Deep Learning Technology for Drunks Detection with Infrared Camera

Pisit Iamudomchai
Biomedical Engineering Department Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
Email: Pisit.bme@gmail.com

Pattanawadee Seelaso
Biomedical Engineering Department Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok, Thailand
Email: Pattanawadee.sls@gmail.com

Satjana Pattanasak
Faculty of Science and Technology,
Valaya Alongkorn Rajabhat University
13180, Thailand
Email: satjanap@hotmail.com

Wibool Piyawattanametha
Biomedical Engineering Department Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang
Bangkok 10520,Thailand
Institute for Quantitative for Health Sciences and
Engineering Michigan State University
East Lansing, MI 48824, USA
Email: wibool@gmail.com

Abstract— **This project manipulates for the purpose to solve the problem of alcohol measurement delays and reduces the issue of contamination that arises from the breath analyzer test. This project focuses on designing a novel infrared (IR) camera-based alcohol detection system with deep learning technology. This system consists of 2 parts. The first part is an infrared camera (FLIR) used for collecting both IR and normal images then the next part is an image processing system for alcohol detection based on deep learning technology operating on an iPhone operating system (iOS) mobile phone. Our handheld IR based detection system achieves an accuracy of 85.10% (135 population) accuracy with 4 levels of classification (sober, 1 glass, 2 glasses, or 3 glasses) and 74.07% with binary identification (Sober or Drunk). Each glass contains 200 ml of beer (5% vol).**

Keywords— Alcohol measurement, Infrared Camera, Deep Learning.

I. Introduction

Most road accidents are often seen. Regularly caused by drunk driving. There is a lot of sufferer either property and life. And they're also causing a lot of damage even other people that are not related to them. The legal of blood alcohol limits for driving in Thailand, which stipulates a blood alcohol concentration value of 0.50 mg%. One of the best methods to curb this accident is to perform an alcohol measurement on drivers on the streets. Typically, alcohol measurements are performed by 2 methods. The first method is performed via a collection of blood to measure alcohol content. This method is the most reliable method and accurate but it is a highly invasive, time-consuming, and expensive price. Next method. Using the blood alcohol level detector by the Breath Analyzer Test is the most popular method at present. Which have the advantages of being fast output, portable, easy, and cheap cost. However, a drawback to this method is a person needs to blow through a plastic mouthpiece each time the measurement is done which is an additional cost of buying a plastic mouthpiece and not environmentally friendly. Most importantly, drivers sometimes lack consciousness, so they try to avoid and refuse to cooperate in alcohol measurement via a breath analyzer. then the resulting is delays.

Based on the above problems, we research on how alcohol affects the human body. We found the temperature of the body is a significant change. Skin blood flow and chest sweat rate in the alcohol session increased significantly after 10 minutes of drinking.[1] Due to skin blood flow and thermal is increased we prefer the infrared camera in this invention. We take a photo of a subject's face, which is probably the most exposed part of the body and any changes of vascularity are readily visible. [2] We also found some research using a thermal image to identify the drunk person [3]

we invented the innovative alcohol level detection with infrared cameras. The work of such innovations consists of receiving images from infrared cameras. And send the image data to the server for processing with Deep learning, save and display via the application. This makes our research have a fast working system, no contamination from the driver. Because we use infrared cameras to store images. So, the driver can't evade. And can shoot even in the dark.

II. Background

A. Infrared Radiation (IR)

Infrared radiation (IR), sometimes called infrared light, is electromagnetic radiation (EMR) with wavelengths longer than those of visible light. Infrared radiation extends from the nominal red edge of the visible spectrum at 700 nanometers (nm) to 1 millimeter (mm) as illustrated in Fig. 1.

Thermal cameras are passive sensors that capture the infrared radiation emitted by all objects with a temperature above absolute zero. Deploying this type of sensor in vision systems eliminates the illumination problems of normal greyscale and RGB cameras[4].

