Classification of Imbalanced leukocytes Dataset using ANN-based Deep Learning

Jameela Ali Alkrimi¹, Rajaa Salih Mohammed Hasin², Ali Zaki Naji³, Loay E. George⁴, Sherna Aziz Tome⁵

¹,³ College of Dentistry, University of Babylon, Babylon, Iraq.
²Al Mansur Institute of Medical Technology, Middle Technical University Baghdad, Iraq
³University of Information Technology & Communications, Iraq
⁴College of medicine, Baghdad University, Baghdad, Iraq
sci.ahmed.ali@uobabylon.edu.iq

ABSTRACT

Nowadays, classification of imbalanced data is a major challenge in the machine learning (ML) algorithms, especially in medical data analysis. In this paper, deep learning algorithm which is the advance artificial neural network (ANN) is used for classifying five white blood cells (WBCs). Different preprocessing image techniques and algorithms are applied to isolate WBCs and segment the nucleus for the cytoplasm. Geometric, statistical and color features are extracted, the principal component analysis technique is applied to select the optimal features. The classification process has been repeated several times to tune the algorithm parameters and to find the best patterns match through the training data in the learning process until achieve best classification accuracy. Multi-class classification results show high accuracy of more than 94% for the five types of WBCs. We evaluate the classification model using the geometric mean, Cohen’s Kappa, Receiver operating characteristic curve, Root mean squared error, relative absolute error and cross-validation techniques. The algorithm model achieves high accuracy and can conduct a multi-class classification of imbalanced datasets in terms of the above-mentioned metrics.

Keywords: Imbalanced data, Deep learning, Leukocytes, Pre-processing image, Artificial neural network

1. INTRODUCTION

Imbalanced classes are a common problem in machine learning (ML) classification. In this classification problem, the classes are unequally represented. Classification of imbalanced datasets is an inevitable problem in medical intelligent diagnosis because medical data have limited samples and high-dimensional features. This condition affects the classification performance of the model and causes inaccurate guidance for the diagnosis of diseases. Classification of imbalanced datasets is a challenging predictive modelling task due to the severely skewed classes distribution. This condition causes an inferior performance of traditional ML models and evaluation metrics that assumed a balanced classes distribution (Zhang, 2018), (Wang, 2016), (Ali, 2015). Predictive modelling of such classification is complex given that most ML algorithms used for classification are considered under the assumption of an equal number of examples for each class (Korfiatis, 2019) (Sun, 2007).

A classification problem may be a little skewed, namely, it has a slight imbalance. Otherwise, the classification problem may have a severe imbalance, in which enormous examples may be present in one class and tens of examples in another class for a given training dataset (Géron, 2019), (Vanhoeyveld, 2018). The consequence of this condition lead to a poor predictive performance of the models, specifically for the minority class. The classification errors are more sensitive for the minority class than the majority class so that, the minority class is more important than the majority class (Yousefi, 2019) (Lemaître, 2017). Classes imbalance occurs when one class is highly represented in the dataset compared with others. Datasets are divided into two categories: balanced and imbalanced datasets. In imbalanced datasets, instances are divided into two sets: majority instances, which are the most frequent ones; and minority instances, which are the less frequent ones (Akbani, 2004). A low number of examples from one class may also mean insufficient information available for any obvious difference pattern between classes to be
discovered by the algorithms (Salehinejad, 2018). Imbalanced data should be given attention due to two reasons: accuracy paradox and algorithm bias towards the highly represented class.

Leukocytes perform an important role in the immune system by protecting the body from infectious diseases. The number of WBCs and the count of different WBCs play a crucial role in clinical diagnoses and tests: they reflect the hidden infection within the body and alert the haematologist. The presence of WBCs in the medical world is also used as an indicator of various types of diseases, such as leukaemia, immunological disorders and certain types of cancer (Fan, 2019), (Yıldırım, 2019), (AL-Dulaimi, 2018). WBCs can be considered into five types: eosinophils, lymphocytes, neutrophils, monocytes and basophils. Neutrophils are the most abundant with 50%–70% of WBCs, and they are responsible for defending against bacteria or fungal infection. Eosinophils occupy around 2%–4% of WBCs and act in response to allergies and parasite infection. Lymphocytes have 25%–45% of WBCs and undertake the task of the specific recognition of foreign agents and the consequent removal from the host. Monocytes are of 3%–8% of WBCs and effective in direct destruction of pathogens and clean-up of debris from infection sites. Basophils are the least common type of WBCs (Negm, 2018) (Gonzalez Viejo, 2019), (AL-Dulaimi, 2018), (Doan, 1954).

