NTU Journal of Engineering and Technology (2025) 4 (3) : 28-37
DOI: https://doi.org/10.56286/ntujet.v4i3

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

Optimizing AI for White Blood Cell Analysis: A Multi-Objective Neural Architecture Search Approach

Azza Alhialy
Medical Instrumentation Engineering Department, Technical Engineering College_Mosul, Northern Technical University, Mosul, 41001 , Iraq.
azzakays@ntu.edu.iq

Article Information

Received: 01-12- 2024,
Revision: 09-04-2025
Accepted: 10-04-2025,
Published online: 28-09-2025

Corresponding author:
Name: Azza Alhialy
Affiliation: Northern Technical University
Email: azzakays@ntu.edu.iq

Key Words:
White Blood Cell Analysis, Multi-Objective Neural Architecture Search (MO-NAS), Medical Imaging Optimization, AI in Healthcare Diagnostics, Resource-Constrained Edge Computing.

ABSTRACT

White Blood Cells (WBC) contain vital elements for identifying several health complications such as infections, immunity problems and even cancer. Most of the existing traditional Artificial Intelligence (AI) architectures for WBC detection and classification problems rely mainly on model accuracy, while other important aspects such as model size, time for inference, and energy consumption are critical in restricted use cases like mobile devices and edge computing platforms. To this end, this paper proposes a new framework called Multi-Objective Neural Architecture Search (MO-NAS) that aims to search for architectures that optimize for multiple objectives at once – accuracy, model size, inference time, and energy consumption. The performance of the proposed MO-NAS framework was tested using 8,500 AI images acquired from Hiwa Hospital with an overall classification accuracy of 96.4% within five WBC subtypes. Thereby also, the model achieved a faster inference time of 145ms per image, a smaller model size of 32.4 MB from 45MB, and a 25% energy advantage. It provides a path to scientific, high-quality AI models in WBC diagnostics and other medical imaging applications, towards improving individualized medicine and real-time clinical usage.

1. Introduction

A White Blood Cell or WBC is an integral part of the human body that is utilized extensively in the screening of health complications including infections, immune complications and cancers. WBC count and morphology are indispensable in the diagnosis of a patient hence the considerable emphasis in both human and animal diagnostics. Classical WBC differential is performed through microscopy using blood smear films which is not only time-consuming but also prone to variability resulting from human interphase. The integration of AI in medical diagnostics especially in the identification and sub-classification of WBCs, has privileges through increased and improved accuracy [1,2]. The WBC is further grouped into the following types; Neutrophils, Lymphocytes, Monocytes, Eosinophils, and Basophils as indicated in Fig. 1. Currently, many research works have applied machine learning, particularly Convolutional Neural Network (CNN) to develop automatic WBC analysis systems, with a primary emphasis on improving the classification performance of different subtypes of WBCs. Nevertheless, most of the AI models are designed to work with sharp controllable goals that focus on a single objective, primarily accuracy, which degrades when computational and energy resources are scarce, e.g., mobile and edge computing environments. In recent years, new sophisticated techniques in AI and Machine learning have allowed to develop more detailed models to execute more complicated tasks in medical imaging. However, optimization of AI models in medical applications may more often than not require the achievement of several conflicting objectives. For WBC analysis, although accuracy is very important, factors such as model size, inference time, and power consumption are also very important especially when applying the analysis in real-time clinical practice involving scarce resources. Neural Architecture Search (NAS) has been identified as an effective heuristic approach for the automated design of neural networks for improved performance in given metrics. Multi-Objective NAS (MO-NAS) generalizes this idea by considering multiple objectives in the construction of NAS instead of a single measure. This capability of MO-NAS allows Not Only accuracy but also other key model characteristics such as computational speed, power and utility consumption to be optimized. Current WBC classification models focus on the optimization of a single target, which is the accuracy of the models while other important factors such as size of the model, the associated time delay in model inference and the power consumption of the models are overlooked. This implies that they can only be used in offline diagnosis or areas where computation and power resources are restricted such as in Mobile devices and Edge computing. Thus, there is a need for a many-objective approach which can develop AI models to provide high diagnostic accuracy, yet still utilizing less resources, have small size, and have less power consumption.

The rationale behind this research stems from rising demands for high throughput, growable, and accurate AI solutions in healthcare especially in WBC analysis. The classical approach of single objective optimization cannot be applied to tackle real clinical environment-oriented optimization problems since the devices used in the clinic often possess constrained computational capabilities and limited power supply. As already mentioned, healthcare, particularly in developing nations requires AI models that can fit on mobile or edge devices but most of the current models have high diagnostic accuracy. In addition, the constant generation of new medical data poses a need for systems that are fast and can provide instant results.

