

ARTIFICIAL INTELLIGENCE THEORY AND APPLICATIONS



RESEARCH ARTICLES

Digital Health Navigator: Preliminary Work on a Personal Health Assistant Software for All Health Literacy Level Users in Turkey

A New Similarity Method for Tourism Recommendation Systems

Prediction of Acute Burn-Induced Coagulopathy Risk with Machine Learning Models

Performance Analysis of Efficient Deep Learning Models for Multi-Label Classification of Fundus Image

Artificial Intelligence Based Chatbot in E-Health System

REVIEW ARTICLE

Modern Computer Tomography with Artificial Intelligence and Deep Learning Applications



Volume 3 – Issue 2

01.10.2023

e-ISSN: 2757-9778

Publisher: İzmir Bakırçay University

www.aitajournal.com

<https://dergipark.org.tr/en/pub/aita>

Journal Board

AITA

Artificial Intelligence Theory and Applications

Founding Editor

Prof. Mustafa Berktaş, M.D.
İzmir Bakırçay University

Editor-in-chief

Prof. Kadir Hızıroğlu, Ph.D.
İzmir Bakırçay University

Editors

Prof. Deniz Kılınc, Ph.D.
İzmir Bakırçay University

Prof. Orhan Er, Ph.D.
İzmir Bakırçay University

Assoc. Prof. Onur Doğan, Ph.D.
İzmir Bakırçay University

Editorial Team

Ali Pişirgen, Publishing Editor, , İzmir Bakırçay University
Ali Mert Erdoğan, Layout Editor, İzmir Bakırçay University
Mehmet Gencer, Content Editor, İzmir Bakırçay University
Mustafa Furkan Aksu, Design Editor, İzmir Bakırçay University
Betül Yürdem, Social Media Editor, İzmir Bakırçay University
Ourania Areta, Phd, Language Editor, , İzmir Bakırçay University

Editorial Board

Adil Alpkoçak, İzmir Bakırçay University
Ahmet Emin Erbaycu, İzmir Bakırçay University
Ahu Pakdemirli, University of Health Sciences
Banu Başok İşbilen, University of Health Sciences
Bedile İrem Tiftikçioğlu, İzmir Bakırçay University
Cem Gök, Pamukkale University
Dilek Orbatu, University of Health Sciences
Elif Güler Kazancı, University of Health Sciences
Feyzullah Temurtaş, Bandırma 17 Eylül University
Gizem Çalıbaşı Koçal, Dokuz Eylül University
Halil İbrahim Cebeci, Sakarya University
Hanefi Özbek, İzmir Bakırçay University
Haşim Özgür Tabakoğlu, İzmir Bakırçay University
Hilal Arslan, Yıldırım Beyazıt University
İbrahim Onur Alici, İzmir Bakırçay University
İhsan Hakan Selvi, Sakarya University
Kadir Gök, İzmir Bakırçay University
Kemal Polat, Abant İzzet Baysal University
Keziban Seçkin Codal, Yıldırım Beyazıt University
Mehmet Kemal Güllü, İzmir Bakırçay University
Mehmet Bakır, Yozgat Bozok University
Mehmet İtik, İzmir Democracy University
Mehmet Sağbaş, İzmir Bakırçay University
Murat Korkmaz, Yozgat Bozok University
Nejat Yumuşak, Sakarya University
Özge Tüzün Özmen, İzmir Bakırçay University
Senem Alkan Özdemir, University of Health Sciences
Ümit Hüseyin Kaynar, İzmir Bakırçay University
Volkan Akdoğan, İskenderun Technical University
Yusuf Murat Erten, İzmir University of Economics

International Advisory Board

Abdulaziz Ahmed, University of Minnesota Crookston, United States
Carlos Fernandez-Llatas, Valencia Polytechnic University, Spain
Erdal Çoşgun, Seattle Microsoft Genomics, United States
Mehmet Emin Aydın, University of the West of England, United Kingdom
M. Akhil Jabbar, Vardhaman College of Engineering, India
Salih Tütün, Institute for Public Health Washington University, United States
Sanju Mishra Tiwari, Universidad Autonoma de Tamaulipas, Mexico
Yadgar I. Abdulkarim, Charmo University, Iraq

International Indexes



Academic Resource Index: Research Bible

<https://www.researchbib.com/view/issn/2757-9778>



Index Copernicus International

<https://journals.indexcopernicus.com/search/details?id=125039&lang=en>

About Journal

Artificial Intelligence Theory and Applications (AITA) provides coverage of the most significant work on principles of artificial intelligence, broadly interpreted. The scope of research we cover encompasses contributions of lasting value to any area of artificial intelligence. To be accepted, a paper must be judged to be truly outstanding in its field. AITA is interested in work in core artificial intelligence and at the boundaries, both the boundaries of sub-disciplines of artificial intelligence and the boundaries between artificial intelligence and other fields.

Scope

The best indicator of the scope of the journal is provided by the areas covered by its Editorial Board in theoretical (artificial intelligence and computing methodologies) and practical (artificial intelligence applications and applied computing) ways. These areas change from time to time, as the field evolves.

Year : 2023
Volume : 3
Issue : 2

Table of Content

Page

Digital Health Navigator: Preliminary Work on a Personal Health Assistant Software for All Health Literacy Level Users in Turkey..... <i>by Nazlı Tokatlı, Muhammed Tayyip Koçak, Seda Kırtay, Fatma Nur Kömür, Gürkan Göztepeli, Beyza Gül, Halis Altun</i>	67-76
A New Similarity Method for Tourism Recommendation Systems..... <i>by Eren Türkel, Adil Alpkoçak</i>	77-91
Prediction of Acute Burn-Induced Coagulopathy Risk with Machine Learning Models..... <i>by Murat Ali Çınar</i>	92-104
Performance Analysis of Efficient Deep Learning Models for Multi-Label Classification of Fundus Image..... <i>by Muhammed Pektaş</i>	105-112
Artificial Intelligence Based Chatbot in E-Health System..... <i>by Kamil Akarsu, Orhan Er</i>	113-122
Modern Computer Tomography with Artificial Intelligence and Deep Learning Applications..... <i>by Çoşkun Deniz</i>	123-136

Digital Health Navigator: Preliminary Work on a Personal Health Assistant Software for All Health Literacy Level Users in Turkey

Nazlı Tokatlı^{a†}, Muhammed Tayyip Koçak^b, Seda Kırtay^b, Fatma Nur Kömür^c
Gürkan Göztepelı^b, Beyza Gül^a, Halis Altun^b

^a Computer Engineering, İstanbul Health and Technology University, İstanbul, Türkiye

^b Software Engineering, İstanbul Health and Technology University, İstanbul, Türkiye

^c Audiology, İstanbul Health and Technology University, İstanbul, Türkiye

[†] nazli.tokatli@istun.edu.tr, corresponding author

RECEIVED MAY 23, 2023

ACCEPTED JULY 26, 2023

CITATION Tokatlı, N., Koçak, M.T., Kırtay, S., Kömür, F.N., Göztepelı, G., Gül, B., Altun, H. (2023). Digital health navigator: Preliminary work on a personal health assistant software for all health literacy level users in Turkey. *Artificial Intelligence Theory and Applications*, 3(2), 67-76.

Abstract

Today's digital health terminology is advanced medical technologies that include computer-assisted therapy, smartphone apps, and wearable technologies. These technologies offer significant potential for improving access to immediate medical care, efficiency, clinical effectiveness, and personalization of many health problem therapies. In this ongoing research paper, we propose the initial design steps of a personalized health assistant application. The proposed application can be classified as a mobile health app and not a telemedicine application. The idea behind this application is to reduce the physicians' workload in hospitals while providing health care to the community with different health literacy levels by easily using the application when general assistance about any health issues or an overall health and wellness improvement is required.

Keywords: digital health; mHealth; eHealth literacy; digital transformation; mobile applications

1. Introduction

Tablets and mobile phones are necessary for today's society. Nearly every country worldwide has begun defining and establishing the foundation of digital transformation in many industries. The healthcare system is among these fields since the health systems have started facing a shortage of healthcare professionals and a continuous workload increase of physicians, especially during the pandemic [1]. In addition, an increase in the amount of medical information may be accessed using the Internet by anybody. All of the aforementioned factors result in the development of a wide range of mobile health applications. Healthcare-focused mobile applications serve as digital conduits of health-related knowledge and research, facilitating both healthcare practitioners and patients in augmenting health treatments and public health. By transforming mobile platforms into regulated medical devices, these applications embody an evolution of technology in healthcare. Superseding traditional technologies, mobile apps have established new avenues for communication with autistic individuals and are increasingly being used as

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than AITA must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from info@aitajournal.com

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2023 İzmir Bakırçay University

digital medical identification tools, integrated within smartphone lock screens [2]. Such applications leverage sensor-collected data to assess an array of health-related parameters, including physical activity, body part imaging, and overall health status [3]. Moreover, mobile sensors can capture crucial health parameters, enabling the identification of daily activities and lifestyle patterns [4].

Governments opened the door for the legalization of the public's use of those apps that can be purchased [5]. Statistics cited by IQVIA in an article published by Medical Device Network in 2021 indicate that more than 350,000 digital health applications are currently accessible on the global market. Moreover, based on GlobalData's recent report Thematic Research: Mobile Health Apps forecasts that the regulated medical apps market will reach \$12.1 billion by 2030[6,7]. These applications can be downloaded through the apple store as well as the google play store [8,9]. Furthermore, Figure 1 depicts the general idea of Miguel et al. research. Their research showed how they classified mobile health applications and categorized them based on application features and usage goals [10]. The use cases range from wellness and fitness services to complicated applications for diagnosis and support in aftercare settings. mhealth applications may also be used for educational purposes, as well as for tracking, monitoring, or managing the chronic conditions of patients by medical professionals. However, it is still exceedingly difficult for patients and healthcare professionals to identify and evaluate high-quality apps in the various app stores. There is currently no common quality standard for mhealth applications [11]. Additionally, Margaret R. Emerson et al.'s study has shown that dissemination of health information through mobile devices (mHealth) increases the amount of information that is available. However, it also leads to challenges in terms of ensuring that the materials are appropriate for and understandable by all community members, regardless of their health literacy levels [12]. A certain need to be considered is the social structure of the community while designing mhealth applications. Otherwise, using little complex mhealth applications by low or basic health literacy level users may increase medical errors, illness, and compromised public health.

Lastly, we can conclude that all mhealth applications are not suitably designed for all community users of all ages and literacy levels. Additionally, they provide or target to solve a specific health problem. Consequently, one of the primary purposes of this study is to build a mhealth application that provides individualized support and early guidance for daily health needs.



Figure 1. The general classification of mHealth applications

2. Related Work

There are a significant number of mobile health applications available across the globe [9]. The scope of these applications extends from advice on leading a healthy lifestyle every day to professionals who give personal health care. There are a variety of approaches that can be learned through mobile health applications, including learning the disease-related process, improving patient care to facilitate treatment and diagnosis, electronic prescription, clinical support, and personal care, and e-learning to increase health literacy [13,14]. These are some of the primary purposes of mobile health applications.

Given the area covered by the current applications, it can basically be divided into two main parts: Applications used by health professionals and individual care applications used by community members. Health professionals use mobile health applications for various documents and encyclopaedia containing detailed medical information [15,16] memory tests related to anatomy [17,18], calculation of drug dosages and reference values [19,20,21], social networking applications that facilitate communication with other healthcare professionals [22,23,24], tracking and follow-up patient care [25,26].

Patient and non-patient society members use mobile health applications, such as monitoring personal care and life changes [27,28], record and store medical information and health history, as well as drug and appointment management [29,30,31,32], together with general health education, first aid, and rehabilitation practices [33,34].

Lastly, although the existing applications differ in terms of their subjects and functions, our PHAS is more comprehensive than the others. For example, our application will be accessible to both health professionals as well as other members of the community. Thus, patient-specific treatment, control, medical follow-up, continuous patient monitoring, and performance evaluation will become easier. In addition, we believe that the proposed PHAS will definitely raise the health literacy of individuals and lessen the strain on healthcare centers since the application provides the user with the right direction to follow about his/her health during the life course.

3. PHAS Technical Structure

The proposed PHAS differentiates itself from its counterpart by trying to cover all areas related to personal health rather than concentrating on one health problem and its solution. Figure 2 shows PHAS's main services. Lastly, in the following sections, we will examine the detailed structure of each service.

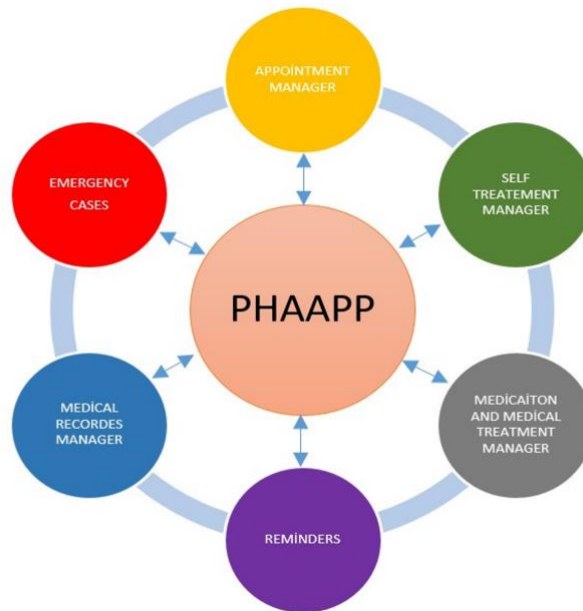


Figure 2. PHAS services

3.1. PHAS Appointment Manager

We believe this phase will significantly alleviate the strain on the healthcare system. Due to varying degrees of health literacy in the community, many patients may schedule appointments with the incorrect hospital department. The physician will then refer the patient to the relevant department to diagnose and treat the health issue. This procedure can waste patient and physician time and prevent other patients from scheduling an appointment with the relevant department.

At this time, we propose implementing an appointment manager in our PHAS to resolve this issue. Figure 3 illustrates the appointment manager's phases. First, we intend to visit specialists from several health departments and compile a comprehensive list of symptoms associated with a certain health concern, such as developing a general list of typical symptoms associated with respiratory system issues and presenting the symptom list to the patient in an understandable and straightforward manner, allowing them to choose from the list. This can be accomplished by integrating a Chatbot into our appointment manager [35,36,37].

Based on the patient's choices from the symptom, our manager will suggest that the patient book an appointment with the right physician, if needed. The appointment booking system flowchart is shown in Figure 4. The parameter `symptom_count_threshold` specifies the minimum number of symptoms from the symptom list that must be selected by the patient to be directed to the correct

appointment system. We intend to use static threshold values for symptom count and X value for the time being and leave the use of dynamic threshold values for future work.

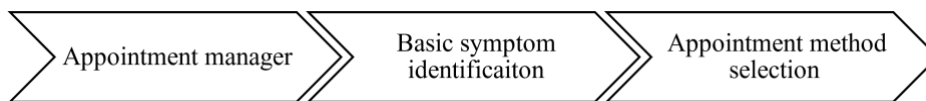


Figure 3. Appointment manager in PHAS

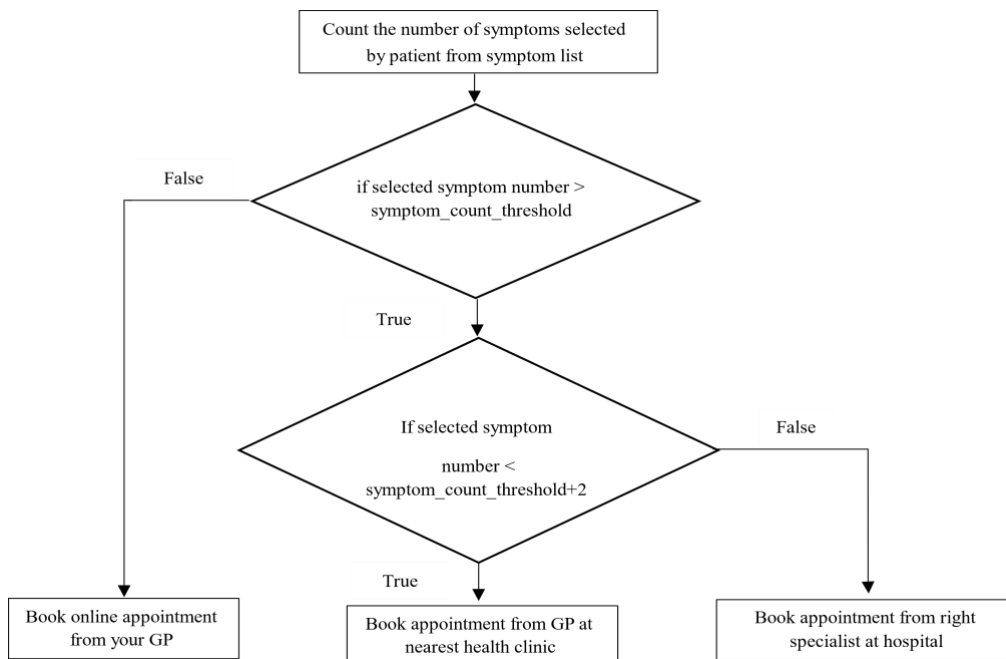


Figure 4. Flowchart for an appointment booking system

3.2. Self-treatment Manager

Many nations encounter challenges in providing access to quality healthcare systems for diverse age groups [38]. The health systems of the world must be resilient. For example, COVID-19 assessed their ability to withstand acute shocks, but they must also withstand long-term trials and threats. Articles published in the Lancet on global and planetary health illustrated the increasing burden of non-communicable diseases that will accompany aging populations due to climate change and a decline in nutritional quality. Furthermore, the most significant future challenges for a healthcare system will be issues with monitoring and addressing unmet mental health burdens in various age groups [39,40]. On the other side, governments start preparing and looking for a solution to decrease the impact of these difficulties on their health system by educating and encouraging the community to utilize mhealth applications and building general-purpose websites approved by the Ministry of Health. All of these applications strive to provide advice and a variety of information regarding learning how to maintain a healthy lifestyle over the course of one's life [41,42].

The proposed self-treatment manager in our PHAS will assist in instructing users of all ages to follow necessary daily actions that will protect them as long as possible from hospitalization and health-related interventions. The PHAS self-treatment manager sections and general working mechanism are shown in Figures 5 and 6, respectively.

Finally, in this paper, we will not dive into the implementation details of these sections, which are left as further work.



Figure 5. PHAS self-treatment manager sections

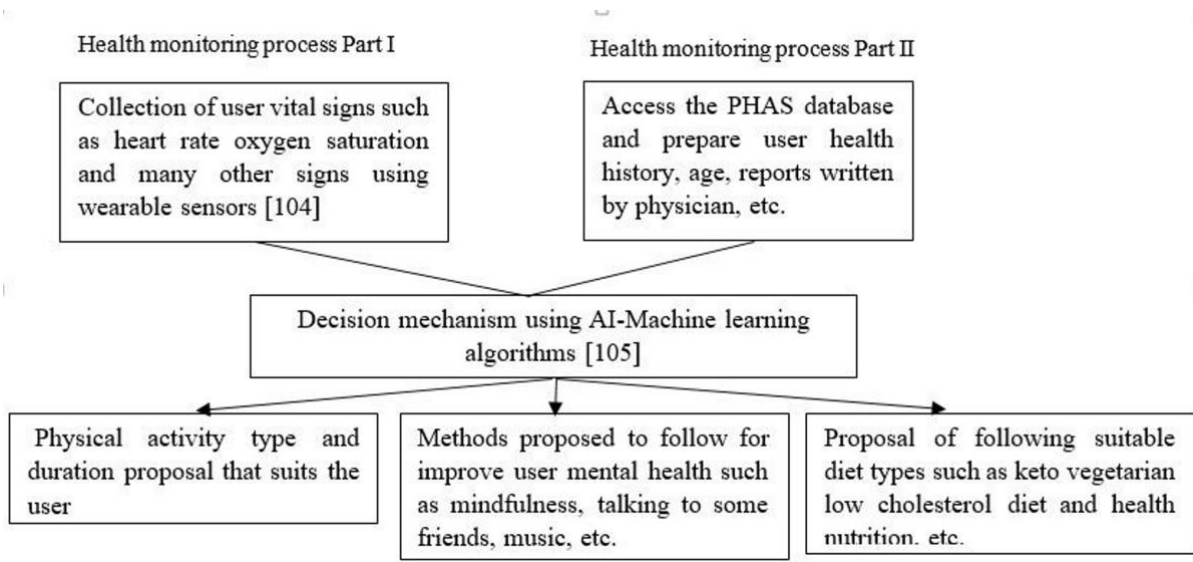


Figure 6. PHAS self-treatment manager working mechanism

3.3. Medical Record and Remainder Manager

Even the implementation of these two sections is not new and is available in many healthcare systems [43]. We believe integrating these managers into PHAS will simplify the access process and save the user time. We plan to save the user time by constructing a separate health database for each user that may later be used by machine learning algorithms from other PHAS sections to analyse the data and give basic feedback about the action type that shall be followed by the user to prevent any complicated health issues (early diagnosis system). Figure 7 and Figure 8 show both remainder and medical records manager services, respectively.

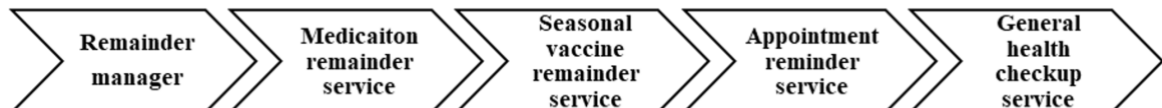


Figure 7. PHAS remainder manager sections

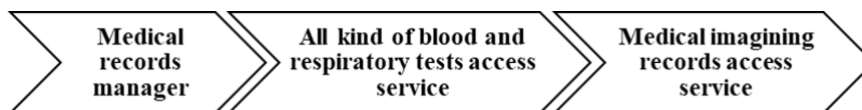


Figure 8. PHAS medical records manager sections

3.4. Medication and Medical Treatment Manager

As mentioned in the previous sections about raising population knowledge in all health-related fields, one of these areas is drugs and medications. In the market, there are few mhealth applications for only providing information about a specific group of drugs rather than all medication types [44,45]. One of the services provided by the PHAS medication manager is to provide general information about any medications (herbal medications included), such as an instruction list. To construct these lists, we plan to work with a team of pharmacists. Also, it allows patients to report to their physicians in case they have or show any side effects while taking the prescribed medication. In addition, allowing the patient to acquire the prescription from the nearest pharmacy to his or her location. The detailed structure for medication and medical treatment is shown in Figure 9. Lastly, there is a need to mention that the prescription uploading service is not new and is already available in the Turkish health system [43].

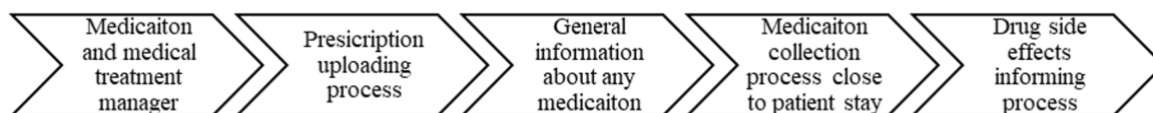


Figure 9. PHAS medication and medical treatment manager

4. Experimental And Implementation Methodologies

We can divide the phas implementation phase to two major parts. The first one using software platforms to design a mobile application that execute on both android and iOS operating systems. For this purpose, using react, flutter will allow creating cross-platform mobile applications with a single codebase [46,47]. The second part is the usage of backend services, many technologies such as Node.js, Ruby on Rails, and Django are used [48,49]. In our research we decide to use Node.js to create a backend server application because of its wide range availability of frameworks and libraries [50]. For data storage and processing, many database technologies are utilized. According to the latest survey research conducted by Stack Overflow in 2022 [51], PostgreSQL and MySQL are the most popular relational databases, whereas MongoDB is the most popular non-relational database. Thus, we use MongoDB due to its capability in managing large amounts of data and has built-in support for horizontal scaling, making it ideally suited for use in modern, high-traffic mobile applications [52]. Also, it is essential to utilize version control systems to monitor changes to the source code and to allow cooperation among other developers. GraphQL will be used as a flexible and efficient data retrieval from a server [53]. Git platform will used as a version control system because it facilitates user collaboration and stores a complete history of commits [54]. Moreover, for providing personalized health recommendations and treatment plans for users, we need to use different data sets collection tools and platforms for training our machine learning models such as ukbiobank, Kaggle, U.S. Department of Health & Human Services (HHS), Centers for Disease Control and Prevention (CDC), National Institutes of Health (NIH). [55,56,57,58,59]. Lastly TensorFlow [60] is an end-to-end

machine learning platform which is very popular and thus will be used to create machine learning models while utilizing the obtained data sets from the aforementioned platforms. The summary of technologies that planned to be used in our paper are described with their advantages in Table 1 and Table 2 respectively.

Table 1. Planned technology

Technologies	Free to use	Is Open Source
React Native	Yes	Yes
Node.js	Yes	Yes
MongoDB	Yes, for the community edition	Yes
GraphQL	Yes	Yes
Git	Yes	Yes
TensorFlow	Yes	Yes

Table 2. Planned technology and their advantages

Technology Name	Description	Advantages
React Native	Enables the development of native apps using React for platforms including Android, iOS, and more	<ul style="list-style-type: none"> – React Native is Community-driven – Maximum code reuse & cost saving – Strong performance for mobile environments – Modular and intuitive architecture similar to React
Node.js	Node.js is an open-source, cross-platform JavaScript runtime environment	<ul style="list-style-type: none"> – Helps in Building Cross-Platform Applications – High-performance for Real-time Applications – Easy Scalability for Modern Applications – Community Support to Simplify Development
MongoDB	MongoDB is a document database with the scalability, flexibility, and necessary querying and indexing	<ul style="list-style-type: none"> – Full cloud-based developer data platform – Flexible document schemas – Cost-effective – Powerful querying and analytics
GraphQL	A runtime that responds to API queries with already-existing data is called GraphQL. Additionally, GraphQL offers flexibility, provides robust developer tools, and makes it simpler to adapt APIs over time	<ul style="list-style-type: none"> – GraphQL provides a flexible structure – Best for complex systems and microservices – No over-fetching and under-fetching problems
Git	Git is a distributed version control system that is free and open source and is made to efficiently and quickly handle projects of all sizes	<ul style="list-style-type: none"> – Git is free and open source – Git is fast compared to others because each developer has access to a local repository with a complete history of commits – Branching and merging operations are simple and cost-effective
TensorFlow	Machine learning models may be easily created with TensorFlow for desktop, mobile, online, and cloud platforms by both beginners and specialists	<ul style="list-style-type: none"> – Easy model building – Robust ML production anywhere – Powerful experimentation for research

5. Conclusion and Further Research

Undoubtedly, the healthcare system will encounter many issues in the future. Countries should strengthen their healthcare systems and alleviate the strain on them. Then, they can offer quality health care to all community levels. By utilizing contemporary technology and smartphones, the design and use of the mobile application in health areas will significantly increase the health literacy level of the population, thereby educating and instructing users to maintain good health and preventing them from visiting hospitals, thus reducing the workload of all healthcare professionals. Most designed and proposed mobile health applications focus on a single health concern rather than assisting the user in all health-related areas. The proposed PHAS will host a wide range of technologies, such as the adoption of wearable sensors, machine learning algorithms used for analysing user data, and many others during its design process. We confidently can say after PHAS implementation completion in the near future, and our software will provide a solid foundation that addresses all types of users' health-related demands and which path they should follow to live a healthy lifestyle to keep them as far away from hospitalization and physician assistance as possible.

