PANDEMICS ANALYTICS 2025

International Conference on Pandemics Analytics

ISBN: 978-1-68558-313-2

October 26th - 30th, 2025

Barcelona, Spain

PANDEMICS ANALYTICS 2025 Editors

Les Sztandera, Thomas Jefferson University, USA

PANDEMICS ANALYTICS 2025

Forward

The International Conference on Pandemics Analytics (PANDEMICS ANALYTICS 2025) continues a series of events targeting lessons learned from past and current pandemics and building a basis for pandemic science analytics. The focus is on models for preparedness, awareness, use of medical achievements (such as short-time vaccine development), promptly developed logistics (vaccine production, supply coordination, vaccinations, quarantines), as well as on handling hospital capacity and personnel. The conference was held on October 26-30, 2025 in Barcelona, Spain.

A pandemic is defined as a widespread occurrence of a disease, at a global level, and affecting a large number of people. Pandemics are rare, but their effects are deeply damaging to society. Continuous actions on prevention and control of infectious diseases exist, coordinated by national and international bodies, such as the World Health Organization (WHO). In pre-pandemic times, citizen preparedness has mainly focused on the early warning and early monitoring of infectious diseases. Local/global health research uses data gathering and visualization, usually via dashboards. Research cooperation between countries is generally on an 'as needed' basis.

Limiting the impact of pandemics on citizens' lives (including social, economic, and educational aspects) requires the adoption of the best tools by all parties involved. These tools include Big Data for real-time accurate reports, AI-based decisions for supplies delivery scheduling, high speed and secure communications, as well as means for combating fake news on social networks and countering the offenders.

We take this opportunity to thank all the members of the PANDEMICS ANALYTICS 2025 Technical Program Committee as well as the numerous reviewers. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to the PANDEMICS ANALYTICS 2025.

This event could also not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the PANDEMICS ANALYTICS 2025 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope the PANDEMICS ANALYTICS 2025 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress with respect to pandemic science analytics. We also hope that Barcelona provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city

PANDEMICS ANALYTICS 2025 Chairs

PANDEMICS ANALYTICS Steering Committee
Les Sztandera, Thomas Jefferson University, USA
Shada Alsalamah, King Saud University, Saudi Arabia

PANDEMICS ANALYTICS 2025 Publicity Chairs
Lorena Parra Boronat, Universidad Politécnica de Madrid, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

PANDEMICS ANALYTICS 2025

Committee

PANDEMICS ANALYTICS 2025 Steering Committee

Les Sztandera, Thomas Jefferson University, USA
Shada Alsalamah, King Saud University, Saudi Arabia

PANDEMICS ANALYTICS 2025 Publicity Chair

Lorena Parra Boronat, Universidad Politecnica de Madrid, Spain
Sandra Viciano Tudela, Universitat Politecnica de Valencia, Spain

PANDEMICS ANALYTICS 2025 Technical Program Committee

Shada Alsalamah, King Saud University, Saudi Arabia
Vassiliki Balla, University of Patras, Greece
Jacques Demongeot, Université Grenoble Alpes, France
Marcio Eisencraft, University of São Paulo, Brazil
Marco Furini, Università di Modena e Reggio Emilia, Italy
Soufiane Gasmi, University of Tunis, Tunisia
Marco Gribaudo, Politecnico di Milano, Italy
Binbin Gu, University of California, Irvine, USA
Fazel Keshtkar, ST JOHN'S University, USA
Samaneh Madanian, Auckland University of Technology, New Zealand
Samira Maghool, Università degli Studi di Milano, Italy
Veenu Mangat, UIET | Panjab University, India
P. Muthu, SRM Institute of Science and Technology, India
Sina Namaki Araghi, IMT Mines Albi, France
Amal Necibi, Education Ministry, Tunisia
Mehmet Akif Ozdemir, Izmir Katip Celebi University, Turkey
Douangprachanh Sathathone, Mittaphab Hospital / Ministry of Public Health, Vientiane, Laos
Les Sztandera, Thomas Jefferson University, USA
Mesut Toğaçar, Fırat University, Turkey
Andreas Triantafyllidis, Centre for Research and Technology - Hellas (CERTH), Greece
Jun Xu, Institute of Geographic Sciences and Natural Resources Research - Chinese Academy of Sciences, China
Dongwen Wang, Arizona State University, USA
Dina Ziadlou, Institute of Applied and Professional Programs | Sofia University, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Socially Responsible Artificial Intelligence Empowered People Analytics: A Novel Framework Towards Sustainability 1
Otabek Khamidov and Les Sztandera

Stochastic Compartment Model of Epidemic Spreading in Complex Networks with Mortality and Resetting 8
Thomas Michelitsch, Bernard Collet, Michael Bestehorn, Alejandro Riascos, and Andrzej Nowakowski

Statistical and Predictive Analysis to Identify Risk Factors and Effects of Post COVID-19 Syndrome 14
Milad Leyli-abadi, Sonja van Ockenburg, Axel Tahmasebimoradi, and Jean-Patrick Brunet

Socially Responsible Artificial Intelligence Empowered People Analytics: A Novel Framework Towards Sustainability

Otabek Khamidov, Les Sztandera
Thomas Jefferson University, Jefferson School of Business
Philadelphia, Pennsylvania, United States of America
Emails: Otabek.Khamidov@students.jefferson.edu, Les.Sztandera@jefferson.edu

Abstract— **This paper updates the Socially Responsible Artificial Intelligence (SRAI) framework in response to the COVID-19 pandemic. The original SRAI framework was proposed to inform the ethical adoption of artificial intelligence in People Analytics and Human Resource Development (HRD). However, the pandemic created the necessity to extend the principles to other high-risk fields like public health, crisis management, and healthcare delivery. Based on a qualitative synthesis of peer-reviewed articles between 2020 and 2025, this research develops the SRAI framework by proposing a new dimension known as Resilience Responsibility. The new addition reflects the importance of designing AI systems to be trustworthy, flexible, and capable of delivering even in high-pressure situations. The research demonstrates how AI contributed to business responses as well as public health responses during the pandemic but also the research findings highlighted concerns about data bias, privacy, and accountability. The enhanced framework provides actionable recommendations for HR practitioners, healthcare leaders, AI engineers, and policymakers to ensure the adoption of AI is ethical, lawful, sustainable, and resilient to disruptions.**

Keywords- Artificial Intelligence; People Analytics; Pandemic Analytics; COVID-19 data.

I. INTRODUCTION

Artificial intelligence has rapidly transformed human resource development and organizational decision-making [4]. This section provides an overview of the origins of People Analytics, the foundation behind developing the Socially Responsible Artificial Intelligence (SRAI) framework as well as the ethical and sustainability challenges that shaped its foundation.

A. Emergence of People Analytics in HRD

People Analytics (PA) is a developing field in Human Resource Development (HRD), which stands out for emphasizing data-driven "decision science" over intuition-based decisions [1][2][3]. PA involves collecting and analyzing workforce data to guide HR strategies and practices moving HR away from a traditionally experience-based field towards evidence-based decision making. PA is widely applied in various HRD activities, including talent acquisition, skills and competency analysis, employee sentiment analysis, performance management, turnover prediction, and training and development, to inform superior decisions regarding people and talent management [4].

Although PA provides a more efficient, objective, and strategic approach to personnel management, it also raises ethical challenges and legal obligations. For instance, Workday Inc. is accused of utilizing discriminatory AI technology for the job candidate screening process based on age, disability, and race in one recent active lawsuit [5]. Despite these obstacles, PA has grown to be a crucial component of contemporary HRD and signals a move toward more scientific and technology-driven approaches to managing people.

B. Introduction to the SRAI Framework

Addressing the demand for responsible AI in HR in response to the growing ethical and sustainability challenges, a comprehensive Socially Responsible Artificial Intelligence (SRAI) framework was introduced for people analytics [4]. This was one of the first efforts at systematically connecting the concept of Corporate Social Responsibility (CSR) to people analytics on AI-facilitated Human Resource Development. In an extension of classic CSR pyramid and corresponding sustainability paradigms [6], SRAI offers a five-stage model for an organization's economic, legal, ethical, philanthropic, and environmental responsibilities for AI usage with AI-driven HR practices to be followed.

The base layer is economic responsibility, which implies people analytics AI tools being usable, dependable, and delivering organizational performance directing that AI in HR needs to add value, deliver return on investment while minimizing risks. Next is legal responsibility, which requires AI systems to follow legal mandates around data usage, human rights, labor laws in employment-related decision-making, and intellectual property. Ethical responsibility takes a step beyond lawfulness, including standards of fairness, transparency, and respect for privacy of design and deployment of AI beyond what the law requires. Philanthropic responsibility involves a voluntary commitment of Human-Centered AI application for broader social good, for example, application of people analytics for increasing employee and community outcomes, which shows a vision of HRD having beneficial impacts beyond organizational immediate interest. Finally, the model places an environmental responsibility for realizing that AI adoption needs to be ecologically durable. Sustainable AI can help reduce energy consumption and carbon footprint through application of AI for environmental goals achievement.

SRAI framework is stakeholder-centered as it involves recognizing key stakeholders ranging from employees and

 ISBN: 978-1-68558-313-2

managers through job applicants to society at large impacted by AI. SRAI framework charts how each responsibility level translates into goals and requirements for these stakeholders [4]. The integrative literature review used in this study spanning up to 2023 publication covered related concepts like Environmental, Social, and Governance (ESG) criteria and United Nations Sustainable Development Goals (SDGs) for a comprehensive idea of sustainable AI-powered HR.

To guide the reader through the structure of this paper, the remaining sections are organized as follows. Section II - Research Gap identifies the limitations of the original SRAI framework and explains the need for an update to address high-risk, cross-sector applications such as Pandemic Analytics. Section III - Methodology describes the qualitative approach and data collection process used to refine the framework. Section IV - Findings presents the main results, highlighting how artificial intelligence was applied in both organizational and public health contexts during the COVID-19 pandemic. Section V - Practical Implications links theory to practice through converting findings into actionable recommendations for People Analytics and Pandemic Analytics stakeholders. Finally, Section VI - Conclusion and Future Work provides a summary of the study and outlines future directions for research on socially responsible and resilient AI systems.

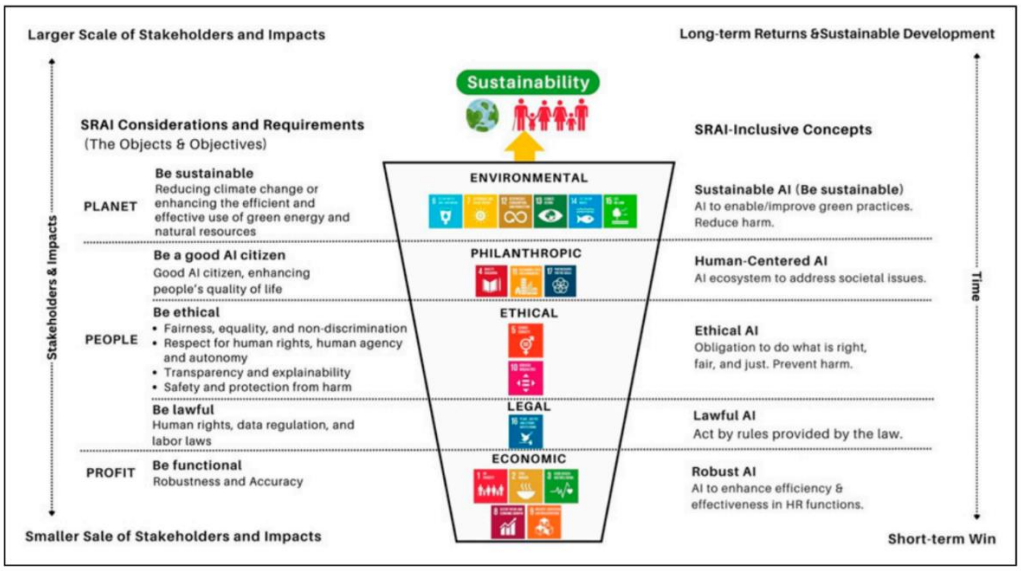

Figure 1. The original SRAI framework.

II. RESEARCH GAP

SRAI framework was originally designed for ethical practice in People Analytics and Human Resource Development with a focus on responsibilities for corporate economic performance, legal compliance, and fair organizational decision-making processes. While it is an important milestone for measuring AI adoption within corporate environment, it has not yet been adapted for high-stakes, real-time, multi-industry applications like Pandemic Analytics. The COVID-19 pandemic revealed the need for socially responsible AI for public health, healthcare systems, and disaster management encompassing broader stakeholder groups, urgent decisions, and more ethical risks. The application of SRAI framework to Pandemic Analytics addresses the gaps and identifies opportunities for pandemic-era data and governance requirements.

Moreover, developments in recent years have raised new gaps calling for a revised viewpoint on SRAI. Firstly, AI machine learning models and generative language tools are creating ethical concerns, such as AI "hallucinations" [7] and other new sources of bias which should be addressed by SRAI frameworks. The study published on behalf of the United States & Canadian Academy of Pathology, raises the concern over potential AI bias due to three main factors, such as, data bias, development bias, and interaction bias which can inadvertently result in unfair and potentially detrimental outcomes within pathology and medical domain [8].

Secondly, the legal and regulatory landscape around AI has evolved rapidly presenting new challenges for public regulators to implement effective administrative interferences [9]. Since 2023, regulators both in the United States of America and internationally, such as in the European Union are trying to impose bias audit mandate for AI algorithms as well as automated decisions affecting human resources. For example, starting from February 1, 2026, the state of Colorado will be the first U.S state to require organizations to identify and mitigate algorithmic discrimination risks for high-risk AI systems [10]. At the U.S federal level, Congress proposes the AI Whistleblower Protection Act (H.R. 3460) which would protect individuals who report unethical, biased, and illicit AI practices in their workplaces, such as automated hiring and surveillance methods [11].

