Multispectral-Based Imaging and Machine Learning for Noninvasive Blood Loss Estimation⋆

Ara Abigail E. Ambita[1] , Catherine S. Co[2] , Laura T. David[3] , Charissa M. Ferrera[3] , and Prospero C. Naval, Jr.[3]

[1] University of the Philippines Visayas, Iloilo, Philippines
[2] University of the Philippines Manila, Manila, Philippines
[3] University of the Philippines Diliman, Quezon City, Philippines
{aeambita, pcnaval,csco2}@up.edu.ph, {ltdavid,cmferrera}@msi.upd.edu.ph

Abstract. Blood loss estimation during surgical operations is crucial in determining the appropriate transfusion decisions. More practical emerging solutions, e.g. the Triton System, use image processing and artificial intelligence (AI) in quantifying blood loss from images of blood-soaked sponges. Triton utilizes an infrared or depth camera that's used to identify the region of color (RGB) image corresponding to a surgical textile. However, calculating depth is computationally expensive and can provide only the shape information. In this research, we propose a multispectral-based imaging and machine learning approach to directly quantify blood loss from images of surgical sponges. Near-infrared (NIR) and Visible (Vis) light sources in conjunction with an RGB imaging sensor without a NIR filter are used. With this, in addition to the improved focus and reduced background interference on the gauze image due to blood's IR absorption capacities, the color as well as the shape information may be utilized. Results show that the multispectral-based imaging approach rendered a +28.30%, +48%, +27.97%, and 25.72% improvement on the MAE, MSE, RMSE, and MAPE, compared to using a single Vis wavelength or RGB image.

Keywords: Multispectral Imaging · Machine Learning · Blood Loss.

1 Introduction

The estimation of blood loss in a patient during surgery is essential in determining the appropriate transfusion decision or it might lead to costly, invasive, and unnecessary treatments. Currently, there is a lack of a standardized approach to practically and accurately measure intraoperative blood loss [16, 19]. The most frequently practiced method used by physicians in determining blood loss is a visual estimation. However, aside from demanding expertise, the visual judgment may have been occluded or affected by other fluids, such as urine, amniotic fluid, or sterile water, that combine with blood [13].

Other alternatives to blood loss estimation is the gravimetric and the photometric/colorimetric method. While the gravimetric method is more accurate than visual estimation, it is time-consuming, laborious, and also heavily affected by the combined nonsanguineous fluids that cause overestimation [12]. On the other hand, the photometric/colorimetric analysis is the reference standard but the complexity of the clinical procedure and increased medical costs also limit the use of this method in clinical practice [15].

More recent blood loss estimation methods use computer vision and artificial intelligence (AI). An example is a US FDA-approved mobile application (Triton System, Gauss Surgical Inc, Los Altos, CA) that captures images of surgical sponges and then uses advanced algorithms to distinguish blood and non-sanguineous fluids [1]. Triton System utilizes an infrared or depth camera in conjunction with a color (RGB) image. From the depth image, an image mask may be generated using to identify a region of the color image that is corresponding to a surgical textile. This information from the color image may then be used to estimate a blood component characteristic [3]. However, calculating the depth as well as transforming the infrared image to the geometry perspective of the color image is computationally expensive. Moreover, the low light performance can be poor, and the accuracy of the depth is influenced by the baseline distance. True enough, in Triton, there are plenty of parameters in the depth image that must be passed in order to process the color image, i.e. perimeter classifier, planarity classifier, normality classifier, distance classifier, texture classifier, and color classifier. But, while there are IR cameras that can generate depth without calculation, they can only provide shape information. Moreover, since Triton estimates the hemoglobin (Hgb) loss per sponge using the patient's pre-procedure Hgb value, systemic biases from rinsing and the Hb analyzer could not be eliminated.

A direct measurement of blood loss on sponge using AI was also conducted (Li et al, 2020 [14]). Their research did not consider blood mixed with other non-sanguineous fluids, which is primarily the issue that impedes an accurate estimation. Moreover, the blood-soaked sponge was fully expanded in order to see all the details in the gauze, which generally costs 5x the time of capturing a folded gauze. As this is both time-consuming and inefficient, this may not be sustainable for the personnel in the long run.

Motivated by the discussed risks of inaccurate estimation and the shortcomings of each existing method to estimate blood loss, we aim to develop a direct, cost-effective approach for the accurate and real-time estimation of intraoperative well as the post-operative blood loss volume absorbed in surgical gauze and sponge. Triton uses optical systems that can provide depth information such as range-gated time-of-flight (ToF) cameras, RF-modulated ToF cameras, pulsed light ToF, and projected light stereo cameras [3]. Whereas, we propose that instead of using an infrared or a depth camera, a typical RGB imaging sensor without an infrared filter is used. With this, in addition to the improved focus and reduced background interference on the gauze image due to blood's IR absorption capacities [6], the color as well as the shape information may be utilized.

We aim to improve (1) blood detection, (2) direct and non-invasive quantification of the blood volume absorbed on the gauze, and (3) distinguishing blood from non-blood samples or fluids.

More specifically, our proposed method is multispectral-based imaging that utilizes light sources with varying wavelengths. We employ an 850 nm near-infrared (NIR) in addition to the white visible (Vis) light, for which we term the *dual Vis-NIR method*. Additionally, we compare its estimation performance to models using only a Visible (white) light or NIR light. Also, because we aim to deploy the application in small devices, we utilize machine learning which requires less computational resources than traditional deep learning methods.

