


Review

# Bibliometrics of Machine Learning Research Using Homomorphic Encryption

Zhigang Chen <sup>1,\*</sup> , Gang Hu <sup>1</sup>, Mengce Zheng <sup>1</sup>, Xinxia Song <sup>2</sup> and Liqun Chen <sup>3</sup>

<sup>1</sup> School of Information and Intelligent Engineering, Zhejiang Wanli University, Ningbo 315100, China; hugang@zwu.edu.cn (G.H.); mczheng@zwu.edu.cn (M.Z.)

<sup>2</sup> School of Junior, Zhejiang Wanli University, Ningbo 315100, China; sxxx@zwu.edu.cn

<sup>3</sup> Department of Computer Science, University of Surrey, Surrey GU2 7XH, UK; liqun.chen@surrey.ac.uk

\* Correspondence: chenzhigang@zwu.edu.cn

**Abstract:** Since the first fully homomorphic encryption scheme was published in 2009, many papers have been published on fully homomorphic encryption and its applications. Machine learning is one of the most interesting applications and has drawn a lot of attention from researchers. To better represent and understand the field of Homomorphic Encryption in Machine Learning (HEML), this paper utilizes automated citation and topic analysis to characterize the HEML research literature over the years and provide the bibliometrics assessments for this burgeoning field. This is conducted by using a bibliometric statistical analysis approach. We make use of web-based literature databases and automated tools to present the development of HEML. This allows us to target several popular topics for in-depth discussion. To achieve these goals, we have chosen the well-established Scopus literature database and analyzed them through keyword counts and Scopus relevance searches. The results show a relative increase in the number of papers published each year that involve both homomorphic cryptography and machine learning. Using text mining of articles titles, we have found that cloud computing is a popular topic in this field, which also includes neural networks, big data, and the Internet of Things. The analysis results show that China, the US, and India have generated almost half of all the research contributions in HEML. The citation statistics, keyword statistics, and topic analyses give us a quick overview of the development of the field, which can be of great help to new researchers. It is also possible to apply our methodology to other research areas, and we see great value in this approach.

**Keywords:** homomorphic encryption; machine learning; privacy; security; bibliometrics



**Citation:** Chen, Z.; Hu, G.; Zheng, M.; Song, X.; Chen, L. Bibliometrics of Machine Learning Research Using Homomorphic Encryption. *Mathematics* **2021**, *9*, 2792. <https://doi.org/10.3390/math9212792>

Academic Editor: Alfonso Mateos Caballero

Received: 24 September 2021

Accepted: 27 October 2021

Published: 3 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Machine learning algorithms based on deep neural networks are gaining increasing attention as a breakthrough in the development of artificial intelligence, which is the mainstream of current artificial intelligence research. These techniques have achieved remarkable results and are widely used for data processing and analysis in areas, such as spam detection, traffic analysis, intrusion detection, medical organization prediction, face recognition, and financial prediction.

Furthermore, with the growing number of cloud services, machine learning services can be run on the facilities of cloud providers, where machine learning models are trained and deployed. In short, this is Machine Learning as a Service (MLaaS), and several such services are offered by many prominent Internet companies, including Microsoft AzureAI [1], Google Prediction API [2], GraphLab [3], and ErsatzLabs [4]. However, machine learning algorithms require access to raw data, which is often privacy-sensitive and poses potential security and privacy risks. In recent years, many scholars have investigated various privacy-preserving methods for sensitive data in different machine learning algorithms, such as logistic regression [5–9] or neural networks [10–13].

Three options for dealing with this type of data privacy security problem are well-known: The first option is federated learning, which trains algorithms across distributed edge devices without exchanging their data samples, maintaining the decentralized and distributed nature of the training data. However, this approach is difficult to check for evaluation and is subject to power, computation and internet fluctuations. The second option is differential privacy, which maximizes the accuracy of data queries from statistical databases while minimizing the chance of identifying their records. However, this option still exposes information about the dataset while retaining information about the individual, and it can lead to reduced accuracy due to the injection of noise. The third option is a fully homomorphic encryption scheme, which is an emerging means of data protection and is the subject of this paper.

Homomorphic encryption (HE) is an encryption system that allows one to perform certain arithmetic operations on encrypted data and to obtain an encrypted result that corresponds to the result of the operation performed in plaintext. The size of the encrypted data and the learning time depend heavily on the number of features, so the performance of large data sets is often less than optimal in terms of storage and computational costs, and existing homomorphic encryption schemes share not only common drawbacks but also their limitations. For example, the CKKS scheme in SEAL [14] and the BGV scheme in HELib [15] are both built on polynomials (ideal lattices) and can be computed by SIMD batching techniques for encrypted “vectors”. The Ttfe [16] scheme, on the other hand, is based on gate-circuit computation, which differs significantly from the previous two in that gate-circuit based homomorphic computation can be implemented for arbitrary forms of arithmetic algorithms.

Our main contributions of this paper are as follows:

- An analysis of the literature of homomorphic encryption in machine learning (HEML) was studied by various researchers.
- An approach to the application of bibliometric statistics to the field of homomorphic encryption.
- Analysis of the application and development of homomorphic encryption in the field of machine learning from a literature perspective.

This paper is organized as follows. Section 2 classifies and compares the literature of HEML and further categorizes which homomorphic encryption schemes have been used in the literature of HEML. We then present in Section 3 the research methodology, the data source and the data extraction process, which we use to prepare the pool of all HEML papers used later for analysis. Section 4 presents the results of the study. Section 5 summarizes the findings and implications. Finally, Section 6 concludes this study and states the future work directions.

## 2. Literature Classification and Comparison

Based on the analysis of the literature, it can be seen that most of the papers related to HEML are on ciphertext prediction with neural networks or machine learning. Only a few of them are related to the training of models for fully homomorphic encryption. For example, the DNN scheme in [13,17], the matrix computation scheme in [12], and the ciphertext space conversion scheme in [18] are representative. The ciphertext conversion algorithm between Ttfe and BGV in [18] is cited in the article [11]. In addition, CryptoNets [13] is also seminal, and many of its efficient methods are still practiced today, making it one of the most cited papers in the field.

The literature summary is given in the following two tables. Table 1 classifies and compares the literature of HEML. Table 2 further categorizes which homomorphic encryption scheme is used in the literature of HEML.

**Table 1.** An overview of some related literature.

Models	Overview
Logistic Regression	<p>Logistic regression is a very common regression model, such as the CKKS scheme in articles [5,7], which achieves good training capability on large data sets and achieves 96.4% accuracy at MNIST. The paper [6] achieves high accuracy and efficiency based on good data packing and the logistic regression algorithm optimized for ciphertext, which further improves the processing capability of large datasets based on the former. Similar to the previous two, the paper [9] implements least-squares approximation of logistic functions to improve accuracy and efficiency (i.e., reduce computational cost), as well as applies new packing and parallelization techniques. The paper [8] proposes a new strategy that combines differential privacy methods and homomorphic encryption to achieve the best of both, achieving a good variance of less than 1%. In addition, we also find that the articles published by related authors in this field are all relatively similar and are based on the CKSS scheme. The paper [19] differs from all the above papers in that it implements a logistic regression prediction process between the cloud and the client based on the semi-honest assumption and the BGV scheme.</p>
DNN, NN	<p>Traditional neural networks, called fully connected layers as the last layer in CNNs, are relatively well implemented and therefore, have a lot of relevant applications. The paper [20] relies mainly on multi-key and coding algorithms (MK-FHE), chunking cryptographic computations before handing them over to the cloud, with high network throughput and latency. The IBM paper [21] makes a considerable contribution by using low polynomial approximation functions in the BGV scheme, averaging pools instead of max pools, and using homomorphic lookup tables, which also take a longer time. The paper [22] uses cryptographic data augmentation for neural network DNN computation and does not use a homomorphic encryption scheme. It is more of an image encoding improvement to achieve privacy protection and high accuracy is achieved on different datasets.</p>
CNN	<p>Convolutional neural networks are now the basis of image-related artificial intelligence, so the research here is numerous and outstanding. Differing from articles [23,24] that provide a polynomial approximation to the ReLU activation function, article [10] uses a hybrid network structure that uses homomorphic encryption for the first propagation and final evaluation computations, with the intermediate process still being conventionally explicit training. Article [25] achieves very high efficiency with the help of batch processing and special data processing packaging. Article [11] uses the BGV scheme of the HELib library and the TFHE scheme to implement complete HEML. The common computations, such as ciphertext multiplication and addition are performed in the BGV scheme, while numerical comparisons (activation functions, such as ReLU) are computed in TFHE by a special ciphertext conversion algorithm. The paper [12] implements the matrix computation of homomorphic ciphertext vectors by a special matrix encoding method, and the paper [13] proposes a scheme of the average pool instead of the maximum pool, polynomial approximation of activation function, and parallel encoding of data.</p>

Table 1. Cont.

