AI in gastronomic tourism

G. Pavlidis[1], S. Markantonatou[1], K. Toraki[1], A. Vacalopoulou[1], C. Strouthopoulos[2], D. Varsamis[2], A. Tsimpiris[2], S. Mouroutsos[3], C. Kiourt[1], V. Sevetlidis[1] and P. Minos[1]

[1]Athena Research Centre, University Campus at Kimmeria, GR-67100, Xanthi, Greece
[2]International Hellenic University, Terma Magnisias, GR-62124, Serres, Greece
[3]Democritus University of Thrace, University Campus at Kimmeria, GR-67100, Xanthi, Greece
Tel.: + 302541078787, fax: + 302541063656
E-mail: gpavlid@athenarc.gr

Summary: Gastronomy is increasingly becoming a decisive flavour of tourists' experience. Tourists' gastronomic experience could be significantly enhanced by AI tools that provide image-based dish recognition and menu translation, thus covering the basic needs of tourists during a visit that involves culinary experiences. This paper presents and discusses solutions explored with the project GRE-Taste that deals with the enhancement of culinary tourism experience in northern Greece, for which no linguistic resources exist and where general-purpose tools fail.

Keywords: dish recognition, menu translation, gastronomic tourism, AI, OCR, CNN, PureFoodNet.

1. Introduction

Food is both an essential commodity and a social and cultural heritage. According to [1] food is as vital to human health and well-being as all other products together and this is the main reason why so much importance is attached to it. This work also suggests that food plays a multi-functional connective role in society and that sustainable food systems support sustainable communities. Food influences peoples' lifestyle, health and habits as well as the design model for land, water, energy, transport and ecosystem services. *Gastronomy* and cooking are gradually becoming more and more important multidimensional societal factors. In 2014, the European Parliament's Committee on Culture and Education adopted a movement on "European culinary heritage: cultural and educational aspects"[1], which recognizes the importance of food and gastronomy as an artistic and cultural expression and proclaims them fundamental pillars of family and social relationships[2].

In the context of events organized by the S3 Platform[3] international experts proclaimed food an element of smart specialization in EU countries and regions and addressed food innovation issues as a driving force for smart regional development [2].

According to [3], *gastronomic tourism is a visit to primary and secondary producers of food and beverages, gastronomy festivals, dining venues and specific locations, where tasting and experience of specialty local food features are a prime motivation for the visit*.

Nowadays, the culinary aspect of tourism is prominent with a growing tendency in the next decade, whereas mass tourism seems to have reached a tipping point, as the quest of individual experiences is taking over [4]. This implies that the tourist will seek the pleasure of discovering tastes along with the cultural traits of foods in their local settings rather than being directed to pre-selected restaurants. To fully enable tourists to savour foods and experience the cultural context in local settings one has to draw on two resources, namely rich visual content [5] and knowledge of details of a foreign language and scripture. This requirement goes even further if knowledge of the materials, dietary properties, cooking techniques and, health issues come into play. Accordingly, two challenges present themselves, namely image-based content recognition and classification—*food image recognition*, and language-based knowledge mining and translation—*menu translation*. This paper presents the path followed to address these challenges within the GRE-Taste project that targets culinary tourism and food tradition in northern Greece[4].

2. Food image classification

Food (dish) image classification is a particular challenge due to the visual and the semantic complexity stemming from the variation in the mixing of ingredients practiced by regional communities [6], [7]. Although a multitude of applications, algorithms and systems is now available, recognizing dishes and their ingredients is a

[1] European gastronomic heritage: cultural and educational aspects, http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P7-TA-2014-0211&language=EN&ring=A7-2014-0127
[2] European gastronomic heritage: cultural and educational aspects, http://www.europarl.europa.eu/sides/getDoc.do?type=TA&reference=P7-TA-2014-0211&language=EN&ring=A7-2014-0127
[3] S3 Platform provides aid in the development and application of smart specialization strategies (RIS3) in Europe, http://s3platform.jrc.ec.europa.eu/s3-platform
[4] GRE-Taste project website @ http://gre-taste.ceti.gr

problem that has not been fully addressed by the machine learning and computer vision communities [6], basically due to the lack of distinctive spatial layout in food images. Ingredients, say of a salad, constitute mixtures that typically come in different shapes and sizes; furthermore, often the nature of a dish is defined by the different colours, shapes and textures of the ingredients [8]. Solutions to this problem have been proposed that exploit a range of innovative ideas; recently deep learning techniques produce remarkably successful results.

