Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 3084~3097
e-ISSN: 2548-6861

Evaluating LSB and MSB Steganography in Retinal Fundus Images Through Image Quality Assessment and VGG19-Based Classification

Gilang Faturrahman [1], Muhammad Naufal [2*], Wahyu Aji Eko Prabowo [3], Sindhu Rakasiwi [4]
[1,3,4]Distance Learning Study Program in Informatics, Dian Nuswantoro University
[2,3,4]Informatics Engineering, Dian Nuswantoro University
118202300036@mhs.dinus.ac.id [1], m.naufal@dsn.dinus.ac.id [2], prabowo@dsn.dinus.ac.id [3], sindhu.rakasiwi@dsn.dinus.ac.id [4]

Article Info

Article history:

Received 2026-06-04
Revised 2026-06-19
Accepted 2026-06-22

Keyword:

Fundus Retina,
LSB,
MSB,
Steganography,
VGG19,

ABSTRACT

The security of medical image data within electronic medical record systems has become a critical issue due to the increasing threat of health data breaches. Steganography is a promising technique for protecting patient information by concealing secret data within medical images without significantly altering their visual appearance. However, the application of steganography to retinal fundus images, which carry high diagnostic value, has never been comprehensively evaluated in terms of image quality or its impact on artificial intelligence-based diagnostic model performance. This study compares Least Significant Bit (LSB) and Most Significant Bit (MSB) steganography methods applied to 3,200 retinal fundus images from the Retinal Fundus Multi-disease Image Dataset (RFMiD) dataset across four payload levels (0.1-0.4 bpp), evaluated using PSNR, SNR, SSIM, and FSIM for image quality, and VGG19 classification accuracy and AUC for diagnostic impact. Results show LSB achieves substantially superior image quality (PSNR: 59.97-65.93 dB; SNR: 49.34-55.30 dB; SSIM: 0.9981-0.9997; FSIM: 0.9999-1.0000) compared to MSB (PSNR: 12.98-18.99 dB; SNR: 2.35-8.37 dB; SSIM: 0.5979-0.9003; FSIM: 0.5342-0.7500), while VGG19 classification accuracy remains stable for both methods (LSB: 0.8938-0.9000; MSB: 0.8953-0.9031) with a maximum difference of 0.62% from baseline. This study demonstrates that LSB is the more appropriate steganography method for retinal fundus images, delivering superior visual quality while preserving VGG19 diagnostic capability.

I. INTRODUCTION

The global health sector is in the midst of digital transformation, marked by a shift from manual record-keeping to computerized Electronic Medical Records (EMR) infrastructure. In Indonesia, this transformation is mandated through the Ministry of Health Regulation Number 24 of 2022, requiring all healthcare institutions to transition to electronic medical record management to protect the three pillars of health information security: confidentiality, integrity, and availability of patient data [1]. This transition expands the attack surface for cybercriminals targeting health data, with The World Health Organization (WHO) noting a consistent year-on-year escalation in medical data breaches, particularly during the transmission and storage phases [2]. Indonesia's experience reflects this risk directly: documented incidents include unlawful access to 230 Covid-19 patients' records in 2020 and the circulation of 720 gigabytes of hospital data on illegal forums in 2022 [1]. Since medical records contain history, diagnostic results, and identity that can be exploited for fraud, discrimination, or extortion if breached, protecting medical data during digital transmission is a security imperative that is no longer optional.

Medical imaging is not a category of data that can be treated the same as ordinary text documents or administrative records. It is a direct visual window into the condition of a patient's organs that supports the entire chain of clinical decision-making. In the context of ophthalmology, no

modality is more informative yet simultaneously more demanding in terms of integrity than retinal fundus images. The depiction of blood vessels, the optic disc, and nerve fiber layers captured by a fundus camera can serve as early visual evidence of conditions such as diabetic retinopathy, glaucoma, or macular degeneration. These conditions, if detected late, can lead to permanent blindness [3], [4]. As a result, if the pixels carrying diagnostic information are disturbed during digital transmission, a clinician might read an image that no longer faithfully represents the patient's original condition without realizing it. This threat becomes even more real when retinal fundus images must be digitally transferred between service facilities within the EMR ecosystem over networks that are just as vulnerable as the networks that have been proven to leak in previously recorded incidents [1], [2]. This situation creates a very specific need. A protection mechanism that not only keeps the patient's identity contained in the image confidential but also must not touch or alter the diagnostic pixel values in the slightest. Fitriyasari's systematic review of medical image security techniques identifies steganography's distinguishing advantage over cryptography in precisely this context: while cryptography secures content, it cannot conceal the fact that a communication is occurring and remains vulnerable to brute-force attacks on unsecured networks [2]. Watermarking, a related technique sometimes proposed for medical image protection, instead embeds ownership or integrity markers into the image to support authentication and tamper detection, a goal oriented toward verifying that an image has not been altered rather than concealing data within it [5]. Access control mechanisms, meanwhile, govern who may retrieve an image but offer no protection once a file leaves the controlled system, the precise scenario realized in Indonesia's documented data breaches. Steganography occupies a distinct niche relative to both: it conceals the very existence of embedded patient data within the image itself, a property neither watermarking nor access control provides, making it a complementary rather than redundant layer of protection for images already secured by access policies. Ramyashree et al. further demonstrated that steganography particularly when applied to medical image formats such as Digital Imaging and Communications in Medicine (DICOM) can embed patient data without introducing extraction errors or compromising the diagnostic metadata that clinicians depend upon [6]. However, the unanswered question is: among the many steganography methods developed, which is truly safe to apply to retinal fundus images without compromising diagnostic quality standards? And that is what drives the need for a systematic comparative study based on empirical evidence.

Several studies have attempted to map steganographic performance, though with limited focus and scope. Among spatial-domain approaches, LSB dominates published research: Rahman et al. found their substitution method outperformed the closest competing technique by roughly 5.5% on signal-fidelity measures across varied image types including RGB, grayscale, textured, and aerial photography, a result consistent with the principle that altering only the least-weighted bit keeps pixel shifts too small for the eye to register [7]. A related refinement by Rustad et al. let the embedding pattern be chosen adaptively rather than fixed in advance, yielding measurably better fidelity than standard LSB, though how reliably this holds across diverse image content remains an open question [8]. Al-Faydi et al. advanced this further with a cover-stego matching mechanism that leaves unnecessary pixels untouched [9]. On the MSB side, a matching strategy proposed by Ali et al. held visual fidelity above a 37 dB threshold on ordinary digital photographs, yet its behavior on medical content was never examined [10]. Ilham and Kirana identified a measurable capacity-fidelity trade-off across pixel-based variants, limited to generic grayscale images [11]. Closest to a clinical setting, Ramyashree et al. tested DICOM-format images and recorded PSNR above 47 dB with SSIM near 0.99 across various compression levels, yet stopped short of checking whether embedding altered an AI model's diagnostic output [6]. This body of work reveals a consistent gap across three dimensions: none of these studies evaluated diagnostically sensitive images such as retinal fundus photographs [7], [8], [9], [11]; none compared LSB and MSB simultaneously on the same medical dataset across multiple payload levels; and even the closest clinical attempt, Ramyashree et al. [6], measured only image quality without assessing AI-based diagnostic impact. This unoccupied space, the joint evaluation of LSB and MSB on retinal fundus images together with their downstream effect on deep learning diagnostic performance, motivates the use of VGG19, a 19-layer convolutional architecture from Oxford's Visual Geometry Group, as the evaluation instrument in this study.

Selecting an evaluation instrument with an established empirical record on retinal fundus images was a deliberate methodological decision, and VGG19 satisfies that criterion. The track record of VGG19 in the domain of retinal fundus images has been documented in several independent studies. Li et al. integrated VGG19 into a multi-label ensemble model framework on the Retinal Fundus Multi-disease Image Dataset (RFMiD), the same dataset used in this study, and noted that VGG19's contribution to the system resulted in an overall accuracy of 97% in the complex retinal fundus disease classification task, confirming its capability to extract rich feature representations from fundus images [3]. Raiaan et al. directly evaluated VGG19 in a diabetic retinopathy classification setting on the Asia Pacific Tele-Ophthalmology Society (APTOS) dataset, Messidor2, and the Indian Diabetic Retinopathy Image Dataset (IDRiD), with hyperparameter settings standardized across all tested models. VGG19 achieved an accuracy of 88.21%, demonstrating consistent, reliable performance as a reference point for evaluation [12]. Meanwhile, Aswathi et al. confirmed VGG19's position as a competitive baseline for diabetic retinopathy classification on the Messidor dataset, where it delivered performance comparable to other architectures under default conditions

without additional parameter adjustments [4]. Saputra et al. further reinforced the belief that the transfer learning paradigm as a whole is a proven effective approach for multi-class retinal disease classification from fundus images, with the transfer learning-based model they developed successfully distinguishing four retinal conditions accurately with an accuracy rate of 95.38% [13]. The convergence of findings from these independent studies provides strong empirical justification for using VGG19 as an appropriate, validated, and reliable proxy to measure the diagnostic impact of steganography on retinal fundus images in this study.