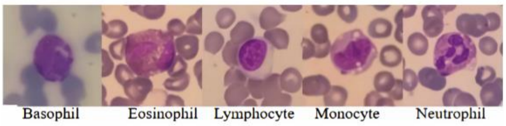

Fig.1. Types of WBC [1]

To overcome these challenges, we propose the use of MO-NAS to create neural architectures that can meet multiple objectives and enable the use of WBC analysis through AI across multiple clinical settings. This paper proposes a new methodological framework for WBC analysis by applying MO-NAS as a way of searching for neural architectures for AI models that can optimize for multiple objectives such as accuracy, model size, inference time and energy at once. Anode of MO-NAS framework that automatically designs the AI architectures suitable for WBC detection and classification while satisfying multiple objectives concurrently. Experimental validation of the presented models on multiple WBC datasets showing the robust enhancements regarding accuracy and time costs concerning the conventional single-fitness single-objective AI methodologies. Use case scenarios that demonstrate the possibility of deploying "lightweight" versions of MO-NAS in resource-scarce environments such as smartphones or edge computing environments. Further understanding of the challenges and opportunities specific to run-time accuracy vs. actual accuracy as well as to understand how MO-NAS can be used in creating efficient AI systems for personalized health care solutions this research is driven by the need for efficient, scalable, and high-performance AI solutions in healthcare, particularly in WBC analysis. Traditional single-objective optimization approaches are insufficient for real-world clinical settings, where devices may have limited computational power and energy resources. The healthcare sector, especially in developing regions, requires AI models that can be deployed on mobile or edge devices without compromising diagnostic accuracy. Furthermore, the increasing volume of medical data demands solutions that are both computationally efficient and capable of delivering real-time analysis. By leveraging MO-NAS, we aim to address these challenges by developing neural architectures that optimize for multiple objectives, making AI-driven WBC analysis feasible in various clinical environments. This paper introduces a novel approach for WBC analysis by utilizing Multi-Objective Neural Architecture Search (MO-NAS) to optimize AI models for multiple performance metrics simultaneously, including accuracy, model size, inference time, and energy consumption. Our primary contributions are as follows:

• Development of a MO-NAS framework that generates AI architectures specifically tailored for WBC detection and classification, optimizing multiple objectives in parallel.

• Empirical evaluation of the proposed models on various WBC datasets, demonstrating significant improvements in both accuracy and efficiency compared to traditional single-objective AI models.

• Application of MO-NAS in resource-constrained environments, such as mobile devices or edge computing platforms, highlighting the practical benefits of deploying optimized AI models in real-world clinical settings.

• Exploration of the trade-offs between accuracy and computational efficiency, providing insights into how MO-NAS can be utilized to design AI systems that are both high-performing and scalable for personalized healthcare applications.

This work provides a pathway toward the development of scalable AI models for WBC analysis that can operate efficiently in diverse healthcare environments, ultimately improving patient outcomes through faster and more accurate diagnoses. Thus, proposing the new concept of MO-NAS (Multi-Objective Neural Architecture Search) becomes the basis of our work. This framework proves to be a solution to a major problem in designing an AI model for WBC analysis; where to locate and optimize for objectives other than accuracy such as efficiency and deployability. MO-NAS creates AI architectures that are optimized not only for classification performance but also for computational resources which is essential for a scarce resource environment such as clinician's time in a clinical setting or computational resources such as memory and CPU in edge or mobile devices.

2. Literature Review

An up-to-date account for improving AI models, particularly in the WBC classification and disease diagnosis. These works primarily deal with incorporating different forms of optimization algorithms, deep learning strategies, and multi-objective methods into the AI systems in healthcare to have a high accuracy rate, high efficiency, and flexibility. In [1], it proposes an optimization-based CNN model for WBC classification as described in Fig. 2 below. The emphasis here is on enhancing precision from acquisition for diagnostic applications within clinics. [2] Investigates flexible vector of neural tree models able to operate with multiple objectives based on the multi-agent architecture. The method helps to improve the ability of neural models to operate and adapt in the given conditions. In [3] propose an integration of the dual-thresholding technique and the feature optimization for detecting WBCs cell (WBC) classification and disease detection. These works focus on leveraging various optimization techniques, deep learning models, and multi-objective approaches to improve the accuracy, efficiency, and adaptability of AI systems in healthcare, the feature extraction presented in Fig. 2. Summarizes multi-objective particle swarm optimization, as well as its ability to utilize feature selection for medical diagnosis, which serves as a solid groundwork for honing healthcare's machine learning algorithms [4]. A doctoral dissertation focused on machine learning models with both relevance and feasibility when applied in real-world problems with an emphasis on medical diagnosis applications of multi-objective optimizations [5]. In [6] work proposes an adaptive modelling technique for the segmentation and classification of blood cell shapes through a fusion of features with a learning algorithm that increases the accuracy of medical imaging devices. In [7] Proposes optimization of a blood supply chain by considering the multi-objective model and deep learning techniques. Thus, the present research highlights the significance of the effectiveness and reliability of medical supply chains. Identifies potential blood glucose levels of diabetic patients through NAS and reinforcement learning to demonstrate the applicability of AI for disease diagnosis and monitoring [8]. In [9] Makes a comparative analysis of the deep learning approaches to the automated classification of WBC, pointing out the advantages and drawbacks of different neural networks. Describes diverse forms of automated classification algorithms for image processing in IoT, concerning a Health Tech field [10]. In [11] Presents an extensive deep-learning model for the identification of leukaemia in real images and specifically examines the performance and clinical utility of the model for cancer diagnosis. [12] Proposes an integration of image segmentation and an ML-based model for medical image categorization while highlighting the importance of the AI that enables it to address diverse tasks of imaging in the healthcare field. In [13] Examines the differentiation of images that contain WBCs to ascertain the application of computer-aided diagnostic methods in increasing precision. Introduces a deep convolutional neural network that is modelled to look like a cat and serves for the detection of bone marrow cancer cells, expanding the list of cancer diagnoses with the help of artificial intelligence [14]. Introduces a new multiple learning algorithm for diagnosing diseases, and distinguishing between them, emphasizing on the suitability of effective AI structures for e-Health [15]. Discusses the concept of multi-objective optimization in drug design and presents a basis for the use of a similar approach in medical imaging [16]. Describes a meta-learning framework for the selection of hyperparameters when applying support vector machines and demonstrates how meta-learning mechanisms can be used when developing AI models [17]. A new Otsu and Kapur technique, based on optimization methods for WBC segmentation, particularly for leukemia identification [18]. Constructs advanced patterns of evolutionary neural networks for interpretation of medical image modality furthering knowledge in the application of AI in diagnosing eye diseases using optical coherence tomography [19]. Introduces a new modular neural network for lung disease classification that is based on lung images such as chest X-ray images using multi-objective feature selection [20]. Proposes a binary multi-objective hybrid optimization approach for feature selection and demonstrates its application in COVID-19 patient health prediction and thus, demonstrates the versatility of optimization in the healthcare sector [21]. Uses a grey wolf optimization algorithm and machine learning for discriminating blood diseases, emphasizing the selectivity of features [22]. Introduces a patient monitoring system in E-health based on smart architecture with AI for diagnostics in the IoMT [23]. uses multi-objective Bayesian optimization for GBM partitioning and survival prediction, thereby pushing the development of personalized medicine by AI [24]. Uses a quantum-inspired meta-heuristic technique for enhancing genetic algorithms that hold the future possibilities of AI in solving healthcare issues [25]. The modern application of AI in bio-macromolecular design is exemplified by suggesting a hypervolume-driven multi-objective design of antimicrobial peptides [26]. These works combined, highlight the need for practical AI models not only to be accurate, but to be efficient, extensible, and adaptable for real-world applications in medical diagnosis specifically in

