2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

The Use of Mobile Thermal Imaging and Deep Learning for Prediction of Surgical Site Infection

Richard Ribón Fletcher, *Member IEEE*, Gabriel Schneider, Laban Bikorimana, Gilbert Rukundo, Anne Niyigena, Elizabeth Miranda, Robert Riviello, Fredrick Kateera, Bethany Hedt-Gauthier

***Abstract*— The ability to detect surgical site infections (SSI) is a critical need for healthcare worldwide, but is especially important in low-income countries, where there is limited access to health facilities and trained clinical staff. In this paper, we present a new method of predicting SSI using a thermal image collected with a smart phone. Machine learning algorithms were developed using images collected as part of a clinical study that included 530 women in rural Rwanda who underwent cesarean section surgery. Thermal images were collected approximately 10 days after surgery, in conjunction with an examination by a trained doctor to determine the status of the wound (infected or not). Of the 530 women, 30 were found to have infected wounds. The data were used to develop two Convolutional Neural Net (CNN) models, with special care taken to avoid overfitting and address the problem of class imbalance in binary classification. The first model, a 6-layer naïve CNN model, demonstrated a median accuracy of AUC=0.84 with sensitivity=71% and specificity=87%. The transfer learning CNN model demonstrated a median accuracy of AUC=0.90 with sensitivity =95% and specificity=84%. To our knowledge, this is the first successful demonstration of a machine learning algorithm to predict surgical infection using thermal images alone.**

***Clinical Relevance*— This work establishes a promising new method for automated detection of surgical site infection.**

I. Introduction and Motivation

A. The Burden and Challenge of Surgical Site Infections

Infections are a critical health concern, which can lead to sepsis, organ failure, or even death. The infection of surgical wounds, also known as surgical site infection (SSI), occurs worldwide. In higher-income countries such as the United States, such infections are responsible for costly readmissions and continued treatment of the patient, accounting for 21.8% of hospital infections and 0.6% of all hospital deaths [1]. In the US, infections occur in approximately 3.5% of all patients undergoing surgical procedures [2].

However, the rate of surgical site infection (SSI) is considerably higher in low-income countries, exacerbated by challenges of shortage of trained clinical staff, underdeveloped healthcare infrastructure, and anti-microbial resistance. In Rwanda, which is the site of this study, 11% of women who have Cesarean section births developed an infection [3].

Research supported by funding from the National Institutes of Health (No R21EB022368) and the Harvard Center for Global Health-Dubai;

R. R. Fletcher and G. Schneider are with the Massachusetts Institute of Technology, 77 Massachusetts, Cambridge, MA. 02139 USA. (phone: 617-694-1428, e-mail: fletcher@media.mit.edu).

B. Hedt-Gauthier, R. Riviello, and L. Miranda are with Harvard Medical School in Boston, MA.

F. Kateera, L. Bikorimana, and A. Niyigena are with Partners In Health, Kigali Rwanda.

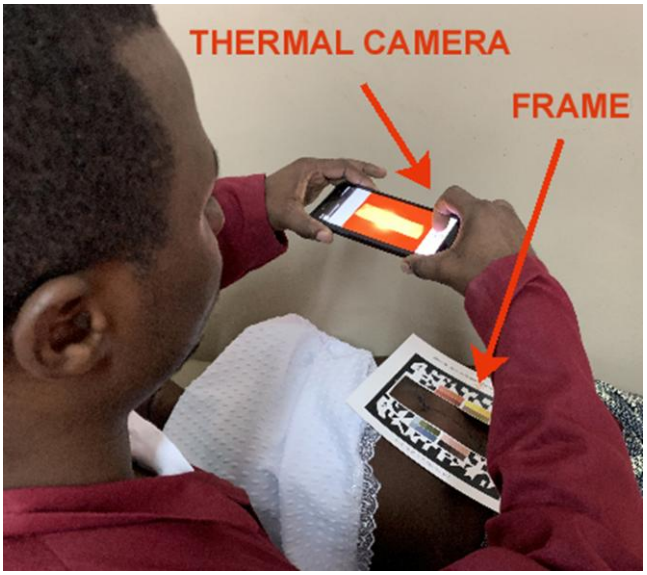

Fig. 1. Health worker standing over surgery patient using a thermal camera module and smart phone to capture an image of the surgical wound. The paper frame target is used for alignment

B. Development of Tools for Detecting Infection

Until recently, the detection of infection has been the responsibility of trained clinical staff, such as doctors and nurses in hospitals. These methods generally rely on subjective indicators including heat, erythema (redness), swelling, and pain. In recent decades, these methods have been improved with better clinical guidelines [4] and quantitative approaches, such as the ASEPSIS score [5], which takes into account the fluid discharge, color, odor, and separation of the wound tissues. While ASEPSIS provides a useful framework with which to assess possible infection, this method requires clinical experience which can be challenging to access in low-resource areas.