Although VGG19 has been shown to be reliable as a classification model for retinal fundus images, and various studies have separately explored the performance of LSB and MSB, a comprehensive review of the existing literature reveals an unanswered question: no study has simultaneously compared LSB and MSB across two evaluation dimensions. The visual quality of stego images and their impact on the diagnostic ability of artificial intelligence (AI) models on clinically validated multi-disease retinal fundus image datasets [2], [6]. Previous studies evaluated the visual quality of steganography without measuring its consequences on AI model accuracy, or conversely, tested classification models without considering the impact of steganography on the integrity of the images they used as input. Although several studies have used the RFMiD dataset to test deep learning-based retinal disease classification models with various architectural approaches [14], [15], [16], [17], none have evaluated the impact of steganography as an analytical dimension. It is this gap that directly motivates this research. The fundamental unanswered question is: which method, LSB or MSB, is safer to apply to retinal fundus images without compromising the diagnostic capability of deep learning-based classification models? This study is designed to answer that question through a comparative evaluation encompassing two dimensions simultaneously: the visual quality of stego images using PSNR, Signal-to-Noise Ratio (SNR), SSIM metrics, Feature Similarity Index (FSIM) metrics and the diagnostic impact, measured by the VGG19 model's performance on the RFMiD dataset. The core contribution of this research is to provide the first empirical evidence of a direct simultaneous comparison between LSB and MSB steganography on retinal fundus images, evaluated across both visual fidelity and diagnostic integrity dimensions using a validated deep learning classification model. As established in the literature review above, prior studies addressed these dimensions in isolation: image quality was evaluated without diagnostic impact measurement, or classification performance was tested without accounting for steganographic distortion of the input images [6], [7], [11]. No prior study has combined both dimensions on retinal fundus images specifically, making this study the first to empirically characterize the joint effect of spatial-domain steganographic embedding on image quality and VGG19 diagnostic capability within a multi-disease ophthalmological dataset. These findings are expected to serve as a scientific foundation for EMR system developers in determining the most appropriate steganography method for the protection of ophthalmic medical images, in accordance with the mandate of the Minister of Health Regulation Number 24 of 2022 on Electronic Medical Records [1] and Law Number 27 of 2023 concerning Personal Data Protection.

II. METHOD

This study uses a purposively designed two-arm comparative experimental design to evaluate and contrast LSB and MSB steganography methods on retinal fundus images, with assessments covering two orthogonal dimensions: the extent to which the visual fidelity of the stego images can be maintained, and the degree to which the embedding process affects the classification capability of VGG19 as a proxy for diagnostic impact. The overall research methodology is divided into five structured stages: (1) dataset preparation and preprocessing, (2) implementation of LSB and MSB steganography embedding at four different payload levels, (3) evaluation of stego image quality, (4) quantification of diagnostic impact based on VGG19, and (5) cross-method comparative analysis. The full implementation of this study was carried out using Python 3.10, with Graphics Processing Unit (GPU)-accelerated computation via the Google Colaboratory (Google Colab) platform [18].

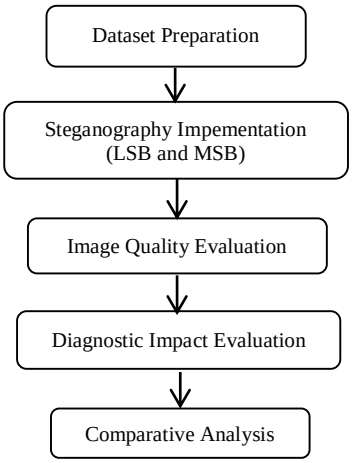

Figure 1. Research Framework

A. Dataset Preparation

This study uses RFMiD, a publicly available, clinically verified dataset of colored retinal fundus images officially published by Pachade et al. [19]. This dataset is widely recognized in the literature as a reference for multi-label retinal disease classification tasks based on deep learning [3]. To ensure the reproducibility of experimental results as well as comparability with previous studies using similar datasets, this study adopts the data split officially established by the RFMiD provider: 1,920 images are allocated as training data, while 640 images each are used as validation and test data,

making a total of 3,200 images distributed across the three partitions. Furthermore, each image is standardized to a 224×224-pixel resolution before being fed into the VGG19 model, in accordance with the architecture's input resolution specifications [3].

The class label assignment in this study is sourced from the Disease_Risk column in the official RFMiD Comma-Separated Values (CSV) dataset, with 0 indicating the Normal class and 1 indicating the Abnormal class. This two-class binarization scheme was chosen because it is the most clinically meaningful formulation, allowing the model to definitively distinguish retinal fundus images that are free of disease signs from images that show at least one clinical condition from all pathological categories covered in the dataset annotations [3], [4]. In the test subset, the class distribution was 134 Normal images and 506 Abnormal images. This distribution asymmetry accurately reflects the class imbalance frequently observed in real-world ophthalmological clinical imaging datasets [4].

B. Steganography Implementation

The steganography embedding in this study was carried out entirely in the spatial domain by applying two methods in parallel, namely LSB and MSB, each tested at four payload levels: 0.1 bpp, 0.2 bpp, 0.3 bpp, and 0.4 bpp. Belagali and Udupi [20] define payload capacity as the maximum volume of secret data embeddable within a cover image measured in bits per pixel (bpp) and note that as embedded volume increases, the magnitude of pixel-value deviation in the cover image grows proportionally, directly affecting stego image quality. Mohsin and Alameen [21] treat this capacity dimension as one of three foundational evaluation criteria in image steganography, alongside imperceptibility and robustness, and use it as the primary optimization target in their edge-area-based method. Mohsin and Alameen [21] formalize the three-factor evaluation framework for image steganography: imperceptibility quantified through PSNR measures whether embedding-induced pixel changes escape human visual detection; robustness measures the payload's ability to survive steganalytic attacks or image processing operations; and capacity defines the ceiling on embeddable data volume. Belagali and Udupi [20] further emphasize that these three properties exist in inherent tension, as increased payload capacity inevitably exerts upward pressure on pixel deviation, a trade-off that is particularly consequential in the retinal fundus imaging context of this study, where even minor pixel disturbance carries diagnostic implications. The spatial domain is prioritized as the implementation domain, given that domain-based methods inherently yield higher imperceptibility values and can accommodate a larger embedding volume than frequency-domain techniques [21].

The selection of four payload levels, 0.1 to 0.4 bpp, reflects a deliberate balance between sufficient embedding capacity for realistic patient data volumes and image quality within clinically acceptable limits. Belagali and Udupi [20] establish that increased payload volume produces a proportional rise in pixel-value deviation, while Mohsin and Alameen [21], position capacity as one of three foundational steganographic criteria alongside imperceptibility and robustness, where gains in one dimension come at the expense of the others. Given this tension, 0.1 bpp was set as the lower bound, sufficient to carry compact patient data such as an identifier or short clinical annotation, while 0.4 bpp was set as the upper bound following Ramyashree et al.'s [22] use of bpp as the governing metric for acceptable degradation, ensuring both methods remain within their native single-bit-per-channel mechanism without requiring multi-bit modification strategies that would alter the embedding approach itself.

Cumulatively, all steganography experiments in this study cover eight different embedding scenarios resulting from the combination of two methods and four payload levels, each applied uniformly to the entire 3,200 images in the dataset.

The secret payload embedded in all experimental scenarios consists of randomly generated data, with the specific form differing between the two methods. For LSB, the payload takes the form of a randomly generated ASCII string, with character count determined by $[(224 \times 224 \times \text{bpp}) / 8]$ bytes, embedded across all three RGB channels via sequential substitution. For MSB, the payload consists of randomly generated binary bits, with total bit count computed as $224 \times 224 \times 3 \times \text{bpp}$, reflecting the number of pixel channel positions modified across the full RGB structure. In both cases, generation is performed without a fixed seed, so payload content varies across images and runs. This non-semantic design is methodologically appropriate, since the study evaluates the effect of embedding on image quality and VGG19 performance, outcomes determined by the proportion and position of bits modified rather than by the semantic content of the embedded data or its recovery accuracy at the receiver side.