Models	Overview
Improved CNN	<p>These models are based on DNNs and CNNs with different degrees of improvement. The article [26] builds a bridge between homomorphic encryption and common machine learning libraries and improves processing power with SIMD batching. Combined with TensorFlow, it has good performance in <math>2^{13}</math> and <math>2^{14}</math> dimensions. The article [17] proposes two solutions to solve the high latency problem of HEML. The first is a special information representation (LoLa) after encoding pixel points, which can significantly reduce latency and memory usage and the second is a deep neural network through transfer learning. Although the accuracy is slightly lower than other models of the same class, the latency is far better than other models, with only single-digit latency, with the LoLa-Small model having a latency of only 0.29. The paper [27] implements cryptographic computation on arbitrary neural networks through the TFHE and TLWE schemes. The core lies in the ciphertext conversion of the two schemes, and the data encoding method. The performance on the ciphertext face dataset is good, and the total time for one update is less than 0.2 s. The paper [28] uses the NTL library to build the FHE scheme and implements face template matching on facenet, which is more efficient because of the shallow ciphertext depth. The paper [29] puts the computation process that can be linearized into homomorphic encryption, and the computation that cannot be linearized and has high overhead is still performed locally. In the paper [18], TFHE and CKKS are ciphertexts transformed with the BGV scheme, and finally, a fully homomorphic encrypted neural network is trained. All computations that cannot be linearized can be performed by binary circuits because of the properties of TFHE. The accuracy of the encrypted ResNet-34 is even higher than that of the classical ResNet-34.</p>

Table 2. Status of homomorphic encryption scheme used.

Schemes	Articles with High Relevance
BGV	[11,18,21,23]
CKKS	[5,6,12,13,17,18,24–26,29]
TFHE	[11,25,27,29]
Other HE Schemes	[10,20,22,28,30]

### 3. Research Methodology and Data Extraction

In this section, we present selection and retrieval methods in the face of multiple literature databases.

#### 3.1. Literature Databases

The goal of this study is to perform a bibliometric analysis of papers related to HEML, with a focus on citations and topics, to better describe and understand the research literature in the field from the researcher's perspective. Based on the above objectives, we aim to answer the following research questions:

1. How many papers related to Homomorphic Encryption Machine Learning (HEML) have been published each year in recent years?
2. What is the citation status of such papers?
3. What are the topics of HEML papers?
4. What is the relationship between machine learning and homomorphic encryption?

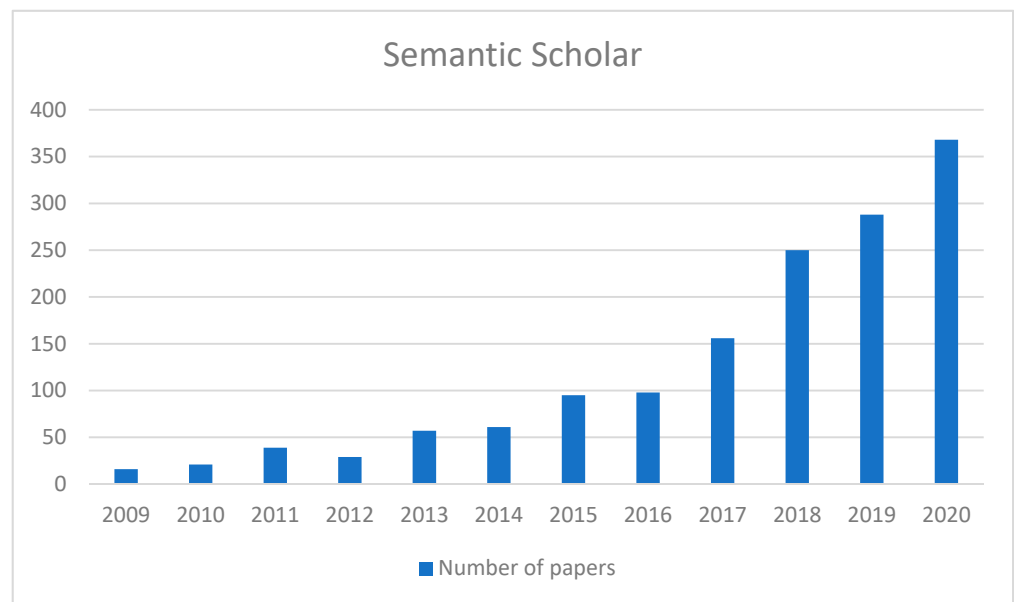
Several important and authoritative literature databases have been identified: Science Direct ([www.sciencedirect.com](http://www.sciencedirect.com), assessed on 5 March 2021), Scopus ([www.scopus.com](http://www.scopus.com), assessed on 18 April 2021), Web of Science ([www.webofknowledge.com](http://www.webofknowledge.com), assessed on 5 March 2021), Google Scholar ([scholar.google.com](http://scholar.google.com), assessed on 4 March 2021), and Semantic Scholar ([www.semanticscholar.org](http://www.semanticscholar.org), assessed on 4 March 2021). These databases are the most common databases used by researchers in various bibliometric studies. We have

analyzed and comprehensively evaluated four of these bibliographic databases through a related study [31], as shown in Table 3.

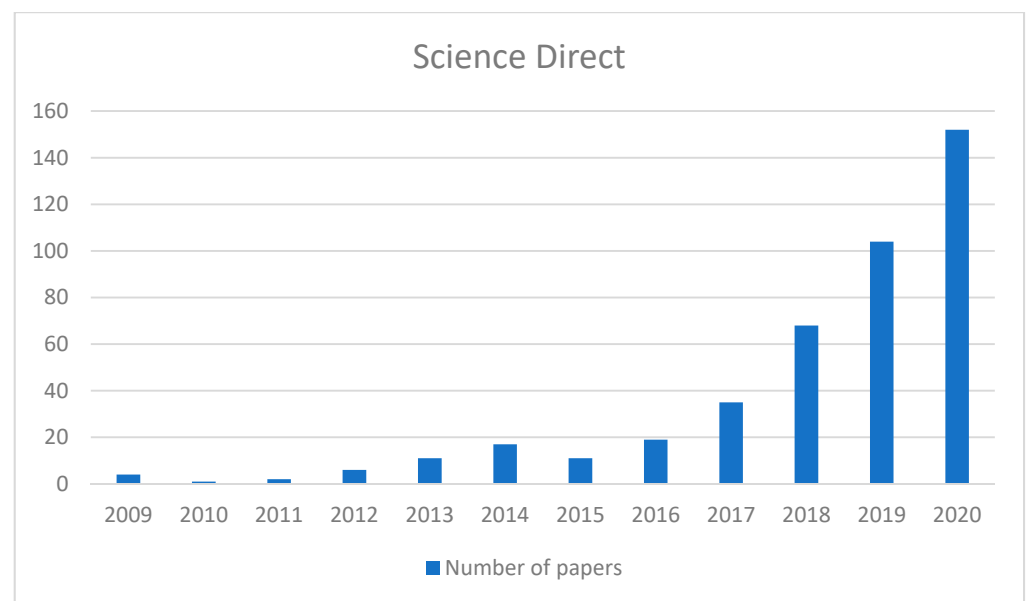
**Table 3.** Literature databases.

Criteria	Literature Databases			
	Scopus	Web of Science	Google Scholar	Semantic Scholar
Quality and reliability	Thanks to Scopus’ feature of searching by “source name” (location name), full coverage quality and reliability of search results can be achieved to a large extent.	Given the nature of homomorphic encryption papers, the quality and reliability of the search results cannot be guaranteed within the full coverage.	Based on Google’s powerful engine, we can search a wide range of literature, but the sources are complicated and duplicated, and the quality is not guaranteed.	There is a good range of literature collection in homomorphic encryption, and the statistics of citation data are clear.
Citation data	Yes	Yes	Yes	Yes
Search interface and output export	Allow saving all exported papers to CSV files	Only allow saving up to 500 results at a time	No	No

Although the above four literature databases have their shortcomings, we still counted the literature on “HEML” since 2009 in two of them manually for the reliability of the statistical results. Among them, Google Scholar cannot be counted by years in the form of search engine results, and hence it is omitted. The number of search results with the keyword HEML is about 14,300 and the data can be summarized in Figures 1 and 2, respectively.



**Figure 1.** Statistics of the number of HEML papers by Semantic Scholar over the years.



**Figure 2.** Statistics of the number of HEML papers by Science Direct over the years.

Many bibliographic databases cannot easily export the list of retrieved papers to a file (unless complex scripts are written) or cannot find any applicable APIs; for example, using Google Scholar to manually analyze a large number of papers would be very time-consuming. Even Web of Science only allows saving the list of retrieved papers to CSV file page by page or select part of it. If the paper search results return 100 pages of papers, we need to go to the export page by page. Only Scopus allows saving the list of all the papers searched to a CSV file. In addition, Scopus has the function to search by “source name” (location name). Thus, using Scopus, it is possible to cover the field of HEML to a large extent, while ensuring the quality and reliability of the search results.

Because fully homomorphic encryption was only developed in 2009, its applications should be considered valid after this period. Moreover, the keyword “HEML” is not fully reflected in many literature titles and keywords, so we need to supplement it with similar words, such as “Neural Network”, “Deep Learning”, “Machine Learning”, and “Bioinformatics”, etc. More details are discussed further in the next section.