2 Methodology

The general flow of the method or research is presented in Figure1. The dataset is composed of surgical gauze RGB images captured using different lighting sources - Visible (Vis) / white light, 850 nm Near Infrared (NIR), and combined white light and NIR light (dual Vis-NIR). Before extracting features, the images are first converted to HSV. These images are the inputs for the model composed of three general steps: feature extraction, machine learning, and performance evaluation.
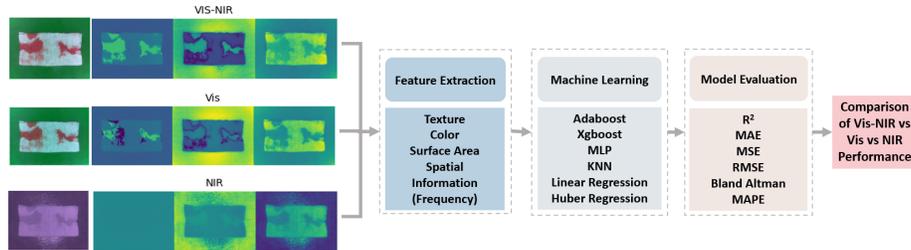

Fig. 1: Proposed Methodology

2.1 Hardware Setup

To lessen the varying lighting conditions, the device is enclosed in a box where no external light can seep in. This creates a controlled environment that has consistent lighting despite the different settings where the device can be used.

Table 1 shows the proposed dual Vis-NIR configuration inside the controlled environment. It contains an Infrared LED (850 nm), White LED (Visible Light), and (4) Imaging Sensor that can switch between an RGB mode and RGB-IR mode.

Table 1: Dual Vis-NIR Imaging Sensor and LED Configurations

Sensor	Conditions		
	Vis	NIR	Vis-NIR
IR (LED 850nm)	On	-	On
Vis (LED)	-	On	On
Vis (Cam)	-	On	-
IR (Cam)	On	-	On

Table 2: Dataset Distribution

Setup	# Images
Pure Blood (Dry)	397
Pure Blood (Wet)	260
Blood + Water	355
Total	1012

2.2 Dataset

During surgical operations, the dry gauze may be soaked in water before use to avoid grazing the internal organs or may be rinsed with water and reused. We prepared pure blood solutions as well as different dilution ratios for the blood+water setup. These setups are summarized in Table 2. For the Pure Blood (Dry Gauze), we varied weights from 0.5-10.5 grams (g), 0.5 intervals, each with 10 repetitions. Since the gauze is folded, there are differences in the spread and color of blood on the two sides of the gauze. To augment our dataset, we take images of both sides as different samples. Meanwhile, in Pure Blood (Wet Gauze), the weights are limited to 7 g. Since the gauze was rinsed with water, it already contains water and therefore limits the blood capacity of the gauze. Lastly, for the Blood + Water (Dry Gauze) setup, we collected samples from 1-10 total g, with a 1 g interval. For each total weight, we took samples with varying blood + water ratio by 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%.
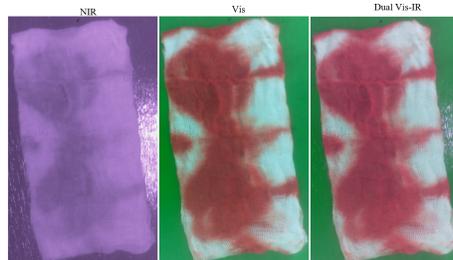

Fig. 2: NIR, Vis, and Dual Vis-NIR of a Gauze (With 2g Blood)

Each gauze sample will have a set of three images of different configurations captured using : (1) a Vis light, (2) an NIR light, and (3) dual Vis-NIR light. An example of a 2g blood is shown in Figure 2. The resulting dataset is composed of gauze images with three channels (H, S, V), each two spatial (x, y) dimensions and one λ wavelength dimension, where the λ = VIS, VIS-NIR, NIR.

2.3   Feature Extraction Techniques (FET)

The color of blood appears desaturated as it gets diluted with more water and the volume is directly proportional to the surface area. In order to differentiate gauze with varying volumes and setups (e.g. dry vs wet gauze, pure blood vs blood + water), we often look at indicators such as color, texture, and as well as the surface area covered by the blood. We exploit color moments to represent colors, local binary patterns represent texture, Fourier transform descriptors for spatial information, and thresholding to compute for the surface area.

Color Moments The color composition of an image can be viewed as a color distribution where the dominant features from this distribution can be extracted. These dominant features, the color moments, are measures that can be used to differentiate images based on their features of colors. The basis of this idea lies in the assumption that the color distribution in an image can be interpreted as a probability distribution, characterized by its moments, resp. central moments. Thus, it follows that if we interpret the color distribution of an image as a probability distribution, the color distribution can be characterized by its moments as well [21].

The feature descriptor will then be composed of the first four (4) moments of an image's color distribution - the mean, standard, deviation, skewness, and kurtosis. We define the mean as the average color of the image and the standard deviation as the square root of the variance. In the third and fourth moments, the skewness and kurtosis both provide an idea about the shape of the color distribution. The skewness measures how asymmetric the color distribution is while the kurtosis measures how extreme the tails are in comparison to the normal distribution. These moments are calculated independently for every color channel. Thus, if we compute for the moments of an HSV image, then will obtain a 12-dimensional feature vector - 4 features for each channel.

