การประมาณการคุณค่าทางโภชนาการกับปริมาณแคลอรี่ของอาหารไทยจากรูปถ่าย

The Autonomous Nutrient and Calorie Analytics from a Thai Food Image

ภาคภูมิ มุกดาสนิท[1*] และ ลวัณกร มุกดาสนิท[2]
Pakpoom Mookdarsanit[1*] and Lawankorn Mookdarsanit[2]

บทคัดย่อ

วัฒนธรรมการกินกับครัวอาหารไทยนั้นสืบเนื่องมาเป็นระยะเวลานานควบคู่กับศิลปะและประวัติศาสตร์ชาติไทย ในปัจจุบันนั้นอาหารไทยมีความอร่อยติดอันดับโลก ซึ่งนับเป็นความภาคภูมิใจของคนไทย โดยที่วัตถุดิบและเครื่องปรุงต่าง ๆ ของข้าวปลาอาหารไทยนั้น ล้วนเป็นผลผลิตทางการเกษตรที่เติบโตมาจากผืนดินประเทศไทยแทบทั้งสิ้น เนื่องจากนักโภชนาการยุคใหม่นั้นสามารถประมาณการคุณค่าทางสารอาหารกับปริมาณแคลอรี่ของอาหารได้จากรูปถ่าย เช่นเดียวกันกับการสร้างครัวอัจฉริยะจากข้อมูลขนาดใหญ่ การประมาณการคุณค่าทางสารอาหารกับปริมาณแคลอรี่ก็สามารถกระทำได้เองอย่างอัตโนมัติจากรูปถ่ายของอาหาร และแสดงผลต่อผู้ใช้งาน โดยใช้หลักการเรียนรู้เชิงลึกแบบเครือข่ายประสาทเทียมคอนโวลูชั่น ซึ่งเป็นกลไกทางปัญญาประดิษฐ์ เครื่องมือการวิเคราะห์นี้ถูกสร้างขึ้นมาจากภาพอาหารไทยตัวอย่างจำนวน 56,258 ภาพ ซึ่งรูปภาพทั้งหมดนี้ถูกบอทคอมพิวเตอร์ทำการดาวน์โหลดและคัดเลือกมาจากเครือข่ายสังคมออนไลน์ โดยมีค่าเฉลี่ยความถูกต้องอยู่ที่ 0.76 ซึ่งรูปภาพทั้งหมดนั้นครอบคลุมข้าวแกงอาหารไทย 15 อย่าง เพื่อใช้เป็นภาพในการสอนปัญญาประดิษฐ์เพื่อสร้างเครื่องมือวิเคราะห์คุณค่าทางสารอาหารและปริมาณแคลอรี่

คำสำคัญ : ครัวอัจฉริยะ, การรู้จำภาพอาหาร, การประมาณการคุณค่าทางโภชนาการ, การประมาณปริมาณแคลอรี่, การเรียนรู้เชิงลึก

[1] วิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏจันทรเกษม
[1] Computer Science, Faculty of Science, Chandrakasem Rajabhat University
[2] คอมพิวเตอร์ธุรกิจ คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏจันทรเกษม
[2] Business Computer, Faculty of Management Science, Chandrakasem Rajabhat Univerisity
[*] ผู้ประสานงานนิพนธ์ E-mail: pakpoom.m@chandra.ac.th

ABSTRACT

Thai traditional gastronomy and food has existed in a long time together with Thailand's art and history. Proudly, Thai food is currently one of the best tastes in the world that most ingredients are agriculturally grown on Thai soil. Those ingredients affect the nutrient substances and calories. Since the nutritionists can estimate the nutrients and calories from a food image by their eyes. Made up of kitchen intelligence from big data, the nutrients and calories can be autonomously estimated using only a Thai food image, based on Convolutional Neural Network (CNN) – one of deep learning family used for pixel data. All nutrients and calories are shown to users. We used the collection of 56,258 Thai food images to create the deep learning model with the global accuracy of model as 0.76. All images cover 15 Thai foods that are crawled by bots and are used to train the deep learning based analytic model.