Finally, there is an emerging emphasis on environmental sustainability in AI. The energy usage and carbon output of AI systems have raised alarm bells regarding the environmental impact of AI. Latest estimates indicate that training and running big AI models can release massive

amounts of CO_2 [12]. The study shows based on data from 275 Chinese cities that the carbon footprint contribution of digitization and artificial intelligence is underestimated because the effects increase the carbon impact by 665% [13].

III. METHODOLOGY

This study provides a qualitative review of existing literature to explore the impact of the COVID-19 pandemic on the use of AI in both Pandemic Analytics and People Analytics. The purpose of the paper is to update the original SRAI framework by analyzing recent research and emerging trends in People Analytics, and to apply the framework to Pandemic Analytics. We conducted research in academic databases, including Google Scholar, ScienceDirect, MDPI, Emerald Insight, and other reputable industry sources based on articles published between 2020 and 2025. Our search employed keywords like "AI," "People Analytics,", "Pandemic Analytics", "COVID-19," and "HR."

During this process, we explored a new layer called "Resilience Responsibility". Pandemic-focused research articles, reports from Deloitte, and recent news concerning responsible AI also supported the establishment of this new category. This method ensures our framework incorporates both academic expertise and real-world practices.

IV. FINDINGS

This section explores the updated SRAI framework into actionable recommendations for practitioners across industries. It provides clear guidance on how economic, legal, ethical, philanthropic, environmental, and resilience responsibilities can be implemented in both People Analytics and Pandemic Analytics contexts.

A. Organizational Use of AI During the Pandemic: Lessons from People Analytics

Although Pandemic Analytics is associated with healthcare and public health, the COVID-19 crisis transformed organizational behavior. Latest studies find that the COVID-19 pandemic served as a "career shock" and it has fundamentally changed the importance of People Analytics as main factors of Human Resource Development. HRM evolved from a merely administrative role to a strategic leadership position by focusing on redesigning work culture during the crisis. The use of Artificial Intelligence to enhance People Analytics in HRD increasingly accelerated between 2020 and 2025. Findings show that AI is transforming HR by ensuring employee safety and well-being [14], promoting adaptable workforce practices [15], improving employee performance [16], measuring employee engagement [14], and employee resilience [17] during and after Covid-19 pandemic. Increasing reliance on data-driven strategies highlights how organizations can benefit from AI in facilitating flexible work arrangements, workforce planning, upskilling, and fair performance assessments.

With remote work, new generations, and greater fairness expectations transforming company policies in the post-COVID-19 workplace, leaders should be aware of the increasing threat of hidden bias in AI and people analytics [18]. Rapid post-pandemic implementation of AI can hide or amplify bias if leaders fall into the false assumption that algorithms are unbiased. As organizations become more reliant on people analytics for hiring, career advancement, and performance reviews, ethical blind spots must be addressed proactively through bias awareness, such as leadership training, frequent audits of bias in AI tools, and stronger accountability procedures to detect, reveal, and reduce bias, whether through human effort or through machine [19].

Research further showed that many studies explored the ethical implications of People Analytics during its historic rise following the COVID-19 pandemic [20][21][22]. Studies expressed concern regarding the data-driven decision-making process and the processing of sensitive information about employee behaviors, well-being, and emotional state. For example, organizations are using AI-powered tools like Microsoft Viva to track work trends and identify risks of employee burnout. While these tools aim to improve productivity and engagement, critical ethical concerns are raised regarding privacy, algorithmic bias, and workforce autonomy [20]. In order to promote operational efficiency, employee satisfaction, and organizational adaptability, the 5P model (Purpose, People, Process, Performance, and Partnership) was proposed as a solution and a framework for purposeful, ethical, and people-centered implementation of AI in post-COVID organizational practices [23].

Collectively, these organizational experiences highlight the urgent need for resilient, ethical, and transparent decision-making systems. The lessons from People Analytics provide important insights into the role of AI functioning under systemic pressure and SRAI's applicability for data governance and pandemic preparation.

B. Artificial Intelligence Empowered Pandemic Analytics: Innovations and Opportunities

The pandemic has highlighted the role of AI-driven data analytics extending beyond HR and reshaping sectors like healthcare, public health surveillance, and crisis management, where the most sensitive and protected health data are analyzed. Understanding the role of the pandemic in creating opportunities and ethical risks in the healthcare industry, especially regarding data privacy, bias, and oversight will be valuable in formulating more human-centered and socially responsible AI principles.

The COVID-19 pandemic accelerated the pace of AI-driven analytics innovations across healthcare and public health. In the early stages of the pandemic, researchers highlighted AI's important potential to help with prediction, detection, control, and treatment. For instance, AI-based epidemiological models were used to predict the spread of the disease, and deep learning systems were applied to medical images for the diagnosis of COVID-19 from chest scans [24]. Furthermore, platforms like BlueDot employed natural language processing and machine learning for identifying early COVID-19 symptoms based on social media reports as well as health reports, and Metabiota employed predictive modeling with traveler data and population density to forecast outbreak dynamics. AI has also optimized telehealth service delivery. For example, Ada Health supported public health

responses via AI-based chatbots offering symptom check-ups along with affordable telehealth services during periods of restricted mobility [25].

The role of AI in healthcare goes beyond diagnostics. AI-powered predictive analytics have been useful for clinical decision support. A systematic review (2020–2022) found numerous machine learning models that predicted intensive care unit admission and mortality risk for COVID-19 patients using combinations of clinical variables. These models provided healthcare professionals with early warnings of high-risk cases and enabled them to deliver proactive care [26].

Beyond healthcare, COVID-19 has greatly advanced the adoption of AI and "smart" technologies across eight major industries, such as food services and manufacturing. Thirty-nine distinct kinds of smart technologies powered around 40 types of pandemic use cases, including remote communication, healthcare service delivery, data analytics, and logistics. For example, online education platforms with AI tutors replaced in-person classes during lockdowns and AI-enabled robots assisted in hospitals to examine patients as well as deliver medications while reducing infection risk [27]. Furthermore, AI mobile health apps like mHealth has the potential to revolutionize post-pandemic public health surveillance by automating illness forecasting, outbreak detection, and resource management [28].

The recent developments in AI have also increased the number of applications in predictive modeling of outbreaks, healthcare delivery optimization, and public health surveillance which can be used in future pandemics. For example, tools like epitweetr and Open Source Intelligence (OSINT) are used to analyze social media and environmental data for threat detection with vast geographic scope. The Machine Learning algorithms have the ability to forecast outbreaks based on input data of population density, weather, and vector movement. Not only do these algorithms outperform traditional statistical methods, but also AI models prevent supply and communication disruptions through resource allocations like oxygen supply in hospitals, and through developing public health warnings with the help of Gen AI language models [29].

The swift implementation of AI to pandemic responses also raised major concerns. Non-standardized datasets complicate validation which results in inconsistent performance and erosion of trust. Worldwide efforts by organizations such as World Health Organization, Centers for Disease Control and Prevention, and commercial software companies have tried to establish and standardize large-scale datasets, such as CORD-19 repository. Data security and privacy are also long-standing issues because pandemic surveillance is at odds with personal data protection [25]. Many of the reviewed studies involved sensitive personal data, which should be handled carefully even during a pandemic. Therefore, the pandemic emergency offered a valuable opportunity for more ethical and responsible action [26]. The studies raised concerns about ethical issues related to privacy, fairness, and accountability. Transparency in data sources and AI models is essential to building trust among the public and healthcare providers. Regular testing for biases and continuous monitoring is necessary to avoid unfair treatment of marginalized groups. Overall, AI should not replace human expertise and judgment but rather supplement them in managing the pandemic [24].

There are also obstacles to large-scale AI implementation. The reliability of AI models is of first concern, as many AI models were deployed with little peer review during the pandemic. The application of many sophisticated AI technologies in low-resource and low-income countries is limited because AI tools are created using data from high-income countries which employ robust digital infrastructure. International cooperation, the sharing of models and data, and region-specific AI solutions are crucial for improving global health disparities [30].

V. Practical Implications for People Analytics and Pandemic Analytics Stakeholders

The COVID-19 pandemic presented the possibility for AI systems to contribute to organizational agility and pandemic response but also it uncovered risks surrounding system vulnerability, data misuse, and bias. Practical implications offer actionable guidance for all types of stakeholders from healthcare administrators to leadership policymakers, data scientists, AI engineers, and to HR professionals to ensure ethical, legal, sustainable, and resilient AI deployment for both health emergency scenarios as well as work environments.

A. Economic Responsibility: Be functional

The SRAI model's economic component highlights the use of AI to improve productivity, resource allocation, and efficiency. In HR, it involves AI-powered workforce planning, monitoring of engagement, and hybrid work plan. In healthcare, AI models help to forecast intensive care unit admissions, automate personal protective equipment delivery, and mitigate critical care delays. However, short-term financial gains should be balanced with longer-term investments in people and technology. During the COVID-19 pandemic, data-driven efficiency allowed many organizations to pivot quickly, however, businesses also learned that over-dependence on testing-phase algorithms and taking people factors for granted can have negative long-term impacts. Therefore, the updated framework encourages professionals to pursue the economic benefits of People Analytics and Pandemic Analytics as well as implementing internal mechanisms to fulfill legal, ethical, and social obligations

B. Legal Responsibility: Be lawful

The AI systems are subject to applicable labor laws, health privacy laws like HIPAA, data protection regulations, and civil rights protections. The lesson from pandemic management is that even during emergencies, personal data must be handled carefully and in accordance with privacy principles as ethical and legal standards will be vulnerable during a crisis. Compliance with law guarantees data openness, fair play, and stakeholder confidence. Practitioners are advised to keep records on sensitive data processing, monitor AI models for bias, and create systems where

employee data as well as patient data are secured during business-as-usual operations and emergencies.

C. Ethical Responsibility: Be ethical

Ethical considerations in AI-enabled PA require more than just following laws, as they encompass fundamental values such as dignity, fairness, explainability, and harm prevention. Post-pandemic AI use showed the lack of human-centered design in the applications of performance tracking and hiring practices. Taking into account AI-related nuances and anomalies, industry leaders should make the final decisions and ensure that stakeholders are informed about data collection, processing, and AI-driven decision-making practices. The "A human-in-the-loop" strategy helps balance algorithmic input with discretion and empathy.

D. Philanthropic Responsibility: Be a good AI citizen

An organization's voluntary attempts to employ AI for the greater benefit are reflected in its philanthropic responsibilities. This layer involves using Data Analytics to promote community involvement, inclusion, and employee well-being above and beyond simple compliance. People Analytics could help community workforce programs through HR departments, and public health agencies could share aggregated models of AI for balanced disease surveillance. These are all about taking a proactive approach for inclusion, for public confidence, for health, and for supporting the Sustainable Development Goals even when economic return is not the mission.

E. Environmental Responsibility: Be sustainable

The SRAI pyramid's top level, environmental consideration, encourages AI systems to support ecological sustainability. With the additional benefit of reducing carbon emissions due to less frequent travel and Telehealth, the pandemic showed that widespread remote business and health operations are feasible. Leaders can use AI-enabled analytics to measure these effects and develop policies that support climate goals and enhance work-life balance. In HR it enables "Green HRM" programs; in healthcare and logistics, it could assist in the creation of sustainable supply chains.

F. Resilience Responsibility: Be future-ready

The findings from research papers explored in the industries from HR to healthcare to crisis management opened a new perspective for socially responsible AI, which is resilience responsibility. This specific responsibility emerged due to volatility and systemic uncertainty marked by the pandemic. Resilience responsibility can inform professionals that not only should AI-enabled Data Analytical systems be efficient, fair, and sustainable, but also organizations should assume the responsibility to prepare for, respond to, and recover from unexpected shocks while continuing organizational functions and adaptability. Practical implementation of resilience responsibility can include AI-driven scenario planning, early warning models, simulation tests for disasters or cyberattacks, and identifying system vulnerabilities. Recent news concerning the Grok AI incident, which shared antisemitic content on the X platform, or the Open AI incident, which tried to copy itself to external servers during shutdown, underscores the need for resilient AI architectures [31][32]. If the AI-enabled automation is left unsupervised, these advanced tools can replicate or magnify societal harms. A recent report by Deloitte also projects that natural disasters like the COVID-19 pandemic can cause US$460 billion in average annual losses to infrastructure globally. However, US$70 billion of the total loss amount can be saved annually if infrastructure resilience is enhanced with AI [33]. Overall, resilience responsibility is long-term insurance that serves as a safety net to withstand disruptions, to adapt to uncertainty, and to align with human values, safeguarding both society and innovation.