### 3.2. Extracting HEML Papers from Scopus

After selecting Scopus as the bibliographic database to search for HEML papers, the next step was to search for these papers. First, we tried to search with “Homomorphic Encryption” as the keyword, and we obtained 4734 results. It is easy to find that Scopus does show all the papers with the search keywords in the title, abstract, and keywords, and several important papers on homomorphic encryption are among them. So, we can also be sure that the search results of Scopus can be used as our dataset.

A simple “HE” search result is not what we want, and we found many practical tips and tricks after a thorough exploration of the search method. For example, when searching in Scopus, the keyword “Homomorphic Encryption” is used in the “TITLE-ABS-KEY” with machine learning related terms (“Neural Network”, “Machine Learning”, etc.) It is an effective way to ensure coverage, but it is also important to ensure that papers that only discuss homomorphic encryption schemes are excluded, as this is a good way to ensure coverage. However, it is also necessary to ensure that papers discussing only homomorphic encryption schemes are excluded, since such papers mostly discuss new homomorphic encryption schemes or optimization methods for homomorphic encryption, etc. This is not an easy task, and we will present the research process one by one.

After the initial summary, we carried on this search and performed several reviews to exclude some irrelevant literature (e.g., Bootstrapping for Approximate Homomorphic



Encryption). It should be noted that the data extraction phase of this study was conducted in March 2021 and the Scopus database engine also needed time to record or import all literature data from other sources. Therefore, the data for 2021 is incomplete. In addition, citations to 2021 papers are relatively low because they are either “in the press” or recently published. For example, our analysis shows that only six of the 57 papers published in 2021 have been cited (more than once), this number is low compared to 106 of the 260 papers published in 2020. Therefore, we will not collect and analyze papers published in 2021 but will use 2020 as the last year of publication.

The initial dataset of 1461 papers available was initially derived from the search results. The results returned by Scopus include all major computer and information-related publications, including top cryptography conferences and journals, such as IEEE Trans, ACM, etc. In addition, we know that some papers may be related only to “Encryption” or “Computing”. For example, the source publication “IEEE Transactions on Parallel and Distributed Systems” is related to parallel systems. Some search results are included with the keyword “Homomorphic Property”, which are mostly irrelevant to the topic of this paper, and we excluded them.

Most of the time, the boundary between homomorphic encryption and machine learning is usually “gray”. Therefore, for this study, we have to draw the line somewhere. In addition, we focus more on scenarios of encrypted data processing or scenarios related to the design phase of HEML, including for example, “Model Training”, “Cryptonets”, etc. In summary, we further optimized the search keywords.

Our dataset has been further refined to include literature related to “HEML” as much as possible. In addition, we also found some papers that are not related to HEML, such as “Computing arbitrary functions of encrypted data”, “Security and Privacy for Cloud-Based IoT: Challenges”, etc. We found that these irrelevant papers are basically brought up by the index keyword “cloud computing”, but unfortunately, we did not find an effective means to exclude all these documents at once during the search. The reason for this is that HEML itself is deeply rooted in HE and is more or less bundled with HE or with other HE-related applications. Therefore, it can only manually exclude such literature. From the dataset, we observed that the total number of citations of this kind of non-HEML related literature was mostly high and ranked higher, so we can safely infer that most of the literature is related to HEML, but the rarity of this kind of literature makes it less frequently cited by researchers. Moreover, it is not difficult to find a few excellent HEML papers with very high citation totals, e.g., [5,13,32], which we will discuss in-depth in Section 4.4.

Scopus collects at least nine types of documents (resources): articles, books, book chapters, conference papers, conference reviews, editorials, notes, reviews, and short surveys. We aim to only include papers related to computer science, so we only include the following types of records: conference papers, journal articles, book chapters, and review articles (Conference Paper, Article, Book Chapter, Review), while the rest are excluded. As mentioned earlier, this work was conducted in March 2021, and we only analyzed literature from 2020 and earlier. At this point, we have further compressed the dataset and obtained optimizing results.

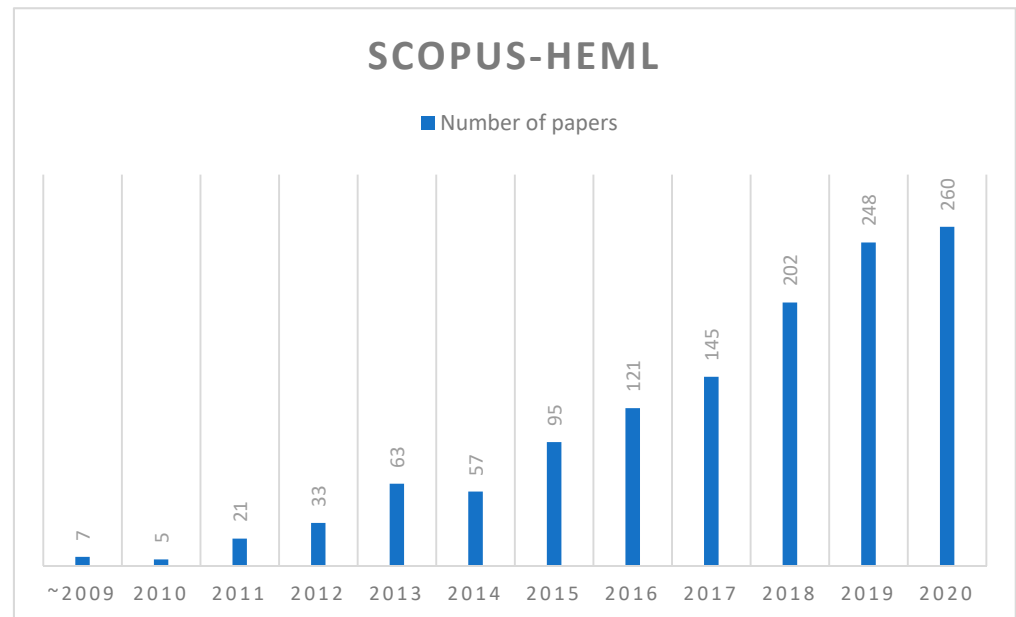
With perfect constraints and optimization, we obtained the final HEML literature data and checked the exported records to ensure their completeness. For example, there were no duplicate records of the literature. Since some of the data exported from Scopus were duplicates, we cleaned the dataset and applied all the above steps, resulting in a total of 1262 papers before the year 2020, and 1257 papers after removing the five papers before 2005 to form our final HEML literature database. To ensure the transparency and portability of our analysis and to enable other researchers to perform other types of analyses, all the raw data of the literature are available as Excel files, which can be downloaded online [33].

#### 4. Analysis and Discussion

In this section, we will discuss in detail the citations of the literature.

#### 4.1. Literature Databases

Regarding the growth of the literature related to HEML, Figure 3 shows the number of HEML-related papers in Scopus over the years. Starting from 2005, only a single-digit number of papers were included in Scopus during 2005–2009. From 2010 onwards, the number of papers in Scopus has gradually accelerated, with an increase in the number of papers each year, reaching a peak of 260 papers last year (2020). It can be confirmed that HEML has been formally developed in the last few years, and its development has been accelerated every year.



**Figure 3.** Statistics of papers over the years.

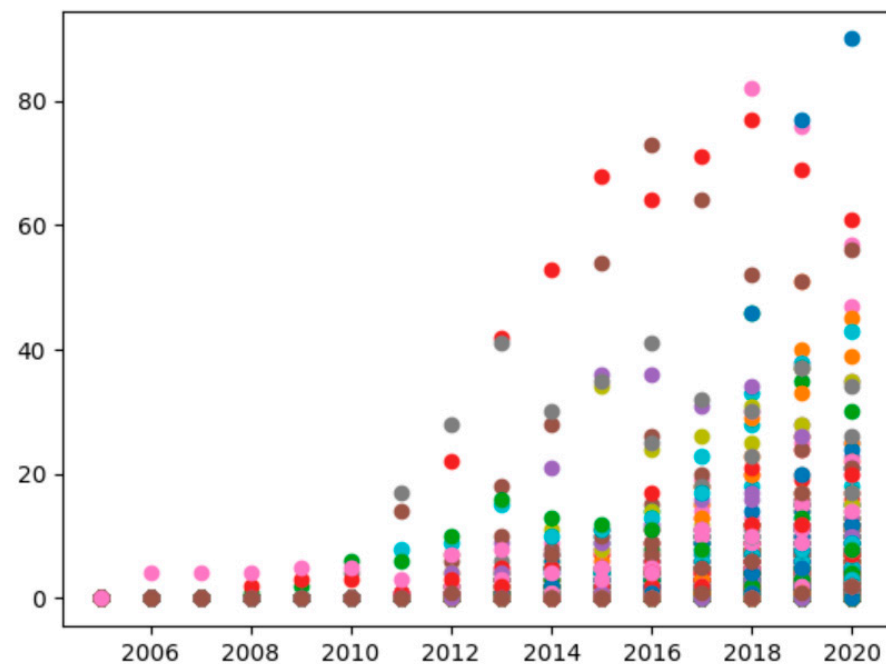
#### 4.2. Citation Analysis

##### 4.2.1. The Overall Situation of Citations

Citations are critical to the positioning of any research and the positioning of other work. Highly cited papers are generally considered to represent their impact. Based on the dataset extracted from Scopus, we present in Figure 4, the HEML citation profile, as a scatter plot of the number of citations for all papers versus year of publication. There are 10,352 points on the graph. Each color point represents a paper and shows the citation profile per year on the graph. We list the papers with high citation rates corresponding to those color points as below. Both the generation method and the data are publicly available in [33].