Currently, Artificial neural networks based Deep learning (ANNs) represents an optimal selection in image applications for medical purposes, such as medical image detection and classification. ANNs achieved preferable results subjected to availability to engh data for deep neural networks training process. Also, it required a high-quality computational resource to train and test the required size of the medical data. The limited data set may affect negatively on the deep neural networks training process from scratch in many cases. In such scenario, transfer learning can be utilized to leverage the power of applied Algorithm and reduce the computational costs (Sahlol, 2020) (Nguyen, 2018) (LeCun, 2015). In this Algorithm, the deep neural network is firstly pre-trained on the various and considerable general image dataset and then has been applied for selected task. These predictive modelling problems are challenging because an enough representative number of examples of each class are required for a model to learn the problems. The difficulty is evident when the number of examples in each class is imbalanced or skewed towards one or a few of the classes have very few examples of other classes. Problems of this type are referred to as imbalanced multi-class classification problems, and they require the careful design of an evaluation metric and test harness and choice of ML models (Das, 2020) (Madasamy, 2017).

An inclusive literature review was applied to addressing class imbalance with deep learning algorithms. In the 1990s, back-propagation algorithm deep neural networks applied to discover the effects of class imbalance by Anand et al. The researchers indicated to the difference length of the gradient component between minority class and majority class as they proved that the minority class has the shortest one (Khan, 2017). Essentially, the majority class has dominating role on the net gradient which in charge of updating the model weights as Ben mentioned in his work in the 1990s, back-propagation algorithm deep neural networks applied to discover the effects of class imbalance by Anand et al. The researchers indicated to the difference length of the gradient component between minority class and majority class as they proved that the minority class has the shortest one (Khan, 2017). Essentially, the majority class has dominating role on the net gradient which in charge of updating the model weights as Ben mentioned in his work in the 1990s. This condition decreases the error of the majority group very rapidly in the precocious repetition but often grows the error of the minority group that lead the network to get stuck in a slow mode of convergence (Anand R, 1993). Khan et al in 2015, showed that during training phase, effective cost-sensitive of the deep learning that learns simultaneously the network weight parameters and class misclassification feedbacks. Johnson et al. studied the problems of deep learning with a class imbalance to enhance understanding the efficacy of deep learning when applied to the imbalanced class datasets (Johnson, 2019). The methods for usage class imbalance in ML can be classify into three groups: algorithm-level methods, data-level techniques and the hybrid method (Polat, 2018) (L, 2016). The data-level techniques try to minimize the level of imbalance over several data sampling approaches. Where the algorithm-level approaches are generally performed with a cost or weight sketch, including adjusting the implied learner or its outcomes to decrease the bias across the majority classes. Hybrid methods integrate the sampling and algorithmic technique in strategic way (Ali H. S., 2019) (Li, 2017).

The most common evaluation method that use to evaluate the classification performance is the accuracy rate. However, in the basis of imbalanced datasets, accuracy is not an efficient measurement method because it does not identify the numbers of rightly classified samples of different classes. This inadequacy may drive to mistaken conclusions. (Tyagi, 2020), (Fernández, 2013) In imbalanced fields, the confusion matrix has been adopted as a significant evaluation method to measure the classification performance of positive and negative classes independently to evaluate the performance of the classifier. (Thabtah, 2020), (García, 2010).