Keywords : Kitchen Intelligence, Food Recognition, Nutrient Analytics, Calorie Analytics, Deep Learning

1. Introduction

Not only Thai food occurred in the historical timeline as a local gastronomy but also became one of the best tastes in the world that most ingredients were agriculturally grown on Thai soil. Thailand is an important natural source of nutrients which has high ecological abundance distribution such as plant cover, fish population and diversity of species, etc. From the historical evidence, one of King Ramkhamhaeng (Thai: พ่อขุนรามคำแหง)'s stones (Thai: ศิลาจารึก) was inscribed in Thai alphabets (Chadchaidee, 1994). Which means there were a plenty of fishes in the waterway; there was abundant rice on the land. (Thai: ในน้ำมีปลา ในนามีข้าว). As referred to the verse of Thai foods and desserts (Thai: กาพย์เห่ชมเครื่องคาวหวาน) by King Buddha Loetla Nabhala (Thai: สมเด็จพระพุทธเลิศหล้านภาลัย) (Roufs & Roufs, 2014). that is an acknowledged literature describing a variety of gastronomic foods from Queen Sri Suriyendra (Thai: สมเด็จพระศรีสุริเยนทราบรมราชินี)'s kitchen. In the reign of King Chulalongkorn (Thai: พระบาทสมเด็จพระจุลจอมเกล้าเจ้าอยู่หัว), the first Thai royal recipe archives (Thai: ตำรับอาหารฉบับชาววัง) (Punyaratabandhu, 2017). were officially collected by Princess Saisavali Bhiromya (Thai: พระวิมาดาเธอ พระองค์เจ้าสายสวลีภิรมย์) Up until now, Thai food has become one of the best tastes in the world by CNN travel (Li, 2019) that seems to be “Thai heritage” as well as Thai tourist attractions (Soimart & Mookdarsanit, 2017b)., Buddhist ecclesiastical patterns (Mookdarsanit & Rattanasiriwongwut, 2017a&c; Mookdarsanit, 2020). , Thai martial arts (Mookdarsanit & Mookdarsanit, 2018a) and Thai written scripts (Mookdarsanit & Mookdarsanit, 2019). Since most ingredients of Thai foods from any street foods are from the agricultural commodities from Thailand. These ingredients totally affect the nutrient substances and calories.

The research question is that what is any information that we can observe from a single food image? How do the nutritionists estimate the food by their eyes? How fortunate that we are able to automatically know the nutrition of a food image? Nowadays, many Thai foods with their nutrients and calories have already been collected by the nutritionist. As well as the nutrients and calories of duplicated foods (as the same recipes) can be approximately estimated by the nutritionist’s eyes. In this paper, we propose an AI-based nutrient and calorie analytics using deep learning from a Thai food image. Technically, our proposed model is constructed from 56,258 Thai food images that are crawled from online social media. All images cover 15 lists. For training, many Thai food images with their information are trained to the deep learning. For testing, an unknown Thai food image is input to the model; the name and ingredients are finally shown to users. Either training or testing has feature extraction for selecting the interesting visual points from the food image and keeping in term of a vector before the learning model as shown in Figure 1.

This paper is organized into 7 parts. The literature is in part 2 as related works. Part 3 and 4 are convolution and batch normalization. Max pooling back-propagation and fully-connected layer with experimental results are in part 5 and 6. Finally, conclusion is in part 7.

