

CAIR-NEPAL

Shaping the Future with AI

Essentials of Research Writing

Dr. Tek Raj Chhetri

Founder & Director CAIR-Nepal

1 February 2025

Research Writing (Online)

Hosted by IEEE Pulchowk Campus & CAIR-Nepal



Outline

1. Introduction
2. Structure of a Research Paper
3. Publication Process
4. Cover Letter For Article Submission
5. Revising the manuscript
6. Rejection
7. Where to Publish?
8. Best Practices

1.

Introduction

1. Introduction

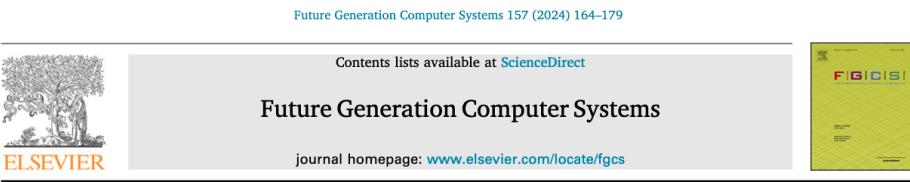
- What is a research paper?



A research paper is a structured document that presents an in-depth study of a specific topic, based on original research or a review of existing literature. It is typically written by scholars, researchers, or students to contribute new knowledge, analyze existing findings, or propose novel theories.

- Writing research paper requires substantial effort and is not easy to generate, (definitely) not in a semester [1].

1. Introduction



Enabling privacy-aware interoperable and quality IoT data sharing with context

Tek Raj Chhetri^{a,b,c,d,*}, Chinmaya Kumar Dehury^e, Blesson Varghese^f, Anna Fensel^{g,h}, Satish Narayana Sriramaⁱ, Rance J. DeLong^j

^a McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

^b Semantic Technology Institute (STI), Department of Computer Science, Universität Innsbruck, Innsbruck, 6020, Austria

^c Center for Artificial Intelligence (AI) Research Nepal, Sundarharaincha-09, Nepal

^d Web and Internet Science Research Group, School of Electronics and Computer Science, The University of Southampton, Southampton SO17 1BJ, United Kingdom

^e Institute of Computer Science, University of Tartu, Tartu, 50090, Estonia

^f School of Computer Science, University of St Andrews, St Andrews, Fife, UK KY16 9SX, United Kingdom

^g Wageningen Data Competence Center, Wageningen University & Research, Wageningen, 6708 PB, The Netherlands

^h Artificial Intelligence Chair Group, Wageningen University & Research, Wageningen, 6708 PB, The Netherlands

ⁱ School of Computer and Information Sciences, University of Hyderabad, Hyderabad, 500046, India

^j The Open Group, Reading, Berkshire, RG1 1AX, United Kingdom



Random Tech Blog

About

AWS Terraform Setup

Feb 1, 2023

Terraform is a tool for building, changing, and versioning infrastructure safely and efficiently...at least according to ChatGPT.

I started messing around with it so I could quickly spin up EC2 instances. Here are some quick instructions to get your workstation ready to terraform.

I utilized the Windows Linux Subsystem, so first, you need to install WSL along with Ubuntu.

From an elevated powershell prompt, run the command `wsl --install -d ubuntu` and it should install both WLS and Ubuntu.

```
wsl --install -d ubuntu  
Copy
```

Once Ubuntu is up and running, we need to install Terraform.

```
sudo apt-get update && sudo apt-get install -y gnupg software-properties-common  
  
wget -O- https://apt.releases.hashicorp.com/gpg | \  
gpg --dearmor | \  
sudo tee /usr/share/keyrings/hashicorp-archive-keyring.gpg  
  
gpg --no-default-keyring \  
--keyring /usr/share/keyrings/hashicorp-archive-keyring.gpg \  
--fingerprint
```

1. Introduction

- What types of research can one conduct?
- Below are some non exhaustive types of research one can conduct.
 - Foundational research: Research that focuses on gaining new knowledge, for example, studying how neurons communicate in brain.
 - Applied research: Research focuses on solving real-world problems, e.g., autonomous vehicles.
 - Translational research: Translating research findings into practical applications to improve human health, e.g., using neuroscience findings to solve Alzheimer's (brain disorder that destroys memory).

2.

Structure of a Research Paper

2. Structure of a Research Paper

- The general structure of a research paper.
 - Title: Should be short and catchy, conveying your research.
 - Abstract
 - Introduction
 - Literature Review
 - Methodology
 - Experiment
 - Results & Discussion
 - Conclusion
 - References/Bibliography

2.1

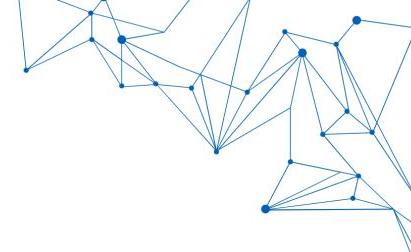
Structure of a Research Paper

Abstract

2.1 Abstract

- Provides the overview of the research paper, summarising the paper and conveys what was done, why, how and what results were achieved.
- Abstract is usually between 150-250 words (unless explicitly specified otherwise).
- Generally abstract doesn't include any references.
- Avoid abbreviations unless it is widely known.
- Based on publishers or fields of research abstracts may vary. Some publishers ask explicitly have sections such as objectives, results ... in abstract.

2.1 Abstract



Towards improving prediction accuracy and user-level explainability using deep learning and knowledge graphs: A study on cassava disease

Tek Raj Chhetri^{a,b,*}, Armin Hohenegger^a, Anna Fensel^{c,d}, Mariam Aramide Kasali^e, Asiru Afeez Adekunle^f

^a Semantic Technology Institute (STI), Department of Computer Science, Universität Innsbruck, Innsbruck, 6020, Austria

^b Center for Artificial Intelligence (AI) Research Nepal, Sundarharancha-09, Nepal

^c Wageningen Data Competence Center, Wageningen University & Research, Wageningen, The Netherlands

^d Consumption and Healthy Lifestyles Chair Group, Wageningen University & Research, Wageningen, The Netherlands

^e Department of Plant Physiology and Crop Production, College of Plant Sciences, Federal University of Agriculture, Abeokuta, Nigeria

^f Kwara Agricultural Network, Ilorin, Kwara, Nigeria

ARTICLE INFO

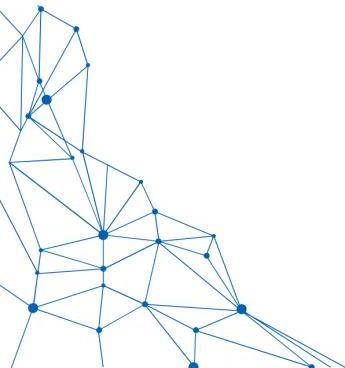
Dataset link: Survey Data: Towards Improving Prediction Accuracy and User-Level Explainability Using Deep Learning and Knowledge Graphs : A Study on Cassava Disease (Original data)

Keywords:

Explainable AI (XAI)
Agricultural sustainability
Knowledge graphs
Deep learning
Cassava

