Data Security Strategies in Digital Health Services: A Bibliometric Analysis

Natália Proença[1], Igor Polezi Munhoz[2(✉)], Alessandra Cristina Santos Akkari Munhoz[3], and Luciana Pereira[1]

[1] Federal University of ABC, São Paulo, Brazil
[2] Federal Institute of São Paulo (IFSP), São Paulo, Brazil
igor.munhoz@ifsp.edu.br
[3] Sul-rio-Grandense Federal Institute (IFSul), Rio Grande do Sul, Brazil

Abstract. An increasing occurrence of facts that threaten public health, such as the COVID-19 pandemic, is driving the adoption of digital technologies, generating a large volume of data, and making the management, protection, and confidentiality of this information challenging. This paper aimed to carry out a bibliometric analysis to map data security strategies and tools in telehealth services and digital health. To this end, an exploratory study was conducted by collecting and analyzing secondary data from the Web of Science. This paper contributes to digital health providers by pointing out vulnerabilities, potentials, and trends in privacy and security of health care systems.

Keywords: Privacy · Data Security · Digital Health · Digital Service

1 Introduction

The COVID-19 pandemic has altered our society [1]. One of the biggest impacts of this crisis has been on the health care system, which, in some cases, has catalyzed innovations in service delivery. An example of this was the expansion of digital health services, more specifically telehealth [2].

Digital health is a concept that has been evolving since 2000, through internet media with an initial focus on improving medical content, connectivity, and commerce. With the expansion of the concept, new scopes were added, such as genomics, artificial intelligence, and mobile applications, also expanding the application in diagnosis, treatment, clinical decision-making process, and medical care [3].

The term telehealth refers to the entire spectrum of activities used to provide care at a distance - without direct physical contact with the user [4]. Despite all the advantages offered by a telehealth service, it is not without risks. One of the most sensitive issues in relation to the telehealth service is related to the protection of the user's confidential data, respecting the right to privacy and security as established by the General Law for the Protection of Personal Data (LGPD-Law no. 13,709 of August 14th, 2018).

Therefore, the same connectivity that makes it possible to offer the telehealth service also creates threats to users. Protecting health information and providing health services

L. Pereira et al. (Eds.): IDEAS 2022, DSI, pp. 34–43, 2023.
https://doi.org/10.1007/978-3-031-29129-6_4

remotely can generate some complications. A question to be answered in this study is what telehealth services need to do to meet the requirements of the LGPD in such a way as to guarantee the security and, therefore, the adequate protection of user data? To provide a safe digital health service, specifically telehealth, it is necessary to establish best practices. For this, based on a qualitative approach, a bibliometric analysis was carried out to map data security strategies and tools in telehealth services. In addition, considering the definition of bibliometrics as the study of the quantitative aspects of the production, dissemination, and use of recorded information [5], this study provides for a bibliometric analysis to assess, in a predetermined time and event, what is being produced in relation to the present research proposal.

The contribution of this paper lies in understanding which are the most stringent security standards that healthcare organizations must follow to minimize the risks to the privacy and security of these systems, in order not to affect people's trust in telehealth services. The content of this paper is organized as follows. In Sect. 2 we presented the methodology. Section 3 explained the results, Sect. 4 we presented the main conclusions.

2 Methodology

To identify the state of the art of research on telehealth data security, a bibliometric analysis, and a construction of "Word Cloud" were performed. The literature review combined the goals of accuracy, reliability, clarity, and brevity to allow the researcher to make an efficient analysis of the state of the art [6].

2.1 Bibliometric Analysis

For the bibliometric study, a visualization structure based on five parameters was used: *Who* (authors), *What* (keywords), *Where* (countries), *When* (years of publication), and *With whom* (authors' affiliation). Furthermore, the searches were performed on the same platforms as those used for the systematic review, given the results obtained from this one. Bibliometrix software was used for scientific mapping, a tool in an R programming environment that allows the conversion, analysis, and elaboration of graphs through data exported from academic databases. Different analyzes were conducted, as well as Alfred Lotka's productivity law and Bradford's law [7]. First, it was necessary to define the research protocol, dividing it into five activities, namely: search strategy, database query, document management, standardization, and document selection.

Thus, in order to develop such a protocol, as a search strategy, the well-known Pagliosa et al. [8] method were used in this study.