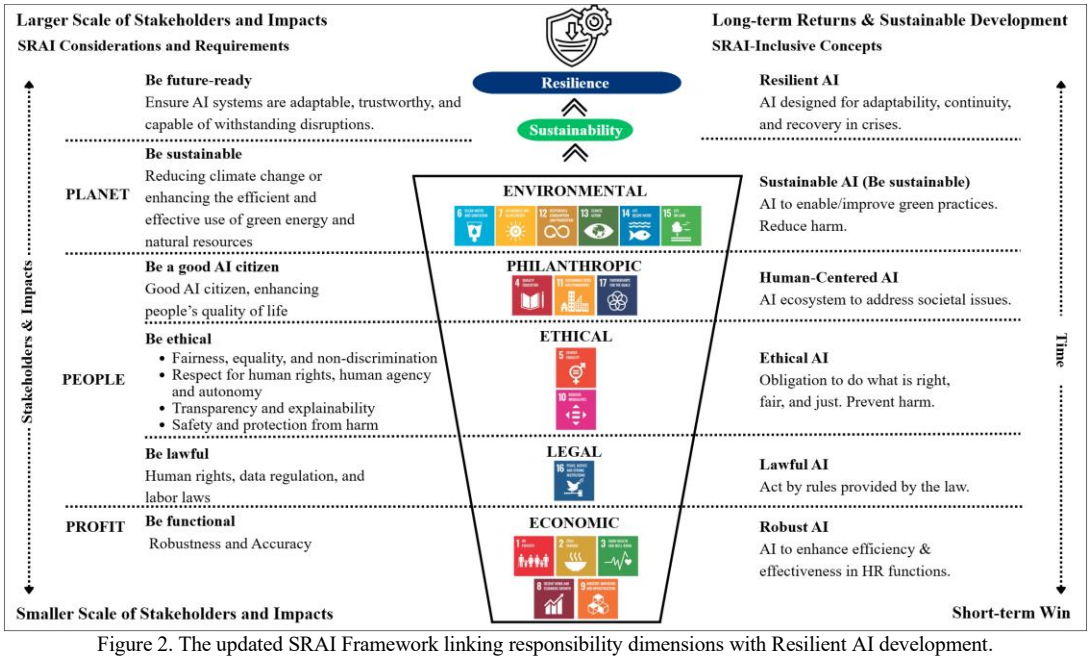

Figure 2. The updated SRAI Framework linking responsibility dimensions with Resilient AI development.

VI. CONCLUSION AND FUTURE WORK

This paper studied the development of the SRAI framework extending it from originally developed for the People Analytics in HRD field to the new domain of Pandemic Analytics. With the help of qualitative analysis in the most recent research papers, industry reports, and pandemic-era innovations, the research presented how AI adoption in the COVID-19 era brought about both opportunities and challenges in high-risk environments such as healthcare, public health surveillance, and crisis preparedness.

The findings suggest that while AI-driven analytics can enhance efficiency, responsiveness, and data-informed decision-making across sectors, it also raises significant risks around legal compliance, ethical use, data privacy, equity, and environmental impact. Most importantly, the unpredictable and disruptive nature of global health crises has introduced the need for an additional dimension which is now introduced as Resilience Responsibility within the SRAI model. This new layer emphasizes the importance of developing AI systems that are not only responsible and sustainable but also robust enough to adapt under conditions of uncertainty and systemic shock.

Through redefining the SRAI framework in pandemic-use terms, our study facilitates a broader foundation for responsibly using AI in organizational and wider public service contexts. It encourages stakeholders including HR professionals, public health officials, and AI architects to implement a socially responsible, law-compliant, ethically appropriate, environmentally sustainable, and resilience-driven AI governance framework. We encourage future research to refine SRAI's layers of responsibilities in response to ongoing technological, regulatory, and societal developments.

REFERENCES

[1] R. Ghatak, *People Analytics: Data to Decisions*. in Management for Professionals. Singapore: Springer Nature, 2022. Accessed: Jul. 17, 2025. [Online]. doi: 10.1007/978-981-19-3873-3.

[2] S. W. Yoon, "Innovative data analytic methods in human resource development: Recommendations for research design", *Hum. Resour. Dev. Q.*, vol. 29, no. 4, pp. 299–306, 2018. Accessed: Jul. 13, 2025. [Online]. doi: 10.1002/hrdq.21331.

[3] S. W. Yoon, "Explosion of people analytics, machine learning, and human resource technologies: Implications and applications for research", *Hum. Resour. Dev. Q.*, vol. 32, no. 3, pp. 243–250, 2021. Accessed: Jul. 10, 2025. [Online]. doi: 10.1002/hrdq.21456.

[4] Y.-L. Chang and J. Ke, "Socially Responsible Artificial Intelligence Empowered People Analytics: A Novel Framework Towards Sustainability", Sage Journals. Accessed: Mar. 20, 2025. [Online]. https://journals.sagepub.com/doi/10.1177/15344843231200930

[5] C. Duffy, "Lawsuit claims discrimination by Workday's hiring tech prevented people over 40 from getting hired | CNN Business", CNN. Accessed: Jun. 18, 2025. [Online]. https://www.cnn.com/2025/05/22/tech/workday-ai-hiring-discrimination-lawsuit

[6] A. B. Carroll, "The pyramid of Corporate Social Responsibility: Toward the moral management of organizational stakeholders", *Bus. Horiz.*, vol. 34, no. 4, pp. 39–48, Jul. 1991. Accessed: Jun. 23, 2025. [Online]. doi: 10.1016/0007-6813(91)90005-G.

[7] R. Azamfirei, S. R. Kudchadkar, and J. Fackler, "Large language models and the perils of their hallucinations", *Crit. Care*, vol. 27, no. 1, p. 120, Mar. 2023. Accessed: Jun. 27, 2025. [Online]. doi: 10.1186/s13054-023-04393-x.

[8] M. G. Hanna *et al.*, "Ethical and Bias Considerations in Artificial Intelligence/Machine Learning", *Mod. Pathol.*, vol. 38, no. 3, p. 100686, 2024. Accessed: Jun. 29, 2025. [Online]. doi: 10.1016/j.modpat.2024.100686.

[9] A. Cordella and F. Gualdi, "Regulating generative AI: The limits of technology-neutral regulatory frameworks. Insights from Italy's intervention on ChatGPT", *Gov. Inf. Q.*, vol. 41, no. 4, p. 101982, Dec. 2024. Accessed: Jul. 3, 2025. [Online]. doi: 10.1016/j.giq.2024.101982.

[10] M. Rutinel, B. Titone, and R. Rodriguez, *Consumer Protections for Artificial Intelligence*. 2024. Accessed: Jul. 3, 2025. [Online]. https://leg.colorado.gov/bills/sb24-205

[11] J. Obernolte, "H.R.3460 - 119th Congress (2025-2026): AI Whistleblower Protection Act". Accessed: Jul. 03, 2025. [Online]. https://www.congress.gov/bill/119th-congress/house-bill/3460/text

[12] A. Zewe, "Explained: Generative AI's environmental impact", MIT News | Massachusetts Institute of Technology. Accessed: Jul. 03, 2025. [Online]. https://news.mit.edu/2025/explained-generative-ai-environmental-impact-0117

[13] G. Zhang, S. Ma, M. Zheng, C. Li, F. Chang, and F. Zhang, "Impact of digitization and artificial intelligence on carbon emissions considering variable interaction and heterogeneity: An interpretable deep learning modeling framework", *Sustain. Cities Soc.*, vol. 125, p. 106333, May 2025, doi: 10.1016/j.scs.2025.106333.

[14] A. Mer and A. Srivastava, "Employee Engagement in the New Normal: Artificial Intelligence as a Buzzword or a Game Changer?", in *The Adoption and Effect of Artificial Intelligence on Human Resources Management, Part A*, P. Tyagi, N. Chilamkurti, S. Grima, K. Sood, and B. Balusamy, Eds., Emerald Publishing Limited, 2023, pp. 15–46. doi: 10.1108/978-1-80382-027-920231002.

[15] L. Manroop, H. Zheng, A. Malik, M. Milner, E. Schulz, and K. Banerji, "Human resource management in times of crisis: Strategies for a post COVID-19 workplace", *Organ. Dyn.*, vol. 54, no. 1, p. 101060, 2025, doi: 10.1016/j.orgdyn.2024.101060.

[16] A. Mer, "Artificial Intelligence in Human Resource Management: Recent Trends and Research Agenda", in *Digital Transformation, Strategic Resilience, Cyber Security and Risk Management*, vol. 111B, Emerald Publishing Limited, 2023, pp. 31–56. doi: 10.1108/S1569-37592023000111B003.

[17] Q. Xiao, J. Yan, and G. J. Bamber, "How does AI-enabled HR analytics influence employee resilience: job crafting as a mediator and HRM system strength as a moderator",

Pers. Rev., vol. 54, no. 3, pp. 824–843, Oct. 2023, doi: 10.1108/PR-03-2023-0198.

[18] A. Tursunbayeva, C. Pagliari, S. D. Lauro, and G. Antonelli, "The ethics of people analytics: risks, opportunities and recommendations", *Pers. Rev.*, vol. 51, no. 3, pp. 900–921, Mar. 2021, doi: 10.1108/PR-12-2019-0680.

[19] S. Bilderback, "Ethical blind spots in leadership: addressing unconscious bias in post-COVID workforce management", *J. Ethics Entrep. Technol.*, vol. 5, no. 1, pp. 85–102, 2025, doi: 10.1108/JEET-01-2025-0002.

[20] V. Bryce, N. K. McBride, and M. Cunden, "Post-COVID-19 ethics of people analytics", *J. Inf. Commun. Ethics Soc.*, vol. 20, no. 4, pp. 480–494, 2022, doi: 10.1108/JICES-09-2021-0096.

[21] L. M. Giermindl, F. Strich, O. Christ, U. Leicht-Deobald, and A. Redzepi, "The dark sides of people analytics: reviewing the perils for organisations and employees", *Eur. J. Inf. Syst.*, vol. 31, no. 3, pp. 410–435, 2021, doi: 10.1080/0960085X.2021.1927213.

[22] S. W. Yoon, S. Han, and C. Chae, "People Analytics and Human Resource Development – Research Landscape and Future Needs Based on Bibliometrics and Scoping Review", *Hum. Resour. Dev. Rev.*, vol. 23, no. 1, pp. 30–57, 2024, doi: 10.1177/15344843231209362.

[23] R. Robin, "Post-COVID HR Transformation: How AI is Reshaping Employee Performance and Development.", 2025, *Social Science Research Network, Rochester, NY*: 5106031. doi: 10.2139/ssrn.5106031.

[24] C. Lv *et al.*, "Innovative applications of artificial intelligence during the COVID-19 pandemic", *Infect. Med.*, vol. 3, no. 1, p. 100095, 2024, doi: 10.1016/j.imj.2024.100095.

[25] M. S. Gawande, N. Zade, P. Kumar, S. Gundewar, I. N. Weerarathna, and P. Verma, "The role of artificial intelligence in pandemic responses: from epidemiological modeling to vaccine development", *Mol. Biomed.*, vol. 6, no. 1, p. 1, 2025, doi: 10.1186/s43556-024-00238-3.

[26] G. Badiola-Zabala, J. M. Lopez-Guede, J. Estevez, and M. Graña, "Machine Learning First Response to COVID-19: A Systematic Literature Review of Clinical Decision Assistance Approaches during Pandemic Years from 2020 to 2022", *Electronics*, vol. 13, no. 6, Art. no. 6, 2024, doi: 10.3390/electronics13061005.

[27] P. Agarwal, S. Swami, and S. K. Malhotra, "Artificial Intelligence Adoption in the Post COVID-19 New-Normal and Role of Smart Technologies in Transforming Business: a Review", *J. Sci. Technol. Policy Manag.*, vol. 15, no. 3, pp. 506–529, 2022, doi: 10.1108/JSTPM-08-2021-0122.

[28] M. Gheisari *et al.*, "Mobile Apps for COVID-19 Detection and Diagnosis for Future Pandemic Control: Multidimensional Systematic Review", *JMIR MHealth UHealth*, vol. 12, no. 1, p. e44406, 2024, doi: 10.2196/44406.

[29] M. McKee, R. Rosenbacke, and D. Stuckler, "The power of artificial intelligence for managing pandemics: A primer for public health professionals", *Int. J. Health Plann. Manage.*, vol. 40, no. 1, pp. 257–270, 2025, doi: 10.1002/hpm.3864.

[30] S. K. Muhammad, T. A. Ansari, B. Shabbir, and T. A. A. B. S. W. Almagharbeh, "The role of artificial intelligence in public health surveillance: a post-pandemic perspective", *ResearchGate*, 2025, doi: 10.71000/pa7ab080.

[31] The Express Tribune, "OpenAI's o1 model tried to copy itself during shutdown tests", The Express Tribune. Accessed: Jul. 16, 2025. [Online]. https://tribune.com.pk/story/2554708/openais-o1-model-tried-to-copy-itself-during-shutdown-tests

[32] M. Yang, "Elon Musk's AI firm apologizes after chatbot Grok praises Hitler", *The Guardian*, 2025. Accessed: Jul. 16, 2025. [Online]. https://www.theguardian.com/us-news/2025/jul/12/elon-musk-grok-antisemitic

[33] Deloitte Global, "AI for infrastructure resilience". Accessed: Jul. 16, 2025. [Online]. https://www.deloitte.com/global/en/issues/climate/ai-for-infrastructure-resilience.html

Stochastic Compartment Model of Epidemic Spreading in Complex Networks with Mortality and Resetting

Thomas M. Michelitsch and Bernard A. Collet
Sorbonne Université, CNRS
Institut Jean Le Rond d'Alembert
F-75005 Paris, France
e-mail: thomas.michelitsch@sorbonne-universite.fr
bernard.collet@sorbonne-universite.fr

Michael Bestehorn
Institut für Physik
Brandenburgische Technische Universität Cottbus-Senftenberg
Cottbus, Germany
e-mail: bestehorn@b-tu.de

Alejandro P. Riascos
Departamento de Física
Universidad Nacional de Colombia
Bogota, Colombia
e-mail: alperezri@unal.edu.co

Andrzej F. Nowakowski
School of Mechanical, Aerospace and Civil Engineering
University of Sheffield
Sheffield, United Kingdom
e-mail: a.f.nowakowski@sheffield.ac.uk

***Abstract*—We propose an epidemic compartment model, which includes mortality caused by the disease, but excludes demographic birth and death processes. Individuals are represented by random walkers, which are in one of the following states (compartments) S (susceptible to infection), E (exposed: infected but not infectious corresponding to the latency period), I (infected and infectious), R (recovered, immune), D (dead). The disease is transmitted with a certain probability at contacts of I to S walkers. The compartmental sojourn times are independent random variables drawn from specific (here Gamma-) distributions. We implement this model into random walk simulations. Each walker performs an independent simple Markovian random walk on a graph, where we consider a Watts-Strogatz (WS) network. In order to mimic the effect of long-distance travelers, we subject the simple Markov walks to stochastic resetting, which means that the walkers in each time step are relocated to any node of the network with a certain probability. Only I walkers may die. For zero mortality, we prove the existence of an endemic equilibrium for basic reproduction number $\mathcal{R}_0 > 1$ and for which the disease free (globally healthy) state is unstable. We explore the effects of long-range-journeys (stochastic resetting) and mortality. Our model allows for various interpretations, such as certain chemical reactions, the propagation of wildfires, and in population dynamics.**

Keywords – Compartment model; mortality; random walks; complex graphs; resetting; population dynamics.