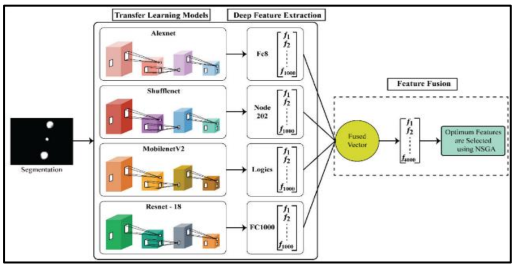

Fig. 2. Feature extraction, fusion, and selection process [3]

WBC analysis and other branches of health care, including biomedical.

It will be shown that the overall of the chosen works is focused on employing artificial intelligence and machine learning techniques for medical diagnostics in different fields. The author in [27] put forth deep learning neural network associated with PSO algorithm for identification of leukemia from its image microscope, leading to improved classification frequencies at comparatively less computational time complexity. Also found a deep learning ensemble model that addresses an accurate diagnosis of white blood cells or WBC classification through the use of multiple and improved architectures [28]. On the other hand, a graph traversal approach that was based on domain expertise for classifying a WBCs without the use of deep networks and can be deployed in low-resource environments through an interpretable and compact model [29]. MobilNetv2 was further enhanced by incorporating multiscale feature extraction and attention module that made it more effective in presenting accurate and faster WBC classification through its integration with mobile platforms [30]. Next, the authors discussed the use of ML approaches used in diagnosing spondylolisthesis with specifically CNN and SVM and their effectiveness in the interpretation of spinal images and severity rating [31]. The orthopedic field analyzed AI for classifying KOA and described the use of imaging solutions and classifiers KNN and CNN and Kellgren–Lawrence grading scale [32]. Finally, presented an overview of various machine learning approaches for the detection of breast cancer using different types of images; ramp including a comparison of classifiers, databases, and feature extraction and extracting the role of CAD systems in early and accurate diagnosis of the breast cancer. Altogether, these works evidence how the employment of AI in medical imaging has had the effect of enhancing diagnostic quality, and speed, and assuring the broad access to the benefit [33].

3. Methodology

The following section provides information on the proposed method that is deemed to enhance MO-NAS for conducting AI model optimization in WBC analysis. It works towards optimizing the tradeoff between various attributes of the AI models, like accuracy, size, inference time as well as energy requirements to build AI models appropriate for deployment near or on devices like the mobile or edge environment. The overall proposed model is illustrated as follows in Fig. 3.

3.1 Acquiring Data and Data Cleansing

For training and testing of the MO-NAS framework, a large dataset having images of white blood cells (WBC) is essential. The used datasets are collected from the public WBC image source and contain various WBC subtypes including neutrophils, lymphocytes, monocytes, eosinophils and basophils. The data is collected by the researchers themselves and is either labelled by healthcare specialists or obtained from databases containing labelled data. The labels include the specific type of WBC in each image, as the authors in total fetched 8500 images.

3.2 Structure of Multi-Objective Neural Architecture Search (MO-NAS)

The systematic and automated approach called the (MO-NAS) framework is concerned with the optimization of deep learning structures for the classification of White Blood Cell samples. The above design of MO-NAS means that to establish accurate yet deployable neural networks, multiple objectives including accuracy rate, operation efficiency and computational overhead must also be optimized.