2.1.1 Search Strategy

This activity aimed to establish the research topics, conducting searches through keywords. To verify the suitability of the search command, an adherence test was performed to validate the keywords. Thus, five articles were selected that had the highest number of citations and whose objectives were aligned with the research topic. In this way, the keywords of the articles were compared with the keywords used in the search command in

order to verify if there was a need to improve the search in the search field. As there was no need to improve the search in the search field, the following keywords were obtained: Privacy, Data Security, Digital Health and Digital Service, highlighted in green in the table.

2.1.2 Database Query

It is essential to emphasize that originally, two databases were selected for the development of this research, Web of Science (WoS) and CAPES journals base. However, despite the CAPES periodical database having great recognition and validation, it constitutes a database of great relevance to the local scientific community, from the geographic region where the researchers are located. In this way, for the purposes of greater recognition and familiarity by the world scientific community, only the results obtained through the WoS database were prioritized.

Through a Boolean search, the selected keywords were combined, and the relationships identified between them. The date of the initial search was recorded, November 9,2021 and, first, 82 results were found in the WoS database, using the adopted keywords. It is important to note that, in this situation, the results had not yet been refined in terms of the period of publication year, free access or language used. Then, the period of analysis of publications was established for the last five years. Thus, the total number of selected documents increased to 58 in the WoS database. In addition, other filters were applied, such as for the language (English), obtaining 56 results in a WoS database. Still at this stage, it was possible to eliminate academic results with restricted access, obtaining 34 open access results in WoS.

2.1.3 Document Management

This activity aimed to organize the documents so that the filtering and analysis processes could be carried out. Within the WoS site itself, after performing the keyword search, it was possible to extract reports containing the main information of the 34 articles obtained from the WoS database. These reports contain information such as publication name, authors, authors' emails and addresses, pages, keywords, publisher, ISSN, eISSN, DOI, place of publication, language, year and month of publication, patent number, number of references and number of citations. After extracting the report from the database where the searches were performed, with the help of the BibTeX software it was possible to create a specific database. Then, the treatment of the report table generated from the WoS database was performed, to exclude columns without data or repeated, making it possible to obtain a more concise and organized report.

2.1.4 Standardization and Selection of Documents

At this stage, the files were selected by type of result, with only those categorized as "Article/Article" being selected. Therefore, 9 results were eliminated, resulting in 25 articles. The analysis for duplicates was performed using filters and commands in the Microsoft Excel platform, in which all cells from each database were selected, all of which were subjected to "conditional formatting", "cell highlighting rules" and then

"duplicate values". It was possible to conclude, through this specific analysis, that there were duplicates in the generated report, resulting in a report of 23 results from this database. Then, the resulting articles were analyzed under three aspects, being (i) title, (ii) keywords and (iii) abstract. Thus, at this stage, articles could be eliminated if they did not present a research topic related to the investigated topic. The remaining articles will be analyzed in their entirety. Firstly, in terms of (i) title, it was possible to discard from the analysis for this study, those results in which the title certainly did not show the slightest relationship with the theme proposed for this paper, data security strategies in security services. Digital health. Consequently, 3 results were disregarded in the report extracted from the WoS database.

2.1.5 Analysis, Synthesis, and Consolidation of Results

This last phase aimed to analyze, develop a synthesis and consolidate the results. After all this process, it was possible to indicate all the research findings.

2.2 Construction of "Word Cloud" and "TreeMap"

At this stage, the "Word Clouds" technique was used to analyze the collected material. The keywords of the articles obtained through the systematic literature review were used as a basis. This method can be understood as a way of visualizing linguistic data, revealing the frequency with which words appear. The image construction technique consists of using different sizes and fonts of letters that vary according to the occurrence of the words, where the most frequent words usually appear in the center of the image with an increased size, while the others appear around with a reduced size. Thus, this technique contributes to the visualization of what is most relevant in the selected articles.

However, in order to generate the word cloud, it was necessary to use R programming and BibTeX formatting, installing 'RStudio', already mentioned in this article. Once installed, the command 'library(bibliometrix)' and 'biblioshiny()' were plotted, thus generating access to the 'Biblioshiny for bibliometrix' platform.