ABSTRACT

Food security is currently a major concern due to the growing global population, the exponential increase in food demand, the deterioration of soil quality, the occurrence of numerous diseases, and the effects of climate change on crop yield. Sustainable agriculture is necessary to solve this food security challenge. Disruptive technologies, such as of artificial intelligence, especially, deep learning techniques can contribute to agricultural sustainability. For example, applying deep learning techniques for early disease classification allows us to take timely action, thereby helping to increase the yield without inflicting unnecessary environmental damage, such as excessive use of fertilisers or pesticides. Several studies have been conducted on agricultural sustainability using deep learning techniques and also semantic web technologies such as ontologies and knowledge graphs. However, the three major challenges remain: (i) the lack of explainability of deep learning-based systems (e.g. disease information), especially to non-experts like farmers; (ii) a lack of contextual information (e.g. soil or plant information) and domain-expert knowledge in deep learning-based systems; and (iii) the lack of pattern learning ability of systems based on the semantic web, despite their ability to incorporate domain knowledge. Therefore, this paper presents the work on disease classification, addressing the challenges as mentioned earlier by combining deep learning and semantic web technologies, namely ontologies and knowledge graphs. The findings are: (i) 0.905 (90.5%) prediction accuracy on large noisy dataset; (ii) ability to generate user-level explanations about disease and incorporate contextual and domain knowledge; (iii) the average prediction latency of 3.8514 s on 5268 samples; (iv) 95% of users finding the explanation of the proposed method useful; and (v) 85% of users being able to understand generated explanations easily—show that the proposed method is superior to the state-of-the-art in terms of performance and explainability and is also suitable for real-world scenarios.



2.1 Abstract



OPEN ACCESS

Citation: Shiely F, Gallagher K, Millar SR (2024) How, and why, science and health researchers read scientific (IMRAD) papers. PLoS ONE 19(1): e0297034. <https://doi.org/10.1371/journal.pone.0297034>

Editor: Bogdan Nadolu, West University of Timisoara: Universitatea de Vest din Timisoara, ROMANIA

Received: February 21, 2023

Accepted: December 26, 2023

Published: January 22, 2024

Copyright: © 2024 Shiely et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The datasets generated and/or analysed during the current study are available <https://osf.io/up4ny/>.

Funding: This study was funded by Teaching and Learning Enhancement Fund at University College Cork for a summer student scholarship. The

RESEARCH ARTICLE

How, and why, science and health researchers read scientific (IMRAD) papers

Frances Shiely^{1,2*}, Kerrie Gallagher^{1,2}, Séan R. Millar²

¹ Trials Research and Methodologies Unit, HRB Clinical Research Facility, University College Cork, Cork, Ireland, ² School of Public Health, University College Cork, Cork, Ireland

* f.shiely@ucc.ie

Abstract

Objectives

The purpose of our study was to determine the order in which science and health researchers read scientific papers, their reasons for doing so and the perceived difficulty and perceived importance of each section.

Study design and setting

An online survey open to science and health academics and researchers distributed via existing research networks, X (formerly Twitter), and LinkedIn.

Results

Almost 90% of respondents self-declared to be experienced in reading research papers. 98.6% of the sample read the abstract first because it provides an overview of the paper and facilitates a decision on continuing to read or not. Seventy-five percent perceived it to be the easiest to read and 62.4% perceived it to be very important (highest rank on a 5-point Likert scale). The majority of respondents did not read a paper in the IMRAD (Introduction, Methods, Results And Discussion) format. Perceived difficulty and perceived importance influenced reading order.

Conclusion

Science and health researchers do not typically read scientific and health research papers in IMRAD format. The more important a respondent perceives a section to be, the more likely they are to read it. The easier a section is perceived, the more likely it will be read. We present recommendations to those teaching the skill of writing scientific papers and reports.

2.2

Structure of a Research Paper

Introduction

2.2 Introduction

- One of the most important part of the paper and it should clearly convey what research you are doing, why you are doing particular research and why does it matter?
- The introduction should follow a ***funnel structure***, starting with broad context, narrowing down to the specific research problem, and concluding with the study's objectives and significance.
- It should be max. 1.5-2 pages long.



Source: Wikimedia.org

2.2 Introduction

- The introduction can be structured into subsections for clarity and logical flow, such as:
 - Motivation provides the broader significance of the problem.
 - Goal should clearly indicate objectives and research questions being explored.
 - Contributions should highlight the key contributions that paper has made. In case of multiple contributions it's suggested to organize using bullet points.



Source: Wikimedia.org

2.2 Introduction

Article

A Combined System Metrics Approach to Cloud Service Reliability Using Artificial Intelligence

Tek Raj Chhetri¹ , Chinmaya Kumar Dehury² , Artjom Lind² , Satish Narayana Srirama³  and Anna Fensel^{1,4,5} 

- 1 Semantic Technology Institute (STI) Innsbruck, Department of Computer Science, University of Innsbruck, 6020 Innsbruck, Austria; tek.raj.chhetri@stibit.at (T.R.C.); anna.fensel@stibit.at (A.F.)
 - 2 Institute of Computer Science, University of Tartu, 30090 Tartu, Estonia; artjom.lind@ut.ee
 - 3 School of Computer and Information Sciences, University of Hyderabad, Hyderabad 500046, India; satish.srirama@uohyd.ac.in
 - 4 Wageningen Data Competence Center, Wageningen University & Research, 6708 PB Wageningen, The Netherlands
 - 5 Consumption and Healthy Lifestyles Chair Group, Wageningen University & Research, 6706 KN Wageningen, The Netherlands
- * Correspondence: chinmaya.dehury@ut.ee

Abstract: Identifying and anticipating potential failures in the cloud is an effective method for increasing cloud reliability and proactive failure management. Many studies have been conducted to predict potential failure, but none have combined SMART (self-monitoring, analysis, and reporting technology) hard drive metrics with other system metrics, such as central processing unit (CPU) utilisation. Therefore, we propose a combined system metrics approach for failure prediction based on artificial intelligence to improve reliability. We tested over 100 cloud servers' data and four artificial intelligence algorithms: random forest, gradient boosting, long short-term memory, and gated recurrent unit, and also performed correlation analysis. Our correlation analysis sheds light on the relationships that exist between system metrics and failure, and the experimental results demonstrate the advantages of combining system metrics, outperforming the state-of-the-art.

Keywords: failure prediction; fault tolerance; cloud computing; artificial intelligence; reliability

1. Introduction

Cloud computing has emerged as the fifth utility over the last decade, and is a backbone to the modern economy [1]. It is a model of computing that allows flexible use of virtual servers, massive scalability, and management services for the delivery of information services. With the low-cost pay-per-use model of on-demand computing [2], the cloud has grown massively over the years, both in terms of size and complexity.