In GRE-Taste the problem is even more challenging, as traditional and regional Greek food and dish images have not been used to develop models in the past. Apparently, since the project targets tourism and mobile device apps, several restrictions apply regarding the storage and processing power availability. Given this context, three alternatives have been explored, (a) the development of a totally new architecture based on CNNs, (b) the application of transfer learning and fine-tuning of existing models, and (c) the usage of online deep learning platforms. The first scenario included the development of a new CNN architecture, the *PureFoodNet*, which consists of three convolutional and one classification blocks (as shown in Fig. 1) [9]. To implement the second scenario, VGG16 [10], InceptionV3 [11], ResNet50 [12], InceptionResNetV2 [13], MobileNetV2 [14], DenseNet121 [15], and NASNetLarge [16] have being used. The third scenario included the experiments on Google Vision AI[5] (VAI), Clarifai[6] (CAI), Amazon Rekognition[7] (AR) and Microsoft Computer Vision[8] (MCV) platforms.

As deep learning approaches rely heavily on existing data, the first issue to be tackled was the selection of an appropriate dataset that would aid in the recognition of Greek dishes; however, such a dataset does not exist. The most popular datasets consist of general/universal food categories, whereas some are only regional (like UECFood100/256). A relevant image collection and photo-shooting task was initiated and has not been completed yet. Thus, the results presented in this study are based on the usage of the popular Food101 dataset, which was created around 2015 and includes 101 food categories with 101,000 food images.

In the first set of experiments (scenarios a and b), PureFoodNet was tested against VGG16, InceptionV3, ResNet50, InceptionResNetV2, MobileNetV2, DenseNet121 and NASNetLarge. Fig. 2 graphically depicts the accuracy attained by each model (left) in terms of categorical top-1 and top-5 performance for training and validation ("val" denotes the validation phase), along with a graph of the number of epochs needed during training respectively (right). Apparently, most of the models performed quite well, with the worst validation accuracy observed for the VGG16 model, which was also among the slowest ones. In addition, although NASNetLarge showed an excellent performance it needed 83,33% more time and around 92% more time than the PureFoodNet.

PureFoodNet performed quite well even though it was not the best, but it was definitely the most light-weight, time/resource consuming approach. Furthermore, due to its low complexity, PureFoodNet is one of the most easily customisable and tuneable models for further studies.

For the second set of experiments (scenario c), two test images were selected (typical Greek salad and a fruit desert) to compare the performance of the online AI platforms; and the task also included reporting on the ingredients of the dish. The performance of each platform is shown in Fig. 3, in which green denotes an acceptable response, red denotes an unacceptable response and white an irrelevant response. Apparently, MCV and AR performed rather poorly, whereas VAI was somewhat better. CAI performed significantly better, basically because it had been designed for the specific task of food recognition.

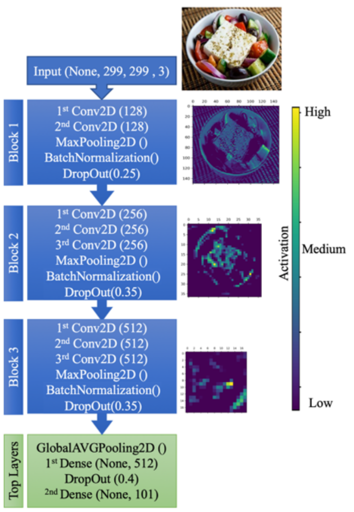

Fig. 1. The PureFoodNet architecture for traditional Greek food recognition.

[5] Google Vision AI @ https://cloud.google.com/vision/
[6] Clarifai @ https://clarifai.com
[7] Amazon Rekognition @ https://aws.amazon.com/rekognition/
[8] Microsoft Computer Vision @ https://www.microsoft.com/cognitive-services

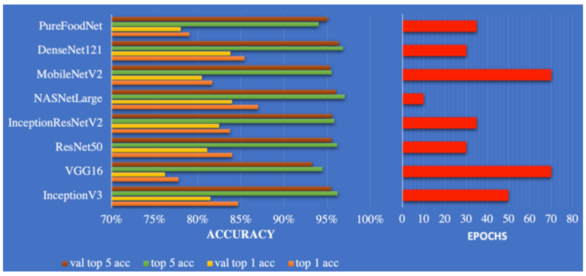

Fig. 2. Comparison of popular deep architectures with PureFoodNet in food recognition.