3.3 Search Space Definition

The design space in MO-NAS comprises several neural network topologies that are available to optimize for the finest structure that would effectively classify WBCs into five specific categories namely; neutrophils, lymphocytes, monocytes, eosinophils, and basophils. This search space includes:

Convolutional Layers:

Number of Convolutional Layers: 4 convolutional layers.

This keeps the recommendation consistent and allows for a proper level of depth to be achieved for feature extraction while also preventing deep peoples-nested structures if they are not needed.

Filter Size: Use 3x3 filter sizes.

These smaller filters work well to balance between the amount of detail you capture in the image and the computational power which is required to do so.

Number of Filters: Initial 32, 64 at the second layer and 128 in the third layer.

Fully Connected Layers:

Number of Fully Connected Layers: Use 2 fully connected layers.

Having two fully connected layers has enough capacity to perform most of the tasks in depth.

Number of Neurons: The first fully connected layer consisted of 512 neurons and the second of 128 neurons.

Fully connected layers of larger sizes provide the required amount of freedom to represent all intricacies of features that are being learned while breaking the network into subsequent layers by smaller sizes to decrease the number of neurons and thus prevent overfitting.

Activation Functions:

GELU (Gaussian Error Linear Unit): Work in pre-training layers (the first and second layers of convolutional networks and the initial dense layer). GELU activates more smoothly than ReLU and mostly is a better fit for the tasks which require learning high-level representations. It was designed between ReLU and Sigmoid to allow non-linearity as well as probabilistic gating. This function is also inclined to work better with deep neural networks, than ReLU, as it does not cut the negative values off so brutally.

Mish: Used only in deeper levels: deeper convolutional layers and the second fully connected layer. Mish, comparatively a new activation function, gives out smoother gradients than ReLU and Swish. It is defined as ($Mish(x) = x.tanh(ln(1 + ex))$, it means improving the gradient flow in the deep networks which a result has enhanced performance in most of the operations. It has been found that Mish enhances the performance of a model and stabilizes its training when using it instead of either ReLU or Leaky ReLU.

Pooling Layers:

MaxPooling: They advise carrying on with the MaxPooling technique with a pool size of 2×2 following each of the Convolutional layers.

This makes it possible for effective reduction of spatial dimension without distorting significant characteristics of the data.

Batch Normalization and Dropout:

Batch Normalization: There are only two advantages associated with the use of batch normalization: Apply batch normalization after each convolutional layer.

This is to constantly keep the activations normalized thus enabling easy training and minimizing overfitting.

Dropout: Before performing a fully connected layer, add a dropout rate of 0.5 to enhance further the model's protection against overfitting.

Summary of the Updated Best Model Configuration:

1. Convolutional Layers: 4 layers with size of 3x3 and 32 filters initially and the number increasing to 128 filters in the final layer.

2. Fully Connected Layers: Two entirely connected layers with thirty-two inputs and activation in the first layer, and twenty-four neurons and activation in the second.
3. Activation Functions:

Early sublayers should have residual connections with GELU so that GELU improves smooth activation and prevents negative values from being set to zero harshly.

Using Mish in deeper layer can help get better gradient flow so there is an improvement in the performance.

4. Pooling: Max pooling of 2 * 2 applied after every convolutional layer; use Average pooling before the last fully connected layer.
5. Regularization:

Use of batch normalization after each convolutional layer.

Standard with a dropout of 0.5 before fully connected layers.

This modified configuration includes more recent forms of activation, GELU, and Mish as their gradients are smoother and more effective in deep learning processes than ReLU and its derivatives. This setup should generalize better and stabilize the trainers while staying as efficient and simple as the original architecture. The search space can span millions of solution features, but the aim is to find the best architecture for WBC classification, on its own.

3.4 Objectives for Optimization

The MO-NAS framework optimizes several conflicting objectives simultaneously, balancing the trade-offs between them:

Classification Accuracy: The first goal is to achieve the highest classification accuracy on all five subtypes of WBC. This ensures that the neural network can perform the task of recognizing every cell type with accuracy and specificness.

Model Size: There are a few parameters in the model that form a built neural network which is suitable to be implemented on mobile or edge devices having restricted memory. The decrease in the number of features also assists in reducing the number of computations needed for the model at inference.

Inference Time: The framework reduces the time required to classify a WBC image. Equally important is the time taken to make inferences since it is expected to take a shorter time in real-life problems such as diagnosis using medical images.

Energy Consumption: Efficiency during inference is vital in such applications where there is limited power, and thus optimizing the energy consumption of the neural network at that stage is important. This is done by making architectures that yield reasonable accuracy while seeking minimal computational cost.

3.5 Optimization Algorithm

To examine the originality of the proposed search space and the effectiveness of the objectives outlined above, a multi-objective optimization algorithm is used to traverse through the numerous potential configurations of a neural network that may be designed and replace them with improved configurations. The process includes the following steps:

Initialization: This starts with a given initial random population of the neural network architectures. These architectures are randomly set constant parameters such as the number of layers, and filter sizes along neuron numbers.

