blood cancer cells and ordinary blood cells.

In the training model for machine learning process of this phase started by selecting a certain number of blood images as training samples. In each training sample, cropping a number of regions was taken from the inside of the blood image for instance, small windows of square shape of different sizes (3x3, 4x4 and others). Figure 3 illustrates the training phase for the segmentation of the blast cells (COI). To truncate small regions, a small tool was implemented from three different classes for instance, truncate region from class 1 (positive class) (or COI), class 2 (Negative class or RBC) (or Non-COI) and class 3 (background). These processes were conducted manually by truncating and labeling those regions for the training phase, where a small basic tool that enables the user to tap on an object in the blood image (tap on the positive regions, negative region, and background) was used and implemented. Consequently, the truncated parts were used to extract the features and choose the relevant features by using genetic algorithm (GA) as described in details in the next section.

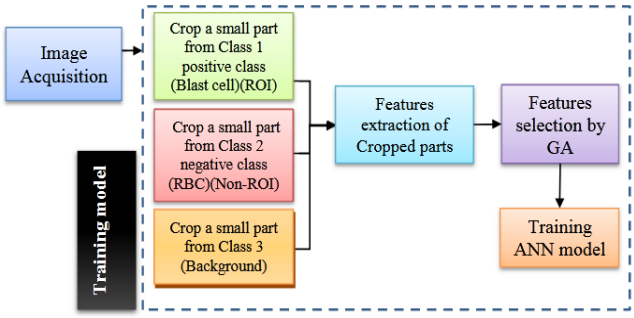

Fig. 3. The training model

The set of texture features were extracted (4-moments with 3-texture features) from the three classes, where those classes have been labeled based on those features. Thus, a range of appropriate features was chosen to label the three classes; therefore, proper extraction of features is essential for the better determination of the COI and Non-COI. Moreover, these extracted features should represent the characteristics (in terms of similarity and difference between the COI and Non-COI) of each object in the blood image as much as possible.

Moreover, this classifier has been used extensively in the classification of medical images within the image-processing techniques and machine-learning approaches. The testing stage of trainable segmentation technique started by scanning microscope blood image with different window sizes (pixel by pixel) to catch all objects in the image and then divided the objects in the image into three classes that involve class 1: positive class (blast cells), class 2: negative class (read blood cells) and class 3 (background) based on the 7-texture features [14]. Then, these classes of information were passed to an artificial neural network classifier to filter out the Non-COI and get the COI.

Figure 5 illustrates the various stages of the proposed testing classification model. The model includes the following steps: (a) blast cell (COI) identification from microscope blood images and separates the overlapping regions, (b) feature extraction from the COI, (c) feature selection using GA, and (d) classification of ROI into class A or class B using an RNN classifier. The main goals of the proposed multi-trainable model were to successfully classification diagnose cancer cells (COI) based on the blood images.

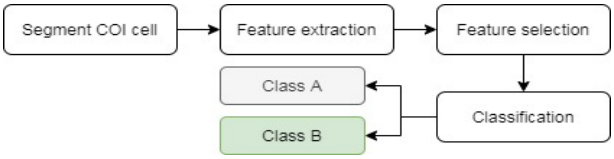

Fig.5. The testing of classification model

The original blood image sample was divided into 10 equal size sub-samples in this phase of the analysis. A single sub-sample was retained from the 10 sub-samples as validation data for evaluating the trainable model, and the remaining 9 sub-samples were used as training data. The method of cross-validation was replicated 10 times (number of folds), with each of the 10 sub-samples used as validation data exactly once. The key benefit of this approach was that both preparation and confirmation were used with all observations [15].