Acknowledgement



This article was published as a proceeding of 14th Turkish Congress of Medical Informatics Association in 2023

References

- [1] Boniol, M., Kunjumen, T., Nair, T. S., Siyam, A., Campbell, J., & Diallo, K. (2022). The global health workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage?. *BMJ global health*, 7(6), e009316.
- [2] Pires, I. M., Marques, G., Garcia, N. M., Flórez-Revuelta, F., Ponciano, V., & Oniani, S. (2020). A research on the classification and applicability of the mobile health applications. *Journal of personalized medicine*, 10(1), 11.
- [3] Marques, G., Pitarma, R., M. Garcia, N., & Pombo, N. (2019). Internet of things architectures, technologies, applications, challenges, and future directions for enhanced living environments and healthcare systems: a review. *Electronics*, 8(10), 1081.
- [4] Hong, Y. J., Kim, I. J., Ahn, S. C., & Kim, H. G. (2010). Mobile health monitoring system based on activity recognition using accelerometer. *Simulation Modelling Practice and Theory*, 18(4), 446-455.
- [5] European mhealth hub — health apps repositories in europe. Retrieved January 10, 2023, from <https://mhealth-hub.org/health-apps-repositories-in-europe>
- [6] GlobalData. (2023, February 10). Major trends for regulated medical apps. *Medical Device Network*. <https://www.medicaldevice-network.com/comment/regulated-medical-apps/>
- [7] Mobile health apps - Thematic intelligence. (2023, May 30). Market Research Reports & Consulting | GlobalData UK Ltd. https://www.globaldata.com/store/report/mobile-health-apps-theme-analysis/?_gl=1*1n3nn7g*_ga*NTk3MjEyNzM1LjE2ODk3ODU4NDU.*_ga_8SZ1HHP33J*MTY4OTc4NTg0Ni4xLjE4MTY4OTc4NjE3NC40Mi4wLjA
- [8] Aitkin, M., Clancy, B., & Nass, D. (2017). The growing value of digital health. *IQVIA Institute for Human Data Science*, 1-76.
- [9] mhealth apps market size & share report. Retrieved January 10, 2023, from <https://www.grandviewresearch.com/industry-analysis/mhealth-app-market>
- [10] Pires, I. M., Marques, G., Garcia, N. M., Flórez-Revuelta, F., Ponciano, V., & Oniani, S. (2020). A research on the classification and applicability of the mobile health applications. *Journal of personalized medicine*, 10(1), 11.
- [11] Gütekriterien-Kernset für mehr Qualitätstransparenz bei digitalen Gesundheitsanwendungen. Retrieved January 5, 2023, from https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Studienbericht_AppQ_191028.pdf
- [12] Emerson, M. R., Buckland, S., Lawlor, M. A., Dinkel, D., Johnson, D. J., Mickles, M. S., ... & Watanabe-Galloway, S. (2022). Addressing and evaluating health literacy in mHealth: a scoping review. *Mhealth*, 8.
- [13] Griffiths, F., Cave, J., Boardman, F., Ren, J., Pawlikowska, T., Ball, R., ... & Cohen, A. (2012). Social networks—the future for health care delivery. *Social science & medicine*, 75(12), 2233-2241.
- [14] Luxton, D. D., June, J. D., Sano, A., & Bickmore, T. (2016). Intelligent mobile, wearable, and ambient technologies for behavioral health care. In *Artificial intelligence in behavioral and mental health care* (pp. 137-162). Academic Press.
- [15] Oxford medical dictionary - apps on google play. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=com.mobisystems.msdict.embedded.wireless.oxford.concisemedical>.
- [16] Vinay, K. V., & Vishal, K. (2013). Smartphone applications for medical students and professionals. *Journal of Health and Allied Sciences NU*, 3(01), 59-62.
- [17] Speed muscles md on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/speed-musclesmd/id315978709>.
- [18] Speed muscles md - apps on google play. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=com.bessiambre.speedMuscles&hl=enUS>

- [19] Mdcalc medical calculator on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/mdcalc-medical-calculators-decision-support/id1001640662>.
- [20] Mdcalc medical calculator. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=com.mdaware.mdcalc>.
- [21] Drug infusion - iv medications on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/drug-infusion-iv-medications/id311096766>.
- [22] Doximity apps on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/developer/dximity/id391582379>.
- [23] Amion - physician calendar - apps on google play. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=com.dximity.amiondroid>.
- [24] DocbookMD. Retrieved on January 10, 2023, from <https://www.docbookmd.com/>.
- [25] T'el'ecardio - apps on google play. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=info.telecardiologie.app&hl=enUS>.
- [26] Apple's top 50 ipad apps for doctors. Retrieved on January 10, 2023, from <https://www.mobihealthnews.com/13638/apples-top-50-ipad-apps-for-doctors/>
- [27] Webmd: Symptom checker - apps on google play. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=com.webmd.android&hl=en>.
- [28] Webmd: Symptom checker on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/webmd-symptoms-doctors-rx/id295076329>.
- [29] Capzule on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/capzule/id386321118>.
- [30] Mera aspataal - apps on google play. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=com.mahifi.myhospital&hl=en>.
- [31] Healthnotes viewer on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/healthnotes-viewer/id917722770>.
- [32] Healthnotes viewer - apps on google play. Retrieved on January 10, 2023, from <https://play.google.com/store/apps/details?id=com.aisle7.kiosviewer&hl=en>.
- [33] drawmd-patient education on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/drawmd-patient-education/id1024211520>.
- [34] Vueme on the app store. Retrieved on January 10, 2023, from <https://apps.apple.com/us/app/vueme/id437850916>.
- [35] Wei, Z., Liu, Q., Peng, B., Tou, H., Chen, T., Huang, X. J., ... & Dai, X. (2018, July). Task-oriented dialogue system for automatic diagnosis. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 201-207).
- [36] Kim, G.W., Lee, D.H., IEEE Access 7, 9419 (2019). DOI 10.1109/ACCESS.2019.2891710.
- [37] Tchango, A. F., Goel, R., Wen, Z., Martel, J., & Ghosn, J. (2022). DDXPlus: A New Dataset For Automatic Medical Diagnosis. Proceedings of the Neural Information Processing Systems-Track on Datasets and Benchmarks, 2.
- [38] Haakenstad, A., Yearwood, J. A., Fullman, N., Bintz, C., Bienhoff, K., Weaver, M. R., ... & Gupta, V. K. (2022). Assessing performance of the Healthcare Access and Quality Index, overall and by select age groups, for 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *The Lancet Global Health*, 10(12), e1715-e1743.
- [39] T.L.G. Health, *The Lancet Global Health* 11(1), e1 (2023). DOI 10.1016/S2214-109X(22)00521-6. URL [https://doi.org/10.1016/S2214-109X\(22\)00521-6](https://doi.org/10.1016/S2214-109X(22)00521-6)
- [40] T.L.P. Health, *The Lancet Planetary Health* 5(7), e395 (2021). DOI 10.1016/S2542-5196(21)00181-9. URL [https://doi.org/10.1016/S2542-5196\(21\)00181-9](https://doi.org/10.1016/S2542-5196(21)00181-9)
- [41] Live well - nhs. Retrieved on January 10, 2023, from <https://www.nhs.uk/live-well/>.
- [42] Healthy lifestyle — hhs.gov. Retrieved on January 10, 2023, <https://www.hhs.gov/programs/prevention-and-wellness/healthy-lifestyle/index.html>.
- [43] e-nabız — t.c. sağlık bakanlığı. Retrieved on January 10, 2023, from <https://enabiz.gov.tr>.
- [44] García-Sánchez, S., Somoza-Fernández, B., De Lorenzo-Pinto, A., Ortega-Navarro, C., Herranz-Alonso, A., & Sanjurjo, M. (2022). Mobile health apps providing information on drugs for adult emergency care: Systematic search on app stores and content analysis. *JMIR mHealth and uHealth*, 10(4), e29985. <https://doi.org/10.2196/29985>
- [45] Kim, B. Y., Sharafoddini, A., Tran, N., Wen, E. Y., & Lee, J. (2018). Consumer mobile apps for potential drug-drug interaction check: Systematic review and content analysis using the mobile app rating scale (MARS). *JMIR mHealth and uHealth*, 6(3), e74. <https://doi.org/10.2196/mhealth.8613>
- [46] React Native. (n.d.). Retrieved July 25, 2023, from <https://reactnative.dev/>
- [47] Flutter. (n.d.). Retrieved July 25, 2023, from <https://flutter.dev/>
- [48] Ruby. (n.d.). Retrieved July 25, 2023, from <https://www.ruby-lang.org/en/>
- [49] Ruby on Rails. (n.d.). Retrieved July 25, 2023, from <https://rubyonrails.org/>
- [50] Node.js. (n.d.). Retrieved July 25, 2023, from <https://nodejs.org/en/>
- [51] Stack Overflow. (n.d.). Retrieved July 25, 2023, from <https://survey.stackoverflow.co/2022/>
- [52] MongoDB. (n.d.). Retrieved July 25, 2023, from <https://www.mongodb.com/>
- [53] GraphQL. (n.d.). Retrieved July 25, 2023, from <https://graphql.org/>
- [54] Chacon, S., & Straub, B. (2014). *Pro Git*. Apress.
- [55] HealthData.gov. (n.d.). Retrieved July 25, 2023, from <https://www.healthdata.gov/>
- [56] Centers for Disease Control and Prevention. (n.d.). Data & Statistics. Retrieved July 24, 2023, from <https://www.cdc.gov/datastatistics/index.html>
- [57] National Institutes of Health. (n.d.). Data Sharing Repositories. Retrieved July 24, 2023, from <https://www.nih.gov/health-information/nih-clinical-research-trials-you/data-sharing-repositories>
- [58] Kaggle. (n.d.). Datasets. Retrieved July 24, 2023, from <https://www.kaggle.com/datasets>
- [59] Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3), e1001779. <https://doi.org/10.1371/journal.pmed.1001779>
- [60] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Retrieved from <https://www.tensorflow.org/>. Software available from tensorflow.org.

A New Similarity Method for Tourism Recommendation Systems

Eren Türkel ^a , Adil Alpoçak ^{at} 

^a Computer Engineering, Dokuz Eylül University, İzmir, Turkey,
[†] eturkel1312@gmail.com , corresponding author

RECEIVED MAY 30, 2023
ACCEPTED AUGUST 10, 2023

CITATION Türkel, E. & Alpoçak, A. (2023). A New Similarity Method A new similarity method for tourism recommendation systems. *Artificial Intelligence Theory and Applications*, 3(2), 77-91.

Abstract

In this paper, we proposed a new similarity method to use in tourism recommendation systems. Recommendation systems highly depend on the existence of a similarity measure used to identify similar items. In tourism products such as hotels, trips, packages are all hard to judge for their similarity. The proposed method is simply based on user defined weights to calculate similarity. First, we represented each product as a vector and then weighted by user defined scores. Then it uses cosine similarity to measure similarity between items. We evaluated our method using a dataset created by the travel expert. Our experimental results indicate that the proposed method achieves a significant improvement in terms of mean average precision (MAP). We conclude that the proposed method is a promising approach for improving the performance of tourism recommendation systems.

Keywords: similarity method; tourism recommendation system; cosine similarity

1. Introduction

Recommendation systems are very common in our daily lives, from e-commerce platforms to social media apps. These systems aim to provide recommendations to users, based on their past behaviour, preferences, or interests. Similarity methods are one of the key components of recommendation systems and they are used to identify items or users that are similar to each other [1]. Similarity methods enable recommendation systems to leverage the wisdom of the crowd and provide relevant recommendations to users, even if they haven't interacted with a particular item before [2]. In this context, similarity methods are at the heart of recommendation systems, as they allow these systems to make accurate and effective predictions. In this article, we will propose a new similarity method used in a tourism recommendation system.

Tourists widely use tourism recommendation systems to assist them in making informed decisions about their travel destinations, accommodations, and activities [3]. One of the key challenges in these systems is to accurately recommend options that suit travellers' preferences and needs, which simple similarity measures such as user ratings or content-based features may not capture [4]. As noted by Kim and Han [5], these methods also do not consider contextual factors such as travel type, budget, and season, which can significantly influence the travel experience. Thus, a new similarity method that can incorporate a variety of contextual factors is needed to generate specific

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than AITA must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from info@aitajournal.com

Artificial Intelligence Theory and Applications, ISSN: 2757-9778. ISBN: 978-605-69730-2-4 © 2023 İzmir Bakırçay University

recommendations for travellers, and recommendation systems can provide more useful support to travellers with such a method.

Tourism systems are widely used to help travellers make informed decisions about their travel destinations, accommodations, and activities. However, according to Li, Liang and Huang [6], existing recommendation methods struggle to accurately capture the diverse and dynamic preferences and needs of travellers. Therefore, there is a need to develop new and effective similarity methods for tourism systems. Simple similarity criteria such as user ratings, which traditional similarity measures rely on, may not capture the dynamic and diverse preferences of tourists. Furthermore, Wang et al. argued that these methods fail to take into account contextual factors such as travel purpose, type, and season, which can significantly impact the travel experience [7]. In this paper, we propose a new similarity method for tourism systems that takes into account the type, content, and season of the travel. We evaluated the proposed method on a dataset obtained from travel experts, and the results show that the proposed similarity method improves recommendation accuracy.

We assign a value to each feature in our method to determine its relevance in finding similarity, which we refer to as the "weight" of the feature. After applying the weights to the features, we form vectors. We use the cosine similarity measure to find the most similar results among these vectors. Developers can adjust the weights of the features over time, which can improve the accuracy of the method by altering the weights.

Main contribution of this paper is to propose a similarity method to use in tourism recommendation systems. This method provides a more dynamic and diverse approach to capturing the preferences and needs of travellers. It allows developers to adjust the weights of features over time, which can help to refine the accuracy of the recommendation system. Furthermore, this method considers contextual factors such as travel purpose, type, and season, which can have a significant impact on the overall travel experience. Overall, the proposed method offers a promising approach to enhancing the effectiveness of tourism recommendation systems, and can benefit both travellers and the tourism industry.

The organisation of this paper is as follows. In Section I, we provide an introduction to the problem of tourism recommendation systems and the need for new similarity methods. In Section II, we present a literature review of existing methods. Section III describes the proposed similarity method, including the weighting of features and the use of cosine similarity measure. In Section IV, we present the experimental setup and results of applying the proposed method to a dataset of travel experts. Finally, Section V discusses the results and implications of the study, including limitations and future directions for research.

2. Related Works

Chen, Wu, and Buhalis [8] observe that the growth of the tourism industry and the increasing demand for personalised travel recommendations have fueled the popularity of tourism recommendation systems. These systems aim to provide tourists with recommendations on various aspects of their trip, such as accommodations, restaurants, attractions, and activities. To achieve this goal, developers of tourism recommendation systems typically rely on collaborative filtering techniques, which use user feedback to make recommendations. According to Koren, Bell, and Volinsky [9], similarity measurement is an important component of collaborative filtering, as it helps to identify users with similar preferences to the target user.

Zhang, Wang, Chen, and Huang [10] point out that significant research has been conducted in the context of tourism recommendation systems to develop new similarity measures that can enhance the accuracy of recommendations. For example, Li, Wang, Zhang, and Liu [11] and Xiang, Du, Ma and Fan [12] have proposed incorporating contextual information such as travel purpose and season into the similarity measures, while Ma, Liu, Li, Huang and Li [13] have utilised social network analysis to capture the social influence among users. In the context of tourism recommendation systems, Liu, Li, Zhou and Li [14] proposed a hybrid similarity measure that combines item-based and user-based approaches to improve the accuracy of recommendations. The proposed method first identifies the most similar items to the target item, and then finds the most similar users to the target user based on their preferences for these items.

Another line of research has focused on incorporating contextual information into similarity measurement. For example, Ye, Yang, Wang and Law [15] proposed a contextual similarity measure that takes into account the temporal context of user preferences. The proposed method computes the similarity between two users based on the overlap between their preferences within a certain time period, rather than considering all preferences equally.

In addition to traditional similarity measures, there have been efforts to develop new similarity measures based on machine learning techniques. For example, Zhang, Zheng and Lyu [16] proposed a deep learning-based similarity measure that uses a neural network to learn the underlying patterns in user preferences. The proposed method achieved higher accuracy than traditional similarity measures in experiments on a real-world dataset.

Feng, Huang, and Zhang [17] proposed a new similarity method for tourism recommendation systems based on the Dirichlet distribution. This method takes into consideration both the direction and length of rating vectors, and uses a Bayesian approach to compute similarity weights for rating pairs. The proposed method also reduces correlation due to chance and potential system bias. Experimental results on six real-world datasets have shown that the method achieves superior accuracy in comparison with traditional similarity measures.

Overall, similarity measurement is a crucial component of tourism recommendation systems, and researchers have proposed various methods to improve its accuracy. These methods range from traditional similarity measures based on cosine similarity and Pearson correlation coefficient to more advanced methods based on machine learning and contextual information. Ma, Wang and Wang [18] introduced a promising direction for future research in the area of tourism recommendation systems with their proposed similarity method based on the Dirichlet distribution.

In addition to the aforementioned approaches, trust and social network analysis have also been investigated in the context of tourism recommendation systems. These methods aim to incorporate information about the relationships between users, such as their friendships, into the recommendation process. For example, Liu, Liu and Lu [19] proposed a trust-based similarity measure that utilises the trust relationships between users to improve recommendation accuracy. Similarly, Li, Wang, Zhang and Chen [20] used social network analysis to identify influential users and incorporate their opinions into the recommendation process.

According to Wang, Zhang, and Liu [21], active learning is a promising approach in the field of tourism recommendation systems. Active learning involves selecting the most

informative instances for feedback to improve the accuracy of the recommendation model. Wang, Zhang and Liu [22] proposed an active learning framework that uses an uncertainty sampling strategy to select the most uncertain instances for feedback. The authors tested their approach on a dataset of tourist attractions in Beijing and found that their active learning framework significantly improved recommendation accuracy with fewer feedback instances.

According to Jin, Xiang, Du and Ma [23], tourism recommendation systems raise concerns for privacy and security as they involve the collection of personal data from users. To address these concerns, researchers have explored various privacy protection methods such as differential privacy and federated learning. In addition, Zhou, Zhang, Liu and Hu [24] highlight the importance of implementing privacy-preserving recommendation systems. Researchers have also proposed security measures such as encryption and access control to prevent unauthorised access to user data. Therefore, protecting users' personal data through these privacy and security measures is crucial in the development of tourism recommendation systems.

3. Method

This paper presents a new similarity method for tourism recommendation systems. The proposed method aims to increase the accuracy and relevance of travel recommendations. The method assigns values called weights to features and detects similar travels using cosine similarity, making it easier for the user to find the desired travels. In this section, we will provide a detailed description of the methodology used in the study, including data collection and processing procedures, as well as the application of the proposed similarity method.

3.1. Data Collection

We collected our data for this study from a tourism system. The data consists of travel packages, and the features of these travel packages include date, program (text), duration, and categories (beach, wildlife, medicine, ecology, culture, adventure, family, and honeymoon).

We received a reference dataset from a travel expert. Reference dataset contains travel ids and most similar travels to them to evaluate the proposed method. We used this system as a benchmark data.

3.2. Data Preparation

We needed to convert the travel packages into vectors in order to implement our proposed similarity method. Each travel package was initially characterised by various features, including the day, month, and year of travel, as well as the latitude and longitude coordinates of the destination. Additionally, the packages were categorised by program type, such as adventure, wildlife, medical, eco, cultural, cruise, family, honeymoon, historical, or beach. The package duration was also a defining feature. However, in order to better represent the packages numerically, we eliminated the latitude and longitude features, as they were not effective in determining similarity between packages. We transformed the program feature into a TF IDF vector, allowing us to capture the presence or absence of the program in the package. By converting the remaining features into numerical values, we were able to create vectors that represent each travel package in a format that could be processed by our proposed similarity method.

Travel packages vary widely in terms of their focus and features, catering to a diverse range of interests and preferences. Adventure travel packages typically offer physically challenging outdoor activities, such as hiking, climbing, or rafting. Wildlife travel packages focus on observing and interacting with animals in their natural habitats, while medical travel packages may offer opportunities for medical treatments or procedures in foreign destinations. Eco travel packages emphasise sustainable tourism practices and environmental conservation efforts. Cultural travel packages offer experiences that immerse travellers in the local customs, traditions, and history of a destination, while cruise packages provide luxurious voyages to various ports of call. Family travel packages are tailored to meet the needs of families with children, while honeymoon packages offer romantic getaways for newlyweds. Historical travel packages focus on exploring significant historical sites and landmarks, while beach packages are centred around relaxation and leisure activities in coastal destinations.

After discussions with travel experts, we have concluded that latitude and longitude values alone are not sufficient to determine similarities between travel packages. Travel packages with similar latitude or longitude values may have significant differences, while packages with very different latitude and longitude values may be quite similar. Therefore, we have eliminated latitude and longitude values in order to more accurately assess the degree of similarity between travel packages. This way, we will be able to offer more relevant recommendations.

We transformed the day, month, and year features into the time difference feature as the seasonal difference between travel packages is a more effective measure of similarity than the difference in dates. The time difference feature is created at the time of package selection. Figure 1 depicts the seasonal difference between each package and the selected package. We displayed the dates of four packages in a circle in Figure 1. We refer to this circle as the seasonal circle. Seasons are continuously following each other, so we decided to show the dates in a circle. For travel packages, the season of the package is more important than the actual date. Therefore, it would be more appropriate to consider only the season and not the year. A travel package with a date in the summer of 2020 is similar to a travel package that will take place in the summer of 2022 regardless of the year. We displayed the dates of packages A, B, C, and the selected package p on the circle in Figure 1. Each point on the circle represents a time in terms of day and month, without showing the year. The circumference of this circle is equal to the number of days in a year. Travel packages taking place in the same season are more similar to each other in terms of time. As shown in Figure 1, the distance between p and B is much shorter than the other distances. Therefore, the package that is most similar to the selected package in terms of time is package B. We expressed the seasonal difference between two packages in Figure 1 using the absolute value. In the seasonal cycle, there are two different distances between the dates of the two travel packages, long and short. We choose the shorter one to show the historical difference.

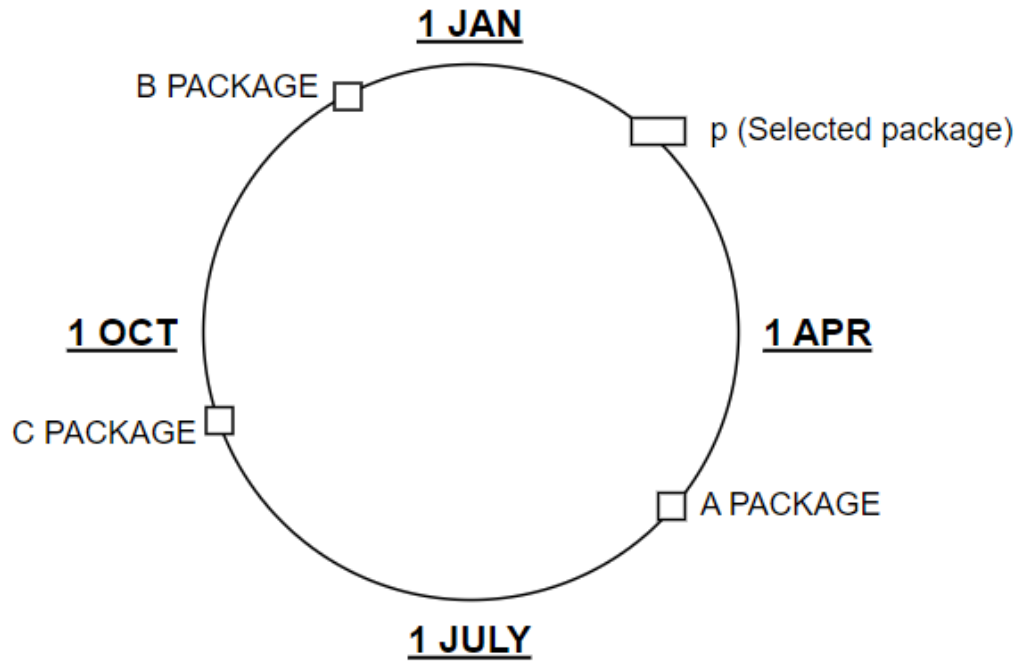


Figure 1. Seasonal circle and package dates

The seasonal difference between travel packages X and Y as $|X - Y|$. $|X - Y|$ represents the number of days between the two packages and can be at most half of the number of days in a year. Determine the day of the year for X and Y to calculate $|X - Y|$. For example, package B's date is the 340th day of the year in Figure 1, the selected package's date is the 30th day of the year, package A's date is the 135th day of the year, and package C's date is the 260th day of the year. The example equation for Figure 1:

$$p = 30, A = 135, B = 340, C = 260$$

$$|p - B| = \begin{cases} -(p - B), & p - B < 0 \\ p - B, & p - B \geq 0 \end{cases} \quad [1]$$

$$|30 - 340| = \begin{cases} -(-310), & -310 < 0 \\ -310, & -310 \geq 0 \end{cases}$$

$$|30 - 340| = 310$$

There are 365 days in a year, and the number of days between two packages on the seasonal circle can be at most 182 days. Therefore, to obtain the number of days between the packages, we subtract from 365 any result that is greater than 182.

$$|p - B| = 365 - 310 = 55$$

$$|p - A| = |30 - 135| = |-95| = 95$$

$$|p - C| = |30 - 260| = |-230| = 230$$

$$|p - C| = 365 - 230 = 135$$

$$|p - B| = 55, \quad |p - A| = 95, \quad |p - C| = 135$$

In this case, the package that is most similar to the selected package in terms of seasonal similarity will be package B, whose date is shown in Figure 1.

We applied the bag-of-words paradigm to digitise the program feature. The program means the text describing what to do in that package, not only what to do, it can also contain explanations about travel. For example, suppose there are three packages with the following programs: "Sunny day today," "Sunny day tomorrow," and "Ship day today." In this case, the vocabulary would be "sunny-day-today-tomorrow-ship." We showed the bag-of-words representation of sample programs in Table 1.

Table 1. Bag-of-words representation of example programs

Program	sunny	day	today	tomorrow	ship
Sunny day today	1	1	1	0	0
Sunny day tomorrow	1	1	0	1	0
ship day today	0	1	1	0	1

After obtaining the Bag-of-Words representation, we calculate the TF IDF value for each travel package, which converts the itinerary of each package into a numerical vector.

IDF (Inverse Document Frequency), is a statistical measure that is commonly used in natural language processing and information retrieval. IDF measures the rarity of a term in a collection of documents. Specifically, it measures how much information a term provides across a collection of documents. The less common a term is across documents, the higher its IDF score will be. IDF is calculated by dividing the total number of documents in a corpus by the number of documents containing the term, and then taking the logarithm of that quotient. The resulting IDF score is used to downweight the importance of terms that occur frequently across documents, and upweight the importance of terms that occur rarely. The goal of IDF weighting is to give more weight to terms that are more informative and less weight to terms that are less informative.

There is an IDF score for each dimension in the corpus. Inverse document frequency is referred to as IDF. To create TF-IDF vectors for packages, we must know the IDF score for each dimension. The IDF score is calculated using the base-2 logarithm as follows:

$$\text{IDF}(\text{word}, \text{corpus}) = \log \left(\frac{\# \text{ of programs in corpus}}{\# \text{ of programs with word in it in corpus}} \right) \quad [2]$$

The TF scores for each dimension were then determined. The acronym TF stands for "term frequency." By dividing the total number of words in the text by the number of times those words appear in the text, total frequency (TF) is calculated.