Rahman et al. [23] establish that LSB substitution's core operating principle replacing the lowest-weight bit position in each pixel channel with payload bits functions precisely because it minimizes the embedding error rate. In their validation across 165 RGB images of varying dimensions and sizes of hidden information, LSB's multi-channel architecture (red, green, blue each contributing one embeddable bit per pixel) consistently demonstrated imperceptibility preservation across all tested quality metrics including PSNR, SSIM, Mean Squared Error (MSE), and Normalized Cross-Correlation (NCC). The arithmetic basis for this behavior is that modifying a single bit at position zero of an 8-bit channel shifts the pixel's decimal intensity by at most ±1 a perturbation of approximately 0.00002% of the representable range remaining reliably below the human visual system's detection boundary [7]. These characteristics make LSB the most widely used method in medical image steganography, supported by its computationally simple architecture, adequate embedding capacity, and its ability to consistently produce high imperceptibility values [2], [6]. Karawia stated that the design of LSB-based steganography algorithms in

medical images must meet one fundamental requirement. The entire embedding process must not cause changes that affect a specialist doctor's ability to make clinical decisions based on the images [24]. Meanwhile, Rustad et al. demonstrated that the inverted LSB variant with an adaptive pattern selection mechanism significantly improved imperceptibility, although the stability of this approach's performance across various image types remains a variable that needs to be examined more systematically [8].

MSB embedding reverses this logic entirely. Inserting payload bits into the highest-weight position forces immediate, large-magnitude intensity shifts; a pixel at maximum intensity (255; binary 11111111) becomes 127 (01111111) upon a single MSB flip, a reduction exceeding half the representable range. Unlike LSB perturbations which are subthreshold, this class of modification produces artifacts that are visually evident to the unaided eye, independent of any steganalysis instrument [7], [10]. The payload level in this study is quantified in units of bpp, where each bpp value represents the proportion of bits modified relative to the total bits available in the image, so applying a payload of 0.1 bpp concretely means that 10% of the total image bits are used as the space for secret data insertion [20].

Data extraction for both methods follows the inverse of the embedding procedure, retrieving the secret payload by reading the same bit positions used during insertion. For LSB, the lowest-weight bit of each modified pixel channel is read sequentially across the RGB structure to reconstruct the original payload bitstream, exploiting the fact that no other bit position is altered during embedding. For MSB, the highest-weight bit of each modified channel is read in the same sequential order. Since both methods modify a fixed, predetermined bit position rather than a content-dependent one, extraction requires no additional key beyond knowledge of the bpp level and embedding order used during insertion, and was verified in this study by confirming bit-for-bit correspondence between the extracted payload and the original randomly generated payload for every stego image produced.

C. Image Quality Evaluation

The assessment of stego image quality in this study is based on four objective metrics that have been tested and consistently used in steganography research [6], [11], [21]. Each resulting stego image is compared with its original cover image. The validity of using PSNR and SSIM as evaluation metrics in this study is supported by Sanjalawe et al., who, in their study of deep learning-based layered steganography, utilized MSE, PSNR, and SSIM as quantitative assessment tools, reflecting the position of PSNR and SSIM as widely recognized benchmarks in the steganography research community for assessing image fidelity and visual integrity [25].

PSNR serves as the primary quantitative indicator of imperceptibility, measuring the ratio between maximum possible pixel signal intensity and the noise power introduced by embedding, a role it plays in prior cover-stego fidelity analysis [9] and in studies contrasting spatial- and frequency-domain steganographic performance [11], expressed through equation (1):

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) dB \quad (1)$$

In the context of equation (1), MAX represents the maximum intensity that can be represented by each pixel, which is 255 for an 8-bit image format. Meanwhile, MSE reflects the magnitude of the mean squared error arising from differences in pixel values between the original image before embedding and the stego image after embedding. From an interpretative perspective, an increase in PSNR indicates a reduction in distortion introduced by the embedding process, thereby better preserving the visual fidelity of the stego image relative to the original. Consistent with the threshold applied by Ilham and Kirana [11] and independently corroborated by Mohsin and Alameen [21], PSNR values below 35 dB in this study are treated as indicative of unacceptable embedding distortion.

SNR serves as a complementary metric that provides a different perspective from PSNR, by quantifying the ratio between the total accumulated signal energy in the original image and the total accumulated noise energy introduced throughout the embedding process [11], as formulated in the following equation (2):

$$SNR = 10\log_{10}\left(\frac{\sum I^2}{\sum(I - I_{stego})^2}\right) dB \quad (2)$$

In equation (2), *I* denotes the intensity value of each pixel in the original image before embedding, while I_stego denotes the intensity value of the corresponding pixel in the stego image after the embedding operation has been completed. From an interpretive perspective, a higher SNR indicates a smaller proportion of noise energy relative to the original image signal's total energy, thereby preserving the fidelity of the stego image signal. Referencing the operational floor established by Ilham and Kirana [11] and corroborated in Naufal et al. [26], SNR values below 30 dB are flagged in this study as falling outside the range of acceptable embedding fidelity.

SSIM emerges as a complementary metric to PSNR, designed to overcome the limitations of purely pixel-based evaluation, by quantifying the level of similarity between the stego image and the original cover image through three dimensions of visual perception simultaneously: luminance, contrast, and structural content [11], [23], as expressed by equation (3):

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

In equation (3), μ_x and μ_y represent the mean intensity values of the original image x and the stego image y, respectively. σ_x² and σ_y² denote the variances of images x and y, while σ_xy denotes the covariance between the two images. C_1 and C_2 are stabilization constants introduced to prevent numerical instability when the denominator approaches zero, defined as $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$, where L is the dynamic range of pixel values (255 for 8-bit images) and $K_1 = 0.01$, $K_2 = 0.03$ [9]. The SSIM value ranges from -1 to 1, where 1 represents perfect identity between the two images, while values approaching -1 indicate increasing structural divergence. Following the convention adopted by Ilham and Kirana [11], SSIM values below 0.90 are treated in this analysis as evidence that structural fidelity has been compromised beyond the acceptable boundary.

A fourth metric, the FSIM, complements the pixel-level and structural assessments of PSNR, SNR, and SSIM with a feature-based perspective, extending structural similarity assessment by additionally accounting for gradient-based features for a more comprehensive quality measure [27]. FSIM combines two feature maps: phase congruency, capturing structural features such as edges independent of contrast variation, and gradient magnitude, capturing local contrast information, as expressed by equation (4):

$$FSIM = \frac{\sum \Omega\, SL(x).PCm(x)}{\sum \Omega\, PCm(x)} \quad (4)$$

In equation 4, SL(x) denotes the combined local similarity at spatial location x, while PCm(x) denotes the maximum phase congruency between the two images at that location, weighting regions with greater structural significance more heavily, and Ω denotes the entire spatial domain [26]. The FSIM value ranges from 0 to 1, where values approaching 1 indicate close feature-level agreement between the stego and cover images, while lower values indicate greater divergence introduced by embedding.

D. Diagnostic Impact Evaluation

This sub-section addresses the second evaluation dimension: the extent to which steganographic embedding affects VGG19's diagnostic performance, quantified through five confusion-matrix-derived metrics, followed by the model architecture and training configuration.

The performance of the VGG19 classification model in this study was evaluated using five quantitative metrics derived from the confusion matrix of test-set predictions [28], [29]. The confusion matrix records four classification outcome conditions: True Positive (TP), which is an Abnormal image correctly identified, True Negative (TN), which is a Normal image correctly identified, False Positive (FP), which is a Normal image incorrectly classified as Abnormal, and False Negative (FN), which is an Abnormal image incorrectly classified as Normal.

Accuracy is the ratio of the number of correct predictions to the total number of samples evaluated, formulated in equation (5) [13], [19], [28], [29]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision measures the accuracy of a model in providing positive labels, that is, the proportion of predictions labeled as Abnormal that are actually Abnormal, as expressed in equation (6) [13], [19], [28]:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Recall or sensitivity reflects the model's ability to capture all actual positive cases in the test data, as shown in equation (7) [13], [19], [28], [29]:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

The F1-Score combines precision and recall into a single value through their harmonic mean, thus providing an overview of the balance between the two, especially in conditions of imbalanced class distribution, as stated in equation (8) [13], [19], [28], [29]:

$$F1 - Score = 2\,x\,\frac{Precision\ x\ Recall}{Precision + Recall} \quad (8)$$

The Area Under the ROC Curve (AUC-ROC) quantifies a model's overall ability to separate Normal and Abnormal classes, regardless of the classification threshold. An AUC value close to 1.0 reflects excellent discrimination ability, whereas a value of 0.5 corresponds to random prediction, as stated in equation (9) [12], [29]:

$$\int_0^1 TPRd(FPR) \quad (9)$$

Where True Positive Rate (TPR) = Recall and False Positive Rate (FPR) = FP / (FP + TN).

These five metrics collectively form the quantitative framework through which the diagnostic impact of LSB and MSB steganography is operationalized in this study, applied to the output of the VGG19 model described in the following section.