Today, almost everyone is connected to the cloud in one way or another. This is because of cost effectiveness with a pay-as-you-go or subscription-based service model for on-demand access to IT resources [1,2]. Industries rely on the cloud for their operations, academics to accelerate and conduct scientific experiments, and ordinary end-users by using cloud-based services knowingly or unknowingly, such as Google Drive, Gmail, Outlook, and so on. Furthermore, the cloud today is more important than yesterday, as it supports smart city construction [3], enterprise business [4], scalable data analysis [5,6], healthcare [7,8] and also new evolving computing paradigms, such as fog and edge com-



Citation: Chhetri, T.R.; Dehury, C.K.; Lind, A.; Srirama, S.N.; Fensel, A. A Combined System Metrics Approach to Cloud Service Reliability Using Artificial Intelligence. *Big Data Cogn. Comput.* **2022**, *6*, 26. <https://doi.org/10.3390/bdcc6010026>

Academic Editor: Domenico Talia
Received: 25 November 2021
Accepted: 24 February 2022
Published: 1 March 2022

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



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland.

1.1. Motivation

Today, the majority of businesses rely on cloud services to run their daily operations. Any failure of cloud services directly impacts the business, and repeated failures result in reputation damage. Reputation is an intangible asset that accounts for 85% of the value of a business [16,17]. Significant effort has been made to increase the reliability of cloud services. For failure prediction, the studies (see Section 2) either use hard drive SMART (self-monitoring, analysis and reporting technology) metrics or other system metrics, such as CPU and memory utilisation. However, despite the demonstrated ability of SMART hard drive metrics and other system metrics, such as CPU and memory utilisation to predict failures, to our knowledge, no study has combined both system metrics. We hypothesised that combining both system metrics will give us more information and, as a result, can help improve reliability further. The identified existing research gap on the combined use of SMART hard drive metrics and other system metrics, which we in our study refer to as combined system metrics, to improve reliability is the main motivation for this study. The other motivation includes the limited number of studies in the direction of server failure prediction, the use of a traditional rule-based tool, such as Prometheus [18]. Additionally, there is little, if any, correlation analysis of such metrics, which could reveal critical patterns for forecasting.

1.2. Goal

The ability to accurately predict failure is an essential factor for reliable performance. This is because it allows us to take action that is proactive. The reliability of a process depends on how well we predict failures. The aim of this study is to improve the reliability of cloud services by improving cloud server failure prediction, using selected AI techniques and the combined system metrics approach.

1.3. Contributions

Based on our motivation and goal to achieve, the main contributions of our study can be summarised as follows.

- A novel approach to server failure prediction based on combined system metrics.
- Use of AI approaches to overcome the disadvantages of rule-based failure prediction.

2.3

Structure of a Research Paper

Literature Review

2.3 Literature Review

- Literature review section provides the overview of the state-of-the-art (SOTA) in relation to the work being done.
- Literature review may be optional in some journals. In computer science, **it's always required!**
- Should contrast existing SOTA, highlighting gaps and advancements.

Tip! Have a table comparing your works with SOTA.

The screenshot shows a web browser window with the URL keapublishing.com/en/journals/artificial-intelligence-in-agriculture/guide-for-authors. The page content is a guide for authors, specifically for a journal on artificial intelligence in agriculture. It includes several sections: Article structure, Subdivision - numbered sections, Introduction, Material and methods, Theory/calculation, Results, Discussion, Conclusions, and Appendices. The browser tabs at the top include links to 14th International..., Maintaining Privacy, GitHub - google/rainbow, Knowledge Graph, Semantic annotation, An Attention Enhancer, SCCH customer p..., BibTeX Style Editor, Home, About us, Journals, Events, Partners, and Forum. The overall theme of the slide is the importance of a well-structured literature review in academic publications, particularly in computer science and engineering fields like AI in agriculture.

2.3 Literature Review

- Group the SOTA works by type.
- If you are doing systematic literature review study, then follow the PRISMA model [2].

mdpi.com/2504-2289/6/I/26

International Journal of Grid and Utility Computing

Maintaining Privacy in Cloud Computing

GitHub - google/ra...

Knowledge Graph...

Semantic annotat...

An Attention Enhanc...

SCCH customer p...

BibTeX Style Exam...

Order Article Reprint

Abstract

Introduction

State-of-the-Art

Materials and Methods

Performance Evaluation

Conclusions

Author Contributions

Funding

Institutional Review Board Statement

Informed Consent Statement

Data Availability Statement

Acknowledgments

Conflicts of Interest

Abbreviations

References

Download PDF

2. State-of-the-Art

In this section, we present a brief survey of recent works in failure prediction in the domain of cloud computing. Section 2.1 provides an overview of the work related to server (or server-level) failure prediction, which our work focuses on. Furthermore, we also provide a brief overview of virtual machine (VM) and task failure prediction in Section 2.2 and Section 2.3, respectively. This is to provide an overview of the research on failure prediction in the cloud domain. Finally, we provide a summary of the literature review in Section 2.4.

2.1. Server-Level Failure Prediction

Mohammed et al. [19], Xu et al. [20], Lai et al. [21], Das et al. [22], Chigurupati et al. [23], Tehrani et al. [24], and Adamu et al. [25] carried out a study on server (or server-level) failure prediction. The research by Mohammed et al. [19] focused on the prediction of containerised high-performance computing (HPC) system failures using failure information, such as hardware, software, network, undetermined, and human error. Furthermore, support vector machine (SVM), RF, K-nearest neighbours (KNN), classification and regression trees (CART), and linear discriminant analysis (LDA) were used in the study. However, we cannot tell if the system failed or if there was human intervention based on information such as human errors. Furthermore, the scope of the unidentified error source is unclear. Unlike Mohammed et al. [19], Xu et al. [20] used a ranking-based machine learning approach and SMART hard disk information for failure prediction in cloud systems to improve the service availability of Microsoft Azure by migrating VMs from failing to healthy nodes.

Similarly to Xu et al. [20], Das et al. [22] also focused on migrating computation from a failing node to a healthy node. However, Das et al. [22] focused on using a deep learning (i.e., LSTM) approach, compared to Xu et al. [20] who used a ranking-based approach. On the other hand, Lai et al. [21] used techniques such as KNN and hard disk data from the SLAC Accelerator Laboratory [26] to predict server failure within 60 days and introduced a derived metric *time_since_prev_failure* for server failure prediction. Furthermore, the study by Lai et al. [21] made use of failure logs that were kept for a period of 10 years. Based on their experience, Lai et al. [21] also recommend using an RNN-based technique, such as LSTM.

Similarly, Chigurupati et al. [23], Tehrani et al. [24], and Adamu et al. [25] used techniques such as SVM failure prediction. While the study by Chigurupati et al. [23] focused on predicting communication hardware failure min ahead, the study by Tehrani et al. [24] focused on failure prediction in cloud systems in a simulated environment using system metrics such as temperature, CPU, RAM, and bandwidth utilisation. Adamu et al. [25], like other previous studies, focused on failure prediction in a cloud environment using data from the National Energy Research Scientific Computing Center's [27] Computer Failure Data Repository. The author separated the failures of a disc, dual in-line memory module (DIMM), the CPU, and other components. However, the scope of the failure, such as other failures in the study, is unclear, and network information was not used, which is another reason for the failure.