After we complete the TF-IDF vectorization process of the programs, all travel packages will be transformed into numerical vectors. The new package features will be as shown in Table 2.

Table 2. Vector representation of travel packages (TD: Time difference between the selected travel package, first row is the selected one)

TD	Duration	word1(TF IDF)	word2(TF IDF)	wordN(TF IDF)	adventure	beach	wildlife
0	5	0.35	0.14	0.78	0.88	0.12	0.02
60	8	0	0.01	0.96	0.56	0.64	0.25
72	15	0.17	0.23	0.25	0.13	0	0
21	6	0.21	0.15	0.65	0.74	0.2	0

Then, we normalise the vectors using the min-max normalisation technique in order to provide a consistent comparison. We apply the weights we describe below to transform the resulting vectors into new vectors. In similarity calculation, we consider the cosine similarity between these vectors and the vector of the selected travel package.

3.3. Technical Definition of Method

In order to identify similarities among travel packages, we converted the packages into numerical vectors. We assigned values to each feature of the packages, which we referred to as weights. We expressed the multiplication of weights and features using the \pm symbol. We denoted the features with the letter "a" and represented the weight options with the letter "w". The \pm symbol indicates the repetition of that feature. The resulting package vectors can be symbolised using the formula below:

$$Px = (a1x \pm w1, a2x \pm w2, \dots, aNx \pm wN) \quad [3]$$

We discussed which characteristics could be effective in finding similarities with the domain experts and optimised the values in the weight matrix accordingly. Weights can take on different values, and we need to determine the weights before using the method. The selected weights are expressed as w_s :

$$ws = (w1, w2, \dots, wn) \quad [4]$$

Each package attributes has its own set of attributes and a typical package attribute vector is defined as follows:

$$A = (a1, a2, \dots, an) \quad [5]$$

The package vector is formed by applying the weights to the features. Let us represent the package vector with P , and the attributes with A .

So the package vector will be:

$$P = A \pm ws \quad [6]$$

We explain it in a trivial example. A is the attribute vector of the package vector. Attribute vector A contains: Time difference, duration, TF IDF vector of the package program and categories (adventure, beach, ..., wildlife). The vector P is formed as a result of applying the weights, i.e., w_s , to the attribute vector. P is referred to as the package vector. P_s is the selected package. P_i , P_j and P_k are other packages in Table 2 and w_s is the selected weight vector.

$$ws = (1, 1, 3, 2, 0, 0, 1, 0)$$

$$Ps = As \pm ws$$

$$As = (0, 5, 0.35, 0.14, 0.78, 0.88, 0.12, 0.02)$$

$$Ps = (0, 5, 0.35, 0.14, 0.78, 0.88, 0.12, 0.02) \pm (1, 1, 3, 2, 0, 0, 1, 0)$$

$$Ps = (0 \pm 1, 5 \pm 1, 0.35 \pm 3, 0.14 \pm 2, 0.78 \pm 0, 0.88 \pm 0, 0.12 \pm 1, 0.02 \pm 0)$$

$$Ps = (0, 5, 0.35, 0.35, 0.35, 0.14, 0.14, 0.12)$$

Then, package vectors with weights for packages in Table 2 are:

$$Pi = Ai \pm ws$$

$$Ai = (60, 8, 0, 0.01, 0.96, 0.56, 0.64, 0.25)$$

$$Pi = (60, 8, 0, 0.01, 0.96, 0.56, 0.64, 0.25) \pm (1, 1, 3, 2, 0, 0, 1, 0)$$

$$Pi = (60 \pm 1, 8 \pm 1, 0 \pm 3, 0.01 \pm 2, 0.96 \pm 0, 0.56 \pm 0, 0.64 \pm 1, 0.25 \pm 0)$$

$$Pi = (60, 8, 0, 0, 0, 0.01, 0.01, 0.64)$$

$$Pj = Aj \pm ws$$

$$Aj = (72, 15, 0.17, 0.23, 0.25, 0.13, 0.0, 0.0)$$

$$Pj = (72, 15, 0.17, 0.23, 0.25, 0.13, 0.0, 0.0) \pm (1, 1, 3, 2, 0, 0, 1, 0)$$

$$Pj = (72 \pm 1, 15 \pm 1, 0.17 \pm 3, 0.23 \pm 2, 0.25 \pm 0, 0.13 \pm 0, 0.0 \pm 1, 0.0 \pm 0)$$

$$Pj = (72, 15, 0.17, 0.17, 0.17, 0.23, 0.23, 0)$$

$$Pk = Ak \pm ws$$

$$Ak = (21, 6, 0.21, 0.15, 0.65, 0.74, 0.2, 0.0)$$

$$Pk = (21, 6, 0.21, 0.15, 0.65, 0.74, 0.2, 0.0) \pm (1, 1, 3, 2, 0, 0, 1, 0)$$

$$Pk = (21 \pm 1, 6 \pm 1, 0.21 \pm 3, 0.15 \pm 2, 0.65 \pm 0, 0.74 \pm 0, 0.2 \pm 1, 0.0 \pm 0)$$

$$Pk = (21, 6, 0.21, 0.21, 0.21, 0.15, 0.15, 0.2)$$

So package vectors are,

$$Ps = (0, 5, 0.35, 0.35, 0.35, 0.14, 0.14, 0.12)$$

$$Pi = (60, 8, 0, 0, 0, 0.01, 0.01, 0.64)$$

$$Pj = (72, 15, 0.17, 0.17, 0.17, 0.23, 0.23, 0)$$

$$Pk = (21, 6, 0.21, 0.21, 0.21, 0.15, 0.15, 0.2)$$

Min-max normalisation, also known as feature scaling, is a technique used to scale and transform numerical data into a standardised range. The purpose of using min-max normalisation is to transform variables so that they are comparable and have equal weights in analysis. The process involves subtracting the minimum value of the variable and dividing the result by the range of the variable (maximum value - minimum value). This results in a new variable with values ranging from 0 to 1, where 0 represents the minimum value and 1 represents the maximum value.

Min-max normalisation is commonly used in machine learning and data analysis, especially when working with models that are sensitive to the scale of the input variables. By scaling the input variables, the model can better identify patterns and relationships in the data, which can lead to more accurate predictions and better performance. Additionally, normalisation can help reduce the impact of outliers and improve the convergence of iterative algorithms. Overall, min-max normalisation is a simple and effective way to standardise data and improve the accuracy of analytical models.

We apply min-max normalisation to obtain comparable results, normalised vectors are:

$$P_s = (0, 0, 1, 1, 1, 0.57, 0.57, 0.19)$$

$$P_i = (0.83, 0.3, 0, 0, 0, 0, 0, 1)$$

$$P_j = (1, 1, 0.49, 0.49, 0.49, 1, 1, 0)$$

$$P_k = (0.3, 0.1, 0.6, 0.6, 0.6, 0.6, 0.6, 0.31)$$

Then, we round it up to 1 floating point for better view, rounded vectors are:

$$P_s = (0, 0, 1, 1, 1, 0.6, 0.6, 0.2)$$

$$P_i = (0.8, 0.3, 0, 0, 0, 0, 0, 1)$$

$$P_j = (1, 1, 0.5, 0.5, 0.5, 1, 1, 0)$$

$$P_k = (0.3, 0.1, 0.6, 0.6, 0.6, 0.6, 0.6, 0.3)$$

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space. It is defined as the cosine of the angle between two vectors and ranges from -1 to 1, where 1 indicates that the vectors are identical and 0 indicates that the vectors are orthogonal. In other words, cosine similarity measures how closely two vectors are aligned with each other. This measure is commonly used in natural language processing and information retrieval applications to compare the similarity of documents or words based on their word frequency vectors. Cosine similarity is a popular metric because it is efficient to compute and does not depend on the length of the vectors, making it useful for comparing documents of different lengths. We use the following formula to calculate the cosine similarity between two vectors:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad [7]$$

Here, A and B represent two different vectors, dot product represents the dot product of the two vectors, and norm represents the length of the vector.

We explain it with an example of calculating the cosine similarity between P_s and P_i :

$$P_s.P_i = (0 * 0.8) + (0 * 0.3) + (1 * 0) + (1 * 0) + (1 * 0) + (0.6 * 0) + (0.6 * 0) + (0.2 * 1) = 0.2$$

$$||P_s|| = \sqrt{0^2 + 0^2 + 1^2 + 1^2 + 1^2 + 0.6^2 + 0.6^2 + 0.2^2} = 1.609$$

$$||P_i|| = \sqrt{0.8^2 + 0.3^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 1^2} = 1.080$$

$$similarity = \cos(\theta) = \frac{P_s.P_i}{||P_s|| ||P_i||} = \frac{0.2}{1.609 * 1.080} = 0.1115$$

Similarly, we can calculate the cosine similarity between all the vectors, similarity between P_s and others:

$$similarity = \cos(\theta) = \frac{P_s.P_j}{||P_s|| ||P_j||} = 0.7497$$

$$similarity = \cos(\theta) = \frac{P_s.P_k}{||P_s|| ||P_k||} = 0.6965$$

The result obtained is a value between 0 and 1. A value of 1 indicates that the two vectors are exactly the same, while a value of 0 indicates no similarity between the two vectors.

4. Experimentation and Results

We have developed a recommendation system to test our method, and Figure 2 depicts the application structure. We used Angular on the front-end and NodeJS on the back-end. Data selection and data mining processes were performed on the NodeJS side, and we established the connection between NodeJS and Python using a node module called python-shell. Our method was implemented using Python.

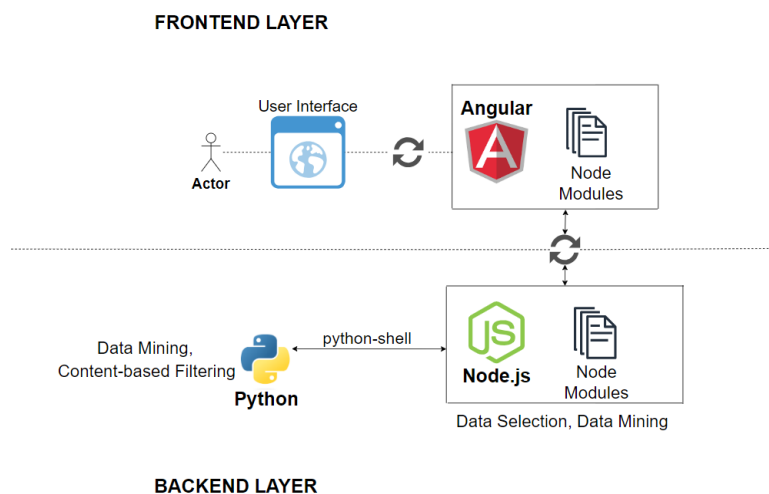


Figure 2. Application architecture

We used a reference dataset to test the accuracy of the recommendations. The reference dataset shows which packages are most similar to the 100 packages we selected. We obtained the reference dataset from a travel expert at a travel agency. We checked if a recommendation was correct by looking at the reference dataset.

The reference dataset contains the 5 most similar packages to each of the hundred packages. There is no ranking among these similar packages. There is no scoring in the dataset, only the packages and the ids of the 5 most similar packages. The travel agency determined the most similar packages to the 100 packages with its own employees. Travel agency employees can know most accurately whether a package is similar to another.

Mean average precision is a useful metric for assessing the performance of a system. In this evaluation, we calculated mean average precision using a macro approach. We obtain the average precision by recommending each travel package. The average of the scores for average precision is known as mean average precision.

However, since the mean average precision is the average of the average precision scores, we can also find the MAP value using just one search. Average precision can only be applied to one query. The recommendations' ranking in the mean sensitivity score has a considerable impact. Each piece of advice will be given an accuracy rating. The presentation of the recommendations will have a significant impact on this rating. The precision scores of the right recommendations are simply averaged to determine average precision. A sensitivity rating is assigned to every recommendation.

First off, if a recommendation is not specified as a similar package in the reference dataset, we consider the recommendation as false. It won't be factored into the calculation, but its position affects the calculation. A recommendation's sensitivity is determined by how many correct recommendations came before it in relation to all other recommendations. We can compute the mean average precision of the system after the precision calculation.

In each query, we applied a different weighting. Time difference and year are affected by date weight (\check{O}). Program is affected by program weight (P). Duration is affected by duration weight (q). Categories are affected by category weight (\mathcal{AE}). Each query implies the process of finding the most similar packages for a given package. The package on the screen is denoted by the letters T_p , and the packages that are most like T_p are T_{p7} , T_{p4} , T_{p1} , T_{p5} , and T_{p3} . The reference list has the ones that are the most comparable. Reference lists, which are helpful for testing, include the packages that are most like a given package. We presented results of 6 queries in Figure 3.

The user selects a travel package and proceeds, with the most accurate recommended travel packages for the selected package being T_{p7} , T_{p4} , T_{p1} , T_{p5} , and T_{p3} , respectively. We refer to each package selection made by the user as a query. To test the method, the user made selections with different weights. The weights can be determined by the person writing the code, as there is no option to set weights on the screen. In Figure 3, date weight is symbolised by \check{O} , program weight by P , duration weight by q , and category weight by \mathcal{AE} . We indicate the weight values with the letter "w" for each query. Based on the queries, we obtained package recommendations and calculated the accuracy, precision, and average precision values for these recommendations.

Table 3. Query results with different weights

Weights	Query 1	Query 2	Query 3	Query 4	Query 5	Query 6
\bar{O}	1	1	1	1	1	1
P	5	2	16	16	16	1
q	1	1	1	1	1	1
\bar{E}	12	16	4	2	16	1
Result	Tp7, Tp4, Tp1, Tp5, Tp8	Tp7, Tp4, Tp1, Tp9, Tp3	Tp7, Tp4, Tp8, Tp2, Tp3	Tp9, Tp2, Tp1, Tp5, Tp8	Tp7, Tp4, Tp1, Tp5, Tp9	Tp7, Tp4, Tp1, Tp2, Tp8
Actual Result	Tp7, Tp4, Tp1, Tp5, Tp3					
Accuracy	4 / 5 = 0.8	4 / 5 = 0.8	3 / 5 = 0.6	2 / 5 = 0.4	4 / 5 = 0.8	3 / 5 = 0.6
Precision	4 / 5 = 0.8	4 / 5 = 0.8	3 / 5 = 0.6	2 / 5 = 0.4	4 / 5 = 0.8	3 / 5 = 0.6
Average Precision	(1+1+1+1) / 4 = 1	(1+1+0.75+0.8) / 4 = 0.8875	(1 + 1 + 0.6) / 3 = 0.86	(0.33 + 0.5) / 2 = 0.415	(1+1+1+1) / 4 = 1	(1+1+1) / 3 = 1

In this experiment, we arrived at a mean average precision score of 0.86. So, our method can be applied in tourism recommendation systems.

5. Conclusion

In this study, we introduced a new similarity approach for tourism recommendation systems that utilises multiple features of travel packages to make recommendations. Our method adopts a content-based technique by implementing a weighting scheme to adjust the importance of different features.

We conducted experiments on a reference dataset to evaluate the performance of our new method. The results show that the accuracy of our method can vary depending on the weights used. To explore the scenario of using different weights, we calculated the mean average precision (MAP) by running queries with different weights. The results demonstrate the effectiveness of our method in finding similarities.

Our study demonstrates that the similarity method we proposed can be used in tourism recommendation systems. First, it emphasises the importance of incorporating various features in the recommendation process to enhance user experience and satisfaction. Second, it suggests that weighting schemes can be used to fine-tune the relevance of different features and optimise recommendation results.

Our study adds to the existing body of research on tourism recommendation systems and presents a potential solution to the difficulties of providing precise and diverse travel recommendations. There is room for further investigation into the extension of our approach to incorporate additional features and datasets, as well as examining its effectiveness in various contexts and scenarios.

In addition to exploring the extension of our methodology, future research could also explore the use of machine learning techniques for selecting and optimising weights from one or more reference datasets. This could enhance the accuracy and effectiveness of the recommendations generated by our approach. Ultimately, our study contributes to the ongoing efforts to improve tourism recommendation systems and advance the field of travel technology.

References

- [1] Kesici, E., & Öztürk, O. (2020). A hybrid movie recommendation system based on item similarity and user preferences. *Journal of Intelligent & Fuzzy Systems*, 39(2), 1603-1612. <https://doi.org/10.3233/JIFS-179469>
- [2] Burke, R. (2018). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 28(4-5), 331-388. <https://doi.org/10.1007/s11257-018-9195-x>
- [3] Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65. doi: 10.1016/j.tourman.2016.10.015.
- [4] Chen, Y., Wu, B., Zhang, X., & Zhang, Z. (2018). A personalized travel recommendation system using long-tail correlations. *Journal of Travel Research*, 57(1), 31-44. doi: 10.1177/0047287517696563.
- [5] Kim, J. H., & Han, J. Y. (2019). Personalization in travel recommendation systems: An exploratory study on the role of contextual factors. *Journal of Travel Research*, 58(1), 85-99. doi: 10.1177/0047287517733559.
- [6] Li, X., Liang, Y., Huang, Z., & Huang, Y. (2019). Deep learning for travel recommendation: An end-to-end neural recommendation approach on multi-source data. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2392-2405. doi: 10.1109/TKDE.2019.2929371.
- [7] Wang, J., Zhang, J., Zhang, X., Liu, Q., & Zhang, X. (2020). A novel travel recommendation algorithm based on deep learning with context awareness. *IEEE Access*, 8, 1-14. doi: 10.1109/ACCESS.2020.2981155.
- [8] Chen, C., Wu, Y., & Buhalis, D. (2021). A critical review of smart tourism destination literature and future research directions. *Journal of Hospitality and Tourism Technology*, 12(2), 202-221. doi: 10.1108/JHTT-08-2019-0131.
- [9] Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30-37. doi: 10.1109/MC.2009.263.
- [10] Zhang, L., Wang, D., Chen, X., & Huang, J. (2019). Travel recommendation: A survey of recent advances and future directions. *Journal of Hospitality and Tourism Research*, 43(3), 421-463. doi: 10.1177/1096348019857687.
- [11] Li, X., Wang, J., Zhang, Y., & Liu, Y. (2019). An intelligent recommendation algorithm based on improved collaborative filtering. *IEEE Access*, 7, 160293-160304. doi: 10.1109/ACCESS.2019.2951277.
- [12] Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65. doi: 10.1016/j.tourman.2016.10.015.
- [13] Ma, S., Liu, X., Li, Y., Huang, X., & Li, G. (2018). Social influence-based collaborative filtering recommendation algorithm. *IEEE Access*, 6, 38885-38893. doi: 10.1109/ACCESS.2018.2859886.
- [14] Liu, X., Li, Y., Zhou, T., & Li, J. (2016). A hybrid similarity measure for collaborative filtering recommender systems in the tourism domain. *Tourism Management*, 54, 87-101. doi: 10.1016/j.tourman.2015.11.012.
- [15] Ye, H., Yang, L., Wang, X., & Law, R. (2011). Contextual collaborative filtering based on Bayes probability model for personalized service recommendation in ubiquitous commerce. *Expert Systems with Applications*, 38(5), 5850-5859. doi: 10.1016/j.eswa.2010.11.129
- [16] Zhang, X., Zheng, Y., & Lyu, M. R. (2018). Deep learning-based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1), 1-38. doi: 10.1145/3137597
- [17] Feng, S., Huang, Y., & Zhang, Y. (2019). A tourism recommendation algorithm based on Dirichlet distribution and item attributes. *IEEE Access*, 7, 41434-41444. doi: 10.1109/ACCESS.2019.2908521
- [18] Ma, X., Lv, P., Wang, P., & Wang, Y. (2021). An improved Dirichlet-based similarity measure for tourism recommendation. *Journal of Ambient Intelligence and Humanized Computing*, 12(9), 9275-9287. doi: 10.1007/s12652-021-03398-1.
- [19] Liu, X., Liu, J., & Lu, X. (2010). A trust-based recommendation algorithm for tourism. *Expert Systems with Applications*, 37(12), 8460-8468.
- [20] Li, X., Wang, D., Zhang, J., & Chen, H. (2014). An intelligent recommendation system for tourism planning based on social network analysis and ontology. *Journal of Computer and System Sciences*, 80(7), 1364-1376.
- [21] Wang, L., Zhang, Z., & Liu, Y. (2012). Active learning for content-based recommendation: An exploratory study. *Journal of Computer Information Systems*, 53(1), 14-23.

- [22] Wang, L., Zhang, Z., & Liu, Y. (2014). Active learning framework for content-based recommendation. In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (pp. 151-156). IEEE.
- [23] Jin, X., Xiang, Z., Du, Q., & Ma, Y. (2019). Privacy protection in tourism big data: a research agenda. *Journal of Travel Research*, 58(6), 963-977. <https://doi.org/10.1177/0047287518809143>
- [24] Zhou, Y., Zhang, Y., Liu, H., & Hu, J. (2019). Privacy preserving recommendation systems: a survey. *Journal of Big Data*, 6(1), 1-29. <https://doi.org/10.1186/s40537-019-0194-2>

Prediction of Acute Burn-Induced Coagulopathy Risk with Machine Learning Models

Murat Ali Çınar ^{a†} 

^a Hasan Kalyoncu University, Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation

[†] muratali.cinar@hku.edu.tr, corresponding author

RECEIVED AUGUST 8, 2023
ACCEPTED SEPTEMBER 7, 2023

CITATION Çınar, M. A. (2023). Prediction of Acute Burn-Induced Coagulopathy Risk with Machine Learning Models. *Artificial Intelligence Theory and Applications*, 3(2), 92-104.

Abstract

In burn patients, lipolysis, proteolysis, glycolysis and many severe hyperdynamic and hypermetabolic responses are seen with high fever. Patients who have these hypermetabolic reactions lose lean muscle mass, their immune systems deteriorate, their coagulopathy worsens, and they eventually die. Due to enhanced vascular permeability, fluid collection in the interstitial space is seen, especially in the first 24 hours following serious burns. If not treated, decreased intravascular volume has an impact on tissue perfusion. Acute burns also raise the risk of coagulopathy. Coagulopathy is one of the leading contributors to death in burn patients. Therefore, the use of machine learning-based decision support systems for quick coagulopathy pre-diagnosis may be crucial for physicians and healthcare executives. In this study, machine learning models were investigated for the estimation of the risk of coagulopathy due to acute burns, using a data set of 1040 burn patients and 35 different biochemical parameters of these patients. The Subspace KNN model showed the highest prediction success compared to other machine learning methods with 100% accuracy.

Keywords: burns; coagulopathy; machine learning

1. Introduction

Burn injuries are traumas that result from sources including fire, hot liquid, radiation, chemical, cold, or electricity and destroy skin and/or organic tissues as a result of energy transfer [1]. After falls, traffic accidents, and wars, burns are the fourth most frequent type of trauma in the globe. According to a 2004 survey, around 11 million people apply to medical facilities annually owing to burn injuries worldwide. The American Burn Association's research from 2016 highlighted the fact that 67% of burns only affect less than 10% of the total body surface area [TVSA] [2]. According to a 2019 study, more than 13,000 people experience severe burns that cover more than 20% of their body surface area every year [3]. Deeply catabolic and hypermetabolic reactions are triggered by burn injuries, and they can endure for weeks, months, or even years. In particular, lipolysis, proteolysis, glycolysis, high temperature, and several severe hyperdynamic and hypermetabolic responses are reported in burn patients who are impacted by more than 20% of TBSA. These hypermetabolic responses seen in patients cause a decrease in lean muscle mass, decrease in functional capacity, delay in wound healing, coagulopathy, and deterioration of fibrinolytic activity, leading to serious mortality [4,5]. After burning, muscle protein breaks down much faster than it is synthesized. Net protein loss leads to loss of lean body mass and severe muscle loss. Protein degradation can

continue for up to 9 months after severe burn injury. Protein catabolism is associated with increases in metabolic rates. For severely burned patients, the resting metabolic rate at thermal neutral temperature [30°C] exceeds 140% of normal and increases to 130% after wounds are fully healed, then to 120% at 6 months and to 110% at 12 months. falls. These changes in metabolic rate vary according to the percentage of TBSA affected by the burn [6].

Burn-Induced Coagulopathy

Coagulopathy is frequently seen in patients with TBSA burns of 20% or more. The development of coagulopathy after burn injury further complicates treatment approaches in these patients. The inflammatory response that occurs after a burn has profound effects, affecting clot formation and leading to burn-induced coagulopathy. Coagulopathy seen in burn patients is similar to coagulopathy seen in patients with sepsis and/or post-traumatic and is characterized by disruption of natural anticoagulant systems, and procoagulant and antifibrinolytic changes [7-8].

Pathophysiology of Burn-Induced Coagulopathy

In patients with severe burns, coagulopathy may cause thromboembolic complications, multi-organ failure, and increased mortality and morbidity. Although there is no accepted definitive pathophysiology, it is thought that the hypermetabolism, inflammatory response, hypothermia and endothelial damage seen after burns cause coagulopathy and anti-fibrinolytic activities (Figure 1) [8].

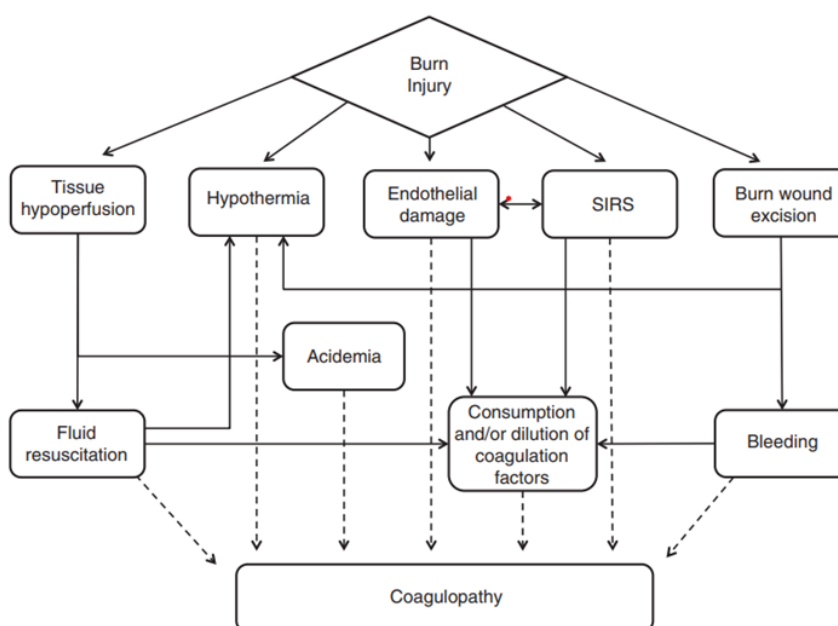


Figure 1. Pathophysiologic mechanisms impacting coagulation in patients with severe burns (The syndrome known as the systemic inflammatory response [8])

The changes that cause coagulopathy in burn patients are quite similar. These changes are:

- Increase in thrombin-antithrombin complex [TAT] levels and FVII [FVIIa] activation,
- Decrease in protein c, protein s levels,

- Seeing markers of thrombin activation such as fibrin degradation products and prothrombin fragment F1+2 [8].