The need for better outpatient services has motivated interest in creating new electronic tools to help surgery patients take care of their wounds and monitor the healing process at home. This technology is also of interest to non-surgical patients who have *chronic* wounds or ulcers, which can result from vascular disease, diabetes, and numerous other health conditions [6].

With the growing prevalence of smart phones, mobile applications, such as Tissue Analytics and the Mobile Post-Operative Wound Evaluator (mPOWER), have been developed to help measure and monitor wounds using a mobile phone image of the wound. However, none of these tools claim to detect infection [7, 8].

Most recently, machine learning methods have been developed by our group and others to attempt prediction of infection using a color mobile phone image of the wound [9,

10], which could be of great value to both community health workers in low resource areas as well as outpatients in wealthier communities around the world.

C. Racial Bias and Generalizability

With emerging use of artificial intelligence and machine learning in medicine, a growing concern within the health research community are the issues of fairness and bias in these algorithms [11]. While the use of mobile phone color images is a promising and useful approach, it is also quite likely that an infection prediction algorithm that depends on color may not work well among racially diverse patient populations having a range of skin colors. For this reason, we present in this paper a new method of predicting surgical site infection making use of thermal imaging alone, which is not impacted by the color of the patient's skin.

II. Clinical Study

A. Study Design

For algorithm development, image data was collected as part of a study led by Harvard Medical School and Partners in Health, in Kigali, Rwanda. In this study, all women who underwent cesarean section at Kirehe District Hospital between September 2019 through February 2020 were prospectively enrolled on postoperative day 1 (POD1). On POD11 (+/- 3 days), patients returned to the hospital for follow-up; a general practitioner then performed a physical exam and inspection of the wound to determine whether or not the patient had a surgical site infection (SSI). Both visible (RGB) and thermal images were collected in our study; however for this paper, we focus on the thermal images, which included 530 women, yielding images of 30 infected wounds and 500 non-infected wounds.

Our clinical study was approved by the Institutional Review Boards (IRBs) of Harvard Medical School, MIT and Kirehe district hospital.

B. Thermal Image Data Collection and Labelling

Images were captured using a Samsung Galaxy J8 smart phone and a low-cost (US$200) thermal camera module (SEEK Thermal Compact). The thermal camera connects to the mobile phone and is also electrically powered from the mobile phone. We developed a custom Android application using the Android SDK for the SEEK Thermal camera to enable the collection of a JPG thermal image of the wound, as well as a separate 2D temperature array. The JPG thermal image is used for real-time image feedback to the user. However, since the dynamic range of each image pixel is limited to only 8-bits, we used the 2D temperature array to create a thermal pseudo-image having 0.1 °C temperature precision per pixel, which translates to roughly 10-bits of precision.

To avoid the need to manually edit the photographs in the data processing stage, we designed a custom mobile app, called *WoundScreener*, with the Vuforia augmented reality library and computer vision software that automatically aligned, rectified, and cropped the image at the time of data collection. As shown in Figure 1, a thick paper frame with a printed "target" pattern and color chart was placed over the surgical wound which enabled the computer vision software to automatically track the image in 3D and correct for parallax distortion. For the thermal camera image, a strip of copper tape served as an optical marker that was used to align the image. Sample visible and thermal images, captured on the same woman, are shown in Figure 2.

