I. INTRODUCTION

Eye contribute about 80% of the brain's knowledge and memory through vision. Maintaining eye health is essential to ensure an optimal quality of life and sustain daily productivity [1]. According to the World Health Organization (WHO), there are 2.2 billion global eye disease cases, with one billion still preventable or treatable. Eye diseases can cause social isolation, depression, and delays in motor, language, emotional, social, and cognitive development. Patients with eye diseases spend over 40% of their income on eye care, leading to severe financial strain [2].

One of the most vital components within the structure of eye is retina. Diameter, density, shape, and tortuosity of retinal blood vessels serve as indicators for diagnosing eye diseases, including glaucoma and cataract [3]. Blood vessels in the retina form a network that supplies nutrients to various parts of eye and can reflect circulatory dynamics and overall health status [4], [5], [6], [7]. Various eye diseases can lead to deformities and bleeding in the retinal blood vessels [8], [9]. Retinal blood vessels can be observed non-invasively with imaging technologies like fundus image segmentation, essential for diagnosing ocular diseases. However, challenges persist, including limited vessel edge visibility [3] and the complex structure of retinal blood vessels, complicating the segmentation process [10]. Retinal blood vessels often have noise, low contrast, and multi-scale structures [11].

Retinal vessel segmentation can be conducted using various approaches, including manual and algorithm-based methods [10]. Manual segmentation is performed by ophthalmologists through manual annotation [11] to mark the dynamics of the blood vessels [10]. This process is costly, time-consuming, labor-intensive, and depends heavily on ophthalmologists' experience [10], [11]. Algorithm-based methods like threshold segmentation, matched filter, and vessel tracking have successfully segmented retinal blood vessels but rely heavily on pixel intensity and vessel morphology, with filters and parameters often manually designed [11].

Deep Learning (DL) offers an alternative to address segmentation challenges and has become the primary method in medical image segmentation due to advancements in DL and related technologies [12], [13], [14]. In recent years, Deep Convolutional Neural Networks (DCNN) have been used as a reliable architecture for image recognition tasks, particularly for classification [15], [16]. However, in many visual tasks, particularly in biomedical image processing, the desired output typically involves the localization of class labels assigned to each pixel. [17].

Several studies related to the development of DCNN for retinal vessel segmentation have been extensively conducted. In 2015, Ronneberger et al. combined CNN and Fully Convolutional Network (FCN) to create the U-Net model, which enhances segmentation performance with localization. U-Net has been successfully implemented, achieving an F1-score of 81.55% [17]. In 2018, Alom et al. proposed two models, namely Recurrent U-Net (RU-Net) and Recurrent Residual U-Net (R2U-Net), by adding recurrent and/or residual layers to the U-Net model. These models successfully performed retinal vessel segmentation with superior performance compared to U-Net. With the same number of parameters, the F1-scores for each model were achieved at 81.80% and 81.87%, respectively [18]. In 2018, Oktay et al. proposed the Attention U-Net model by adding attention gates to the U-Net's skip connections, achieving an F1-score of 80.03% [19]. In 2021, Zuo et al. proposed the R2AU-Net model by combining the Attention U-Net and R2U-Net models, achieving an F1-score of 82.13% [20].

Researchers have also extensively developed encoder and decoder blocks to enhance the performance of U-Net-based architectures. In 2023, Ryu et al. proposed an encoder block consisting of a Feature Extraction and Embedding (FEE) block and a Deep Feature Magnification (DFM) block, which can attribute dynamic morphology and amplify thin retinal vessels. Additionally, they introduced the Feature Precision and Interference Block (FPI) and the Denser Multiscale Feature Fusion Block (DMFF) as decoder blocks to aid in reconstructing images through precise spatial features, achieving an F1-score of 80.97% [10]. In 2024, Liu et al. proposed an encoder block consisting of a specialized convolution block and also created a decoder block that combines with Attention Pooling Fusion (APF), introducing the architecture as IMFF-Net, which achieved an F1-score of 79.77% [21]. In 2024, Ma et al. proposed an encoder block called the Decoder Fusion Module (DFM), which enhances feature extraction and reduces loss caused by convolutional operations. The Context Squeeze and Excitation Module (CSE) was introduced as a decoder block to avoid loss in thin retinal vessels, achieving an F1-score of 82.98% [22].

Many researchers have developed new encoder and decoder blocks and integrated them into U-Net-based architectures, yet there remains uncertainty in identifying the best encoder and decoder for each level. Meta-heuristic approaches offer innovative strategies for identifying optimal encoder and decoder configurations within U-Net-based architectures. One of meta-heuristic algorithm that can be employed to find the optimal encoder and decoder is Komodo Mlipir Algorithm (KMA) [23]. KMA method is chosen due to its high exploitation movement through the exploration of male komodos, which helps avoid getting trapped in local optima [23]. IMFF-Net was chosen because it demonstrates superior performance compared to conventional methods and U-Net in terms of the F1-score, sensitivity, specificity, accuracy, and AUC metrics [20]. Therefore, this study proposes the implementation of KMA to optimize the type combination of encoder-decoder.