2. Related Works

Due to the vision analytics (Soimart & Mookdarsanit, 2017a). that were successful in many autonomous applications: tourism recognition (Mookdarsanit & Mookdarsanit, 2018c)., landmark retrieval (Mookdarsanit & Ketcham, 2016; Mookdarsanit & Rattanasiriwongwut, 2017b)., animal recognition (Taheri & Toygar, 2018; Mookdarsanit & Mookdarsanit, 2019a)., dance recognition (Mookdarsanit & Mookdarsanit, 2018b)., silk pattern recognition (Khamket & Surinta, 2020)., facial attribute recognition (Soimart & Mookdarsanit, 2016)., handwriting recognition (Surinta, Karaaka, Schomaker & Wiering, 2015). plant recognition (Mookdarsanit & Mookdarsanit, 2019b)., remote sensing (Li, Wan, Cheng, Meng, & Han, 2019; Soimart & Ketcham, 2016a&b)., bio-cell recogntion (Loresco & Neyra, 2019)., reverse Turing test (Mookdarsanit & Mookdarsanit, 2020a). and medical image (Sutthaluang, 2018; Zhao, Xue, & Li, 2018)., etc.

Some Thai food recognition researches (Termritthikun, Muneesawang, & Kanprachar, 2017; Turmchokkasam & Chamnongthai, 2018; Mookdarsanit & Mookdarsanit, 2018d). are available and categorized into handcrafted and deep learning recognition. Since AlexNet (Krizhevsky, Sutskever, & Hinton, 2012). shouted out the world, the handcrafted based model was shown to be defeated by deep learning in term of correctness (Raksaard & Surinta, 2018)., either animal (Okafor, Pawara, Karaaba, Surinta, Codreanu, Schomaker & Wiering, 2016). or plant (Pawara, Okafor, Surinta, Schomaker & Wiering, 2017). recognition. Particularly deep learning works well on big data – that can find the deep insights (Mookdarsanit & Moorkdarsanit, 2020b). from the large-scale data. Extraordinarily, our deep learning is deeper but fewer parameters and shows the great advance detection in various scales, respectively.

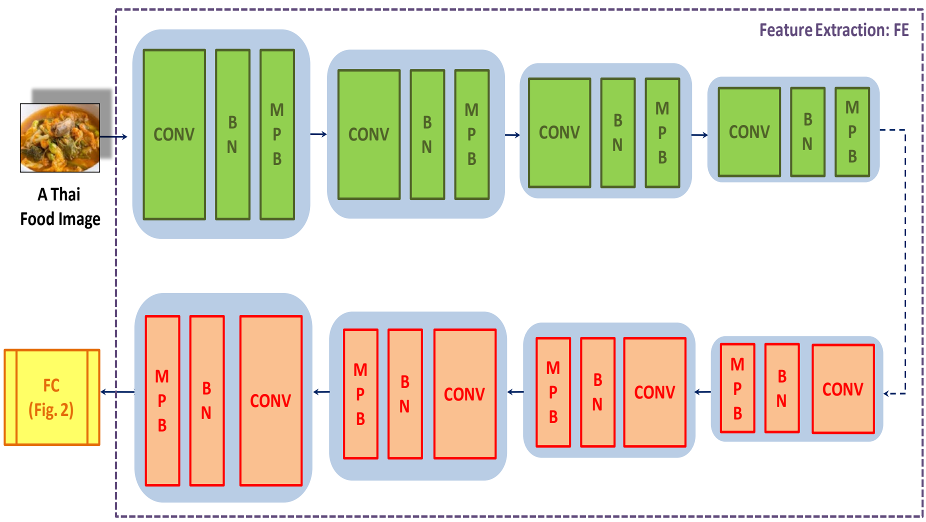

Figure 1 Feature extraction of a food image

3. Convolution (CONV)

In view of signal processing, convolution $CONV(I_{TH\ food}, w_{ij})$ is the convolutional operation (Zou, Shi, Guo & Yi, 2019). between the Fourier transform of a Thai food image and one of a shared weight (a.k.a. kernel or filter).

$$CONV(I_{TH\ food}, w_{ij}) = \left| I_{TH\ food} * w_{ij} \right| \tag{1}$$

where $*$ is convolutional operation, $I_{TH\ food}$ is an input image and w_{ij} is the *j-th* weight of the *i-th* layer.