The diagnostic impact was measured in this study using the VGG19 architecture, pre-trained on ImageNet, and applying the transfer learning paradigm [4]. The selection of VGG19 is based on its consistent track record in delivering reliable performance on medical image classification tasks, as well as its well-established position as a benchmark architecture in retinal fundus image classification research

[3], [12]. The transfer learning approach is prioritized because empirical evidence consistently demonstrates its superiority over training from scratch, particularly in medical image classification, where training data volume is relatively limited [4]. Raiaan et al. specifically reported that VGG19 demonstrated competitive classification performance on retinal fundus images, achieving a test accuracy of 88.21% on the combined dataset [12]. No data augmentation was applied during training, as the objective of this study is to isolate the diagnostic effect of steganographic embedding itself, and introducing augmentation-induced pixel transformations could confound the comparison between baseline, LSB, and MSB conditions. During implementation, all convolutional layers of VGG19 were frozen, so the pre-trained ImageNet weights did not update during training. As a replacement for VGG19's original fully connected layers, a custom classification head was added, consisting of a Global Average Pooling layer to reduce feature dimensions, followed by a Dense layer of 256 units with Retrified Linear Unit (ReLU) activation and a Dropout layer (rate = 0.5), then a Dense layer of 128 units with ReLU activation and a Dropout layer (rate = 0.3), and ending with a single-unit Dense layer with sigmoid activation to accommodate the binary classification task of Normal and Abnormal. The model architecture used is illustrated in Figure 2.

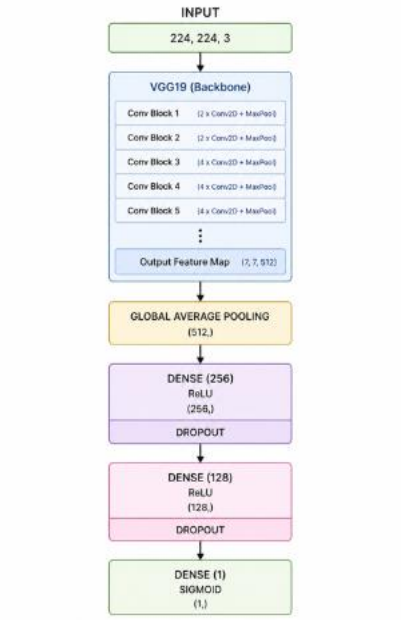

Figure 2. VGG19 Model Architecture.

This model has 20,188,737 parameters, totaling 77.01 MB, of which 164,353 (642 KB) are trainable, and 20,024,384 (76.39 MB) are non-trainable. The model training hyperparameter configuration is set as follows: batch size of 32, Binary Cross-Entropy loss function, Adam optimizer with a learning rate of 1×10^{-4}, and an early stopping mechanism with a patience value of 10 epochs to prevent overfitting. With this configuration, the training process automatically ends at the 13th epoch. The computing infrastructure used is the same as that applied in the steganography experiment stage, namely the Google Colab platform.

Overall, the series of VGG19 experiments in this study encompasses nine different classification scenarios: one baseline scenario using the original RFMiD images without steganography insertion, four scenarios with the LSB method at payload levels of 0.1, 0.2, 0.3, and 0.4 bpp, and four scenarios with the MSB method at the same four payload levels. All scenarios use the same data split as set in Section A: 1,920 training images, 640 validation images, and 640 test images to ensure consistency and comparability across scenarios. Model performance on the test set was evaluated using five quantitative metrics: accuracy, Area Under the ROC Curve (AUC-ROC), precision, recall, and F1-score [28].

E. Comparative Analysis

The comparison between LSB and MSB follows a two-dimensional framework. The first dimension compares average PSNR, SNR, SSIM, and FSIM values from all 3,200 stego images between the two methods at each payload level, with results organized into summary tables and visualized as bar charts for side-by-side reading [11]. The second dimension compares VGG19 accuracy and AUC between LSB and MSB pairs against the baseline scenario. Trend tracking across the four payload levels further assesses whether embedding's impact on quality and diagnostic performance follows a consistent, directional pattern.

III. RESULT AND DISCUSSION

A. Image Quality Evaluation Result

This section presents results from evaluating the stego-image quality produced by the LSB and MSB methods on 3,200 RFMiD retinal fundus images using PSNR, SNR, SSIM, and FSIM metrics. Table I summarizes the average values of the four metrics for both methods at each tested payload level.

The mean and standard deviation in Table I reflect how consistently each method performs across payload levels, with a smaller spread indicating more stable quality regardless of embedded volume. LSB's narrow spread relative to MSB reflects this consistency: its PSNR ranges from 59.9659 to 65.9259 dB, exceeding the 35 dB threshold by nearly 25 dB even at the highest payload [11]. Its SNR of 49.3401-55.3001 dB likewise clears the 30 dB minimum with almost 20 dB to spare [11], [26]. SSIM follows the same pattern, holding within a narrow band of 0.9981 to 0.9997 and sitting well above the 0.90 feasibility mark [9], [11].

TABLE I.

COMPARISON OF IMAGE QUALITY FROM LSB AND MSB STEGANOGRAPHY ON THE RFMID DATASET

Payload (bpp)	LSB PSNR (dB)	LSB SNR (dB)	LSB SSIM	LSB FSIM	MSB PSNR (dB)	MSB SNR (dB)	MSB SSIM	MSB FSIM
0.1	65.9259	55.3001	0.9997	1.0000	18.9972	8.3714	0.9003	0.7500
0.2	62.9566	52.3308	0.9992	1.0000	15.9867	5.3610	0.7994	0.6449
0.3	61.2102	50.5845	0.9986	0.9999	14.2256	3.5998	0.6987	0.5729
0.4	59.9659	49.3401	0.9981	0.9999	12.9759	2.3501	0.5979	0.5342
Mean ± SD	62.51 ± 2.58	51.89 ± 2.58	1.00 ± 0.00	1.00 ± 0.00	15.55 ± 2.61	4.92 ± 2.61	0.75 ± 0.13	0.63 ± 0.09

The comparison of PSNR values between the two methods at all payload levels is visualized in Figure 3.

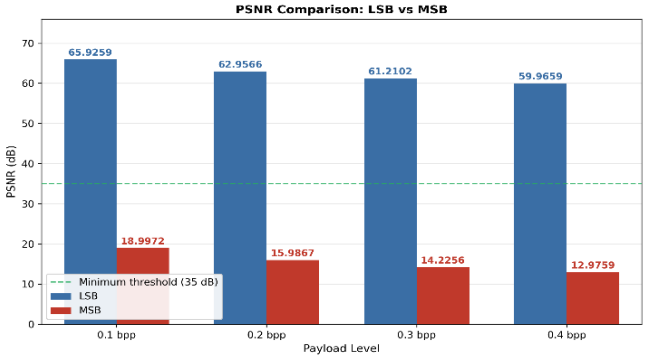

Figure 3. PSNR Comparison of LSB and MSB at Each Payload Level (dashed line = 35 dB threshold).

In contrast to LSB's tight spread, MSB's wider standard deviation across payload levels signals less consistent quality, exacerbated by its already low baseline values. MSB PSNR values are only in the range of 12.9759 dB to 18.9972 dB, all falling below the minimum threshold of 35 dB and more than 16 dB below this standard, even at the lightest payload condition. MSB SNR values were recorded between 2.3501 dB and 8.3714 dB, failing to reach the minimum limit of 30 dB at each tested payload level. As shown in Figure 4.

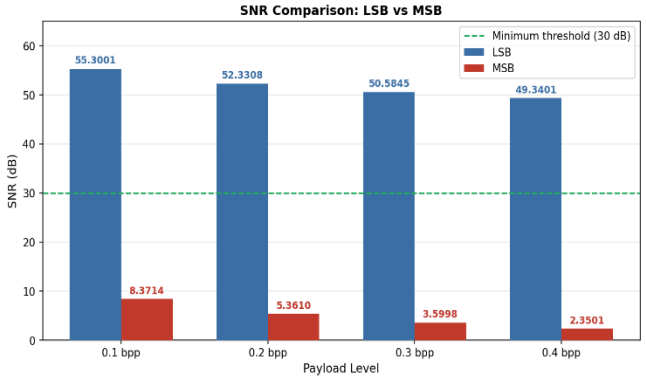

Figure 4. SNR Comparison of LSB and MSB at Each Payload Level (dashed line = 30 dB threshold).

As for the MSB SSIM values, they range from 0.5979 to 0.9003, where only the 0.1 bpp condition managed to slightly reach the 0.90 threshold, while the other three payload levels were below it. As shown in Figure 5.

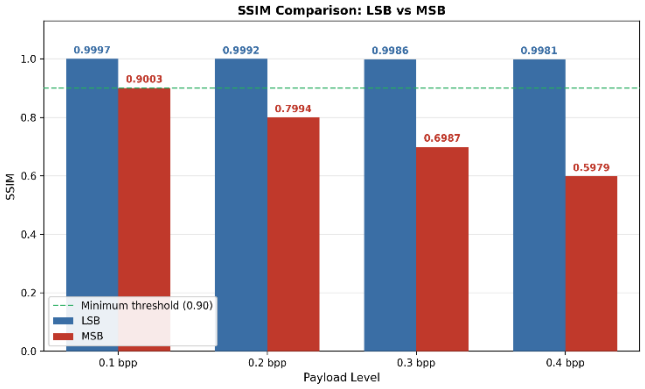

Figure 5. SSIM Comparison of LSB and MSB at Each Payload Level (dashed line = 0.90 threshold).