2.2. VM-Level Failure Prediction

A study on VM failure prediction was carried out by Meenakumari et al. [28], Alkaesem et al. [29], Qasim et al. [30], Liu et al. [31] and Rawat et al. [32]. The study by Meenakumari et al. [28] employed a dynamic thresholding approach to predict failure based on system metrics such as CPU utilisation, CPU usage, bandwidth, temperature, and memory. Similar to Meenakumari et al. [28], Alkaesem et al. [29] also focused on VM fail prediction. The study by Alkaesem et al. [29] focused on the VM startup failure problem by using system metrics such as CPU utilisation, memory usage, network overhead, and IO (input/output) storage usage. Alkaesem et al. [29] used Apache Spark [33] streaming together with Native Bayes (NB). Both Qasem et al. [30] and Liu et al. [31] investigated VM failure using recurrent neural networks (RNN). However, Qasem et al. [30] used simulated data from CloudSim [34] to evaluate their proposed model.

2.3 Literature Review

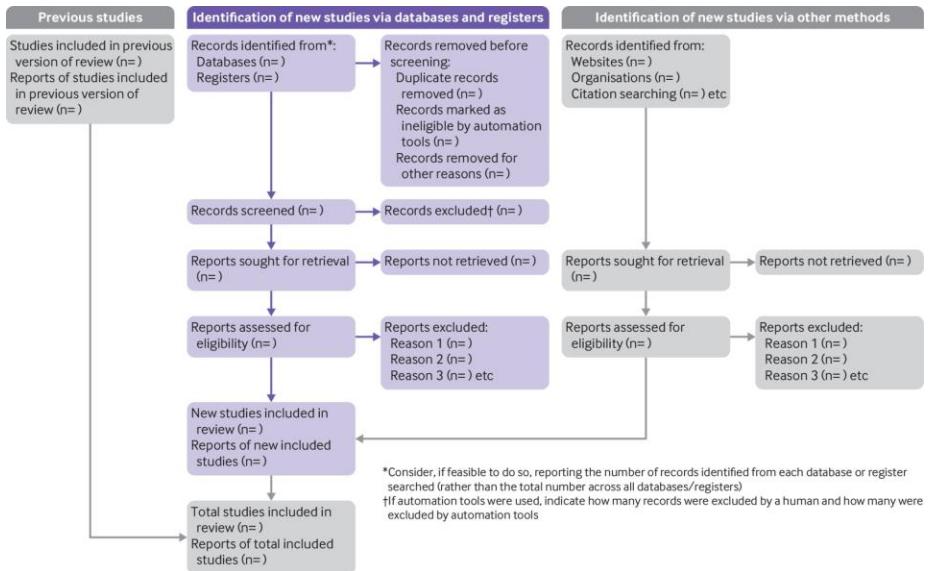


Figure 1: PRISMA 2020 flow diagram template for systematic reviews [2].

2.4

Structure of a Research Paper

Methodology

2.4 Methodology

- The central component of the section that highlights ***overall strategy and rationale behind your approach (or method)*** [3].
- Not to be confused, **methods and methodology are not same!**
- (Research) Methodology provides a theoretical framework, defines the research design and justification for specific method. It is a systematic way to solve problem [3] and dictates how a problem is approached.
- (Research) Methods are the procedures (or techniques) by which you conduct research on a subject or topic [3].

2.4 Methodology

- One can calculate mean, median and apply in a research, but the consideration for which one would be suitable for the research constitute a methodology.
- Methodology is focused on explaining questions such as:
 - Why particular method is used?
 - What types of data were collected and how?
- The methodology differs from the type of research and also the field of research and can include subsections.
 - Despite variations in structure across disciplines, the **central component remains the same.**

2.4 Methodology

- **Careful!** incorrect or unclear methodology (or methods) can lead to rejection of the paper.
- Suggested reading
 - Goundar, S., 2012. Research methodology and research method. Victoria University of Wellington, 1(1), pp.1-47.

2.4 Methodology

- Unclear methodology can lead to irreproducible research.

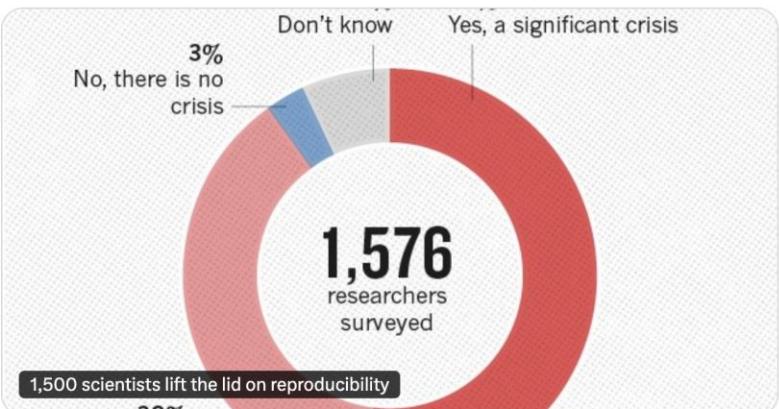


Tek Raj Chhetri
@tekraj_14

Promote

...

70% of research isn't **#reproducible**, and 50% of scientists can't even reproduce their own work! (Ref: nature.com/articles/53345...) 🔈 Stay tuned for our accepted article at **#TheWebConference** to learn more! 🔎 **#Neuroscience #ComputationalBiology #KnowledgeGraphs #OpenScience**



From nature.com

2.4 Methodology

mdpi.com/2504-2289/6/1/26

Maintaining Privacy... GitHub - google/ra... Knowledge Graph... Semantic annotati... An Attention Enh... SCCH customer p... BibTeX Style Exam... Stud...

Abstract
Introduction
State-of-the-Art
Materials and Methods
Performance Evaluation
Conclusions
Author Contributions
Funding
Institutional Review Board Statement
Informed Consent Statement
Data Availability Statement
Acknowledgments
Conflicts of Interest
Abbreviations
References

Download PDF Order Article Reprints

3. Materials and Methods

This section details our approach. Section 3.1 contains information about data collection; Section 3.2 contains information about the data used in our study; and Section 3.3 contains information about data preprocessing. Similarly, Section 3.4, Section 3.5 and Section 3.6 discuss our implementation of RF, GB, and LSTM and GRU algorithms. Before going into detail about the implementation, Section 3.4, Section 3.5 and Section 3.6 provide an overview of the algorithms RF, GB, and LSTM and GRU, as well as the rationale for their selection.

3.1. Data Collection

There are various datasets available, including the Google job failure dataset [45], and the SMART hard drive dataset [46]. However, these datasets either contain information for job failure or hard drive failure and lack information on other system metrics, such as CPU utilisation for failure, necessitating data collection. Our data collection step entails downloading data from Prometheus [47]—monitored University of Tartu High-Performance Centre [48] cloud servers. A Python [49] script converts the downloaded JSON [50] data to CSV (Comma-separated values) format, the source code for which is available at [51].

