applied sciences

MDPI

Project Report

PROMISE CLIP Project: A Retrospective, Multicenter Study for Prostate Cancer that Integrates Clinical, Imaging and Pathology Data

Jihwan Park [1,2], Mi Jung Rho [1], Yong Hyun Park [1,3], Chan Kwon Jung [4], Yosep Chong [5], Choung-Soo Kim [6], Heounjeong Go [7], Seong Soo Jeon [8], Minyong Kang [8], Hak Jong Lee [9], Sung Il Hwang [9] and Ji Youl Lee [1,3,*]

1 Catholic Cancer Research Institute, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea
2 Department of Biomedicine & Health Sciences, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea
3 Department of Urology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea
4 Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 06591, Korea
5 Department of Hospital Pathology, Yeouido St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul 07345, Korea
6 Department of Urology, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, Korea
7 Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, Korea
8 Department of Urology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 06351, Korea
9 Department of Radiology, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam 13620, Korea
* Correspondence: uroljy@catholic.ac.kr

Received: 12 June 2019; Accepted: 23 July 2019; Published: 25 July 2019

Abstract: There are many medical demands that still need to be resolved for prostate cancer (PCa), including better diagnosis and predictive medicine. For this to be accomplished, diverse medical data need to be integrated with the development of intelligent software (SW) based on various types of medical data. Various types of information technology have been used to address these medical demands of PCa. We initiated the PROstate Medical Intelligence System Enterprise-Clinical, Imaging, and Pathology (PROMISE CLIP) and a multicenter, big data study to develop PCa SW for patients with PCa and clinicians. We integrated the clinical data of 7257 patients, 610 patients' imaging data, and 39,000 cores of pathology digital scanning data from four tertiary hospitals in South Korea. We developed the PROMISE CLIP registry based on integrated clinical, imaging, and pathology data. Related intelligent SW has been developed for helping patients and clinicians decide on the best treatment option. The PROMISE CLIP study directs guidelines for intelligent SW development to solve medical demands for PCa. The PROMISE CLIP registry plays an important role in advancing PCa research and care.

Keywords: artificial intelligence; prostate cancer; clinical data; imaging data; pathology data

1. Introduction

Prostate cancer (PCa) is the most commonly diagnosed cancer among males, accounting for almost 1 in 5 new diagnoses, and is the second leading cause of cancer-related death in men, with more than 3.3 million men living with PCa in the United States [1]. In addition, according to the cancer

statistics report of Korea 2019, PCa is the fourth most common male cancer, and its incidence continues to increase in Korea; PCa's crude incidence rate (CR) is 46.2 per 100,000 [2].

In PCa, there are diverse medical demands that remain to be resolved for all stages of treatment, such as the diagnosis of PCa, interpretation of imaging or pathology, and prediction of clinical outcomes. Various information technologies have been used to address these medical demands.

Information technologies, such as artificial intelligence and machine learning, are advantageous in PCa. Regnier-Coudert et al. used an artificial neural network(ANN) and a Bayesian network to improve pathological staging of PCa [3]. Zupan et al. used machine learning for survival analysis on recurrence of PCa [4]. Cruz et al. surveyed research of ANN and other machine learning methods for diagnosis and predictive medicine. PCa is a large focus of predictive medicine research using machine learning including ANN [5]. Fakoor et al. applied deep learning techniques to the detection and classification of PCa based on gene expression data [6].

However, these previous studies were based on Western patients with PCa. There are genetic and clinical differences between Koreans and other races in the world. Large-scale registries for PCa are playing a growing role in advancing PCa research and care [7]. Thus, we need to develop large-scale Korean registries for PCa care and research. In addition, we need to develop intelligent SW technology based on large-scale registries to solve these medical demands.

Accordingly, we have designed the PROstate Medical Intelligence System Enterprise-Clinical, Imaging, and Pathology (PROMISE CLIP) project. The goal of PROMISE CLIP is to convert large amounts of raw medical data into meaningful information to address the diverse medical demands of PCa.

2. Patients and Methods

2.1. Study Design

PROMISE CLIP was designed to solve the medical demands of PCa. The PROMISE CLIP registry consists of clinical, imaging, and pathology data. The study period spans three years from 1 April 2018 to 31 December 2020 (Figure 1). We collected electronic medical records (EMR) data, magnetic resonance imaging (MRI) images, and biopsy slides from four hospitals to develop this registry.