This work aims to apply ANN-based deep learning with imbalanced class for multi-class classification of normal and abnormal WBC datasets. Loss functions in different forms have been suggested to ensure achieving high
classification accuracy by making learning algorithm more sensitive to the minority class. Presently, the mostly adopted loss function in the deep learning algorithms is the mean squared error.

2. Materials and Methodology

The first step in this study is data collection. Thereafter, we suggest a classification scheme with four main stages, as shown in Figure 1. The first stage is data collection. The images are preprocessed to segment WBCs from the whole image. After WBCs are segmented, feature extraction and selection are performed in the second stage. In the third stage, ANN as the ML algorithm is applied using Waikato Environment for Knowledge Analysis (Weka) software to classify five types of WBCs. The classification performance is evaluated using different evaluation metrics in the fourth stage.

Figure 1. the classification WBCs scheme

2.1 Preprocessing

Our dataset is composed of 180 blood smear images; each image includes at least one WBC, as shown in Figure 2. The images are obtained from the Department of Laboratories at Al-Hilla Teaching Hospital under the approval of the consultant pathologist Dr. Ali Zaki Naji. Several image preprocessing steps include one or more algorithms and techniques to isolate the individual WBCs, as shown in Figure 2. The perfect segmentation process is performed following [40]. The segmentation process aims to isolate the nucleus for the cytoplasm for extracting its features. Each step includes several methods and algorithms; segmentation is the most important step in processing medical images [41]. Failure of the above-mentioned step affects the sensitivity level. The sensitivity lost in this step cannot be recovered in the later steps [42]. The main processes are gap filling, noise cleaning, removal of small black and white cut outs using seed filling, image resizing using bilinear interpolation method and edge smoothing using mean filter technique. A total of 1000 individual WBCs are obtained. Each image has a size of 354 × 362. The images have been hand labelled by a pathologist and are collected from an existing dataset.

Figure 2. blood smear images with different number of WBCs.

2.2 Identification of WBCs

The process obtains 1000 WBCs for five classes, and each class includes different numbers of cells, as shown in Table 1. WBCs contain nuclei and cytoplasms and can be categorised into five classes, which are mentioned earlier. The idea of most existing methods is to identify the nuclei firstly because they are more noticeable than other components. The nucleus of each type has a unique shape, which is the most important feature of it [43][44]. The nucleus of neutrophils has a ‘U’ shape or a curved rod before segmentation. Basophils are the smallest circulating granulocytes, and their granules are large and very numerous. Eosinophil are simply distinguished in stained smears by their massive granules. The nucleus of eosinophil often has two lobes connected by a band of nuclear materials. Their granules grab onto a large amount of basic dyes and have deep blue purple colour. Monocytes represent the biggest’s kind of WBCs, but they have only single rarely or barely lobed nucleus. Folds can be observed in the nucleus, which can be of different shapes as round, lobular and kidney like. The nucleus of lymphocytes is a bit oval or round and stained dark [45][46].
3. Feature Extraction

The process of features extraction determines a group of features or characteristics that sampling to the significant information required for classification task [47]. The features extraction has an essential role of the automatic WBC classification system performance. The main WBC features extracted from WBCs are the number of nuclei, the colour of cytoplasm and the entire cell, as shown in Table 1. Three types of features, namely, geometrical, statistical and colour features, are extracted for classification. Most existing feature methods extract geometrical features, including WBC size, nucleus cell and nucleus shape [48]. Textural features comprise statistical features, such as momentum and contrast. Colour features use variance and histogram. A total of 34 of these features are extracted in this study from three-colour (RGB) and grey images.

\begin{equation}
F_r(n) = \frac{T}{2\pi} \sum_{i} \Delta \delta_{i} \cos \left( \frac{2\pi}{T} \sum_{j} \Delta \delta_{j} \right) - \sin \left( \frac{2\pi}{T} \sum_{j} \Delta \delta_{j} \right)
\end{equation}