3.2. Dataset Description

The dataset used in this study contains information about the system metrics collected through the use of Prometheus and is publicly available at [52]. The collected system metrics were chosen following a thorough assessment and analysis of their impact on the system failure. Table 1 summarises the selected system metrics. In addition to the selected system metrics, the dataset also contains the timestamp and anonymised server information. The dataset contains a total of 7,371,203 samples. Furthermore, details on target label generation and preprocessing are available in Section 3.3. Additionally, Table 2 summarises the dataset's statistical information in terms of mean, standard deviation (std.), and counts (or sample size). From Table 2, looking at the mean and the standard deviation, we can observe that the data are not uniformly distributed. Furthermore, Table 2 demonstrates the uneven distribution of sample sizes for various characteristics (or selected system metrics). This uneven distribution of sample sizes results in an empty value, which has an effect on the learning of machine learning algorithms. Additionally, Figure 1, Figure 2 and Figure 3 show the visualisation of the data distribution of the system metrics memory utilisation, SMART 194, and SMART 3 (selected randomly). Figure 1, Figure 2 and Figure 3 demonstrate a non-uniform data distribution as indicated by the means and standard deviations. The detailed description (or analysis) of the data is present in Section 4.4.1.

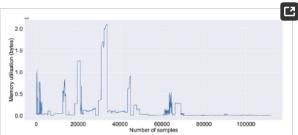


Figure 1. Memory utilisation data distribution visualisation.

2.4 Methodology

it Convert E-Sign

it Convert E-Sign

Find text or tools

Future Generation Computer Systems 157 (2024) 164–179



Contents lists available at ScienceDirect
Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Enabling privacy-aware interoperable and quality IoT data sharing with context

Tek Raj Chhetri^{a,b,c,d,*}, Chimayya Kumar Dehury^e, Blessen Varghese^f, Anna Fensel^{g,h},
Satish Narayana Srivamaⁱ, Rance J. DeLong^j

^a McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

^b Semantic Technology Institute (STI), Department of Computer Science, Universität Innsbruck, Innsbruck, 6020, Austria

^c Center for Artificial Intelligence (AI) Research Nepal, Sandharkantika-09, Nepal

^d Web and Internet Science Research Group, School of Electronics and Computer Science, The University of Southampton, University Road, Highfield Campus Southampton SO17 1BJ, United Kingdom

^e Institute of Computer Science, University of Tartu, Tartu, 50009, Estonia

^f School of Computer Science, University of St Andrews, St Andrews, Fife, UK KY16 9SX, United Kingdom

^g Wageningen Data Composite Center, Wageningen University & Research, Wageningen, 6706 PB, The Netherlands

^h Artificial Intelligence Chair Group, Wageningen University & Research, Wageningen, 6706 PB, The Netherlands

ⁱ School of Computer and Information Sciences, University of Hyderabad, Hyderabad, 500046, India

^j The Open Group, Reading, Berkshire, RG1 1AX, United Kingdom

ARTICLE INFO

ABSTRACT

Keywords:
Sharing Internet of Things (IoT) data across different sectors, such as in smart cities, become heterogeneity. This poses challenges related to a lack of interoperability, data quality issues, information, and a lack of data veracity (or accuracy). In addition, there are privacy concerns may contain personally identifiable information. To address the above challenges, this paper semantic technology-based framework that enables data sharing in a GDPR-compliant manner that the shared data is interoperable, contains required context information, is of acceptable accurate and trustworthy. The proposed framework also accounts for the edge/fog, an up paradigm for the IoT to support real-time decisions. We evaluate the performance of the pro

Sharing Internet of Things (IoT) data across different sectors, such as in smart cities, become heterogeneity. This poses challenges related to a lack of interoperability, data quality issues, information, and a lack of data veracity (or accuracy). In addition, there are privacy concerns may contain personally identifiable information. To address the above challenges, this paper semantic technology-based framework that enables data sharing in a GDPR-compliant manner that the shared data is interoperable, contains required context information, is of acceptable accurate and trustworthy. The proposed framework also accounts for the edge/fog, an up paradigm for the IoT to support real-time decisions. We evaluate the performance of the pro

racy, and quality aspects, our work also considers the intelligence at the edge or fog levels to support real-time (or near real-time) decision making. For example, the increasing use of IoT for monitoring, such as in power plants, necessitates real-time decision making. This can be accomplished by the use of edge or fog computing. However, the reviewed studies fail to take into account this essential factor that our work does.

4. Methodology

This section provides details about the methodology that is adopted for ontology modeling and the trust metric (also referred to as the trust score or trustworthiness score). The methodology that is employed in our work is applicable in other contexts. Furthermore, the ontology enables the integration of heterogeneous data sources, such as in the case of smart cities [40]. Moreover, this section also provides an overview of the functional and non-functional requirements. The preliminaries of the concepts are available as supplementary material (see Appendix A).

4.1. Ontology modeling

The ontology is the core of our systems, as it facilitates interoperability and further enables reasoning capabilities. We followed the standard ontology modeling practice [41], wherein the first step is to define the scope of the ontology (or ontology requirements), such as what the ontology should model (or cover). Since this research is concerned with the sharing of IoT data, the ontology's scope should be established in a similar manner. The scope of our ontology is outlined below. The ontology as well as the source code can be accessed at [42].

1. The ontology should be able to represent sensor information and observation results, as well as context such as units of measurement describing observation results.

2. Based on what the sensors see, the ontology should be able to support reasoning (or analytics) operations, like creating an alert in our case.

According to our scope (i.e., ontology requirements), the ontology must be able to answer the competency questions (CQs) listed below. These CQs are derived from our proposed framework, which is aimed at addressing the challenges of IoT data sharing, such as data quality

CQ 5. What type of edge reasoning is carried out?

Fig. 2 shows the ontology used in our study that was developed following our CQs. The ontology used in our study is small, i.e., it only includes a limited number of concepts pertaining to sensors and their observation. Our study's small ontology is due to the different focus of our work, which is to demonstrate the use of ontology to address challenges such as interoperability, as opposed to ontology engineering that focuses on creating large ontologies. However, the methodology adopted for our ontology modeling can be used to model larger ontologies in a collaborative setting or extend the scope of the ontology used in our study. The ontology used in this study reuses the concepts from the two different ontologies, namely SOSA [43] and the OM (Ontology of Units of Measure) [44]. We consider SOSA [43] ontology compared to other ontologies like DogOnt [45] due to its wide adoption. In Fig. 2, the reused classes and properties from the OM and SOSA ontologies are denoted by the prefixes *om* and *sosa*, respectively. The prefix *sricats* shows that the concepts are unique to this study.