Local Binary Patterns Local Binary Patterns (LBP) are based on the assumption that texture is based on a pattern and its strength, as proposed by Ojala et. al [17,18]. To compute for the LBP, a neighborhood of size r surrounding a center pixel is defined. Originally, LBP is defined in a 3x3 neighborhood with the gray value of the center pixel set as a threshold. Neighbors with intensities higher or equal to the value of the center pixel are given a value of 1, otherwise, they are set as 0. The thresholded values (0 or 1), are weighted, and by summing up the result, an LBP code that contains information about the local features of the texture of the image is obtained [20].

Surface Area We define the surface area as the area of the gauze covered by the blood. In order to compute the area, we follow the process in Figure 3 where we apply thresholding to segment the blood from the unused sections of the gauze and the background. We first compute the "C" (chroma) channel by finding the difference between the largest and smallest of the RGB values (for each pixel independently). A simple threshold of the chroma leads to finding the regions with blood.

Fig. 3: Surface Area
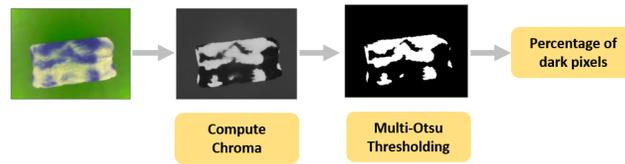
However, the issue with Otsu thresholding is it cannot filter the regions with only a minimal amount of blood and has a lighter shade of red. To overcome this problem, the threshold value returned by the Otsu binarization algorithm is multiplied by a certain factor or bias before using it in regular binary thresholding. After the thresholding, we have an image with black pixels that coincide with the blood and white pixels which we consider as the background. We simply compute for the number of the black pixels divided by the total number of pixels to compute for the surface area.
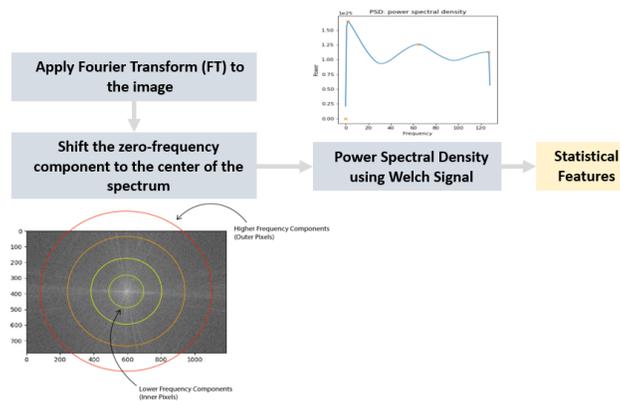

Fig. 4: Fourier Transform

Fourier Transform Descriptors We compute the two-dimensional fast Fourier transform of the input image. Since the result is a complex number array that is difficult to visualize directly, we take the absolute values of the output to be displayed as an image. We, therefore, transform it into a 2-dimensional spectrum. The output frequency domain image tells us how much each frequency component is included in the original image. The 2d may be converted to a summarized 1d by computing the Welch signal (Power Spectral Density). From this simplified representation, we can easily obtain statistical features such as the mean, standard deviation, skewness, kurtosis, minimum, and maximum value. The flowchart for computing these features is displayed in Figure 4.

FET Keywords As a standard, the keywords and acronyms used to represent features are listed in Table 3.

Table 3: Feature Summary. n refers to the number of components extracted per channel

Feature	Keyword	Components	n
Local Binary Patterns	LBP	-	40
Color Moments	CM	mean, std, skew, kurtosis	4
Power Spectral Density	PSD	mean, std, skew, kurtosis, min, max	6
Surface Area	SA	-	1
LBP of Fast Fourier Transform	LBP-FFT	mean, std, skew, kurtosis, min, max	6

2.4   Regression Modelling

Regression predictive modeling involves predicting a numeric variable given some input, often numerical input. Primarily, XGBoost was used as the model to predict the blood absorbed in surgical sponges using features discussed in Section 2.3. XGboost is an optimized scalable machine learning algorithm that uses a gradient boosting framework, like Adaboost [4]. Gradient boosting [9] refers to an ensemble of many decision trees, each of which is a weak learner because it only learns from several attributes from the dataset [8]. The boosted regression will obtain a strong predictor from this ensemble of multiple weak learners [23]. Each weak learner is based on random subsamples of the training set through several iterations, created one by one so that each subsequent learner is trained using the residuals of the previous learner. In other words, the new learner corrects the errors made previously by the previous learner and then predicts the outcome. In XGboost, each ensemble uses the sum of K functions to predict an output y_i using Eq. 1.

$$\hat{y_i} = \theta(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{1}$$

where f_k is the kth independent decision tree in the sample, $f_k(x_i)$ represents the prediction score generated by that tree for the *ith* sample. To train the weak learner, the set of $f_k s$ is then used to minimize the objective function (loss function and regularization) at iteration t. Loss function, e.g. mean squared error (MSE) for regression measures the difference between the observed response and predicted response. Meanwhile, the tree pruning parameter that regulates the depth of the tree reduces the size and complexity of the decision tree, hence, preventing overfitting.

Its performance is compared to other state-of-the-art machine learning algorithms such as multilayer perceptron (MLP), K nearest neighbors (KNN), support vector machines (SVM), random forest (RF), linear regression, Huber regression, and AdaBoost.