3 Results and Discussion

3.1 Bibliometric Analysis

First, analysis was performed under the parameter *Who* (authors), which shows the most relevant authors. As a result of this analysis, the authors of the article "Alavi A" were one of the authors of the article "A scalable, secure, and interoperable platform for deep data-driven health management" and "Sheikh A", one of the authors of the article's main highlight. "Health information technology and digital innovation for national health systems and learning care" and the article "Technology and Universal Health Coverage: Examining the role of digital health".

Then, analysis was performed under the parameter *What* (keywords), which shows the most relevant keywords. From this analysis extracted from the Bibliometrix platform, it was revealed that the words "care" and "security" were the most relevant. Such analysis meets the objective and theme of this article, in which data security strategies are placed

under study, inserted in the scope of health care, since such strategies are related to digital health.

Soon after, analysis was performed under the parameter *Where* (countries). With this, it was possible to affirm the responsibility and scientific relevance, on the subject of this study, of countries such as the United States, United Kingdom, India, Australia and China, mainly, and in descending order of relevance.

An analysis was also performed under the parameter *When* (years of publications), under the Bibliometrix platform.

Under such analyzes generated through the Bibliometrix platform, it was possible to perceive that such scientific productions reached peaks of relevance in 2020 and 2021, revealing the importance of such topics in question, in the context of a pandemic and the importance of technology.

Finally, analysis was performed under the parameter *With whom* (affiliations), which shows the most relevant affiliations. In this way, under the generation of such an analysis, it was possible to perceive that Schwarz, Stanford University and Johns Hopkins University, all North American universities, stood out in terms of the theme of this article.

Furthermore, analyzes were performed according to Alfred Lotka's Law of Productivity and Bradford's Law [7].

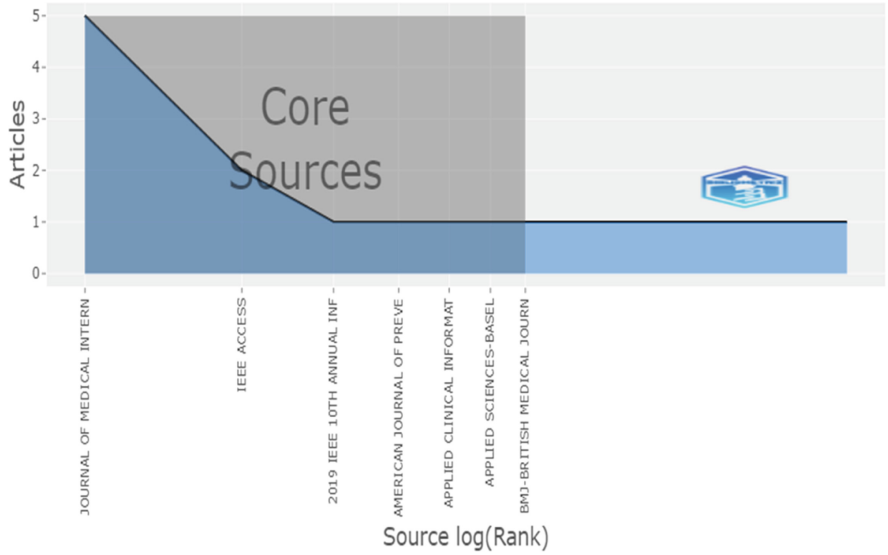

Fig. 1. Analysis generated in the Bibliometrix platform of results extracted from the WoS database, for Bradford's Law.

As Bradford's Law makes it possible to estimate the degree of relevance of journals, it was possible to analyze that (Fig. 1), under such generated images, about the results extracted from the WoS database, the journals "Journal of Medical Inter" and "IEEE Access" were the most relevant.