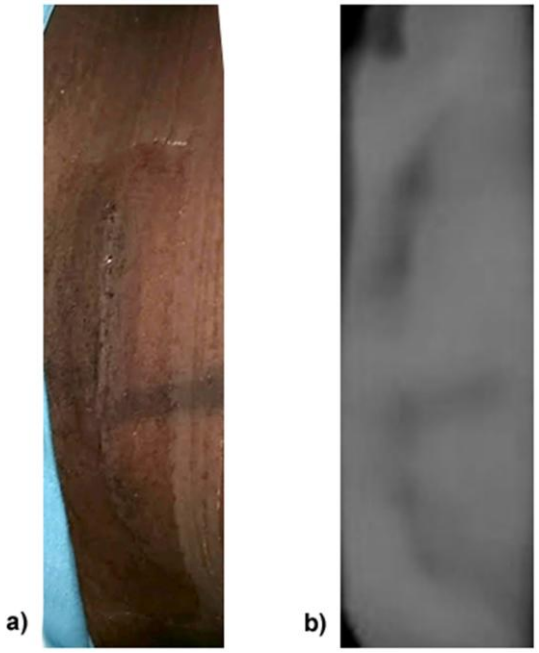

Fig. 2. Sample images taken of the same wound: a) standard RGB phone camera image and b) thermal camera image.

For the purpose of this short paper, we focus on the algorithm development using the thermal imaging data.

C. Data Labelling and Preprocessing

As mentioned above, the SSI diagnosis was performed by the doctor, and this assessment was used to label all the data for the purpose of conducting supervised machine learning.

Starting with the raw thermal image of 206x156 pixels, we then applied a custom computer vision script, with a bounding box algorithm and polygon model to automatically find the outline of each wound frame and crop the remainder of the image. The resulting wound thermal images were standardized to 160x100 pixels (width x height).

In order to reduce overfitting and improve generalizability, we also added a random image rotation, image flip, and batch normalization as a pre-processing step for our CNN model testing and optimization.

III. Naïve CNN Model

A. Algorithm Design and Implementation

As our initial model, we developed a few-layer naïve convolutional neural net (CNN) model. For development, we used the Keras Tensorflow library to compile and train the CNN models in Python. The architecture consisted of 3 convolutional layers with ReLU activation layers added between the convolutional layers, followed by a pooling layer. For reducing overfitting, and improve generalization, we also used a node dropout rate of 0.5, with a final fully-connected binary layer with sigmoid activation—giving us the labels "0" for non-infected or "1" for infected wound images.

B. Addressing Class Imbalance

Given the relatively low proportion (1:9) of infected images, it was necessary to implement methods to address the significant class imbalance. Without addressing this issue, the standard machine learning optimization would tend to produce a model that would trivially predict all images as non-infected, which would yield a fairly high classification accuracy, but would result in a very low sensitivity (true positive rate).

The methods we explored are described below:

Creating a custom cost/loss function: We first created a custom loss function for optimization, which increased the cost of false negative classification. This custom loss function was effective in increasing the number of true positives. However, the custom loss function negatively impacted the rate of convergence and significantly increased the time to train the model, so the following two methods were adopted instead.

Data Synthesis: A second method we explored to address class imbalance was to expand the minority class and reduce class imbalance. Although we found this method to be very effective at increasing our true positive rate, we limited the maximum amount of data synthesis to 2X, which is generally recommended to avoid overfitting.

Class Weights: A further method of compensating for class imbalance was adjustment of the class weight inside the model. We used the Keras library to adjust the weights and explored a range of weight ratios ranging from 5 to 15.

C. Model Training

For model training, three separate optimization algorithms were evaluated (Stochastic Gradient Descent (SGD), Adam, and RMSprop) using a range of training epochs, from 30 and 50, which was sufficient to achieve convergence.

As part of our parameter optimization, several hundred different combinations of weights, optimizer functions, and training epochs were evaluated. For each combination of parameters, we trained a separate model using k-fold cross validation. For this task, we split our input data into sizes 1/k and (k-1)/k for validation and training sets, respectively. We explored both k = 5 splits and 10 splits, which corresponds to13 and 6 infected images, respectively. As expected, the 5-fold validation model did a better job of generalizing to more data and producing smoother ROC curves, with less overfitting than the 10–split validation model.

D. Model Optimization and Prediction Results

For each model created in the parameter search, a ROC curve was generated as well as the standard metrics of Area Under the Curve (AUC), sensitivity and specificity. We also calculated the Matthews correlation coefficient (MCC), which is a preferred metric for binary classification that tracks false negatives as well as false positives.

Since the optimization of our model depends on the chosen objective of our project, it is important to state that our objective was to create a screening tool for surgical site infection. As a general criteria for screening tools, we placed higher importance on *sensitivity* rather than *specificity*, and rejected any model with sensitivity<0.7. In the context of global health, this implies that our model would detect as many infections as possible, at the expense of allowing more false positives.