2.5   Model Evaluation

To evaluate the performance of the regression model, the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R^2, and Mean Absolute Percentage Error (MAPE) are computed. The correlation only measures the linear association between the association of two sets of observations. Since this technique may be inadequate and misleading when assessing an agreement between two methods, the Bland-Altman method [10] is also included as an evaluation metric.

3   Results and Discussion

In this experiment, the performance of every single wavelength (visible (Vis), infrared (IR), and dual Vis-IR) is evaluated in order to determine the most effective wavelength for the detection and estimation of blood volume absorbed in gauze. Figure 1 displays the general overview of the setup involving a single wavelength training. Note that each dataset with a corresponding wavelength is trained independently. For each wavelength, different feature combinations (listed in Table 3) are extracted that serve as input to machine learning regressors. Five-fold cross-validation on the dataset was applied where we computed the R^2, MAE, and RMSE per split. Each train-test subset is comprised of 1012 and 203 images, respectively.

The result of running this experiment is displayed in Table 4 and Figure 5. Note that we prefer a higher R^2 and a lower result for other metrics (MAE, MSE, RMSE, MAPE). Consistent on all the evaluation metrics and regardless of the feature set used, combining dual Vis-NIR wavelengths has dramatically reduced the errors in the prediction, as seen in the large difference between the dual Vis-NIR and just using the Vis or IR wavelength. For instance, simply using the CM (or the color moments extracted) as a feature, the improvements recorded a +28.30% improvement on the MAE, +48% on the MSE, +27.97% on RMSE, and 25.72% when using the dual-wavelength approach compared to just using visible (white) lighting.

(a) R^2

(b) MAE
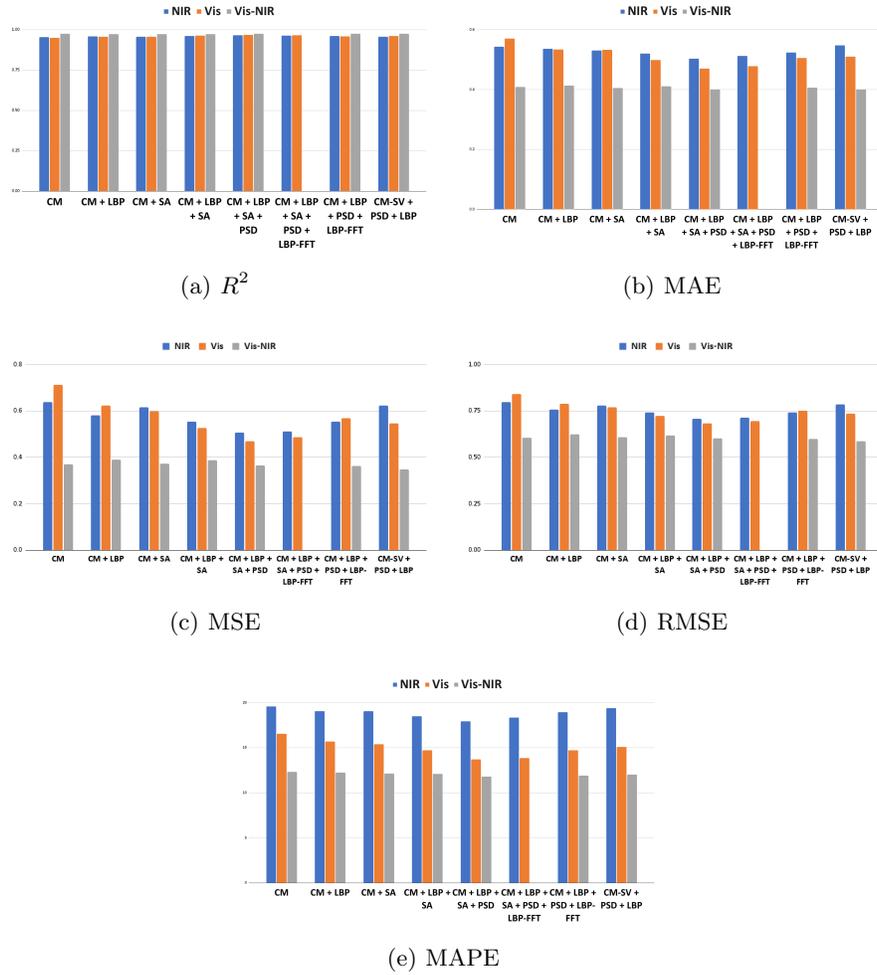

(c) MSE

(d) RMSE

(e) MAPE

Fig. 5: Performance Comparison of Light Sources with Varying Wavelengths

From a spectroscopy perspective, specific molecular absorptions in a certain chemical or fluid cause NIR absorptions and therefore provide more information about a sample's chemical structure [7]. Since NIR wavelength then has been widely explored in blood-related medical or forensic applications, the result of this experiment further confirms the effectiveness of infrared lighting in medical imaging. Specifically, the proposed dual-wavelength technique that takes advantage of infrared lighting was proven to improve blood detection and estimation.

In addition, lower volumes and overly diluted blood are a source of spectral variations, where the absorbed light may be dominated by the background, hampering the quality of Vis images. The background interference may be re-

Table 4: Performance Comparison of Different Feature Combinations and Wavelengths on the Merged Dataset. Values in red indicate the best result for each metric. Highlighted in bold indicates the best wavelength.