\text{where,}

\begin{align*}
\{x_1, x_2, x_3, ..., x_n\} & \text{ are the } x \text{-coordinates of external boundary points.} \\
\{y_1, y_2, y_3, ..., y_n\} & \text{ are the } y \text{-coordinates of external boundary points.} \\
M & \text{ is the number of boundary points.} \\
\Delta \delta_i & = x(i) - x(i-1) \\
\Delta \delta_j & = y(i) - y(i-1) \\
\Delta \delta & = \sqrt{\Delta \delta_i^2 + \Delta \delta_j^2} \\
T & = \sum \Delta \delta(i)
\end{align*}

\begin{equation}
\text{Contrast} \quad Cnt(n) = \frac{Vr(n + 1)}{R}, \quad n = 1, 2, 3 \quad \text{Eq. (2)}
\end{equation}

\begin{equation}
\text{Moment} \quad Mom(n) = \frac{1}{k} \sum \left( x_i + jy_j \right)^n, \quad n = 1, 2, 3, 4 \quad \text{Eq. (3)}
\end{equation}

\begin{equation}
\text{Variance} \quad VrX(n) = \frac{1}{k} \sum x_i^n, \quad n = 1, 2, 3, 4 \quad \text{Eq. (4)}
\end{equation}

4. Classification Task

The sampling process was applied after the preprocessing stage then the deep neural classification model was implemented. One run of 10-fold cross-validation is executed for each experiment to ensure the model learns from all observations. An algorithm with hyper-parameter ‘balanced’ class weight is used as an algorithm-level technique because of its simplicity and common application in credit risk. This research applies ANN algorithms with several
steps for imbalanced datasets to classify WBC types. Dataset noise is cleaned because the set of features used for classification is insufficient to draw class boundaries, data examples are mislabelled and data acquisition processes are poor [49][50]. Thus, training ANN for learning is not ideal. SPSS version 20 is used to clean the dataset to achieve a chisquare statistic of 10.73 and a Sig of 0.030. Thus, the assumption is satisfied. Thereafter, feature reduction is performed using principal component analysis (PCA) to reduce poor-quality features. As a result, only four components, namely, the size of WBCs, the shape and number of nuclei and the colour of cytoplasm (bluish and reddish), are input to the ANN algorithm. The classification task is performed using Weka software. The classification task requires adjusting the parameters, including the number of clusters, clustering seed and the ridge parameters for the linear regression. These parameters are experimentally determined. The ridge parameter is tested and has a value of $1 \times 10^{-8}$.

5. RESULTS

The classification results of five types of WBCs under an imbalanced dataset using ANN-based deep learning algorithm show a high classification accuracy of WBCs of more than 94% after several adjustments to the parameters, namely, the number of clusters, clustering seed and the ridge parameter, as shown in Table 3. The best accuracy is achieved after 100 attempts to balance the parameters of ANN algorithms, as shown in Figure 3 and Table 2. Figure 4 shows the correct and incorrect classification of WBC types.

| Experiment | Number of Clusters | Clustering Seed |
|------------|--------------------|-----------------|
|            | 1                  | 1               |
|            | 2                  | 2               |
|            | 4                  | 4               |
|            | 6                  | 6               |
|            | 10                 | 10              |
|            | 20                 | 1               |
|            | 30                 | 2               |
|            | 40                 | 4               |
|            | 70                 | 6               |
|            | 100                | 10              |

| Experiment | Number of Clusters | Clustering Seed |
|------------|--------------------|-----------------|
|            | 1                  | 56.01%          |
|            | 2                  | 60.50%          |
|            | 4                  | 60.40%          |
|            | 6                  | 72.05%          |
|            | 10                 | 76.30%          |
|            | 20                 | 80.10%          |
|            | 30                 | 83.70%          |
|            | 40                 | 87.40%          |
|            | 70                 | 90.05%          |
|            | 100                | 94.04%          |

Figure 3. The correct classification rate with different number of clusters
6. Evaluation of Algorithm model

The performance of the classification model is evaluated using the main types of metrics for evaluating classifier models, namely, geometric mean (GM) of true rates, threshold and probability. Threshold metrics quantify classification prediction errors. In this study, the relative absolute error (RAE) and the root mean squared error (RMES) are used for evaluation.