4. Batch Normalization (BN)

After all convolutions in each layer, the previous normalized parameters are essentially necessary link to the next layer of neural network that can be computed by batch normalization. Rectified Linear Unit (ReLU: $BN_{\mathrm{Re}LU}(.)$) is used in this paper, by (2).

$$BN_{\mathrm{Re}LU}(p_{(x,y)}) = \begin{cases} p_{(x,y)} & if\ p_{(x,y)} > 0 \\ 0 & if\ p_{(x,y)} \leq 0 \end{cases} \tag{2}$$

where $p_{(x,y)}$ is a pixel in x and y position within the output of convolution

5. Max Pooling Backpropagation (MPB)

An output of batch normalization is stored in form of the “feature map ($FeatureMap_{ij}$)”, as (3).

$$FeatureMap_{ij} = \{p_{(x,y)} / p_{(x,y)} \in (Layer_i \cap Node_j)\} \tag{3}$$

By (4), max pooling ($MaxPooling_{ij}$) is used to reduce the dimensionality of feature map. The feature map is first divided into k windows. Later, the max value of pixel within a window is selected to the pool.

$$MaxPooling_{ij} = \arg_{p_{(x,y)}} \max(p_{(x,y)} \mid FeatureMap_{ij\ at\ k}) \tag{4}$$

where i is the *i-th* layer, j as the *j-th weight* and k as the *k-th* window

6. Fully-connected Layer (FC)

Fully-connected layer is the final part of deep learning as “AI-brain” with the parameterized architecture as Visual Geometry Group (VGG-16). It can be categorized into 2 sub-ways: training (as red-lines in Figure 2) – Thai food images with their information are trained to teach the computer model and testing (as blue-lines in Figure 2) – the computer model classifies the unknown Thai food image. After the food name recognition, the nutrients and calories are retrieved from the database as an external knowledge by its food’s name, in Figure 2. The ratio between training and testing is 80:20

6.1 Training

For image collection, we crawl all 56,258 Thai food images from social media and add them with some information like nutrient and calorie as our collection. Overall images covers 15 Thai food types that are used to create the learning model e.g., 1) Tom Yum Goong, 2) Gaeng Som, 3) Chicken Galangal Soup, 4) Tom Saap, 5) Tom Jap Chai, 6) Fish Kidney Curry, 7) Khanom Jeen Nam Ya, 8) Geang Keow Wan, 9) Gaeng Hanglay, 10) Hor Mok, 11) Panang Curry, 12) Pad Sataw, 13) Massaman, 14)
Gaeng Taypo and 15) Gaeng Kee Lek, respectively. These images are input to feature extraction (as shown in Figure 1) and trained to the model as Figure 2

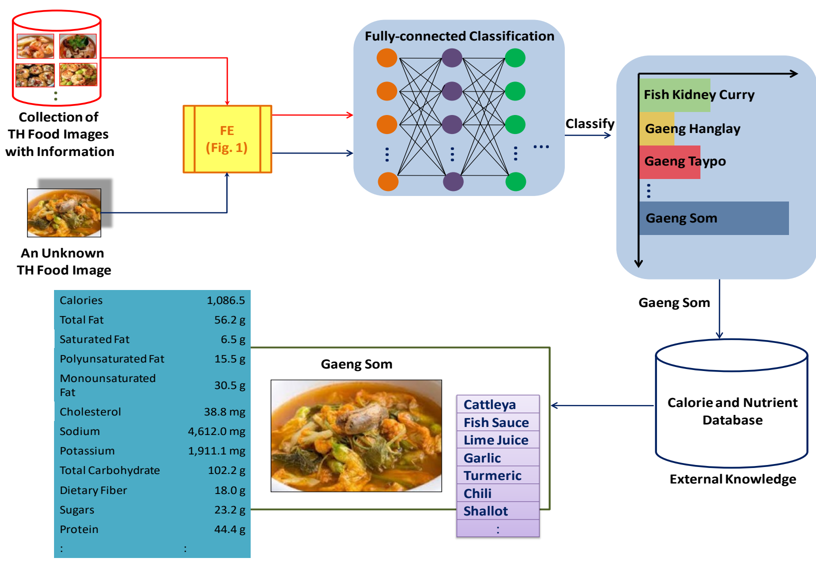

Figure 2 Fully-connected layer: training (as red-lines) and testing (as blue-lines)