With CQ 1, CQ 2, and CQ 3, the ontology captures information about the sensor observation result and also context information describing the observation, such as units in which observation was made and the time when observation was made. The *sosa:Sensor* class and property *sosamadeBySensor* provides an answer to CQ 1, while *sosc:Observation* class and *sosc:resultTime* property provides an answer to CQ 3. In a similar manner, classes *sosc:Observation*, *sosc:ObservationProperty* and *om:Unit*, as well as their respective properties *onhasUnit*, *sosa:hasSimpleResult*, and *sosc:observedProperty* provides an answer to our CQ 2. Moreover, the ontology can be easily expanded to include additional context information, such as location, based on the requirements. In addition, this incorporation of context information into the ontology enables us to validate using SHACL (Shapes constraint language) as part of the data quality, which is one of the limitations highlighted by [5] and one of our contributions.

The properties *sricats:hasTrustabilityScore* and *sricats:hasBlockchainHash* provides an answer to our CQ 4 and the property *sricats:hasEdgeReasoningType* as well as the classes *sricats:TemperatureAlert* and *sricats:HumidityAlert* answers CQ 5. The *sricats:TemperatureAlert* and *sricats:HumidityAlert* are classes representing alert operations specific to the humidity and temperature observations, respectively. The reason for this separation is to make concepts explicit and follows the ontology definition. These concepts, namely *sricats:TemperatureAlert*

Bookmarks

Enabling privacy-aware interoperable IoT data sharing with context

Introduction

Background, Motivation & Contr

Motivation

Contribution

Related work

Semantic technology-based st

Non-semantic technological st

Summary

Methodology

Ontology modeling

Trust metric

Functional and non-functional

Proposed system

Data source

Smart contract

Data receiver

Data transformation engine

Legal engine

Validation engine

2.5

Structure of a Research Paper

Experiment

2.5 Experiment

- Describes experiment component of the research, e.g., what dataset was used, hardware and how was experiment conducted.
- It also describes the tools used and the metrics used to evaluation.
- In computer science, particularly, in machine learning works, you will describe your training and testing in experiment section.

2.6

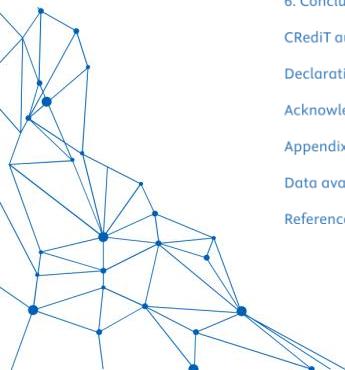
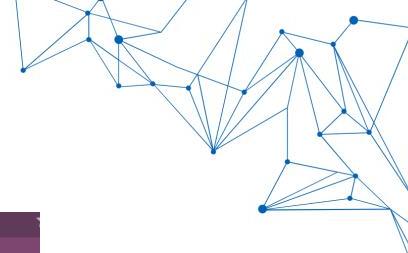
Structure of a Research Paper

Results & Discussion

2.6 Results & Discussion

- Present the results/findings of your study clearly. Use tables, figures and graphs to support your findings.
- Make sure you interpret your results correctly.
- **Tip!** For computer science research, compare your result with SOTA.

2.6 Results & Discussion



Outline	Study	Model/Methods	Accuracy	User-level explanation
Highlights				
Abstract	(Emmanuel et al., 2023)	MobileNet V2	0.901	No
Keywords				
1. Introduction	(Ahishakiye et al., 2023)	Ensemble of GLVQ, GMLVQ and LGMLVQ	0.82	No
2. Related work	(Kumar et al., 2023)	Ensemble of EfficientNet, SEResNeXt, ViT, DeiT and MobileNetV3	0.9075	No
3. Materials and methods	(Paiva-Peredo, 2023)	DenseNet169	0.7477	No
4. Experiment				
5. Results and discussion	(Chen et al., 2022)	Smooth-Taylor CE	0.893	No
6. Conclusion	(Ravi et al., 2022)	A_L_EfficientNet	0.8708	No
CRediT authorship contribution statement	(Riaz et al., 2022)	EfficiennetB3	0.8303	No
Declaration of competing interest	(Anitha & Saranya, 2022)	CNN with data augmentation	0.90	No
Acknowledgements				
Appendix.	The proposed approach	Fusion of DL and semantic technology	0.905	Yes
Data availability				
References				

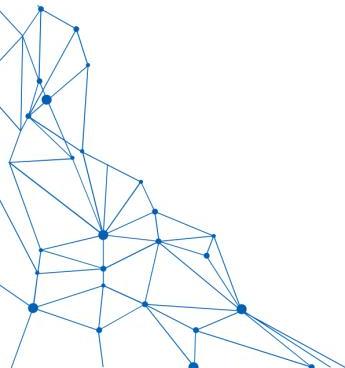
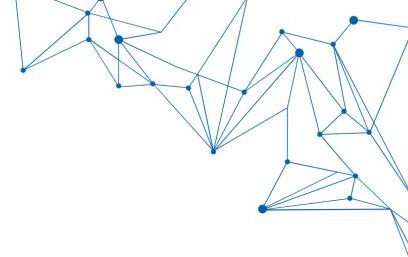
2.7

Structure of a Research Paper

Conclusion

2.7 Conclusion

- Restate research problem and objective as well as key findings but avoid repetitions.
- Explain significance and implications of the findings.
- Highlight the limitations and provide the future directions.



3. Publication Process

3 Publication Process

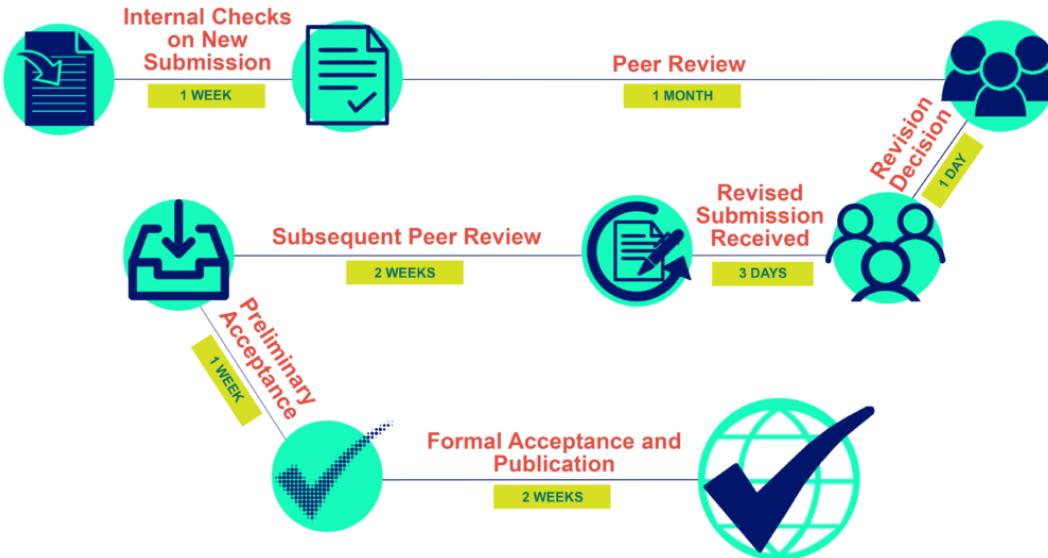
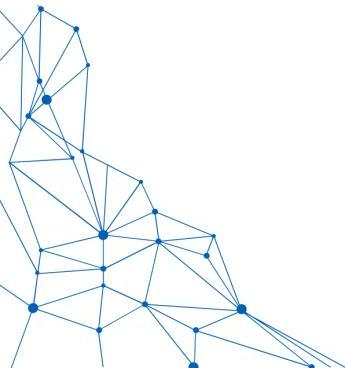
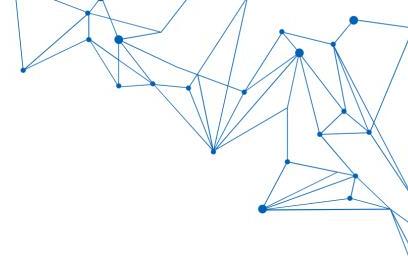


Figure 1: Publication process [4].