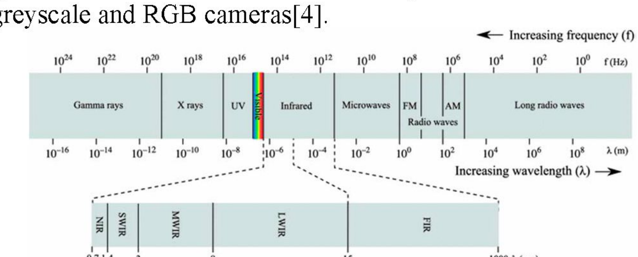

Fig. 1 The electromagnetic spectrum with a sub-divided infrared spectrum [4].

978-1-7281-9978-8/20/$31.00 ©2020 IEEE

B. *Facial analysis*

The human face has a distinctive feature that is any changes in vascularity are visible [2]. Since the face is normally not covered by clothes, a thermal camera can capture the direct skin temperature of the face [5].Face recognition using thermal cameras eliminates the effects of illumination changes and eases the segmentation step, but it can also introduce some challenges due to the different heat patterns of a subject, caused by different activity levels or emotions such as anxiety.

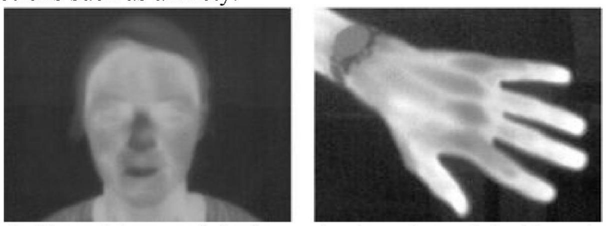

Fig. 2 Thermal images of the face or hand can be used for biometric identification [4].

III. METHOD

The drunks detection system with an infrared camera consists of two processes. The first process is collecting image data from samples. All the infrared images used in this work were acquired by the thermal cameras for smartphones FLIR ONE The thermal images have a resolution 160 x 120 px and Visual resolution 1440 x 1080 px with 17 µm thermal pixel size. The operating wavelengths are from 8-15 µm. This is the bandwidth of most of the heat emitted by radiation from humans and the environment around us. The infrared camera is connected to the iPhone operating system. In the second step Thermal images will be used in the application. Then the classification will be distinguished with a deep learning model. Finally, the results will appear on the screen.

Fig. 3 The infrared camera with an iPhone operating system.

A. *Collecting data*

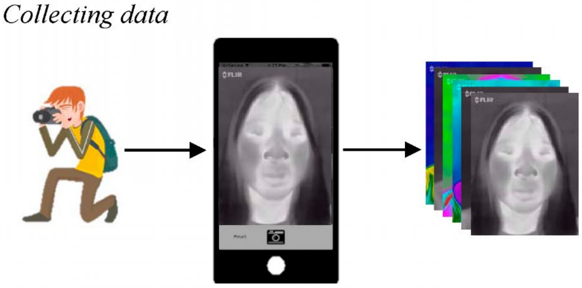

Fig. 4 Collecting image.

We collect image data from infrared cameras for training, with 50-unisex volunteers that are Thai nationality. Each volunteer drinking 200 mL for 3 glasses of beer. Take pictures of volunteers before drinking. After drinking the first glass of beer for 15 minutes and then photographed. Then repeat all process 3 times. So, each volunteer will get 4 pictures, no drinking, drinking 1 glass, drinking 2 glasses, and drinking 3 glasses in Fig. 4.

Our infrared camera has a lot of filters such as coldest hottest contrast and greyscale shown in Fig. 5. We choose only the filter which represents an obvious feature and gives us the highest accuracy. We choose the contrast filter in this work, an example of the pictures set shown in Fig. 4.