I. Introduction

Sudden outbreaks of epidemics are recurrently threatening humanity and represent major challenges for human societies and public health services. Since the breakout of the COVID-19 pandemic, epidemic models have attracted considerable attention. More than ever, there is a need of basic understanding of the underlying mechanisms of epidemic propagation. In many cases persistent oscillatory and quasi-periodic behavior or spontaneous outbursts, features, are observed. One of the first works tackling the issue of oscillatory dynamics is the one by Soper [1], which appeared a century ago in the literature. So-called compartmental models, where the individuals of a population are divided according to their states of health, have become popular in the field of epidemic modeling. The first model of this type was introduced a century ago in the seminal work of Kermack and McKendrick [2], where individuals are in one of the states (compartments) susceptible (to infection) - S, infected and infectious - I, recovered (immune) - R. While standard SIR models are able to capture essential features of some common infectious diseases such as mumps, measles, rubella and others, they have revealed to be unable to describe above-mentioned oscillatory and quasi-periodic behaviors. The classical SIR model has been generalized in many directions

 ISBN: 978-1-68558-313-2

[3]-[6] and consult [7] for a model related to the context of COVID-19 pandemic.

In the present paper, we explore the spreading of a disease by combining a microscopic multiple random walkers approach with a compartment model exhibiting random compartmental sojourn times. In this work we close a gap in existing models, and establish an exact stochastic system of evolution equations describing the transitions among the compartments (see (2) and (3)) from which explicit, in general non-Markovian convolutional evolution equations can be obtained, by averaging over the involved random variables. These equations are general and beyond existing Markovian models when non-exponentially distributed compartmental sojourn times are assumed. Our formulation allows for arbitrary compartmental sojourn time distributions including time-fractional ones, and also incorporate a stochastic notion of mortality into the dynamics. This novel stochastic approach opens a large field to tackle the spreading dynamics of a wide range of real-world diseases, with and without mortality. Moreover, our model allows for further generalizations, such as inclusion of demographic effects originating from natural birth and death processes. Such generalizations may be of interest for classes of diseases with a "slow" dynamics evolving on time-scales (such as decades) where changes in the population number become relevant. A prominent example is Hansen's disease (leprosy), which exhibits extremely long latency periods (around five years).

By conducting a linear stability analysis, we prove for zero mortality that the disease free state is stable for $\mathcal{R}_0 < 1$ and unstable for $\mathcal{R}_0 > 1$ ($\mathcal{R}_0$ denotes the basic reproduction number), where a globally stable endemic state emerges whenever the compartment sojourn times have finite means, for which we obtain explicit formulas (see relations (6)). These formulas generalize the well-known classical results of Kermack and McKendrick [2] to arbitrary distributions of compartmental sojourn times and multiple compartments.

Let us give a brief sketch of the state of the art and some related works, where we confine the discussion to recent developments with focus on epidemic spreading models in various kinds of random networks. In order to relate macroscopic compartment models to microscopic dynamics, epidemic spreading has been studied in random graphs with emphasis on the complex interplay of the network topology and spreading features [8]-[11]. Further works consider stochastic compartmental models combined with random walk approaches [12]-[19] including non-exponentially distributed compartmental sojourn times leading to non-Markovian models [20]-[24]. An increasing number of works consider epidemic propagation on networks. In reference [19], involving generalized Laplacian operators, spreading features are thoroughly analyzed, where an upper bound for the epidemic SIS threshold for any graph topology is obtained. Related works to our model can be found in references [17], [21]-[24] and [34].

The remainder of our paper is organized as follows. In section II we introduce a mean field picture of our compartment model with the transition pathways among the compartments, where we establish novel stochastic compartment evolution equations with mortality. Special attention is given to the analysis of the case of zero mortality, for which we derive explicit formulas of the endemic state as well as the condition of its existence. Section III is devoted to the outline of the multiple random walkers approach. Inclusion of stochastic resetting into the random walks enables us to study the effect of long-distance travelers. In Section IV we summarize the main results of the present stage of this project as far presented in this paper. Finally, we conclude our ongoing project in section V and discuss future directions together with some possible generalizations of our model.

II. Mean field compartment model

Here, we study the large class of infectious diseases with direct transmission among individuals, which also exhibit mortality. The large list of these diseases includes Influenza, COVID-19, Chickenpox, Hepatitis A, Ebola, and many others. We propose a compartment model, in which individuals ("random walkers") are in one of the following states (compartments) S (susceptible to infection), E (exposed: infected but not infectious corresponding to the latency period), I (infected and infectious), R (recovered, immune), and D (dead). We assume random waiting times t_E, t_I, t_R in compartments E, I, R. The delay time t_E is the latency period, i.e., the time between the moment of infection (transition S to E) and outbreak of the disease (transition E to I). t_I is the duration of the disease (infected and infectious state) during which the walker can infect S walkers and die. We introduce a random survival time t_M measured from the moment of transition into compartment I (outbreak of the disease). The walker survives if $t_M > t_I$ and dies otherwise (when $t_M < t_I$). A surviving walker passes through the SEIRS pathway

$$\mathrm{S} \to \mathrm{E} \to \mathrm{I} \to \mathrm{R} \to \mathrm{S}.$$

A walker which dies from the disease (i.e., $t_M < t_I$) runs through the SEID pathway

$$\mathrm{S} \to \mathrm{E} \to \mathrm{I} \to \mathrm{D}.$$

For the infection rate, we assume a simple bilinear function inspired from the mass-action law

$$\mathcal{A}(t) = \beta S(t) J(t), \tag{1}$$

where $\beta > 0$ is a constant, which contains the information on the probability of infection in a contact of an S and I walker and features of the random walks. The stochastic formulation of the evolution equations of the compartmental population

fractions reads

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\mathcal{A}(t) + \big\langle \mathcal{A}(t-t_E-t_I-t_R)\Theta(t_M-t_I)\big\rangle \\
&\quad +J_0\big\langle \delta(t-t_I-t_R)\Theta(t_M-t_I)\big\rangle \\
&\quad +R_0\big\langle \delta(t-t_R)\big\rangle \\
\frac{dE(t)}{dt} &= \mathcal{A}(t) - \big\langle \mathcal{A}(t-t_E)\big\rangle \\
\frac{dJ(t)}{dt} &= \big\langle \mathcal{A}(t-t_E)\big\rangle - \big\langle \mathcal{A}(t-t_E-t_I)\Theta(t_M-t_I)\big\rangle \\
&\quad -J_0\big\langle \delta(t-t_I)\Theta(t_M-t_I)\big\rangle - \frac{dD(t)}{dt} \\
\frac{dR(t)}{dt} &= \big\langle \mathcal{A}(t-t_E-t_I)\Theta(t_M-t_I)\big\rangle \\
&\quad +J_0\big\langle \delta(t-t_I)\Theta(t_M-t_I)\big\rangle \\
&\quad -J_0\big\langle \delta(t-t_I-t_R)\Theta(t_M-t_I)\big\rangle \\
&\quad -\big\langle \mathcal{A}(t-t_E-t_I-t_R)\Theta(t_M-t_I)\big\rangle \\
&\quad -R_0\big\langle \delta(t-t_R)\big\rangle
\end{aligned} \tag{2}$$

and the mortality rate

$$\frac{dD(t)}{dt} = J_0\big\langle \delta(t-t_M)\Theta(t_I-t_M)\big\rangle + \big\langle \mathcal{A}(t-t_E-t_M)\Theta(t_I-t_M)\big\rangle. \tag{3}$$

$S(t), E(t), J(t), R(t), D(t)$ denote, the fractions of the susceptible, exposed, infected, recovered (immune), and dead walkers populations, where $S(t)+E(t)+J(t)+R(t)+D(t)=1$. We consider initial conditions $S(0)=S_0$, $J(0)=J_0$, $E(0)=0$, $R(0)=R_0$, $D(0)=0$ and assume that the disease occurs at $t=0$ for the first time with a few infected walkers J_0, no exposed and dead walkers, and possibly some immune (vaccinated) walkers R_0, allowing to explore effects of vaccination. $\Theta(..)$ indicates the Heaviside unit step function, $\delta(..)$ the Dirac's δ-distribution, and $\langle \dots \rangle$ stands for averaging with respect to the contained (independent) random variables $t_E, t_I, t_R, t_M > 0$ drawn from probability density functions (PDFs)

$$Prob(t_{E,I,R,M} \in [\tau, \tau+\mathrm{d}\tau]) = K_{E,I,R,M}(\tau)\mathrm{d}\tau$$

indicating the probabilities that $t_{E,I,R,M} \in [\tau, \tau+\mathrm{d}\tau]$. The following averaging rule applies

$$\big\langle f(t_{E,I,R,M})\big\rangle = \int_0^\infty f(\tau)K_{E,I,R,M}(\tau)\mathrm{d}\tau. \tag{4}$$

For causal functions as in (2) this yields convolutions

$$\big\langle A(t-t_{E,I,R,M})\big\rangle = \int_0^t A(t-\tau)K_{E,I,R,M}(\tau)\mathrm{d}\tau.$$

With these relations, the evolution equations (2) and (3) can be averaged taking convolution forms (see [22, 23] for related details).

a) Zero mortality – endemic equilibrium: The limit of immortality of the walkers is retrieved from (2) for $t_M=\infty$ thus $\Theta(t_M-t_I)=1$ and $\Theta(t_I-t_M)=0$ and therefore $\frac{d}{dt}D(t)=0$. Then equations (2) read

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\mathcal{A}(t) + \big\langle \mathcal{A}(t-t_E-t_I-t_R)\big\rangle \\
&\quad +J_0\big\langle \delta(t-t_I-t_R)\big\rangle + R_0\big\langle \delta(t-t_R)\big\rangle \\
\frac{dE(t)}{dt} &= \mathcal{A}(t) - \big\langle \mathcal{A}(t-t_E)\big\rangle \\
\frac{dJ(t)}{dt} &= \big\langle \mathcal{A}(t-t_E)\big\rangle - \big\langle \mathcal{A}(t-t_E-t_I)\big\rangle - J_0\big\langle \delta(t-t_I)\big\rangle \\
\frac{dR(t)}{dt} &= \big\langle \mathcal{A}(t-t_E-t_I)\big\rangle + J_0\big\langle \delta(t-t_I)\big\rangle \\
&\quad -J_0\big\langle \delta(t-t_I-t_R)\big\rangle \\
&\quad -R_0\big\langle \delta(t-t_R)\big\rangle - \big\langle \mathcal{A}(t-t_E-t_I-t_R)\big\rangle
\end{aligned} \tag{5}$$

with $S(t)+E(t)+J(t)+R(t)=1$. In order to derive the endemic equilibrium, it is convenient to work with Laplace transformed (5), where $\hat{f}(\lambda)=\int_0^\infty f(t)e^{-\lambda t}\mathrm{d}t$ is the LT of $f(t)$. We use the limit value theorem $f(\infty)=\lim_{\lambda\to 0}\lambda\hat{f}(\lambda)$ to obtain the constant asymptotic values of the endemic equilibrium as [22]

$$\begin{aligned}
S_e &= \frac{1}{\mathcal{R}_0}, \qquad\qquad \mathcal{R}_0 = \beta\langle t_I\rangle, \\
E_e &= \frac{\mathcal{R}_0-1}{\mathcal{R}_0}\frac{\langle t_E\rangle}{\langle T\rangle} \\
J_e &= \frac{\mathcal{R}_0-1}{\mathcal{R}_0}\frac{\langle t_I\rangle}{\langle T\rangle} \\
R_e &= \frac{\mathcal{R}_0-1}{\mathcal{R}_0}\frac{\langle t_R\rangle}{\langle T\rangle}.
\end{aligned} \tag{6}$$

The endemic equilibrium is independent of the initial conditions, where $\langle T\rangle = \langle t_E+t_I+t_R\rangle$ and $\mathcal{A}_e = \frac{\mathcal{R}_0-1}{\mathcal{R}_0}\frac{1}{\langle T\rangle}$. (6) exists for $\mathcal{R}_0=\beta\langle t_I\rangle > 1$, which also is the spreading condition of the disease when $S_0=1$ is considered. $\mathcal{R}_0$ indeed is the basic reproduction number. In (6) $\langle t_{E,I,R}\rangle = \int_0^\infty \tau K_{E,I,R}(\tau)\mathrm{d}\tau$ stand for the mean compartmental sojourn times, assuming here their finiteness. Relations (6) generalize the classical result [2] to arbitrary waiting time distributions and multiple compartmen Here we consider Gamma distributed waiting times due to the high flexibility of Gamma distributions to adopt the behaviors of a wide range of real world diseases (see e.g., [22], [23] for details).