- Blue: Security and Privacy for Cloud-Based IoT: Challenges
- Pink: Multi-key privacy-preserving deep learning in cloud computing
- Red: Can homomorphic encryption be practical?
- Brown: On-the-fly multiparty computation on the cloud via multikey fully homomorphic encryption
- Grey: Computing arbitrary functions of encrypted data
- Green: Privacy preserving error resilient DNA searching through oblivious automata





**Figure 4.** Citation statistics of papers over the years.

In terms of year of publication, the majority of papers were published in more recent years. For example, between 2005 and 2020, nearly 85% of the papers were published in the last 5 years (2015–2020), while the remaining 15% were published earlier. This indicates that the volume of HEML papers is experiencing tremendous growth over the years. Besides, we observe that a few papers are highly cited, such as “Can homomorphic encryption be practical?” [32], which is important for the application area of homomorphic encryption. In addition, a large number of papers can be found clustered at the bottom, which means that they have never been cited.

Among the 1257 HEML papers retrieved in the Scopus database, 455 papers (about 36%) have never been cited (zero citations), while 176 papers (about 14%) have been cited only once. In total, 802 papers (64% of the total) were cited two times or more, and the total number of citations was 10,352. Therefore, the average number of citations per paper is about eight. The HEML paper with the highest number of citations [32] was 527 (to be discussed in detail in Section 4.2.2).

Considering the special nature of HEML, which itself is in between the two fields of homomorphic encryption and machine learning, many citations also come from these two fields and probably do not use the related papers previously published in this field, which is probably also a phenomenon at the beginning of the birth of a new field.

#### 4.2.2. Highly Cited Papers

To analyze these most cited papers, we used two metrics: the absolute number of citations to the paper from the year of publication to 2020 and the average annual number of citations to a given paper. This measure normalizing the effect of the year of publication (age) on the total number of citations has been used in many bibliometric studies. The top 10 “most influential” articles using each of these two methods, ranging from 2009 to 2020, can be found in Tables 4 and 5.

**Table 4.** Top 10 papers cited in total.

#	Paper Title	Publishing Year	Number of Citations
1	Can homomorphic encryption be practical? [32]	2011	542
2	Computing arbitrary functions of encrypted data [34]	2010	332
3	Multi-key privacy-preserving deep learning in cloud computing [20]	2017	222
4	Privacy-preserving ridge regression on hundreds of millions of records [35]	2013	174
5	Privacy-Preserving Deep Learning via Additively Homomorphic Encryption [36]	2018	168
6	ML confidential: Machine learning on encrypted data [37]	2013	152
7	Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy [13]	2016	123
8	Toward secure multi-keyword top-k retrieval over encrypted cloud data [38]	2013	122
9	Privacy-preserving outsourced classification in cloud computing [39]	2018	114
10	GAZELLE: A low latency framework for secure neural network inference [40]	2018	103

**Table 5.** Top 10 papers with average annual citations since publication.

#	Paper Title	Publishing Year	Average Number of Citations	Total Number of Citations
1	Privacy-Preserving Deep Learning via Additively Homomorphic Encryption [36]	2018	56	168
2	Multi-key privacy-preserving deep learning in cloud computing [20]	2017	55.5	222
3	Can homomorphic encryption be practical? [32]	2011	54.2	542
4	Privacy-preserving outsourced classification in cloud computing [39]	2018	40.3	121
5	GAZELLE: A low latency framework for secure neural network inference [40]	2018	34.3	103
6	Cloud-Based Approximate Constrained Shortest Distance Queries over Encrypted Graphs with Privacy Protection [41]	2018	27.3	82
7	Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy [13]	2016	24.6	123
8	Chameleon: A hybrid secure computation framework for machine learning applications [42]	2018	22.3	67
9	Privacy-preserving ridge regression on hundreds of millions of records [35]	2013	21.8	174
10	ML confidential: Machine learning on encrypted data [37]	2013	18.5	152

We found eight duplicates in the above two Top 10 citation count tables. The most representative paper “Can homomorphic encryption be practical?” [32] appears unsurprisingly at the top of the two tables, which is indeed one of the earliest papers on HEML. It is easy to see that the highly cited papers are the most representative and constructive in the field of homomorphic encryption and are widely cited by scholars. It can be seen that the progress in the last 5 years is much better than the previous years. Many papers published in the last few years have already received high citations. The contributions presented in these papers will likely be developed further in the future. We speculate that this situation is closely related to the development of big data and artificial intelligence applications on the Internet, both of which are closely related to data privacy and security issues. Table 5 also highlights the fact that homomorphic cryptographic machine learning has entered

a rapid development phase since 2017, with a considerable number of high-quality papers being published, related to the fields of “cloud computing”, “neural networks” and “deep learning”.

It is well-known that the identification and classification of highly cited papers are regularly reported in various scientific fields, such as biology, medicine, ecology, and social sciences. For example, the once prestigious October 2014 issue of Nature reported the top 100 papers in all scientific fields under the cover of the “Top 100 papers” [43]. The study reported that of the 58 million articles in Thomson Reuter’s Web of Science, only 14,499 papers were cited more than 1000 times. The top three papers identified in the article [43] were cited 305,148, 213,005, and 15,530 times, respectively, all of which were in “biological laboratory technology”.

#### 4.2.3. Statistics on the Number and Citations of Different Publication Types

As described in Section 3.2, Scopus stores at least nine types of documents (resources) in the domain (HEML) database: journal articles, books, book chapters, conference papers, conference reviews, commentaries, notes, review articles, and short surveys. We wanted to include only papers in the scientific field, so we included only the following four types of records: conference papers, journal articles, book chapters, and review articles (Conference Paper, Article, Book Chapter, Review), excluding other types of records.

Six statistics for different types of documents are calculated, as shown in Table 6, and as a percentage of the number of the four main document types in Figure 5. In terms of the percentage of papers, conference papers and journal articles had the highest percentage, 58.0%, and 38.8%, respectively. In terms of the average number of citations per document type, conference papers (e.g., reports of cryptographic academic conferences) and journal articles have an average of 8.5 and 10.0, respectively. Surprisingly, review articles and book chapters also have a 50% citation rate, with the major contributions coming from ACM and IEEE. There are also about 142 never-cited papers from journal articles and 236 from conference papers.