The receiver operating characteristic (ROC) curve is a ranking metric parameter and summarises the behaviour of a model by calculating the false and true positive rates. Kappa refers to a fundamental scale that measure the performance of classification system, in particular the imbalanced datasets. It shows the powerful classifier performance compared with the hypothesis of target distribution. Table 4 shows the evaluation metrics of ANN classification model. Figure 5 shows the ROC curves of classification of WBC types. GM, ROC and Kappa can be defined as the following equations.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad \text{Eq. (5)}
\]

\[
\text{Geometric mean} \quad GM = \sqrt{\frac{TP}{(TP + FP)} \cdot \frac{TN}{(FP + TN)}} \quad \text{Eq. (6)}
\]

Where;
TP= True positive, TN= True negative, FP= False positive, FN= False negative rate.

\[
\text{Kappa Statistic} \quad K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad \text{Eq. (7)}
\]

Where;
Pr (a) represents the actual observed agreement
Pr (e) represents chance agreement.

| Classes of WBCs | Accuracy% | Kappa Statistic | RMSE  | RAE   | GM   |
|-----------------|-----------|-----------------|-------|-------|------|
| Neutrophils     | 96.6      | 0.989           | 0.053 | 3.201 | 0.771|
| lymphocytes     | 94.5      | 0.930           | 0.986 | 9.0124| 0.760|
| Eosinophils     | 91.5      | 0.906           | 0.123 | 14.107| 0.899|
| Monocytes       | 93.78     | 0.927           | 0.100 | 9.257 | 0.795|

Table 2. The evaluation of classification ANN algorithm
Figure 5. The ROC curves of classification WBCs type

7. Statistical analysis

SPSS version 20 are adopted for statistical analysis. Kolmogorov–Smirnov test is used for normality distribution test research data. A Pearson’s correlation testing is conducted to evaluate whether a statistical evidence exists for a linear relevance amongst the component features, namely, the size of WBC cells, the shape and number of nuclei and the colour of cytoplasm (bluish and reddish). Table 6 shows the analysis of the relationship amongst the component features of WBCs.

Figure 6. The correlation relation between features

|                | Size WBCs | No. nucleus | Size of nucleus | Bluish color | Reddish color |
|----------------|-----------|-------------|-----------------|--------------|---------------|
| Size WBCs      | 1         | 0.637       | 0.16            | 0.681        | 0.902         |
| No. nucleus    | 0.637     | 1           | 0.884           | 0.81         | 0.913         |
| Size of nucleus| 0.16      | 0.884       | 1               | 0.188        | 0.093         |
| Bluish color   | 0.681     | 0.81        | 0.188           | 1            | 0.142         |
| Reddish color  | 0.902     | 0.913       | 0.093           | 0.142        | 1             |

8. Conclusion

Classification of WBCs is an important task in medical diagnosis because it can signal the presence of diseases, especially cancerous diseases. Each type of WBCs causes diseases that differ from the others. Therefore, pathologists need to distinguish amongst the types of WBCs to diagnose diseases. This research classifies five types of WBCs using imbalanced data samples by ANN-based deep leaning as an ML algorithm classifier. The results show that the classification task performs effectively by 10-fold cross-validation and tuning of parameters, namely, number of clusters and clustering seed. The classification accuracy is average for the five types more than 94%. The analysis of statistical features indicates a strong negative correlation relationship between the number of nuclei and the size of WBCs. A negative relationship also exists between the number of nuclei and the color of cytoplasm (bluish and reddish). The classification accuracy achieves a GM of 0.806. The obtained Kappa value 0.92 shows that the classification model can conduct multiclass classification of imbalanced class problems effectively. The RMSE 0.275 shows that the model can relatively predict the data accurately with an acceptable RAE 9.50.
References

[1] Zhang, C. T. (2018). A cost-sensitive deep belief network for imbalanced classification. IEEE transactions on neural networks and learning systems., 30(1), 109-122.

[2] Wang, S. L. (2016). Training deep neural networks on imbalanced data sets. , international joint conference on neural networks (IJCNN). , (pp. 4368-4374). IEEE.