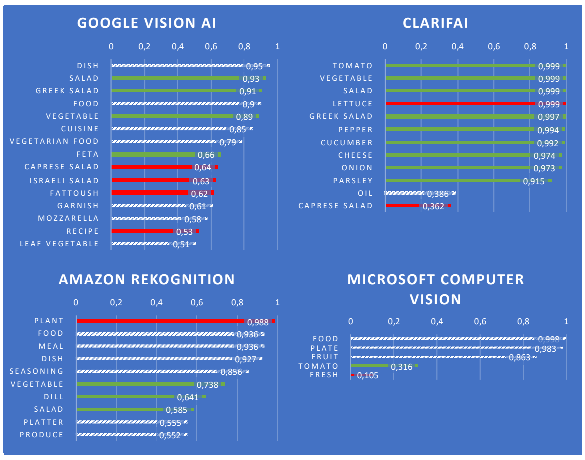

Fig. 3. Results on food recognition image A by the reviewed platforms.

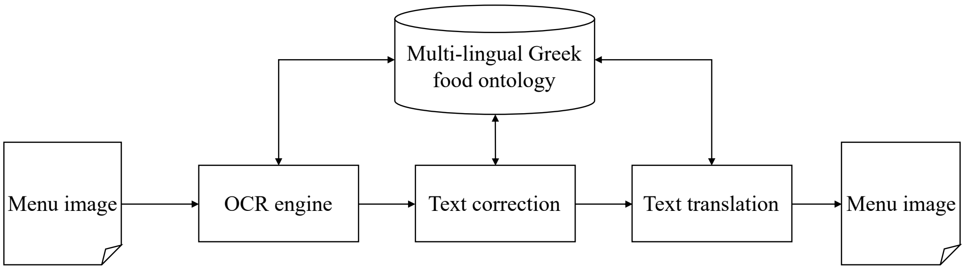

Fig. 4. The menu translation workflow in GRE-Taste.

3. Menu translation

Greece boasts a wealth of local foods and drinks that form part of a rich cultural context and contribute to the gastronomic aspect of the local tourism industry. Greek, however, is not a widely spoken language and has its own alphabet that renders menu understanding difficult for the average tourist. At the moment, only some guided tours are available that focus on the gastronomy aspect while gastronomy-related tours enhanced with the cultural traits of food are even rarer. To date, no relevant electronic applications are available.

Current solutions to the linguistic problem of culinary tourist experience focus (a) on providing restaurants with on-demand menu translations [17], or (b) on machine translation of menus, such as the Word Lens [18] and the Purdue Menu Translator [19]. Nevertheless, it should be stressed that despite any technical details, machine translation approaches require special databases for each language and none of them includes Greek (yet). Eventually, Greek have to make do with solutions like Google Images, Google Translate, BabelNet and Wikipedia. In addition, understanding a menu is not a terminological problem only, it has to do with the way menus are written and on the contextual knowledge required to fully appreciate a type of food.

In GRE-Taste the workflow for the menu translation consists of two main parts, one that deals with the optical character recognition (OCR) in menu photos and another that deals with the content translation (Fig. 4). Currently, the OCR part of the workflow is explored using (a) online services, APIs and engines, like Google Vision AI, Tesseract and ABBYY and (b) in-house research. The translation part involves the development of a tri-lingual thesaurus of foods, nutritional and cultural information and the translation engine itself. The Greek food ontology will stand on top of the system to aid in the correct multi-lingual food translation, although it is currently in a form of a multi-lingual thesaurus.