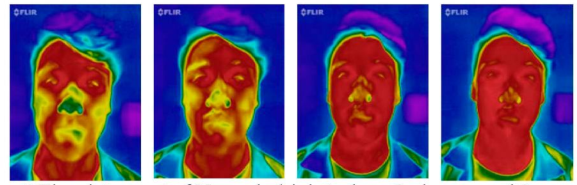

Fig. 4 The picture set of Normal, drink 1 glass, 2 glasses, and 3 glasses, respectively.

Fig. 5 The example of filter contrast greyscale and antarctic, respectively.

B. *Data Preprocessing*

All data set is rearranged and sorted. First, we have duplicated all the pictures and separate them into 2 groups that are binary classification and 4 levels of classifications. From then, we label both, for the first group (binary classification), label with Sober and Drunk, while the second group (4 levels of classifications) was labeled with Normal, drink 1 glass, 2 glasses, and 3 glasses. The process diagram is demonstrated as Fig. 6.

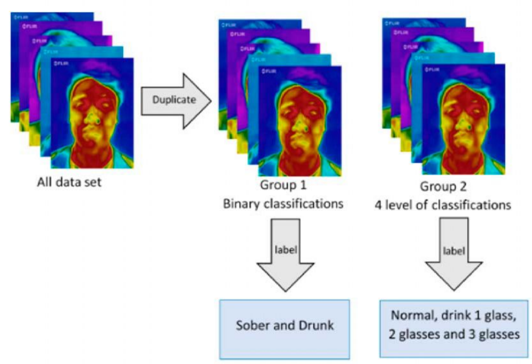

Fig. 6 The data preprocessing diagram.

C. *Deep learning*

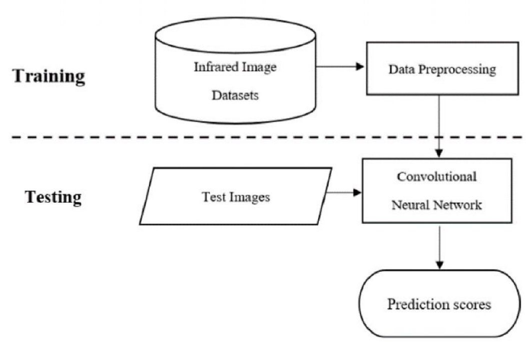

Fig. 7 Overview of the proposed method using Deep Learning for classifying infrared image datasets.

In the part of Deep learning. There are include a training and testing stage, as shown in Fig. 7. In the training stage, data processing is applied to the infrared image datasets. Next, we bring the pre-trained network parameters from the example model, these parameters are used to initialize a new ConvNet. In the testing stage, the preprocessed testing images are fed into the fine-tuned ConvNet. The prediction score is obtained by aggregating the ConvNet's output values. Further details are described below.

- Data Preprocessing

Add the labels to the images based on the image name. Then divided all image datasets separate into 4 groups are normal, drink 1 glass of beer, drink 2 glasses of beer, and 3 glasses of beer. Next, we reshape images into 128*128 because this needs to be flattened to be passed through the convolution layer.

- Convolutional Neural Networks

Convolutional Neural Networks are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product, and optionally follows it with a non-linearity. The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. And they still have a loss function (e.g. SVM/Softmax) on the last (fully connected) layer and all the tips/tricks we developed for learning regular Neural Networks still apply.

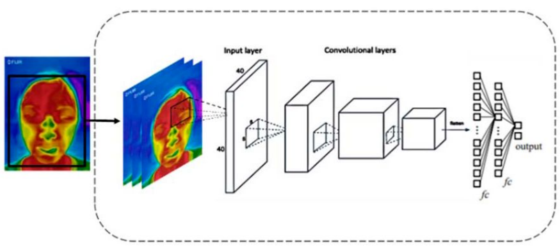

Fig. 8 Convolutional Neural Networks.