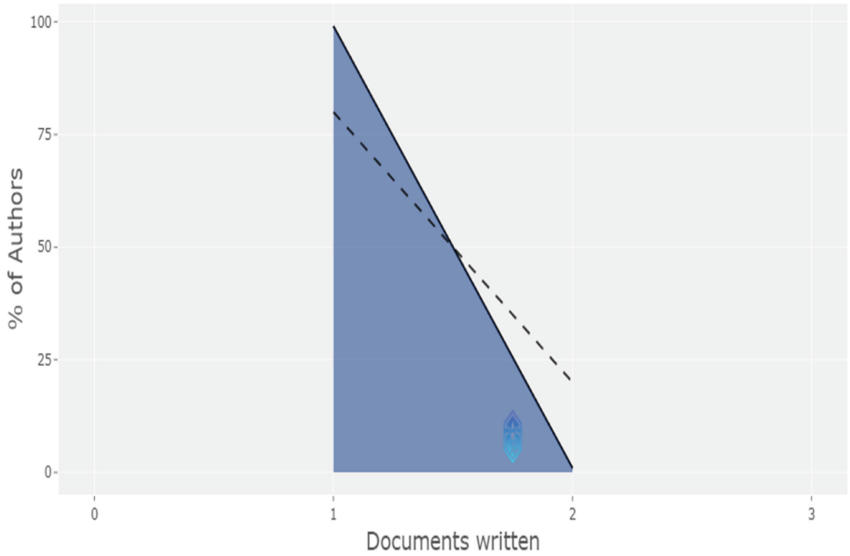

Fig. 2. Analysis generated in the Bibliometrix platform of results extracted from the WoS database, for Lotka's Productivity Law.

To conclude this stage of bibliometric analysis, an analysis of the generated image was performed (Fig. 2). Thus, with such an analysis under the parameter of Lotka's Productivity Law, it was possible to infer the level of productivity of the researchers, and little expression in terms of research and researchers was revealed. This demonstrated that the theme of this study is current, new, and has studies that are still not relevant in terms of its researchers, which, even so, does not discredit the relevance of the theme for society and the scientific community.

3.2 "Word Cloud" and "TreeMap"

Word Cloud (Fig. 3) was generated on the 'Biblioshiny for bibliometrix' platform, containing the 50 most frequent words in the 50 most relevant articles analyzed, from the Web of Science database.

Word clouds were generated, refined to single words (unigrams) as well as to two words (bigrams), generating expression, as in 'data security', for example. The following is the unigram word cloud.

In addition, 'TreeMaps' were also generated (Fig. 4), to complement the analysis in question, by revealing the percentages of each frequency of words. This consists of a hierarchical data visualization technique, using grouped rectangles, structured in a tree, justifying its name.

Fig. 3. Image of the first Word Cloud generated, from BibTeX software, with data extracted from WoS database and refinement to unigram words.

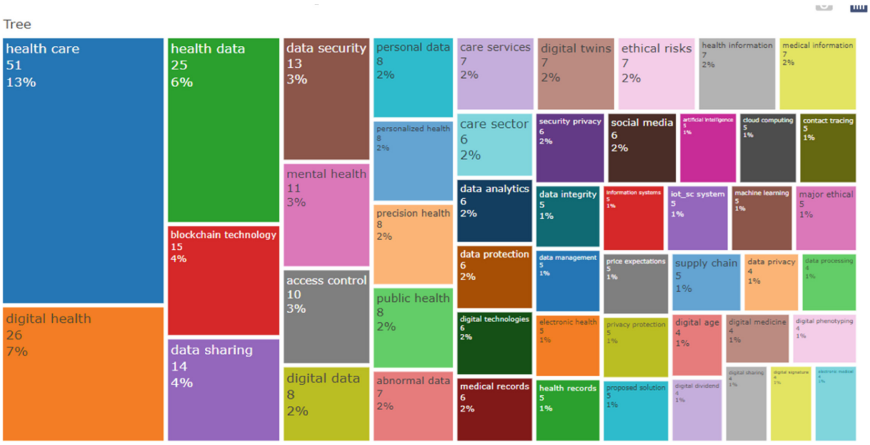

Fig. 4. Image of second 'TreeMap', generated using BibTeX software, with data extracted from WoS database (bigrams).

From Figs. 3 and 4 were possible to provide specific analysis about the results extracted from the database. Such analysis reveals that the most frequent and, therefore, most relevant words, such as 'digital health', 'data security' or 'privacy', express the central and most important themes of the articles. Such techniques thus reveal that the articles analyzed are well aligned with the line of research and the problem in question.

In terms of data privacy and security, these are constantly threatened, and need protection means to generate efficient and reliable service to users in the health sector. Currently, several techniques are suggested capable of mitigating privacy threats as much as possible, such as the master key management approach and the multi-key server approach [9]. However, this study suggests a security subject to the speed of access to data, constituting only an alternative, not a solution to the source of the problem.