The OCR problem relates with the identification of text in images. There is a rich bibliography on this challenge and various text detection approaches have been developed that successfully identify texts of interest in complex images [20]. Recently, this research was expanded to accommodate the needs of food and meal applications by recognising text in menu images[9] [21]–[23]. In GRE-Taste the focus is on identifying text in menu photos shot with any mobile phone cameras, in any lighting conditions. This may include under/over-exposed photos with a partially captured menu in an arbitrary orientation, which, in many cases, includes heavy combinations of graphics, photos, printed and handwritten text. Since this is something like a worst-case usage scenario in OCR, the GRE-Taste team focused the efforts and conducted a number of experiments in this direction of devising an image preprocessing strategy whose output will feed a state-of-the-art OCR system, like the Tesseract[10] [24]. A number of filtering techniques have been compared, including several thresholding approaches, morphological operations, global and adaptive gradient approaches, gaussian techniques and wiener filtering [25]–[30]. Preliminary experiments have been conducted on a mixture of general-purpose text images and Greek menu images collected and annotated by the GRE-Taste team[11]. Fig.5 shows the cosine, Jaccard, CER and WER indices [31]–[33] as the average performance results of Tesseract OCR empowered with the pre-processing methods on a subset of 80 menu images. Horizontal lines denote the corresponding index value attained without any pre-filtering. Apparently, there is a slight improvement in the results for the case of erosion and median blurring in almost all indices.

Repeating these experiments in the Google Vision platform for the 11 pre-filtering strategies, the average results shown in Fig. 6 were obtained, again in terms of the 1-CERi, cosine, Jaccard and 1-WERi indices. It is evident that gradient-based filtering and contrast-limited adaptive histogram equalization gave a slight improvement at least in the 1-CERi index. It is also evident in both experiments that for a very limited number of pre-filtering cases only there was an improvement, basically due to the inner processes in Tesseract and Google Vision AI.

It should be mentioned that, although Google Vision AI achieved about 20% better performance than the customized Tesseract application, its usage in relevant applications may be prohibited because of the charging policy adopted by its developers.

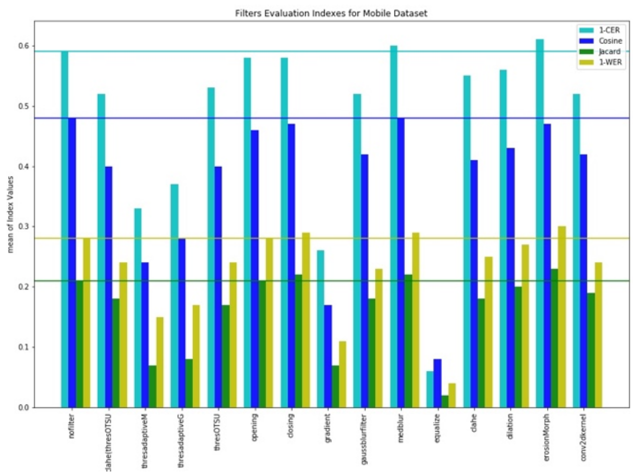

Fig. 5. Average performance results for 15 pre-filtering methods in 1-CERi, 1-WERi, cosine and Jaccard indices.

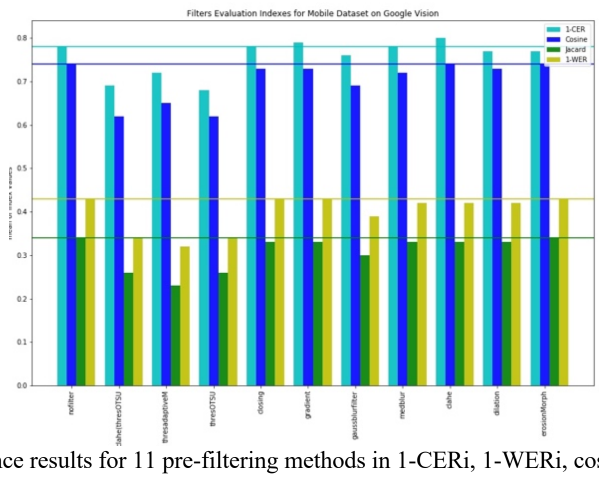

Fig. 6. Average performance results for 11 pre-filtering methods in 1-CERi, 1-WERi, cosine and Jaccard indices using the Google Vison AI.

[9] See for example https://apkpure.com/foodtranslator/com.foodwandering.android.foodtranslator, https://apkpure.com/foodie-for-all/com.uos.kyungimlee.myapplication based on Google Vision https://cloud.google.com/vision/.

[10] https://github.com/tesseract-ocr/

[11] It should be stressed that these tests only included a limited dataset as the data collection process is still underway.

