AI-BASED AND DATA-DRIVEN ANALYSIS OF PANDEMIC DATA

Sobhan Moazemi

AI & Data Science Team

Fraunhofer SCAI

March 14, 2024 - University of Koblenz

sobhan.moazemi@scai.fraunhofer.de holger.froehlich@scai.fraunhofer.de

AI & DATA SCIENCE GROUP

Mission: Bringing Better Treatments to the Right Patients

Data

- Omics
- Longitudinal clinical
- Real World Data
  - EHRs
  - Device data

>150 publications

Methodologies, e.g.

- Hybrid AI / Knowledge Infusion
- (Generative) time series models
- Multi-modal, multi-scale data fusion

AI & Data Science

Applications

- Better Targets
- Precision Medicine: The Right Patients
- Better Trials
  - Enrichment trials
  - Synthetic controls

Markets

- Pharma
- Biotech
- Public Health

https://www.scai.fraunhofer.de/de/geschaeftsfelder/bioinformatik/forschungsfelder/AI-DAS.html

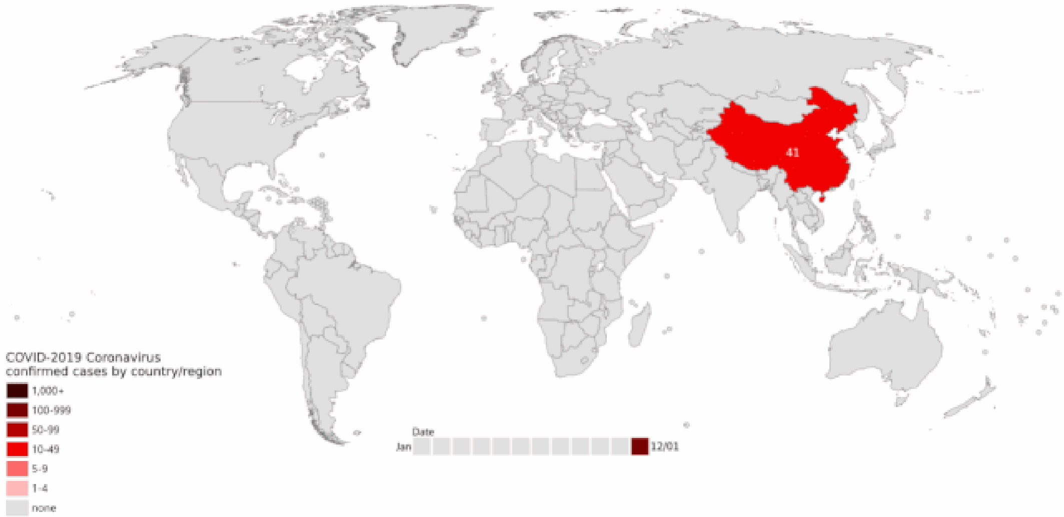
Fraunhofer SCAI

Introduction

Background

- COVID-19 pandemic

COVID-19 Worldwide deaths



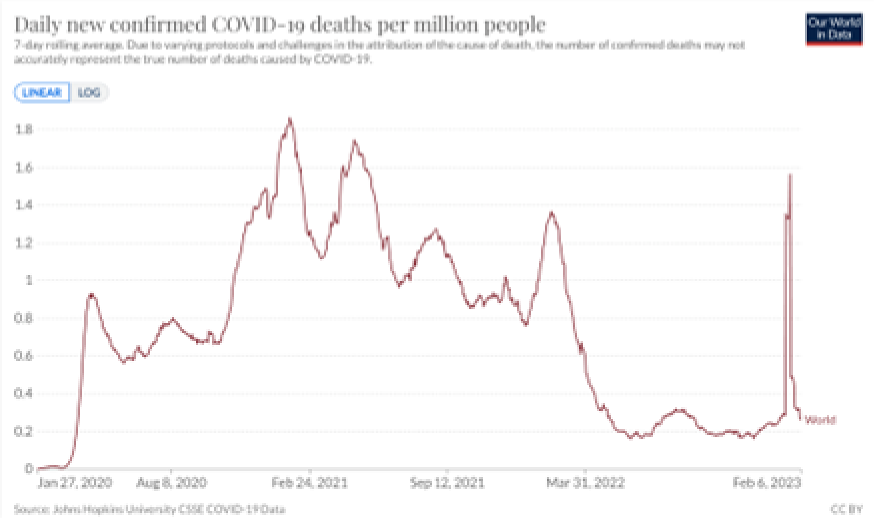


Figure 1. Outbreak of COVID-19 pandemic and number of worldwide deaths[1, 2]

- Lack of vaccination and effective treatment.
- Reporting delays of surveillance data (confirmed cases, hospitalization and deaths) for tracking pandemic waves; Countries with lack of surveillance capacity
- COVID-19 pandemic has affected healthcare around the world.
- The availability of a reliable, real-time indicator of pandemic would aid in timing public health interventions.