[3] Ali, Aida, Siti Mariyam Shamsuddin, and Anca L. Ralescu (2015):. "Classification with class imbalance problem: a review." Int. J. Advance Soft Compu. Appl 7, no. 3 176-204.

[4] Korfiatis, P. K. (2019). Predictive modeling, machine learning, and statistical issues. Radiomics and Radiogenomics,Technical Basis and Clinical Applications, 151.

[5] Sun, Y. K. (2007). Cost-sensitive boosting for classification of imbalanced data. Pattern Recognition., 40(12), 3358-337.

[6] Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Tokyo: O'Reilly Media.

[7] Vanhoeyveld, J. &. (2018). Imbalanced classification in sparse and large behaviour datasets. Data Mining and Knowledge Discovery, 32(1), 25-82.

[8] Yousefi, N. A. (2019). A Comprehensive Survey on Machine Learning Techniques and User Authentication Approaches for Credit Card Fraud Detection. arXiv preprint arXiv:, 1912.02629.

[9] Lemaître, G. N. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. The Journal of Machine Learning Research., 18(1), 559-563.

[10] Fan, H. Z. (2019). LeukocyteMask: An automated localization and segmentation method for leukocyte in blood smearimages using deep neural networks. Journal of Biophotonics, 12(7):

[11] Salehinejad, H. V. (2018). Generalization of deep neural networks for chest pathology classification in x-rays using generative adversarial networks. In 2018 IEEE International Conference on Acoustics.

[12] AL-Dulaimi, K. B.-R. (2018). Classification of white blood cell types from microscope images Techniques and challenges. Last Approaches on Educational Programs and Applied Res.

[13] Yildirim, M. &. (2019). Classification of White Blood Cells by Deep Learning Methods for Diagnosing Disease . international information and engenerning technologyassocation.e201800488,https://onlinelibrary.wiley.com/doi/abs/10.1002/jbio.201800488.

[14] Negm, A. S. (2018). A decision support system for Acute Leukaemia classification based on digital microscopic images. Alexandria engineering journal, 57(4), 2319-2332.

[15] Gonzalez Viejo, C. T. (2019). Emerging Technologies Based on Artificial Intelligence to Assess the Quality and Consumer Preference of Beverages. Beverages, 5(4), 62.

[16] AL-Dulaimi, K. B.-R. (2018). Classification of white blood cell types from microscope images Techniques and challenges. Last Approaches on Educational Programs and Applied Res.

[17] Doan, C. A. ( 1954). The white blood cells in health and disease. Bulletin of the New York Academy of Medicine, 30(6), 415.

[18] Dal Pozzolo, A. C. ( 2015). Calibrating probability with undersampling for unbalanced classification. IEEE Symposium Series on Computational Intelligence, (pp. 159-166.

[19] Mollineda, R. A. (2007). The class imbalance problem in pattern classification and learning. In II Congreso Espanol de Informática (CEDI 2007), ISBN (pp. 978-84).

[20] Bauder RA, K. T. (2018). The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data. Health Inf Sci Syst, Health Inf Sci Syst., 6(1):9. https://doi.org/10.1007/s13755-018-0051-3.

[21] Kutlukeyan, M. P. (2020). Interpolative Leishman-Stained transformation invariant deep pattern classification for white blood cells. . Soft Computing., 1-11.