6.2 Testing and Evaluation

For an unknown Thai food image that is firstly input to feature extraction (as well as training). Then, all features are considered by the computer model. Finally, the name of food with their information (nutrients and calories) is shown to users as shown in Figure 2.

For the model evaluation, we use the global accuracy of those food classes that can be computed by (5), under the ratio of training: testing as 80:20

Table 1 the global accuracy of each food class

Class	Accuracy
Tom Yum Goong	0.74
Gaeng Som	0.90
Chicken Galangal Soup	0.94
Tom Saap	0.71
Tom Jap Chai	0.73
Fish Kidney Curry	0.76
Khanom Jeen Nam Ya	0.84
Geang Keow Wan	0.63
Gaeng Hanglay	0.68
Hor Mok	0.96
Panang Curry	0.63
Pad Sataw	0.84
Massaman	0.66
Gaeng Taypo	0.78
Gaeng Kee Lek	0.66
Average Accuracy	**0.76**

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

where TP is the actual class as food A and the correctly predicted class as food A, TN is not the actual class as food A and the correctly predicted class as not food A, FP is not the actual class as food A but the wrongly predicted class as food A, FN is the actual class as food A but the wrongly predicted class as not food A.

The results are shown in Table 1. Hor Mok, Chicken Galangal Soup and Gaeng Som have the Accuracy beyond 0.9 because of their distinctiveness that affect the recognition accuracy. Meanwhile, Massaman, Panang Curry and Geang Keow Wan sometimes are looked similar that make the lower accuracy. Hence, the average accuracy is 0.76.

7. Conclusion

Thai food is one of the best tastes in the world that are gradually inherited from the long history as "Thai Heritage". This paper proposes an AI-based deep learning model to analyze the nutrient and calorie from a single Thai food image as one of the kitchen intelligence applications. The model is constructed from 56,258 Thai food images which are categorized into 15 Thai food types: 1) Tom Yum Goong, 2) Gaeng Som, 3) Chicken Galangal Soup, 4) Tom Saap, 5) Tom Jap Chai, 6) Fish Kidney Curry, 7) Khanom Jeen Nam Ya, 8) Geang Keow Wan, 9) Gaeng Hanglay, 10) Hor Mok, 11) Panang Curry, 12) Pad Sataw, 13) Massaman, 14) Gaeng Taypo and 15) Gaeng Kee Lek. The recognition algorithm for these Thai food names are done by Convolutional Neural Network (CNN) based on Visual Geometry Group (VGG-16). Coupled with CNN, the nutrients and calories of the food are retrieved from the external database by its food's name. The average accuracy is 0.76. As there is a large number of food images shared on social media, the autonomous nutrient and calorie analytics will soon be one of the new features of social media like Facebook or Instragram. Many million food images are surely enough for AI-based computer model creation. For future work, many augmentation models can be used to generate many food images that increase the correctness of model. Together with Recurrent Neural Network (e.g., Gate Recurrent Unit, Long-short-term-memory), this can be applied to autonomous food textual captioning to a single image.

8. Reference

Chadchaidee, T. (1994). *Essays on Thailand.* Bangkok, Thailand: D.K. Today.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *The 6th Conference on Neural Information Processing Systems*, (pp. 1-9). Lake Tahoe, Nevada, USA.