Processes of Burn-Induced Coagulopathy

Very Early Stage After Burn – Before Hospitalization

It is unknown when coagulopathy first manifests in patients with severe burns, however it is generally accepted that coagulation indicators and systemic alterations appear soon after burn damage. Although there aren't any studies on coagulopathy in the very early stages following burns, it is believed that these changes start to occur right away and even before hospitalization [7-8].

First Days After Burn

On the first day following a serious burn, significant activation of both coagulation and fibrinolysis is observed. During this time, these patients are regarded as hyperfibrinolytic, and tissue-type plasminogen activator [t-PA] levels also rise [8].

Chronic Period After Burn

After a burn injury, coagulopathy is characterized by changes to the natural anticoagulation system that are procoagulant and antifibrinolytic in nature. Within 7 days following a burn injury, this hypercoagulation often returns to more stable levels, unless a condition that causes coagulopathy also manifests. At various stages in patients with severe burns, a number of potential pathophysiological factors, alone or in combination, may cause and/or exacerbate coagulopathy. Aggressive fluid resuscitation techniques used to prevent shock [burn-induced shock] that also trigger hypothermia in later days may exacerbate or induce coagulopathy. In the early stage, burn injury directly triggers coagulopathy through tissue hypoperfusion, hypothermia, endothelial damage, or a systemic inflammatory response. Coagulation disorder may also be brought on by burn-related complications such sepsis, bleeding from prolonged inactivity during the hospital stay, repeated wound excision, and other surgical procedures [8].

Treatment of Burn-Induced Coagulopathy

There are no specific suggestions or tested guidelines for the management of postburn coagulopathy. Most people are aware of how to avoid situations that will worsen or cause coagulopathy. These measures are can be listed as [8];

- Preventing tissue hypoperfusion
- Preventing hypothermia during fluid resuscitation
- Use of anticoagulants

The use of low-molecular heparin or antithrombin at a protective dose against increased coagulopathy in burn patients is part of the standard treatment, but has a low inhibitory effect on microvascular thrombus formation. The use of drugs such as streptokinase or urokinase may be more effective in preventing coagulopathy. However, these drugs are not suitable for use in traumas where there are many open wounds such as burns and where there is a risk of infection. In the literature, randomized controlled studies for the treatment of coagulopathy in patients with severe burns were described as insufficient [7-8].

Parameters Used in Monitoring Coagulation and Fibrinolytic System

Clinical coagulation tests are functional assays that evaluate the rate of clot formation from the moment the coagulation cascade is activated. Prothrombin time [PT], activated partial thromboplastin time [APTT], fibrinogen, thrombocyte, d-dimer tests are used.

These tests are commonly used to identify disorders in the intrinsic and extrinsic pathways of coagulation [9].

Prothrombin Time [PT]

This test, which gives information about the functions of the extrinsic pathways of coagulation, has been used since 1935. Following the addition of thromboplastin and calcium to the plasma, its normal values are stated as 10-14 seconds. It is a parameter generally used in the monitoring of anticoagulant treatments. It is generally used in the evaluation of acute coagulopathy in burn patients. [9-10]. It was determined that coagulopathy was observed in patients with a value above 14.6 sec [9-10].

Activated Partial Thromboplastin Time [APTT]

The APTT is a test used to evaluate the intrinsic and common pathways of coagulation. It is especially used in the evaluation of acute coagulopathy in burn patients. Reference values vary between 22.0-35.0 sec. A value >45.0 sec in burn patients indicates the risk of acute coagulopathy [11]. Coagulopathy may be an early indicator of mortality in patient groups with a very high mortality rate, such as burns. Therefore, the aim of this study is to predict the risk of coagulopathy associated with acute burns with machine learning models. Developing artificial intelligence-based interfaces that can be used by clinicians working in the field of burns and creating a sample data set needed for academicians who want to work in the field of artificial intelligence in burns are other purposes of the study.

2. Related Works

The incidence of coagulopathy associated with acute burns has been defined as 37% in the literature [12,13]. Although its incidence is low, a coagulopathy that starts within 24 hours increases the risk of mortality for burn patients and causes disruption of treatment processes. Although there are studies in the literature such as predicting mortality with machine learning models in burn patients, studies evaluating the risk of coagulopathy due to acute burns have not been found. Similar studies are seen in other patient groups other than burns.

Fengping et al. In 2020, they looked examined patients who had spontaneous intracerebral hemorrhages for coagulation abnormalities. Aspartate transaminase, alanine transaminase, hemoglobin, platelet count, white blood cell count, neutrophil percentage, systolic and diastolic pressure, albumin/globulin ratio, neutrophil count, lymphocyte percentage, aspartate transaminase, and aspartate transaminase were employed in this investigation. These figures indicate that the accuracy rate is 931%. They have developed models that have the predictive ability [14].

Another study conducted in 2021 developed machine learning models to dynamically predict the risk of sepsis-induced coagulopathy. In this study, in which open-access data was used, it was stated that two models were developed that could better predict the risk of coagulopathy in septic patients than Logistic Regression and SIC scores [15].

In 2020, Hasegawa et al. In another study, machine learning models that can predict sepsis-induced coagulopathy were compared. A total of 17 parameters in 1017 patients were included in the study. Machine learning methods such as random forests [RF], support vector machines [SVM] and neural networks [NN] are used. It was reported that the prediction accuracy of multiple linear regression, RF, SVM, and NN models were 63.7%, 67.0%, 64.4%, and 59.8%, respectively [16].

Kaiyuan et al. In a study in which they evaluated acute coagulopathy due to trauma in 2020, they used the data of 818 patients. They stated that traditional logistic regression models may be better than the predictive power of machine learning models [17].

When all these studies were examined, it was observed that machine learning models were tried to predict acute coagulopathy in cases such as trauma such as sepsis, but no such study was performed in burn patients. In addition, the data used in the studies in the literature are not as large as the data set in our study. We believe that we can make important contributions to the literature in this study, which can predict coagulopathy due to acute burns and uses more algorithms.

3. Materials Methods

3.1. Study Population

A total of 1040 patients hospitalized in the intensive care and/or service units between January 2016 and June 2023 in Turkey 25 December State Hospital Burn Center were included in the study. The data of the patients were obtained retrospectively from the web Karmed database of 25 December State Hospital. Demographic information of all patients and blood test results at the time they were first admitted to the hospital were recorded. Permission was obtained from Hasan Kalyoncu University Health Sciences Ethics Committee to conduct the study [ethical approval numbered 2023/65]. Informed consent form was signed, and consent was obtained in order to use the data of all patients included in the study. The Declaration of Helsinki was followed when conducting the study.

3.2. Data Collection

A data set was created with hemogram and biochemistry parameters of all patients: age, gender, burn type, burn percentage, presence of inhalation burn, duration of hospitalization in intensive care unit, length of hospital stay, total hospitalization time (Table-1).

Table 1. Biochemical parameters and abbreviations

Name	Abbreviation	Name	Abbreviation
White Blood Cell	WBC	Creatine Kinase	CK
Hemoglobin	HGB	Sodium	-
Hemorectitis	HCT	Aspartate Aminotransferase	AST
Platelet	PLT	Phosphorus	-
Neutrophil	NEU%	C-Reactive Protein	CRP
Lymphocyte	LYM%	Alanine Aminotransferase	ALT
Monocyte	MONO%	Very Low Density Lipoprotein	VLDL
Urea	-	Potassium	-
Blood Urea Nitrogen	BUN	T-Protein	-
Creatinine	-	Glucose	-
Triglyceride	-	Albumin	-
Albumin	-	Egfr	-
Calcium	-	Active Partial Thromboplastin Time	APTT
Magnesium	-	Alkaline Phosphatase	-
Bilirubin	-		

Among the patients, those with PT tests over 14.6 seconds were considered as coagulopathy due to acute burn [12]. Patients with values between 10-14.5 were recorded as normal. While coagulation due to acute burn was observed in 334 of the

patients, this did not occur in 706. The data set was randomly divided into two separate clusters, 70% to be used in ANN training and 30% to test model success. The data set consisting of 40 input variables and 1 outcome variable for each patient is given in Table 2

Table 2. Dataset of features collected from patients.

Features and Outcome
Sex
Age
Nationality
Type of Burn
Percentage of Burn
Inhalation Burn
Duration of Intensive Care Unit
Duration of Ward Hospitalization
Length of Hospitalization [total]
WBC-1
HGB-1
HCT-1
PLT-1
NEU%-1
LYM%-1
MONO%-1
Ure-1
BUN-1
Kreatinin-1
Trigliserit-1
Albumin-1
Kalsiyum-1
Magnezyum-1
Bilurubin-1
ALP-1
CK-1
Sodyum-1
AST-1
Fosfor-1
CRP-1
ALT-1
VLDL-1
Potasyum-1
T-protein-1
Glukoz-1
Albumin[g/L]-1
eGFR-1
Procalsitonin-1
Discharge (ex or recovery)
APTT-1
Outcome- Acute Burn-Induced Coagulopathy

3.3. Machine Learning Algorithms

By creating algorithms that effectively describe a dataset, machine learning focuses on the learning component of artificial intelligence. Or, to put it another way, machine learning algorithms are algorithms that employ a range of statistical, probabilistic, and optimization techniques to draw lessons from the past and identify insightful patterns in huge, complicated datasets [18]. These days, research is done on the issue and these algorithms are commonly utilized in the early diagnosis of disease or in the calculation of risk factors. [18,19].

3.3.1. Artificial neural networks

One of the most popular machine learning techniques, artificial neural networks (ANNs), are made up of a network of simple information processing units called neurons. High-level relationships can be recognized and nonlinear relationships between variables can be learned using ANNs. Instead of using intricate mathematical models to represent complicated relationships, artificial neural networks (ANNs) make use of an interactive construction of many basic neurons. Algorithms for supervised, unsupervised, and reinforcement learning can all be completely implemented using ANNs. [20].

3.3.2. Decision Tree

Decision trees are a supervised classification method that produces a model in the form of a tree data structure consisting of decision-making nodes that direct the classes according to the characteristics of the data used and the classes of the classification to be made, and leaf nodes marked as class information. The algorithm that creates the decision tree model is divided into small pieces of the data set to be analyzed and the tree model is developed. In the decision tree model, a black decision node may contain a single or more than one sub-branch [21-27].

3.3.3. K-Nearest Neighbor [KNN]

KNN Nearest Neighbor Algorithm is a supervised and non-parametric classification method that classifies data based on the proximity of training samples in the data set. The algorithm maintains its popularity among machine learning methods for years with its advantages such as not requiring training, being easy to implement, adaptable to local information and being resistant to noisy training data. In order to classify a new data, the algorithm operates by looking at the closeness of k of the training data known to belong to which class. The k value is defined intuitively at the beginning of the algorithm. For the proximity calculation, distance equations such as Manhattan, Hamming, Euclid and Minkowski are used [21-27].

3.3.4. Naive Bayes

Naive Bayes Classifier, named after British mathematician Thomas Bayes, is a simple probabilistic classification method with high performance in diagnosis based on Bayesian decision theory. In the Naive Bayes classifier, it is accepted that all the features are equally important and that all the features are independent of each other. For the Bayesian classifier using a tutorial learning method, taught data with known classes is presented to the system. Probability calculations are made on the taught data. According to the probability values obtained, it is tried to calculate how probable the test data is included in which class [21-27].

3.3.5. Support Vector Machine

Vladimir Vapnik created the Support Vector Machines [SVM] approach in 1992. It is based on statistical learning theory. Problems involving classification, regression analysis, and nonlinear function approaches are resolved using SVM, a supervised learning method. It offers strong classification and high generalization performance in text, voice, text, object, and image recognition difficulties when handling bioinformatics challenges [21-27]. For a small set of learning patterns, SVM seeks to achieve good generalization. It uses a helpful learning algorithm to identify patterns in challenging-to-analyze complex data sets. The approach uses classification learning to separate

samples for classification estimation from previously unobserved data. The objective of SVM is to construct an n-dimensional hyperplane that optimally separates the data into various [21-27].

3.4. Success Evaluation Criteria

The current dataset is split into two groups in order to assess the system's effectiveness in each group. One is used for training, and the other serves as a test set to simulate potential examples that the system is completely unaware of. With the chosen training algorithm, the system gains knowledge from the training set. The trained system's performance is then measured against the test set. [28-29].

3.4.1. Confusion Matrix

Machine learning and especially those with statistical classification problems A complexity matrix is a tabular layout that visualizes the performance of an algorithm. If the predicted variable (dependent variable, target, target, output, output) is in binary format, the accuracy is evaluated with the evaluation complexity matrix [28-29].

Using this matrix, sensitivity, specificity, precision, negative predicted value, Many performance measures such as 24 npv), accuracy (accuracy) and f1-score (f1 score) can be calculated. In the evaluation of the success of the prediction model, it is generally accuracy formula is used [28-29].

3.4.2. ROC (Receiver Operating Characteristic) Curve

It allows the determination of appropriate cut-off points to determine the optimum sensitivity and optimum specificity of a medical test. A reference is needed to determine the appropriate cut-off point via the ROC curve. The ROC curve method uses values such as Sensitivity, Specificity, Accuracy Rates. Quantitative data of the variable measured in clinically or pathologically ill or healthy individuals are used to determine the appropriate cut-off point. In order to create the ROC curve, the variable values measured from patients and healthy individuals and used in diagnosis are ranked in order of magnitude. Each value from the sorted array is taken as the cutoff point in sequence. If the height of the measured quantitative variable indicates the disease, if the value above this value is below the patient, the categories are determined as robust. is created [28-29]. The number of patients and healthy individuals in these categories is 2x2. displayed in the tables. According to the value determined as the cut-off point in these tables, rates are determined such as how many of the real patients are qualified as sick, how many of them are determined as healthy even though they are sick, and they are considered sick according to the cut-off point even though they are healthy. Sensitivity ratios y obtained from the 2x2 tables created by taking each value as the cutoff point to obtain the ROC curve. axis, a coordinate axis is created to show 1-Specific ratios on the x-axis. The ratios calculated for each cut point are marked on the coordinate. A curve is obtained by connecting the points. This is a concave curve and is called the ROC curve. The area under the curve (AUC) is an important indicator in evaluating the accuracy of the test. How much is the ROC AUC The higher the discriminative power of the medical test, the higher the discriminative power will be [28-29].

3.4.3. Cross Validation

A common technique to gauge a classification algorithm's performance or compare the performance of two classification algorithms in a dataset is cross-validation. A dataset is randomly divided into k parts of roughly equal size, and each piece is used to test a classification technique and the model made up of the other pieces.

By averaging the k accuracies from cross validation, the classification algorithm's performance is assessed [28-29].

4. Experimental Study

In this study, 25 different machine learning methods (have been used. Here, each algorithm has utilized different activation functions, optimization algorithms, and loss functions, the details of which are provided below (with or without PCA) (Table-3). All these algorithms are developed using the machine learning toolbox of the MATLAB programming language. On a Windows 10 computer with an Intel processor, MATLAB R2021b was used to produce all results.

Table 3. Machine learning techniques for comparison (accuracy %)

Machine Learning Models	TV		3-Fold CV		5-Fold CV		10-Fold CV	
	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable	PCA Disable	PCA Enable
Fine Tree	99.9%	80.5%	99.9%	61.3%	99.9%	62.9%	99.8%	63.8%
Medium Tree	99.9%	70.9%	99.9%	66.2%	99.9%	65.5%	99.8%	66.1%
Coarse Tree	99.9%	68.7%	99.9%	66.9%	99.9%	67.7%	99.8%	67.1%
Linear Discriminant	87.2%	68.0%	85.4%	67.9%	84.4%	67.9%	85.6%	67.9%
Quadratic Discriminant	79.1%	67.7%	64.3	66.77%	69.4%	67.6%	failed	67.4%
Logistic Regression	93.4%	68.0%	91.71%	67.9%	90.0%	67.9%	90.5%	67.9%
Gaussian Naive Bayes	72.3%	67.4%	89.2%	67.23%	69.6%	67.4%	69.8%	67.2%
Kernel Naive Bayes	91.75%	70.1%	62.6%	67.5%	85.19%	67.5%	82.9%	66.7%
Linear SVM	91.00%	68.0%	87.7%	67.6%	83.1%	64.3%	89.3%	66.2%
Quadratic SVM	93.2%	47.2%	87.1%	53.6%	88.7%	53.7%	88.2%	52.5%
Cubic SVM	93.4%	41.8%	85.3%	91.71%	87.2%	48.3%	85.9%	49.1%
Fine Gaussian SVM	93.4%	68.00%	69.7%	50.1%	70.1%	67.9%	70.3%	67.8%
Medium Gaussian SVM	91.1%	68.00%	83.7%	67.9%	85.4%	67.9%	85.3%	67.9%
Coarse Gaussian SVM	77.4%	68.00%	72.9%	67.9%	74.4%	67.9%	75.4%	67.9%
Fine KNN	93.4%	93.4%	73.7%	64.2%	73.0%	62.5%	73.7%	63.6%
Medium KNN	78.8%	70.2%	74.1%	65.1%	75.3%	64.5%	75.2%	66.6%
Coarse KNN	72.6%	67.9%	69.9%	67.9%	71.2%	67.9%	70.8%	67.9%
Cosine KNN	79.1%	67.0%	74.6%	56.6%	74.6%	59.0%	74.7%	62.1%
Cubic KNN	77.5%	70.2%	72.6%	65.4%	72.0%	64.7%	73.7%	66.6%
Weighted KNN	93.4%	93.4%	76.3%	64.5%	76.5%	64.1%	77.8%	65.5%
Boosted Trees	67.9%	71.3%	67.9%	66.8%	67.9%	68.3%	67.9%	67.9%
Bagged Trees	99.7%	93.1%	96.6%	65.0%	99.7%	64.6%	99.7%	66.3%
Subspace Discriminant	86.7%	68.0%	85.7%	67.9%	84.9%	67.9%	85.6%	67.9%
Subspace KNN	100%	93.4%	70.6%	62.2%	71.6%	64.3%	72.3%	65.5%
RUS Boosted Trees	90.9%	74.4%	89.3%	60.6%	80.8%	61.6%	77.4%	59.2%

Subspace KNN [1009%, Fine Tree, Medium Tree, Coarse Tree [99%]] showed the best performance in estimating the risk of coagulopathy due to acute burns. In addition, the PCA method was used to test different parameter variations and the best results were found in Table-3. A confusion matrix represents the most successful and unsuccessful algorithms implemented. Machine learning uses a confusion matrix to interpret the performance of the classification model used. Figure 2 and 3 show a confusion matrix comparing the predicted and actual values of the target feature.

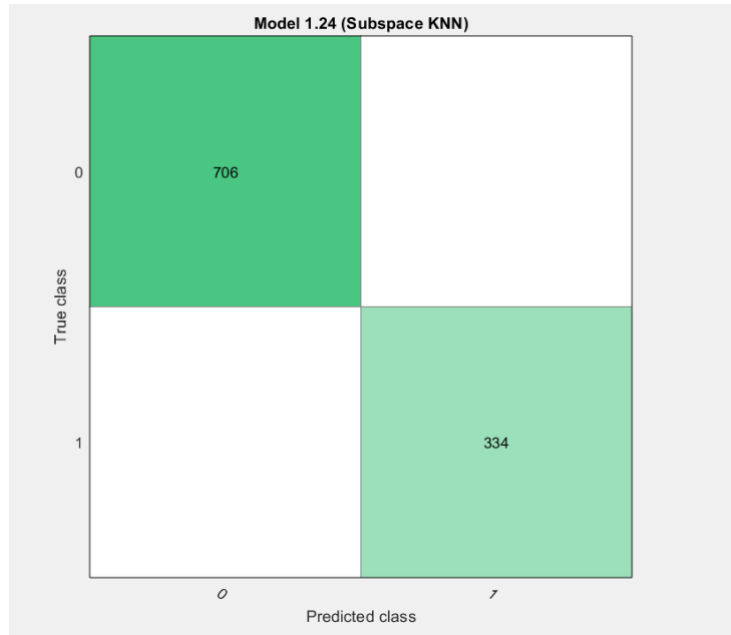


Figure 2. Confusion matrix for the most successful subspace KNN algorithm

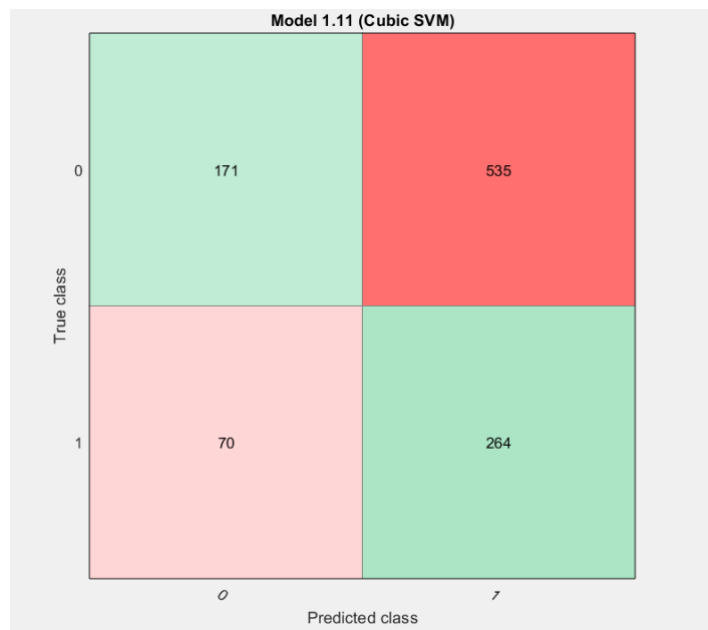


Figure 3. Confusion matrix for the worst successful cubic SVM algorithm

An illustration of a binary classifier system's classification performance as the discrimination threshold varies is the receiver operating characteristic (ROC) curve. ROC curve obtained after evaluating the effectiveness of the algorithm. Figures 4 and 5 illustrate, respectively, the ROC curves for the best-performing Subspace KNN method and the worst-performing Cubic SVM technique.

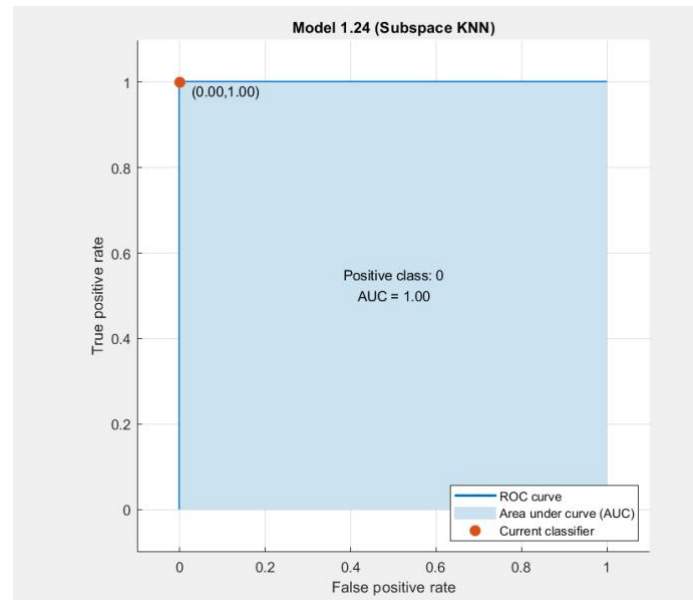


Figure 4. ROC curve for the best performing subspace KNN algorithm.

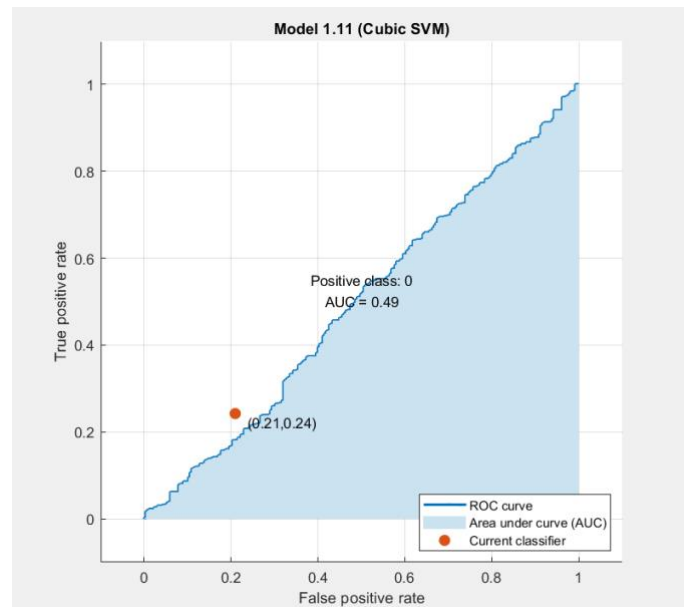


Figure 5. ROC curve for the worst-performing cubic SVM algorithm.

5. Conclusion

In this study, we developed an innovative clinical decision support system that can be used in various processes for clinicians. The Subspace KNN model proposed in this study showed the highest prediction success with 100% accuracy compared to other machine learning methods. We believe that this model, which was created by using 35 different biochemical markers, which predicts the risk of coagulopathy, which can be seen in the first 24 hours in burn patients and is difficult to diagnose, will contribute to the literature.

In this study, the application of machine learning-based decision support systems for rapid pre-diagnosis of coagulopathy in burn patients may be important for clinicians and healthcare administrators. By using these systems, healthcare providers can improve the efficiency and accuracy of the initial assessment and treatment process, especially in non-burn clinics. This can lead to better allocation of resources, including medical personnel, equipment and treatments. In addition, such systems can contribute to the standardization of care in different healthcare facilities, ensuring that patients receive consistent, high-quality care regardless of their location. Ultimately, the integration of these decision support systems into burn care management can help improve patient outcomes, reduce the burden on healthcare professionals, and guide policy makers in developing well-informed strategies to address the challenges associated with burn injuries. We also state that multicenter studies can contribute more to the literature.

References

- [1] Jeschke, M. G., van Baar, M. E., Choudhry, M. A., Chung, K. K., Gibran, N. S., Logsetty, S. (2020). Burn injury. *Nature Reviews Disease Primers*, 6(1), 1-25.
- [2] Greenhalgh, D. G. (2019). Management of burns. *New England Journal of Medicine*, 380(24), 2349-2359.
- [3] Romero, S. A., Moralez, G., Jaffery, M. F., Huang, M., Cramer, M. N., Romain, N., Crandall, C. G. (2019). Progressive exercise training improves maximal aerobic capacity in individuals with well-healed burn injuries. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 317(4), R563-R570.
- [4] Mandell, S. P., Gibran, N. S. (2014). Early enteral nutrition for burn injury. *Advances in wound care*, 3(1), 64-70.
- [5] Martini, W. Z., Holcomb, J. B., Yu, Y. M., Wolf, S. E., Cancio, L. C., Pusateri, A. E., Dubick, M. A. (2020). Hypercoagulation and Hypermetabolism of Fibrinogen in Severely Burned Adults. *Journal of Burn Care & Research*, 41(1), 23-29.
- [6] Williams, F. N., Jeschke, M. G., Chinkes, D. L., Suman, O. E., Branski, L. K., Herndon, D. N. (2009). Modulation of the hypermetabolic response to trauma: temperature, nutrition, and drugs. *Journal of the American College of Surgeons*, 208(4), 489-502.
- [7] Ball, R. L., Keyloun, J. W., Brummel-Ziedins, K., Orfeo, T., Palmieri, T. L., Johnson, L. S., Shupp, J. W. (2020). Burn-induced coagulopathies: a comprehensive review. *Shock* Augusta, Ga., 54(2), 154.
- [8] Glas, G. J., Levi, M., Schultz, M. J. (2016). Coagulopathy and its management in patients with severe burns. *Journal of Thrombosis and Haemostasis*, 14(5), 865-874.
- [9] Winter, W. E., Flax, S. D., Harris, N. S. (2017). Coagulation testing in the core laboratory. *Laboratory Medicine*, 48(4), 295-313.
- [10] Geng, K., Liu, Y., Yang, Y., Ding, X., Tian, X., Liu, H., Yan, H. (2020). Incidence and Prognostic Value of Acute Coagulopathy After Extensive Severe Burns. *Journal of Burn Care & Research*, 41(3), 544-549.
- [11] Lippi, G., Favaloro, E. J. (2008, October). Activated partial thromboplastin time: new tricks for an old dogma. In *Seminars in thrombosis and hemostasis* 34(07), (604-611).
- [12] Sherren, P. B., Hussey, J., Martin, R., Kundishora, T., Parker, M., & Emerson, B. (2013). Acute burn induced coagulopathy. *Burns*, 39(6), 1157-1161.
- [13] Mitra, B., Wasiak, J., Cameron, P. A., O'Reilly, G., Dobson, H., & Cleland, H. (2013). Early coagulopathy of major burns. *Injury*, 44(1), 40-43.