Thus, having the notion of respect for user rights in relation to their personal data, such as access, modification and forgetting, it is feasible to make use of technologies

such as Blockchain (BC) [10], in which sensitive data of users would be stored in trusted databases managed by the BC network. In this way, an entire state-of-the-art architecture would be developed, so that users' personal data would never be sent to service providers. Finally, BC technology proves to be promising, in addition to being an excellent option for the purpose of combating the threats of technology to the privacy of individuals, since it is a decentralized technology and eliminates several intermediaries from formal procedures, which, in turn, contribute to the vulnerability of personal data [11]. However, like all technology, it has its limitations, more specifically, in data storage and computing, which are limited and costly, impacting security, privacy, and, consequently, reliability and efficiency.

Maria Stoyanova et al., in their study about vulnerability problems in IoT systems by forensic logic, suggests a methodology for segregating processed data into two categories, which allowed, within the forensic perspective, data and privacy to be preserved [12]. This methodology, although used under a specific scenario by the author, is valid in a context of telehealth, external to the forensic theme.

To complement the analyzes carried out, a "Word Growth" chart (Fig. 5) was generated on the Bibliometrix platform, responsible for revealing the occurrence of the most relevant terms over the years. This graph is represented in the following image.

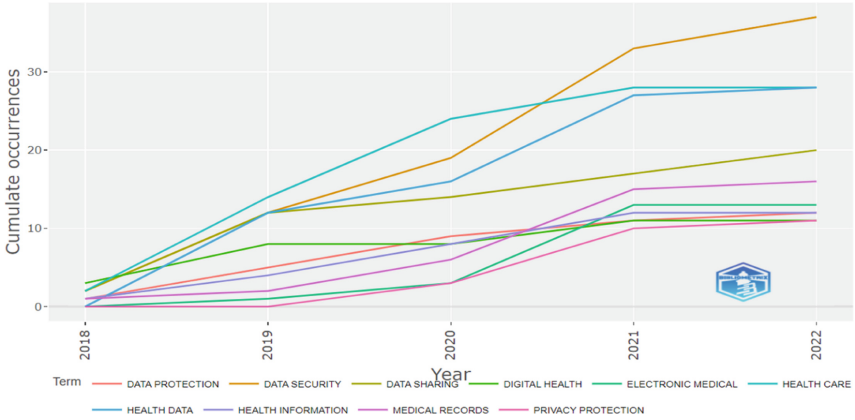

Fig. 5. "Word Growth" chart image, generated using BibTeX software.

Thus, although generated under the aspect of "bigrams", it was extremely important to analyze the considerable increase in the terms "Data Security", "Health Data", "Data Sharing", in addition to the terms "Privacy Protection". This situation demonstrated how the importance of topics such as health data security and privacy, dealt with in this paper, have gained relevance in recent years, especially after the first pandemic outbreak, revealing the degree of relevance of this study.

In "Word Growth" chart (Fig. 5) was possible to perceive that the dates of the articles obtained are summarized in the year 2020 to date, which demonstrates, as already evidenced in certain results in this research, the current relevance of this research topic

for society. This even confirms the current scenario of increasing incidents involving security and privacy of sensitive data.

Comprising the results, 16 scientific articles were obtained as a final amount of research to be analyzed and discussed in its entirety, all with a high probability of high synergy with the theme proposed in this paper.

4 Conclusion

To conclude, it was possible to infer, from bibliometric analysis and literature review, that the theme proposed in this study, besides being current and still having little expression in the scientific community, has a process of natural evolution of its knowledge. And its solutions, consequently generating, as one of its many results, the method of organizing the conceptual framework in order of publication, thus demonstrating the evolution of scientific thinking.

In addition, according to bibliometric analyzes carried out, it was possible to highlight the greatest scientific advance on this subject in the United States, as well as the main affiliations, all from North American universities, such as Stanford University and John's Hopkins University. In addition, other generated analyzes led to conclusions such as the fact that the words "care" and "security" are the most influential and relevant in this study, in addition to the fact that the journal "Journal of Medical Inter" has the greatest relevance, even among the two bases. of analyzed data.