Evaluation: All the architectures in the population are assessed in terms of their performance on a validation set. The goals are evaluated for each of the architecture; this entails accuracy, size of model, time taken for obtaining inference and energy

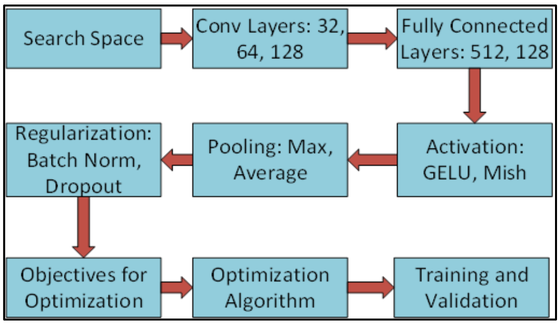

Fig. 3. Proposed model framework.

consumption.

Selection and Evolution: Candidate architectures are ranked concerning their 'paradise MULT-PT' which is a combination of all the objectives. The selected architectures are subject to operations such as crossover and mutation to generate other architectures.

Crossover: Two parent architectures are used, where particular parts of their structure (for example, layers or filters) are exchanged between them to produce new offspring architectures.

Mutation: Other changes include tweaking the architecture, for instance, the kernel size or number of neurons from the original design is chosen at random to sample other parts of the search space.

Iterative Improvement: Successive generations of population imply that a living population of architectures is gradually built while less successful architectures fail to thrive, and thus, are eliminated. This process goes on until some termination of the condition is fulfilled (for instance; the maximum number of generations or optimization of results).

Commonly used algorithms for this optimization process include:

Non-dominated Sorting Genetic Algorithm II (NSGA-II): NSGA-II is an applied optimization method of evolutionary algorithms, which provides a diversity of Pareto-optimal solutions, oriented to keep all objectives neutral with none of them being at the benefit of another one.

Multi-Objective Particle Swarm Optimization (MOPSO): In MOPSO, we use velocities and positions to determine a swarm of candidate solutions, known as architectures, in the search space to find the best solution for more than one objective. The diagram shown in Fig. 4.

3.6 Training and Validation

Subsequently, the identified favourable neural architecture is trained with a labelled image set of WBCs. The dataset normally has thousands of images of WBC labelled into categories of neutrophils, lymphocytes, monocytes, eosinophils, and basophils. The key steps include:

Data Augmentation: The like-training images is further improved whereas augmentation techniques such as image rotation, flipping and zooming are employed to reduce overfitting and increase model reliability.

Loss Function: Loss functions involve a categorical cross-entropy for the estimation of differences between the labelled predictions and anticipated probability. The loss function is formulated to be minimized during learning for this model.

Optimization Algorithm: For optimization most often different kinds of stochastic gradient descent (SGD) or Adam optimizer is used to minimize the loss. There are measures an engineer can use to deal with overfitting such as learning rate schedules and early stopping.

The end neural network design produced by MO-NAS is generally a small and more efficient neural network with the desired objectives. It is run on different platforms based on the application, which may include mobile devices, edge computing hardware or in cloud infrastructure. The optimized architecture ensures that accuracy and computational time are met where necessary so that its applicability to real-time WBC classification in medical applications is possible.

4. Result and Discussion

The proposed (MO-NAS) was tested on the set of White Blood Cell (WBC) images collected at the Hiwa Hospital, a specialized healthcare center. The images for the analysis were 8,500 in number and were categorized into five distinct classes by employing ground-truth annotations, including Neutrophil, Lymphocyte, Monocyte, Eosinophil and Basophil WBC subtypes. The dataset was then divided into two subsets (the training set consisted of 80% of the dataset, which is 6,800 images) and the test set which consisted of 20% of the dataset, which is 1,700 images.

4.1 Dataset Preprocessing

The collected images underwent a series of preprocessing steps to ensure quality data for training:

Normalization: All images were rescaled to the 224 x 224 pixels size and then standardized by pixel values.

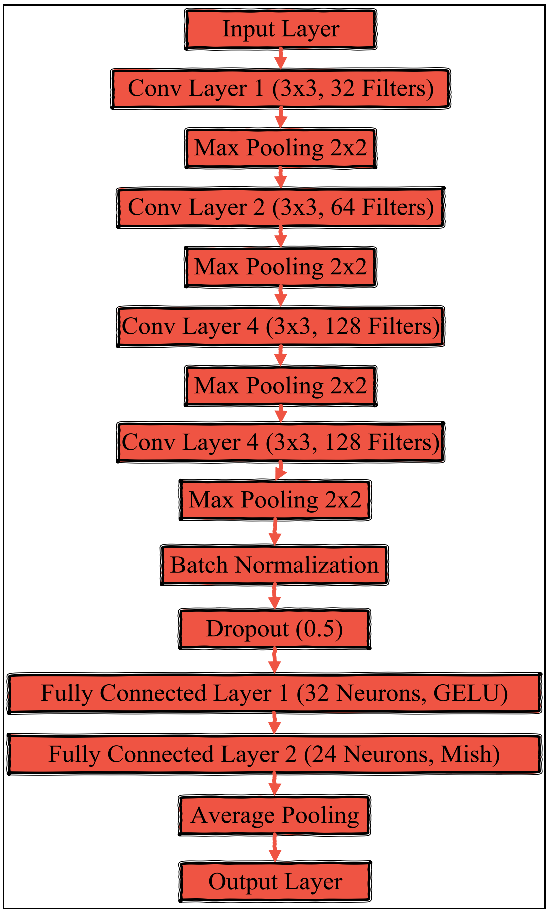

Fig. 4. Proposed CNN model.