Metric	W	Feature Combination							
		CM	CM+LBP	CM+SA	CM+LBP+SA	CM+LBP+SA+PSD	CM+LBP+SA+PSD+LBP-FFT	CM+LBP+PSD+LBP-FFT	CM+PSD+LBP-FFT
R^2	NIR	0.954	0.959	0.956	0.961	0.964	0.964	0.961	0.956
	Vis	0.948	0.955	0.957	0.962	0.967	0.966	0.959	0.961
	Vis-NIR	**0.974**	**0.972**	**0.973**	**0.972**	**0.974**	**0.976**	**0.974**	**0.975**
MAE	NIR	0.543	0.536	0.530	0.520	0.504	0.512	0.524	0.547
	Vis	0.570	0.534	0.533	0.499	0.470	0.478	0.506	0.510
	Vis-NIR	**0.409**	**0.414**	**0.405**	**0.411**	**0.401**	**0.399**	**0.406**	**0.401**
MSE	NIR	0.638	0.580	0.617	0.552	0.506	0.510	0.554	0.623
	Vis	0.713	0.623	0.598	0.526	0.470	0.487	0.569	0.545
	Vis-NIR	**0.371**	**0.390**	**0.372**	**0.388**	**0.366**	**0.351**	**0.363**	**0.347**
RMSE	NIR	0.796	0.758	0.779	0.741	0.707	0.713	0.740	0.786
	Vis	0.842	0.788	0.770	0.723	0.682	0.696	0.752	0.736
	Vis-NIR	**0.607**	**0.623**	**0.607**	**0.619**	**0.601**	**0.590**	**0.600**	**0.586**
MAPE	NIR	19.560	19.044	19.052	18.472	17.935	18.342	18.939	19.379
	Vis	16.562	15.689	15.392	14.723	13.718	13.836	14.713	15.100
	Vis-NIR	**12.302**	**12.250**	**12.136**	**12.102**	**11.799**	**11.846**	**11.902**	**12.014**

duced with the addition of infrared lighting [6, 7]. Since blood absorbs infrared light, it then appears darker when compared to using Vis (white) light alone, as observed in the samples. The contrast between the blood and the background (unused sections of the gauze and the green platform) is more pronounced, essentially due to blood/water absorbing the IR light than the platform. Therefore, a small amount of blood or diluted blood (with high water content) absorbed in the gauze, that is hardly visible to the naked eye becomes more evident.

3.1 FET Results

We also evaluate the performance of local binary pattern (LBP), color moments (CM), statistical features (SF) from the power spectral density (PSD), and statistical features (SF) from the Fourier-transformed LBP (LBP-FFT) image used as the feature input. The features are extracted from at least one channel of an HSV image and concatenated to form one feature vector.

Both the experiments trained independently using the Vis (white) or the infrared wavelengths have achieved the best results with *CM + LBP + SA + PSD*, which are features that represent the color, texture, surface area, and spatial information in the image. The dual Vis-NIR configuration on the other hand has produced the best results with a similar feature set but with the addition of *LBP-FFT*.
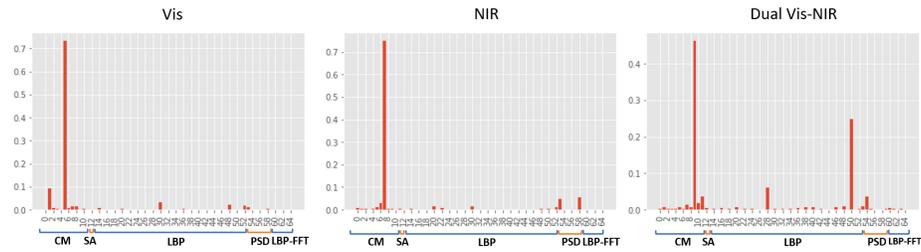
Fig. 6: Feature Importance Plots

While color representations are the primary indicators of blood detection and estimation, other features have also largely improved the estimations. Volume changes also result to change in the thickness of the blood which consequently influences the spectral variation [24]. However, based on our sample observations, volume changes also results to change in the pattern of blood spread. Spatial information then can also reveal useful information when estimating blood volume. This further highlights another contribution where we combine spectroscopy with imaging, thereby obtaining both spatial and spectral information in our gauze images.

In the Triton System approach, the infrared image is only used to generate depth that will be used as a mask to determine the surgical textile in an RGB image. But in our method, we utilize an RGB-IR image, taken under both white and IR lighting. Essentially, not only do we take advantage of the shape information, but also of the enhanced blood color due to the addition of the IR light. This claim is apparent in Figure 6 where we visualize the importance scores for each feature. The score provides how useful or valuable each feature was in the construction of the boosted decision trees within the model. The more a feature is used to make key decisions with decision trees, the higher its relative importance. The importance score is calculated explicitly for each feature, allowing the features to be ranked and compared to each other. From Figure 6, we plot the importance scores of each model independently trained on a specific wavelength. We cite a similar observation on all the wavelengths, i.e. the color moments (CM), which have highly contributed to the prediction. But, in the dual Vis-NIR setup, the model also takes advantage of texture and spatial features such as PSD - in contrast to Vis and NIR which focus on color as their primary indicator.