The core of the menu translation challenge consists in rendering the text the menus into a predetermined language; this complicated problem involves translation of the information that is explicit in the menus, namely names of dishes and their ingredients, and provision of dietary and cultural information that is implicit in the menus. The menu translation problem is mostly a terminological one and, precisely for this reason, GRE-Taste has focused on the development of a trilingual thesaurus of foods served in restaurants and taverns of northern Greece. A dedicated web terminographical environment was developed[12] that (a) accommodates texts retrieved from 120 restaurant/tavern/pastry shop menus of the study area and (b) enables the development of a thesaurus that models information on dishes, ingredients, courses in a meal, source of food and part of it, cooking methods in addition to nutritional and cultural information. Currently, the thesaurus contains 3072 Greek and English terms designating 1405 concepts that are connected with 2660 relation instances.

Overall, GRE-Taste will result in a multi-faceted thesaurus for Greek gastronomy for the first time. Some of the issues GRE-Taste takes into account are:

- varieties in culinary language related with dialectic forms, local specialties, etc;
- dish variations, leading to the definition of specific types of dishes, e.g. "stifado" and "rabbit/beef/cuttlefish stifado";
- synecdoche and ambiguity, i.e. when the same word denotes both the food and its source, e.g. "lettuce", which may be used to name the vegetable as a plant, the vegetable as an ingredient of a salad or the lettuce salad;
- categorization of dishes, i.e. some dishes may need to be classified under more than one general category (facet), e.g. "papoutsakia", literary small shoes, "eggplant with minced meat" may be classified both as meat dishes and vegetable dishes.

Dealing with the above, mainly terminological, issues ensures that users will be able to select dishes according to their liking and their specific needs and peculiarities.

The translation procedure strongly relies on the thesaurus; it also uses existing home-made multilingual dictionaries of the general language, part of speech recognition and lemmatization performed with the ILSP tools for Greek[13], whereas automata implementing simple grammar rules are currently being developed. The first translation toy experiments[14] suggest that the adopted approach has a significant potential to outperform Google Translate only by using its lexical resources and lemmatising facilities (no other linguistic knowledge) provided that the thesaurus can cover the terminological demands of the menu.

4. Conclusions

Culinary and gastronomic tourism are increasingly gaining momentum in the tourism industry as food is an important factor of social life and culture. Generic AI tools, such as machine translation and image-based recognition, can provide some aid to the tourist but still with significant limitations and high levels of generalization that leads to fails in traditional, ethnic or regional settings.

GRE-Taste is a project focusing on developing new workflows and AI tools to address the main issues in this domain for a particular region of interest with rich cultural context. New methods for dish recognition and menu translation have been designed with promising results. The PureFoodNet system that was developed is able to provide accurate dish recognition with low processing and storage demands[15]. In addition, the lack of language resources has been addressed with the development of a completely new multi-faceted thesaurus for the Greek gastronomy in the general cultural context.

Acknowledgements

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-02015). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for the experiments of this research.

---

[12] http://gretaste.ilsp.gr/
[13] http://nlp.ilsp.gr/soaplab2-axis/
[14] These experiments included the selection of a menu, the lemmatisation of the Greek text and the verbatim translation.
[15] A cross-platform online version of the system can be found in http://gre-taste.ceti.gr/webapp/. The system can be used on a mobile device as a service and can be easily ported to a native mobile app.

References

[1] K. Morgan, “Local and green, global and fair: the ethical foodscape and the politics of care,” *Environment and Planning A*, vol. 42, no. 8, pp. 1852–1867, 2010.

[2] A. Cavicchi and K. C. Stancova, “Food and gastronomy as elements of regional innovation strategies,” *Institute for Prospective and Technological Studies, Joint Research Centre, Sevilla[Links]*, 2016.

[3] C. Mitchel Hall, L. Sharples, R. Mitchel, N. Macionis, and B. Cambourne, *Food Tourism around the world: development, management and markets*. Elsevier Ltd, Burligton USA, 2003.

[4] Skift Team - The Ontario Culinary Tourism Alliance, “Free Report: The Rise of Food Tourism,” *Free Report: The Rise of Food Tourism*, 2015. [Online]. Available: https://skift.com/insight/free-report-the-rise-of-food-tourism/. [Accessed: 18-Jan-2020].