Primary endpoints of the PROMISE CLIP project are as follows. (1) Prediction of pathologic outcomes that help patients and clinicians choose the best treatment option. (2) Prediction of treatment outcomes after definitive surgery that helps to define the ideal patient population for aggressive follow-up or early postoperative ancillary treatment. (3) Accurate interpretation of multiparametric MRI. (4) Accurate digital pathology of PCa to improve accuracy, reduce human error, and increase reproducibility.

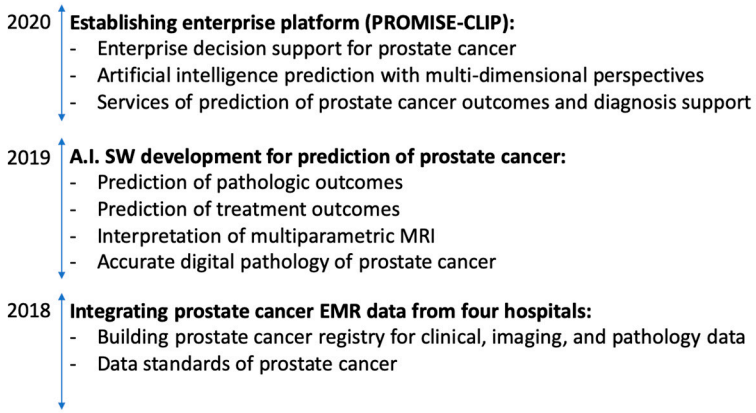

Figure 1. The PROstate Medical Intelligence System Enterprise-Clinical, Imaging, and Pathology (PROMISE CLIP) project.

2.2. Study Organization

PROMISE CLIP is a multicenter study attended by four hospitals: Seoul St. Mary's Hospital of the Catholic University, Seoul National University Bundang Hospital, Samsung Medical Center, and Asan Medical Center. We developed the PROMISE CLIP registry. The four participating hospitals are tertiary hospitals located in Seoul and Gyeonggi-do Province (capital area). The number of hospital beds is 1355 in Seoul St. Mary's Hospital of the Catholic University, 1339 in Seoul National University Bundang Hospital, 1979 in Samsung Medical Center, and 2704 in Asan Medical Center.

2.3. Inclusion and Exclusion Criteria

For clinical data, we collected PCa patient data on radical prostatectomy occurring between 1 January 2010 and 31 December 2017. We excluded patients treated with chemotherapy for other malignant tumors within one year.

For imaging data, we collected patient 3T multiparametric MRI data from PCa patients who underwent radical prostatectomy between 1 January 2010 and 31 December 2017. We excluded patients treated with chemotherapy for other malignant tumors within one year.

For pathology data, patients must have hematoxylin and eosin stained (H&E) slides of transrectal prostate biopsy among PCa patients who underwent radical prostatectomy between 1 January 2010 and 31 December 2017. Exclusion criteria included chemotherapy for other malignant tumors within one-year, medical history of neoadjuvant treatment, such as radiation therapy or androgen deprivation therapy, and medication history of 5α reductase inhibitor, such as finasteride and dutasteride, for benign prostatic hyperplasia.

2.4. Data Acquisition

For clinical data (PROMISE-CL), 7,257 patients with PCa treated with radical prostatectomy were included from each participant hospital's EMR data. In addition, we used the multicenter Korean Prostate Cancer Database (K-CaP) and the Asian Prostate Cancer (A-CaP) to refer to select data fields [8,9]. The K-CaP database is an observational longitudinal database of Korean patients with biopsy-proven PCa enrolled from five hospitals throughout Korea. The K-CaP provides 220 items for PCa. We set up the rules for exclusion and inclusion prior to collecting data from all of the PCa patients of each participating hospital. Final data were collected by discussions with physicians [10]. We are collecting prospective data, and the number of data groups continues to grow.

In PROMISE-I, a total of 610 patients with PCa treated with radical prostatectomy were included, and clinical data for all patients were collected. 610 patients' multiparametric MRI images obtained from 3.0 T MRI were collected.

In PROMISE-P, 39,160 previously diagnosed PCa glass slides were de-identified and scanned into SVS file whole slide images at 400× magnification (Aperio AT2). A total of 39,160 whole slide images (10–12 needle biopsy cores per patient) were then annotated according to criteria determined by four independent experienced pathologists.

Before loading these data into the research areas, each data group in PROMISE-CL, PROMISE-I, and PROMISE-P was reviewed by clinicians, radiologists, and pathologists. Participant companies are working on developing diagnosis and predictive models using these confirmed data.