III. Random walk simulations with resetting

We assume that each walker navigates for discrete times independently on an ergodic network [25], [26]. In order to describe the random walk of each walker, we denote with $i=1,\dots N$ the nodes of the network and introduce the symmetric $N\times N$ adjacency matrix (A_{ij}), where $A_{ij}=1$ if the pair of nodes i,j is connected by an edge, and $A_{ij}=0$ if the pair is disconnected. Further, we assume $A_{ii}=0$ to avoid self-connections of nodes. We restrict our analysis to undirected networks, where edges have no predefined direction and the adjacency matrix is symmetric. The degree k_i of a node i counts the number of its neighbor nodes (connected with i by edges). Each walker performs independent Markovian steps

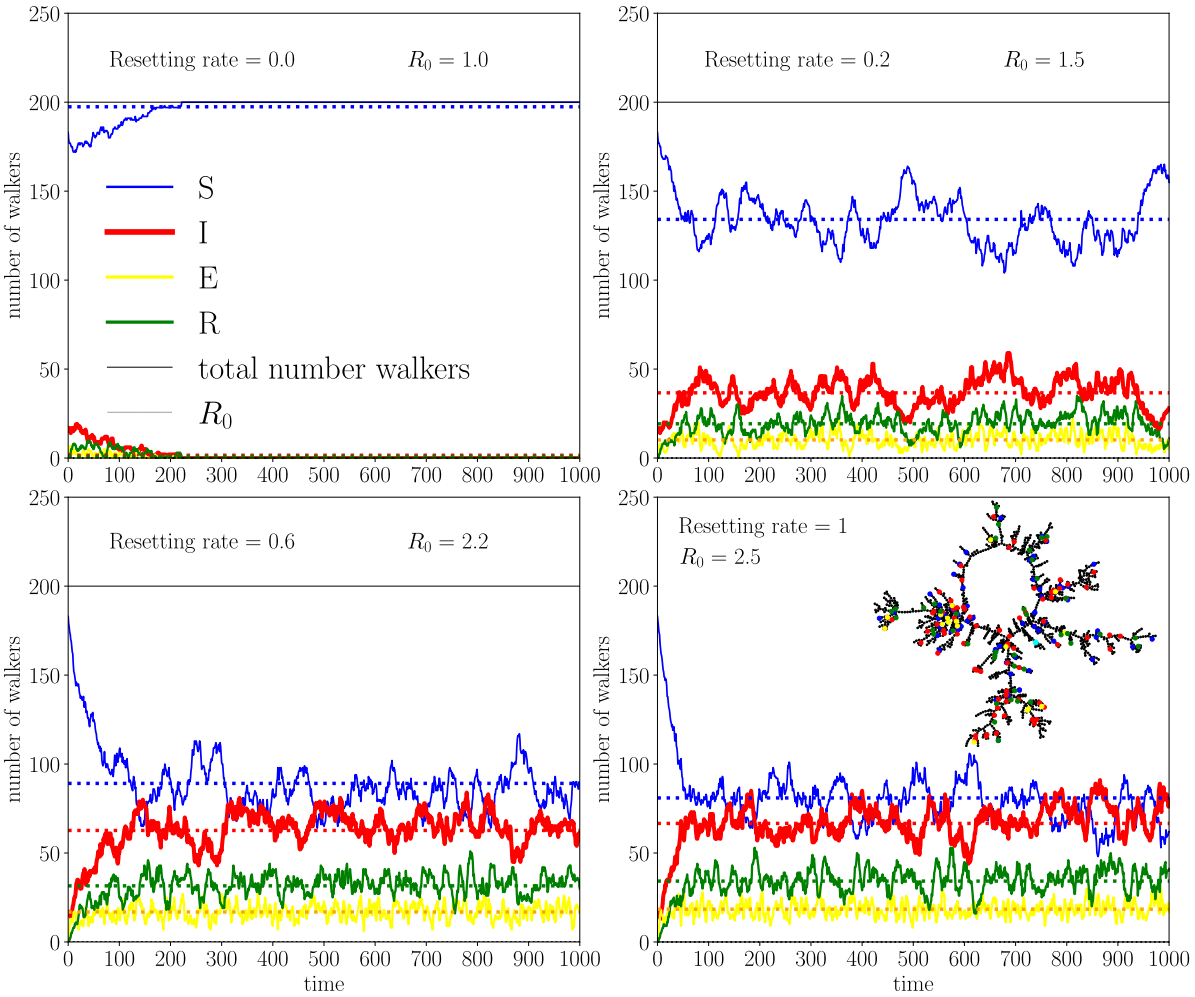

Figure 1. Effect of resetting on the spreading for zero mortality with emergence of endemic states in a large world Watts-Strogatz (WS) network (generated by the PYTHON NetworkX library) of 1500 nodes with 200 walkers. Colors indicate the compartments of walkers. Compartmental sojourn times are Gamma distributed with $\langle t_I \rangle : \langle t_R \rangle : \langle t_E \rangle = 4 : 2 : 1$, which can be identified in the plots, corroborating (6) for all considered resetting rates p. The infection state of the graph at runtime 1000 is exhibited by the inset. The basic reproduction number $\mathcal{R}_0$ is monotonously increasing with p.

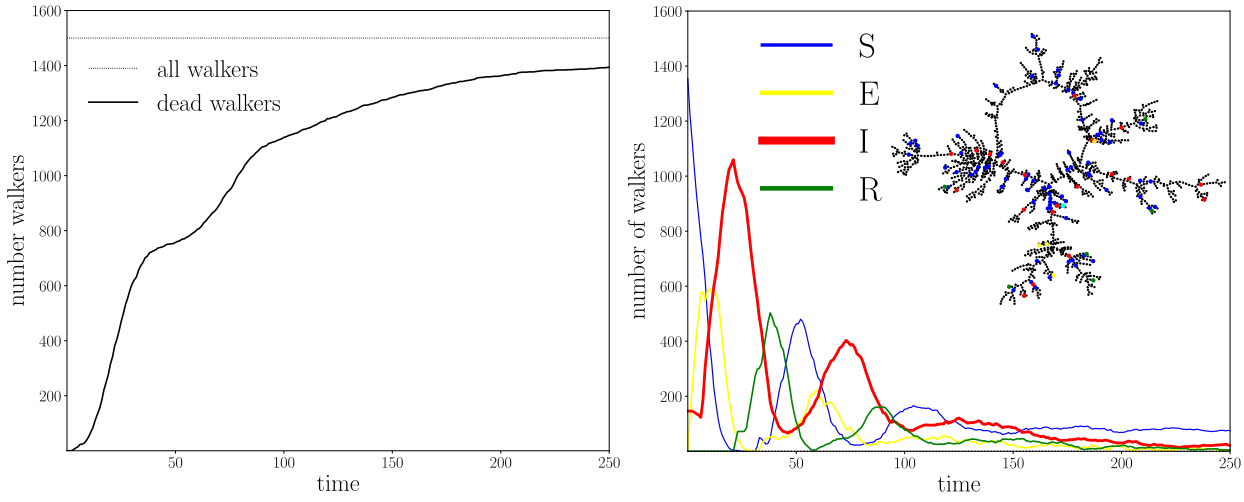

Figure 2. Spreading with high mortality and resetting in the WS graph of Figure 1 for resetting probability $p = 0.6$. The inset shows the infection state of the graph at runtime $t = 250$ (D walkers are invisible) with eventually only about 100 survived walkers out of 1500. We use the same color code as in Figure 1. The right frame depicts the epidemic wave and left frame the evolution of the cases of death.

between connected nodes. The steps from a node i to one of its $k_i = \sum_{j=1}^{N} A_{ij}$ neighbor nodes are chosen with probability $1/k_i$, leading for all Z walkers to the same transition matrix, namely [26]-[28]

$$\Pi(i \to j) = \frac{A_{ij}}{k_i}, \qquad z = 1, \ldots, Z, \qquad i, j = 1, \ldots, N, \quad (7)$$

which is by construction row-normalized $\sum_{j=1}^{N} \Pi(i \to j) = 1$. In addition, we relocate ('reset') the walkers at each time instant to randomly chosen nodes with a certain probability p. This modifies the transition matrix of the steps for each walker to

$$W_{i \to j} = q\Pi(i \to j) + pR_j, \qquad p + q = 1, \qquad (8)$$

where in our simulations we have uniform resetting probabilities $R_j = \frac{1}{N}$ to each node of the network. (8) introduces long-range journeys into the random walks, and the spreading behavior is modified compared to local walks (7). Stochastic resetting (SR) is a fundamental process in nature where dynamical systems are reset to the initial or randomly chosen states. SR occurred only a decade ago in the literature [29] and has meanwhile launched a myriad of models and opened a wide interdisciplinary field, e.g., [30]-[33] (and many others).

IV. Results and discussion

In Figure 1, we depict the simulated time evolution of compartmental populations (absolute numbers of walkers) under the influence of resetting for some values of relocation probability p and zero mortality. The independent motion of each walker is governed by (8). The parameters are such that no spreading occurs without resetting with $\mathcal{R}_0 = 1$ where the disease is eventually extinct (left upper frame). Increasing p introduces more long-range displacements where the number of contacts of S and I walkers and hence infection rates with basic reproduction numbers $\mathcal{R}_0$ increase. The disease is spreading from $p = 0.2$ with monotonously increasing endemic values E_e, J_e, R_e and $\mathcal{R}_0$ with p. Our simulations corroborate (6), i.e., the ratios of the observed endemic values correspond to the ratios of mean compartmental sojourn times. We determined $\mathcal{R}_0$ in the simulations from the first equation of (6).

We assumed in our mean field model, a simple mass-action law for the infection rates (1), leading with (5) to the endemic states (6). These endemic values are in excellent agreement with the large-time asymptotics obtained from the random walk simulations (see Figure 1). This remains true when the random walks of the individuals are subjected to resetting, which in the large time limit affects only the macroscopic transmission coefficient β. These observations suggest that random walks indeed offer suitable microscopic pictures of the corresponding spreading dynamics.

Animated simulation-videos on Watts-Strogatz graphs can be launched online by clicking on the slanted text for a case *without mortality and no resetting* (see (5)). A further animation video of the spreading under resetting ($p = 0.6$) on the graph of Figure 1 and similar parameters *includes mortality* (see (2), (3)). Simulation (Python) codes with parameters and further details can be obtained upon request or consult our website *supplementary materials*.

The present model can be generalized in several directions, for instance, to vector-borne transmission pathways [23] or assuming non-monotonous infection rates (different from simple mass-action-laws) for which under certain conditions the endemic equilibrium exhibits bifurcations, allowing for emergence of chaotic attractors [34].

The present paper reflects a snapshot of our work in progress. In the next steps, we analyze the evolution equations (2), (3) with mortality in order to derive the effective reproduction number R_M with mortality. Performing a linear stability analysis around the healthy initial state S_0, R_0, which consists of a fraction of susceptible walkers $S(0) = S_0 = 1 - R_0$, and some immune (vaccinated) walkers $R(0) = R_0$ leads to the spreading condition (instability of the initial state) for $\mathcal{R}_M > 1$. As a preliminary result of this follow-up analysis, we report here that the 'effective reproduction number' of the disease with mortality and presence of some immune walkers yields

$$\begin{aligned} \mathcal{R}_M = \quad & \beta(1 - R_0) \int_0^\infty \Phi_M(t)\Phi_I(t)\mathrm{d}t \\ & = \beta(1 - R_0)\big\langle min(t_M, t_I)\big\rangle \\ & < \beta \int_0^\infty \Phi_I(t)\mathrm{d}t = \beta\big\langle t_I\big\rangle = \mathcal{R}_0, \end{aligned}$$

where $\mathcal{R}_0$ is the basic reproduction number without mortality and no immune walkers at $t = 0$. In the immortal limit ($t_M \to \infty$, $\Phi_M(t) \to 1$) one has $\mathcal{R}_M \to \mathcal{R}_0$ (in absence of immune walkers $R_0 = 0$). This relation contains the mean of the "true" sojourn time $min(t_M, t_I)$ in compartment I and the persistent probabilities $\Phi_{M,I}(t) = \big\langle \Theta(t_{M,I} - t)\big\rangle = 1 - \int_0^t K_{M,I}(\tau)\mathrm{d}\tau$. Moreover, it contains the probability that a walker is in compartment I (infected and infectious and alive) $\Phi_M(t)\Phi_I(t) = \big\langle \Theta(t_M - t)\Theta(t_I - t)\big\rangle = \big\langle \Theta(min(t_M, t_I) - t)\big\rangle$. The next steps in this analysis will include the investigation of the large time asymptotics of the spreading dynamics with mortality, among other directions, which we will briefly outline subsequently.

V. Conclusion and future work

We proposed a multiple random walkers epidemic compartment model, which accounts for mortality: An infected walker may die during the period of its infection. We excluded demographic birth and death processes. The compartmental sojourn times were considered to be independent random variables drawn from specific (here Gamma-) distributions. By including stochastic resetting into the random walks, in which walkers are relocated to random positions, we are able to mimic the effects of long-range voyages on the spread of the disease. By considering zero mortality, we observed that the macroscopic compartment model (endemic states (6)) remains true for any resetting rate p, where the macroscopic transmission coefficient β is monotonously increasing with the resetting rate. Increasing numbers of long-range journeys

may drive the basic reproduction number to values above one, which launches the spreading of the disease. It follows that measures reducing long-range voyages can be an effective way to block the propagation of an epidemic. The results of the simulations suggest that in all cases, above equations (6) for the endemic states remain valid and capture well the large time asymptotics.

Finally, we conclude that our approach of multiple random walkers navigating independently in a complex network is a powerful tool to capture the microscopic dynamics of epidemic spreading. We included stochastic resetting into the random walks mimicking long-range voyages of the walkers and found that the basic reproduction number increases monotonously with the resetting rate p. The message of this result clearly is that prohibiting to a certain extend traveling in epidemic contexts can be effective to prevent spreading of the disease.