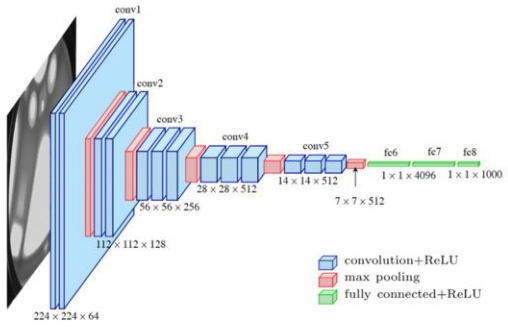

Fig. 3. The architecture of the ResNet50 neural net. Notice the many convolutions and fully-connected layers.

Using these criteria, we selected the best performing model, which is shown in Figure 4, including the ROC curve for all five folds, with median AUC=0.84, sensitivity=0.71, specificity = 0.87, and MCC = 0.37, as summarized in Table 1. The optimal class weight ratio was 15:1 with the Adam optimizer function.

IV. Transfer Learning Model

A. Algorithm Design and Implementation

Our next approach was to create a CNN model using *transfer learning*, which makes use of a pre-trained, highly sensitive neural network commonly used for image classification. For this task, we used the ResNet50 model, which has a fairly complex architecture (Fig 3), with many convolutions and transformations. Since the input to the transfer learning model required a specific size image, we implemented preprocessing to scale the image pixel size to 224 x 224, which also helped to reduce the training time.

As the final layer, we added a dense layer to classify the output images into two classes, with a sigmoid activation. To reduce overfitting, a node dropout rate of 20% was also used. As in the previous naïve CNN model, a randomized flip, random rotation, and batch normalization, was also used in the image pre-processing step.

B. Class Imbalance and Model Training

For this transfer learning model, we addressed the class imbalance using the same methods as described previously. For model training, we again used 5-fold cross validation, and once again did a broad parameter search, with several hundred parameter combinations, varying the class weights, optimizer function and number of epochs.

C. Prediction Results from Transfer Learning

The resulting best transfer learning model is shown in Figure 5, in terms of the ROC curve for all 5 folds, with mean AUC = 0.90, sensitivity=0.95, specificity = 0.84, and MCC=0.44. the optimal class weight ratio was 15:1, with the RMS-Prop optimizer function.

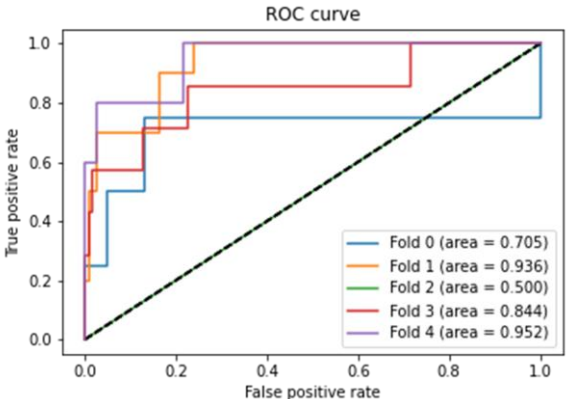

Fig. 4. Representative folds for naïve CNN infection model with a median AUC= 0.84.

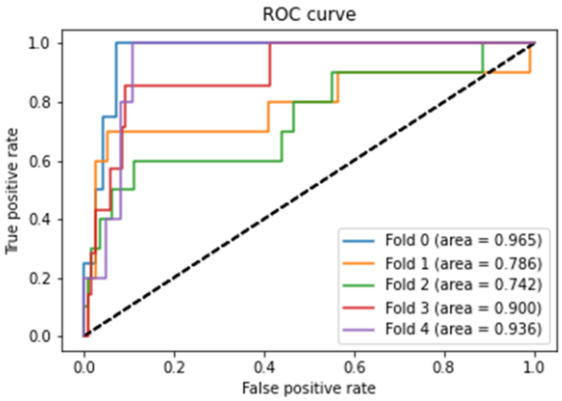

Fig.5. Representative folds for transfer learning CNN model with a median AUC= 0.90.

V. Discussion

Both the naïve CNN and transfer learning CNN produced reasonably good results for the prediction of infection. We have taken care to use conservative methods to minimize overfitting and to address the issue of class imbalance. However, additional testing is needed using a larger dataset in order to verify these results.