[5] D. Jurafsky, *The language of food: A linguist reads the menu*. WW Norton & Company, 2014.

[6] M. Weiqing, J. Shuqiang, L. Linhu, R. Yong, and J. Ramesh, *A Survey on Food Computing*. ACM Computing Surveys, 2019.

[7] S. Mezgec and S. Koroušić, “NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment,” *Nutrients*, vol. 9, no. 657., 2017.

[8] G. Ciocca, P. Napoletano, and R. Schettini, *CNN-based features for retrieval and classification of food images*. Computer Vision and Image Understanding, 2018.

[9] C. Kiourt, G. Pavlidis, and S. Markantonatou, “Deep learning approaches in food recognition,” in *Machine Learning Paradigms - Advances in Theory and Applications of Deep Learning*, Springer, 2020.

[10] K. Simonyan and A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 3rd IAPR Asian Conference on Pattern Recognition (ACPR), 2015.

[11] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.

[12] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[13] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning.,” in *AAAI*, 2017, vol. 4, p. 12.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, C. L, and C., *MobileNetV2: Inverted Residuals and Linear Bottlenecks*. The IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[15] G. Huang, Z. Liu, L. van der Maaten, and K. Weinberger, *Densely connected convolutional networks*. IEEE Conference on Pattern Recognition and Computer Vision, 2017.

[16] B. Zoph, V. Vasudevan, J. Shlens, and V. Le, *Learning Transferable Architectures for Scalable Image Recognition*. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[17] TranslateMyMenu, “Translate My Menu – Menu translation service,” 2014. [Online]. Available: https://translatemymenu.com/en_US/. [Accessed: 18-Jan-2020].

[18] Wikipedia, “Word Lens,” *Wikipedia*. 20-Oct-2019.

[19] Purdue University, “New translator app makes sense of foreign-language food menus,” 2011. [Online]. Available: https://www.purdue.edu/newsroom/research/2011/110908BoutinMenutranslate.html. [Accessed: 18-Jan-2020].

[20] X. Zhou *et al.*, “EAST: an efficient and accurate scene text detector,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 5551–5560.

[21] W. Swastika, H. Setiawan, and M. Subianto, “Android based application for recognition of Indonesian restaurant menus using convolution neural network,” in *Sustainable Information Engineering and Technology (SIET), 2017 International Conference on*, 2017, pp. 30–34.

[22] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, “A Survey on Food Computing,” *arXiv preprint arXiv:1808.07202*, 2018.

[23] R. Ullah, A. Sohani, A. Rai, F. Ali, and R. Messier, “OCR Engine to Extract Food-Items, Prices, Quantity, Units from Receipt Images, Heuristics Rules Based Approach,” *International Journal of Scientific & Engineering Research*, vol. 9, no. 2, pp. 1334–1341, 2018.

[24] R. Smith, “Tesseract Open Source OCR Engine,” *GitHub*, 2018. [Online]. Available: https://github.com/tesseract-ocr/tesseract. [Accessed: 09-Jan-2019].

[25] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[26] B. Hunt, “A matrix theory proof of the discrete convolution theorem,” *IEEE Transactions on Audio and Electroacoustics*, vol. 19, no. 4, pp. 285–288, 1971.

[27] C. A. Glasbey and G. W. Horgan, *Image analysis for the biological sciences*, vol. 1. Wiley Chichester, 1995.

[28] F. Orieux, J.-F. Giovannelli, and T. Rodet, “Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution,” *JOSA A*, vol. 27, no. 7, pp. 1593–1607, 2010.

[29] E. D. Pisano *et al.*, “Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms,” *Journal of Digital imaging*, vol. 11, no. 4, p. 193, 1998.

[30] K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics gems IV*, 1994, pp. 474–485.

[31] S. V. Rice, J. Kanai, and T. A. Nartker, “An evaluation of OCR accuracy,” *Information Science Research Institute, 1993 Annual Research Report*, vol. 9, p. 20, 1993.

[32] J. W. Ratcliff and D. E. Metzener, “Pattern-matching-the gestalt approach,” *Dr Dobbs Journal*, vol. 13, no. 7, p. 46, 1988.

[33] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals,” in *Soviet physics doklady*, 1966, vol. 10, pp. 707–710.