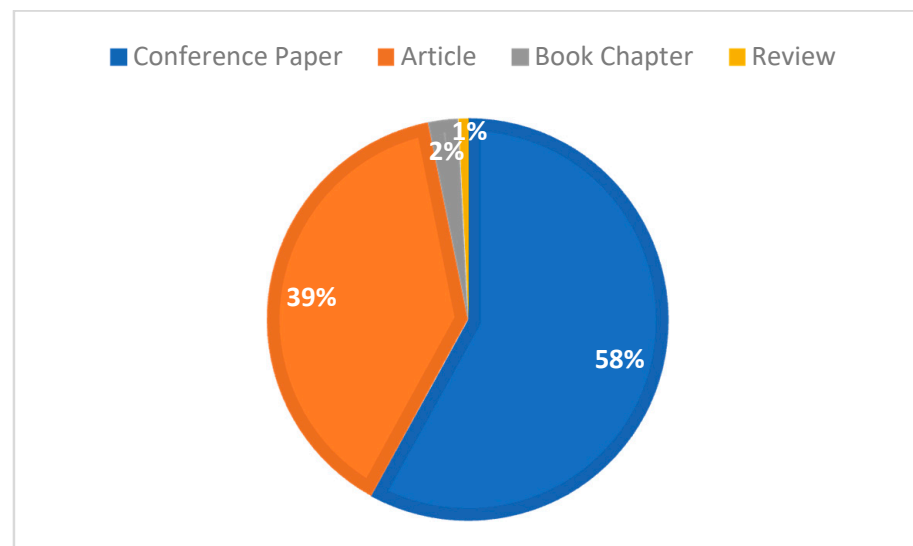


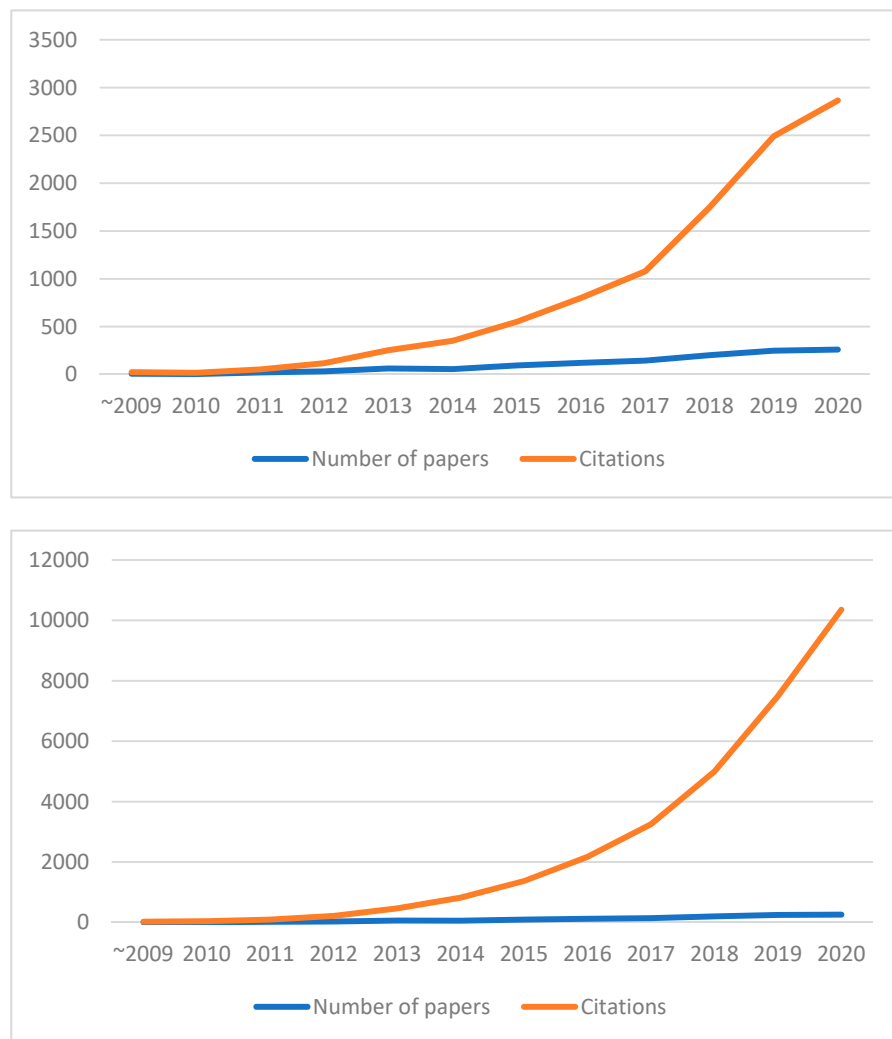
Figure 5. Proportion of the four main document types.

**Table 6.** Statistics based on the number of citations of different document types.

Data Statistics	Publication Types			
	Conference Paper	Article	Book Chapter	Review
Total number of articles	729	488	30	10
Percentage%	58.0	38.8	2.4	0.8
Total number of citations	6202	4912	59	14
Average number of citations	8.5	10.0	3.9	1.4
Number of never cited	236	142	15	5
At least one citation %	67.6% (493)	70.9% (346)	50% (15)	50% (5)

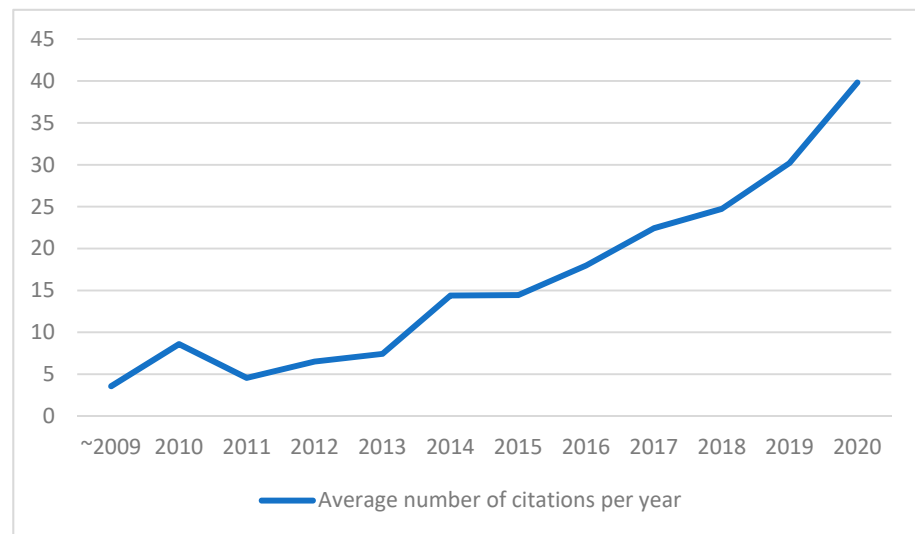
4.2.4. Annual Citation Analysis

Figure 6 shows the number of papers and citations by years, in which both annual values and cumulative values are shown. Figure 6 also shows that 2017 began as a high growth phase for the field, although there was a slight lull since 2019 because new papers need more time to reach sufficient exposure.



**Figure 6.** Annual number of papers and citations (**above:** each year, **bottom:** cumulative annual).

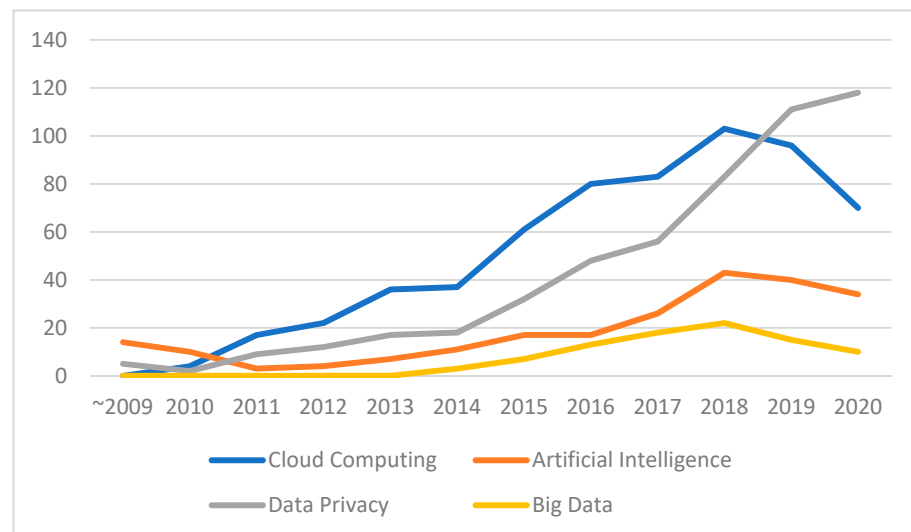
Next, we want to see how the number of citations differs across years of publication. Figure 7 shows the average citation trend for papers published in different years, which is the result of dividing the values in the lower panel of Figure 6.



**Figure 7.** Average number of citations per year.

#### 4.2.5. Number of Papers and Citations in the HEML Subfield

Our dataset [33] is comprehensive and supports other types of analyses, in addition to the ones that were discussed above. This allows us to make another analysis, in which we grouped papers in different HEML subfields. Our purpose is to distinguish four representative HEML subfields, namely: “Cloud Computing”, “Artificial Intelligence (Machine Learning)”, “Data Privacy”, and “Big Data”. To do this, we search for these subfield names as keywords in the papers. Our result of the publication status is shown in Figure 8. It is important to note that this simple textual analysis is limited and does not include phrases with similar meanings to a topic.



**Figure 8.** Annual number of papers published in the subfield.

#### 4.3. Focused Keywords

HEML is based on the combination of homomorphic encryption schemes and related practices in machine learning. The work of the HEML researchers can be with different research focuses, so we want to catch this feature. Just as with analyzing the four subfields discussed in the previous section, this time we pooled the keywords of all papers in our dataset and used WordArt (wordart.com, accessed on 24 September 2021) to generate a







**Table 9.** LDA Topic Modeling in NLP.

#	Theme Model
1	0.013*“ml” + 0.012*“data” + 0.010*“service” + 0.009*“security” + 0.009*“solution” + 0.008*“privacy”
2	0.042*“data” + 0.020*“cloud” + 0.017*“encryption” + 0.015*“image” + 0.015*“homomorphic” + 0.013*“algorithm”
3	0.038*“learning” + 0.029*“model” + 0.024*“network” + 0.019*“privacy” + 0.019*“neural” + 0.017*“machine”
4	0.038*“cloud” + 0.028*“data” + 0.026*“computing” + 0.017*“security” + 0.011*“paper” + 0.009*“encryption”
5	0.041*“scheme” + 0.039*“homomorphic” + 0.032*“encryption” + 0.027*“fhe” + 0.021*“fully” + 0.014*“implementation”
6	0.017*“information” + 0.009*“security” + 0.007*“encryption” + 0.007*“content” + 0.005*“codeword” + 0.005*“homomorphic”
7	0.022*“scheme” + 0.020*“homomorphic” + 0.013*“image” + 0.013*“encryption” + 0.011*“secure” + 0.009*“algorithm”
8	0.032*“voting” + 0.022*“electronic” + 0.013*“scheme” + 0.012*“homomorphic” + 0.009*“encryption” + 0.008*“proposed”
9	0.011*“fhe” + 0.009*“computation” + 0.009*“problem” + 0.008*“tree” + 0.008*“private” + 0.008*“character”
10	0.050*“data” + 0.049*“cloud” + 0.025*“encryption” + 0.021*“user” + 0.018*“security” + 0.018*“computing”
11	0.031*“data” + 0.024*“encryption” + 0.021*“homomorphic” + 0.020*“cloud” + 0.017*“scheme” + 0.013*“computing”
12	0.053*“data” + 0.013*“encrypted” + 0.012*“privacy” + 0.012*“cloud” + 0.010*“encryption” + 0.009*“homomorphic”
13	0.030*“image” + 0.018*“encryption” + 0.014*“privacy” + 0.011*“homomorphic” + 0.010*“cloud” + 0.010*“proposed”
14	0.012*“crm” + 0.007*“device” + 0.006*“city” + 0.005*“stealthy” + 0.005*“edge” + 0.005*“deeper”
15	0.048*“data” + 0.022*“cloud” + 0.014*“encrypted” + 0.014*“scheme” + 0.014*“search” + 0.013*“encryption”
16	0.021*“data” + 0.010*“cloud” + 0.009*“log” + 0.008*“solution” + 0.007*“privacy” + 0.007*“algorithm”
17	0.022*“encryption” + 0.014*“data” + 0.012*“homomorphic” + 0.010*“privacy” + 0.009*“scheme” + 0.008*“network”
18	0.032*“learning” + 0.029*“privacy” + 0.026*“data” + 0.019*“model” + 0.013*“machine” + 0.008*“federated”
19	0.023*“data” + 0.019*“computation” + 0.016*“query” + 0.015*“encrypted” + 0.013*“cloud” + 0.012*“scheme”
20	0.035*“computation” + 0.026*“protocol” + 0.017*“secure” + 0.017*“scheme” + 0.016*“cloud” + 0.014*“client”

