# Editorial on Machine Learning, AI and Big Data Methods and Findings for COVID-19

Victor Chang[1] · Carole Goble[2] · Muthu Ramachandran[3] · Lazarus Jegatha Deborah[4] · Reinhold Behringer[5,6]

## 1 Introduction

COVID-19 has become the most significant challenge human beings have encountered since World War 2 (WW2). In the USA, more COVID related deaths have been reported than the combined fatalities of the Pearl Harbor War and the September 11 terror attacks. COVID-19 itself is highly infectious and the speed at which it can mutate is rapid, leading to multiple varieties and strands of active coronaviruses, and rapid increases in the numbers of infected cases and deaths. The virus has infected more than 217 million of the population worldwide before the end of August 2021, yet in early March 2020, the total infected cases were still not reaching 100,000 (WHO, 2021).

The global challenge has changed how we live, with social distancing measures and face masks mandates, and has shone a spotlight on our healthcare capabilities, highlighting shortages of medical resources and capacity, and raising issues of equitable access to vaccines and drugs.

✉ Victor Chang
victorchang.research@gmail.com; V.Chang@tees.ac.uk

Carole Goble
carole.goble@manchester.ac.uk

Muthu Ramachandran
muthuram@ieee.org

Lazarus Jegatha Deborah
blessedjeny@gmail.com; deborah@aucet.in

Reinhold Behringer
reinhold.behringer@gmail.com;
reinhold.behringer@knorr-bremse.com

[1]  Teesside University, Middlesbrough, UK

[2]  University of Manchester, Manchester, UK

[3]  Integrated Cloud Solutions, Glasgow, UK

[4]  Anna University, Chennai, India

[5]  Knorr-Bremse GmbH, Munich, Germany

[6]  Leeds Beckett University, Leeds, UK

Governments of different countries have been challenged to deal with the crisis using limited resources and have emphasized different policies suiting their national strategies (Ecke, 2020; Ranney et al., 2020).

Machine Learning, AI and Big Data research, scientists can offer recommendations, new discoveries and predictive methods in addressing both societal impact (Gupta et al., 2018) and medical services (van der Sommen et al., 2020). Big Data and cloud computing (Chang, 2014; Hosseinian-Far et al., 2018), the Internet of Things (Sicari et al., 2016), and Artificial Intelligence techniques (Vaishya et al., 2020) have significantly contributed to developing an understanding of the virus, identifying and developing treatments and managing and tracking the social and economic impact of the pandemic, and predicting trends (Tuli et al., 2020). Scientists can better understand the structure and weaknesses of virus variants through simulations offered by ML algorithms, leading to vaccine development with higher efficacy (Rotondo et al., 2021). Challenges addressed range from COVID-19 diagnosis and infection confirmation based on a real-time reverse-transcriptase polymerase chain reaction (Kim et al., 2020) to privacy protection and secure transmission of messages among medical professionals (Wang et al., 2020).

## 2 Selection of Papers

We have sought unpublished and high-quality work based on unique Machine Learning, AI and Big Data methods and findings. Best paper winners and top authors from IIoT-BDSC 2020 have also been invited. We applied rigorous review processes, selecting six papers out of twenty-five submissions, which we briefly summarise.

The use of AI for patient screening and disease detection is addressed by two articles. Elakkiya et al. (2021) *demonstrate 'COVID_SCREENET: COVID-19 Screening in Chest Radiography Images Using Deep Transfer Stacking', which*
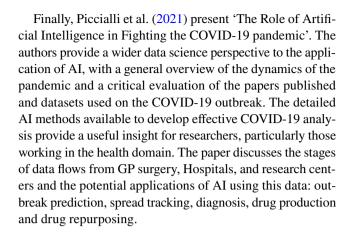
proposes bifold COVID_SCREENET, a novel software architecture to extract features of scanned images and use Deep Learning models that work under multi-class classification. Convolutional Neural Networks (CNNs) are used to analyze images and identify features for faster diagnostics, and the authors claim 100% accuracy in performing multi-class classification for 2216 chest X-ray images collected between April and May 2020.

Singh et al. (2021) develop 'A Novel Ensemble-based Classifier for Detecting the COVID-19 Disease for Infected Patients', which describes a method for detecting COVID-19 using a synthetic dataset with COVID-19 aimed at determining if a patient has the disease. 15 COVID-19 symptoms feed the classifier and the result is a decision about the patient's illness. A comparison to other classifiers shows improved accuracy on prior approaches. The approach is potentially applicable to other pandemic diseases and particularly promising for use in telemedicine, where direct access to patients is not possible.