The number of RNN hidden layers was selected as one at this point of the proposed model, while the hyperbolic tangent function was used as an activation function; the hyperbolic tangent function

obtained better performance results compared to the Sigmoid function. The overall number of epochs (training cycle) will be 1000, where the precision of validation was taken every 100 training cycles, while the learning rate was 0.001, 0.01, and 0.1. Learning rates have been checked for success. However, the number of neurons in the hidden layer was 2, 4, 6, 8, 10, and 12 as shown above in a new technique used to identify the cells, whether they are ABC or Non-ABC, was created after the blast cell segmentation design was completed. The testing classification stage is shown in Figure 5.

To resolve the blast cells, color and texture characteristics are extracted the features. Therefore, two forms of methods of feature analysis were used, namely: GLCM texture and histogram-based color derived from RGB channel grey scale levels. Figure 6 shows the extraction mechanism of blast cells.

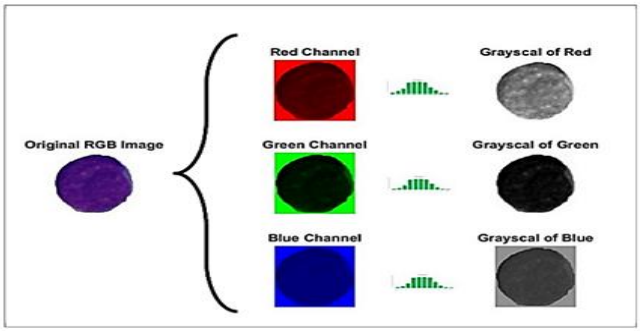

Fig.6. Feature extraction of blast cells GLCM based on histogram color.

The optimal features were selected using GA after the texture and histogram-based color features had been extracted from the segmented cell. The aim of selecting features is to choose the most important features that are useful for the stage of learning. It is one of the significant steps in the classification process because it can lead to uncertainty by maintaining unimportant features.

The key objectives of using the two-stage feature selection method, namely the segmentation and classification phases of this work, are to increase the accuracy of the classification and decrease the number of features. In the previous step, GA had chosen the appropriate features, so the next step in this suggested trainable model was to transfer this subset of features to the RNN classifier to identify the segmented cells and decide if they are ABC or Non-ABC. Two processes such as training and testing were included in the RNN classifier in this phase.

In this work, five metrics were used to evaluate the performance of the proposed A Multi-Trainable Model technique, which includes the specificity, sensitivity, precision, accuracy, and F-measure, as described in details in the next sections. Firstly, the specificity refers to the rate of true negative that is predicated as correctly negative [16,19]. The specificity is illustrated in Equation 1.

$$\text{Specificity } \frac{TN}{TN+FP} \quad (1)$$

Secondly, the sensitivity refers to the rate of true positive which is predicated as correctly positive. The sensitivity is illustrated in Equation 2.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

Thirdly, the precision refers to the consistency of the results. The precision is illustrated in Equation 3.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

Fourthly, the accuracy refers to the overall correctness of proposed classifier. The accuracy is illustrated in Equation 4.

$$\text{Accuracy} = \frac{(TP+TN)}{TP+FN+FP+TN} \quad (4)$$

Finally, the F-measure refers to the measure of tests accuracy. The F-measure is illustrated in Equation 5.

$$\text{F- Measure} = \frac{2*precision * sensitivity}{precision + sensitivity} \quad (5)$$

Moreover, the previous examination was performed through visual human standard measurement techniques. In addition, our work was evaluated by comparing results with the manual examination implemented by some experts in the field of detection of blood cancer. The proposed trainable classification technique was compared with the manual model.

V. RESULTS AND DISCUSSION

ALL-IDB1 consists of 108 images in JPG format with 24-bit color depth, and these images were captured with a microscope consisted of a Canon camera (PowerShot G5) with 2592 ×1944 resolution. Other images were captured with a microscope consisted of an Olympus camera (C2500L) with 1712 ×1368 resolution.

The blood image samples were captured using different ranging of camera enlargements ranging from 300x to 500x. Figure 7 shows some of the blood image samples taken from ALL-IDB1.