This substantial quality degradation aligns with the technical consequences of the MSB working mechanism, in which data insertion into the highest-weighted bit of each pixel causes a drastic shift in its numerical value, resulting in visible artifacts that can be identified by the human visual system without instruments [7], [10].

A consistent, downward trend is observed across all three metrics (PSNR, SNR, and SSIM) for both methods as the payload level increases from 0.1 to 0.4 bpp, consistent with the theoretical predictions of spatial-domain-based steganography [11], [20]. The FSIM results reinforce the divergence already evident in the PSNR, SNR, and SSIM findings. The LSB method maintains an FSIM score between 0.9999 and 1.0000 across all four payload levels, indicating near-identical structural and contrast features to the original cover image regardless of embedding capacity. The MSB method, in contrast, declines from 0.7500 at 0.1 bpp to 0.5342 at 0.4 bpp, confirming that feature-level structural integrity deteriorates progressively as embedding capacity increases. This trend mirrors the SSIM results, providing convergent evidence from an independent feature-based perspective that MSB introduces structural distortion beyond what pixel-level metrics alone capture [27], as shown in Figure 6.

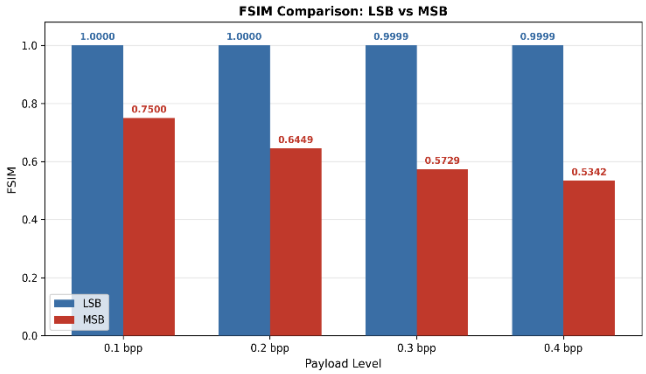

Figure 6. FSIM Comparison of LSB and MSB at Each Payload Level.

A visual comparison between the original retinal fundus image, the LSB stego image, and the MSB stego image at each payload level can be directly seen in Figure 7.

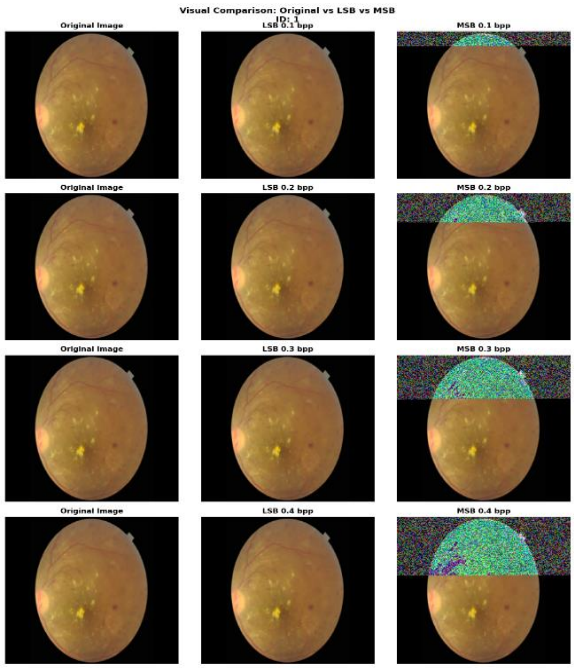

Figure 7. Visual Comparison of Original, LSB, and MSB Stego Images at 0.1-0.4 bpp.

B. Diagnostic Impact Evaluation Result - VGG19

B.1 Baseline Performance of VGG19

Table II summarizes the classification performance of VGG19 on the original RFMiD dataset without steganographic embedding, serving as the reference condition for all experimental scenarios.

TABLE I

BASIC PERFORMANCE OF VGG19 CLASSIFICATION

Class	Precision	Recall	F1-Score	Support
Normal	0.74	0.80	0.77	134
Abnormal	0.95	0.93	0.94	506
Overall	0.90	0.90	0.90	640
Accuracy	0.9000			640
AUC-ROC	0.9321			640

Under baseline conditions, VGG19 achieved an overall accuracy of 90.00% and an AUC-ROC of 0.9321. The Abnormal class achieved an F1-score of 0.94, while the Normal class achieved 0.77, a difference attributed to the asymmetric distribution of the test data, with 134 Normal images and 506 Abnormal images [3]. The training curve shown in Figure 8 indicates that the validation accuracy stabilized early in training, without any signs of overfitting, confirming that the training process remained stable.

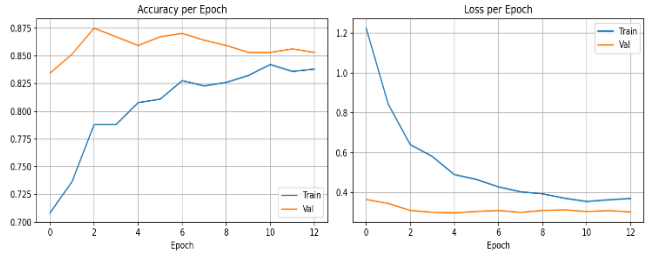

Figure 8. VGG19 Training History on Original RFMiD Images (Baseline).

As shown in Figure 9, out of a total of 640 test images, the model correctly classified 107 Normal images (TN) and 469 Abnormal images (TP), while 27 Normal images were misclassified as Abnormal (FP) and 37 Abnormal images were incorrectly classified as Normal (FN).

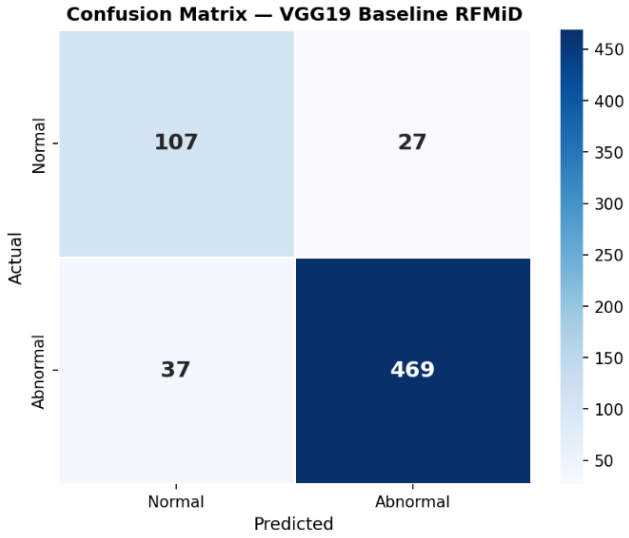

Figure 9. Confusion Matrix of Baseline VGG19 on RFMiD Test Set (640 images).

B.2. VGG19 Performance on LSB Stego Images

Table III presents the confusion matrix and VGG19 accuracy for all LSB conditions. The ΔAcc column represents the change in accuracy relative to the baseline condition, with a positive value indicating an increase and a negative value indicating a decrease from the baseline accuracy of 0.9000.

TABLE II

CONFUSION MATRIX LSB

Payload	TN	FP	FN	TP	Accuracy	ΔAcc
0.1 bpp	114	20	48	458	0.8938	-0.0062
0.2 bpp	113	21	43	463	0.9000	0.0000
0.3 bpp	101	33	34	472	0.8953	-0.0047
0.4 bpp	109	25	41	465	0.8969	-0.0031

Table IV presents the corresponding classification report, including Precision, Recall, F1-Score, and AUC-ROC for each LSB condition.

TABLE III

LSB CLASSIFICATION REPORT

Payload	Precision	Recall	F1-Score	AUC-ROC
0.1 bpp	0.8309	0.8779	0.8506	0.9434
0.2 bpp	0.8405	0.8792	0.8573	0.9549
0.3 bpp	0.8414	0.8433	0.8423	0.9511
0.4 bpp	0.8378	0.8662	0.8507	0.9553

Referring to Table III and Table IV, the accuracy of VGG19 across all LSB embedding conditions ranges from 0.8938 to 0.9000, with a maximum deviation of only 0.0062 points (equivalent to 0.62%) compared to the baseline condition. The LSB 0.2 bpp condition, in particular, achieves accuracy equivalent to the baseline (0.9000) and has the overall best performance among the four tested LSB payload levels. The confusion matrix for the 0.2 bpp LSB condition is shown in Figure 10.

Furthermore, the AUC-ROC values across all LSB conditions (ranging from 0.9434 to 0.9553) consistently exceed the baseline AUC value of 0.9321 at every payload level, indicating that LSB embedding does not erode the discriminative ability of the VGG19 model in distinguishing between Normal and Abnormal classes.