- [14] Zhu, F., Pan, Z., Tang, Y., Fu, P., Cheng, S., Hou, W., ... & Sun, Y. (2021). Machine learning models predict coagulopathy in spontaneous intracerebral hemorrhage patients in ER. *CNS Neuroscience & Therapeutics*, 27(1), 92-100.
- [15] Zhao, Q. Y., Liu, L. P., Luo, J. C., Luo, Y. W., Wang, H., Zhang, YJ, Luo, Z. (2021). A machine-learning approach for dynamic prediction of sepsis-induced coagulopathy in critically ill patients with sepsis. *Frontiers in Medicine*, 7, 637434.
- [16] Hasegawa, D., Yamakawa, K., Nishida, K., Okada, N., Murao, S., Nishida, O. (2020). Comparative analysis of three machine-learning techniques and conventional techniques for predicting sepsis-induced coagulopathy progression. *Journal of clinical medicine*, 9(7), 2113.
- [17] Li, K., Wu, H., Pan, F., Chen, L., Feng, C., Liu, Y., Li, T. (2020). A machine learning-based model to predict acute traumatic coagulopathy in trauma patients upon emergency hospitalization. *Clinical and Applied Thrombosis/Hemostasis*, 26, 1076029619897827.
- [18] Uddin, S., Khan, A., Hossain, M. E., Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1), 1-16.
- [19] Choi, RY., Coyner, A. S., Kalpathy-Cramer, J., Chiang, M. F., Campbell, J. P. (2020). Introduction to machine learning, neural networks, and deep learning. *Translational vision science & technology*, 9(2), 14-14.
- [20] AKSOY, PK., Erdemir, F., KILINÇ, D., & Orhan, E. R. (2022). A Decision Support System on Artificial Intelligence Based Early Diagnosis of Sepsis. *Artificial Intelligence Theory and Applications*, 2(1), 14-26.
- [21] Safavian, S. R., and Landgrebe, D. (1991). A survey of decision tree classifier methodology. *Institute of Electrical and Electronics Engineers Transactions on Systems, Man and Cybernetics*, 21(3), 660-674.
- [22] Loh, W. Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14-23.
- [23] Kavzoğlu, T. ve Çölkesen, İ. (2010). Karar ağaçları ile uydu görüntülerinin sınıflandırılması: Kocaeli örneği. *Harita Teknolojileri Elektronik Dergisi*, 2(1), 36-45.
- [24] Raj, R., Nehemiah, H. K., Elizabeth, D. S. and Kannan, A. (2018). A novel featuresignificance based k-nearest neighbour classification approach for computer aided diagnosis of lung disorders. *Current Medical Imaging*, 14(2), 289-300.
- [25] Fan, G. F., Guo, Y. H., Zheng, J. M. and Hong, W. C. (2019). Application of the weighted k-nearest neighbor algorithm for short-term load forecasting. *Energies*, 12(5), 916. 46
- [26] Jaber, M. M., Abd, S. K., Shakeel, P. M., Burhanuddin, M. A., Mohammed, M. A. And Yussof, S. (2020). A telemedicine tool framework for lung sounds classification using ensemble classifier algorithms. *Measurement*, 162, 107883.
- [27] Başer, F. ve Apaydın, A. (2015). Sınıflandırma amaçlı destek vektör makinelerinin lojistik regresyon ile karşılaştırılması. *Anadolu University of Sciences & Technology-Theoretical Sciences*, 3(2), 53-65.
- [28] Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote sensing of Environment*, 62(1), 77-89.
- [29] Çelik, Ö., Altunaydın, S. S. (2018). A Research on Machine Learning Methods and Its Applications. *Online Learning*, 1(3).

Performance Analysis of Efficient Deep Learning Models for Multi-Label Classification of Fundus Image

Muhammed Pektaş ^{at} 

^a Izmir Bakırçay University, Izmir, Turkey

[†] mhmdpkts@gmail.com, corresponding author

RECEIVED AUGUST 8, 2023
ACCEPTED SEPTEMBER 7, 2023

CITATION Pektaş, M. (2023). Performance analysis of efficient deep learning models for multi-label classification of fundus image. *Artificial Intelligence Theory and Applications*, 3(2), 105-112.

Abstract

Convolutional Neural Networks (CNNs) have demonstrated significant advancements in the domain of fundus images owing to their exceptional capability to learn meaningful features. By appropriately processing and analyzing fundus images, computer-aided diagnosis systems can furnish healthcare practitioners with valuable reference information for clinical diagnosis or screening purposes. Nevertheless, prior investigations have predominantly concentrated on detecting individual fundus diseases, while the simultaneous diagnosis of multiple fundus diseases continues to pose substantial challenges. Furthermore, the majority of previous studies have prioritized diagnostic accuracy as their main focus. Efficient Deep Learning constitutes a crucial concept that enables the utilization of deep learning models on edge devices, thereby reducing the computational carbon footprint. Facilitating the cost-effective diagnosis of eye diseases from fundus images on edge devices holds significance for researchers aiming to deploy these vital healthcare models into practical use. This study focuses on assessing the performance of well-known efficient deep learning models in addressing the multi-label classification problem of fundus images. The models underwent training and testing using the dataset provided by international competition on ocular disease intelligent recognition in 2019. The experimental findings demonstrate that the efficientnetb3 model outperforms the other models, exhibiting the highest level of performance. And also, when applying standard data augmentation techniques to the current dataset, we observe decreasing in f1-score and accuracy. Code is available at <https://github.com/m-pektas/Performance-Analysis-of-Efficient-Deep-Learning-Models-for-MultiLabel-Classification-of-Fundus-Image>

Keywords: artificial intelligence; deep learning; efficient deep learning; multi-label classification; eye disease diagnosis; fundus images

1. Introduction

Diseases of the fundus can lead to loss of vision, which is the leading cause of blindness [1]. There are some common diseases that can be diagnosed from fundus images such as cataracts, diabetes, etc. For example, cataract disease is one of the important diseases. An estimated 95 million people worldwide are affected by cataract, according to the World Health Organization (WHO) [2]. In middle and low-income countries, cataract remains the leading cause of blindness. The development of late-stage fundus diseases often has a serious impact on the visual function of patients, and there is no specific treatment for such diseases. Diabetes patients are now one of the largest

disease groups in the world. There are no obvious abnormal symptoms in the early stages of the disease. However, it eventually leads to blindness. It is one of the four most common blinding diseases. [3]. If the disease is caught in the early stages, it is still treatable. If diagnosed at a late stage, the visual prognosis is very poor and the cost of treating is high, even if surgery is successful [4]. The patient's vision will rapidly deteriorate within a short period of time, resulting in irreversible visual impairment, if prompt and effective treatment is not provided [5].

Early detection and treatment of fundus diseases are important. Artificial Intelligent technology has the potential to assist primary care ophthalmologists in diagnosing ocular diseases by leveraging comprehensive medical data, thereby offering new strategies to improve the ability to diagnose and treat in primary care clinics. Integrating artificial intelligence into ophthalmic practice promises to meet the practical needs of many patients suffering from fundus-related diseases. Therefore, developing and deploying efficient and accurate machine learning models plays a crucial role in enabling early diagnosis, improving the availability of healthcare services, and making them more cost-effective.

In previous works, researchers designed and implemented deep learning models to diagnose eye diseases in the early stage of the treatments. In this paper, we analyze the efficient CNN models' performance on the ocular disease intelligent recognition (ODIR-5K) dataset due to care about cost and efficiency in health applications to increase accessibility of health technologies.

In this paper, we analyze previous works about eye disease diagnosis from fundus image in Section 2. In Section 3, we give information about dataset and methods that we used. In Section 4, we share our experiment results.

2. Literature Review

Tham et al. work on diagnosing cataract disease. They compare AI algorithm performance and clinician performance on this problem. Using more than 25,000 images from population-based studies, they report the development and validation of a retinal photograph-based deep learning algorithm for automated detection of visually significant cataract [6]. Wang et al, work on the ODIR-5K dataset and focus on diagnosing multi-disease classification from fundus image problems about the eye. They propose a method that uses the EfficientNets [7] model with ensemble learning [8]. In Zhou et al's work, they focus on diabetic retinopathy. As they claim that, most researchers can't obtain satisfactory performance on diabetic retinopathy due to less training data that is without consistent annotation. Therefore, they demonstrate a dataset that has consisted of pixel and image level annotations [9]. He et al, propose a new method that uses attention and different fusion techniques. They achieve the best performance using Resnet-101 with their proposed fusion technique [10]. Bhati et al, propose a method called Discriminative Kernel Convolution Network (DKCNet) that explores discriminative region-wise features without adding extra computational cost [11]. Wang et al, combine convolutional neural networks and self-attention called MBSaNet for fundus disease identification. Their experimental results show that MBSaNet achieves state-of-the-art performance with fewer parameters than current methods [12].

3. Material and Methods

3.1. Dataset

The ODIR-5K dataset [13] was obtained from the "International Competition on Ocular Disease Intelligent Recognition" sponsored by the Peking University. Dataset consists of "real" patient data from Shangong Medical Technology Co. Ltd. from different hospitals/medical centers in China. The training set is a structured ophthalmology database containing 3,500 patients, their ages, left and right eye color fundus images, and doctors' diagnosis codes. Color fundus photographs of 500 patients without age and sex were used as the test set. The fundus images are taken with different image resolutions by various cameras on the market, such as Canon, Zeiss, and Kowa. Thanks to this dataset, the patients can be classified into eight labels. Class names (normal, diabetic retinopathy, glaucoma, cataract, AMD, hypertensive retinopathy, myopia, and other diseases) and distributions are shown in Figure 1.

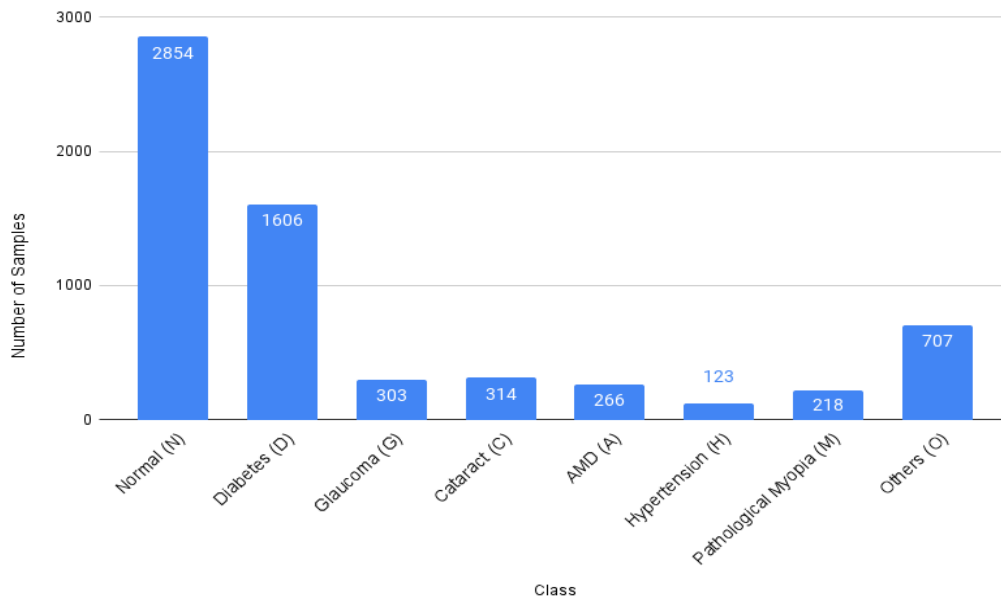


Figure 1. Class names and distribution of ODIR-5K dataset.

3.2. Data Pre-processing

In dataset pre-processing step, we used the already pre-processed version of the dataset [14]. The images in this version of the dataset are already resized to 512x512 and cropped by providing 1:1 aspect ratio. The dataset divided into a training set and a test set in a 7:3 ratio, and then resize the images to 224 x 224. This version of the dataset is our baseline.

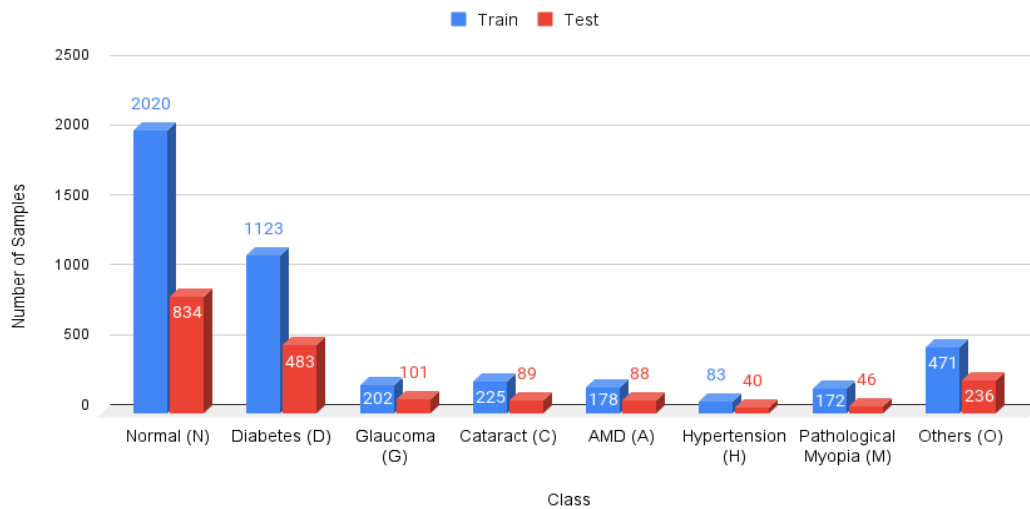


Figure 2. Train and test set distribution of our baseline version of the dataset.

3.3. Methods

Mobilenet, EfficientNet, and SqueezeNet versions are well-known architectures in the efficient deep learning community. We use these models and their versions in our experiments. We described these models in the following paragraphs.

Mobilenets: MobileNets are developed for resource-constrained use cases. Therefore, it is designed with the focus on making it a small, low-latency, low-power solution. Similar to other popular large-scale models, such as Inception, they can be used for classification, detection, embedding, and segmentation [15].

EfficientNets: The baseline network is highly dependent on the effectiveness of model scaling. To further improve performance, they also designed a new baseline network through neural architecture search using the AutoML MNAS framework, which optimizes both accuracy and efficiency (FLOPS). The Efficientnets uses mobile inverted bottleneck convolution as a main module, similar to MobileNetV2 and MnasNet, but is slightly larger due to an increased FLOP budget. As a result, the basic network is then scaled up to obtain a family of networks known as EfficientNets [7, 16].

SqueezeNets: SqueezeNet is an algorithm designed for small size and high accuracy. This compact and efficient CNN model was proposed in 2016 by researchers from DeepScale, UC Berkeley, and Stanford University. SqueezeNet balances high accuracy with low complexity, making it ideal for resource-constrained edge devices such as mobile phones. SqueezeNet uses a special type of convolution layer called the fire module, a combination of 1x1 and 3x3 filters, to reduce parameters while maintaining high accuracy, making it efficient for resource-limited devices. [17].

We train these models and their versions on our baseline dataset and we select the best architecture by accuracy and f1-score metrics. After that, we do experiments with different data augmentation and different hyperparameters to obtain the best results. At the end of these experiments, we compare the selected models' accuracy and f1-score performances. We demonstrate that the best performer option is EfficientnetB3 without data augmentation.

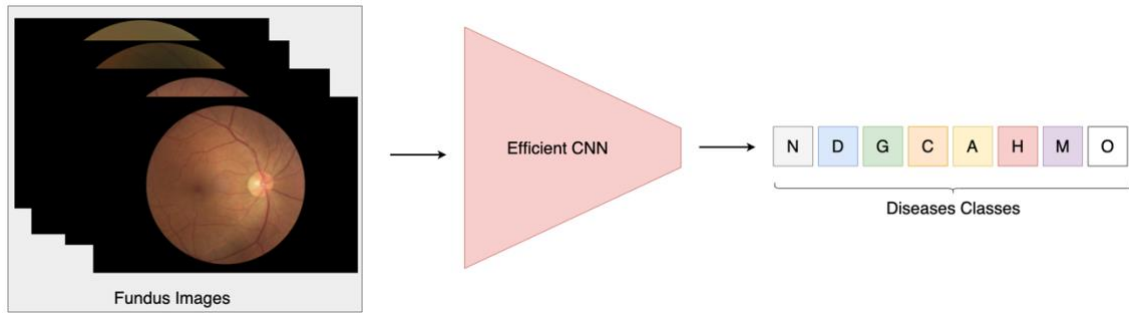


Figure 3. A simple pipeline for multi-label classification of eye diseases with an efficient CNN architecture.

4. Experiments and Results

4.1. Configurations

During these experiments, we use a computer that has Intel® Core™ i7-7700HQ CPU @ 2.80GHz × 8, NVIDIA GeForce GTX 1050 Ti, and 8 GB ram.

4.2. Implementation Details

In these trainings, we use Adam optimizer with 0.001 learning rate. Also, we use ReduceLRonPlateau learning rate scheduler to adjust the learning rate during the training process. The batch size and early stopping patient set 15 and 20 respectively. All hyperparameters that we used are shown in Table 1.

Table 1. Hyperparameter configurations.

Configuration	Value
Optimization Function	Adam
Maximum Epoch	100
Batch Size	15
Learning Rate	0.001
Learning Rate Scheduler	ReduceLRonPlateau
Early Stopping Patience	20

4.3. Selection of Best Architecture

Firstly, we use our baseline dataset that split with 7:3 train/test ratio during our selection of best efficient architecture experiments. Training set has 4474, testing set 1917 samples with this splitting ratio. In these experiments, we train 10 different efficient deep learning models with the same dataset and hyperparameters. We achieve worst accuracy performances with SqueezeNet versions. Our EfficientNet and MobileNet experiments have competitive accuracy with each other, but best performer model is EfficientNetB3 with 90.12% accuracy.

Based on these results, we select EfficientNetB3 to our further experiments. Experiment results are shown in Table 2.

Table 2. Selection of best architecture experiment results on test dataset.

Method	Train/Test Split	Best Test Loss	Accuracy
EfficientNetB0	0.7	0.4629	88.12
EfficientNetB1	0.7	0.4825	87.96
EfficientNetB2	0.7	0.4604	87.89
EfficientNetB3	0.7	0.4297	90.14
EfficientNet_V2_S	0.7	0.4754	88.26
MobilenetV2	0.7	0.5835	85.03
MobilenetV3_S	0.7	0.5798	83.55
MobilenetV3_L	0.7	0.4627	87.34
Squeezenet_10	0.7	1.258	51.00
Squeezenet_11	0.7	1.2854	49.59

As shown in Table 2, We obtain the best results with EfficientNetB3 and the worst result with Squeezenet_11. Most of these models have higher than 85% accuracy on test dataset. The top 3 results belong to EfficientNetB3, EfficientNet_V2_S, and EfficientNetB0. As our experiments, the EfficientNet model's versions have better performance than other methods such as MobilenetV2, MobilenetV3_L, and Squeezenet_10.

4.4. Optimizing the First Results

We select the best performer model as EfficientNetB3 on eye disease diagnosis problem based on the Table 2. We try different train/test split ratio to achieve best performance. Our experiment results that we try different dataset split ratio are shown in Table 3.

Method	Train/Test Split	Best Test Loss	Accuracy	F1-Score
<i>EfficientNetB3</i>	<i>0.7</i>	<i>0.4297</i>	<i>90.14</i>	<i>0.6619</i>
<i>EfficientNetB3</i>	<i>0.8</i>	<i>0.2656</i>	<i>0.9387</i>	<i>0.6596</i>
EfficientNetB3	0.9	0.1606	0.9664	0.6870

Table 3: Train/Test split ratio experiment results.

As shown in Table 3, when we used 0.9 as the dataset split ratio, we obtain the best accuracy and f1-score. These results give us some insights about the model that needs more data to learn diagnosing eye diseases.

Table 4. Data augmentation experiments' results.

Method	Train/Test Split	Data Augmentation	Best Test Loss	Accuracy	F1-Score
<i>EfficientNetB3</i>	0.9	No	0.1606	0.9664	0.6870
<i>EfficientNetB3</i>	0.9	Imagenet	0.3753	0.8857	0.5430
<i>EfficientNetB3</i>	0.9	Cifar10	0.3288	0.8868	0.6056
<i>EfficientNetB3</i>	0.9	Simple	0.2092	0.9441	0.6338

We select 0.9 as the best train/test split ratio based on Table 3. We used this ratio in our data augmentation experiments. In our experiments, data augmentation have 4 different values as No, Imagenet, Cifar10, and Simple. Imagenet and Cifar10 are the default data augmentation processes for Cifar10 and Imagenet datasets [18]. Simple is defined by us. It contains only different color jittering with brightness 0.5 and hue 0.3 values. As shown in Table 4, applying data augmentation caused the decreasing model performance. Also, by increasing the data augmentation effect, the model performances are getting decrease. According to Table 3 and 4, we can say that, the model need to more data and training.

Table 5. Comparison of CNN performances between Wang et al [8] and our methods.

Method	Accuracy	Parameter Count (M)
EfficientNetB3 (Ours)	0.96	12
VGG 16	0.86	134.7
VGG 19	0.86	143.7
InceptionV3	0.87	6.23
EfficientB3	0.90	12
Resnet50	0.86	23.9

As shown in Table 5, we compare ours and Wang et al [8] results. We achieve better accuracy with 96% on EfficientNetB3 then others. In addition, EfficientNet has lower parameter count then most of other methods while has higher accuracy. These results support to our less costly and more accessible eye disease diagnosis health applications goal.

5. Conclusion

Late-stage fundus diseases often cause severe visual impairment and there is no specific treatment to treat such diseases. In the early stages of most diseases, there are no obvious abnormalities, but eventually, it will cause blindness. It is still treatable if caught in the early stages. When diagnosed in late stages, even with successful surgery, the visual prognosis is very poor and the cost of treatment is high. Therefore, it is very important to detect and treat fundus diseases at an early stage. Combining artificial intelligence and ophthalmic medical treatment is expected to meet the practical needs of many patients suffering from fundus diseases. Thus, efficient and accurate machine learning models have an important role in terms of early diagnosis, increasing the accessibility of health services and making

them cheaper. In this work, we analyze the efficient deep CNN model's performance on ODIR-5K dataset for the identification of eye diseases from fundus images. According to our experiment results, EfficientNetB3 best performer among the well-known current efficient CNN models. Also, we make empirical studies on EfficientNetB3 using different hyperparameters and data augmentation techniques. In our experiments, our best results 96.94% accuracy and 0.6870 f1-Score. According to us, these findings can make contribution to accurate, efficient, and more accessible health applications. In future works, we use an automatic hyperparameter tuning process to obtain possible better results and generate synthetic data to make the dataset balanced and diverse.

References

- [1] J. B. Jonas, R. R. A. Bourne, R. A. White, S. R. Flaxman, J. Keeffe, J. Leasher, K. Naidoo, K. Pesudovs, H. Price, T. Y. Wong, S. Resnikoff, and H. R. Taylor, "Visual impairment and blindness due to macular diseases globally: A systematic review and meta-analysis," *Amer. J. Ophthalmol.*, vol. 158, no. 4, pp. 808–815, Oct. 2014.
- [2] Liu, Y. C., Wilkins, M., Kim, T., Malyugin, B., & Mehta, J. S. (2017). Cataracts. *The Lancet*, 390(10094), 600-612.
- [3] J. L. Leasher, R. R. A. Bourne, S. R. Flaxman, J. B. Jonas, J. Keeffe, K. Naidoo, K. Pesudovs, H. Price, R. A. White, T. Y. Wong, S. Resnikoff, and H. R. Taylor, "Global estimates on the number of people blind or visually impaired by diabetic retinopathy: A meta-analysis from 1990 to 2010," *Diabetes Care*, vol. 39, pp. 1643–1649, Sep. 2016, doi: 10.2337/dc15-2171
- [4] O. B. Walton, R. B. Garoon, C. Y. Weng, J. Gross, A. K. Young, K. A. Camero, H. Jin, P. E. Carvounis, R. E. Coffee, and Y. I. Chu, "Evaluation of automated teleretinal screening program for diabetic retinopathy," *JAMA Ophthalmol.*, vol. 134, no. 2, pp. 204–209, Feb. 2016, doi: 10.1001/jamaophthalmol.2015.5083.
- [5] H. Ye, Q. Zhang, X. Liu, X. Cai, W. Yu, S. Yu, T. Wang, W. Lu, X. Li, H. Jin, Y. Hu, X. Kang, and P. Zhao, "Prevalence of age-related macular degeneration in an elderly urban Chinese population in China: The Jiangning eye study," *Investigative Ophthalmol. Vis. Sci.*, vol. 55, no. 10, pp. 6374–6380, Sep. 2014, doi: 10.1167/iovs.14-14899.
- [6] Tham, Y. C., Goh, J. H. L., Anees, A., Lei, X., Rim, T. H., Chee, M. L., ... & Cheng, C. Y. (2022). Detecting visually significant cataract using retinal photograph-based deep learning. *Nature Aging*, 2(3), 264-271.
- [7] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PMLR.
- [8] Wang, J., Yang, L., Huo, Z., He, W., & Luo, J. (2020). Multi-label classification of fundus images with efficientnet. *IEEE Access*, 8, 212499-212508.
- [9] Zhou, Y., Wang, B., Huang, L., Cui, S., & Shao, L. (2020). A benchmark for studying diabetic retinopathy: segmentation, grading, and transferability. *IEEE Transactions on Medical Imaging*, 40(3), 818-828.
- [10] He, J., Li, C., Ye, J., Wang, S., Qiao, Y., & Gu, L. (2020, April). Classification of ocular diseases employing attention-based unilateral and bilateral feature weighting and fusion. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (pp. 1258-1261). IEEE.
- [11] Bhati, A., Gour, N., Khanna, P., & Ojha, A. (2023). Discriminative kernel convolution network for multi-label ophthalmic disease detection on imbalanced fundus image dataset. *Computers in Biology and Medicine*, 106519.
- [12] Wang, K., Xu, C., Li, G., Zhang, Y., Zheng, Y., & Sun, C. (2023). Combining convolutional neural networks and self-attention for fundus diseases identification. *Scientific Reports*, 13(1), 76.
- [13] Peking University International Competition on Ocular Disease Intelligent Recognition (ODIR-2019), [online] Available: <https://odir2019.grand-challenge.org/dataset/>.
- [14] Kaggle. "Ocular Disease Recognition", Access Date: 11.06.2023. <https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k>
- [15] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- [16] Google Research. "EfficientNet: Improving Accuracy and Efficiency through AutoML and Model Scaling", Access Date: 11.06.2023. <https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html>
- [17] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *arXiv preprint arXiv:1602.07360*.
- [18] Pytorch. "AutoAugment", Access Date: 11.06.2023. <https://pytorch.org/vision/master/generated/torchvision.transforms.AutoAugment.html#torchvision.transforms.AutoAugment>

Artificial Intelligence Based Chatbot in E-Health System

Kamil AKARSU ^{at} , Orhan Er ^b 

^a Department of Computer Engineering, İzmir Bakırçay University, İzmir, Turkey

[†] kamilakarsu94@gmail.com, corresponding author

RECEIVED JUNE 8, 2023
ACCEPTED SEPTEMBER 26, 2023

CITATION Akarsu, K. & Er, O. (2023). Artificial intelligence based chatbot in e-health system. *Artificial Intelligence Theory and Applications*, 3(2), 113-122.