At this point, we believe readers have got a good idea of the topics and research directions in this field. It is easy to see that these topics are a combination of homomorphic encryption and machine learning, with the general purpose of protecting data security and solving privacy problems.

#### 4.5. In-Depth Discussion

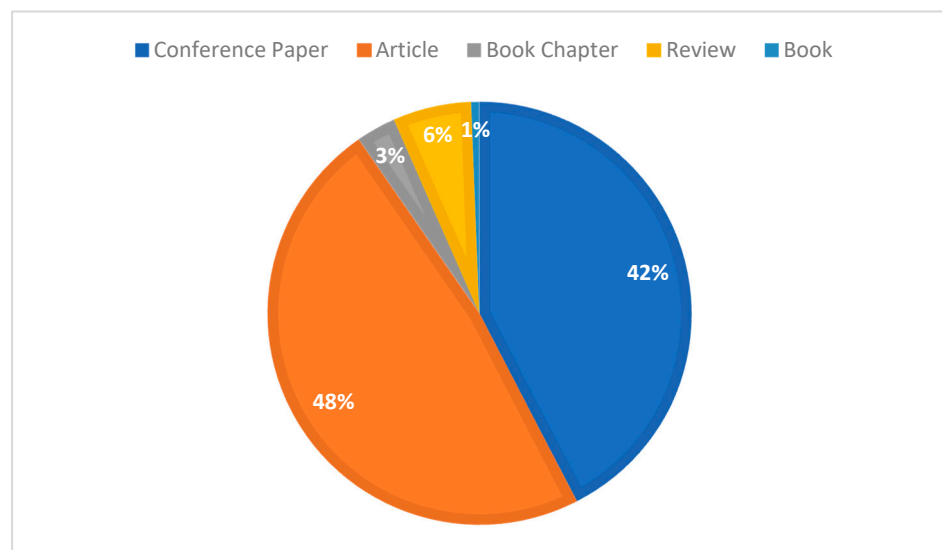
During our research, we discovered a very meaningful search in Scopus, namely the “related literature” at the bottom of each search. We made use of this feature for two well-known articles, “Logistic Regression Model Training based on the Approximate Homomorphic Encryption” [5] and “Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy” [13] to discuss the in-depth and strong relationship to the topic.

First, looking at the literature related to [5], Scopus gives us a considerable result. It is not difficult to analyze that these are of some relevance, such as titles, keywords, abstracts, etc. There is much more than the papers in the field of homomorphic encryption, and the results are listed in Table 10.

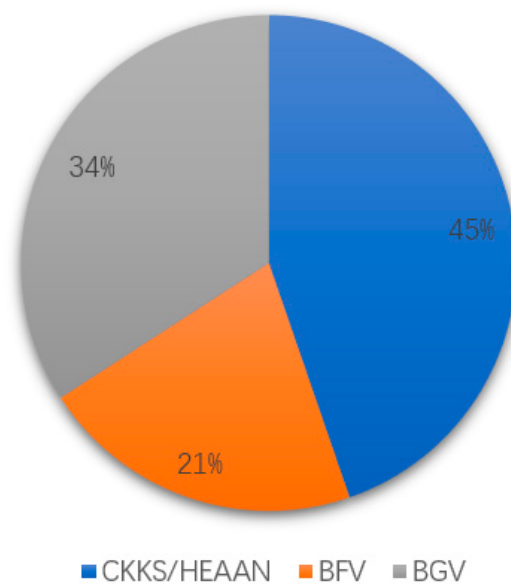
From Table 10, we notice that the literature related to logistic regression or homomorphic encryption is at the top of the list, although the number of citations is not large, this does not affect the connection between them. Based on the type of literature, it is similar to what we have discussed above, and the related literature is mainly concentrated in journal articles and conference papers, see Figure 11. Based on this, we would like to further understand the proportion of the three mainstream homomorphic encryption schemes, and our statistical analysis results are shown in Figure 12.

**Table 10.** Related literature on homomorphic encryption logistic regression.

#	Paper Title	Publishing Year	Number of Citations
1	Secure and Differentially Private Logistic Regression for Horizontally Distributed Data [8]	2020	4
2	Logistic regression on homomorphically encrypted data at scale [7]	2019	3
3	Privacy-Preserving Classification of Personal Data with Fully Homomorphic Encryption: An Application to High-Quality Ionospheric Data Prediction [19]	2020	0
4	Secure logistic regression based on homomorphic encryption: Design and evaluation [9]	2018	42
5	Secure outsourced matrix computation and application to neural networks [12]	2018	42



**Figure 11.** Proportion of types of literature data sets related to [5].



**Figure 12.** Proportion of homomorphic encryption schemes used in different documents.

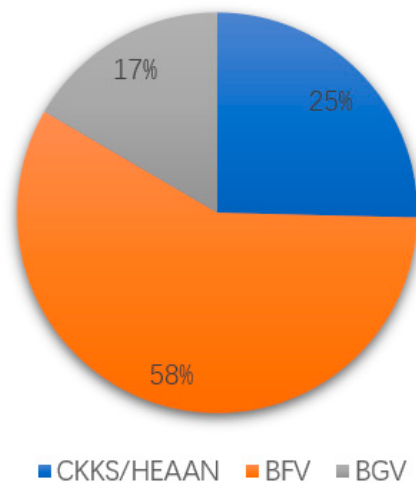
We believe that the use of the CKKS scheme is more predominant in this area and is also consistent with the relevance of the article [5]. Although the search of literature databases is not always reliable and a considerable number of papers does not appear, the

main reason for which we speculate is that some of the relevant literature does not state the homomorphic scheme they use in the abstract or keywords. We then looked at the literature related to the article [13], whose related literature is shown in Table 11.

**Table 11.** Cryptonets related literature.

#	Paper Title	Publishing Year	Number of Citations
1	Application of homomorphic encryption on neural network in the prediction of acute lymphoid Leukemia [46]	2020	0
2	Toward practical homomorphic evaluation of block ciphers using prince [47]	2014	23
3	Depth optimized efficient homomorphic sorting [48]	2015	18
4	PPolyNets: Achieving High Prediction Accuracy and Efficiency with Parametric Polynomial Activations [49]	2018	4
5	A general design method of constructing fully homomorphic encryption with ciphertext matrix [50]	2019	0

As it can be seen from Table 11, the citations are similar to the literature related to the article [5], with a low number of citations, but they all show similar characteristics. The types of literature are mainly focused on conference papers and journal articles. However, the number of related literature is also less than half, which can be inferred that the development of logistic regression is earlier and more comprehensive compared with neural networks. The distribution of the literature types of article [13] and article [5] are relatively similar, so instead of showing their distribution here, we can look at the percentage of homomorphic encryption schemes, as shown in Figure 13.