Additional work is also being done to visualize the activation layers of the CNN model in order to explore what parts of the wound contribute the most to the infection prediction. Based on our preliminary analysis, it seems that the image pixels near the wound incision have a large influence on the prediction; this is consistent with our knowledge that SSI's most commonly begin at the site of the incision, where bacteria have been introduced to the tissue. Since we know that the progression of the infection and the severity of infection can depend on the specific type of bacteria that caused the infection, other variables are being examined as well.

VI. Conclusion

Based on our results from 530 surgical C-section wounds (30 infected), we have successfully developed two different deep learning models to predict surgical site infection. To the best of our knowledge, this is the first reported successful demonstration of using a machine learning algorithm to perform a prediction of SSI using thermal images alone. While these results are very encouraging, it is prudent to verify these results with additional data.

Nevertheless, the availability of low-cost (US$200) thermal imaging camera modules for smart phones, now provide a new opportunity to help identify wound infection without a dependence on skin color or racial differences.

Model	Max. Accuracy	Sensitivity	Specificity	Median AUC (IQR)
Naïve CNN	0.86	0.71	0.87	0.844 (0.34)
Transfer Learning CNN	0.94	0.84	0.95	0.900 (0.19)

TABLE I.

VII. Acknowledgements

We would like to acknowledge Rich Redemske for development of the augmented reality WoundScreener mobile app, and Kaveri Nadhumani for early work with image analysis. We would also like to acknowledge the great work of the field staff in Rwanda that enabled the collection of patient data, and we would like to thank the patients themselves who contributed their data to this study.

References

[1] de Lissovoy, G., et al., "Surgical site infection: Incidence and impact on hospital utilization and treatment costs". Am J Infect Control, 37(5): (2009): 387-97.

[2] Berríos-Torres, S.I., Umscheid, C.A., Bratzler, D.W., Leas, B., Stone, E.C., Kelz, R.R., Reinke, C.E., Morgan, S., Solomkin, J.S., Mazuski, J.E. and Dellinger, E.P., 2017. Centers for disease control and prevention guideline for the prevention of surgical site infection, 2017. *JAMA surgery*, *152*(8), pp.784-791.

[3] Nkurunziza, T., Kateera, F., Sonderman, K., Gruendl, M., et al., 2019. Prevalence and predictors of surgical-site infection after caesarean section at a rural district hospital in Rwanda. *BJS,106*(2), pp.e121-e128.

[4] Horan, T.C., Andrus, M. and Dudeck, M.A., 2008. CDC/NHSN surveillance definition of health care–associated infection and criteria for specific types of infections in the acute care setting.*American journal of infection control*, *36*(5), pp.309-332.

[5] Wilson, A.P.R., Weavill, C., Burridge, J. and Kelsey, M.C., 1990. The use of the wound scoring method 'ASEPSIS'in postoperative wound surveillance. *Journal of Hospital Infection*,*16*(4), pp.297-309.

[6] Frykberg, R.G. and Banks, J., 2015. Challenges in the treatment of chronic wounds. *Advances in wound care*, *4*(9), pp.560-582.

[7] https://www.tissue-analytics.com

[8] Sood, R.F., Wright, A.S., Nilsen, H., Whitney, J.D., Lober, W.B. and Evans, H.L., 2017. Use of the Mobile Post-Operative Wound Evaluator in the Management of Deep Surgical Site Infection after Abdominal Wall Reconstruction. *Surgical Infections Case Reports*, *2*(1), pp.80-84.

[9] Wu, J.M., Tsai, C.J., Ho, T.W., Lai, F., Tai, H.C. and Lin, M.T., 2020. A Unified Framework for Automatic Detection of Wound Infection with Artificial Intelligence. *Applied Sciences*,*10*(15), p.5353.

[10] Fletcher, R.R., Olubeko, O., Sonthalia, H., Kateera, F., Nkurunziza, T., Ashby, J.L., Riviello, R. and Hedt-Gauthier, B., 2019, July. Application of machine learning to prediction of surgical site infection. *IEEE Engineering in Medicine and Biology Society Conference (EMBC) 2019*.

[11] Fletcher, R.R., Nakeshimana, A. and Olubeko, O., 2020. Addressing Fairness, Bias, and Appropriate Use of Artificial Intelligence and Machine Learning in Global Health. *Frontiers in Artificial Intelligence*, *3*, p.116.