[22] Kutlu, H. A. (2020). White blood cells detection and classification based on regional convolutional neural networks. Medical hypotheses, 135, 109472.
[23] Sahlol, A. T. (2020). Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features. Scientific Reports, 10(1), 1-11.
[24] LCun, Y. B. (2015). Deep learning. nature. Nature, 521(7553), 436-444.
[25] Nguyen, L. D. (2018). Deep CNNs for microscopic image classification by exploiting transfer learning and feature concatenation. IEEE International Symposium on Circuits and Systems (ISCAS) , (pp. 1-5). IEEE.
[26] Das, H. N. (2020). An Experimental Analysis of Machine Learning Classification Algorithms on Biomedical Data. In Proceedings of the 2nd International Conference on Communication, Devices and Computing Springer, Singa, (pp. 525-539).
[27] Madasamy, K. &. (2017). Data Imbalance and Classifiers. Impact and Solutions from a Big Data Perspective. IJCIR,, 13, 2267-2281.
[28] Anand R, M. K. (1993). An improved algorithm for neural network classification of imbalanced training sets. IEEE Trans Neural Netw. , 4(6):962–9. https://doi.org/10.1109/72.286891.
[29] Khan, S. H. (2017). Cost-sensitive learning of deep feature representations from imbalanced data. IEEE transactions on neural networks and learning systems, 3573-3587.
[30] Ben, N. T. (2019). Demystifying parallel and distributed deep learning: An in-depth concurrency analysis. ACM Computing Surveys (CSUR), 52(4), 1-43.
[31] Johnson, J. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data. Journal of Big Data,, 6(1), 27.
[32] Polat, K. (2018). Similarity-based attribute weighting methods via clustering algorithms in the classification of imbalanced medical datasets. Neural Computing and Applications,30(3), 987-1013.
[33] Krawczyk, Bartosz. "Learning from imbalanced data: open challenges and future directions." Progress in Artificial Intelligence 5, no. 4 (2016): 221-232.
[34] Ali, Haseeb, Mohd Najib Mohd Salleh, Rohmat Saedudin, Kashif Hussain, and Muhammad Faheem Mushtaq(2019). "Imbalance class problems in data mining: A review." Indonesian Journal of Electrical Engineering and Computer Science 14, no. 10.11591.
[35] Jinyan, Li, Lian-sheng Liu, Simon Fong, Raymond K. Wong, Sabah Mohammed, Jinan Fiaidhi, Yunsick Sung, and Kelvin KL Wong(2017). "Adaptive swarm balancing algorithms for rare-event prediction in imbalanced healthcare data." PloS one 12, no. 7 e0180830.
[36] Fernández, A. L. (2013). Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. Knowledge-based systems, . , 42, 97-110.
[37] Tyagi, S. &. (2020). Sampling Approaches for Imbalanced Data Classification Problem in Machine Learning. Springer, Cham., In Proceedings of ICRIC 2019 (pp. 209-221).
[38] Thabtah, F. H. (2020). Data imbalance in classification: Experimental evaluation. Information Sciences, 513, 429-441.
[39] García, V. M. (2010). Theoretical analysis of a performance measure for imbalanced data. In 2010 20th International Conference on Pattern Recognition (pp. 617-620). IEEE., 617-620.
[40] Akrimi, J. A. (2014). Classification red blood cells using support vector machine. " In Proceedings of the 6th International Conference on Information Technology and Multimedia,E., pp. 265-269. IEEE.
[41] Kelain, M. J. (2020). Compatibility of Enhancement and Segmentation of Digital Image Processing in Medical Applications. Journal of Southwest Jiaotong University, 55.
[42] Al-Kilidar, S. H. (2020). Texture Classification Using Gradient Features with Artificial Neural Network. Journal of Southwest Jiaotong University, 55(1).
[43] Banik, P. P. (2020). An Automatic Nucleus Segmentation and CNN Model based Classification Method of White Blood Cell. Expert Systems with Applications, 149, 113211.
[44] Putzu, L. &. (2013). White blood cells identification and classification from leukemic blood image. In International Work-Conference on Bioinformatics and Biomedical Engineering , Copiceentro Editorial., 99-106.
[45] Blanco, R. A. (2020). The use of a hematology analyzer with a new generation of software as an alternative to flow cytometry for enumerating residual white blood ce. Transfusion, 60(1),, 155-164.
[46] AL-Dulaimi, K. B.-R. (2018). Classification of white blood cell types from microscope images. Techniques and challenges. Last Approaches on Educational Programs and Applied Res.

[47] J. A. Alkrimi, S. A. (2020). Using Knowledge Discovery to Enhance Classification Techniques for Detect Malaria-Infected Red Blood Cells. International Journal of Scientific Research in Network Security and Communication, Vol.8(1).