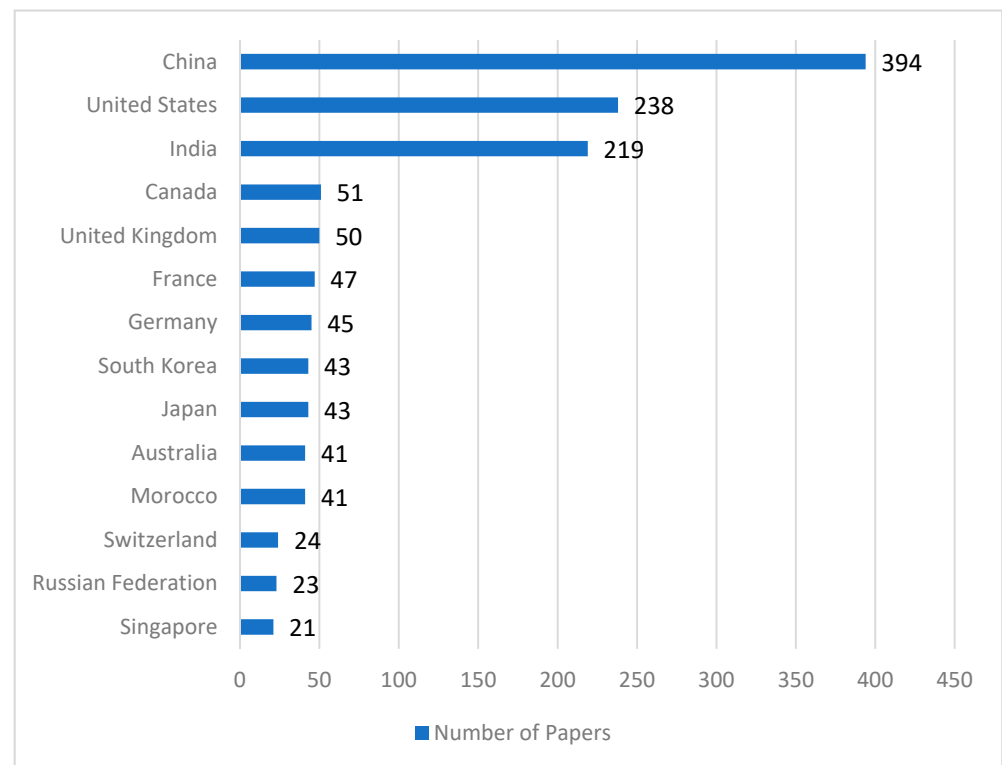


**Figure 13.** Proportion of homomorphic encryption schemes used in different documents.

More literature related to article [13], chose the BFV scheme, perhaps due to the guiding effect of Microsoft SEAL [14] that led to this phenomenon. Through the above analysis, it has been proved that using the relevant indexes of Scopus can help us find the corresponding papers or materials conveniently and efficiently. We hope that our method can enlighten more researchers, particularly any new researchers.

#### 4.6. Number of Contributed Papers by Country

After each search query, Scopus provides the country/region information of the authors of a paper. This information has also been included in our database. Figure 14 shows the count ranking of researchers by country. The top three countries (China, the USA, and India) account for a large part of the ranking.



**Figure 14.** Ranking of contributions by different countries.

## 5. Further Discussion

In this section, we summarize our findings and discuss the development of the field.

### 5.1. Research Findings, Trends, and Insights

Our study shows that the number of papers on HEML has been gradually increasing since 2009. During 2005–2015, less than 100 HEML papers were published each year. However, after 2015, the number of publications per year has increased significantly, reaching between 200 and 500 in recent years. This fact clearly shows that it is desirable to perform an automated bibliometric analysis, as we have conducted in this paper. Otherwise, it is impossible to develop an objective and comprehensive view of our scientific community and the topics under study.

The citation analysis reveals that a significant number of papers in HEML (about 36% of the dataset) have not been cited at all. This may indicate that the work is of low quality in terms of research or reporting. Another explanation is that the academic community has little interest in these papers. The “publish or perish” culture of scholarship encourages scholars to despise papers that are of low or lesser quality, such as quantity over quality.

Throughout the HEML citation landscape, the bulk of its development is dependent on the development of homomorphic encryption schemes. For example, if we do not limit the scope, the most cited papers, which we discussed in Section 4.2.2, are not in HEML but in FHE, e.g., “Fully Homomorphic Encryption Using Ideal Lattices”. The citation count itself is a bit problematic, and it cannot be ruled out that some researchers in some countries may need to swipe a large number of citations for some reason.

An analysis of paper types shows that about two-thirds of the papers in our dataset are published in conferences, while only one-third of the papers are published in journals. This is typical in computer science, but unlike in the natural sciences, where most papers are published in journals. When comparing citations, we find that conference papers also have the highest number of citations, but that the average number of citations in journal articles is higher. Perhaps this is because the cryptography development community has always been more closely aligned.

Then, the keyword analysis shows that the HEML field involves only a few topics: cloud computing, neural networks, data privacy, bioinformatics, etc. They are closely related to the current hot topic “artificial intelligence”, which is the same as our expectation.

Finally, the analysis of countries shows that the top three countries (China, USA, and India) account for almost half of all papers published. If we divide these data by the population size of the country, it is easy to assume that some Western European countries and small Asian countries may have higher scores. Most of these countries are economically developed regions with relatively high access and quality of education.

### 5.2. Potential Limitations and Effectiveness

In Section 3.1, we discussed several major bibliographic databases. We chose Scopus, because it is the most convenient database for exporting and analyzing data. However, this does not necessarily mean that the final dataset that we obtained from Scopus is the most comprehensive and complete. We carefully studied the search terms and indexing methods to ensure that the process would be reproducible by readers. During our research, we found that using a combination of the keywords “Homomorphic Encryption” and machine learning related terms was an effective way to ensure coverage, but we also need to ensure that papers related to homomorphic encryption schemes only were excluded. We cursorily reviewed the obtained literature dataset, which meets our expectations and has no duplication. In addition, the screening of keywords may not directly represent the topic of the field, sometimes there are often keywords from another field, which is either related or included in the same field. It is difficult to exclude them directly by a single indexing method (more manual screening is needed). In this case, what we can do is not miss those articles that are valuable. With proper indexing, we can learn what we want from these literature databases. Finally, because the amount of literature in the field of homomorphic cryptography is so small compared to other fields that have grown significantly, we do not intend to use the latest machine learning methods of topic modeling for this analysis, as it requires a significant amount of sample data. Perhaps in the future, when the field of HEML has grown and a large dataset is available, it will be more appropriate to analyze the statistics with a thematic model.

## 6. Conclusions and Future Work

This paper presents a preliminary bibliometric analysis of the HEML research literature. As expected, HEML is evolving and the number of papers in this field is increasing every year. To date, however, approximately one-third (36%) of the papers in the field have not been cited. This raises the question: Why is the proportion of papers that have never been cited in this field so large? How does this trend compare to other scientific fields? Is it because we have too many little-known venues where papers are published that are invisible by other researchers? Does it have to do with the quality of the paper or conference? Or is it because this is a new field that has only developed in the last 10 years? Perhaps we should further investigate the HEML literature in the future.

Our dataset [33] can be used to perform other thematic and quantitative statistical analyses in HEML and its subfields. For example, we show the developments in the field from 2009 to 2020 and the most cited research topics. This bibliometric approach can be repeated periodically to analyze the growth and trends in the field in the coming years and to compare the future trends with the findings of this study. From our point of view, the methodology of this paper is general in nature.

**Funding:** This research was funded by Ningbo Natural Science Foundation, grant number 202003N4320 and 202003N4321. This research was partially supported by the National Natural Science Foundation of China, grant number 62002335. This research was partially supported by Zhejiang Province Public Welfare Technology Application Research, grant number GF22F026173.

**Data Availability Statement:** Data results can be found at <https://drive.google.com/file/d/1IwD8Pr6Fs4LeUW7WJcGukJfv02zf6Gg1/view?usp=sharing>.