3.2   Machine Learning Algorithms Performance

We also evaluate the performance of other machine learning algorithms. As seen in Table 7, XGboost has outperformed other ML models on all the metrics. While we have shown only performances of the models trained on the dual Vis-NIR, we note that this result is consistent regardless of the wavelength and feature

combination used. Similarly, apparent in Figure 7, there's a substantial difference between XGboost and other models.
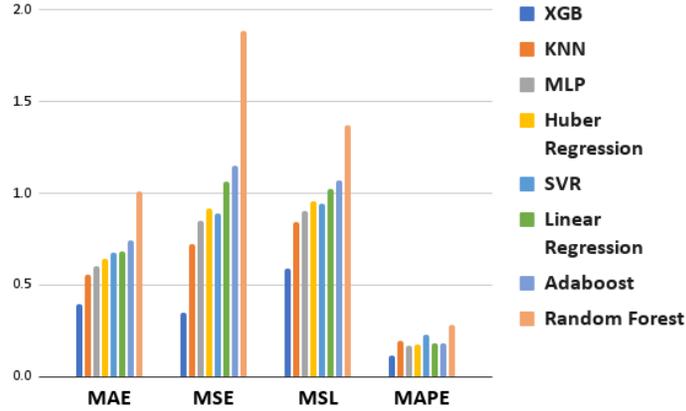

Fig. 7: Performance of Machine Learning Models using LBP+ CM+ PSD SF + LBP-FFT as Feature and dual Vis-NIR as the Selected Wavelength

Table 5: Performance of Machine Learning Models using LBP+ CM+ PSD + LBP-FFT as Feature and dual Vis-NIR as the Selected Wavelength

Model	R^2	MAE	MSE	RMSE	MAPE
XGB	**0.97526***	**0.39946***	**0.35082***	**0.59018***	**11.84578***
Adaboost	0.947	0.745	1.149	1.069	18.620
Random Forest	0.866	1.008	1.883	1.36	28.339
SVR	0.935	0.674	0.892	0.942	22.979
Linear Regression	0.932	0.683	1.061	1.024	18.198
Huber Regression	0.939	0.645	0.919	0.954	17.871
KNN	0.947	0.554	0.722	0.846	19.798
MLP	0.943	0.601	0.851	0.905	16.930

3.3 Sample Results

Lastly, we examine sample results using LBP+ CM+ PSD + LBP-FFT as features and XGBoost as predictor on the five (5) test sets generated using cross-validation. Each test set is comprised of at most 204 instances, with varying

weights from 0.5 to 10.5 from varying blood/gauze setups. The average difference between the actual and predicted test sets is -6.6182 g which signifies that the majority of test sets resulted in an underestimation.

# samples	204
Total (Actual)	747.7
Total (Predicted)	739.588
Difference	-8.112
% of Error	1.09%

(a) Test Set 1

# samples	204
Total (Actual)	794.0
Total (Predicted)	791.84
Difference	-2.16
% of Error	0.27%

(b) Test Set 2

# samples	203
Total (Actual)	811.1
Total (Predicted)	817.109
Difference	6.009
% of Error	0.74%

(c) Test Set 3

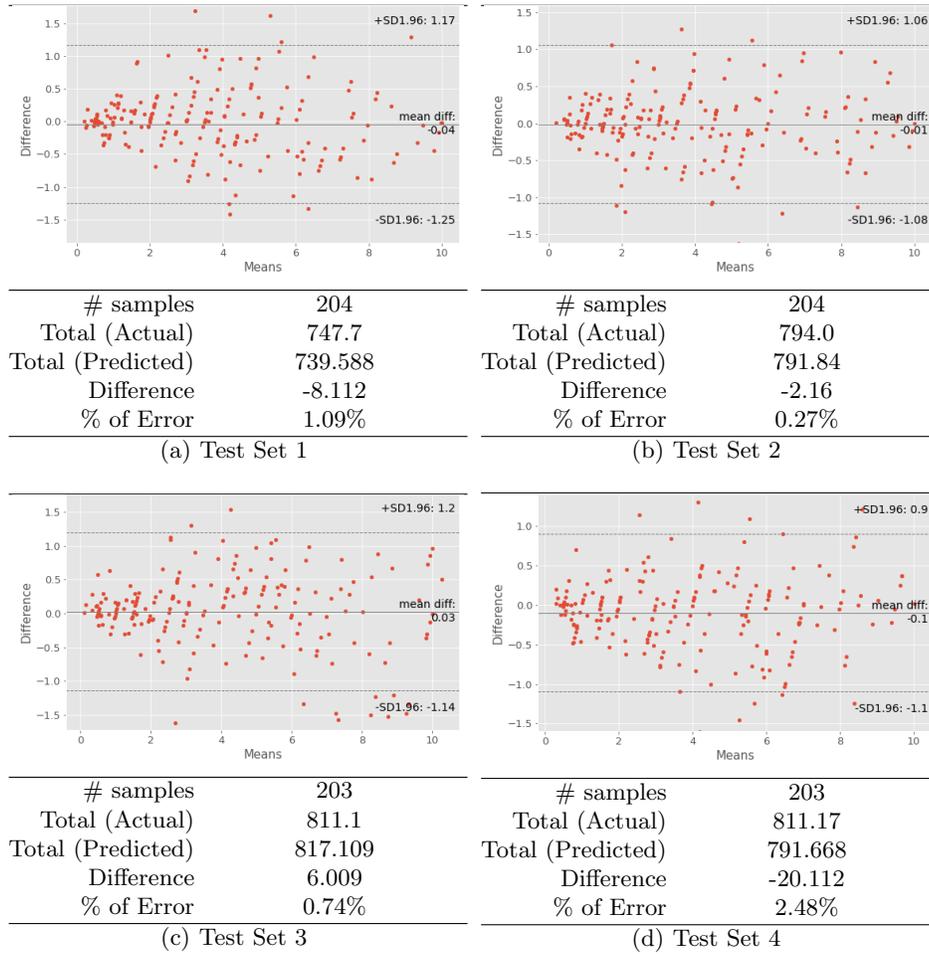| # samples	203
Total (Actual)	811.17
Total (Predicted)	791.668
Difference	-20.112
% of Error	2.48%