Patient privacy cannot be sacrificed in the rush to analyze data. Zhang et al. (2021) address privacy issues arising from the training of Generative Adversarial Networks (GAN) to recognize COVID-19 pneumonia. Instead Federated Learning, which preserves the privacy of patients, is proposed as an alternative. Comparison of the algorithm's performance against other systems is encouraging for its use in contexts where patient privacy is important. The approach is already used across several hospitals for collective training of COVID-19 models without sharing the original data.

Two papers address the well-being of the population under the pressure of the pandemic. Both use Twitter as their cohort but use different classification techniques. Kaur et al. (2021) present 'A Proposed Sentiment Analysis Deep Learning Algorithm for Analyzing COVID-19 Tweets'. They analyze the mental state of twitter users by applying a Hybrid Heterogeneous Support Vector Machine (H-SVM) algorithm for accurate, rapid and large-scale Twitter sentiment analysis, classifying Twitter hashtags as positive, negative and neutral. The proposed algorithm measured by precision, recall, f-measure and accuracy metrics outperform Recurrent Neural Networks and Support Vector Machines.

Choudrie et al. (2021) analyze the financial and emotional state of Twitter users. An advanced deep learning technique of Transfer Learning combined with a Robustly Optimized BERT Pretraining Approach (RoBERTa) is used to analyze 2 million tweets collected during February–June 2020. A multi-class emotion classifier system was formed using RoBERTa and the collated Twitter dataset. The performance analysis shows that the proposed work gives the classification accuracy of 80.33% and an average MCC score of 0.78, which is better than existing AI-based emotion classification methods.

Finally, Piccialli et al. (2021) present 'The Role of Artificial Intelligence in Fighting the COVID-19 pandemic'. The authors provide a wider data science perspective to the application of AI, with a general overview of the dynamics of the pandemic and a critical evaluation of the papers published and datasets used on the COVID-19 outbreak. The detailed AI methods available to develop effective COVID-19 analysis provide a useful insight for researchers, particularly those working in the health domain. The paper discusses the stages of data flows from GP surgery, Hospitals, and research centers and the potential applications of AI using this data: outbreak prediction, spread tracking, diagnosis, drug production and drug repurposing.

# 3 Open Research Direction

There are many open research directions on this topic, and here we highlight three.

## 3.1 More and Better COVID-19 Diagnosis

With the advancement of pioneering AI and Data Science techniques, more and better diagnoses will be available. These may include innovative methods to improve the accuracy of medical analyses, such as PCR diagnosis, or totally new methods, such as alternative solutions similar to COVID_SCREENET and FedDPGAN.

## 3.2 A Better Prediction of Health and State of Well-Being

As the general public is fully aware of their health and the state of well-being, AI and Data Science can make better predictions in real-time. This can be used to collect the users' data through smart devices, and intelligent algorithms can compute different predicted outcomes based on users' needs. Medical institutes and hospitals can also develop their own services to monitor the health and state of well-being of patients contracting with COVID-19 and other diseases.

## 3.3 Sentiment Analysis of Citizens of Particular Countries and the World in General

Measurement of sentiment analysis has become increasingly important because it provides valuable insights to understand how people feel towards new policies and new rules related to the rapid updates in COVID-19 transmission. The general public's views on key agendas, such as economic recovery plans, employment, education, food supply and rules for social events, can also be analyzed through sentiment analysis.

# 4 Conclusion

We are delighted to present these six selected papers that demonstrate the use of intelligent AI methods to conduct COVID-19 and related analyses, and AI and Data Science methods will continue to enhance the abilities to investigate impacts, analyses and current interesting research topics related to the immediate pandemic and beyond.

Finally, we are grateful to the Editors-in-Chief of Information Systems Frontiers for giving us the opportunity to lead and organize this special issue. We look forward to future opportunities to serve the research community.

## References

Chang, V. (2014). Cloud bioinformatics in a private cloud deployment. In: *Advancing medical practice through technology: Applications for healthcare delivery, management, and quality* (pp. 205-220). IGI Global.

Choudrie, J., Kotecha, K., Matta, N., & Pappas, I. (2021). Applying and understanding an advanced, novel deep learning approach: A Covid 19, text based. *Emotions Analysis Study. Information Systems Frontiers, 23*(6). https://doi.org/10.1007/s10796-021-10152-6

Ecke, J. (2020). Labor issues, social movement studies, social economics, politics & government, Anarchist Studies. *Anarchist Studies*.