**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Microsoft Azure, AI. Available online: <https://azure.microsoft.com/zh-cn/services/machine-learning/> (accessed on 20 December 2020).
2. Google Prediction API. Available online: <https://cloud.google.com/ai-platform> (accessed on 20 December 2020).
3. Low, Y.; Gonzalez, J.; Kyrola, A.; Bickson, D.; Guestrin, C.; Hellerstein, J. GraphLab: A new framework for parallel machine learning. In Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, UAI 2010, Catalina Island, CA, USA, 8–11 July 2010; pp. 340–349.
4. Ersatz Labs. Available online: <https://www.ersatzlabs.com/> (accessed on 1 February 2021).
5. Kim, A.; Song, Y.; Kim, M.; Lee, K.; Cheon, J.H. Logistic Regression Model Training based on the Approximate Homomorphic Encryption. In Proceedings of the 6th iDASH Privacy and Security Workshop 2017, Orlando, FL, USA, 14 October 2017.
6. Han, K.; Hong, S.; Cheon, J.H.; Park, D. Efficient Logistic Regression on Large Encrypted. In Proceedings of the 33th AAAI Conference on Artificial Intelligence, AAAI 2019, Honolulu, HI, USA, 27 January–1 February 2019; pp. 9466–9471.
7. Han, K.; Hong, S.; Cheon, J.H.; Park, D. Logistic regression on homomorphic encrypted data at scale. In Proceedings of the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, HI, USA, 28–29 January 2019; pp. 9466–9471.
8. Kim, M.; Lee, J.; Ohno-Machado, L.; Jiang, X. Secure and Differentially Private Logistic Regression for Horizontally Distributed Data. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 695–710. [CrossRef]
9. Kim, M.; Song, Y.; Wang, S.; Xia, Y.; Jiang, X. Secure logistic regression based on homomorphic encryption: Design and evaluation. *JMIR Med. Inform.* **2018**, *6*, e8805. [CrossRef] [PubMed]
10. Xu, R.; Joshi, J.; Li, C. CryptoNN: Training Neural Networks over Encrypted Data. In Proceedings of the International Conference on Distributed Computing Systems 2019, Dallas, TX, USA, 7–9 July 2019; Volume 8885038, pp. 1199–1209.
11. Lou, Q.; Feng, B.; Fox, G.C.; Jiang, L. Glyph: Fast and Accurately Training Deep Neural Networks on Encrypted Data. In Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, BC, Canada, 6–12 December 2020. *arXiv* **2020**, arXiv:1911.07101.
12. Jiang, X.; Lauter, K.; Kim, M.; Song, Y. Secure outsourced matrix computation and application to neural networks. In Proceedings of the ACM Conference on Computer and Communications Security, Toronto, ON, Canada, 15–19 October 2018; pp. 1209–1222.
13. Dowlin, N.; Gilad-Bachrach, R.; Laine, K.; Lauter, K.; Naehrig, M.; Wernsing, J. CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy. In Proceedings of the 33th International Conference on Machine Learning, ICML 2016, New York, NY, USA, 19–24 June 2016; Volume 1, pp. 342–351.
14. SEAL. Available online: <https://github.com/microsoft/SEAL> (accessed on 25 September 2021).
15. HELib. Available online: <https://github.com/homenc/HELlib> (accessed on 7 February 2021).
16. TFHE. Available online: <https://tfhe.github.io/tfhe/> (accessed on 5 May 2021).
17. Brutzkus, A.; Elisha, O.; Gilad-Bachrach, R. Low Latency Privacy Preserving Inference. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 1295–1304.
18. Boura, C.; Gama, N.; Georgieva, M.; Jetchev, D. CHIMERA: Combining Ring-LWE-based Fully Homomorphic Encryption Schemes. *J. Math. Cryptol.* **2020**, *14*, 316–338. [CrossRef]
19. Li, Z.; Sun, M. Privacy-Preserving Classification of Personal Data with Fully Homomorphic Encryption: An Application to High-Quality Ionospheric Data Prediction. In *2020 Lecture Notes in Computer Science 12486 LNCS*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 437–446.
20. Li, P.; Li, J.; Huang, Z.; Li, T.; Gao, C.; You, S.; Chen, K. Multi-key privacy-preserving deep learning in cloud computing. In *Future Generation Computer Systems*; Elsevier: Amsterdam, The Netherlands, 2017; Volume 74, pp. 76–85.
21. Nandakumar, K.; Ratha, N.; Pankanti, S.; Halevi, S. Towards Deep Neural Network Training on Encrypted Data. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*; CVF: Long Beach, CA, USA, 2019; pp. 40–48.
22. Sirichotedumrong, W.; Maekawa, T.; Kinoshita, Y.; Kiya, H. Privacy-Preserving Deep Neural Networks with Pixel-Based Image Encryption Considering Data Augmentation in the Encrypted Domain. In Proceedings of the International Conference on Image Processing, ICIP, Taipei, Taiwan, 22–25 September 2019; pp. 674–678.
23. Hesamifard, E.; Takabi, H.; Ghasemi, M. CryptoDL: Deep Neural Networks over Encrypted Data. *arXiv* **2017**, arXiv:1711.05189.
24. Hesamifard, E.; Takabi, H.; Ghasemi, M. Deep Neural Networks Classification over Encrypted Data. In Proceedings of the 9th ACM Conference on Data and Application Security and Privacy (CODASPY), Dallas, TX, USA, 25–27 March 2019; pp. 97–108.
25. Boemer, F.; Costache, A.; Cammarota, R.; Wierzynski, C. nGraph-HE2: A High-Throughput Framework for Neural Network Inference on Encrypted Data. In Proceedings of the ACM Conference on Computer and Communications Security, Toronto, ON, Canada, 15–19 October 2018; pp. 45–56.
26. Boemer, F.; Lao, Y.; Wierzynski, C. NGRAPH-HE: A Graph Compiler for Deep Learning on Homomorphically Encrypted Data. *arXiv* **2018**, arXiv:1810.10121v1.
27. Izabachène, M.; Sirdey, R.; Zuber, M. Practical fully homomorphic encryption for fully masked neural networks. In *Lecture Notes in Computer Science 11829 LNCS*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 24–36.

28. Boddeti, V. Secure Face Matching Using Fully Homomorphic Encryption. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems, BTAS, Radondo Beach, CA, USA, 22–25 October 2018; Volume 8698601.
29. Suh, J.; Tanaka, T. Reinforcement Learning over Fully Homomorphic Encryption, 7 pages, 2 figures, submitted to SICE ISCS 2021. *arXiv* **2020**, arXiv:2002.00506.
30. Shortell, T.; Shokoufandeh, A. Secure Convolutional Neural Networks using FHE. *arXiv* **2018**, arXiv:1808.03819.
31. Garousi, V.; Mäntylä, M.V. Bibliometrics of SE literature. *Computer Science Review. J. Inf. Sci.* **2016**, *19*, 56–77.
32. Lauter, K.; Naehrig, M.; Vaikuntanathan, V. Can homomorphic encryption be practical? In Proceedings of the ACM Conference on Computer and Communications Security, Toronto, ON, Canada, 15–19 October 2018; 2011; pp. 113–124.
33. HEML. Available online: <https://drive.google.com/file/d/1IwD8Pr6Fs4LeUW7WJcGukJfv02zf6Gg1/view?usp=sharing> (accessed on 9 May 2021).
34. Gentry, C. Computing arbitrary functions of encrypted data. In *Communications of the ACM*; ACM: New York, NY, USA, 2010; Volume 53, pp. 97–105.
35. Nikolaenko, V.; Weinsberg, U.; Ioannidis, S.; Joye, M.; Boneh, D.; Taft, N. Privacy-preserving ridge regression on hundreds of millions of records. In Proceedings of the 2013 IEEE Symposium on Security and Privacy, Berkeley, CA, USA, 19–22 May 2013; pp. 334–348.
36. Phong, L.T.; Aono, Y.; Hayashi, T.; Wang, L.; Moriai, S. Privacy-Preserving Deep Learning via Additively Homomorphic Encryption. In *IEEE Transactions on Information Forensics and Security*; IEEE: Piscataway, NJ, USA, 2017; Volume 13, pp. 1333–1345.
37. Graepel, T.; Lauter, K.; Naehrig, M. ML confidential: Machine learning on encrypted data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7839 LNCS; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 1–21.
38. Yu, J.; Lu, P.; Zhu, Y.; Xue, G.; Li, M. Toward secure multikeyword top-k retrieval over encrypted cloud data. In *IEEE Transactions on Dependable and Secure Computing*; IEEE: Piscataway, NJ, USA, 2017; Volume 10, pp. 239–250.
39. Li, P.; Li, J.; Huang, Z.; Gao, C.-Z.; Chen, W.-B.; Chen, K. *Privacy-Preserving Outsourced Classification in Cloud Computing, Cluster Computing*; Springer: Berlin/Heidelberg, Germany, 2017; Volume 21, pp. 277–286.
40. Juvekar, C.; Vaikuntanathan, V.; Chandrakasan, A. GAZELLE: A low latency framework for secure neural network inference. In Proceedings of the 27th USENIX Security Symposium, Baltimore, Maryland, 15–17 August 2018; pp. 1651–1668.
41. Shen, M.; Ma, B.; Zhu, L.; Mijumbi, R.; Du, X.; Hu, J. Cloud-Based Approximate Constrained Shortest Distance Queries over Encrypted Graphs with Privacy Protection. In *IEEE Transactions on Information Forensics and Security*; IEEE: Piscataway, NJ, USA, 2017; Volume 13, pp. 940–953.
42. Riazzi, M.; Weinert, C.; Tkachenko, O.; Songhori, E.M.; Schneider, T.; Koushanfar, F. Chameleon: A Hybrid Secure Computation Framework for Machine Learning Applications. In Proceedings of the 2018 on Asia Conference on Computer and Communications Security, New York, NY, USA, 4 June 2018.
43. Noorden, R.V.; Maher, B.; Nuzzo, R. The top 100 papers. *Nature* **2014**, *514*, 550–553. [[CrossRef](#)] [[PubMed](#)]
44. Topic Model. Available online: [https://en.wikipedia.org/wiki/Topic\\_model](https://en.wikipedia.org/wiki/Topic_model) (accessed on 18 April 2021).
45. LDA. Available online: [https://en.wikipedia.org/wiki/Latent\\_Dirichlet\\_allocation](https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation) (accessed on 18 April 2021).
46. Khilji, I.Q.; Saha, K.; Shonon, J.A.; Hossain, M.I. Application of homomorphic encryption on neural network in prediction of acute lymphoid Leukemia. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 350–360. [[CrossRef](#)]
47. Doröz, Y.; Shahverdi, A.; Eisenbarth, T.; Sunar, B. Toward practical homomorphic evaluation of block ciphers using prince. In *Lecture Notes in Computer Science* 8438; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 208–220.
48. Çetin, G.S.; Doröz, Y.; Sunar, B.; Savaş, E. Depth optimized efficient homomorphic sorting. In *Lecture Notes in Computer Science* 9230; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 61–80.
49. Wu, W.; Liu, J.; Wang, H.; Tang, F.; Xian, M. PPolyNets: Achieving High Prediction Accuracy and Efficiency with Parametric Polynomial Activations. In *IEEE Access* 6; IEEE: Piscataway, NJ, USA, 2018; pp. 72814–72823.
50. Song, X.; Chen, Z. A general design method of constructing fully homomorphic encryption with ciphertext matrix. In *KSII Transactions on Internet and Information Systems* 13; Elsevier: Amsterdam, The Netherlands, 2019; pp. 2629–2650.