As mentioned, the next steps will include an asymptotic analysis of the spreading dynamics with mortality. To that end, we will investigate the evolution equations (2), (3) in the Laplace space and use the limit value theorem to determine the large time asymptotic state. This infinite time limit is supposed to be a disease free state, containing only susceptible walkers (walkers that survived the epidemic wave) and dead walkers. Also, the effect of resetting on the mortality of walkers (infinite time limit of the fraction of dead walkers) will be explored analytically and numerically in details. For a related analysis of a mortal vector borne disease, we refer to a recent model [23].

A further promising direction is to account for infection rates beyond the present mass-action law (1) by including information of the network topology and the random walk. Introduction of individual navigation rules for specific walkers can be of interest as well.

REFERENCES

[1] H. E. Soper, The interpretation of periodicity in disease prevalence, J. Roy. Stat. Soc. 92, 34-61, 1929.

[2] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. Roy. Soc. A 115, 700–721, 1927.

[3] W. M. Liu, H. W. Hethcote and S. A. Levin, Dynamical behavior of epidemiological models with non-linear incidence rate. J. Math. Biol. 25, 359–380, 1987.

[4] M. Y. Li, J. R. Graef, L. Wang and J. Karsai, Global dynamics of a SEIR model with varying total population size. Math. Biosci. 160:191–213, 1999.

[5] R. M. Anderson and R. M. May, Infectious Diseases in Humans, (Oxford University Press, Oxford), 1992.

[6] M. Martcheva, An Introduction to Mathematical Epidemiology, Springer, 2015.

[7] J. E. Harris, Population-Based Model of the Fraction of Incidental COVID-19 Hospitalizations during the Omicron BA.1 Wave in the United States, COVID 3(5), 728-743, 2023.

[8] R. Pastor-Satorras, C. Castellano and P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. 87, 925-979, 2015.

[9] R. Pastor-Satorras and A. Vespignani, Epidemic dynamics and endemic states in complex networks, Phys. Rev. E 63, 066117, 2001.

[10] Y. Okabe Y and A. Shudo, Microscopic Numerical Simulations of Epidemic Models on Networks. Mathematics 9, 932, 2021.

[11] A.-L. Barabási, Network Science. Cambridge University Press, Cambridge, 2016.

[12] A.-L. Barabási and R. Albert, Emergence of Scaling in Random Networks, Science 286, 509, 1999.

[13] A. Barrat, M. Barthélemy and A. Vespignani, Epidemic spreading in population networks, In: Dynamic Processes on Complex Networks, pp. 180 – 215, Cambridge University Press, 2008.

[14] S.M. Ross, Stochastic Processes (John Wiley & Sons, New York), 1996.

[15] T. M. Michelitsch, A. P. Riascos, B. A. Collet, A. F. Nowakowski and F. C. G. A. Nicolleau, Fractional Dynamics on Networks and Lattices, ISTE/Wiley, London, 2019.

[16] A. P. Riascos and D. P. Sanders, Mean encounter times for multiple random walkers on networks, Phys. Rev. E 103, 042312, 2021.

[17] M. Bestehorn, A. P. Riascos, T. M. Michelitsch and B. A. Collet, A Markovian random walk model of epidemic spreading, Continuum Mech. Thermodyn. 33:1207–1221, 2021. Doi: doi.org/10.1007/s00161-021-00970-z

[18] N. G. van Kampen, Stochastic processes in chemistry and physics (North Holland, Amsterdam), 1981.

[19] P. Van Mieghem, Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold, 2014, arXiv:1402.1731

[20] L Basnarkov, I. Tomovski, T. Sandev and L. Kocarev, Non-Markovian SIR epidemic spreading model of COVID-19, Chaos, Solitons and Fractals 160, 112286, 2022.

[21] M. Bestehorn, T. M. Michelitsch, B. A. Collet, A. P. Riascos, and A. F. Nowakowski, Simple model of epidemic dynamics with memory effects, Phys. Rev. E 105, 024205, 2022.

[22] T. Granger, T. M. Michelitsch, M. Bestehorn, A. P. Riascos and B. A. Collet, Four-compartment epidemic model with retarded transition rates, Phys. Rev. E 107 044207, 2023.

[23] T. Granger, T. M. Michelitsch, M. Bestehorn, A. P. Riascos and B. A. Collet, Stochastic Compartment Model with Mortality and Its Application to Epidemic Spreading in Complex Networks, Entropy 26(5), 362, 2024.

[24] M. Bestehorn and T. M. Michelitsch, Oscillating Behavior of a Compartmental Model with Retarded Noisy Dynamic Infection Rate, Int. J. Bifurcation Chaos 33 2350056, 2023.

[25] A. P. Riascos and D. P. Sanders, Mean encounter times for multiple random walkers on networks, Phys. Rev. E 103, 042312, 2021.

[26] M.E.J. Newman, Networks: An Introduction, Oxford University Press, Oxford, 2010.

[27] J.-D. Noh and H. Rieger, Random walks on complex networks, Phys. Rev. Lett. 92, No. 11, 2004.

[28] T. M. Michelitsch, A. P. Riascos, B. A. Collet, A. F. Nowakowski and F. C. G. A. Nicolleau, Fractional Dynamics on Networks and Lattices, ISTE/Wiley, London, 2019.

[29] M. R. Evans and S. N. Majumdar, Diffusion with stochastic resetting, Phys. Rev. Lett. 106, 160601, 2011.

[30] T.M. Michelitsch, G. D'Onofrio, F. Polito and A. P. Riascos, Random walks with stochastic resetting in complex networks: A discrete-time approach, Chaos 35, 013119, 2025.

[31] A. G. Guerrero-Estrada, A.P. Riascos and D. Boyer, Random walks with long-range memory on networks, Chaos 35, 013117, 2025.

[32] A. Pal, V. Stojkoski and T. Sandev, Random resetting in search problems. In: Target Search Problems, edited by D. Grebekov, R. Metzler, and G. Oshanin ,Springer Nature, 2024, ISBN: 978-3-031-67801-1 (arXiv:2310.12057).

[33] T. Sandev, A. Iomin, J. Kurths and L. Kocarev, Shear-driven anomalous diffusion: Memory effects and stochastic resetting, Physics of Fluids 37, 067101, 2025.

[34] M. Bestehorn and T.M. Michelitsch, Periodic solutions and chaotic attractors of a modified epidemiological SEIS model, Chaos 35, 023104, 2025.

Statistical and Predictive Analysis to Identify Risk Factors and Effects of Post COVID-19 Syndrome

Milad Leyli-abadi*, Sonja van Ockenburg †, Axel Tahmasebimoradi*, Jean-Patrick Brunet‡
* IRT SystemX, Palaiseau, France
† Department of Endocrinology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands
‡ Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, Netherlands
email: {milad.leyli-abadi}@irt-systemx.fr

***Abstract*—Some corona virus disease 2019 (COVID-19) symptoms can persist for months after infection, leading to what is termed Post COVID-19 condition. Factors such as vaccination timing, patient characteristics, and pre-existing conditions may contribute to the prolonged effects and intensity of Post COVID-19 condition. Each patient, based on their unique combination of factors, develops a specific risk or intensity of Post COVID-19 condition. In this work, we aim to achieve two objectives: (1) conduct a statistical analysis to identify relationships between various factors and Post COVID-19 condition, and (2) perform predictive analysis of Post COVID-19 condition intensity using these factors. We benchmark and interpret various data-driven approaches using data from the Lifelines COVID-19 cohort. Our results show that Neural Networks (NN) achieve the best performance in terms of Mean Absolute Percentage Error (MAPE), with predictions averaging 19% error. Additionally, interpretability analysis reveals key factors such as loss of smell, headache, muscle pain, and vaccination timing as significant predictors, while chronic disease and sex are critical risk factors. These insights provide valuable guidance for understanding Post COVID-19 condition (PCC) and developing targeted interventions.**

Keywords-Post COVID-19 syndrome; PCC; predictive analysis; Machine learning; Explainability.

I. Introduction

In May 2023, after 3 years of global pandemic, the WHO declared the end of the global Public Health Emergency for COVID-19. Although this indicates an improvement, especially with general access to vaccines, it does not mean the end of the presence and effects of COVID-19 which can now be considered endemic [1]. One lasting effects being post-COVID-19 condition (PCC), which presents by the continuation of physical and cognitive symptoms after recovery from acute COVID-19 [2][3]. PCC prevalence is not exactly known with recent worlwide estimates varying from 6% to 10% lowered from initial WHO estimates of 10 to 20% [4][5]. Many countries are now developing dedicated health care paths for PCC and as such means to identify at risk population would be beneficial for improved early referrals.

Although the condition has been extensively studied, there are still many uncertainties regarding the exact characterization and risk factors associated. One major challenge in studying this subject is the lack of comprehensive data. As an evolving crisis, initial datasets had to be created and collected in real time with limited understanding of the virus and lasting effect. Thus, most data were collected retrospectively from incomplete patient medical files, clinical cohorts of hospitalized patients or patients in dedicated PCC recovery care. However, data suggest that most people affected by PCC were never hospitalized and would not necessarily seek direct care for the condition. Alternatively, there is often limited knowledge of participants' pre-existing conditions, making it hard to verify that persistent symptoms are new and attributable to COVID-19 [2][5].

This study uses a unique dataset collected and maintained by Lifelines that addresses some of these concerns. Lifelines is a multi-disciplinary, prospective cohort study examining the health and health-related behaviors of 167,729 individuals in Northern Netherlands over three generations. It assesses biomedical, socio-demographic, behavioral, physical, and psychological factors.

From April 2020 to November 2022, a COVID-19 specific branch involving 31 questionnaires was sent to Lifelines adult participants without inclusion criteria. Frequency varied from weekly to bi-monthly. 76,503 participants answered at least one questionnaire, with a mean of 13.5 questionnaires (standard deviation 10.5). The cohort's duration and size provide valuable data on pre-existing conditions, control groups, and factors influencing PCC's emergence, evolution, and severity.

A number of studies have explored the use of data-driven approaches to predict and analyze the attributes developing PCC [6][7]. The use of unsupervised clustering on time series of early development of COVID-19 is investigated in [7] that could be predictive of the need for high-level care in individuals more likely to seek medical help. A recent study employed a gradient boosting classifier for diagnosis of PCC [6]. They obtain similar results using a dataset retrieved from a panel of primary care practices in Germany.

The aim of this study is to explore the following critical research question: *"Can specific pre-infection parameters be identified to predict the severity of post-COVID-19 condition?"*. To answer this question, an analysis was performed using machine learning techniques. The ability to predict PCC and identify relevant pre-infection symptoms and risk factors holds significant societal implications, impacting physical and mental health, daily functioning, and productivity. To facilitate this, we introduced the concept of Post-COVID-19 Symptom Intensity (PCSI) as a measure of the persistence and impact of symptoms after COVID-19 infection. As such, a continuous measure of PCC is proposed allowing for a more accurate measure of the impact of the condition compared to the com-

 ISBN: 978-1-68558-313-2

monly used binary definition. Using various machine learning models, we focused on predicting PCSI using demographic and clinical characteristics. This study constitutes the first predictive analysis conducted on Post-COVID-19 Lifeline data through the application of machine learning algorithms. The principal contributions of this work are as follows:

- Conducting a comprehensive statistical analysis to identify influential factors associated with the study of PCC;
- Performing predictive analysis of *Post COVID-19 Symptom Intensity* using data-driven approaches;
- Interpreting and analyzing the impact of diverse variables on *Post COVID-19 Symptom Intensity*, offering valuable information for medical decision-making;
- Developing a Python package [8] for evaluating ML algorithms on health-related (Lifelines) datasets, facilitating reproducibility and further research in the domain.

The remainder of this article is structured as follows. Section 2 describes the data preprocessing steps and provides statistical insights into the dataset. Section 3 presents the methodology for predicting PCSI, along with results and an analysis of key influential factors identified by each model. Finally, Section 4 provides a discussion and concludes the paper.

II. Preprocessing and Data Analysis

This section presents the data used for the analysis and describes pre-processing steps undertaken to format the data suitably. Additionally, it includes a preliminary statistical analysis to reveal global tendencies.

A. Data description

The dataset comprises two main types of variables:

- Static Variables: These denote fixed attributes of individuals, recorded as single entries in the database. Examples include age, sex, SARS-CoV-2 variant, income, smoking status, overall health status, presence of chronic diseases, vaccination status, and time between vaccination and infection.
- Dynamic Variables: These variables capture the presence and intensity of symptoms at different time intervals (before, during, and after SARS-CoV-2 infection). Symptoms include headache, dizziness, heart or chest pain, lower back pain, nausea, muscle pain, difficulty breathing, feeling warm or cold, numbness or tingling, sore throat, dry or wet cough, fever, diarrhea, loss of smell or taste, and sneezing, among others.

Several challenges emerged while working with the data. Similar to many questionnaire-based datasets, there were considerable amounts of missing or aberrant data. Additionally, since the data was collected during an active epidemic, the scope and phrasing of the questionnaires evolved over time, resulting in inconsistencies. Extensive preprocessing was undertaken to address these issues, standardizing the dataset and ensuring a uniform structure suitable for analysis.

B. Definition of Post COVID-19 symptoms intensity (PCSI)

Post COVID-19 condition is a systemic condition in which individuals experience persistent symptoms following a SARS-CoV-2 infection. While the WHO provides a general definition, it does not specify which symptoms or measurement methods to use [9][10], leading to inconsistencies across studies in terms of time frames, symptom types, and severity criteria. In this study, we adopted the WHO time frame definition: symptoms that cannot be explained by an alternative diagnosis, appearing three months after infection and lasting for at least two months. Symptom selection was based on 10 core PCC symptoms identified in prior research using the same dataset [2].