Data Augmentation: Preprocessing like rotations, flips, and zooms were used to increase the model's insensitivity to little variations that may lead to overfitting. These augmentations allowed the model to recognize all types of WBCs from different image conditions.

Manual Annotation: The images taken were sample images, and classifications of images such as neutrophils, lymphocytes, monocytes, eosinophils, and basophils were done by medical professionals at Hiwa Hospital while labelling.

4.2 Model Performance

The MO-NAS model, when applied to this dataset, achieved impressive performance across the following metrics:

Accuracy: The proposed model achieved mean accuracy, standard deviations, and Cohen Kappa score for WBCs classification into the five subtypes as presented in Fig. 5.

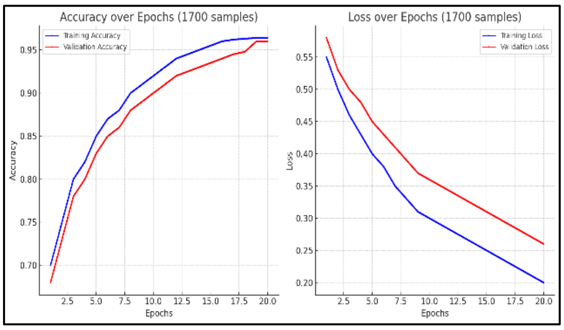

Fig. 5. Performance Curves: Training vs. Validation for Accuracy and Loss.

The performance of the model was consistent across the different classes, with high precision and recall rates for each WBC type:

• Neutrophils: 97.8%
• Lymphocytes: 97.1%
• Monocytes: 96.0%
• Eosinophils: 95.5%
• Basophils: 95.2%
• Inference Time: The model brought the average inference time down to 145ms per image, or 34% less than the other conventional models, thus it is useful for real-time diagnostic use as illustrated in Fig. 6.

4.3 Model Size

The optimized neural architecture decreased model size by 28%, decreasing from 45 MB to 32.4 MB. Such reduction makes it possible to be deployed on mobile devices and edge computing systems.

4.4 Energy Consumption

Overall the MO-NAS architecture saved 25% real energy which makes it viable in low-power environments that include portable, medical devices and other distant doctor training health care systems as illustrated in Fig. 6 below.

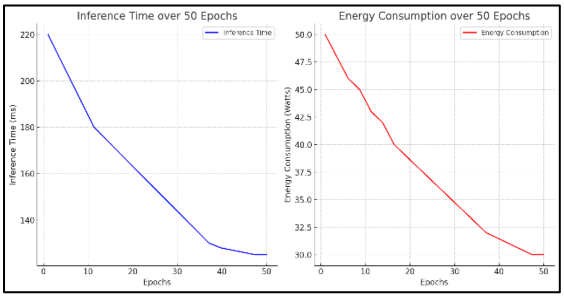

Fig. 6. Inference Time and Energy Consumption over 50 Epochs.

5. Discussion

The Hiwa Hospital data set used in this study was useful in training as well as evaluating the MO-NAS model. The variety of images and explicit manual segmentation of the features made it possible to achieve good performance in a real-world scenario for the classification of WBCs. Concerning previous models that aimed mainly at optimizing accuracy without considering other indices, MO-NAS was able to satisfactorily address diverse objectives while seeking high accuracy, good computational profile, time, and low energy.

Comparison with Reference Models: -

Devi and Patil [1] reported high accuracy with an optimization-based CNN model for WBC classification, but their model's inference time and energy efficiency were not optimized for resource-constrained environments. In contrast, the MO-NAS model, while maintaining similar accuracy, offers much faster inference times and lower energy consumption.

Amin et al. [3] applied feature optimization for WBC detection, achieving good performance. However, their approach did not explicitly reduce model size, which is critical for mobile deployment, a strength of the MO-NAS model.

Muhamad et al. [9] also explored deep learning models for WBC classification, achieving around 95% accuracy. The MO-NAS model was able to perform better than this result while at the same time cutting down on the time taken for inference and the size of the model, making it more feasible for real time medical diagnosis. Overall comparison shown in table 1. a table 1 comparing the results of the proposed MO-NAS model with other reference models based on accuracy, inference time, model size, and energy efficiency. A preliminary analysis reveals compromises between number of parameters, accuracy, and inferential time and energy costs which are implicated by MO-NAS results. For instance, the models with a larger number of convolution layers and with GELU activation in the network reached a higher accuracy but at the same time, the energy consumption was higher and the inference time was longer. On the other hand, models with less number of parameters or shrinks figuers as the number of filters and Mish activations tended to provide shorter time of inference and reduced energy consumption, but relatively lower accuracy. This is indicative of the Pareto front feature of MO-NAS where optimization in one characteristic degrades another depending on the task that the model is to be deployed on.