[48] Ewaid, S.H., Abed, S.A., 2017. Water quality index for Al-Gharraf river, southern Iraq. Egypt J. Aquatic Res. 43 (2), 117–122. http://dx.doi.org/10.1016/j.ejar.201703001.

[49] Ewaid, S.H.; Abed, S.A.; Al-Ansari, N. Crop Water Requirements and Irrigation Schedules for Some Major Crops in Southern Iraq. Water 2019, 11, 756.

[50] Ewaid, S.H.; Abed, S.A.; Al-Ansari, N. Water Footprint of Wheat in Iraq. Water 2019, 11, 535.

[51] Ewaid, S.H.; Abed, S.A.; Al-Ansari, N. Assessment of Main Cereal Crop Trade Impacts on Water and Land Security in Iraq. Agronomy 2020, 10, 98.

[52] Ewaid, S.H.; Abed, S.A.; Al-Ansari, N.; Salih, R.M. Development and Evaluation of a Water Quality Index for the Iraqi Rivers. Hydrology 2020, 7, 67.

[53] Ali, J. A. (2013). Red blood cell recognition using geometrical features. International Journal of Computer Science Issues (IJCSI), 10(1), 90.

[54] Ahmad, M. W. (2012). A survey: soft computing in intelligent information retrieval systems. In 2012 12th International Conference on Computational Science and Its Applications., (pp. 26-34). IEEE.

[55] Jeatrakul, P. W. (2010). Data cleaning for classification using misclassification analysis. Journal of Advanced Computational Intelligence and Intelligent Informatics, 14(3), 297-302.

[56] Hussein Al-abedi, K. J., & Abd Al-Mayahi, F. (2019). Molecular detection of metallo-β-lactamase genes in carbapenem-resistant isolates of Pseudomonas aeruginosa recovered from patients in Al-Diwaniyah province, Iraq. Al-Qadisiyah Journal Of Pure Science, 24(2)

[57] Alramathy, S. khoman, & karim, N. abdul. (2019). Evaluation of serum levels of IL-10 and IL-6 in patients with HCV at Diwaniyah Teaching Hospitals, Iraq. Al-Qadisiyah Journal Of Pure Science, 24 (2).

[58] Kareem, F., & Thamer, W. (2019). Study of some immunological effects of methotrexate in people with rheumatoid arthritis. Al-Qadisiyah Journal Of Pure Science, 24 (2).

[59] Jwad Sahar, Y., & Shamran Mohammed, H. (2019). Synthesis and characterization some complexes of azo dye of pyrimidinyl and evaluating their biological activity. Al-Qadisiyah Journal Of Pure Science, 24(3).

[60] Salah, A. (2020). The New Combination of Semi-Analytical Iterative Method and Elzaki Transform for Solving Some Korteweg-de Vries Equations. Al-Qadisiyah Journal Of Pure Science, 25(1), Math. 23 -26.

[61] Ali , W., & R.Anon, M. (2020). Biological Effective of organic solvent extracts of Mirabilis jalapa Leaves in the Non-cumulative for mortality of Immature stages Culex quinquefasciatus Say ( Diptera : Culicidae ). Al-Qadisiyah Journal Of Pure Science, 25(1), Bio 1-6.

[62] Ewaid, S.H., Al-Farhani, B.F., Abed, S.A. & Al-Ansari, N. (2020). Modeling of trihalomethane compounds formation in Baghdad water supply network. Scientific Review Engineering and Environmental Sciences, 29 (2), 136-144. doi: 10.22630/PNIKS.2020.29.2.12.

[63] Salam Hussein Ewaid et al 2020 J. Phys.: Conf. Ser. 1664 012143.

[64] Sami Abd ali , mohammed, Shaker Hussein, A., & mohammed hadi, H. (2020). Study The Current Density-Voltage (J-V) Characteristics of α-Fe2O3 Thin Film Prepared by Spray Pyrolysis Technique. Al-Qadisiyah Journal Of Pure Science, 25 (1), Phys 1-7.