Elakkiya, R., Vijayakumar, P., & Karuppiah, M. (2021). COVID_SCREENET: COVID-19 screening in chest radiography images using deep transfer stacking. *Information Systems Frontiers, 23*(6). https://doi.org/10.1007/s10796-021-10123-x

Gupta, A., Deokar, A., Iyer, L., Sharda, R., & Schrader, D. (2018). Big data & analytics for societal impact: Recent research and trends. *Information Systems Frontiers, 20*(2), 185–194.

Hosseinian-Far, A., Ramachandran, M., & Slack, C. L. (2018). Emerging trends in cloud computing, big data, fog computing, IoT and smart living. In: *Technology for Smart Futures* (pp. 29–40). Springer.

Kaur, H., Ahsaan, S. U., Alankar, B., & Chang, V. (2021). A proposed sentiment analysis deep learning algorithm for analyzing COVID-19 tweets. *Information Systems Frontiers, 23*(6). https://doi.org/10.1007/s10796-021-10135-7

Kim, H., Hong, H., & Yoon, S. H. (2020). Diagnostic performance of CT and reverse transcriptase-polymerase chain reaction for coronavirus disease 2019: A meta-analysis. *Radiology*, 201343.

Piccialli, F., di Cola, V. S., Giampaolo, F., & Cuomo, S. (2021). The role of artificial intelligence in fighting the COVID-19 pandemic. *Information Systems Frontiers, 23*(6). https://doi.org/10.1007/s10796-021-10131-x

Ranney, M. L., Griffeth, V., & Jha, A. K. (2020). Critical supply shortages—The need for ventilators and personal protective equipment during the Covid-19 pandemic. *New England Journal of Medicine*.

Rotondo, J. C., Martini, F., Maritati, M., Mazziotta, C., Di Mauro, G., Lanzillotti, C., ... & Contini, C. (2021). SARS-CoV-2 Infection: New Molecular, Phylogenetic, and Pathogenetic Insights. Efficacy of Current Vaccines and the Potential Risk of Variants. *Viruses, 13*(9), 1687.

Sicari, S., Cappiello, C., De Pellegrini, F., Miorandi, D., & Coen-Porisini, A. (2016). A security-and quality-aware system architecture for internet of things. *Information Systems Frontiers, 18*(4), 665–677.

Singh, P. D., Kaur, R., Singh, K. D., & Dhiman, G. (2021). A novel ensemble-based classifier for detecting the COVID-19 disease for infected patients. Information systems. *Frontiers, 23*(6). https://doi.org/10.1007/s10796-021-10132-w

Tuli, S., Tuli, S., Tuli, R., & Gill, S. S. (2020). Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. *Internet of Things*, 100222.

Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial intelligence (AI) applications for COVID-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*.

van der Sommen, F., de Groof, J., Struyvenberg, M., van der Putten, J., Boers, T., Fockens, K., et al. (2020). Machine learning in GI endoscopy: Practical guidance in how to interpret a novel field. *Gut, 69*(11), 2035–2045.

Wang, C. J., Ng, C. Y., & Brook, R. H. (2020). Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing. *JAMA, 323*(14), 1341–1342.

World Health Organization, Coronavirus disease (COVID-19) pandemic, https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed 19 May 2021.

Zhang, L., Shen, N., Barnawi, A., Xi, S., Kumar, M., & Wu, Y. (2021). FedDPGAN: Federated differentially private generative adversarial networks framework for the detection of COVID-19 pneumonia. Information systems. *Frontiers, 23*(6). https://doi.org/10.1007/s10796-021-10144-6
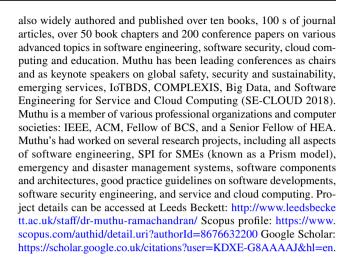
**Victor Chang** is currently a Full Professor of Data Science and Information Systems at the School of Computing, Engineering and Digital Technologies, Teesside University, Middlesbrough, UK. He leads the Artificial Intelligence and Information Systems Research Group at Teesside University, UK. He was a Senior Associate Professor, Director of Ph.D., Director of MRes and Interim Director of BSc IMIS at Xi'an Jiaotong-Liverpool University (XJTLU), Suzhou, China, between June 2016 and August 2019. He was an Honorary Associate Professor at the University of Liverpool. Previously he was a Senior Lecturer at Leeds Beckett University, UK, between Sep 2012 and May 2016. Within four years, he completed his Ph.D. (CS, Southampton) and PGCert (Higher Education, Fellow, Greenwich) while working on several projects at the same time. Prof Chang has been involved in funding, with a total of £14 million in Europe and Asia. Before becoming an academic, he has achieved 97% on average in 27 IT certifications. He won a European Award on Cloud Migration in 2011, IEEE Outstanding Service Award in 2015, best papers in 2012, 2015 and 2018, the 2016 European award, Outstanding Young Scientist 2017, Data Science special Award 2017, 5 INSTICC Service Awards 2017–2021, Outstanding Reviewer Awards 2018 and 2019, etc. He is a visiting scholar/Ph.D. examiner at several universities, an Editor-in-Chief of IJOCI & OJBD journals, Editor of FGCS (Oct 2014- Feb 2020), Springer Nature Open, Information Fusion and IDD journal, Associate Editor of IEEE TII, JGIM and IJBSR journals, founding chair of two international workshops and founding Conference Chair of IoTBDS and COMPLEXIS since the Year 2015–2016. He is the founding Conference Chair for FEMIB since the Year 2018–2019 and a founding Conference Chair for IIoT-BDSC since 2019–2020. He published three books as sole authors and the editor of 2 books on Cloud Computing and related technologies. He has received Outstanding Reviewer Awards from several top journals and the Outstanding Editor Award from FGCS. He gave 26 keynotes at international conferences. He has pioneering work for this research and has been invited for several keynotes. Prof. Chang is in the top 2% of