PROMISE-CL, PROMISE-I, and PROMISE-P were connected by research identification numbers (RIDs), which were generated by de-identification tools developed by participant development companies (Figure 2). For instance, pathology slide de-identification SW removes pathologist's marks and meta-information on the slide scanned data. When it was necessary to review the original details of research data, for security reasons to protect private information, only an authorized clinician accessed the RID and patient ID mapping information to verify the data.

For the external validation of diagnosis and predictive models, we separated training and validating data groups by hospital. We are collecting prospective data to compare to the retrospective model.

Collected data are available only for participants of the PROMISE CLIP project. SW developed from these data is going to be used by participant hospitals for the initial implementation test followed by spreading across the entire country.

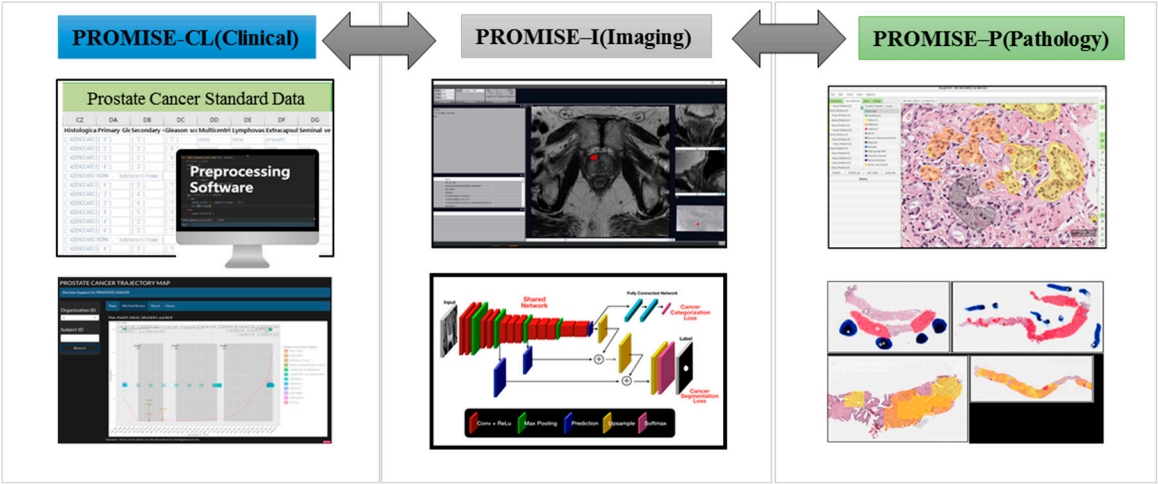

Figure 2. PROMISE CLIP Clinical, Imaging, and Pathology data.

2.5. Preprocessing Methods

The PROMISE CLIP registry contains the following diverse data from PCa patients who underwent radical prostatectomy: age at diagnosis, comorbidity, BMI, Gleason score, MRI results, and so on. We developed data preprocessing SW to handle multicenter unstructured data (Figure 3).

We developed Natural Language Processing (NLP) SW for preprocessing free text data, including pathologic results and clinicians' notes. NLP SW processes free text into clear terms with related values, such as Gleason's score and presence of metastasis, followed by converting these data into standard formats of the PROMISE CLIP registry.

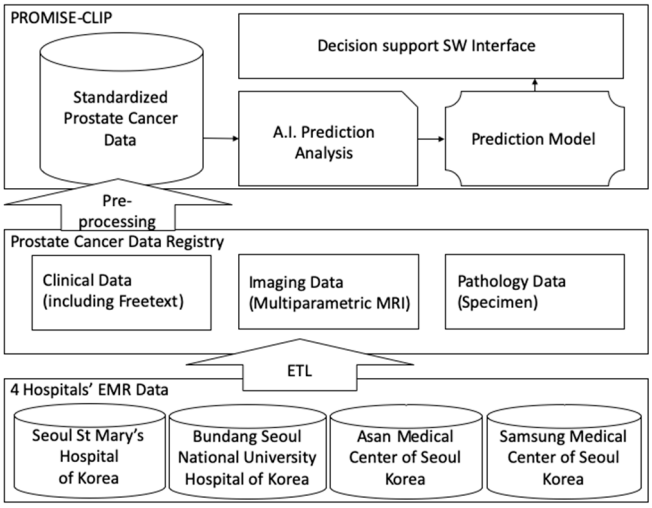

Figure 3. Conceptual diagram of PROMISE CLIP.

