1. INTRODUCTION

The human body needs blood to transport oxygen and other nutrients. Blood, a red body fluid, is formed by respiratory proteins containing iron and hemoglobin. In addition, blood is divided into different types, each performing certain bodily functions. In addition, blood is grouped into different types, each performing a specific task for the body. White blood cells are one of the body's most important types of blood cells[1]. White blood cells, also known as leukocytes, are important to the body's defense system and fight infectious microorganisms, tumor cells, and dangerous foreign substances. Leukocytes consist of monocytes, lymphocytes, basophils, eosinophils, and neutrophils[2].

This paper suggests a computer-based automated system for detecting white blood cells using transformer analysis and white blood cell YOLOv8 in digital images of blood cells. Analyzing white blood cells using digital images requires only peripheral blood sample images, so they are not subjective and, therefore, faster and more accurat[3]. The machine learning component known as deep learning has developed into a powerful tool for classifying images[4]. Convolutional neural networks (CNNs) and transformer-based object identification models may be divided into categories based on the model's structure. Transformer employs self-attentive methods, whereas CNN extracts and processes information via convolutional operations. There are two types of object detection networks: one-stage and two-stage, with benefits and drawbacks.[5]. In stage one, use You Only Look on v8 (YOLOv8). They use the sigmoid function as an activation function for object scores, which indicates the possibility that the bounding box contains objects. They also use the SoftMax function for class probabilities, which indicates the possibility that objects belong to every possible class[6]. In natural language processing, transformer-based models have been widely used.

Vision transformation mode (ViT)[7]. In natural language processing, transformer-based models have been widely used. Google's proposed vision transformer (ViT) model demonstrates the prospect of transformer applications in the field and begins in-depth research on transformers. For accurate and promising detection,transformer-based network backbones such as Swin transformers can be combined with classic CNN detection. Transformers submitted by Facebook based on end-to-end object detection network[8], [9]. In this study, the authors performed image processing of white blood cells using a roboflow platform using the YOLOv8 model. Roboflow is a platform for annotating and labeling and can produce results using the same format as COCO[10]. A large-scale image dataset with high complexity, Common Objects in Text Dataset (COCO) displays complex scene images with many small objects and footnotes with very detailed outlines. COCO was also designed to enable the study of interaction matters[11]. There are research questions in this study that are relevant to the research.

Why should DETR use the COCO dataset?(RQ1)

What factors can affect the detection results of yolov8 with roboflow tools?

The research questions will be addressed later in the discussion section.