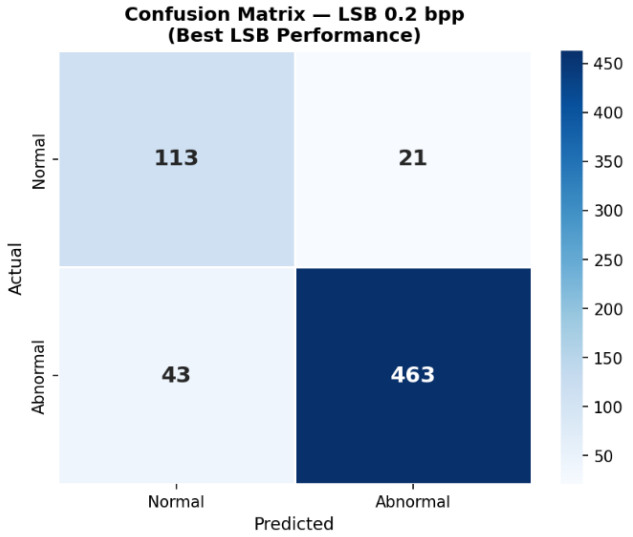

Figure 10. Confusion Matrix of VGG19 on LSB 0.2 bpp (Best LSB). Acc = 0.9000; AUC = 0.9549.

B.3. VGG19 Performance on MSB Stego Images

Table V presents the VGG19 classification performance for all MSB steganography conditions at four payload levels.

TABLE IV

VGG19 CONFUSION MATRIX ON MSB STEGO IMAGES

Payload	TN	FP	FN	TP	Accuracy	ΔAcc
0.1 bpp	108	26	41	465	0.8953	-0.0047
0.2 bpp	108	26	36	470	0.9031	+0.0031
0.3 bpp	95	39	24	482	0.9016	+0.0016
0.4 bpp	100	34	31	475	0.8984	-0.0016

Table VI presents the complete classification report covering Precision, Recall, F1-Score, and AUC-ROC for each MSB condition.

TABLE V

VGG19 CLASSIFICATION REPORT ON MSB STEGO IMAGE

Payload	Precision	Recall	F1-Score	AUC-ROC
0.1 bpp	0.9005	0.8953	0.8973	0.9491
0.2 bpp	0.9062	0.9031	0.9044	0.9548
0.3 bpp	0.8986	0.9016	0.8994	0.9337
0.4 bpp	0.8976	0.8984	0.8980	0.9303

Referring to Tables V and VI, the accuracy values of VGG19 across all MSB embedding conditions range from 0.8953 to 0.9031, with the maximum deviation from the baseline not exceeding 0.0047 (0.47%). The MSB 0.2 bpp

condition recorded the best accuracy achievement among the MSB conditions at 0.9031, slightly above the baseline, as illustrated in the confusion matrix in Figure 11.

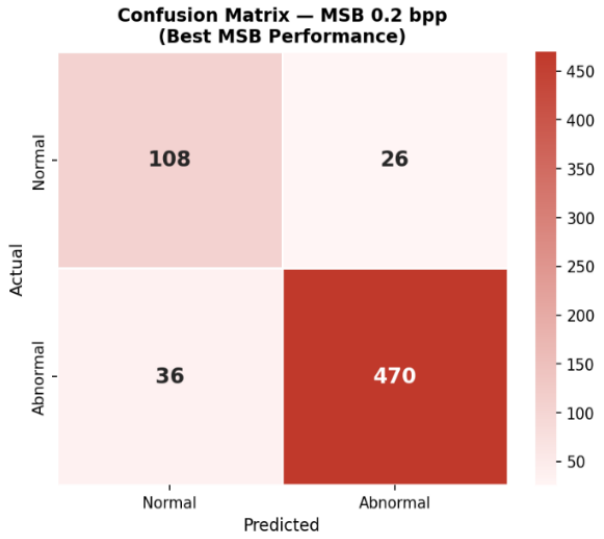

Figure 11. Confusion Matrix of VGG19 on MSB 0.2 bpp (Best MSB). Acc = 0.9031; AUC = 0.9548.

Although MSB embedding induces substantial visual degradation in stego images, VGG19 has been shown to consistently maintain its diagnostic classification capability across all tested conditions. A comprehensive comparison of accuracy and AUC-ROC values across all Baseline, LSB, and MSB scenarios at the four payload levels is presented in Figures 12 and 13.

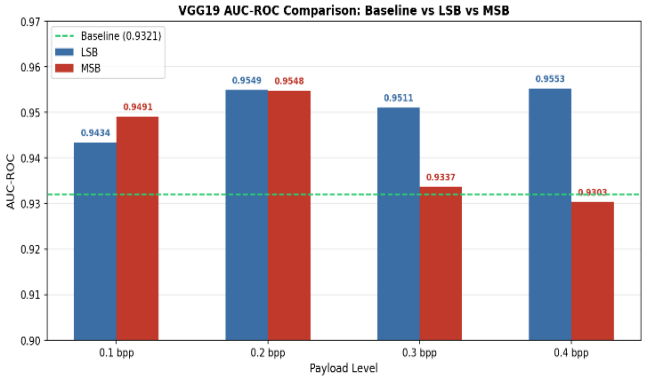

Figure 12. VGG19 Accuracy Comparison Across Baseline, LSB, and MSB Conditions (dashed line = 0.9000).

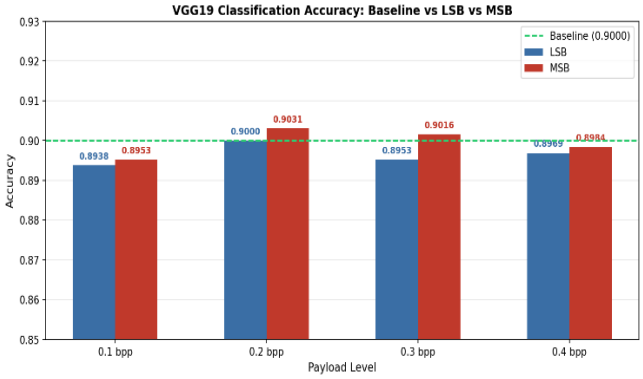

Figure 13. VGG19 AUC-ROC Comparison Across Baseline, LSB, and MSB Conditions.

C. Discussion

C.1. Comparison of Image Quality: LSB vs MSB

Across all payload conditions tested, the empirical record unambiguously positions LSB as the superior method when visual fidelity on retinal fundus images is the evaluative criterion. LSB PSNR values, ranging from 59.9659 dB to 65.9259 dB across all payload levels, exceed the minimum feasibility threshold of 35 dB by a very substantial margin. Even at the highest payload condition (0.4 bpp), LSB PSNR values still exceed the minimum limit by nearly 25 dB [11], [20]. This achievement confirms that 1-bit LSB substitution in retinal fundus images produces a stego quality that is nearly lossless and perceptually safe for medical personnel during diagnostic interpretation.

Conversely, MSB can only produce PSNR in the range of 12.9759 dB to 18.9972 dB, all of which are below the feasibility threshold, reflecting the inevitable consequences of embedding in the highest weighted bits, which triggers drastic pixel intensity shifts and produces visual artifacts that are immediately identifiable by the human visual system, as documented in Figure 5 [7], [10]. This empirically establishes that MSB is intrinsically incompatible with the needs of medical imaging, where visual integrity is an absolute prerequisite in the diagnostic process [7], [10]. To place the image quality achievements and diagnostic performance of VGG19 obtained in this study in a broader context, a comparison is presented with previous studies that used the RFMiD dataset as the test dataset.

C.2. Comparison with Previous Studies on the RFMiD

To contextualize the performance of the VGG19 model used in this study, baseline classification results are compared with those of several previous studies that also used the RFMiD dataset as the basis for evaluation. This comparison is not a direct apple-to-apple one due to differences in the number of classes, model architectures, and augmentation strategies used in each study. However, collectively, it provides an overview of the performance range achievable on the same dataset. A summary of the comparison is presented in Table VII.

TABLE VII.
COMPARISON OF CLASSIFICATION RESULTS ON THE RFMID DATASET

Year	Author(s)	Method	Result
2021	Kumar & Bindu [15]	Ensemble CNN	F1-score: 94.32%
2022	Ho et al. [17]	Deep Ensemble Learning	AUC: 0.9613
2023	Sengar et al. [14]	EyeDeep-Net	Accuracy: 76.04%
2023	Li et al. [3]	Ensemble VGG19 + RestNet50 + Grad-CAM	Accuracy: 97%
2024	Pandey et al. [16]	CNN	Accuracy: 88.72%
2026	This Study	VGG19 Transfer Learning	Accuracy: 90.00%; AUC: 0.9321; F1-score: 90%

The baseline VGG19 results in this study, 90.00% accuracy and AUC of 0.9321 on binary Normal-versus-Abnormal classification, surpass the accuracy levels reported by Sengar et al. [14], and Pandey et al. [16], though both worked with different class structures. Ho et al. instead pooled several CNN architectures into one ensemble to separate healthy from pathological retinal scans, arriving at a discrimination score of 0.9613, roughly three percentage points above the single-network result obtained here, a gap explainable by the architectural difference between combining multiple models and relying on one [17]. A separate ensemble effort by Kumar and Bindu, built to distinguish among 29 distinct retinal conditions rather than a binary outcome, reported an F1 measure of 94.32%; the disparity in both the scoring metric and the sheer number of categories involved limits how directly its result can be set against this study's findings [15]. The 97% figure from Li et al. likewise stems from a differently structured problem, a multi-label setup strengthened by data augmentation rather than the binary task addressed here [3]. Unlike these studies, which optimized models purely for classification accuracy, this study uses VGG19 as an evaluation instrument to assess the diagnostic impact of LSB and MSB steganography, an analytical purpose absent from prior comparative work.