Slide credits: Danqi Wang

AI-DAS's Contributions

- **Early alert model for pandemic situation**

SCIENTIFIC REPORTS
natureresearch

Sci Rep. 2023; 13: 20780. PMCID: PMC10682010
Published online 2023 Nov 27. doi: 10.1038/s41598-023-48096-3 PMID: 38012282

Development of an early alert model for pandemic situations in Germany

Danqi Wang,[1] Manuel Lentzen,[1,2] Jonas Botz,[1,2] Diego Valderrama,[1,2] Lucille Deplante,[3] Jules Perrio,[3] Marie Génin,[3] Edward Thommes,[4] Laurent Coudeville,[4] and Holger Fröhlich[1,2]

▸ Author information ▸ Article notes ▸ Copyright and License information PMC Disclaimer

- **ExMedBERT: a Transformer-based model for structured EHRs**

IEEE J Biomed Health Inform. 2023 Sep;27(9):4548-4558. doi: 10.1109/JBHI.2023.3288768. Epub 2023 Sep 6.

A Transformer-Based Model Trained on Large Scale Claims Data for Prediction of Severe COVID-19 Disease Progression

Manuel Lentzen, Thomas Linden, Sai Veeranki, Sumit Madan, Diether Kramer, Werner Leodolter, Holger Frohlich

PMID: 37347632 DOI: 10.1109/JBHI.2023.3288768

- **Drug repositioning for COVID-19**

ELSEVIER
Artificial Intelligence in the Life Sciences
Volume 1, December 2021, 100020

Research Article

Machine Learning Based Prediction of COVID-19 Mortality Suggests Repositioning of Anticancer Drug for Treating Severe Cases

Thomas Linden [a b #], Frank Hanses [c q #], Daniel Domingo-Fernández [a], Lauren Nicole DeLong [a b], Alpha Tom Kodamullil [a], Jochen Schneider [d], Maria J.G.T. Vehreschild [e], Julia Lanznaster [f], Maria Madeleine Ruethrich [g], Stefan Borgmann [h], Martin Hower [i], Kai Wille [j], Torsten Feldt [k], Siegbert Rieg [l], Bernd Hertenstein [l], Christoph Wyen [m], Christoph Roemmele [n], Jörg Janne Vehreschild [e], Carolin E.M. Jakob [o], Melanie Stecher [p]...Holger Fröhlich [a b]

Fraunhofer SCAI

Introduction

Related work

❑ Social media posts: creating an early warning system [3-15]

Proxy	Techniques	Disease	Country	Training period (COVID-19)
Google Trends; Twitter; News paper feeds;	• Text mining and sentiment analysis (Keyword extraction, TF-IDF, COVID-twitter-BERT, BM, SVM, Naïve Bayes) • Anomaly detection (SH_ESD) • Statistical tests (Kolmogorov-Sminrvov, Anderson-Darling, Pearson correlations) • Linear regression • LSTM	Influenza, dengue, Zika, COVID-19	US, Canada, Australia, Japan, India, Colombia, Iran	Winter seasons: December 15, 2018 to January 21, 2019, December 15, 2019 to January 21, 2020; March 1, 2020 to August 21, 2020; January 2020 to March 2020; February 15, 2020 to March 18, 2020; January 22, 2020 to April 3, 2020; Separate COVID-19 pandemic waves