- Convolutional layer

The convolution layer takes local rectangular patches across (with offset by stride and with/without spatial preservation by padding) the input image (for the first layer) or feature maps (for the sub-sequent layers) as input, on which 2D convolution with a filter is performed. The sum x of the resulting convolutions is fed into a non-linearity function, specifically a rectified linear unit (ReLu) $f(x) = \max(0,x)$, to increase the speed of training. In a given layer, the same filter is shared in a feature map, while different filters are used for different feature maps. This property of filter sharing in the convolution layer allows for detecting the same pattern in different locations of the feature map.

- Pooling layer

The pooling operation down-samples the feature map by summarizing feature responses in each non-overlapping local patch, often by computing the maximum activations (max-pooling). This yield features invariant to minor translations in the data.

- Fully connected layer

convolution and pool generate feature maps of smaller dimensions than the input image, which are then passed through several function layers. The first few function layers fuse these feature maps into a feature vector. The last function layer contains two neurons that compute the classification probability for each class using softmax regression. To reduce overfitting, "dropout" [6] is used to constrain the fully-connected layers.

- Testing

To classify an unseen image, we combine random-view aggregation and multiple crop testing [7] to produce the final prediction score.

IV. Results

From collecting images from Infrared camera. Get approximately 200 pictures were received from a total of 50 volunteers. Each volunteer had 4 pictures, normal, drink 1 glass, 2 glasses, and 3 glasses. All pictures that used in this work use contrast filter as shown in Fig. 4.

The results from the deep learning model will be shown below in the model's accuracy. For this accuracy test we use a set of test images that have 135 images that include both train and never train the model.

TABLE 1. The accuracy and image size of each model for 4 levels of classification.

Models	Image Size	Accuracy
NasnetMobile	224,224	85.10 %
EfficientNet-B1	240,240	74.71 %
mobilenetV2	224,224	73.88 %
InceptionV4	299,299	71.85 %

For 4 levels of classifications, the top 4 highest accuracies were shown in table 1. The model with the highest accuracy is NasnetMobile with 85.10 %. The confusion matrix is shown in Fig. 9. And some class statistics in Table 2.

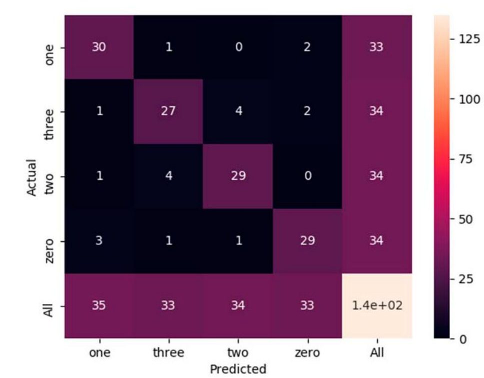

Fig 9 The confusion matric of NasnetMobile with 135 population.

TABLE 2. some class statistics of NasnetMobile.

Classes	Zero	One	Two	Three
True Positive	29	30	29	27
True Negative	97	97	96	95
False Positive	4	5	5	6
False Negative	5	3	5	7
Sensitivity	85.29%	90.90%	85.29%	79.41%
Specificity	96.03%	95.09%	95.04%	94.05%
Accuracy	93.33%	94.07%	92.59%	90.37%

The top 4 highest accuracies of binary classification were shown in table 3.

TABLE 3. The accuracy and image size of each model for binary classification.

Models	Image Size	Accuracy
MobileNet	224,224	74.07%
EfficientNet-B1	240,240	72.59%
ResnetV2_50	224,224	71.11%
InceptionV4	299,299	60.74%

In binary classification the sober class is not drinking, and the drunk class is drinking 1 glass of beer, 2 glasses, and 3 glasses. The model with the highest accuracy is Mobilenet with 74.07%. The confusion matric showed in Fig. 10. And some class statistics in table 4.