2.6. Ethics

The PROMISE CLIP procedures were performed in accordance with the Declaration of Helsinki and were approved by the Institutional Review Board of Catholic University (IRB number: KC18SNDI0512), Samsung Medical Center (IRB number: SMC201807069001), Bundang Seoul University Hospital (IRB number: B1808486102), and Asan Medical Center (IRB number: 2018-0963). Participant data were

de-identified and uploaded to the virtual machine in the cloud service: NAVER CLOUD PLATFORM (https://www.ncloud.com/). This platform has a role-based access control policy. Only permitted users can access the cloud platform.

3. Discussion

We designed PROMISE CLIP and initiated a multicenter, big data study to develop PCa SW for patients and physicians. Based on our findings, we drew the following conclusions.

First, PROMISE CLIP is a large-scale project for developing a Korean precision medicine service. PROMISE CLIP is one project of Intelligent SW Technology Development for Medical Data Analysis by National IT Industry Promotion Agency (NIPA): Dr. Answer. It established a consortium of Korean Data and Software-driven Hospital Consortium (K-Dash) in 2018 (http://dranswer.kr). K-Dash consists of 25 hospitals and 19 companies for eight diseases: cardiocerebrovascular disease, cardiac disorder, breast cancer, colorectal cancer, PCa, dementia, epilepsy, and childhood genetic and rare diseases. Among the eight diseases, PROMISE CLIP is meaningful as a multicenter project for PCa (http://dranswer.kr/disease/cancer1.php?tab=2).

Second, PROMISE CLIP developed a large-scale registry based on clinical, imaging, and pathology data from four tertiary hospitals. We put significant effort into gathering large-scale clinical, imaging, and pathology data. It is difficult to develop a registry with controlled clinical, imaging, and pathology data. Integration of data sources in hospitals is important and has challenges that need to be overcome [11]. We cooperated in harmony to develop the PROMISE CLIP registry. Collection of biopsy and prostatectomy biospecimens is an ideal characteristic of PCa registries [7]. The Cancer of the Prostate Strategic Urologic Research Endeavor (CaPSURE) had plans to collect biospecimens from both prostate biopsies and radical prostatectomies [12]. Diverse registries for PCa have been published and exert a growing role in advancing PCa research and care: CaPSURE [12], the Michigan Urological Surgery Improvement Collaborative (MUSIC) [13], and the Victorian Prostate Cancer Registry (PCR) [14]. In addition, the PROMISE CLIP registry has sufficient follow-up data for clinically relevant endpoints. Accordingly, the PROMISE CLIP registry has advantages for diverse AI projects and unsolved medical demands in PCa.

Third, the PROMISE CLIP project is meaningful for both PCa patients and clinicians because we expect intelligent SW for PCa. These SW help patients and clinicians decide the best treatment option. (1) The first SW is about the prediction of pathologic outcomes and treatment outcomes after definitive surgery that helps patients and clinicians decide the best treatment option. (2) The second SW is about visualization of treatment course with valuable markers that help to define the ideal patient population for aggressive follow-up or early postoperative ancillary treatment. (3) The third SW automatically generates accurate interpretation of multiparametric MRI. (4) The last SW is about accurate digital pathology of PCa to improve accuracy, reduce human error, and increase reproducibility. These SWs are described in detail below.

The first SW provides prediction of clinical stage based on machine learning from Korean EMR data. The SW will show the prediction of pathologic outcomes, biochemical recurrence, and survival rate using visualization library technology. We have plans to develop an Open API to provide a related service. LifeSemantics Corp. has contributed to the development of the first SW, who developed an mHealth management platform for patients [15,16].

The second SW provides visualization and calculation functionalities for an optimized view of PCa patient information. Clinicians access the most recent information at the right time to make treatment decisions. This SW is going to include treatment suggestions using machine learning algorithms. Seoul St. Mary's Hospital has contributed to the development of the second SW.

The third SW automatically generates diagnosis information from MRI images. The core technology involved Convolutional Neural Network (CNN) for registration, segmentation, lesion detection and characterization, and prognosis. Generative Adversarial Nets (GAN) Image augmentation for learning, Linear Regression, Random Forest, and Gradient Boosting Machine are included for prognosis.