(d) Test Set 4

Fig. 8: Cross-validation bland alt-man plots Best Features (CM + LBP + SA + PSD + LBP-FFT) and Best ML Model (XGBoost)

The Bland-Altman plots of actual and the predicted volume (in g) per cross-validation subset are also shown in Figure 8. This plot is a graphical method to analyze the quality of the predictions based on a bias between the mean differences and an agreement interval, within which 95% of the differences of the predicted, compared to the actual, fall [10]. In Figure 8, *md* is the mean

of the differences, and the limits of agreements are expressed in terms of the standard deviation (sd) of those differences, plotted as md sd_limit *sd. We would expect most of the differences to lie between md 2sd, or more precisely, 95% of differences will be between md 1.96sd [2].

As observed, the test sets have a mean difference close to 0 and the limits are very narrow. The highest upper and lower limit of agreement so far is generated by Test Set 5 with md -0.05 + sd*1.12 and Test Set 1 with md -0.05 - sd*1.22, respectively. While the majority of the test sets have underestimated with an average of 1.25% error difference, Test Set 3 has resulted in an overestimation with a 0.74% difference between the actual and predicted total values.

As in previous studies, an acceptance criterion of 30 g of Hgb per case was set a priori as the clinically acceptable maximum bias. This difference represents approximately 5% of the total blood volume of an average adult (Hgb content of 250 mL [approximately 1/2 unit] of whole blood) [5]. Moreover, while we have recorded an average of 11.48% Mean Absolute Percentage (%) Error (MAPE) for all the test sets, it is still lower than the recorded underestimations of 46-75%, 40-49%, and 32% (using visual estimation) in several literatures [11]. Also, our result may be acceptable given that the maximum allowable error is 20% in some studies [22]. However, as our objective is to have results as close as possible to the standard method, the spectrophotometric analysis having a 10 % error, we still have to make re-adjustments with our model as well as data collection. Our results show that the blood loss estimations on the 4x8 gauze do not go over the acceptance threshold.

4 Conclusion

We have evaluated a dual-wavelength approach to estimating the blood volume absorbed in a 4x8 surgical gauze. Specifically, we compared the performances of classic feature extraction techniques and machine learning models trained with different lighting configurations. Results show that combining visible (white) light with infrared (IR) light outperforms the predictive capacities of machine learning models trained on images with just white or IR light.

While we have achieved a considerable performance, the proposed solution must be further validated with the addition of more datasets with varying setups. We want to also perform experiments further, e.g., combining features of images with different lighting configurations, fine-tuning the machine learning models, investigate possible overfitting and how to curb these, and optimizing the feature extraction process.

References

1. Obstetric   Hemorrhage,   https://www.gausssurgical.com/obstetric-hemorrhage/

2. statsmodels.graphics.agreement.mean_diff_plot   —   statsmodels, `https://www.statsmodels.org/dev/generated/statsmodels.graphics.agreement.mean_diff_plot.html`

3. US Patent for Method for estimating blood component quantities in surgical textiles Patent (Patent # 11,282,194 issued March 22, 2022) - Justia Patents Search, `https://patents.justia.com/patent/11282194`

4. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794. ACM, San Francisco California USA (Aug 2016). https://doi.org/10.1145/2939672.2939785, `https://dl.acm.org/doi/10.1145/2939672.2939785`

5. Doctorvaladan, S., Jelks, A., Hsieh, E.W., Thurer, R.L., Zakowski, M.I., Lagrew, D.C.: Accuracy of blood loss measurement during cesarean delivery. AJP Reports **7**, e93 – e100 (2017)

6. Edelman, G.J., van Leeuwen, T.G., Aalders, M.C.G.: Hyperspectral imaging of the crime scene for detection and identification of blood stains. p. 87430A. Baltimore, Maryland, USA (May 2013). https://doi.org/10.1117/12.2021509, `http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2021509`

7. Edelman, G.J., Roos, M., Bolck, A., Aalders, M.C.: Practical Implementation of Blood Stain Age Estimation Using Spectroscopy. IEEE Journal of Selected Topics in Quantum Electronics **22**(3), 415–421 (May 2016). https://doi.org/10.1109/JSTQE.2016.2536655, `http://ieeexplore.ieee.org/document/7422679/`

8. Freund, Y., Schapire, R.E.: A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. Journal of Computer and System Sciences **55**(1), 119–139 (Aug 1997). https://doi.org/10.1006/jcss.1997.1504, `https://linkinghub.elsevier.com/retrieve/pii/S002200009791504X`