❑ Highlights of the study

Multiproxies
- Google Trends
- Twitter

Multivariables
- COVID-19 related German symptoms

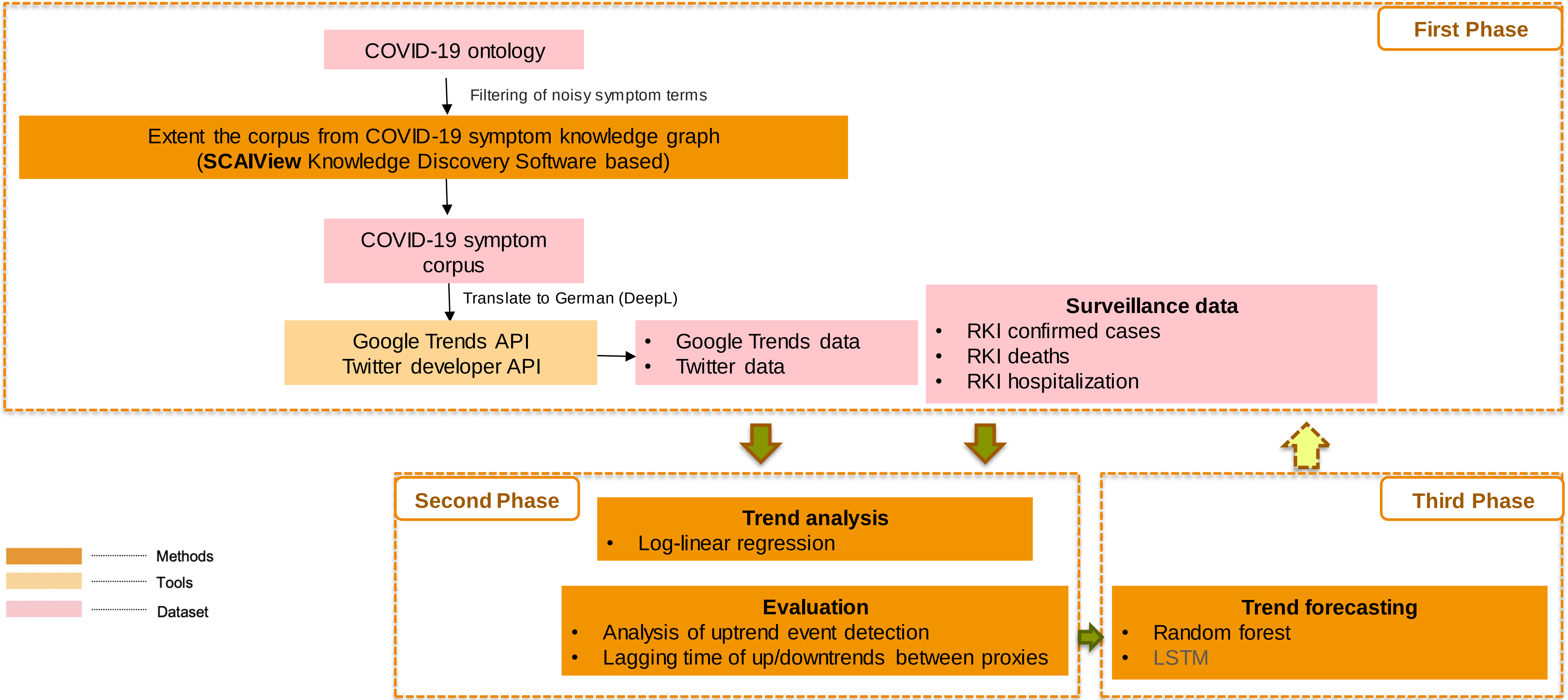
Longer training period
- January 2020 to June 2022