In addition, we will conduct radiomics for analysis, including deep multimodal feature analysis. We can provide decoding services for PCa MRI, biopsy guides for PCa MRI/fusion, and prediction services for PCa. VUNO Inc. has contributed to the development of the second SW. VUNO Inc. has performed many projects in deep learning in medical image analysis [17–20].

The last SW demarcates the cancer area and calculates the Gleason's score from specimen scan images to designate the risk level of certain tissues from specimen scan images. This SW will use deep learning to classify patches according to Gleason pattern 3, 4, 5, etc. In addition, the SW will use weakly supervised learning and classic machine learning to predict PCa and Gleason score using results and characteristics of the patch. This SW is a support system for PCa diagnosis and is able to screen normal slides or notice risky areas before reading the specimen. It is possible to extend the SW using imaging data to obtain precise results. DeepBio Inc. plays a role in the development of this SW. DeepBio has conducted AI projects in both pathology and PCa [21].

The PROMISE CLIP study has plans to validate four intelligent SW both internally and externally. We are going to obtain additional large-scale medical data from four hospitals for a duration of two years, including clinical, imaging, and pathology data.

There are a few limitations of this study. First, we collected EMR data retrospectively. The PROMISE CLIP registry was unable to include all patient data. However, PROMISE CLIP has the potential to use diverse AI projects to tackle unsolved medical demands in PCa. Second, the PROMISE CLIP registry is for the Korean population. Future projects will need to collect multinational data. Third, we integrated EMR data from four hospitals to develop the registry. Future projects will need to collect addition data from diverse hospitals.

Although there are limitations, the PROMISE CLIP project will direct guidelines for intelligent SW development to solve challenging medical demands in PCa. The PROMISE CLIP registry plays an important role in advancing PCa research and care.

Author Contributions: J.P. wrote the article and contributed in building the project. M.J.R. also supported in writing the article and building the project. Y.H.P. gave medical advices and developing the standard of database of the proposed system. C.K.J., Y.C., H.G. gave medical advices and building the pathology database of the system. Y.H.P. and M.K. gave medical advices and building the clinical database of the system. S.I.H. gave medical advices and building the image database of the system. J.Y.L., C.-S.K., and H.J.L. supervised research and the project. J.Y.L., C.-S.K., and H.J.L. also gave medical advices for building the project. All twelve authors substantially contributed in each of their expertise.

Funding: This work was supported by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (2018-2-00861, Intelligent SW Technology Development for Medical Data Analysis).This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (NRF-2017M3A9B8069577).

Acknowledgments: Thank you Soo Jeong Nam at department of pathology, Asan Medical Center for the contribution for building the pathology database of the system. Thank you all of each hospital members who contributed their effort for the project.

Conflicts of Interest: There are no competing financial interests. The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **2018**, *68*, 7–30. [CrossRef] [PubMed]
2. Jung, K.-W.; Won, Y.-J.; Kong, H.-J.; Lee, E.S.; The Community of Population-based Regional. Cancer Registries Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2016. *Cancer Res. Treat.* **2019**, *51*, 417–430. [CrossRef] [PubMed]
3. Regnier-Coudert, O.; McCall, J.; Lothian, R.; Lam, T.; McClinton, S.; N'Dow, J. Machine learning for improved pathological staging of prostate cancer: A performance comparison on a range of classifiers. *Artif. Intell. Med.* **2012**, *55*, 25–35. [CrossRef] [PubMed]