Abstract

The healthcare sector is undergoing a digital revolution due to the rapid growth of technology, and AI technologies are becoming more commonplace in the sector. Chatbots have become useful resources for people to get advice and information about their health issues. The creation and implementation of an AI-based chatbot, integrated with an e-health system, is the main topic of this article. This paper explains the development and creation of chatbots. The chatbot's language comprehension and response capabilities are enhanced through the use of AI techniques such as machine learning and natural language processing (NLP). In addition, the chatbot's user interaction procedure and data security precautions are covered. The paper also examines how the developed chatbot can be integrated into an e-health platform and provides the results of user testing. These evaluations focus on the chatbot's ability to provide accurate and insightful answers, understand user requirements, and provide useful advice. The test results show favourable user evaluations and indicate how well the AI-based chatbot performs in providing healthcare services.

Keywords: chatbot; chatbot in e-health system; sentence matching with deep learning.

1. Introduction

Developed by combining artificial intelligence and natural language processing technology, chatbots are robotic voice systems that can converse with humans in normal language. Using this technology, people can now access healthcare services more quickly and easily, revolutionizing the healthcare sector [1]. The development of chatbots has kept pace with developments in artificial intelligence and computer science. The original chatbots were primarily able to respond to users' simple requests for information, as they were built on a simple question-and-answer basis. However, with the help of deep learning methods and increasingly complex natural language processing algorithms, chatbots have become more sophisticated and intelligent over time.

The use of chatbots in healthcare has the potential to significantly improve patient access to care and increase service productivity. Many healthcare organizations and practitioners are using chatbots in areas such as advice, appointment scheduling, health information dissemination and symptom assessment [2].

There are several advantages to using chatbots in healthcare [3]. Firstly, because chatbots are available 24/7, patients can access healthcare in emergencies or out of

hours. Secondly, chatbots can protect users' confidential health information, alleviating privacy concerns. Chatbots can also answer common questions and help doctors focus on more complicated cases. However, there are significant challenges to using chatbots in healthcare. In order for chatbots to provide accurate and trustworthy answers, they must first be properly trained and provided with health data. This requires access to large and accurate datasets of high-quality training data. In addition, chatbots cannot always take on the role of human doctors and are not always accurate in identifying important medical issues. As a result, the decision to use chatbots should be carefully considered and implemented under the guidance of real doctors [4].

As a result, the use of chatbots in healthcare has the potential to improve patient access to care, increase healthcare productivity, and reduce the workload of healthcare workers. However, care must be taken to design, train and implement the technology correctly. Chatbots can help patients and healthcare professionals communicate and collaborate more effectively and provide patients with a more personalized and accessible experience with healthcare services [5].

2. Literature Review

By providing users with instant access to support, chatbots are essential in the health sector. They help users make informed decisions by providing personalized and reliable information on a range of medical conditions, symptoms and treatments. By freeing up healthcare professionals to focus on more important activities, chatbots can improve the effectiveness of healthcare systems. They enable people to access healthcare resources and assistance at any time, even in remote locations. By analyzing user input and providing relevant information, chatbots can also help in the early detection of health conditions, potentially leading to better health outcomes.

In a study conducted by M. Huang and colleagues [6] aimed to explain how to chatbot is a digital tool that uses text or voice to simulate human conversation and enable interactive communication. It can be accessed through smartphones or computers. Utilizing a chatbot for the follow-up of cancer patients during treatment can be a beneficial solution that also helps healthcare providers save time.

In their study, R. Joshua and colleagues [7] aimed to evaluate whether a multilingual chatbot could effectively engage patients with limited English proficiency (LEP) and improve their outcomes after total joint arthroplasty (TJA). They recognized that language barriers can make the delivery of perioperative instructions challenging for LEP patients, and thus sought to assess the effectiveness of a multilingual chatbot in overcoming these barriers and enhancing patient outcomes in the context of TJA.

In other related study, J. Montenegro and colleagues [8] aimed to further explore the use of chatbots in assisting pregnant women during the prenatal and postnatal periods in Brazil. Similar to the previous study, a pilot study was conducted using a mixed-method design involving healthcare professionals and pregnant women. The participants interacted with the chatbot for seven days and then completed a survey. The results of this study aligned with the previous findings, indicating that pregnant women perceived the chatbot as educational and believed their physicians would approve of its use. The chatbot's performance expectation received positive ratings, while facilitating conditions had a relatively lower influence. Healthcare providers emphasized the benefits of clear language and comprehensive information for pregnant women. The study reaffirmed the viability and usefulness of the presented chatbot and highlighted the potential for future

research to focus on enhancing conversation aspects through Natural Language Processing (NLP) techniques [9]

In a separate study, S. Pandey and colleagues [10] conducted research on the association between increased screen time and its potential health impacts, particularly the detrimental effects on mental health. They employed Deep Learning (DL) and Machine Learning (ML) techniques to examine the influence of technological obsessions on health outcomes. The deployment of chatbots in various industries has proven to be a transformative innovation. The study focused on the development of conversational Artificial Intelligence (AI) systems that enable operators to engage in conversations with machines, resembling human interactions. Two types of chatbots were designed and developed: retrieval-based and generative-based, with each type consisting of six different designs. The accuracy rates for the retrieval-based chatbots varied, with the highest accuracy achieved by the Bidirectional LSTM (Bi-LSTM) [11] at 91.57%. In comparison, the generative-based chatbots, with encoder-decoder designs, exhibited an accuracy rate of 94.45%. A notable distinction between the two types of chatbots is that generative-based chatbots [12] have the capability to generate new text, whereas retrieval-based chatbots are limited to responding based on existing knowledge and outputs.

In a study conducted by Y. Liu and colleagues [13] a comparative analysis was performed to investigate the factors influencing user satisfaction and usage intention in the context of chatbot adoption and development. The study focused on the contrasting experiences of mainland China, where chatbot services have become integrated into daily life, and Hong Kong, where challenges persist in enhancing and promoting chatbot offerings. Through qualitative exploration, critical factors such as perceived quality and privacy concerns were identified. The quantitative findings shed light on the distinct roles of these antecedents across the two regions. Notably, satisfaction was found to positively influence usage intention, with relevance, completeness, pleasure, and assurance emerging as significant factors in both regions.

3. Materials and Methods

3.1. Dataset

In this study, the Instructor Doctor-200k dataset was used to train the deep learning model. The Instructor Doctor-200k dataset, developed by Google AI and Stanford University, is a large collection of human-generated medical questions and answers. It contains more than 200,000 instruction, questions and answers developed by instructors and doctors in the field of medical education [14]. The purpose of this dataset is to develop and test natural language processing models for answering medical questions. The wide range of topics makes it suitable for building models that can answer a wide range of medical questions. Researchers and developers can use the dataset, which can be downloaded from the Google AI Research website, to improve medical question answering systems and increase access to medical information.

There are training and test sets in the Instructor Doctor 200k dataset. While the test set contains only 50,000 instruction, questions and answers, the training set contains 150,000. The questions and answers in the dataset are restricted to being between one and one hundred words long. The dataset also contains medical terms annotated using the Unified Medical Language System (UMLS). Medical software systems often use the UMLS, a complete lexicon of medical terminology. The dataset contains 3 features by structure. These; instruction, input and output [14].

Table 1. Sample data in dataset

Number	input (string)	output (string)	instruction (string)
1	"i had what feels like a muscle cramp about an hour ago under the left bottom rib. it lasted about a minute and then went away. i had no other pains or difficulties since. could this have been a symptom of a minor heart attack. do heart attack symptoms come one at a time or are there more than one symptom when they occur?"	"No this is not a symptom of great attack.... it is normal after some stressful activity. No need to worry. If same thing happen again let me know"	"If you are a doctor, please answer the medical questions based on the patient's description."
2	"My baby has been pooing 5-6 times a day for a week. In the last few days it has increased to 7 and they are very watery with green stringy bits in them. He does not seem unwell i.e no temperature and still eating. He now has a very bad nappy rash from the pooing ...help!"	"Hi... Thank you for consulting in Chat Doctor. It seems your kid is having viral diarrhea. Once it starts it will take 5-7 days to completely get better. Unless the kids having low urine output or very dull or excessively sleepy or blood in motion or green bilious vomiting...you need not worry. There is no need to use antibiotics unless there is blood in the motion. Antibiotics might worsen if unnecessarily used causing antibiotic associated diarrhea. I suggest you use zinc supplements (Z&D Chat Doctor)."	"If you are a doctor, please answer the medical questions based on the patient's description."
3	"my sone has left sided abd pain..It pelvic and rt pelvic pain in the groin area.. can only stand for a short time and sitting impossible for the pain..ct showed enlarged lymph nodes..slight elevation in wt count. he has been 10 weeks and unable to get any relief other than rest and pain meds.. he is 32 6ft 3..approx 220 lb"	"Hi. If there is no relief, it is mandatory to get the biopsy of the node done ASAP to get a correct diagnosis. It may be a serious problem. As a rule - anything not getting OK within 1 to 3 weeks maximum should be removed and tested. Noe-a-days we can do this by laparoscopy too, one day care surgery..."	"If you are a doctor, please answer the medical questions based on the patient's description."
4	"Hi, I have a bump on the top of my helix where usually a piercing would be, but i never had a piercing before, just informing you where this bump is. It looks like a blister, but it doesn t hurt. I don t know how long it s been there. If you could please tell me what it might be that would be great! Thank You."	"Hi, Dear,Welcome with your query to Chat Doctor. Studied your query in full depth of its details. Reviewed it in context of your health concerns. Based On the facts, Dear, you need to consult Surgeon for checking other causes of such a lump. But I would alert you to rule out TB with Erythema Nodes which may be missed as a cause of such a lump on the helix. Fibro-chon Chat Doctor. This reply would help you to plan further treatment soon with your treating doctors. Best of Luck and early recovery. Welcome any further query in this regard, which would be replied in next session. Good Day!!"	"If you are a doctor, please answer the medical questions based on the patient's description."

5 "My baby is 3 month old. She cannot sleep at night.she feels sleepy, so when she is on our lap she is sleeping ,,but when we are trying to put her down in bed , she wakes up.,please let me know the solution of this. We have also checked whether her nose are blocked or not . But she is not having blocked nose."

"DearWelcome to ChatDoctorWe understand your concerns went through your details. 3 year child only understands security and care. She clings to anything that provides security and care. Naturally, she is clinging to you. Let her do it. While putting her down, you also lie down with her. Nursing habit also can be imparted while laying down. Such a change of habit, will change her way of looking at security and care. These are related to training. You need to use alternate techniques of training. If you require more of my help in this aspect, please use this URL. <http://goo.gl/aYW2pR>. Make sure that you include every minute details possible. Hope this answers your query. Available for further clarifications. Good luck."

"If you are a doctor, please answer the medical questions based on the patient's description."

3.2. Model Fine Tuning

The dataset must be transformed into a format compatible with the BERT-Base [15] model before fine-tuning can begin. In this step, the texts are divided into tokens and each token is compared to an index in the dictionary used by BERT. Training on the dataset updates the BERT-Base model as part of the fine-tuning process [16]. Optimization approaches are used to reduce the loss of the model during training. These steps improve the BERT-Base model for the chatbot task. After the fine-tuning process, the BERT-Base model is adapted to better handle the work of building chatbots. The model takes user queries, embeds them in representations, and then generates the best responses.

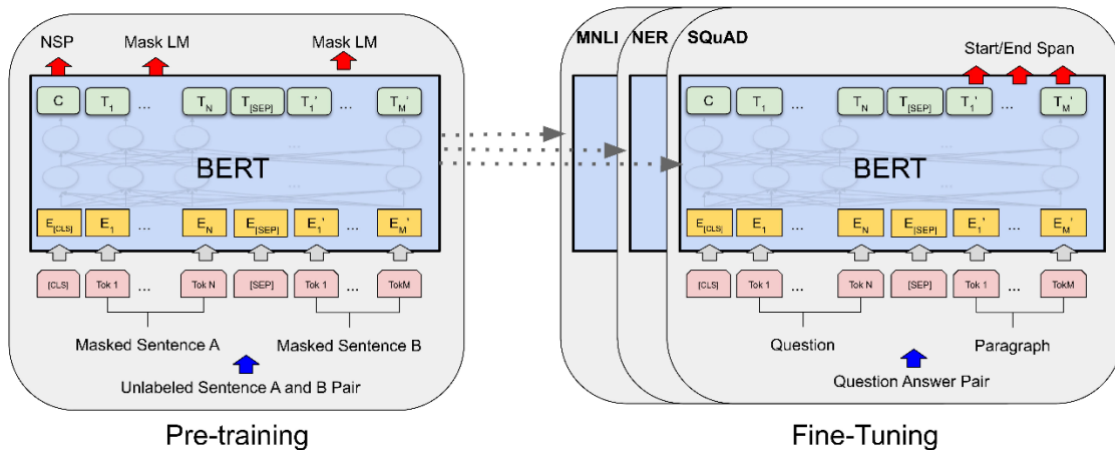


Figure 1. Base and fine-tuned BERT model architecture

We need to use the tokenizer provided by the library in order to use the pre-trained BERT. This is because the model has a specific, fixed vocabulary, and the BERT tokenizer has a special way of handling words that are not in the vocabulary.

In addition, we are required to add special tokens at the beginning and end of each sentence, pad & truncate all sentences to a single constant length, and explicitly specify which tokens are padded with the "attention mask".

3.2. Train The Model

BERT-base consists of 12 transformer layers. Each transformer layer takes a list of token embeddings and produces the same number of embeddings with the same hidden size (or dimensions) on the output. The output of the last transformer layer of the [CLS] token is used as sequence features to feed a classifier.

The Transformers library has a class called BertForSequenceClassification. This class is designed for classification tasks. However, we are going to create a new class so that we can specify our own chosen classifiers.

Below, we will create a BertClassifier class with a BERT model to extract the last hidden layer of the [CLS] token and a single hidden layer feedforward classifier.

3.3. Optimizer and Learning Rate Scheduler

To fine-tune our Bert classifier, we need to create an optimizer. The authors recommend the following hyperparameters:

- Batch size: 32
- Learning rate (Adam): 5e-5
- Number of epochs: 30

We will train our Bert classifier for 30 epochs. In each epoch, we will do training of our model and evaluation of its performance on the validation set.

3.4. Evaluation Criteria

Several performance measures [17] were compared to assess the performance of each model. Accuracy, Recall, Precision, F1 score rate are among the metrics used [18].

For binary classification problems, the discrimination score of the best (optimal) solution during classification training can be defined based on the confusion matrix as shown in Table 2. The row of the table represents the predicted class, and the column represents the actual class. From this confusion matrix, True Positive (TP) and True Negative (TN) indicate the number of positive and negative examples that were correctly classified. Meanwhile, False Negative (FN) and False Positive (FP) indicate the number of misclassified negative and positive examples, respectively. From Table 2, several commonly used metrics can be constructed to evaluate the performance of the classifier with different evaluation foci, as shown in Table 3. Due to multi-class problems, several of the metrics listed in Table 3 have been extended for multi-class classification evaluations.

Table 2. Confusion matrix for binary classification

	Real Positive Class	Real Negative Class
Predicted Positive Class	True Positive (TP)	False Negative (FN)
Predicted Negative Class	False Positive (FP)	True Negative (TN)

Accuracy is the most commonly used evaluation metric in practice for both binary and multi-class classification problems. Accuracy evaluates the quality of the generated solution based on the percentage of correct predictions over the total number of

examples. The complementary accuracy metric is the error rate, which evaluates the generated solution based on the percentage of incorrect predictions. Both metrics have been widely used in practice by researchers to discriminate and select the optimal solution.

The advantages of accuracy or error rate are that this metric is easy to compute with less complexity, valid for multi-class and multi-label problems, easy to use scoring, and easy for humans to understand. As many studies point out, accuracy metrics have limitations in evaluation and discrimination processes. One of the main limitations of accuracy is that it produces less discriminative and less distinguishable values. As a result, it leads to less discriminative power for accuracy in selecting and determining the optimal classifier.

Table 3. Evaluation metrics

Metric	Formula	Evaluation
Accuracy (ACC)	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	In general, the accuracy metric measures the ratio of correct predictions to the total number of samples evaluated.
Error Rate (ERR)	$Error Rate = \frac{FP + FN}{TP + TN + FP + FN}$	Misclassification error measures the ratio of incorrect predictions to the total number of samples evaluated.
Recall	$Recall = \frac{TP}{TP + FN}$	This metric is used to measure the fraction of correctly classified positive patterns.
Precision	$Precision = \frac{TP}{TP + FP}$	Precision is used to measure accurately predicted positive patterns out of the total predicted patterns in a positive class.
F-Measure (FM)	$f_{\beta} - measure = \frac{(1 + \beta^2) \times (recall * precision)}{\beta^2 + recall * precision}$	This metric is obtained by taking the harmonic mean of the precision and rating values.

3.5. Experimental Results

Table 4. Performance values of model

Accuracy	Precision	Recall	F1 Score
0.88	0.89	0.66	0.77

The classifier results of our fine-tuned Bert model are shown in Table 4. The 88% Accuracy rate is actually at a debatable high level of performance for fine tunings made with data sets with such a high number of samples. We believe that we can increase the Accuracy and Precision rate by constantly changing the main elements that affect performance, i.e. the fine-tuning hyperparameters.

In Figure 2-3-4-5, we see the results of the derived classification performance metrics.

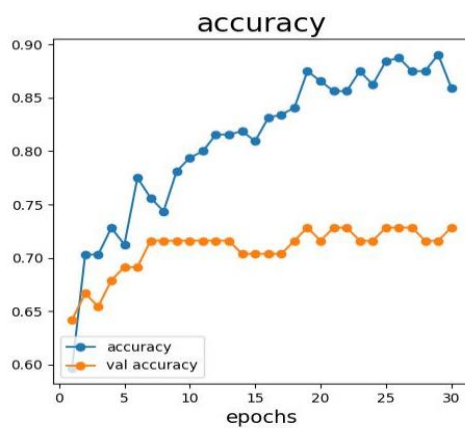


Figure 2. Accuracy graph

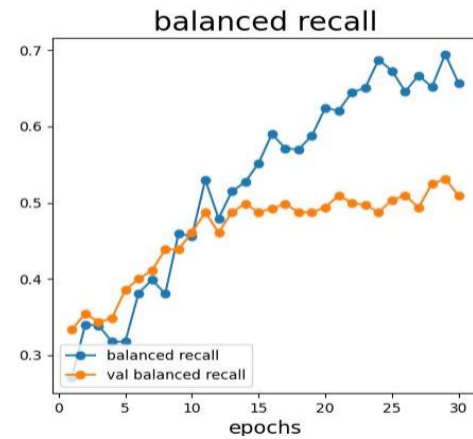


Figure 3. Balanced recall graph

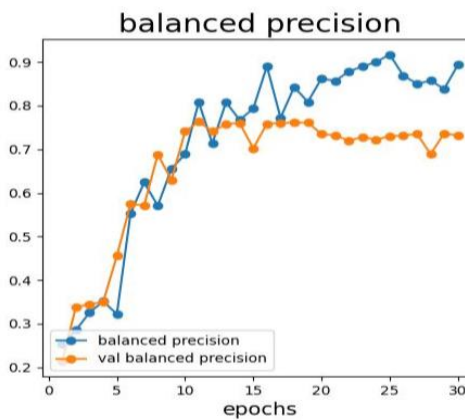


Figure 4. Balanced F1 score graph

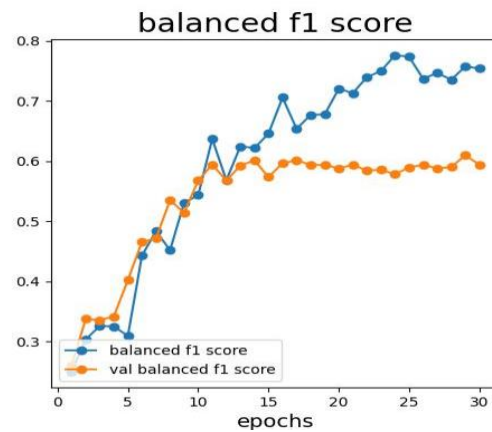


Figure 5. Balanced precision graph

4. Conclusion and Future Works

In conclusion, the integration of artificial intelligence (AI) in the form of a chatbot into the e-health system has the potential to completely transform the way healthcare is delivered. The AI-based chatbot has proven to be an effective tool in providing personalized and timely healthcare information, assisting in triage, and improving patient engagement and satisfaction. The chatbot's performance in understanding user queries and providing accurate responses has been enhanced using machine learning algorithms, including deep learning models such as BERT. The use of chatbots in the e-health system has shown promising results in reducing healthcare costs, increasing accessibility, and improving overall healthcare outcomes.

Despite the progress and success of AI-based chatbots for e-health, there are several areas that warrant further exploration and improvement. First, the integration of more advanced natural language processing techniques, such as sentiment analysis and emotion detection, can enable the chatbot to better understand and respond to the user's emotional state, thereby providing more empathetic and personalized interactions. In addition, the integration of multimodal capabilities, such as speech and image

recognition, can further enhance the functionality of the chatbot and enable it to provide more comprehensive support in the diagnosis and monitoring of health conditions. In addition, continuous refinement and expansion of the chatbot's knowledge base through regular updates and integration with reliable medical databases and resources will ensure that it remains up to date with the latest medical information. Integration with electronic health record (EHR) systems and interoperability with other healthcare platforms can also facilitate seamless information sharing and enable the chatbot to provide more personalized recommendations based on an individual's medical history. Ethical considerations and privacy are also critical areas that require attention. Ensuring that the chatbot complies with privacy regulations and maintains data security is paramount to building user trust. Ongoing research and development should focus on implementing robust privacy measures and conducting regular audits to address potential vulnerabilities. In conclusion, the future of AI-based chatbots in e-health is promising. By addressing the aforementioned areas for improvement and exploring new avenues, we can unlock the full potential of chatbots to revolutionize healthcare delivery, improve the patient experience and ultimately improve overall healthcare outcomes.

References

- [1] C. Zielinski *et al.*, "Chatbots, ChatGPT, and Scholarly Manuscripts - WAME Recommendations on ChatGPT and Chatbots in Relation to Scholarly Publications," *Afro-Egyptian Journal of Infectious and Endemic Diseases*, vol. 13, no. 1, pp. 75–79, Mar. 2023, doi: 10.21608/AEJI.2023.282936.
- [2] N. Bhirud, S. Tatale, S. Randive, S. Tataale, and S. Nahar, "A Literature Review On Chatbots In Healthcare Domain Computational Feasibility of Paninian Grammar for Indian Languages' Analyses View project Machine Learning View project A Literature Review On Chatbots In Healthcare Domain," *International Journal Of Scientific & Technology Research*, vol. 8, p. 7, 2019, Accessed: Jun. 06, 2023. [Online]. Available: www.ijstr.org
- [3] S. Laumer, C. Maier, F. Tobias Gubler, and F. Tobias, "Chatbot Acceptance In Healthcare: Explaining User Adoption Of Conversational Agents For Disease Diagnosis," 2019, Accessed: Jun. 06, 2023. [Online]. Available: https://aisel.aisnet.org/ecis2019_rp/88
- [4] "View of Doctor Recommendation Chatbot: A research study." <https://sabapub.com/index.php/jaai/article/view/310/240> (accessed Jun. 06, 2023).
- [5] Y. Windiatmoko, R. Rahmadi, A. F. Hidayatullah, R. Pradhan, J. Shukla, and M. Bansal, "K-Bot' Knowledge Enabled Personalized Healthcare Chatbot", doi: 10.1088/1757-899X/1116/1/012185.
- [6] M.-Y. Huang, C.-S. Weng, H.-L. Kuo, and Y.-C. Su, "Using a chatbot to reduce emergency department visits and unscheduled hospitalizations among patients with gynecologic malignancies during chemotherapy: A retrospective cohort study," 2023, doi: 10.1016/j.heliyon.2023.e15798.
- [7] J. P. Rainey *et al.*, "A Multilingual Chatbot Can Effectively Engage Arthroplasty Patients With Limited English Proficiency," 2023, doi: 10.1016/j.arth.2023.04.014.
- [8] J. Luis, Z. Montenegro, C. André Da Costa, and L. P. Janssen, "Evaluating the use of chatbot during pregnancy: A usability study," *Healthcare Analytics*, vol. 2, p. 100072, 2022, doi: 10.1016/j.health.2022.100072.
- [9] E. D. Liddy, "Natural Language Processing Natural Language Processing Natural Language Processing 1," 2001, Accessed: Jun. 06, 2023. [Online]. Available: <https://surface.syr.edu/istpub>
- [10] S. Pandey and S. Sharma, "A comparative study of retrieval-based and generative-based chatbots using Deep Learning and Machine Learning," *Healthcare Analytics*, vol. 3, p. 100198, 2023, doi: 10.1016/j.health.2023.100198.
- [11] A. Graves, N. Jaitly, and A.-R. Mohamed, "Hybrid Speech Recognition With Deep Bidirectional LSTM".
- [12] J. Kapoči and "Ut' E-Dzikien", "A Domain-Specific Generative Chatbot Trained from Little Data", doi: 10.3390/app10072221.
- [13] Y.-L. Liu, B. Hu, W. Yan, and Z. Lin, "Can chatbots satisfy me? A mixed-method comparative study of satisfaction with task-oriented chatbots in mainland China and Hong Kong," 2023, doi: 10.1016/j.chb.2023.107716.
- [14] "z111/ChatDoctor · Hugging Face." <https://huggingface.co/z111/ChatDoctor> (accessed Jun. 06, 2023).
- [15] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", Accessed: Jun. 06, 2023. [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [16] H. Wang, C. Focke, R. Sylvester, N. Mishra, and W. Wang, "Fine-tune Bert for DocRED with Two-step Process", Accessed: Jun. 06, 2023. [Online]. Available: <https://github.com/>

- [17] X. Deng, Q. Liu, Y. Deng, and S. Mahadevan, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Inf Sci (N Y)*, pp. 250–261, 2016, doi: 10.1016/j.ins.2016.01.033.
- [18] "Classification Model Evaluation Metrics", doi: 10.14569/IJACSA.2021.0120670.