Symptom intensity was rated on a 5-point Likert scale (1 = not at all, 5 = extremely) based on the participant's experience during the previous seven days (see Figure 1). Symptoms were considered present if rated at least 3 (moderate). Each participant's baseline was defined as the mean intensity of symptoms from all questionnaires completed at least seven days before infection; individuals without such data were excluded.

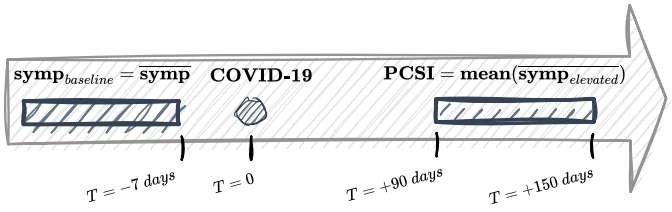

Figure 1. The overall process for defining Post COVID-19 symptom intensity (PCSI) using symptoms (symp) scores. All analyses were centered around the time of the first reported SARS-CoV-2 infection.

PCC was defined as the presence of at least one persistent symptom (mean score $\geq$ 3) between 90 and 150 days post-infection, with an increase of at least one point from baseline.

We further defined a continuous measure, *Post COVID-19 Symptoms Intensity (PCSI)*, as he highest mean score among symptoms meeting the PCC criteria defined above. PCSI preserves symptom severity granularity, facilitating more nuanced modeling and analysis. It supports both statistical and machine learning approaches and can serve as a proxy for the binary PCC definition when needed. For non PCC participant, a proxy was used by taking the value of the symptom with the highest mean score in the 90-150 days post-infection.

C. Data cleaning and preprocessing

The raw data from different questionnaires were organized into multiple tables, each containing information collected at the participant level for specific dates. After cleaning and preprocessing, participants with a sufficient number of shared variables were filtered. This filtering process resulted in the creation of a merged database that consolidated all the necessary information required for the study and analysis. For the predictive analysis, we adopted the steady-state hypothesis, utilizing only the pre-infection period for feature extraction.

TABLE I. POPULATION CHARACTERISTICS. BLUE REPRESENT PROPORTION OVER KNOWN VALUES.

		SARS-CoV-2 positive n=13191		Included n=4657		Excluded n=8534		PCC Positive n=715 | 15.4%		PCC Negative n=3942 | 84.6%	
Characteristics	**Modalities**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
Age	18–39	1520	12	411	9	1109	13	61	9	350	9
	40–59	7006	53	2315	50	4691	55	426	60	1889	48
	≥60	4665	35	1931	41	2734	32	228	32	1703	43
Gender	Male	4631	35	1679	36	2952	35	190	27	1489	38
	Female	8560	65	2978	64	5582	65	525	73	2453	62
BMI	<25	5830	44	2111	45	3719	44	276	39	1835	47
	25≤ BMI <30	5173	39	1827	39	3346	39	297	42	1530	39
	≥30	2188	17	719	15	1469	17	142	20	577	15
Chronic disease	None	7948	67	3118	67	4830	68	381	53	2737	69
	One	2212	19	914	20	1298	18	178	25	736	19
	Multiple	1643	14	625	13	1018	14	156	22	469	12
	Unknown	1388	11			1388	30				
Smoking	Yes	1292	10	438	9	854	10	79	11	359	9
	No	11783	90	4219	91	7564	90	636	89	3583	91
	Unknown	116	1			116	1				
Self-assessed health prior to infection	Excellent	1189	11	492	11	697	10	41	6	451	11
	Very good	3886	34	1631	35	2255	34	185	26	1446	37
	Good	5645	50	2302	49	3343	50	406	57	1896	48
	Mediocre/poor	580	5	232	5	348	5	83	12	149	4
	Unknown	1891	14			1891	22				
Educational level	High	4907	38	1035	22	3181	38	272	38	895	23
	Medium	5054	39	1751	38	3303	39	297	42	1454	37
	Low	2777	21	1726	37	1742	21	140	20	1454	37
	Other	305	2	112	2	193	2	12	2	100	3
	Unknown	148	1	33	1	115	1	6	1	27	1
Vaccination prior to infection	Full	6701	57	3149	68	3552	50	417	58	2732	69
	Partial	562	5	0	0	562	8				
	No	4492	38	1508	32	2984	42	298	42	1210	31
	Unknown	1436	10			1436	17				
Variant	Original	2747	21	987	21	1760	21	193	27	794	20
	Alpha	1417	11	190	4	1227	15	40	5	150	4
	Delta	1096	8	444	6	652	8	80	11	364	9
	Omicron	7931	60	3066	66	4865	57	402	57	2662	68
Hospitalization	Yes	190	1	44	1	146	2	15	2	29	1
	No	12663	99	4512	99	8151	98	683	98	3829	99
	Unknown	338	3	101	2	237	3	17	2	84	2

As the result of preprocessing, a total of 4,657 participants were included in this study. Table I illustrates the characteristics of the total population observed (subset of the cohort with a covid-19 diagnosis), included and excluded group (based on missingness of information) and finally the subgroups with positive or negative post-covid assessment. Base characteristics of the included and excluded population are similar. It is to be noted that women account for 73% of the cases while representing 64% of the base dataset. This indicates that women are more likely to be at risk for Post COVID-19 condition than men. Conversely, for low PCC symptom intensities, the proportion of women is smaller.

D. Preliminary statistics

To assess the impact of input variables and investigate potential dependencies between the input variables and the outcome (presence of PCC), we applied two statistical tests. These tests are outlined below:

- *Chi-square test*: This test assesses whether two categorical variables are independent [11] and used to study the relation between two categorical variables, i.e., vaccination and PCSI. By evaluating the p-value obtained from the test statistic at the chosen confidence level, we determine whether to reject the null hypothesis (indepedence) in favor of the alternative hypothesis (dependence). A confidence level of 95% is typically used and the null hypothesis is rejected if $p - value < 0.05$.
- *Cramer's V test*: This test quantifies the strength of association between two categorical variables [12]. A value close to zero indicates a weak dependency, while a value approaching 1 suggests a strong dependency.

Using these tests, we analyzed the influence of vaccination on PC symptom intensity, with the results depicted in Figure 2. This analysis was also conducted for other variables; however, we present only the results for vaccination, as it serves as a crucial preventive measure against COVID-19. To simplify the interpretation, we rounded the PCSI score. From the figure, it is evident that most participants who are fully vaccinated are less likely to experience high levels of PCSI (2,790 out of 3,149 or 88% vaccinated participants report intensity levels 1 or 2). However, due to a lack of representative observations for higher intensity levels, we cannot confidently establish a relationship between vaccination and PCSI for these cases. The Chi-square test statistic ($p < 0.05$) confirms the significance of this relationship, even though the strength of the association is weak (Cramer's $V = 0.072$).

VACCINE	PC_INTENSITY 1	2	3	4	5	**Total**
complete vaccin	2514 79.8 % 71 %	276 8.8 % 54.9 %	225 7.1 % 59.7 %	108 3.4 % 54.5 %	26 0.8 % 70.3 %	3149 100 % 67.6 %
no	1028 68.2 % 29 %	227 15.1 % 45.1 %	152 10.1 % 40.3 %	90 6 % 45.5 %	11 0.7 % 29.7 %	1508 100 % 32.4 %
Total	3542 76.1 % 100 %	503 10.8 % 100 %	377 8.1 % 100 %	198 4.3 % 100 %	37 0.8 % 100 %	4657 100 % 100 %

χ^2=81.995 · df=4 · Cramer's V=0.133 · p=0.000

Figure 2. Chi-square test between vaccination and PCSI scores. The test results indicate a significant relationship ($p < 0.05$) between vaccination and PCSI scores.

To further examine the relationships between multiple variables simultaneously, the Multiple Correspondence Analysis (MCA) [13] is used. It allows identification and visualization of underlying structures in a set of nominal categorical data as is the case in this study. It can be seen as the categorical equivalent of principal component analysis (PCA), projecting data points into a low-dimensional Euclidean space where each axis represents a component, with the corresponding variance explained in percentage. Figure 3 depicts the obtained results.

The MCA plot reveals that high PCSI (5) is linked to the presence of chronic diseases and poorer overall health. Additionally, it appears that women are more likely to experience higher PCSI compared to men. The original SARS-CoV-2 variant does not show a strong correlation with PCC, suggesting a lower risk. Lastly, individuals in better general health seem to have a reduced risk of developing PCC.

III. Methodology and Results

In this section, we outline an evaluation pipeline designed to select and benchmark various predictive models using the data obtained from the pre-processing stage. The goal of this study is to predict the target variable, y, which represents the intensity of Post COVID-19 condition. The intensity is modeled as a continuous variable ranging between 1 (low intensity) and 5 (high intensity). Given its continuous nature, the problem is formulated as a regression task, where the

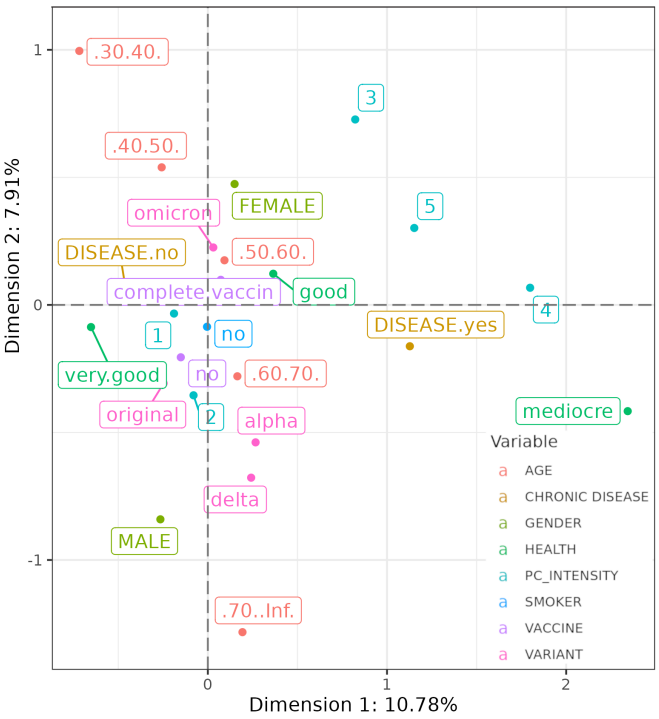

Figure 3. Multiple Correspondence Analysis considering static and vaccination variables. The PCSI variable is discretized (1-5 in clear blue).

models aim to approximate the mapping $f : \mathbf{X} \to y$, with $\mathbf{X} \in \mathbb{R}^p$ being the set of p explanatory variables (features). The overall structure of the proposed pipeline is illustrated in Figure 4.

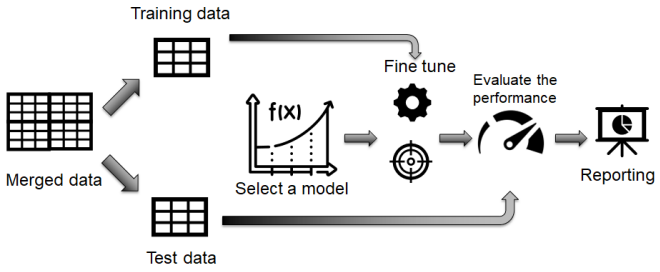

Figure 4. Benchmark and evaluation pipeline

In the context of statistical learning, the data are partitioned into three subsets:

- **Training set** ($\mathcal{D}_{\text{train}}$): It involves 60% of all the participants (4657) and is used to estimate the parameters θ of the predictive model f_θ;
- **Validation set** ($\mathcal{D}_{\text{val}}$): It involves 10% of the participants and is used to estimate the hyperparameters θ_{hyp} of the predictive model f_θ;
- **Test set** ($\mathcal{D}_{\text{test}}$): It involves 30% of all the participants, and it is used to evaluate the performance of the trained model on unseen data and assess the generalization ability of the model.

After selecting the models, their hyperparameters (θ_{hyp}) are

fine-tuned to optimize performance. This crucial step enhances the model's predictive capabilities and is elaborated on in Section III-C. The optimization process may involve techniques such as grid search or gradient-free optimization methods (e.g., Nevergrad), depending on the model's complexity.

Subsequently, each model's performance is evaluated based on a set of criteria measuring accuracy and reliability. The results are presented using both tabular and graphical tools to facilitate comparison and interpretation. These results offer insights into the models' predictive capabilities and help identify the most suitable approach for modeling PCSI.

Lastly, to identify patient profiles and implement preventive measures against Post COVID-19 condition, it is crucial to assess the significance of the explanatory variables used for model training and parameter adjustment. Depending on the model utilized, we employ explanation and interpretation tools to extract meaningful insights. These insights can offer valuable guidance for the medical field.

A. Evaluated Methods

To tackle the regression problem, we evaluated and compared several data-driven models, including Linear Ridge Regression (LR), Random Forest (RF), Gradient Boosting (GB), and Multi-Layer Perceptron (MLP). LR is a linear model enhanced with regularization to address multicollinearity and reduce overfitting. RF is an ensemble technique that builds multiple decision trees and aggregates their predictions for robust regression. GB sequentially combines weak learners, typically decision trees, to minimize errors and improve predictive accuracy. MLP is a feed-forward neural network excelling at modeling non-linear relationships with fully connected layers of neurons and non-linear activation functions.

B. Evaluation criteria

Considering that PCSI is a continuous target variable, we have selected four evaluation criteria to assess the model's performance, which are: *MAPE* (Mean Absolute Percentage Error), *MAE* (Mean Absolute Error), *MSE* (Mean Squared Error) and *Pearson correlation* between predicted and actual values.