4. Zupan, B.; Demšar, J.; Kattan, M.W.; Beck, J.; Bratko, I. Machine learning for survival analysis: A case study on recurrence of prostate cancer. *Artif. Intell. Med.* **2000**, *20*, 59–75. [CrossRef]
5. Cruz, J.A.; Wishart, D.S. Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Inform.* **2006**, *2*, 59–77. [CrossRef]
6. Fakoor, R.; Ladhak, F.; Nazi, A.; Huber, M. Using deep learning to enhance cancer diagnosis and classification. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013.
7. Gandaglia, G.; Bray, F.; Cooperberg, M.R.; Karnes, R.J.; Leveridge, M.J.; Moretti, K.; Murphy, D.G.; Penson, D.F.; Miller, D.C. Prostate Cancer Registries: Current Status and Future Directions. *Eur. Urol.* **2016**, *69*, 998–1012. [CrossRef] [PubMed]
8. Lee, D.H.; Lee, S.H.; Rha, K.H.; Choi, I.Y.; Lee, J.Y.; Kim, S.W.; Lee, S.; Hong, S.K.; Byun, S.-S.; Jeong, I.G.; et al. The establishment of k-cap (the multicenter korean prostate cancer database). *Korean J. Urol.* **2013**, *54*, 229–233. [CrossRef] [PubMed]
9. Kim, C.-S.; Lee, J.Y.; Chung, B.H.; Kim, W.-J.; Fai, N.C.; Hakim, L.; Umbas, R.; Ong, T.A.; Lim, J.; Letran, J.L.; et al. Report of the second asian prostate cancer (a-cap) study meeting. *Prostate Int.* **2017**, *5*, 95–103. [CrossRef] [PubMed]
10. Choi, I.Y.; Park, S.; Park, B.; Chung, B.H.; Kim, C.-S.; Lee, H.M.; Byun, S.-S.; Lee, J.Y. Development of prostate cancer research database with the clinical data warehouse technology for direct linkage with electronic medical record system. *Prostate Int.* **2013**, *1*, 59–64. [CrossRef]
11. Li, B.; Li, J.; Jiang, Y.; Lan, X. Experience and reflection from China's Xiangya medical big data project. *J. Biomed. Inform.* **2019**, *93*, 103149. [CrossRef]
12. Porten, S.P.; Cooperberg, M.R.; Konety, B.R.; Carroll, P.R. The example of CaPSURE: Lessons learned from a national disease registry. *World J. Urol.* **2011**, *29*, 265–271. [CrossRef] [PubMed]
13. Montie, J.E.; Linsell, S.M.; Miller, D.C. Quality of Care in Urology and the Michigan Urological Surgery Improvement Collaborative. *Urol. Pract.* **2014**, *1*, 74–78. [CrossRef]
14. Evans, S.M.; Millar, J.L.; Wood, J.M.; Davis, I.D.; Bolton, D.; Giles, G.G.; Frydenberg, M.; Frauman, A.; Costello, A.; McNeil, J.J.; et al. The prostate cancer registry: Monitoring patterns and quality of care for men diagnosed with prostate cancer. *BJU Int.* **2013**, *111*, E158–E166. [CrossRef] [PubMed]
15. Kwon, H.; Lee, S.; Jung, E.J.; Kim, S.; Lee, J.-K.; Kim, D.K.; Kim, T.-H.; Lee, S.H.; Lee, M.K.; Song, S.J.; et al. An mhealth management platform for patients with chronic obstructive pulmonary disease (efil breath): Randomized controlled trial. *JMIR Mhealth Uhealth* **2018**, *6*, e10502. [CrossRef] [PubMed]
16. Eysenbach, G.; Lee, J.-H.; Shin, S.-Y.; Soh, J.Y.; Cha, W.C.; Chang, D.K.; Hwang, J.H.; Kim, K.; Rha, M.; Kwon, H.; et al. Development and Validation of a Multidisciplinary Mobile Care System for Patients With Advanced Gastrointestinal Cancer: Interventional Observation Study. *JMIR mHealth uHealth* **2018**, *6*, e115.
17. Jung, K.-H.; Park, H.; Hwang, W. Deep Learning for Medical Image Analysis: Applications to Computed Tomography and Magnetic Resonance Imaging. *Hanyang Med. Rev.* **2017**, *37*, 61–70. [CrossRef]
18. Bae, B.-U.; Bae, W.; Jung, K.-H. Improved deep learning model for bone age assessment using triplet ranking loss. In Proceedings of the 1st Conference on Medical Imaging with Deep learning (MIDL 2018), Amsterdam, The Netherlands, 4–6 July 2018.
19. Kwon, J.-M.; Lee, Y.; Lee, Y.; Lee, S.; Park, H.; Park, J. Validation of deep-learning-based triage and acuity score using a large national dataset. *PLoS ONE* **2018**, *13*, e0205836. [CrossRef] [PubMed]
20. Park, S.; Hwang, W.; Jung, K.-H. Integrating Reinforcement Learning to Self Training for Pulmonary Nodule Segmentation in Chest X-rays. *arXiv* **2018**, arXiv:1811.07216.
21. Chang, H.Y.; Jung, C.K.; Woo, J.I.; Lee, S.; Cho, J.; Kim, S.W.; Kwak, T.-Y. Artificial Intelligence in Pathology. *J. Pathol. Transl. Med.* **2018**, *53*, 1–12. [CrossRef] [PubMed]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).