Scientists in 2017 and 2019 and is the most productive AI-based Data Scientist between 2010 and 2019. He is widely regarded as one of the most active and influential young scientists and experts in IoT/Data Science/Cloud/Security/AI/IS, as he has the experience to develop ten different services for multiple disciplines.

**Carole Goble** CBE FREng, is a Full Professor of Computer Science at the University of Manchester. She founded and led the e-Science Group for over 20 years. She has spent over 30 years working at the interface of advances in knowledge and computational technologies (semantic web, linked data, computational workflows, research data management, social computing) and their application in Research Infrastructures (cyberInfrastructure) used by researchers ranging from astrophysics to social science, but primarily in Biomedical science, Systems and Industrial Biotechnology and Biodiversity. She pioneered one of the first open computational workflow systems (Taverna), and the first open workflow sharing platform (myExperiment). Her current research interests are in FAIR and open digital objects, reproducible research, asset curation and preservation, computational workflows, semantic interoperability, knowledge exchange between scientists and new models of scholarly communication. She has published over 350 articles in her fields of interest, including the influential FAIR Data Principles Nature paper, and has an h-index of 83 and has won over 90 funding awards in her career totaling $30million to her institute and considerably more to the full consortiums. She currently serves on 8 EU projects, including the BY-COVID Horizon Europe project. Currently she is co-founder of the UK's Software Sustainability Institute, chairs the FAIRDOM Consortium, which produces asset sharing platform with over 140 installations and supports the COVID-19 Disease Maps consortium; is Joint Head of Node of ELIXIR-UK the national node of ELIXIR, European Research Infrastructure for Life Science data; directs the digital infrastructure for the IBISBA Research Infrastructure for Industrial biotechnology; and is a partner of the BioExcel EU Centre of Excellence for HPC. She co-leads developments of the Workflow Collaboratory for the European Open Science Cloud-Life cluster, developing the WorkflowHub for sharing COVID-19 workflows, and co-leads the ResearchObject.org initiative, which promotes the methods for packaging and exchanging Research Objects across Research infrastructures. She serves on numerous committees and panels and is currently the UK Government expert representative on the G7 Open Science Working Group. In 2008 she was awarded the Microsoft Jim Gray award for outstanding contributions to e-Science and in 2010 was elected a Fellow of the Royal Academy of Engineering. In 2014 she was made Commander of the Order of the British Empire by HM the Queen for her Services to Science.

**Muthu Ramachandran** is currently an Independent educational and research consultant, Integrated Cloud Solutions, UK. He was previouslu a Principal Lecturer (Associate Professor) in the School of Computing, Creative Technologies, and Engineering at Leeds Beckett University in the UK. Prior to this, he spent eight years in industrial research at Philips Research Labs and subsequently at Volantis Systems Ltd., Surrey, the UK where Muthu has worked on various research projects for Software Engineering Applications for Consumer Electronics, including software architecture design and reuse for large scale telecommunication, multi-media and car navigation systems. Currently, Muthu is leading research in the areas of Cloud Software Engineering, Big Data Software Engineering, IoT Software Engineering, Software Security Engineering, SOA, Cloud Computing, and in the main areas of Software Engineering on RE, CBSE, software architecture, reuse, quality and testing. Muthu's first career started as a research scientist, where Muthu has worked on large-scale high integrity real-time systems for Aerospace industrial applications. Muthu is the author of books: Software Components: Guidelines and Applications (Nova Publishers, NY, USA, 2008) and Software Security Engineering: Design and Applications (Nova Publishers, NY, USA, 2011). He has