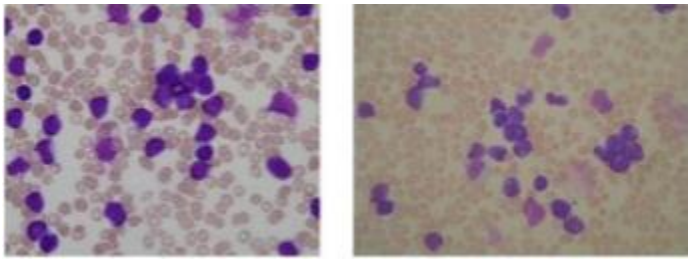

Fig.7. Blood image samples taken from ALL-IDB.

ALL-IDB2 consists of 260 images in TIF format with 24-bit color depth, and these images were captured with a microscope consisted of a camera with 257 ×257 resolution.

Many researchers have used the ALL-IDB2 datasets for the blast cells segmentation and classification [17,18]. Figure 8 shows some blood image samples from ALLIDB2.

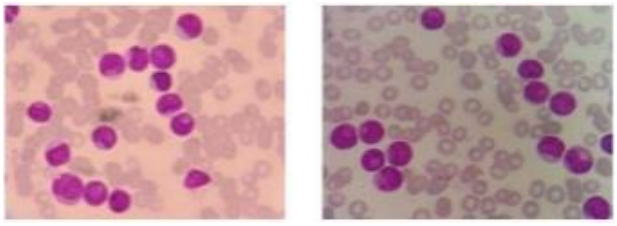

Figure 8. Blood image samples taken from ALL-IDB2.

Several experiments have been used to assess the efficiency of the RNN model classification. The best performance results of the classification (RNN) model were derived with the following parameters: 35 neurons as the input layer, 0.1 of the learning rate, four neurons at the hidden layer, and after 800 training cycles were chosen for subsequent experiments and met the objective in this work.

To test the performance of the RNN model for classifying segmented cell types (normal or abnormal), a total of 368 samples were used to train and test the classification model. Furthermore, each segmented cell (using a train able segmentation model) extracted a vector of features that consisted of 35 features computed by the GLCM texture with histogram-based color features, and selected using GA. On the other hand, another classification experiment have been implemented without using the features selection stage (using the full feature set) to test the performance of the classification (RNN) model. The classification performance results of the dataset for testing as illustrated in Table II.

TABLE II CLASSIFICATION PERFORMANCE RESULTS FOR TRAINING AND TESTING OF DATASET.

Classification performance	RNN with FS	RNN without FS
TP	53	50
TN	55	48
FP	1	4
FN	2	8
Sensitivity	0.964	0.862
Specificity	0.981	0.925
Precision	0.988	0.928
Accuracy	0.972	0.890
F-measure	0.973	0.893

Several directions have been analyzed for further advancement of in these kinds of work. Therefore, the basis for future work as proposed trainable segmentation technique used the artificial neural network as a classifier to classify the regions in the blood image. However, the classification stage in the proposed segmentation technique needs to be tested on other classifiers, for instance, another classifier can be utilized to classify the regions in the image such as Decision tree, Support Vector Machine (SVM), Fuzzy Measure and others.

VI. CONCLUSION

RNN reported that recognize a human medical such as blood cancer disease and its diagnosis depends on counting the percentages of blasts in the bone marrow or blood by used a Multi-Trainable machine learning techniques. Our contribution of Multi-trainable machine learning technique was used to classify. The experiments achieved the classification performance results as the following: 96.4% for sensitivity, 98.1% for specificity, 98.8% for precision, 97.2% for accuracy, and 97.3% for F-measure.

ACKNOWLEDGMEN

REFERENCES

[1] Parra JJ, Alvarado MC, Monsalve P, Costa AL, Montesinos GA, Parra PA. Oral health in children with acute lymphoblastic leukaemia: before and after chemotherapy treatment. European Archives of Paediatric Dentistry. 2020 Feb;21(1):129-36.

[2] Russ JC, Russ JC. Introduction to image processing and analysis. CRC press; 2017 Dec 19.