Khamket, T. & Surinta, O. (2020). Comparative Study between Texture Feature and Local Feature Descriptors for Silk Fabric Pattern Image Recognition. *The 3rd International Conference on Information Science and System (ICISS)* (pp.93-98). Cambridge, UK: ACM.

Li, K., Wan, G., Cheng, G., Meng, L., & Han, J. (2019). Object Detection in Optical Remote Sensing Images: A Survey. *arXiv:1909.00133* .

Li, Z. (2019). *Which country has the best food?* Retrieved October 25, 2019, from CNN Travel: https://edition.cnn.com/travel/article/world-best-food-cultures/index.html?fbclid=IwAR2aql9u6U6bCRQ8XFH64CKSD7YlymowNJhd6D4HIZD-Hzo8HplFcfJ5rQQ

Loresco, P. J., & Neyra, R. Q. (2019). Filipino Braille One-Cell Contractions Recognition Using Machine Vision. *The 2019 IEEE Region 10 Conference* (pp. 2408-2412). Kochi, India: IEEE.

Mookdarsanit, L. (2020). The Intelligent Genuine Validation beyond Online Buddhist Amulet Market. *International Journal of Applied Computer Technology and Information Systems* , 7-11.

Mookdarsanit, L., & Mookdarsanit, P. (2019a). SiamFishNet: The Deep Investigation of Siamese Fighting Fishes. *International Journal of Applied Computer Technology and Information Systems* , 40-46.

Mookdarsanit, L., & Mookdarsanit, P. (2019b). Thai Herb Identification with Medicinal Properties Using Convolutional Neural Network. *Suan Sunandha Science and Technology Journal* , 34-40.

Mookdarsanit, L., & Mookdarsanit, P. (2020a). An Adversarial Perturbation Technique against reCaptcha Image Attacks. *Journal of Science and Technology Buriram Rajabhat University*.

Mookdarsanit, L., & Mookdarsanit, P. (2020b). The Insights in Computer Literacy toward HR Intelligence: Some Associative Patterns Between It Subjects and Job Positions. *Journal of Science and Technology RMUTSB*.

Mookdarsanit, P., & Ketcham, M. (2016). Image Location Estimation of Well-known Places from Multi-source based Information. *The 11th International Symposium on Natural Language Processing*, Ayutthaya, Thailand: AIAT.

Mookdarsanit, P., & Mookdarsanit, L. (2018a). A Content-based Image Retrieval of Muay-Thai Folklores by Salient Region Matching. *International Journal of Applied Computer Technology and Information Systems* , 21-26.

Mookdarsanit, P., & Mookdarsanit, L. (2018b). An Automatic Image Tagging of Thai Dance's Gestures. *Joint Conference on ACTIS and NCOBA*, (pp. 76-80). Ayutthaya, Thailand.

Mookdarsanit, P., & Mookdarsanit, L. (2018c). Contextual Image Classification towards Metadata Annotation of Thai-tourist Attractions. *ITM Transactions on Information Technology Management* , 32-40.

Mookdarsanit, P., & Mookdarsanit, L. (2018d). Name and Recipe Estimation of Thai-desserts beyond Image Tagging. *Kasem Bundit Engineering Journal* , 193-203.

Mookdarsanit, P., & Mookdarsanit, L. (2019). TGF-GRU: A Cyber-bullying Autonomous Detector of Lexical Thai across Social Media. *NKRAFA Journal of Science and Technology*, 50-58.

Mookdarsanit, P., & Rattanasiriwongwut, M. (2017a). GPS Determination of Thai-temple Arts from a Single Photo. *The 11th International Conference on Applied Computer Technology and Information Systems*. Samutprakan: Southeast Bangkok College.

Mookdarsanit, P., & Rattanasiriwongwut, M. (2017b). Location Estimation of a Photo: A Geo-signature MapReduce Workflow. *Engineering Journal*, 295-308.