9. Friedman, J.H.: Stochastic gradient boosting. Computational Statistics & Data Analysis **38**(4), 367–378 (Feb 2002). https://doi.org/10.1016/S0167-9473(01)00065-2, `https://linkinghub.elsevier.com/retrieve/pii/S0167947301000652`

10. Giavarina, D.: Understanding Bland Altman analysis. Biochemia Medica **25**(2), 141–151 (2015). https://doi.org/10.11613/BM.2015.015, `http://www.biochemia-medica.com/en/journal/25/2/10.11613/BM.2015.015`

11. Hancock, A., Weeks, A.D., Lavender, D.T.: Is accurate and reliable blood loss estimation the 'crucial step' in early detection of postpartum haemorrhage: an integrative review of the literature. BMC Pregnancy and Childbirth **15**(1), 230 (Dec 2015). https://doi.org/10.1186/s12884-015-0653-6, `http://bmcpregnancychildbirth.biomedcentral.com/articles/10.1186/s12884-015-0653-6`

12. Johar, R.S., Smith, R.P.: Assessing Gravimetric Estimation of Intraoperative Blood Loss. Journal of Gynecologic Surgery **9**(3), 151–154 (Jan 1993). https://doi.org/10.1089/gyn.1993.9.151, `http://www.liebertpub.com/doi/10.1089/gyn.1993.9.151`

13. Kollberg, S.E., Häggström, A.C.E., Lingehall, H.C., Olofsson, B.: Accuracy of Visually Estimated Blood Loss in Surgical Sponges by Members of the Surgical Team. Nurse Anesthesiology **87**(4) (Mar 2020)

14. Li, Y.J., Zhang, L.G., Zhi, H.Y., Zhong, K.H., He, W.Q., Chen, Y., Yang, Z.Y., Chen, L., Bai, X.H., Qin, X.L., Li, D.F., Wang, D.D., Gu, J.T., Ning, J.L.,

Lu, K.Z., Zhang, J., Xia, Z.Y., Chen, Y.W., Yi, B.: A better method for the dynamic, precise estimating of blood/ haemoglobin loss based on deep learning of artificial intelligence. Annals of Translational Medicine **8**(19), 1219–1219 (Oct 2020). https://doi.org/10.21037/atm-20-1806, `http://atm.amegroups.com/article/view/52195/html`

15. Liumbruno, G.M., Bennardello, F., Lattanzio, A., Piccoli, P.L., Rossetti, G.: Recommendations for the transfusion of red blood cells. Blood Transfusion (2008). https://doi.org/10.2450/2008.0020-08, `https://doi.org/10.2450/2008.0020-08`

16. Nowicki, P.D., Ndika, A., Kemppainen, J., Cassidy, J., Forness, M., Satish, S., Hassan, N.: Measurement of Intraoperative Blood Loss in Pediatric Orthopaedic Patients: Evaluation of a New Method. JAAOS: Global Research and Reviews **2**(5), e014 (May 2018). https://doi.org/10.5435/JAAOSGlobal-D-18-00014, `https://journals.lww.com/01979360-201805000-00002`

17. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition **29**(1), 51–59 (Jan 1996). https://doi.org/10.1016/0031-3203(95)00067-4, `https://linkinghub.elsevier.com/retrieve/pii/0031320395000674`

18. Pietikäinen, M., Zhao, G.: Two decades of local binary patterns. In: Advances in Independent Component Analysis and Learning Machines, pp. 175–210. Elsevier (2015). https://doi.org/10.1016/B978-0-12-802806-3.00009-9, `https://linkinghub.elsevier.com/retrieve/pii/B9780128028063000099`

19. Sharareh, B., Woolwine, S., Satish, S., Abraham, P., Schwarzkopf, R.: Real time intraoperative monitoring of blood loss with a novel tablet application. The open orthopaedics journal **9**, 422–6 (09 2015). https://doi.org/10.2174/1874325001509010422

20. Song, K.C., Yan, Y.H., Chen, W.H., Zhang, X.: Research and Perspective on Local Binary Pattern. Acta Automatica Sinica **39**(6), 730–744 (Jun 2013). https://doi.org/10.1016/S1874-1029(13)60051-8, `https://linkinghub.elsevier.com/retrieve/pii/S1874102913600518`

21. Stricker, M.A., Orengo, M.: Similarity of color images. p. 381. San Jose, CA (Mar 1995). https://doi.org/10.1117/12.205308, `http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.205308`

22. Sukprasert, M., Choktanasiri, W., Ayudhya, N.I.N., Promsonthi, P., O-Prasertsawat, P.: Increase accuracy of visual estimation of blood loss from education programme. Journal of the Medical Association of Thailand = Chotmaihet Thangphaet **89 Suppl 4**, S54–59 (Oct 2006)

23. Wu, S., Nagahashi, H.: Analysis of Generalization Ability for Different AdaBoost Variants Based on Classification and Regression Trees. Journal of Electrical and Computer Engineering **2015**, 1–17 (2015). https://doi.org/10.1155/2015/835357, `http://www.hindawi.com/journals/jece/2015/835357/`

24. Yang, J., Mathew, J.J., Dube, R.R., Messinger, D.W.: Spectral feature characterization methods for blood stain detection in crime scene backgrounds. p. 98400E. Baltimore, Maryland, United States (May 2016). https://doi.org/10.1117/12.2224099, `http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2224099`