C.3. Diagnostic Impact of Steganography on VGG19 Performance

Neither method inflicted meaningful damage on VGG19's capacity to discriminate Normal from Abnormal presentations, with maximum accuracy deviations of 0.62% for LSB and 0.47% for MSB, both within the margin of natural training variability and not clinically significant. This directly addresses the primary research gap: steganographic embedding does not erode VGG19's diagnostic capability regardless of method, consistent with Karawia's principle that modifications introduced by medical image steganography must remain invisible to the diagnostic judgment exercised by the reading specialist [24].

Furthermore, the AUC-ROC values across all LSB conditions (0.9434-0.9553) consistently exceeded the baseline AUC of 0.9321 (Figure 11), with a similar pattern observed for MSB in the 0.1 and 0.2 bpp conditions. This indicates that pixel disturbances due to embedding do not interfere with VGG19's learned feature-space separation, consistent with the robustness of this architecture in retinal fundus image classification [3], [12].

This insensitivity reflects both how VGG19 extracts features and how MSB distortion is structured. Convolutional architectures such as VGG19 build classification decisions on hierarchical patterns of edges, textures, and spatial arrangement rather than raw pixel brightness, a property previously confirmed for this network when applied to fundus photographs [3]. MSB produces large-magnitude, visually destructive shifts, but because they apply uniformly and randomly rather than targeting diagnostically informative regions such as the optic disc, blood vessels, or macula, the global distributional characteristics of VGG19's feature maps remain preserved enough for correct discrimination, consistent with a decision boundary that tolerates this distortion at the payload levels tested.

A second consideration is dataset-level bias: the RFMiD test subset's 134 Normal versus 506 Abnormal images skew toward the Abnormal class, allowing the model to sustain high accuracy on the dominant class even under substantial input distortion, since majority-class prediction alone yields accuracy well above chance. This imbalance may inflate the apparent stability of classification performance under distortion, a limitation future studies should isolate using a class-balanced dataset.

The confusion matrices in Figures 7-10 reinforce this: under the best LSB condition (0.2 bpp), correctly classified images approach baseline (TN=113 vs 107; TP=463 vs 469), with a similar pattern for the best MSB condition (TN=108; TP=470). These minor fluctuations reflect natural training variability rather than a systematic diagnostic impact.

C.4. Overall Comparison: LSB vs MSB

The capacity-quality trade-off across the four payload levels carries direct weight against real patient data storage needs. LSB sustains a PSNR above 59 dB even at the highest tested payload of 0.4 bpp, while MSB's PSNR collapses to below 13 dB at the same payload, a gap of roughly 47 dB holding consistently across all levels. LSB therefore faces no meaningful capacity ceiling before quality becomes unacceptable, whereas MSB crosses into clinically unacceptable quality at every tested level, including the lowest, leaving an EMR architect free to scale LSB payload upward for richer metadata without compromising fidelity, a margin MSB never offers. This reframes the trade-off

described by Belagali and Udupi [20]: it functions not as a property to be balanced, but as a decision boundary determining which method remains viable once minimum diagnostic quality is treated as a hard constraint.

Based on a comprehensive evaluation across both dimensions, LSB is clearly the superior method for steganography of retina fundus images. Although both methods maintain VGG19 diagnostic performance at a similar level (Figure 12), LSB achieves much higher visual quality, with an average PSNR difference of around 47 dB, making it the only method that simultaneously meets image quality standards and diagnostic security requirements within the two dimensions evaluated here, a scope that does not extend to resistance against steganalytic detection. This aligns with Rahman et al., who documented that modifications to low-weight bits (LSB) result in much smaller pixel change impacts compared to modifications to high-weight bits [7], as well as Rustad et al., who demonstrated that the appropriate pattern selection strategy for LSB substitution yields significantly better imperceptibility values in medical images [8].

While VGG19's diagnostic performance remains stable under MSB's severe visual degradation, this must not be interpreted as evidence that MSB is clinically safe for human interpretation. Karawia emphasizes that medical image steganography must not affect a specialist physician's diagnostic accuracy [24]. An automated model and a human clinician evaluate image quality through different lenses: VGG19 separates classes based on statistical feature patterns extracted across the entire image, whereas a specialist relies on direct visual inspection of fine-grained detail. Lerch et al. confirm this divergence empirically, finding that trained radiologists tolerate noise-induced quality loss far better than convolutional classifiers do, maintaining their diagnostic accuracy even where the models' performance deteriorates despite augmentation-based training [30]. MSB's gross, immediately visible distortions would plausibly undermine a clinician's confidence in visual interpretation even where the underlying signal remains sufficient for correct AI classification. Steganographic acceptability for clinical deployment must therefore be judged against human visual tolerance as well as AI accuracy, as demonstrated here by the gap between MSB's preserved VGG19 accuracy and its unacceptable PSNR, SSIM, and FSIM values.

This reinforces LSB's suitability for securing retinal fundus images in EMR systems, in line with the implementation mandate of Indonesian Ministry of Health Regulation No. 24 of 2022 [1].

C.5. Research Limitations

Several limitations should be noted. First, embedding was applied uniformly across all pixels without considering the region of interest, treating diagnostic areas such as the optic disc and retinal blood vessels the same as non-informative background. Second, only binary classification (Normal vs. Abnormal) was evaluated, leaving multi-class disease classification unexplored. Third, resistance to steganalysis has not been tested. Fourth, all experiments relied exclusively on RFMiD, whose camera models, acquisition protocols, class distribution, and disease composition may not generalize to other clinical sources with differing resolution, illumination, or noise characteristics. Fifth, diagnostic impact was assessed using a single classification architecture, VGG19, so the conclusions drawn here cannot be generalized to other deep learning models that may exhibit different sensitivity to steganographic distortion. Sixth, the image quality and classification metrics used in this study, while standard in computational evaluation, were not validated against the subjective judgment of a practicing ophthalmologist, leaving open the question of whether image changes deemed acceptable by these metrics would also be judged acceptable in actual clinical reading. Seventh, while the accuracy and AUC differences across payload levels follow a consistent directional trend that lends indirect support to their stability, formal statistical significance testing comparing baseline, LSB, and MSB outcomes was not performed in this study. The PSNR, SNR, SSIM, FSIM, and classification outcomes reported here should therefore be interpreted as specific to RFMiD rather than as universal properties of LSB and MSB steganography on retinal fundus images.

Future research should integrate ROI-based insertion strategies, assess their impact on multi-class retinal disease classification, examine their resistance to steganalysis detection methods, and validate these findings across additional retinal fundus datasets with differing acquisition characteristics [25], extend the evaluation to other deep learning architectures, and incorporate subjective assessment by ophthalmology specialists to confirm clinical acceptability, and apply formal statistical significance testing, such as McNemar's test, to confirm whether the observed differences in classification performance across methods are statistically meaningful rather than attributable to natural training variability.

IV. CONCLUSION

This study demonstrates that LSB and MSB have fundamentally different impacts on the visual integrity of retinal fundus images, yet neither significantly disrupts the diagnostic capability of VGG19. From the perspective of visual quality, LSB consistently excels with PSNR values of 59.97-65.93 dB, SNR of 49.34-55.30 dB, SSIM of 0.9981-0.9997, and FSIM of 0.9999-1.0000 across all payload levels, all exceeding the established minimum feasibility threshold. Conversely, MSB produces PSNR values of 12.98-18.99 dB and SNR of 2.35-8.37 dB, all below standard, with SSIM only passing at the lowest payload (0.9003 at 0.1 bpp) but plummeting to 0.5979 at the highest payload, while FSIM declines from 0.7500 to 0.5342 across the same range, jointly indicating significant structural damage to the image that is unacceptable in the context of medical imaging. From a diagnostic perspective, both methods proved safe, with maximum deviations from baseline of 0.62% for LSB and

0.47% for MSB, while AUC-ROC values across all scenarios remained above 0.93, well above the threshold for meaningful clinical decline. Within the image quality and VGG19 diagnostic dimensions evaluated here, LSB is established as the more suitable method for retinal fundus images in the EMR ecosystem, though this recommendation does not extend to resistance against steganalytic detection, which remains untested in this study.