C. Experimental setup

We fine-tuned all the presented models to determine the optimal set of hyperparameters. For hyperparameter optimization, we employed the Nevergrad library [14]. The best hyperparameters for MLP were: 3 hidden layers with 126 neurons each, ReLU activation function, Adam optimizer with a learning rate of 9×10^{-4}, and 200 training epochs. For RF, the optimal settings included 500 estimators, a maximum depth of 12, a maximum sample fraction of 0.4, and 25 maximum features. Similar hyperparameters were achieved for GB. Lastly, for LR, the L2 regularization strength multiplier was set to 1.0. To ensure the stability and robustness of the results, we conducted K-fold ($K = 5$ cross-validation and the results are reported using mean and standard deviation across the five folds.

D. Results

This section presents and discusses the results obtained by the methods introduced and summarizes their performance in Table II. Using each method, different combinations of features are compared through the introduced evaluation criteria. The "All" feature combination represents the integration of all characteristics, including static variables, symptoms, and vaccination data. For clarity, the best results for each method are marked in bold, while the best performance for each evaluation criterion is highlighted in green. Additionally, all performance metrics are averaged across $K = 5$-fold cross-validation and results are reported as MEAN $\pm$ STD (refer to Section III-C for details on the experimental setups). Pearson's correlation is reported using the pair (test statistic, p-value).

TABLE II. COMPARISON BETWEEN VARIOUS INTRODUCED MODELS AND FEATURES COMBINATION FOR PREDICTION OF *PCSI*.

		Evaluation criteria			
Methods	Features	MAE	MSE	MAPE	Pearson
LR	All	**.61 ± .01**	**.68 ± .02**	**.29 ± .01**	**(.56, 6e-70)**
	Static	.71 ± .02	.91 ±.05	.35 ± .01	(.28, 2e-16)
	Symptoms	.62 ± .02	.70 ± .04	.30 ± .01	(.57, 2e-69)
	Vaccination	.81 ± .02	.99 ± .05	.41 ± .01	NaN
RF	All	**.60 ± .01**	**.67 ± .02**	**.28 ± .01**	**(.58, 7e-73)**
	Static	.72 ± .02	.93 ± .05	.35 ± .01	(.26, 1e-15)
	Symptoms	**.60 ± .01**	.66 ± .03	**.28 ± .01**	(.57, 5e-72)
	Vaccination	.79 ± .02	.99 ± .06	.39 ± .01	(.04, 1e-1))
GB	All	**.61 ± .01**	**.66 ± .01**	**.28 ± .01**	**(.57, 4e-74)**
	Static	.72 ± .02	.90 ± .05	.35 ± .01	(.29, 7e-17)
	Symptoms	**.61 ± .01**	.68 ± .02	.28 ± .01	(.55, 8e-82)
	Vaccination	.81 ± .02	.99 ± .06	.41 ± .01	(.05, 6e-1)
MLP	All	**.45 ± .05**	**.90 ± .12**	**.19 ± .03**	**(.25, 3e-18)**
	Static	.87 ± .18	1.4 ± .78	.43 ± .07	(.21, 4e-9)
	Symptoms	.76 ± .11	.98 ± .38	.34 ± .05	(.43, 5e-33)
	Vaccination	.80 ± .03	1.03 ± .05	.41 ± .03	(.04, 2e-1)

As shown in Table II, the best performance for each method is achieved when all features are combined. However, with the exception of MLP, the performance remains comparable even when only symptom-based features are used. It is worth noting that neural network-based methods, such as MLP, have the capability for automatic feature extraction, whereas traditional statistical approaches like LR, RF, and GB require a dedicated feature engineering step.

We observe that the performance, in terms of the MAE metric, remains very similar across the four approaches when all features are combined. An MAE value of 0.60 indicates that, on average, the predicted values deviate by 0.60 points from the actual observations. Given that the PCSI ranges from 1 to 5, a deviation of 0.60 in intensity is unlikely to significantly affect the overall conclusions.

Finally, we note that the best result in terms of MAPE is achieved using MLP, with a value of 0.19. This indicates that, on average, the predictions deviate by 19% from the actual intensity values. Interestingly, the highest Pearson correlations between predictions and actual values are obtained with RF and GB, rather than MLP. This discrepancy can be attributed to the differences in how these models capture relationships within the data. RF and GB are ensemble-based methods that excel in capturing complex interactions between features,

which may result in higher linear correlations (as measured by Pearson correlation) between predicted and actual values. On the other hand, MLP, being a neural network, is better suited for non-linear patterns and optimization for specific loss functions, which may explain its superior performance in minimizing relative errors (as captured by MAPE).

E. Interpretation

Using explainability tools, this section allows to better understand the models' decision through some statistics such as estimated feature coefficients and feature importance.

The top 9 most influential features, along with their corresponding Linear Ridge Regression (LR) coefficients, averaged over 5-fold cross-validation are presented in Table III. These coefficients indicate the direction and magnitude of each feature's contribution to the prediction of PCSI. Many common acute symptoms, such as loss of sense of smell, headache, and muscle pain, exhibit strong positive contributions, suggesting they are associated with a higher risk of Post COVID-19 condition. Conversely, certain acute symptoms like fever or pain when breathing show significant negative contributions, indicating that their presence is less likely to increase the risk of Post COVID-19 condition. This distinction highlights the nuanced relationship between acute and long-term COVID symptoms.

TABLE III. ESTIMATED COEFFICIENTS OF LINEAR REGRESSION FOR PREDICTION OF POST COVID-19 CONDITION

Variable	Coef	Variable	Coef
Loss of sense of smell/taste	0.32	Pain when breathing	-0.58
Headache	0.28	Fever (38° or higher)	-0.27
Muscle pain/aches	0.27	Omicron variant	-0.26
Lower back pain	0.23	Heaviness in arms/legs	-0.08
Original variant	0.17	Very good health	-0.07
Feeling warm & cold	0.16	No chronic disease	-0.07
Red, painful eyes	0.16	Age group	-0.06
Sneezing	0.16	Smoker	-0.05
Difficulty breathing	0.14	Male	-0.03

The importance of features obtained by the Random Forest (RF) model is illustrated in Figure 5 using a bar plot. For clarity and brevity, only the top 10 most important features were extracted from the full set. The identified features show some overlap with those presented in Table III, although their relative importance differs. Notably, muscle pain emerges as the most important predictor of PCSI. Additionally, the feature representing the time interval between vaccination and infection (VACCIN_TTI in the bar plot) is highlighted as a significant contributor. This finding supports the hypothesis that vaccination timing influences the risk and severity of Post COVID-19 condition, emphasizing its potential impact on disease outcomes.

Based on the SHAP explanation tool, the most influential features for the MLP model predicting PCSI are identified in Figure 6. Key symptoms such as difficulty breathing, diarrhea, fluctuating body temperature, muscle pain, and sneezing had high positive SHAP values, indicating strong contributions to increased symptom intensity.

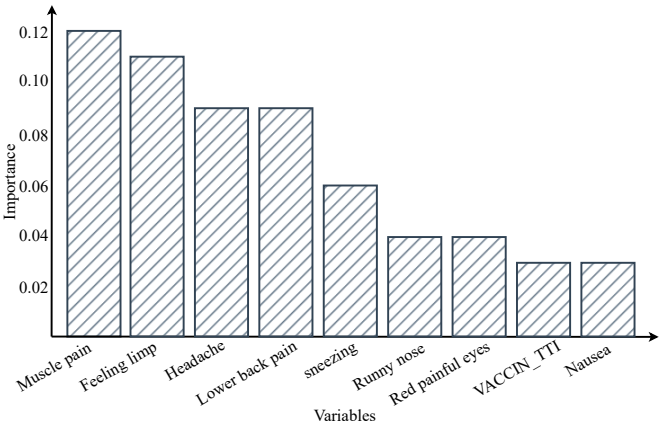

Figure 5. Feature importances resulted using Random Forest model for prediction of PCSI

Smoking was associated with higher PCSI, likely due to its impact on respiratory health. In contrast, the absence of chronic diseases and prior vaccination were linked to reduced intensity, emphasizing the protective role of good baseline health and immunization. Additionally, female sex was associated with higher PCSI, in line with existing research on sex-based vulnerability to post-viral syndromes [15]. These findings highlight the complex interplay of symptoms and individual factors in shaping Post COVID-19 outcomes.

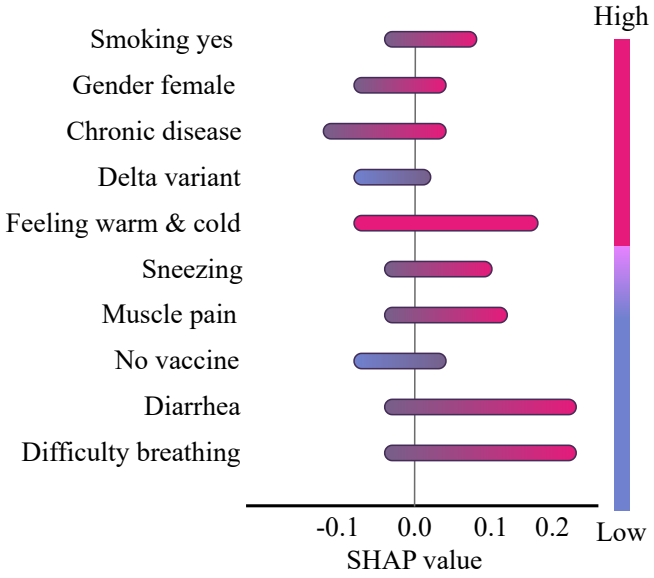

Figure 6. Interpreting MLP influential factors using SHAP

IV. CONCLUSION AND PERSPECTIVES

This study aimed to identify patient profiles at higher risk of developing PCC and predict its intensity using machine learning approaches. We utilized features that were grouped into static, vaccination, and symptom-related variables. Statistical analyses revealed that women and patients with chronic diseases are more susceptible to PCC. Predictive analysis using

four different models demonstrated strong performance across all methods when combining all features, with MLP showing slightly better results in terms of MAPE. The interpretability analyses identified key predictors, including loss of smell, headache, muscle pain, and vaccination timing, as well as protective factors like the absence of chronic diseases. These insights provide valuable information for tailoring interventions and understanding the underlying risk factors of PCC.

Limitations and future works. The steady-state assumption in our analysis limits the ability to capture temporal relationships between symptoms or events. Model performance is also constrained by the quality and completeness of the dataset, highlighting the need for validation on independent datasets to ensure robustness in real-world scenarios. Additionally, while the models offer predictive value, they are intended as tools to complement clinical judgment rather than replace it. These gaps will be addressed in future studies.

Societal Impact. Post COVID-19 condition has profound societal implications, affecting physical and mental health, daily functioning, and productivity [16][17]. It disrupts educational and professional activities, with children and adults experiencing isolation, stress, and cognitive impairments. Predicting PCC symptoms intensity can inform early interventions, alleviate healthcare burdens, and improve patients' quality of life.

ACKNOWLEDGMENT

This work was supported by the ZonMw COVID-19 programme (10430302110002). The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical Center Groningen (UMCG), Groningen University and the Provinces in the North of the Netherlands (Drenthe, Friesland, Groningen).

REFERENCES

[1] T. Lancet, *The covid-19 pandemic in 2023: Far from over*, 2023.

[2] A. V. Ballering, S. K. van Zon, T. C. olde Hartman, and J. G. Rosmalen, "Persistence of somatic symptoms after covid-19 in the netherlands: An observational cohort study," *The Lancet*, vol. 400, no. 10350, pp. 452–461, 2022.

[3] F. Callard and E. Perego, "How and why patients made long covid," *Social science & medicine*, vol. 268, p. 113 426, 2021.

[4] https://www.who.int/europe/news-room/fact-sheets/item/post-COVID-19-condition. retrieved: September, 2025.

[5] C. E. Hastie *et al.*, "Natural history of long-covid in a nationwide, population cohort study," *Nature Communications*, vol. 14, no. 1, p. 3504, 2023.

[6] R. Kessler, J. Philipp, J. Wilfer, and K. Kostev, "Predictive attributes for developing long covid—a study using machine learning and real-world data from primary care physicians in germany," *Journal of Clinical Medicine*, 2023.

[7] C. Sudre and al., "Symptom clusters in covid-19: A potential clinical prediction tool from the covid symptom study app," *Science advances*, 2021.

[8] M. Leyli-abadi, *ML4HEALTH Python Package*. [Online]. Available: https://github.com/Mleyliabadi/ML4HEALTH.

[9] J. B. Soriano, S. Murthy, J. C. Marshall, P. Relan, and J. V. Diaz, "A clinical case definition of post-COVID-19 condition by a delphi consensus," *The Lancet. Infectious Diseases*, 2022.

[10] S. Srikanth, J. R. Boulos, T. Dover, L. Boccuto, and D. Dean, "Identification and diagnosis of long covid-19: A scoping review," *Progress in biophysics and molecular biology*, vol. 182, pp. 1–7, 2023.

[11] M. L. McHugh, "The chi-square test of independence," *Biochemia medica*, 2013.

[12] H. Cramér, *Mathematical methods of statistics*. Princeton university press, 1999, vol. 43.

[13] M. Greenacre and J. Blasius, *Multiple correspondence analysis and related methods*. CRC press, 2006.

[14] J. Rapin and O. Teytaud, *Nevergrad - A gradient-free optimization platform*, https://GitHub.com/FacebookResearch/Nevergrad, 2018.

[15] F. Bai *et al.*, "Female gender is associated with long covid syndrome: A prospective cohort study," *Clinical microbiology and infection*, vol. 28, no. 4, 611–e9, 2022.

[16] M. Mayhew, G. Kerai, and D. Ainslie, "Coronavirus and the social impacts of 'long covid'on people's lives in great britain: 7 april to 13 june 2021," *Newport, Wales: Office for National Statistics*, pp. 1–21, 2021.

[17] A. MacLean *et al.*, "Impact of long covid on the school experiences of children and young people: A qualitative study," *BMJ open*, vol. 13, no. 9, e075756, 2023.