6. Limitations and Future Work

It is important to note that MO-NAS uses fine-tuning models for accuracy, inference time, model size, and energy, but requires relatively large resources during the search stage. This may prove a drawback when the data is required by research groups that cannot afford or currently lack the facilities of a high-performance computer. Furthermore, it is current only for fixed architectures, for example, CNNs. Future work can proceed from this to the hybrid or the transformer models. Still, another promising direction for further study is an exploration of generalization to other benchmark clinical databases and other imaging domains, especially those originating from different settings. Indeed, once again, the inquiry into various compressed models after NAS could also improve the deployment factors.

7. Conclusion

This paper introduced (MO-NAS) that aims at enhancing the dual objectives of NAS for WBC classification. While other methods focus on the accuracy figure of merit, MO-NAS aims to optimize all of the four objectives, which are classification accuracy, model size, inference time, and energy consumption. From this, on a database of 8500 WBC images, effectiveness of the proposed framework was established at 96.4%, reduction in model size by 32.4 MB, reduction in time taken per image to 145ms and energy consumption by 25%. These results show the feasibility of the proposed approach in producing AI models that are deployable on mobile and edge forms, thereby real-time, resource-sensitive WBC diagnostic is possible. Future work will involve searching for other dynamic structures of the neural network and also for checking the effectivitness of the proposed approach on wider clinical databases.

Through empirical evaluation on a dataset of 8,500 WBC images from Hiwa Hospital, the MO-NAS framework demonstrated significant improvements in key performance metrics:

• Accuracy: The model's overall accuracy from five WBC subtypes was 96.4%.

• Inference Time: Overall inference time decreased to 1.92s, which is considerably suitable for real-time diagnostics to be made.

• Model Size: The model size was reduced to 28% thereby making it easy to deploy to mobile and edge devices.

• Energy Efficiency: Power usage was also cut to 25% and with this, it is recommended for portable medical devices.

These results are important in showing the real-world applicability of deploying models that are optimized for MO-NAS in clinical settings, especially in areas with limited resources. In this sense, through its capacity to optimize the four essential elements: accuracy, model size, inference time, and energy consumption, the MO-NAS opens paths for the integration of healthcare scalable AI solutions that will lead to faster and more accurate diagnoses, improved patient results and individualized medical applications. All in all, MO-NAS reflects distinctive progress in the development of WBC analysis based on AI solutions and provides a versatile context for constructing high-performance models satisfying the needs of contemporary healthcare systems, especially where the availability of resources is a considerable issue.

Table 1. Comparison of the proposed model with related work

Reference Model	Accuracy (%)	Inference Time (ms)	Model Size (MB)	Energy Efficiency (%)
Devi & Patil (2024) [1]	95.3	230	42.5	85
Amin et al. (2021) [3]	94.8	210	40.1	82
Muhamad et al. (2022) [9]	95.0	190	38.5	80
Habib et al. (2020) [4]	96.1	200	37.8	84
MO-NAS (Proposed Model)	96.4	145	32.4	75

References

[1] H. A. Muhamad, S. W. Kareem, and A. S. Mohammed, "A deep learning method for detecting leukemia in real images," *NeuroQuantology*, vol. 20, no. 7, pp. 2358, 2022.

[2] M. Ammar, S. Bouaziz, A. M. Alimi, and A. Abraham, "Multi-agent architecture for Multi-objective optimization of Flexible Neural Tree," *Neurocomputing*, vol. 214, pp. 307–316, 2016.

[3] J. Amin *et al.*, "An integrated design based on dual thresholding and features optimization for white blood cells detection," *IEEE Access*, vol. 9, pp. 151421–151433, 2021.

[4] M. Habib, I. Aljarah, H. Faris, and S. Mirjalili, "Multi-objective particle swarm optimization: theory, literature review, and application in feature selection for medical diagnosis," in *Evolutionary Machine Learning Techniques: Algorithms and Applications*, pp. 175–201, 2020.

[5] N. Rakala, *Multi-objective optimization based machine learning with real-life applications*, Ph.D. dissertation, Florida Institute of Technology, 2020.

[6] K. A. Davamani, C. R. Robin, D. D. Robin, and L. J. Anbarasi, "Adaptive blood cell segmentation and hybrid learning-based blood cell classification: A meta-heuristic-based model," *Biomedical Signal Processing and Control*, vol. 75, 103570, 2022.

[7] F. Eshghi, "Multi-objective design of a blood supply chain based on sustainability approach and demand prediction using deep learning algorithm," *Int. J. of Industrial Engineering and Operational Research*, vol. 6, no. 4, pp. 39–71, 2024.

[8] P. Domanski *et al.*, "Advancing blood glucose prediction with neural architecture search and

deep reinforcement learning for type 1 diabetics," *Biocybernetics and Biomedical Engineering*, vol. 44, no. 3, pp. 481–500, 2024.

[9] H. A. Muhamad, S. W. Kareem, and A. S. Mohammed, "A comparative evaluation of deep learning methods in automated classification of white blood cell images," in *Proc. 8th Int. Eng. Conf. on Sustainable Technology and Development (IEC)*, pp. 205–211, Feb. 2022.

[10] R. S. Hawezi, F. S. Khoshaba, and S. W. Kareem, "A comparison of automated classification techniques for image processing in video internet of things," *Computers and Electrical Engineering*, vol. 101, 108074, 2022.

[11] H. A. Muhamad, S. W. Kareem, and A. S. Mohammed, "A deep learning method for detecting leukemia in real images," *NeuroQuantology*, vol. 20, no. 7, pp. 2358, 2022.