Even so, regarding the developed analyses, it was possible to highlight, in terms of "Word Cloud" and "TreeMaps", the relevance of words such as "da-ta", "privacy", "health", "security", with discrete changes in order of relevance between the two bases. This reveals that the reports used to generate such analyzes are in great harmony with the proposal of this study, "Data Security Strategies in Digital Health".

Furthermore, through the bibliometric analysis of "Word Grouth", a considerable increase in the terms "Data Security", "Health Data", "Data Sharing", in addition to "Privacy Protection" was noticed. This situation demonstrated how the importance of topics such as health data security and privacy, dealt with in this article, have gained relevance in recent years, especially after the first pandemic outbreak, thus revealing the degree of relevance of this study, both for the society and the scientific community.

More strictly, when dealing with the process of analyzing the impasse in question, it was possible to conclude that there is a need for constant improvement of the technologies in use, since their vulnerability is inherent to their existence, with the specific objective of preserving the privacy and security of users of health applications, thus generating consequent reliability on the part of society, whether user or server. However, one should not only prioritize the development of a new cutting-edge, innovative and multifaceted architecture, without a precise guideline. This development, in turn, should be directed towards a process of combining technologies, aiming at and considering the advantages of each technology, in the midst of a context of elaboration and improvement of recent technologies.

Finally, the solution to the problem in question is not simple, nor is the impasse, as the field of technologies involved is recent and under development, as well as the rights of users must be fundamentally respected, concomitantly. Seeking a balance between

respecting the rights of individuals and the development of technology, even if focused on services to the community, is undeniably a challenge that can only be overcome with the mutual effort of all the sectors involved.

Acknowledgement. This paper is partially funded by the EP32280009/3006.0082.03/202–41. The first author is funded by PIBIC/CNPq.

References

1. Wilder-Smith, A, Freedman, D.O. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J. Travel Med. **27**(2) (2020)
2. Silva, A.B., Carneiro, A.C.M.G., Sindico, S.R.F.: Regras do governo brasileiro sobre serviços de telessaúde: revisão integrativa. Planejamento e Políticas públicas, (44) (2015)
3. Mathews, S.C., McShea, M.J., Hanely, C.L., Ravitz, A., Labrique, A.B., Cohen, A.B.: Digital health: a path to validation. NPJ Digit. Med. **2**, 38 (2019)
4. Patel, P.D., et al.: Rapid development of telehealth capabilities within pediatric patient portal infrastructure for COVID-19 care: barriers, solutions, results. J. Am. Med. Inform. Assoc. **27**(7), 1116–1120 (2020)
5. Vanti, N.A.P.: Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. Revista Ciência da Informação **31**(2), 152–162 (2002)
6. Hart, C.: Doing a Literature Review: Releasing the Social Science Research Imagination. Sage (1998)
7. Arsenova, I.: New application of bibliometrics. Procedia – Soc. Behav. Sci. **73**, 678–682 (2013)
8. Pagliosa, M., Tortorella, G., Ferreira, J.C.E.: Industry 4.0 and lean manufacturing: a systematic literature review and future research directions. J. Manufact. Technol. Manag. **32**(3), 543–569 (2019)
9. Perumal, A.M., Nadar, E.R.S.: Architectural framework of a group key management system for enhancing e-healthcare data security. Healthcare Technol. Lett. [S.L.] **7**(1), 13–17, 26 November 2019. Institution of Engineering and Technology (IET)
10. Tomás, R., Bordel, B., Alcarria, R., Sanchez-de-Rivera, D.: Enabling trustworthy personal data protection in eHealth and well-being services through privacy-by-design. Int. J. Distrib. Sens. Netw. [S.L.] **16**(5), 155014772091211, maio (2020)
11. Houtan, B., Hafid, A.S., Makrakis, D.: A survey on blockchain-based self-sovereign patient identity in healthcare. IEEE Access, [S.L.] **8**, 90478–90494 (2020). Institute of Electrical and Electronics Engineers (IEEE)
12. Stoyanova, M., Nikoloudakis, Y., Panagiotakis, S., Pallis, E., Markakis, E.K.: A Survey on the Internet of Things (IoT) Forensics: challenges, approaches, and open issues. IEEE Commun. Surv. Tutorials, [S.L.], **22**(2), 1191–1221. Institute of Electrical and Electronics Engineers (IEEE) (2020)