BIBLIOGRAPHY

[1] E. T. Ardianto, Sabran, dan L. Nurjanah, "Analisis Aspek Keamanan Data Pasien dalam Implementasi Rekam Medis Elektronik di Rumah Sakit X," Jul 2024.

[2] M. Fitriyasari, "Tren dan Inovasi dalam Image Hiding untuk Keamanan Informasi Medis: Tinjauan Literatur," RIGGS, vol. 4, no. 2, hlm. 2375–2381, Jun 2025, doi: 10.31004/riggs.v4i2.854.

[3] Z. Li, M. Xu, X. Yang, Y. Han, dan J. Wang, "A Multi-Label Detection Deep Learning Model with Attention-Guided Image Enhancement for Retinal Images," Micromachines, vol. 14, no. 3, hlm. 705, Mar 2023, doi: 10.3390/mi14030705.

[4] T. Aswathi, T. R. Swapna, dan S. Padmavathi, "Transfer Learning approach for grading of Diabetic Retinopathy," J. Phys.: Conf. Ser., vol. 1767, no. 1, hlm. 012033, Feb 2021, doi: 10.1088/1742-6596/1767/1/012033.

[5] R. Taj, F. Tao, S. Kanwal, A. Almogren, dan A. U. Rehman, "A SURF and SVD-based robust zero-watermarking for medical image integrity," PLoS ONE, vol. 19, no. 9, hlm. e0307619, Sep 2024, doi: 10.1371/journal.pone.0307619.

[6] Ramyashree, P. S. Venugopala, S. Raghavendra, dan V. S. Kubihal, "Enhancing Secure Medical Data Communication Through Integration of LSB and DCT for Robust Analysis in Image Steganography," IEEE Access, vol. 13, hlm. 1566–1580, 2025, doi: 10.1109/ACCESS.2024.3522957.

[7] S. Rahman, J. Uddin, H. U. Khan, H. Hussain, A. A. Khan, dan M. Zakarya, "A Novel Steganography Technique for Digital Images Using the Least Significant Bit Substitution Method," IEEE Access, vol. 10, hlm. 124053–124075, 2022, doi: 10.1109/ACCESS.2022.3224745.

[8] S. Rustad, D. R. I. M. Setiadi, A. Syukur, dan P. N. Andono, "Inverted LSB image steganography using adaptive pattern to improve imperceptibility," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 6, hlm. 3559–3568, Jun 2022, doi: 10.1016/j.jksuci.2020.12.017.

[9] S. N. M. Al-Faydi, S. K. Ahmed, dan H. N. Y. Al-Talb, "Improved LSB image steganography with high imperceptibility based on cover-stego matching," IET Image Processing, vol. 17, no. 7, hlm. 2072–2082, Mei 2023, doi: 10.1049/ipr2.12773.

[10] M. Z. Ali, O. Riaz, H. M. Hasnain, W. Sharif, T. Ali, dan G. S. Choi, "Elevating Image Steganography: A Fusion of MSB Matching and LSB Substitution for Enhanced Concealment Capabilities," CMC, vol. 79, no. 2, hlm. 2923–2943, 2024, doi: 10.32604/cmc.2024.049139.

[11] M. Ilham dan C. Kirana, "Perbandingan Kualitas Citra Grayscale Steganografi Metode LSB dan DCT Berdasarkan PSNR dan SSIM," vol. 7, no. 2, 2026, doi: https://doi.org/10.24076/joism.2026v7i2.2492.

[12] M. A. K. Raiaan dkk., "A Lightweight Robust Deep Learning Model Gained High Accuracy in Classifying a Wide Range of Diabetic Retinopathy Images," IEEE Access, vol. 11, hlm. 42361–42388, 2023, doi: 10.1109/ACCESS.2023.3272228.

[13] K. A. Saputra dkk., "Multi-Disease Retinal Classification Using EfficientNet-B3 and Targeted Albumentations: A Benchmark on Kaggle Retinal Fundus Images Dataset," SinkrOn, vol. 10, no. 1, hlm. 232–241, Jan 2026, doi: 10.33395/sinkron.v10i1.15530.

[14] N. Sengar, R. C. Joshi, M. K. Dutta, dan R. Burget, "EyeDeep-Net: a multi-class diagnosis of retinal diseases using deep neural network," Neural Comput & Applic, vol. 35, no. 14, hlm. 10551–10571, Mei 2023, doi: 10.1007/s00521-023-08249-x.

[15] E. S. Kumar dan C. S. Bindu, "MDCF: Multi-Disease Classification Framework on Fundus Image Using Ensemble CNN Models," vol. 40, no. 09, hlm. 35–45, 2021, doi: 10.17605/OSF.IO/ZHA9C.

[16] P. U. Pandey dkk., "Ensemble of deep convolutional neural networks is more accurate and reliable than board-certified ophthalmologists at detecting multiple diseases in retinal fundus photographs," Br J Ophthalmol, vol. 108, no. 3, hlm. 417–423, Mar 2024, doi: 10.1136/bjo-2022-322183.

[17] E. Ho dkk., "Deep Ensemble Learning for Retinal Image Classification," Trans. Vis. Sci. Tech., vol. 11, no. 10, hlm. 39, Okt 2022, doi: 10.1167/tvst.11.10.39.

[18] M. Kim, Y. Cho, H. Park, dan G. Qu, "ASIGM: An Innovative Adversarial Stego Image Generation Method for Fooling Convolutional Neural Network-Based Image Steganalysis Models," Electronics, vol. 14, no. 4, hlm. 764, Feb 2025, doi: 10.3390/electronics14040764.

[19] S. Pachade dkk., "Retinal Fundus Multi-Disease Image Dataset (RFMiD): A Dataset for Multi-Disease Detection Research," Data, vol. 6, no. 2, hlm. 14, Feb 2021, doi: 10.3390/data6020014.

[20] Pooja Belagali, Dr. V. R. Udupi, "Robust Image Steganography Based on Hybrid Edge Detection," tjjpt, vol. 44, no. 3, hlm. 1509–1521, Okt 2023, doi: 10.52783/tjjpt.v44.i3.531.

[21] N. A. Mohsin dan H. A. Alameen, "A Hybrid Method for Payload Enhancement in Image Steganography Based on Edge Area Detection," Cybernetics and Information Technologies, vol. 21, no. 3, hlm. 97–107, Sep 2021, doi: 10.2478/cait-2021-0032.

[22] Ramyashree, P. S. Venugopala, S. Raghavendra, dan B. Ashwini, "CrypticCare: A Strategic Approach to Telemedicine Security Using LSB and DCT Steganography for Enhancing the Patient Data Protection," IEEE Access, vol. 12, hlm. 101166–101183, 2024, doi: 10.1109/ACCESS.2024.3430546.

[23] S. Rahman dkk., "A novel and efficient digital image steganography technique using least significant bit substitution," Sci Rep, vol. 15, no. 1, hlm. 107, Jan 2025, doi: 10.1038/s41598-024-83147-3.

[24] A. A. Karawia, "Medical image steganographic algorithm via modified LSB method and chaotic map," IET Image Processing, vol. 15, no. 11, hlm. 2580–2590, Sep 2021, doi: 10.1049/ipr2.12246.

[25] Y. Sanjalawe, S. Al-E'mari, S. Fraihat, M. Abualhaj, dan E. Alzubi, "A deep learning-driven multi-layered steganographic approach for enhanced data security," Sci Rep, vol. 15, no. 1, hlm. 4761, Feb 2025, doi: 10.1038/s41598-025-89189-5.

[26] M. Naufal, H. A. Azies, F. A. Zami, dan R. M. Brilianto, "Optimizing Driver Drowsiness Detection: Evaluating CLAHE and AHE Enhancement Techniques," vol. 15.

[27] Y. Al Najjar, "Comparative Analysis of Image Quality Assessment Metrics: MSE, PSNR, SSIM and FSIM," IJSR, vol. 13, no. 3, hlm. 110–114, Mar 2024, doi: 10.21275/SR24302013533.

[28] A. M. Almhlbdi, N. D. Altowairqi, A. O. Alshutayri, dan R. K. Qarout, "Deep Learning-Based Multi-Class Detection of LSB Steganography in Digital Images," IEEE Access, vol. 13, hlm. 191543–191553, 2025, doi: 10.1109/ACCESS.2025.3628784.

[29] M. F. Alexander, V. R. T. Adrian, L. R. Esprayenduo, dan M. Naufal, "Klasifikasi Penyakit Mata Menggunakan Random Forest Dengan Optimasi Hyperparameter RandomSearchCV," vol. 5, no. 2, 2026.

[30] L. Lerch dkk., "DreamOn: a data augmentation strategy to narrow the robustness gap between expert radiologists and deep learning classifiers," *Front Radiol*, vol. 4, hlm. 1420545, 2024, doi: 10.3389/fradi.2024.1420545.