Modern Computer Tomography with Artificial Intelligence and Deep Learning Applications

Coşkun Deniz ^{a †} 

^aDepartment of Electrical and Electronics Engineering, Aydın Adnan Menderes University, Aydın, Türkiye

[†]cdeniz@adu.edu.tr, corresponding author

RECEIVED JUNE 20, 2023
ACCEPTED AUGUST 31, 2023

CITATION Deniz, C. (2023). Modern computer tomography with artificial intelligence and deep learning applications. *Artificial Intelligence Theory and Applications*, 3(2), 123-136.

Abstract

X-ray computed tomography (CT) aims production of 2-dimensional mass-density (or X-ray attenuation coefficient) maps of the sliced interior body by using directed X-rays through it to construct 3D CT images from the collection of these sliced 2D maps. Since the CT scan provides us with the interior structure of the body without any cut or physical damage, it is indispensable in our modern medical applications. However, since the X-rays involve ionizing radiation, it is dangerous for living organisms and it brings about the ALARA (as low as reasonably achievable) principle in medical applications emphasizing as high-quality CT images (with the highest possible resolution) as possible by using as little X-ray exposure of the body under scan as possible. This challenging task along with the correct interpretation of these CT images to lead a correct diagnosis and treatment plan brings about designing various fan geometries, scanning styles, and advanced image reconstruction techniques in the evolution of X-ray CT scans. We can see that X-ray CT scans have been evolved enormously since the first discovery in the early 1970s and it continues today with applications of the artificial intelligence (AI) and deep learning (DL) in our modern CT with promising successful results. In this work, a pedagogical study of our modern X-ray CT with the related review of literature regarding *i-scanning geometry*, *ii-reconstruction techniques*, and *iii-AI&DL applications* is being presented hoping to be useful as a quick reference especially for the scholars and researchers in the field.

Keywords: X-ray computed tomography; X-ray CT; artificial intelligence; deep learning; DL.

1. Introduction

Goal of the computed tomography (CT) is for sure to produce mass-density maps (or maps of X-ray attenuation coefficient distribution as will be explained here soon afterwards) of the internal structure of the 3D objects, i.e., a human body (or parts of the body), as sliced 2D graphs to constitute a 3D CT-scan graph by using directed beams of waves (or rays) through it without cutting or looking from any internal point of it. Ultrasound or X-rays can be used as directed beams through the body to achieve this purpose and we study only the X-ray CT here in this work. As we all are familiar in our modern life today, the CT has become an indispensable tool in our modern medical applications where these produced CT images are frequently used for diagnosis and patient treatment plan by the physicians. Although there is no cut or any means of physical damage to the body under scan to pay for such wonderful images, which also

makes it indispensable in the practical applications of today's health sciences, there is an ALARA (as low as reasonably achievable) principle in medical applications emphasizing as high-quality CT images (with the highest possible resolution) as possible by using as little X-ray exposure of the body under scan as possible [1-9]. To maintain this principle, along with the other demands such as scanning speed, image quality, performance, computational speed, etc., CT scanning technique has been evolved with revolutionary improvement since the first discovery in 1970s [10] and it continues with the promising applications of today's artificial intelligence (AI) and deep learning (DL) techniques [1,2,4,6-9,11]. This work deals with a brief review of it in a classification and dimensional analyses of the literature regarding the associated principles and techniques in the field.

In this work, we focus on components, development, and literature review of our modern image reconstruction (IR) techniques in the X-ray CT with the conventionally and commercially used analytic image reconstruction (AIR) and iterative image reconstruction (IIR) techniques within the seven-generations of scanning geometries along with the emphasis on recent applications of modern AI and DL in the field. In this work, we shortly study classification of literature in the modern X-ray CT with the dimensional analyses to be as a quick reference guide hoping to be useful for the scholars and the researchers in the field. Main abbreviations for the X-ray CT concept we used in this work (as also conventionally used in the literature) are listed as shown in Table 1.

Table 1. Abbreviations for the X-ray CT concept we used in this work

Name of the category	Name of the category
AI: Artificial intelligence	DLIR-M: Medium-dose DL (applied) image reconstruction
AIR: Analytic image reconstruction	FBP: Filtered back-projection
AR: Analytic (image) reconstruction	IR: Image reconstruction
ART: Algebraic Reconstruction Technique	IIR: Iterative (image) reconstruction
CT: Computer tomography	SART: Simultaneous Algebraic Reconstruction Technique
DL: Deep learning	SIIRT: Simultaneous Iterative Reconstr. Technique (SIIRT)
DLIR: Deep learning (applied) image reconstruction	MLEM: Maximum Likelihood Expectation Maximization
DLIR-H: High-dose DL (applied) image reconstruction	PL: Penalized Likelihood
DLIR-L: Low-dose DL (applied) image reconstruction	

2. Components and development of X-ray CT

From literature, we can say that processes in medical applications of X-ray CT involve the following steps [1,2,4-6,8,9,11-13,16]: i) Scanning the body of the patient by the X-rays and recording the measured intensities of the receding X-rays from the patient (by sensors), which can be called as construction of image data, ii) Acquiring the 2D or 3D graphs of internal structure from scan data, which is called as image reconstruction (IR), and iii) Further image processing of the image reconstructed X-ray CT-scan graphs, such as segmentation, classification, deblurring, etc., which can be called as post IR processing or simply image processing (IP), and iv) Interpreting the image reconstructed CT-scan graphs or post IR-processed graphs by the physicians, i.e., diagnosis, treatment plan, etc., which can be called as clinical (or medical routine) as shown schematically in Fig.1. Artificial Intelligence (AI) and Deep Learning (DL) based parts where we see in IR and post IR-processing steps are marked in red in Fig.1.

We see that major improvement of X-ray CT-scan in history has been in two fields: i) in scanning process and ii) in IR-process. *Regarding the scanning process*, chronologically we see that there are *seven generations of CT-scans* where design of scanning

apparatus and scanning geometry have been improved with revolutionary designs so far [2,4,10,12,14,15,17-21]. Various scanning geometries are actually the pencil-beam approximation of the early generation X-ray CT scans where discrete scanning angle (θ) and separation of sensors (s) are the two-fundamental parameters forming the dataset of the measured output intensity, in the form, say $I(s, \theta)$, from which the internal mass-density map of the related slice of the body, $f(x, y)$, is formed (or in CT terms, reconstructed) [12,14,17,22]. Regarding the image re-construction process, since the various scanning geometries are basically the pencil-beam approximated early generation CT-scans, there are basically two fundamental reconstruction techniques inherent from the early generation scans: i) *Analytic image reconstruction (AIR)* [4,23-26] and ii) *iterative image reconstruction (IIR) techniques* [5,24-26]. Note that, IIR methods are sometimes referred to as (or related with) algebraic IR as in [4,12,13,27]. The AIR involving (and also frequently used for AIR interchangeably) the conventional *Filtered back propagation (FBP)* method was the originally used image reconstruction technique used in the early stage of the discovery of the CT-scan through the 1990s and it involves the Fourier-slice theorem where the Fourier transform and reverse Fourier transform are involved with the filtered back-projection (FBP) applications [4,23-26]. Although conventional FBP technique provides theoretically exact solutions assuming noiseless measurement of output X-rays as a line integral of attenuation (ray-sum) through the patient in the IRC process, reconstructed images are normally blurred in practical applications but they can be enhanced by the use of some filters to some extent [4,24-26]. On the other hand, IIR techniques basically rely on the fundamentals of Kaczmarz's algorithm, which was originally proposed by Kaczmarz in 1937 [28], which technically enables better images in greater performance but it could not have found clinical applicability by 2009 at which substantial advances in the hardware technology occurred [13,26].

Most of the recent X-ray CT scans (those without AI and DL applications) use some advanced IR techniques such as Simultaneous Algebraic Reconstruction Technique (SART), Maximum Likelihood Expectation Maximization (MLEM), and Penalized Likelihood (PL), etc., since being in higher performance thanks to the modern advanced hardware technology [5,23,24]. However, the ALARA principle causes restrictions in practical medical (clinical) applications especially in pediatric or other risky applications due to the danger of the exposure of high-dose X-rays for such patients since production of high-quality X-ray CT images require high dose X-ray applications in both AIR and IIR methods [2,17,22]. But this problem seems being solvable through today's advances in the AI and DL fields [1,2,4,6-9,11]. We can see that AI and DL applications to the IRs can provide high quality X-ray CT images with promising results as being used in today's modern X-ray CT scans frequently. We study it in Sect. 5.

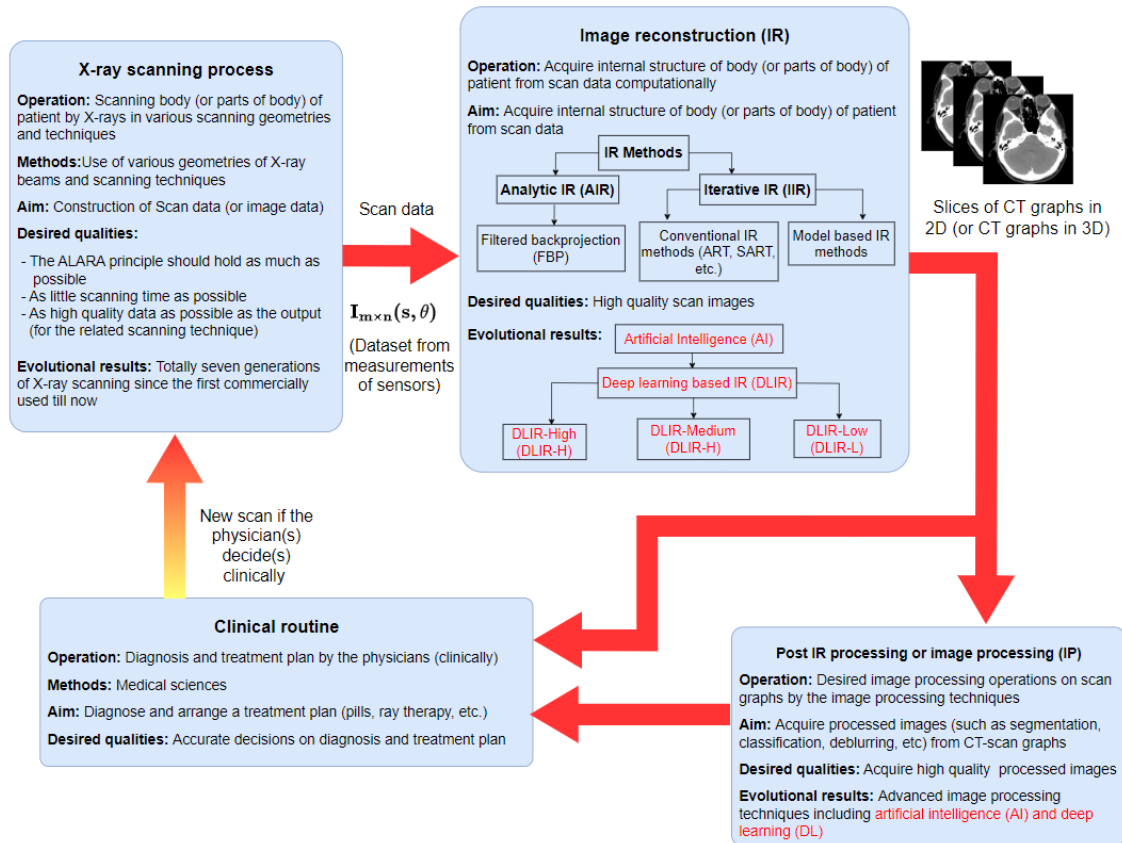


Figure 1. Steps of medical applications of modern X-ray computed tomography

3. Pencil Beam approximation and the Image Reconstruction (IR) Techniques

A schematic sketch of the “*pencil-beam approximated X-ray CT scan*” given in Fig. 1 describes the fundamental model where parallel beams of X-rays are collimated in the form of a so-called “*pencil-beam shape*” where the beams are assumed to be directed from top of a pencil to its tip as an analogy through the patient’s body. This model is actually originated from the first-generation scanners which is actually entirely different in comparison to today’s scanners where beam shapes (being in fan, conic, etc.) and scanning styles (being in stationary or rotary multi-sensor array in the gantry, helical scan, etc.) are completely different as discussed in the next section. But majority of the scanning techniques can mainly be modelled by this model to explain the fundamental principles of the conventional IR techniques in the X-ray CT [12,14]. We also note that calculations regarding IRs of beam shapes other than pencil-beam approximation are studied in details in [12]. As can be seen from Fig.2, parallel pencil-beam shaped X-rays are very thin (about 1 mm to 10 mm) and they are incident on an outer surface of a 3D object (such as a patient’s body or some part of it) to pass through it with an attenuation as a measure of the mass-density distribution of the interior region on the path in the longitudinal direction. The detectors on the other side of the body detect and measure this attenuated output intensities receding from the patient’s body when they hit the sensors. Intensities of the parallel input beams (incident on the patient’s body) are set to a constant value, say I_0 , and the attenuated output beams (from the patient’s body) for each scan angle θ is detected and measured by the parallel sensors separated by a distance of s , say $I(s, \theta)$. An entire scan of a slice of a body completes with a total of n numbers of scans between start and end scanning angles θ_1 and θ_n , namely in $\theta_1 \leq$

$\theta_i \leq \theta_n$, as the pairs of sensor and detector arrays attached to the gantry rotates in discrete $\theta \rightarrow \theta_i$ angles as shown in Fig. 1. Since X-rays interact with matter with a longitudinal attenuation depending on the mass-density of the matter at that point via the *Beer-Lambert law* [14,18,23,29-31], measured output intensity of each beam (from the patient's body) is attenuated as a sum of attenuations of each tiny-sliced cubes (or rectangular prisms), called as *voxels* (as an implication of three-dimensional pixels), in comparison to the input intensity as shown in Fig.1. Consequently, each output beam involves sum of attenuations of each voxel, in other words, sum of mass-density distribution of the body between its end points in the direction of the related X-ray beam through its path. Once a complete scan of a slice is completed for all angles, a 2D graph regarding the measured output intensities for all these scan angles at all the sensor locations, called as a *sinogram*, i.e., [1,12,24], is produced, then the internal structure of the body in the form of mass-density function, say $f(x, z)$, is computationally produced. To illustrate, bones have higher mass-densities hence greater attenuation coefficients in comparison to the fatty regions and a 2D graph of $f(x, z)$ is obviously informative to the physicians showing how the internal structure of the patient's body is.

Fundamental IR techniques, which can be modelled as the pencil-beam approximated model in Fig. 2 as explained above, can be classified conventionally in two groups, namely, analytic image reconstruction (AIR) techniques and iterative image reconstruction (IIR) techniques as being studied in the next sections.

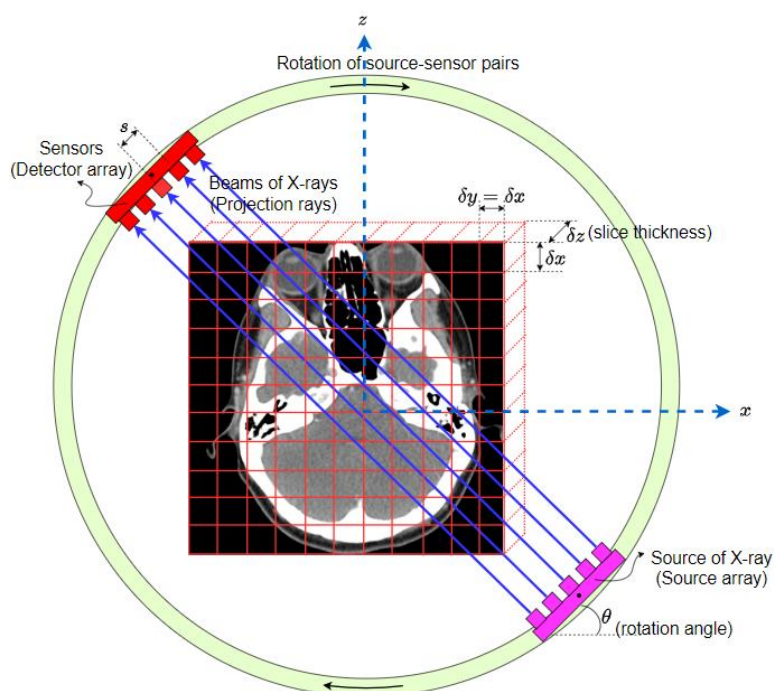


Figure 2. A schematic sketch of pencil-beam approximated X-ray CT-scan.

3.1. A short Review of Analytical Image Reconstruction (AIR) Methods

Most commonly used analytical reconstruction methods on commercial CT scanners are all in the form of filtered back-projection (FBP), which uses a one-dimensional filter on the projection data before back-projecting them onto the image space in 2D [12,17,19,24]. From the conventional pencil-beam approximated model in Fig. 2, we see that both source arrays and the detector arrays (in the opposite side) attached to each

other rotate together in the gantry by an angle of θ for each detector measurement and the process repeats n times in $\theta_1 \leq \theta_i \leq \theta_n$ where i is the index of scan angle (or simply scan index) to complete the whole scan of one slice in total n -scans. Since the intensity of the X-ray detected at the j -th detector (for a specific i -th scan) is attenuated by the longitudinal density of the body under study (along the propagation direction of the X-ray beam), intensities measured by the j -th detector is the line integral of the attenuated intensities (via the Beer-Lambert law) along the path from source to the detector. Another expression involves mathematically a line integral of the mass-density distribution of that region (or voxel) via Dirac Delta function $\delta(x\cos\theta_i + z\sin\theta_i - s)$ to select the correct path on the line $x\cos\theta + z\sin\theta$ (See Fig.1), namely,

$$p_i(s, \theta_i) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, z) \delta(x\cos\theta_i + z\sin\theta_i - s) dx dz. \quad [1]$$

Consequently, projection $p_i(s, \theta)$ represents a line integral whose input is the detector readings for the projection angle $\theta \leftarrow \theta_i$ given in Fig. 2, which is referred to as *mass Radon transform* (or simply Radon transform) [6,12,13,31-33], Now the problem is the inverse process to find $f(x, y)$ from $p_i(s, \theta)$, which is referred to as *inverse Radon transform*. Back projection is not exactly an inverse Radon transform but a kind of a conjugate process to assign a point (x, y) in the object coordinates. Reconstruction process involves the solution of the measured intensity involving measured angle and linear shift position of the related X-ray tube in terms of integral equations by inversion, which is known as the *Back projection*. Traditionally, a filter, called as Ram-Lak filter to compensate the low-pass blur due to the different numbers of projections passing through the center and periphery of the image region is used in the FBP method [6,24,31-33]. FBP involves the Radon transform described above and the Fourier Slice Theorem (FST) which states that the one-dimensional Fourier transform $p(\omega, \theta)$ of a projection $P(s, \theta)$ in parallel-beam geometry for a fixed rotation angle θ is identical to the one-dimensional profile through the origin of the 2-D Fourier transform $F(\omega\cos\theta, \omega\sin\theta)$ of the irradiated object element in (x, y) .

3.2. A short Review of Iterative Image Reconstruction (IIR) Methods

Conventional IIR methods involve the following: Algebraic Reconstruction Technique (ART), Simultaneous Algebraic Reconstruction Technique (SART), Maximum Likelihood Expectation Maximization (MLEM), Penalized Likelihood (PL), etc. [13,23,30]. In this work, we focus on the ART technique, which is the mostly studied technique originally introduced by Kaczmarz in 1937 for the first time [28]. In the ARTs, related matrix obtained by the detector measurement for various angle scans have an M by N equation system as follows [13, pp.276-278]:

$$W.F = P \Rightarrow \sum_{j=1}^n w_{ij} \vec{f}_j = p_i, \quad s.t. \quad i = 1, 2, \dots, M; j = 1, 2, \dots, N \quad [2]$$

where M is the total number of rays (in all projections), p_i is the ray-sum measured for the i -th ray, w_{ij} is the weighing parameter where Kaczmarz's method is associated. Here value of $f(x, z)$ as a square grid on the related voxel is considered to be constant (since the size of the voxel is too small) and f_j denotes this constant value in the j th cell where N is the total number of cells in the path of the related beam. The problem then becomes finding the unknown \vec{f}_j values from the system of M equations in N unknowns

(where W becomes the matrix of constant coefficients) by the iterative methods as being presented in [13]. To illustrate it, let us suppose a standard 255x255 image, then it means: $N \approx 65.000 \approx M$ with $W: M \times N$ matrix (with many zeros due to the voxels not in the beam path). Such a hard work by iterative methods became applicable thanks to the advances in hardware technology around 1990s.

Each linear equation in the system forms a line and the solution is the point of their intersections. Computation starts with an initial guess then it follows its projection firstly on the first line and secondly its projection to the second line to continue the same procedure for all lines. Then it repeats by taking it as the initial guess of the second tour to converge to the intersection point. Finally, it stops when the desired accuracy (or iteration number) has been reached. In each iteration with index i , these guess points have a displacement vector: $\vec{f}^{(i)}$ which is composed of N components: $\vec{f}_1^{(i)}, \vec{f}_2^{(i)}, \dots, \vec{f}_N^{(i)}$ in the N dimensional space. For example, our initial guess with $\vec{f}^{(0)}$ is firstly projected on the related hyperplane to give $\vec{f}^{(1)}$, then it is projected on the second equation to give $\vec{f}^{(2)}$, and the process continues similarly. In order to get $\vec{f}^{(i)}$ from the $\vec{f}^{(i-1)}$, Kaczmarz's algorithm has the following operation mathematically [13]:

$$\vec{f}^{(i)} = \vec{f}^{(i-1)} - \frac{[\vec{f}^{(i-1)} \cdot \vec{w}_i - p_i]}{\vec{w}_i \cdot \vec{w}_i} \cdot \vec{w}_i \quad [3]$$

where $\vec{w}_i = (w_{i1}, w_{i2}, \dots, w_{iN})$ and $\vec{w}_i \cdot \vec{w}_i$ is the dot product. Proofs and details including the convergence conditions and illustrations for higher dimensions are available in [13,31,34]. Fundamental iterative image reconstruction techniques such as Algebraic Reconstruction Techniques (ART), Simultaneous Iterative Reconstruction Technique (SIIRT), Simultaneous Algebraic Reconstruction Technique (SART), etc. works on this principle.

4. Scanning Geometry and evolution of CT scans

Since the first model of X-ray CT scanner, scanning geometry (and hence scanning mechanism and scanning technology) has been evolved enormously and each revolutionary design has been referred to as a new generation scan. There are totally *seven-generations of X-ray CT-scanners* developed and used so far, all being completely different from each other [2,4,10,12,14,15,17-21]. Although scanning mechanisms could be different, image reconstruction techniques can be studied via the pencil-beam approximated X-ray CT-scan whose schematic sketch is given in Fig.2 since they can be approximated and modelled by it as explained above.

The first and second-generation CT-scanners being of the only translational types among them are as shown schematically in Fig.3. First generation scanner shown Fig.2a has a unique X-ray source and sensor pair and it starts scanning translationally along the x -axis for angle $\theta \rightarrow \theta_A = 0^\circ$ in time t as $t_{A0} \leq t \leq t_{An}$ then angle changes from $\theta_A = 0^\circ$ to θ_B by rotating around the y -axis and the translational scan starts again but in the reverse direction (along the $-x'$ -axis now) in the rotated coordinate system. When it completes, angle then changes from θ_B to θ_C and the translational motion but now being in the reverse direction (along the $+x'$ -axis) in the rotated coordinate system starts again. Consequently, first-generation X-ray CT scanners scan in "*translational-rotational*" modes [10,12,14,18]. Second generation scanner shown Fig.2b has the same "*translational-rotational modes*" like the first generation but it includes many sensors as a sensor array in comparison to the first generation. We see that scanning speed is

increased in the second generation. The first and the second-generation CT scanners are the only scanners having translational scanning modes and no more used in today's medical applications.

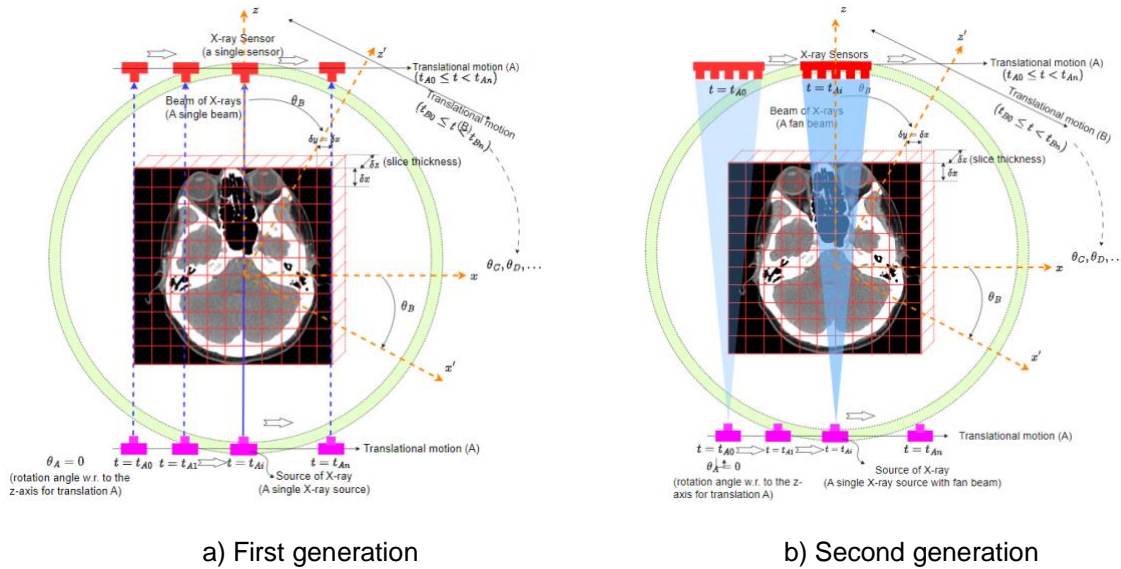


Figure 3. A schematic sketch of the first and second-generation X-ray CT-scanners (not in scale-patient size is exaggerated).

The third and fourth-generation CT-scanners being of the first only-rotary types are as shown schematically in Fig.4. We note here that size of the patient's body (or the organ of the patient under study) in the schematic sketches in Fig. 3&4 are exaggerated with respect to the size of the girth since the beams shown in blue are approximately paraxial (and they would actually fill the entire body if the size of the body were not exaggerated). The third-generation scanner shown in Fig. 4a has a multisensory array attached to a single X-ray source rotates around the y-axis on the gantry. Scanning principle of the fourth-generation CT scanner shown in Fig. 4b is similar to the third generation but the sensor array is stationary and fills the entire inner surface of the gantry [12,18]. The third and fourth-generation CT scanners are the "only rotary mode" scanners (no translational mode).

Several other CT scanning geometries have been developed and marketed afterwards but none of them fits precisely the conventional CT categories studied above. Remaining generations are technologically advanced scanners. Since the scanning time of the first four generations are too long for some certain applications such as cardiac-scans, the fifth generation X-ray CT scanners were developed. They involve an electron beam accelerated by the applied high voltages and deflected by the deflection coils to a stationary tungsten-arc target attached to the inner surface of the gantry running entirely in a *stationary mode* [4,12,14,17]. In the fifth-generation scan, there is no translational or rotational part relative to the gantry, and hence it can be counted as a "*stationary mode*" scanner. As the accelerated and deflected high-energy electrons hit the tungsten arc, X-rays are produced at the point of impact. In other words, the tungsten arc acts as an instantly triggerable X-ray source at the desired position by the high energy electrons. Since the accelerated and deflected electron beams hit the tungsten arc in a very fast rotary scanning nature, scanning speed increases enormously for the applications

requiring high-speed scan like cardiac applications. It is capable of ultra-fast scans, i.e., 50-millisecond scan time and production of 17 CT slices/second, for such special and advanced applications. Its use in cardiac tomographic imaging is frequently referred to as “*cine CT*”.