Multiple Models
- log linear regression model
- Random Forest/ LSTM

Slide credits: Danqi Wang

Fraunhofer SCAI

Workflow



Slide credits: Danqi Wang

Fraunhofer SCAI

Result – Country-level trend analysis

Visualization of the up and downtrends

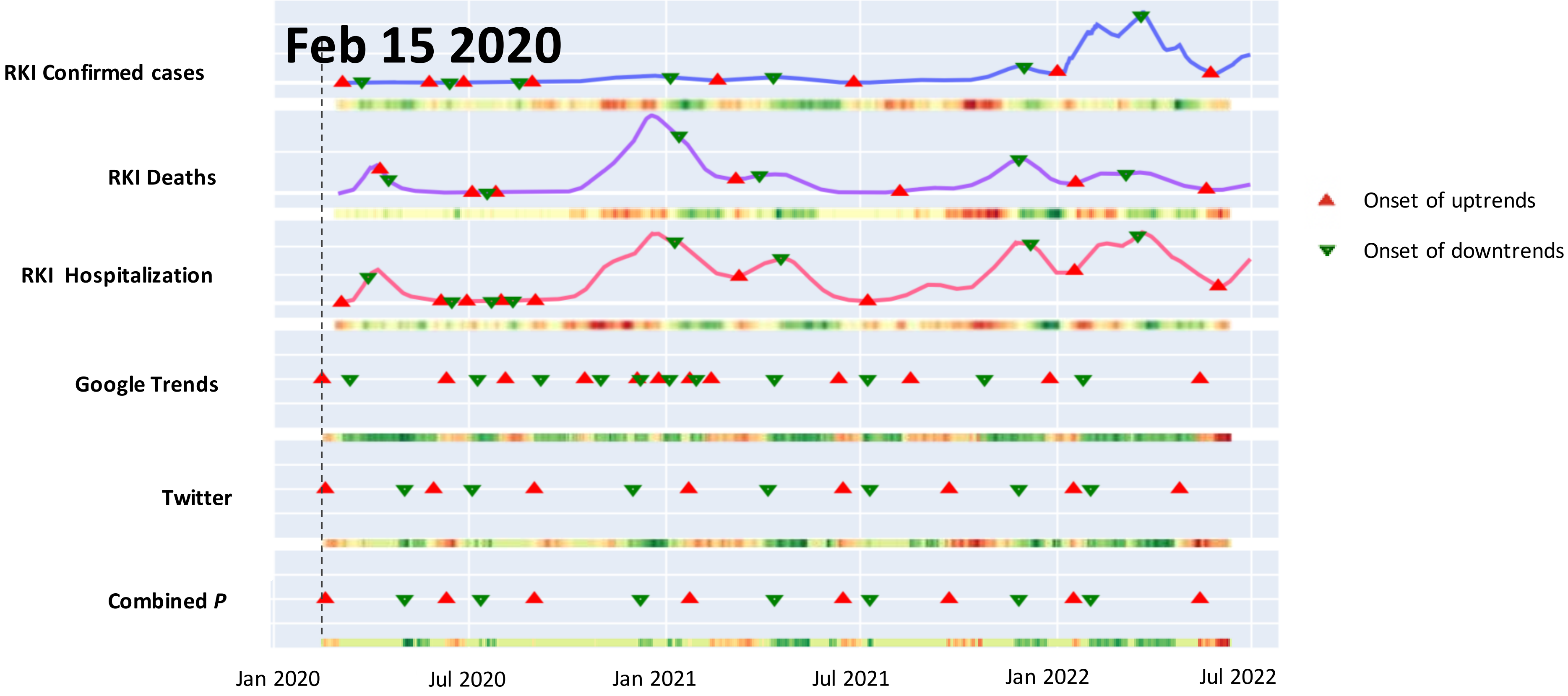


Figure 4. Visualization of the event detection procedure applied to Google Trends, Twitter search and Combined *P*.

Slide credits: Danqi Wang

Fraunhofer SCAI

Result – Country-level trend analysis

Pairwise event comparison

- Lagging correlation between proxies and gold standard

Figure 5. Up/down trend event detection results for pairwise comparisons between COVID-19 proxies(Google Trends, Twitter, Combined *P*) and gold standards A) RKI confirmed cases; B) RKI hospitalization; C) RKI deaths in Germany.

Slide credits: Danqi Wang

Fraunhofer SCAI

Result – Country-level trend analysis

Up trend events performance metrics – Combined proxies

Proxy names	Metric	RKI Cases	RKI Deaths	RKI Hospitalization
Twitter	Sensitivity	0.63	0.29	0.56
	Precision	0.63	0.25	0.63
	F1 Score	0.63	0.27	0.59
Google Trends	Sensitivity	0.88	0.57	0.78
	Precision	0.67	0.33	0.58
	F1 Score	0.76	0.42	0.67
Combined *P*	Sensitivity	0.63	0.43	0.67
	Precision	0.63	0.38	0.75
	F1 Score	0.63	0.40	0.71

Table 3. Sensitivity ($\frac{TP}{TP+TN}$) and precision ($\frac{TP}{TP+FP}$) rates for certain symptom from different proxies as an early indicator for an uptrend in three different COVID-19 gold standards (RKI confirmed cases, deaths, and hospitalization)

Slide credits: Danqi Wang

Alert model generation

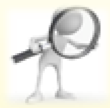
Trend forecasting (Random Forest & LSTM)

Dataset preparation

- **Feature:** Google Trends and Twitter slope data
- **Label:** Trend of RKI Confirmed Cases/ RKI Hospitalization/ RKI death