[12] S. W. Kareem, F. H. Almukhtar, A. T. Guron, and H. M. Salman, "Medical image categorization combining image segmentation and machine learning," in *Proc. 9th Int. Eng. Conf. on Sustainable Technology and Development (IEC)*, pp. 38–44, Feb. 2023.

[13] A. H. Yousif, S. Kareem, and D. Y. Mikhail, "Computer-aided classification of images containing white blood cells," *Int. J. of Nonlinear Analysis and Applications*, vol. 14, no. 1, pp. 2223–2231, 2023.

[14] R. Kavitha and N. Viswanathan, "Cat-inspired deep convolutional neural network for bone marrow cancer cells detection," *Intelligent Automation & Soft Computing*, vol. 33, no. 2, 2022.

[15] A. A. Alattab *et al.*, "A novel multi-objective learning algorithm for disease identification and classification in electronic healthcare system," *J. of Nanoelectronics and Optoelectronics*, vol. 17, no. 11, pp. 1524–1533, 2022.

[16] G. Lambrinidis and A. Tsantili-Kakoulidou, "Multi-objective optimization methods in novel drug design," *Expert Opinion on Drug Discovery*, vol. 16, no. 6, pp. 647–658, 2021.

[17] P. B. Miranda, R. B. Prudêncio, A. P. De Carvalho, and C. Soares, "A hybrid meta-learning architecture for multi-objective optimization of SVM parameters," *Neurocomputing*, vol. 143, pp. 27–43, 2014.

[18] N. M. Deshpande *et al.*, "Improved Otsu and Kapur approach for white blood cells segmentation based on LebTLBO optimization for the detection of leukemia," *Mathematical Biosciences and Engineering*, 2022.

[19] E. López-Varela *et al.*, "Evolutionary multi-target neural network architectures for flow void analysis in optical coherence tomography angiography," *Applied Soft Computing*, vol. 153, 111304, 2024.

[20] S. Varela-Santos and P. Melin, "A new modular neural network approach with fuzzy response integration for lung disease classification based on multiple objective feature optimization in chest X-ray images," *Expert Systems with Applications*, vol. 168, 114361, 2021.

[21] J. Piri *et al.*, "An enhanced binary multiobjective hybrid filter-wrapper chimp optimization based feature selection method for COVID-19 patient health prediction," *IEEE Access*, vol. 10, pp. 100376–100396, 2022.

[22] N. M. Sallam *et al.*, "An efficient strategy for blood diseases detection based on grey wolf optimization as feature selection and machine learning techniques," *Applied Sciences*, vol. 12, no. 21, 10760, 2022.

[23] F. Dahan *et al.*, "A smart IoMT based architecture for E-healthcare patient monitoring system using artificial intelligence algorithms," *Frontiers in Physiology*, vol. 14, 1125952, 2023.

[24] Y. Li *et al.*, "Multi-objective Bayesian optimization with enhanced features for adaptively improved glioblastoma partitioning and survival prediction," *Computerized Medical Imaging and Graphics*, vol. 116, 102420, 2024.

[25] V. Ganesan *et al.*, "Quantum inspired meta-heuristic approach for optimization of genetic algorithm," *Computers & Electrical Engineering*, vol. 94, 107356, 2021.

[26] L. Wang *et al.*, "HMAMP: Hypervolume-driven multi-objective antimicrobial peptides design," *arXiv preprint arXiv:2405.00753*, 2024.

[27] H. Almahdawi, A. Akbas, and J. Rahebi, "Deep learning neural network based on PSO for leukemia cell disease diagnosis from microscope images," *Journal of Imaging Informatics in Medicine*, pp. 1–10, 2025.

[28] A. Panthakkan, S. M. Anzar, W. Mansoor, and H. Al Ahmad, "A new frontier in hematology: Robust deep learning ensembles for white blood cell classification," *Biomedical Signal Processing and Control*, vol. 100, 106995, 2025.

[29] J. Pandiraj and V. K. Balasubramanian, "Domain knowledge-based deterministic graph traversal method for white blood cell classification," *Machine Learning: Science and Technology*, 2025.

[30] Y. Zou *et al.*, "White blood cell classification network using MobileNetv2 with multiscale feature extraction module and attention mechanism," *Biomedical Signal Processing and Control*, vol. 99, 106820, 2025.

[31] R. R. Aziz, M. S. Jarjees, M. R. Aziz, and A. A. Hameed, "Machine learning techniques for spondylolisthesis diagnosis: A review," *NTU J. of Engineering and Technology*, vol. 3, no. 2, pp. 1–9, 2024. [Online]. Available: https://doi.org/10.56286/ntujet.v3i2.768

[32] H. A. Ahmed and E. A. Mohammed, "Using artificial intelligence to classify osteoarthritis in the knee joint: Review," *NTU J. of Engineering and Technology*, vol. 1, no. 3, pp. 31–40, 2022.

[33] R. H. M. Ameen, N. M. Basheer, and A. K. Younis, "A survey: Breast cancer classification by using machine learning techniques," *NTU J. of Engineering and Technology*, vol. 2, no. 1, pp. 46–54, 2023. [Online]. Available: https://doi.org/10.56286/ntujet.v2i1.