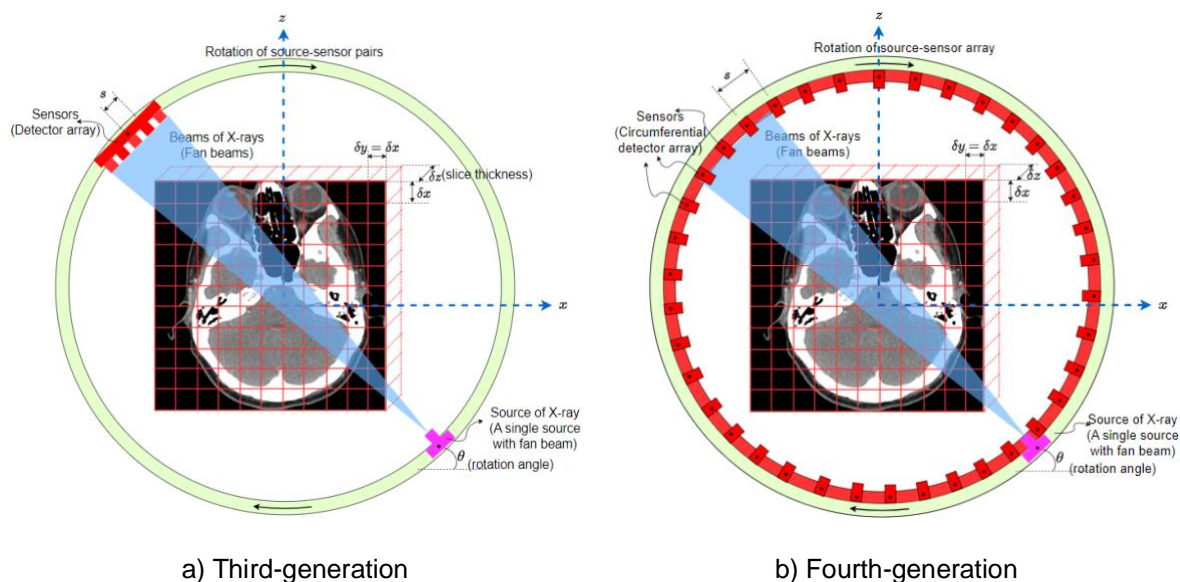


Figure 4. A schematic sketch of the third and fourth-generation X-ray CT-scanners (not in scale-patient size is exaggerated).

Although the fifth-generation scanners present much more desired benefits, they are too expensive and consequently, a doughnut-shaped sixth-generation CT scanner was developed. It is frequently referred to as “*spiral/helical CT*” and it involves “*X-ray source rotation*” and “*exam-table translation*” perpendicular to it to form a spiral (or helical) scan around the patient [4,10,12,14,17,18]. In effect, the patient lying on the exam table passes through a doughnut shaped rotating scanner or in other words, x-ray tube rotates in helical path. In conventional scanners, scanning time could be around 10-15 minutes but in some instances, entire scan can take a single breath-hold time in the 6th generation CTs. It is also called as “*volumetric scanners*” since having a single fan-beam source and a stationary multidetector array in the gantry to fill a volume of X-rayed tissue. It provides high image resolution and improved image quality. Due to the helical path, a three-dimensional data set is obtained and they are used for image reconstruction into sequential images for a stack. In the sixth generation, the tube is energized continuously, and data are also collected continuously. The gantry also rotates continuously. It involves a slip-ring technology and there are three slip-rings attached to the gantry as terminal connections to the X-ray tube, detector and control sensors which enables a high-speed rotation. Slip-ring arrangement enables a rotation speed around 5s/rot a rotation angle more than 360° as a continuous rotation with continuous data acquisition. As to the seventh-generation scanners, it involves multi-sliced detectors allowing acquisition of multiple slices in single row and this makes it to be frequently referred to as “*MS (multi-slice) CT-scanners*” [2,14,17,19]. They are the most advanced X-ray scanners developed so far. It generally uses third generation CT shown in Fig.4a with helical scanning and low voltage slip rings. In the seventh generation CTs, A body section can be scanned faster with a multiple row of detectors system with multiple fan beams scanning simultaneously. It also uses slip-ring technology for switching power and image

data. It involves a conical shaped fan-beam and a fast rotation speed (about 0.5 to 0.8 sec per rotation) which reduces the examination time. The image quality is mostly similar to that of single slice scanners. It enables high spatial resolution.

5. Artificial Intelligence (AI) and Deep Learning (DL) in X-ray CT

We have seen that scan techniques (or geometries) and IR techniques in X-ray CT have evolved to maintain the ALARA principle to produce CT images in high qualities besides other needs such as high performance, high scan and IR speed, etc. Now we continue with that application of AI and DL in the IR stage, as marked in red in Fig. 1, being *closely related to the scan technique and the IR techniques* we studied above, can give promising results. We will also see that, beside the IR stage, AI and DL can also be applied as post-IR applications for some specific purposes of image processing, like segmentation, classification, deblurring, etc. as their classical uses to maintain clinical needs. In X-ray CT, any result which might seem perfect in X-ray CT graphs should also be tested and studied clinically [1,35]. Hence, we also review the related articles in this section.

DL is actually a sub-class of the Machine learning (ML) which is also a sub-class of the AI as shown in Fig. 5 [1,35]. The Artificial neural networks (ANNs) are one of the methods in the ML and Deep neural networks (DNN) are some special types and more advanced form of the ANNs. DNNs involve the Convolutional neural networks (CNNs) which involves advanced network architecture and their principles are available in [1,11,35] as well as ordinary conventional AI/DL textbooks.

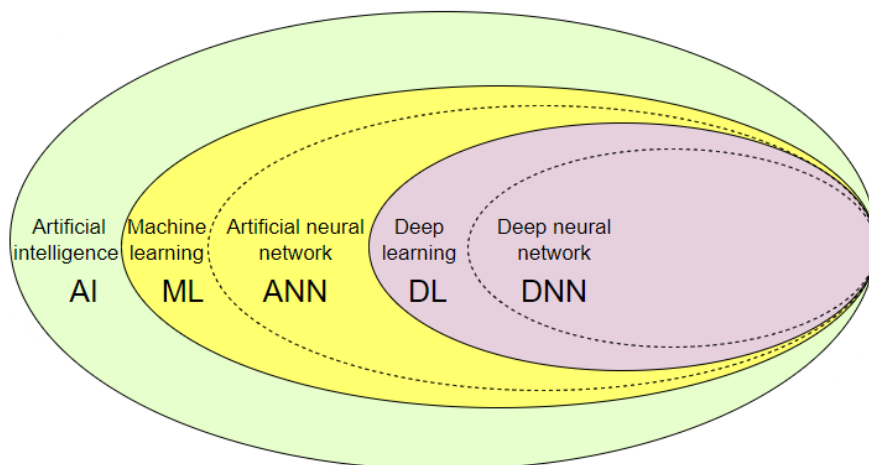


Figure 5. Schematic sketch of clusters of Artificial intelligence

We have seen above that IR techniques of most of the scan geometries (for all seven-generations of CT-scans) can be studied in the conventional Pencil-beam approximated model where the sensor readings in eqn. (1) are $p_i(s, \theta_i)$ for the i -th scan index in a single slice-scan. Then the IR part works out to reconstruct the mass-density function (or the X-ray attenuation function of the body) by computing function f in (1) or (2) analytically (for the AIR) or iteratively (for the IIR). From literature we know that the FBP, which is the conventional and fundamental of the AIR methods, gives better results than the IIR methods when it is used with the necessary filters (also including the AI/DL applications as filters), though IIR enables advanced modifications for greater performance [1,5,7,8,11,20,22,36-38] (See also Fig. 5). However, there is a serious

problem that the superior FBP method (with respect to the IIR techniques) requires application of high-dose X-ray beams to the body under scan for higher image qualities, which violates the ALARA principle [5,7,8,11,36]. This problem and the solution suggested by one of the vendors of AI/DL-based CT-scan systems (GE Healthcare) regarding the effect of the application of AI/DL to the IR process is shown in Fig. 6 [36].

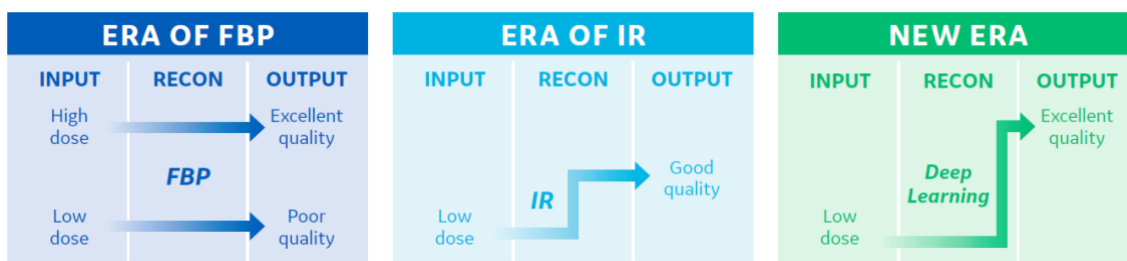


Figure 6. Comparison of the constructed image results of AIR (or FBP) and IIR techniques with the so-called new area, that is, AI/DL-applied IR (image reconstruction) [36]).

Today, we see from literature that success of the AI and DL applied X-ray CT are being tested and reported in various medical applications, such as scans in pediatry, brain-trauma, lung, etc., in the health sciences, i.e., [1,38-54]. Such practical applications involve commercially available in three forms: *high-dose* deep learning (based) image reconstruction (DLIR-H), *medium-dose* deep learning (based) image reconstruction (DLIR-M), and *low-dose* deep learning (based) image reconstruction (LDLIR-H) modules commercially available by two vendors (to our knowledge), namely, TrueFidelity by GE Healthcare and AiCE by Canon Medical Systems [1,35-54]. Mechanism of AI/DL applied X-ray CT given in [36,39] relies on using FBP-reconstructed high-quality images (obtained by high-dose X-ray application) as *ground truth images* in supervised learning. Here, ground truth images are the images of our training dataset used in the related DNN designed [1]. Some of the clinical results of AI/DL-based X-ray CT work with apparently promising results are given in Table 2 where Image qualities obtained after the reconstruction techniques under study (AI/DL-applied or other classical IRs) are evaluated and compared with each other by either qualitatively, or quantitatively, such as in [38,41,46]. In these papers, as being standard routine, qualitative evaluations and comparisons are made by the experienced physicians contrary to the conventional quantitative work in the field of image processing where image parameters such as Signal to noise ratio (SNR) and Contrast to noise ratio (CNR) are evaluated and compared [1,2,11].

Table 2. Dimensional analyses of some AI/DL based clinical CT applications by GE Healthcare’s TrueFidelity™.

Ref. #	Organ, CT type, experimental (phantom, patient, etc.)	Feature and/or aim of the work	Properties and outcomes
[38]	Head CT in trauma	Image optimization	DLIR outperformed ASIR-V both qualitatively and quantitatively
[39]	Pediatric, head CT	-Image optimization: <i>improve image quality</i> -Detection: <i>lesion detection</i>	DL-H reduces noise and improves quality of the images
[40]	Chest CT (with aorta, lung tissue, subscapularis muscle, liver, and vertebrae), 50 patients but 48 selected (22 male, 26 female)	Image optimization: <i>Comparison of SNR and CNR values obtained by DLIR-M, DLIR-H, and ASIR-V 50% by ANOVA (or the Friedman) test</i>	-All three IRs do not change significantly for $p < 0.001$ but differ in soft tissue for $p > 0.05$. -Order of best noise reduction (by increasing CNR and CNR) without

[41]	Abdominal CT, 50 patients (62%F; 56.74±17.05 years)	<p>Image optimization: <i>Comparison of image quality of IRs obtained by ASIR-V at 40%, DLIR-L, DLIR-M, and DLIR-H qualitatively and quantitatively (200 datasets, each IR technique being 50 data).</i></p> <p>-Qualitative analyses: by a point-based Likert scale adapted from the European guidelines on quality criteria for CT.</p> <p>-Quantitative analyses: by CNR for portal vein and liver via ANOVA (or the Friedman test).</p>	<p>distorting the image texture: DLIR-H>DLIR-M>ASIR-V%</p> <p>→ Qualitatively: -DLIRs had better absolute scores than ASIR-V. -DLIR-H demonstrated the best image quality performance compared to ASIR-V for all metrics. - Figure of merit (FOM) analysis also demonstrated that qualitative improvements were amplified by the DLIR</p> <p>→Quantitatively: - All DLIRs had better qualitative scores than ASIR-V. -Compared to ASIR-V, DLIRs had a lower image noise (2.86 vs 1.40–2.29), better image contrast (2.55 vs 1.41–1.96), finer small structure visibility (2.34 vs 1.51–1.90), and improved image sharpness (2.01 vs 1.60–1.86). -Among DLIRs, DLIR-H had the best scores followed by DLIR-M and DLIR-L. -Mean time to reconstruct images with DLIR was longer than with ASIR-V.</p>
[45]	Phantom study with 7 dose levels: CTDI _{vol} : 15/10/7.5/5/2.5/1/0.5mGy	<p>Image optimization: <i>Comparison of image quality of FBP, ASIR-V50%(AV50), ASIR-V100%(AV100), DLIR-L, DLIR-M, and DLIR-H via noise-power spectrum (NPS) and task-based transfer function (TTF)</i></p>	<p>The new TrueFidelity™ deep learning image reconstruction algorithm reduced noise magnitude and improved spatial resolution and detectability without changing noise texture relative to FBP.</p>

6. Conclusion and Discussion

In this work, we have studied a general review of X-ray CT focusing on scanning technique (or scan geometry) and IR techniques along with their relationship with the ALARA principle pedagogically. We see that, regarding the scan geometry, there are seven generations of CT-scans so far and two-main types of image reconstruction techniques: i) Analytical image reconstruction (AIR) and ii) Iterative image reconstruction (IIR). Here we have shortly reviewed the main principles about IR and CT-scanning techniques with the application of AI/DL via the considerations of the ALARA principle. We see that clinical results show promising results for the application of AI/DL in X-ray CT applications. We hope that this work becomes useful as a quick and brief reference for the scholars and researches in the field of AI/DL and X-ray CT.

References

- [1] Arndt, C. et al. (2021). Deep Learning CT Image Reconstruction in Clinical Practice. *Fortschr Röntgenstr*, (193), 252–261.
- [2] Booiij, R., et al. (2020). Technological developments of X-ray computed tomography over half a century: User's influence on protocol optimization. *European Journal of Radiology*, (131), 109261.
- [3] Bueno et al. (2018). Development of a New Cone-Beam Computed Tomography Software for Endodontic Diagnosis. *Brazilian Dental Journal*, 29(6). 517-529. <http://dx.doi.org/10.1590/0103-6440201802455>.
- [4] Liu, Y. (2018). Ch. 5: Research Status and Prospect for CT Imaging of book: "State of the Art in Nanobioimaging". IntechOpen, 73-93. <https://www.intechopen.com/chapters/58660>.
- [5] Padole, A., Khawaja, R., D., A., Kalra, M., K., Singh, S. (2015). CT Radiation Dose and Iterative Reconstruction Techniques, Residents' Section-Structured Review, *AJR*, (204), W384–W392.
- [6] Samei, E. & Pelc, N. (2020). *Computed Tomography Approaches, Applications, and Operations: Approaches, Applications, and Operations*. Springer Nature, 2020, Switzerland. Doi: 10.1007/978-3-030-26957-9 (e-book).
- [7] Xing, R. (2020). Deep Learning Based CT Image Reconstruction, MSc Thesis, University of Washington.

- [8] Sun J, Li H, Wang B, Li J, Li M, Zhou Z, Peng Y. (2021). Application of a deep learning image reconstruction (DLIR) algorithm in head CT imaging for children to improve image quality and lesion detection. *BMC Med Imaging*, 21(1), 108. doi: 10.1186/s12880-021-00637-w. PMID: 34238229; PMCID: PMC8268450.
- [9] Zhong J, Wang L, Shen H, Li J, Lu W, Shi X, Xing Y, Hu Y, Ge X, Ding D, Yan F, Du L, Yao W, Zhang H. (2023). Improving lesion conspicuity in abdominal dual-energy CT with deep learning image reconstruction: a prospective study with five readers. *Eur Radiol.*, doi: 10.1007/s00330-023-09556-6. Epub ahead of print. PMID: 36976337.
- [10] Kalender, W. A. (2006). X-ray computed tomography, *Phys. Med. Biol.*, 51, R29–R43.
- [11] Lee T, Seeram E. (2020). The use of artificial intelligence in computed tomography image reconstruction: A systematic review. *Radiol Open J.*, 4(2): 30-38. doi: 10.17140/ROJ-4-129.
- [12] Buzug, T., M. (2008). *Computed Tomography: from Photon Statistics to Modern Cone-Beam CT*, Springer, Berlin, Heidelberg.
- [13] Kak, C., A., Slany, M., *Principles of computerized tomographic imaging*, IEEE press, 1999.
- [14] Michael, G. (2001). X-ray computed tomography, *Phys. Educ.*, 442-451.
- [15] Romans, L., E. (2019). *Computed tomography for technologists: A comprehensive text*, 2nd ed., Wolters Kluwer.
- [16] Venkatesh, E., Elluru, S., V. (2017). Cone beam computed tomography: basics and applications in dentistry, *J Istanbul Univ Fac Dent.*, 51 (3 Suppl 1), S102-S121.
- [17] Cunningham, I. A., Philip, F. J.. (2000). *Computed Tomography*, CRC Press LLC.
- [18] Ginat, D. T., Gupta, R. (2014). Advances in Computed Tomography Imaging Technology, *Annu. Rev. Biomed. Eng.*, 16, 431–453.
- [19] Goldman, L., W. (2007). Principles of CT and CT Technology, *J Nucl Med Technol.*, 35, 115–128. DOI: 10.2967/jnmt.107.042978.
- [20] Lell, M. M, Wildberger, J. E., Alkadhi, H., Damilakis, J., Kachelriess, M. (2015). Evolution in computed tomography: the battle for speed and dose. *Investigative Radiology*, 50(9), 629-644. DOI: <https://doi.org/10.1097/RLI.0000000000000172>.
- [21] Akagi, M., Nakamura, Y., et al. 2019. Deep learning reconstruction improves image quality of abdominal ultra-high-resolution CT. *Eur Radiol*, 29(11), 6163–6171. (Erratum in: *Eur Radiol*. 2019 May 27, PMID: 30976831).
- [22] Hsieh, J., Nett, B., Yu, Z. et al. (2013). Recent Advances in CT Image Reconstruction. *Curr Radiol Rep*, 1, 39–51. <https://doi.org/10.1007/s40134-012-0003-7>
- [23] Assili, S. (17 Sep 2018). A Review of Tomographic Reconstruction Techniques for Computed Tomography, arXiv:1808.09172v2 [physics.med-ph], 1-5.
- [24] Geyer, L., L., Schoepf, U., J., et al. (2015). State of the Art: Iterative CT Reconstruction Techniques, *Radiology*. 276(2), 339-357.
- [25] Ji, D., Qu, G., Liu, B. (2016). Simultaneous algebraic reconstruction technique based on guided image filtering, *Optics Express*, 24(14), 15897-15911.
- [26] Willemink, M. J., Noel, P. B. (2019). The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence, *Computed Tomography, European Radiology*, 29, 2185–2195. <https://doi.org/10.1007/s00330-018-5810-7>.
- [27] Rubin, G. D. (2014). Computed Tomography: Revolutionizing the Practice of Medicine for 40 Years, *Radiology*, 273(2 Suppl.), (Suppl), 545-74. doi: 10.1148/radiol.14141356. PMID: 25340438.
- [28] Kaczmarz, S. (1937). Angenäherte auflösung von systemen linearer Gleichungen, *Bull. Int. Acad. Sci. Pologne*, 35, 355–357.
- [29] Yusoff, M. S. M., Sulaiman R., Shafinah, K.. (2012). Image Reconstruction for CT Scanner by Using Filtered Back projection Approach, *Journal of American Science*; 8(6), 797-803.
- [30] Jorgensen, J. S. (2013). Sparse Image Reconstruction in computed tomography, Technical University of Denmark PHD-2013 No. 293.
- [31] Kharfi, F. (2013). Ch. 4: Mathematics and Physics of Computed Tomography (CT): Demonstrations and Practical Examples of book: "Imaging and Radioanalytical Techniques in Interdisciplinary Research", IntechOpen, 81-106.
- [32] Onur, T., Ö. (2021). An application of filtered back projection method for computed tomography images, *International Review of Applied Sciences and Engineering*, 12, (2), 194-200.
- [33] Siahaan, F. M. (2008). Computed Tomography (CT) image reconstruction using Matlab programming, Master thesis in University of Indonesia.
- [34] Golsher, J., G. (1977). Iterative Three-Dimensional Image Reconstruction from Tomographic Projections. *Computer Graphics and Image Processing*, 6, 513-537.
- [35] Do S, Song KD, Chung JW. (2020). Basics of Deep Learning: A Radiologist's Guide to Understanding Published Radiology Articles on Deep Learning. *Korean J Radiol.*, 21(1), 33-41. <https://doi.org/10.3348/kjr.2019.0312>
- [36] Hsieh, J., et al. (2019). A new era of image reconstruction: True Fidelity, Technical white paper on deep learning image reconstruction. www.gehealthcare.com.
- [37] Kim, Y., et al. (2015). Ultra-Low-Dose CT of the Thorax Using Iterative Reconstruction: Evaluation of Image Quality and Radiation Dose Reduction. *AJR*, 204, 1197-1202.

- [38] Alagic, Z., Diaz Cardenas, J., Halldorsson, K. et al. (2022). Deep learning versus iterative image reconstruction algorithm for head CT in trauma. *Emerg Radiol*, 29, 339–352. <https://doi.org/10.1007/s10140-021-02012-2>
- [39] Dominik, C., B., et al. (2020). Validation of deep-learning image reconstruction for coronary computed tomography angiography: Impact on noise, image quality and diagnostic accuracy, 14(5), 444–451.
- [40] Jiang, J.-M., Miao, L., Liang, X., Liu, Z.-H., Zhang, L., Li, M. (2022). The Value of Deep Learning Image Reconstruction in Improving the Quality of Low-Dose Chest CT Images. *Diagnostics*, 12(10), 2560. <https://doi.org/10.3390/diagnostics12102560>.
- [41] Parakh, A., Cao, J., Pierce, T., T. et al. (2021). Sinogram-based deep learning image reconstruction technique in abdominal CT: image quality considerations. *Eur Radiol*, 31, 8342–8353. <https://doi.org/10.1007/s00330-021-07952-4>
- [42] Bie, Y., Yang, S., Li, X., Zhao, K., Zhang, C., Zhong, H. (2022). Impact of deep learning-based image reconstruction on image quality compared with adaptive statistical iterative reconstruction-Veo in renal and adrenal computed tomography. *J Xray Sci Technol.*, 30(3), 409–418. doi: 10.3233/XST-211105. PMID: 35124575; PMCID: PMC9108564.
- [43] Kim, J., H., Yoon, H., J., Lee, E., Kim, I., Cha, Y., K., Bak, S., H. (2021). Validation of Deep-Learning Image Reconstruction for Low-Dose Chest Computed Tomography Scan: Emphasis on Image Quality and Noise. *Korean J Radiol.*, 22(1), 131–138. doi: 10.3348/kjr.2020.0116. Epub 2020 Jul 27. PMID: 32729277; PMCID: PMC7772377.
- [44] Tian, Q., Li, X., Li, J., Cheng, Y., Niu, X., Zhu, S., Xu, W., Guo, J. (2022). Image quality improvement in low-dose chest CT with deep learning image reconstruction. *J Appl Clin Med Phys.*, 23(12), e13796. doi: 10.1002/acm2.13796. Epub 2022 Oct 9. PMID: 36210060; PMCID: PMC9797160.
- [45] Greffier, J., Hamard, A., Pereira, F., Barrau, C., Pasquier, H., Beregi, J., P., Frandon, J. (2020). Image quality and dose reduction opportunity of deep learning image reconstruction algorithm for CT: a phantom study. *Eur Radiol.*, 30(7), 3951–3959. doi: 10.1007/s00330-020-06724-w. Epub 2020 Feb 25. PMID: 3210009.
- [46] Jiang, C., Jin, D., Liu, Z. et al. (2022). Deep learning image reconstruction algorithm for carotid dual-energy computed tomography angiography: evaluation of image quality and diagnostic performance. *Insights Imaging*, 13, 182. <https://doi.org/10.1186/s13244-022-01308-2>.
- [47] Noda, Y., Iritani, Y., Kawai, N. et al. (2021). Deep learning image reconstruction for pancreatic low-dose computed tomography: comparison with hybrid iterative reconstruction. *Abdom Radiol*, 46, 4238–4244. <https://doi.org/10.1007/s00261-021-03111-x>.
- [48] Sun, J., Li, H., Wang, B., Li, J., Li, M., Zhou, Z., Peng, Y. (2021). Application of a deep learning image reconstruction (DLIR) algorithm in head CT imaging for children to improve image quality and lesion detection. *BMC Med Imaging*. 21(1), 108. doi: 10.1186/s12880-021-00637-w. PMID: 34238229; PMCID: PMC8268450.
- [49] Sun, J., Li, H., Li, H., Li, M., Gao, Y., Zhou, Z., Peng, Y. (2022). Application of deep learning image reconstruction algorithm to improve image quality in CT angiography of children with Takayasu arteritis. *J Xray Sci Technol*, 30(1), 177–184. doi: 10.3233/XST-211033. PMID: 34806646.
- [50] Yoo, Y., J., Choi, I., Y., Yeom, S., K., Cha, S., H., Jung, Y., Han, H. J., Shim, E. (2022). Evaluation of Abdominal CT Obtained Using a Deep Learning-Based Image Reconstruction Engine Compared with CT Using Adaptive Statistical Iterative Reconstruction. *J Belg Soc Radiol*, 106(1), 15. doi: 10.5334/jbsr.2638. PMID: 35480337; PMCID: PMC8992765.
- [51] Li, Y., Jiang, Y., Yu, X., Ren, B., Wang, C., Chen, S., Ma, D., Su, D., Liu, H., Ren, X., Yang, X., Gao, J., Wu, Y. (2022). Deep-learning image reconstruction for image quality evaluation and accurate bone mineral density measurement on quantitative CT: A phantom-patient study. *Front Endocrinol (Lausanne)*. 13, 884306. doi: 10.3389/fendo.2022.884306. PMID: 36034436; PMCID: PMC9403270.
- [52] Nakamura, Y., Higaki, T., et al. (2019). Deep Learning-based CT Image Reconstruction: Initial Evaluation Targeting Hypovascular Hepatic Metastases. *Radiology: Artificial Intelligence*, 1, e180011.
- [53] Higaki, T., Nakamura, Y., Zhou, J., et al. (2020). Deep Learning Reconstruction at CT: Phantom Study of the Image Characteristics. *Academic Radiology*, 27, 82–87.
- [54] Narita, K., Nakamura, Y., Higaki, T., et al. (2020). Deep learning reconstruction of drip-infusion cholangiography acquired with ultra-high-resolution computed tomography. *Abdom Radiol.*, 45, 2698–2704. <http://link.springer.com/10.1007/s00261-020-02508-4>