4.

Cover Letter for Article Submission

4. Cover Letter for Article Submission

- Should be concise, max 1-1.5 page.
- Use the formal and professional language. Address the letter to the editor in chief.
- Ensure clarity and the logical flow.
- Emphasize impact and novelty of your research, i.e., why your research is suitable for that particular journal or why should it be published.
 - Note: Cover letter is used by editor to screen the submission.

Avoid redundant or unnecessary details.

4. Cover Letter for Article Submission

- Cover letter for initial submission.

Date: 01.30.2023

Dear Prof. Michela Taufer,

I am writing to submit our manuscript, “*Enabling privacy-aware interoperable and quality IoT data sharing with context*,” as a research article in the journal, *Future Generation Computer Systems*. Our research addresses the key challenges including the data quality and accuracy issues in IoT (Internet of Things) data sharing of a heterogeneous and multi-source sensor data in a complex environment like smart cities. Moreover, our work focuses on emerging edge and fog computing paradigms that are currently being deployed in smart cities. In addition, this work considers privacy, specifically GDPR (General Data Protection Regulation). For this, a new framework is proposed. The following are the highlights of the proposed framework:

- The framework facilitates interoperability and enables the sharing of high-quality IoT data as per defined quality criteria, with the guarantee that the accuracy of the shared data can be determined.
- Our findings—an average CPU (central processing unit) utilization, at the edge and fog, respectively, of 33.61% and 15.62% for analytics operations, and 12.35% and 1.24% for data transformation and migration operations; and (ii) 51 megabytes of memory utilization—from the performance evaluation demonstrate the practicality of the proposed framework for resource-constrained IoT devices with low CPU and memory overhead.
- Our findings in data quality and accuracy evaluation from 180 different observations from edge and edge/fog experiment confirm that the proposed framework can ensure the quality criteria and also maintain data accuracy.
- A new trust metric is proposed as intrinsic to the shared data.
- The use of semantic technology allows easy integration of multi-source and heterogeneous sensors data while maintaining other aspects of the data, like quality.

I would like to declare that the authors have no conflicts of interest and that all authors have approved the manuscript and agree with its submission to *Future Generation Computer Systems*.

I would like to confirm that this work is original. Additionally, I would like to confirm that the manuscript is not being considered for publication elsewhere.

Kindly direct all correspondence regarding this manuscript to me (corresponding author) at Tek-Raj.Chhetri@uibk.ac.at.

Best Regards,
Tek Raj Chhetri (corresponding author)
Semantic Technology Institute (STI), Department of Computer Science, University of Innsbruck, Innsbruck, Austria

4. Cover Letter for Article Submission

- Example cover letter for revised submission.

Date: 09.09.2023

Dear Prof. Michela Taufer,

I am writing to submit our revised manuscript, “*Enabling privacy-aware interoperable and quality IoT data sharing with context*,” as a research article in the journal, *Future Generation Computer Systems*. The following are the major changes made in response to the reviewer comment.

- We have updated and restructured the introduction into newer sections: background, motivation, and contributions.
- The manuscript is updated to include the research questions.
- Added a new section to include the functional and non-functional requirements.
- Added a new section describing our evaluation criteria.
- Updated the evaluation section to include evaluation about the interoperability and analytics operation.

I would like to declare that the authors have no conflicts of interest and that all authors have approved the manuscript and agree with its submission to *Future Generation Computer Systems*.

I would like to confirm that this work is original. Additionally, I would like to confirm that the manuscript is not being considered for publication elsewhere.

Kindly direct all correspondence regarding this manuscript to me (corresponding author) at Tek-Raj.Chhetri@uibk.ac.at.

Best Regards,

Tek Raj Chhetri (corresponding author)

Semantic Technology Institute (STI), Department of Computer Science, University of Innsbruck, Innsbruck, Austria

(On behalf of the co-authors)

5.

Revising Manuscript

5. Revising Manuscript

- Read through the reviewers comment carefully and revise the manuscript accordingly.
 - In case of the major revision, you must address all reviewers comment/concerns.
- While submitting revised article, you should follow the same process as the initial submission, e.g., submitting cover letter.
- Be constructive even if you disagree with reviewer. Being angry or rude will not help and is not professional. Use the formal and professional language.

Not addressing reviewer comment can lead to **rejection** of the paper.

6.

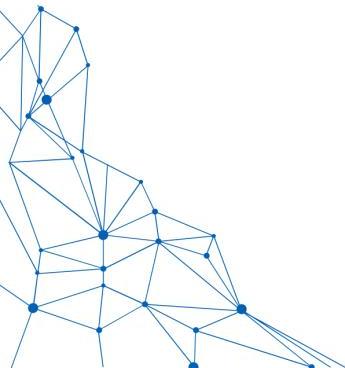
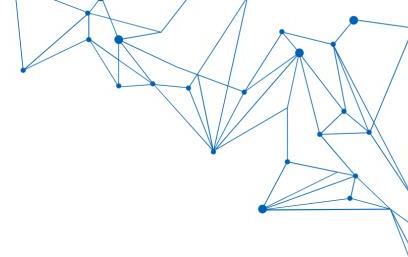
Rejection

6. Rejection

- Desk rejection: The manuscript is rejected by editor even before sending it to the reviewer.
 - Cause: Submitted article doesn't fit journal scope or is not prepared according to guidelines.
- Poor writing (or presentation): The article doesn't convey the message properly due to
 - Lack of clarity, coherence, or organization, e.g., poorly structured abstract, introduction.
 - Grammar, spelling, and formatting issues, e.g., figures are not readable.

6. Rejection

- Methodological flaws, e.g., inappropriate methodology or lack of validation.
- Insufficient results or the results are not convincing enough to support the research questions.
- Unclear research questions, i.e., problem the paper is addressing and lack of novelty?
- Not addressing reviewers comment.
- Ethical issues, e.g., plagiarism or not following other compliance guidelines when research involves human or animal subjects.



7.

Where to Publish?

1. Where to Publish?

- Publications are important, however, if they are not published in right place, it means nothing.
- Always publish in good places, either conferences or journals that are reputable and are not predatory.