Hyperparameter optimization

- Optuna (TPE sampler)
- Time series cross-validation
- Trial size: 90



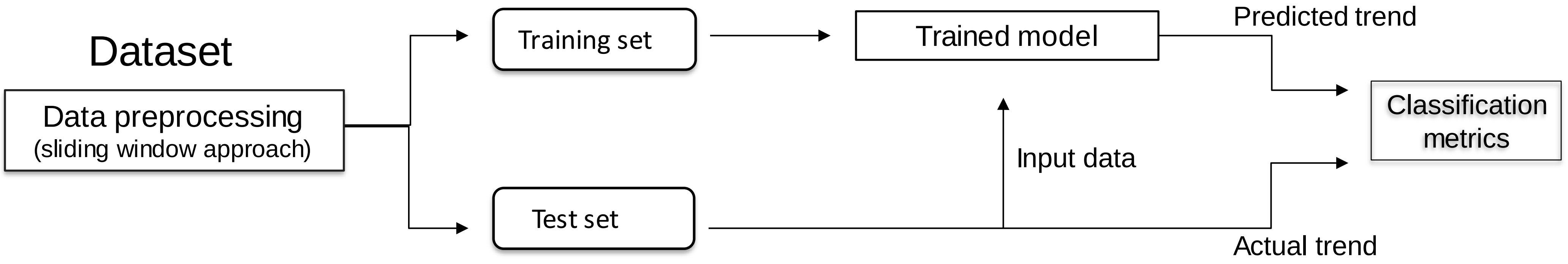


Figure 6. The workflow of trend forecasting

Slide credits: Danqi Wang

Fraunhofer SCAI

Result – Country-level trend forecasting
Uptrend forecasting

- Comparison of alert models

A. **RKI confirmed cases**

Model	Metrics	Uptrends		
		Google Trends	Twitter	Combined
Random Forest	sensitivity	0.71	1	1
	precision	1	0.6	0.67
	F1 score	0.83	0.75	0.8

B. **RKI hospitalization**

Model	Metrics	Uptrends		
		Google Trends	Twitter	Combined
Random Forest	sensitivity	0.88	0.71	0.67
	precision	1	0.55	0.91
	F1 score	0.94	0.62	0.77

Table 5. Evaluation metrics of forecasting models (Random Forest) for forecasting uptrend events (out-of-sample period in 2022) of A) RKI confirmed cases and B) RKI hospitalization

Slide credits: Danqi Wang

Result – Country-level trend forecasting

Downtrend forecasting

- Comparison of alert models

A. **RKI confirmed cases**

Model	Metrics	Downtrends		
		Google Trends	Twitter	Combined
Random Forest	sensitivity	1	0.71	0.86
	precision	0.66	0.76	0.81
	F1 score	0.8	0.74	0.84

B. **RKI hospitalization**

Model	Metrics	Downtrends		
		Google Trends	Twitter	Combined
Random Forest	sensitivity	1	0.88	1
	precision	0.88	0.88	0.95
	F1 score	0.93	0.88	0.97

Table 6. Evaluation metrics of forecasting models (Random Forest) for forecasting downtrend events (out-of-sample period in 2022) of A) RKI confirmed cases and B) RKI hospitalization

Slide credits: Danqi Wang

Fraunhofer SCAI

EARLY ALERT MODEL FOR PANDEMIC SITUATION: CONCLUSION

- **Country- and State-level COVID-19 early warning model in Germany**
  - Assessment the utility of Google Trends and Twitter (German symptom based time-series data), and their combined indicator.
  - Models: Probabilistic based log linear regression model (Trend analysis)
  - Random Forest (Trend Forecasting)
- **Country-level trend analysis**
  - Compared with the result of Kogan *et al.* [3] in U.S. (an uptrend in COVID-19 infections could be predicted up to 7 days in advance with an accuracy of ~75%), Google Trends got F1 scores of 0.76 for tracking RKI confirmed cases.
  - Google Trends can predate an increase in RKI confirmed cases and RKI hospitalization by a median of 16 and 19 days.
- **Country-level trend forecasting**
  - RKI confirmed cases: Random Forest-Google Trends (uptrend: F1 score of 0.83; downtrend: F1 score of 0.8)
  - RKI hospitalization: Random Forest-Google Trends (uptrend: F1 score of 0.94; downtrend: F1 score of 0.93)
- Predictive symptoms for predicting up and down trend events of RKI confirmed cases and RKI hospitalization.
- Google Trends contains important information for effectively predicting disease incidence.