Mookdarsanit, P., & Rattanasiriwongwut, M. (2017c). MONTEAN Framework: A Magnificent Outstanding Native-Thai and Ecclesiastical Art Network. *International Journal of Applied Computer Technology and Information Systems* , 17-22.

Okafor, E., Pawara, P., Karaaba, F., Surinta, O., Codreanu, V., Schomaker, L. & Wiering, M. (2016). Comparative study between deep learning and bag of visual words for wild-animal recognition. *The 2016 IEEE Symposium Series on Computational Intelligence* (pp. 1-8). Athens, Greece: IEEE.

Pawara, P., Okafor, E., Surinta, O., Schomaker, L. & Wiering, M. (2017). Comparing Local Descriptors and Bags of Visual Words to Deep Convolutional Neural

Networks for Plant Recognition. *The 6th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2017)* (pp. 479-486). Porto, Portugal: SCITEPRESS.

Punyaratabandhu, L. (2017). *Bangkok: Recipes and Stories from the Heart of Thailand.* Berkeley, California, United States: Ten Speed Press.

Raksaard, N. & Surinta, O. (2018). Comparative Study between Local Descriptors and Deep Learning for Silk Pattern Image Retrieval. *Science and Technology Mahasarakham University*, 736-746. [written in Thai]

Roufs, T. G., & Roufs, K. S. (2014). *Sweet Treats around the World: An Encyclopedia of Food and Culture.* Santa Barbara, California, United States: ABC-CLIO.

Soimart, L., & Mookdarsanit, P. (2016). Gender Estimation of a Portrait: Asian Facial-significance Framework. *The 6th International Conference on Sciences and Social Sciences*. Mahasarakham: Rajabhat Mahasarakham University.

Soimart, L., & Ketcham, M. (2016a). An efficient algorithm for earth surface interpretation from satellite imagery. *Engineering Journal* , 215-228.

Soimart, L., & Ketcham, M. (2016b). Hybrid of Pixel-based and Region-based Segmentation for Geology Exploration from Multi-spectral Remote Sensing. *The 11th International Symposium on Natural Language Processing*, Ayutthaya, Thailand: AIAT.

Soimart, L, & Mookdarsanit, P. (2017a). Ingredients estimation and recommendation of Thai-foods. *SNRU Journal of Science and Technology* , 509-520.

Soimart, L., & Mookdarsanit, P. (2017b). Name with GPS Auto-tagging of Thai-tourist Attractions from An Image. *The 2nd Technology Innovation Management and Engineering Science International Conference* (pp. 211-217). Nakhon Pathom: Mahidol University.

Surinta, O., Karaaka, M., Schomaker, L. & Wiering, M. (2015). Recognition of handwritten characters using local gradient feature descriptors. *Engineering Applications of Artificial Intelligence,* 405-414.

Sutthaluang, N. (2018). The impact of asbestos exposure on lung disease. *The 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology* (pp. 353-355). Chiang Rai, Thailand: IEEE.

Taheri, S., & Toygar, Ö. (2018). Animal classification using facial images with score-level fusion. *IET Computer Vision*, 679-685.

Termritthikun, C., Muneesawang, P., & Kanprachar, S. (2017). NU-InNet: Thai Food Image Recognition Using Convolutional Neural Networks on Smartphone. *Journal of Telecommunication, Electronic and Computer Engineering* , 63-67.

Turmchokkasam, S., & Chamnongthai, K. (2018). The Design and Implementation of an Ingredient-Based Food Calorie Estimation System Using Nutrition Knowledge and Fusion of Brightness and Heat Information. *IEEE Access* , 46863-46876.

Zhao, Y., Xue, W., & Li, Q. (2018). A Multi-Scale CRNN Model for Chinese Papery Medical Document Recognition. *The 4th International Conference on Multimedia Big Data* (pp. 1-5). Xi'an, China: IEEE.

Zou, Z., Shi, Z., Guo, Y., & Ye, J. (2019). Object Detection in 20 Years: A Survey. arXiv:1905.05055.