- For journals check

<https://www.scimagojr.com/journalrank.php>

- For conference check <http://www.conferenceranks.com>

- https://scholar.google.com/citations?view_op=top_venues

The image shows a digital copy of a research paper titled "Predatory Journals: What They Are and How to Avoid Them" by Susan A. Elmore and Eleanor H. Weston. The journal is "Toxicologic Pathology", volume 40(8), pages 1029-1030. The DOI is 10.1177/0192633310392039. The article discusses predatory journals and provides guidance on how to identify them.

Abstract
Predatory journals—also called fraudulent, deceptive, or pseudo-journals—are publications that claim to be legitimate scholarly journals but misrepresent their publishing practices. Some common forms of predatory publishing practices include falsely claiming to provide peer review, hiding information about article processing charges, misrepresenting members of the journal's editorial board, and other violations of copyright or scholarly ethics. Because of their increasing prevalence, this article aims to provide helpful information for authors on how to identify and avoid predatory journals.

Keywords
fake peer review, deceptive journal, predatory publishing, predatory journals, scholarly communications, publishing ethics

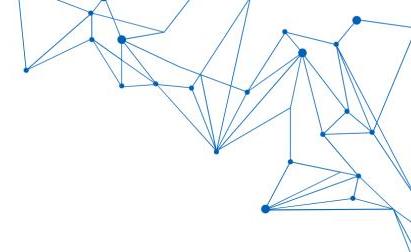
What Are Predatory Journals?
Beall, a librarian at the University of Colorado in Denver, is often credited as coining the term “predatory publishers,” which he described as organizations that “publish counterfeit journals to exploit the open-access model in which the author pays.” These predatory publishers are dishonest and lack transparency. They aim to dupe researchers, especially those inexperienced in scholarly communication.¹ Beall has hosted a list of publishers believed to be predatory on his blog, Scholarly Open Access, from 2012 to 2017.² Since the popular “Beall’s List” was taken down, several other tools have been created to make it easier to identify which journals are reliable and which are potentially predatory.^{3,4}

The main goal of predatory journals is profit. They attempt

make promises to publish anything the author submits, or promise review and publication on an unusually fast time line. It may be tempting for authors to submit to these journals, particularly if they are not sure if it may be a scam, or if they are in a hurry to be published.⁵ However, submitting articles to predatory journals may have serious negative consequences such as the ones listed below.

Fake Peer Review Undermines the Scientific Conversation
One of the most common forms of fraudulent publishing practices is for predatory journals to claim to provide true peer review when they do not. This practice leads to many problems in the broader context of scientific progress. Articles with flawed research or communication issues do not receive the

1. Where to Publish?



Predatory publisher

Tek Chhetri | Enable you to share research

LA

Lisa Arya, Editorial Manager <article-submit@journ4u.org>

Yesterday

To: Tek Chhetri

[Contemporary Mathematics, Real-time IF 2.15] Fast Publication with Special Discount



wayne@wiserpub.com via SurveyMonkey <member@surveym...

Thursday, January 23, 2025

J. Biomed. Res. Environ. Sci. | ISSN: 2766-2276

ICV 2022: 83.03 | NLM ID: 9918470779906676 | (IIF): 6.74 | DOI: 10.37871/jbres

Start Year: 2020 | Published Articles: 1000+

To protect your privacy, some external images in this message were not downloaded. [Go to Settings](#) [Download exte...](#)

[Contemporary Mathematics, Real-time IF 2.15] Fast Publication with Special Discount

Dear Author,

On behalf of the Editorial Board of Contemporary Mathematics (CM, <https://ojs.wiserpub.com/index.php/CM/>), I am pleased to extend an invitation to you and your colleagues to submit an original research article or a review paper to our journal. CM is an esteemed open-access publication by Universal Wiser Publisher (UWP, <https://wiserpub.com>).

We welcome contributions across a wide range of mathematical disciplines, including but not limited to:

- Algebra, Geometry and Topology
- Computational and Applied Mathematics
- Difference and Differential Equations
- Dynamical Systems
- Engineering Mathematics
- Financial Mathematics
- Functional Interpolation
- Fuzzy Sets, Systems and Decision Making
- Mathematical Biology
- Mathematical Physics

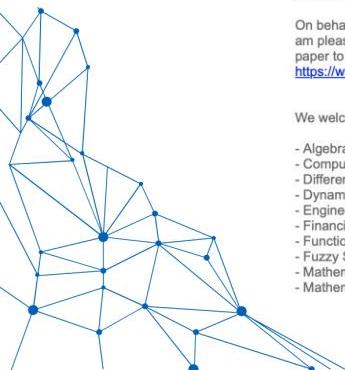
Dear Tek Chhetri,

We've made repeated attempts to reach out to you, but without a response.

We are eager to publish your article by February 14, 2025.. We hope you will consider contributing an article to our publication.

Elevate your research with professional publication services for just \$1199 (original price: \$2095).

If you are prepared to submit your paper, please attach it to this email. Please let us know if you accept this invitation.



8.

Best Practices

8. Best Practices

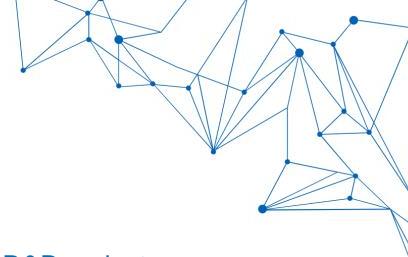
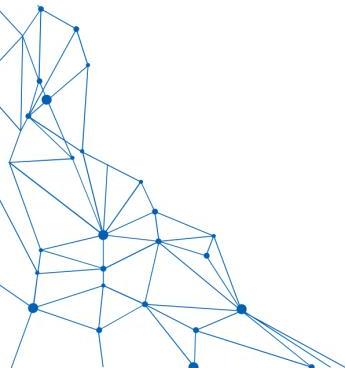
- Follow a funnel structure.
- Always check the guidelines, as some journals require specific section structures for organizing the paper.
- Always spell out acronyms the first time they appear before using their abbreviated form.
- Always proofread and make sure that paper is free from issues such as grammatical errors and typos.
- Make sure your methodology and methods are clear and understandable.

8. Best Practices

- Record your experimental setup and experiments in detail.
- For referencing follow the publisher guidelines, different journals use different styles, e.g., IEEE, APA.
- Make sure your references are complete, i.e., it includes doi, conference venue, time...
- Whenever possible try to compare with the SOTA.

Thank you!

References

- 
- 
- [1] Milutinović, V., Stankovic, S.V., Jovic, A., Draskovic, D., Misic, M. and Furundzic, D., 2017. A new course on R&D project management in computer science and engineering: subjects taught, rationales behind, and lessons learned. In Advances in Computers (Vol. 106, pp. 1-19). Elsevier.
 - [2] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *bmj*, 372
 - [3] Goundar, S., 2012. Research methodology and research method. Victoria University of Wellington, 1(1), pp.1-47.
 - [4] PLOS, n.d. Understanding the publishing process. Available at: <https://plos.org/resource/understanding-the-publishing-process/> [Accessed 31 Jan. 2025].