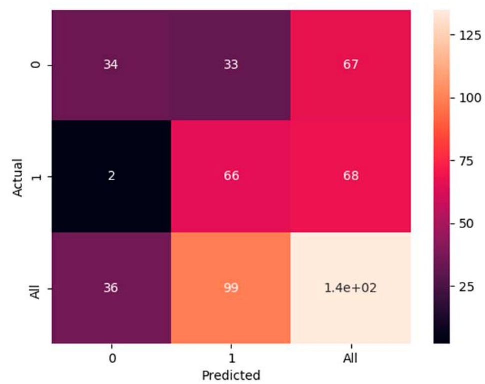

Fig. 10. The confusion matric of MobileNet with 135 population (0 as Sober and 1 as Drunk).

TABLE 4. some class statistics of MobileNet.

Class	
True Positive	66
True Negative	34
False Positive	33
False Negative	2
Sensitivity	97.05%
Specificity	50.74%
Accuracy	74.07%

This binary classification model has a 74.07% accuracy which has a very unsatisfactory result, the model still cannot accurately predict. Because we have little pictures to train models and continue to adjust the parameters accordingly. But if to predict 4 levels classification the model can be classified with an accuracy of 85.10%. From the results, we are still studying and continuing to work. To be able to identify drunken people because it would lead to noninvasive systems which can be beneficial to society.

V. Discussion and Conclusions

In Binary classification has a dissatisfy accuracy. If we look at table 4, we can observe that the positive test (Drunk) have a high value 99 (66+33), while the negative test (Sober) have a lower value 36 (34+2) which is because the model of binary classification distributes between drinking alcohol and not drinking alcohol. The data set that is trained is divided into not drinking (0 glass) and drinking (1 glass, 2 glasses, and 3 glasses). Therefore, the classification of dunks people is better because there is a lot of data taken to train the model

After experimenting and inventing this work, we found that our work has the following restriction. First, Deep Learning Technology for Drunks Detection with Infrared Camera which does not work like a standard alcohol analyzer Breath analyzer or invasive method Test. Our invention can predict how many glasses that subject to drink or not drink at all. Next our invention based on the infrared camera, which involves the heat radiation emitted by the subject. Therefore, using this tool may not be effective if the subject is unusual body temperature such as fever or exercise. Finally, from our dataset, we have observed that normal people generally do not have much difference between people who drink 1 glass, 2 glasses, 3 glasses, and do not drink alcohol. Which is seen as shown in Fig.4 is the people who are alcohol flush, Therefore, if we have more picture, it may be better to identify drunk people by extracting the specific features that are obvious to classify.

Acknowledgment

This study is supported in part grants from KMITL Research Fund; Thailand Science Research and Innovation; National Research Council of Thailand; Newton Fund from British Council, UK; and the Fraunhofer-Bessel Research Award 2015 from the Alexander von Humboldt Foundation, Germany.

References

[1] T. Yoda, et al., *Effects of alcohol on thermoregulation during mild heat exposure in humans.* Alcohol, 2005. 36(3): p. 195-200.

[2] T. von Arx, et al., *The Face – A Vascular Perspective. A literature review.* Swiss dental journal, 2018. 128: p. 382-392.

[3] G. Koukiou, G. Panagopoulos, and V. Anastassopoulos. *Drunk person identification using thermal infrared images.* in *2009 16th International Conference on Digital Signal Processing*. 2009.

[4] R. Gade and T. B. Moeslund, *Thermal cameras and applications: a survey.* Mach. Vision Appl., 2014. 25(1): p. 245–262.

[5] J. Mekyska, V. Espinosa-Duró and M. Faundez-Zanuy. *Face segmentation: A comparison between visible and thermal images.* in *44th Annual 2010 IEEE International Carnahan Conference on Security Technology*. 2010.

[6] G. Hinton, et al., *Improving neural networks by preventing co-adaptation of feature detectors.* arXiv preprint, 2012. arXiv.

[7] H. R. Roth, et al., *Improving Computer-Aided Detection Using Convolutional Neural Networks and Random View Aggregation.* IEEE Transactions on Medical Imaging, 2016. 35(5): p. 1170-1181.