also widely authored and published over ten books, 100 s of journal articles, over 50 book chapters and 200 conference papers on various advanced topics in software engineering, software security, cloud computing and education. Muthu has been leading conferences as chairs and as keynote speakers on global safety, security and sustainability, emerging services, IoTBDS, COMPLEXIS, Big Data, and Software Engineering for Service and Cloud Computing (SE-CLOUD 2018). Muthu is a member of various professional organizations and computer societies: IEEE, ACM, Fellow of BCS, and a Senior Fellow of HEA. Muthu's had worked on several research projects, including all aspects of software engineering, SPI for SMEs (known as a Prism model), emergency and disaster management systems, software components and architectures, good practice guidelines on software developments, software security engineering, and service and cloud computing. Project details can be accessed at Leeds Beckett: http://www.leedsbeckett.ac.uk/staff/dr-muthu-ramachandran/ Scopus profile: https://www.scopus.com/authid/detail.uri?authorId=8676632200 Google Scholar: https://scholar.google.co.uk/citations?user=KDXE-G8AAAAJ&hl=en.

**Lazarus Jegatha Deborah** completed her Ph.D. in Computer Science and Engineering at Anna University Chennai in the year 2013 and completed her Master of Engineering in the field of Computer Science and Engineering in Karunya Institute of Technology in the year 2005. She completed her Bachelor of Engineering at Madurai Kamarajar University, Madurai, in the year 2002. She is presently working as an Assistant Professor at Anna University Chennai (University College of Engineering, Tindivanam), Chennai, India. She is also acting as Head (in charge) of the CSE Department at University College of Engineering, Tindivanam since 2008. She has published many quality papers in reputed journals. She had been to many countries like the United States of America and Malaysia to present her research works in reputed conferences. She had also visited China to deliver a guest lecture to the research students working in Artificial Intelligence in the School of Cyber Engineering, Xidian University, Xidian, China. She is a life member of ISTE. She is a doctoral committee member for many research scholars in various universities like VIT, Sathyabama University. Her key areas of interest include Data Mining, Machine Learning, Natural Language Processing, Deep learning for e-learning applications. She completed a minor research project funded by Anna University Chennai to develop an e-learning platform as a learning and management system. She had been acting as the Guest Editor during the past years for special issues in various reputed journals like Elsevier, The Online Journal of Distance Education and E-Learning, International Journal of Internet Technology and Secured Transactions. She is an active researcher and had taken initiatives in conducting international conferences, faculty development programs and symposiums at the student level. She is currently working on developing Recommendation Systems using deep learning architectures based on psychological perspectives. Her other work includes reinforcement-based learning concepts used in vehicular ad-hoc networks.

**Reinhold Behringer** is working at Knorr Bremse GmbH on the development of autonomous truck systems since April 2019. He previously (7/2017–3/2019) has been employed by Daimler-Protics in the development of the latest Augmented Reality information system for the Mercedes S-Class. He is also a Visiting Professor at Leeds Beckett University, where he was a Professor of Creative Technology from 2005 to 2017. His R&D focus in this academic position was on prototyping applications for eLearning, location-based application development and mobile devices. He has two degrees in Physics (1988: MA, SUNY Buffalo, USA. 1990: Diplom in Physics, University Würzburg, Germany) and a Ph.D. in Engineering (1996: Dr.-Ing., UniBwM München, Germany). His main professional expertise is in autonomous road vehicles and real-time computer vision systems. He participated in the first US DARPA Grand Challenge (2004) and earlier in the development of the very first autonomous road vehicle driving on public roads in Germany

(1995) on Autobahn at speeds of 130 km/h. For this vehicle, he has developed a real-time computer vision system for road/lane detection/following and ego-state estimation, based on recursive Kalman filtering. During his following work at Rockwell Scientific (Thousand Oaks, USA) (1996–2005), he has developed several prototype systems for Augmented Reality demonstrators and multi-modal Human-Computer Interaction demonstrations, which did employ real-time computer vision approaches for scene detection and motion/orientation tracking. In his current role at Knorr-Bremse GmbH, he is involved in the development of neural network concepts for utilizing data from various sensors to enable autonomously driving trucks in SAE level 4. He has a substantial publication record and is regularly active in reviewing research proposals for the European Commission and National Funding agencies of EU member states.