[3] Begum, A.J. and Razak, T.A., Diagnosing Leukaemia from Microscopic Images Using Image Analysis and Processing Techniques. Computing and Communication Technologies (WCCCT), 2017 World Congress on. 2017 IEEE, pp. 227–230.

[4] Liu, H., Cao, H. and Song, E., 2019. Bone Marrow Cells Detection: A Technique for the Microscopic Image Analysis. Journal of medical systems, 43(4), p.82.

[5] Liu JE, An FP. Image classification algorithm based on deep learning-kernel function. Scientific programming. 2020 Jan 31;2020.

[6] Blindar VN, Nikitaev VG, Polyakov EV, Matveeva II. Method of automating of the separation of blasts and lymphocytes in the diagnosis of acute myeloid leukemia. InJournal of Physics: Conference Series 2017 (Vol. 798, No. 1, p. 012129). IOP Publishing.

[7] Su J, Liu S, Song J. A segmentation method based on HMRF for the aided diagnosis of acute myeloid leukemia. Computer methods and programs in biomedicine. 2017 Dec 1;152:115-23.

[8] Mohammed, M.A. et al. Trainable model for segmenting and identifying Nasopharyngeal carcinoma. Computers & Electrical Engineering, 71, 2018, pp.372–387.

[9] Abdi MJ, Hosseini SM, Rezghi M. A novel weighted support vector machine based on particle swarm optimization for gene selection and tumor classification. Computational and mathematical methods in medicine. 2012 Jan 1;2012.

[10] Huang L, Ye C, Gao J, Shih PC, Mngumi F, Mei X. Personnel Scheduling Problem under Hierarchical Management Based on Intelligent Algorithm. Complexity. 2021 Jan 12;2021.

[11] Sirinukunwattana K, Raza SE, Tsang YW, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE transactions on medical imaging. 2016 Feb 4;35(5):1196-206.

[12] Singh, G., Bathla, G. and Kaur, S., 2016. Design of New Architecture to detect leukaemia cancer from medical images. Int J Appl Eng Res, 11(10), pp.7087–7094.

[13] Wang J, Gao J, Tan T, Wang J, Ma M. Event-Triggered Filtering for Markovian Jump Neural Networks under Random Missing Measurements and Deception Attacks. Complexity. 2020 Dec 29;2020.

[14] Al-jaboriy SS, Sjarif NN, Chuprat S, Abduallah WM. Acute lymphoblastic leukemia segmentation using local pixel information. Pattern Recognition Letters. 2019 Jul 1;125:85-90.

[15] Fukunaga K. Introduction to statistical pattern recognition. Elsevier; 2013 Oct 22.

[16] Solomon ML, Bryan KJ, Smith KE, Clauter DA, Smith AO, Peter AM. Infrasound threat classification: a statistical comparison of deep learning architectures. InChemical, Biological, Radiological, Nuclear, and Explosives (CBRNE) Sensing XIX 2018 May 16 (Vol. 10629, p. 1062917). International Society for Optics and Photonics.

[17] Wen S, Xiao S, Yang Y, Yan Z, Zeng Z, Huang T. Adjusting learning rate of memristor-based multilayer neural networks via fuzzy method. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2018 May 9;38(6):1084-94.

[18] Sulong, Ghazali, And Ammar Mohammedali. Recognition Of Human Activities From Still Image Using Novel Classifier. Journal of Theoretical & Applied Information Technology 71.1 (2015).

[19] Atiyha, Baqer Turki, Salahaldain Aljabbar, Ammar Ali, and Abdullah Jaber.. An improved cost estimation for unit commitment using back propagation algorithm. Malaysian Journal of Fundamental and Applied Sciences 15.2 (2019): 243-248.

[20] Sulong, Ghazali, and Ammar Mohammedali. Human Activities Recognition Via Features Extraction from Skeleton. Journal of Theoretical & Applied Information Technology 68.3 (2014).