Slide credits: Danqi Wang

Fraunhofer SCAI

AI-DAS's Contributions

- **Early alert model for pandemic situation**

SCIENTIFIC REPORTS
natureresearch

Sci Rep. 2023; 13: 20780.
Published online 2023 Nov 27. doi: 10.1038/s41598-023-48096-3
PMCID: PMC10682010
PMID: 38012282

Development of an early alert model for pandemic situations in Germany

Danqi Wang,[1] Manuel Lentzen,[1,2] Jonas Botz,[1,2] Diego Valderrama,[1,2] Lucille Deplante,[3] Jules Perrio,[3] Marie Génin,[3] Edward Thommes,[4] Laurent Coudeville,[4] and Holger Fröhlich[1,2]

▸ Author information ▸ Article notes ▸ Copyright and License information PMC Disclaimer

- **ExMedBERT: a Transformer-based model for structured EHRs**

IEEE J Biomed Health Inform. 2023 Sep;27(9):4548-4558. doi: 10.1109/JBHI.2023.3288768.
Epub 2023 Sep 6.

A Transformer-Based Model Trained on Large Scale Claims Data for Prediction of Severe COVID-19 Disease Progression

Manuel Lentzen, Thomas Linden, Sai Veeranki, Sumit Madan, Diether Kramer, Werner Leodolter, Holger Frohlich

PMID: 37347632 DOI: 10.1109/JBHI.2023.3288768

- **Drug repositioning for COVID-19**

ELSEVIER
Artificial Intelligence in the Life Sciences
Volume 1, December 2021, 100020

Research Article

Machine Learning Based Prediction of COVID-19 Mortality Suggests Repositioning of Anticancer Drug for Treating Severe Cases

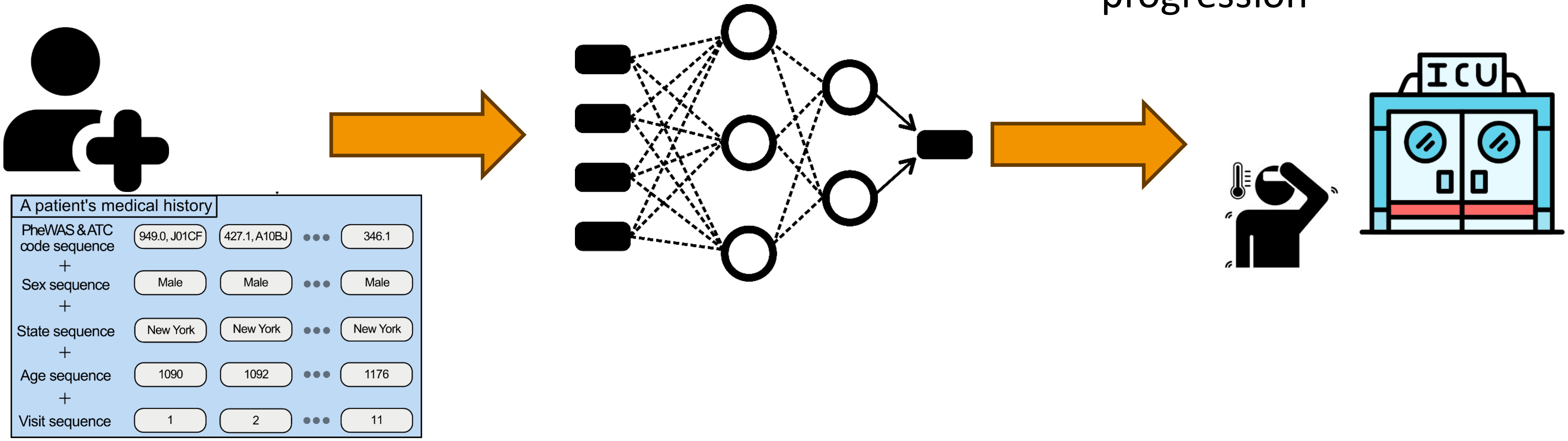
Thomas Linden[a b #], Frank Hanses[c q #], Daniel Domingo-Fernández[a], Lauren Nicole DeLong[a b], Alpha Tom Kodamullil[a], Jochen Schneider[d], Maria J.G.T. Vehreschild[e], Julia Lanznaster[f], Maria Madeleine Ruethrich[g], Stefan Borgmann[h], Martin Hower[i], Kai Wille[j], Torsten Feldt[k], Siegbert Rieg[l], Bernd Hertenstein[l], Christoph Wyen[m], Christoph Roemmele[n], Jörg Janne Vehreschild[e], Carolin E.M. Jakob[o], Melanie Stecher[p]...Holger Fröhlich[a b]

Fraunhofer SCAI

Leveraging Clinical Routine Data to Predict Risk of Severe COVID-19 Disease Progression

Large language model

Prediction of severe COVID-19 disease progression

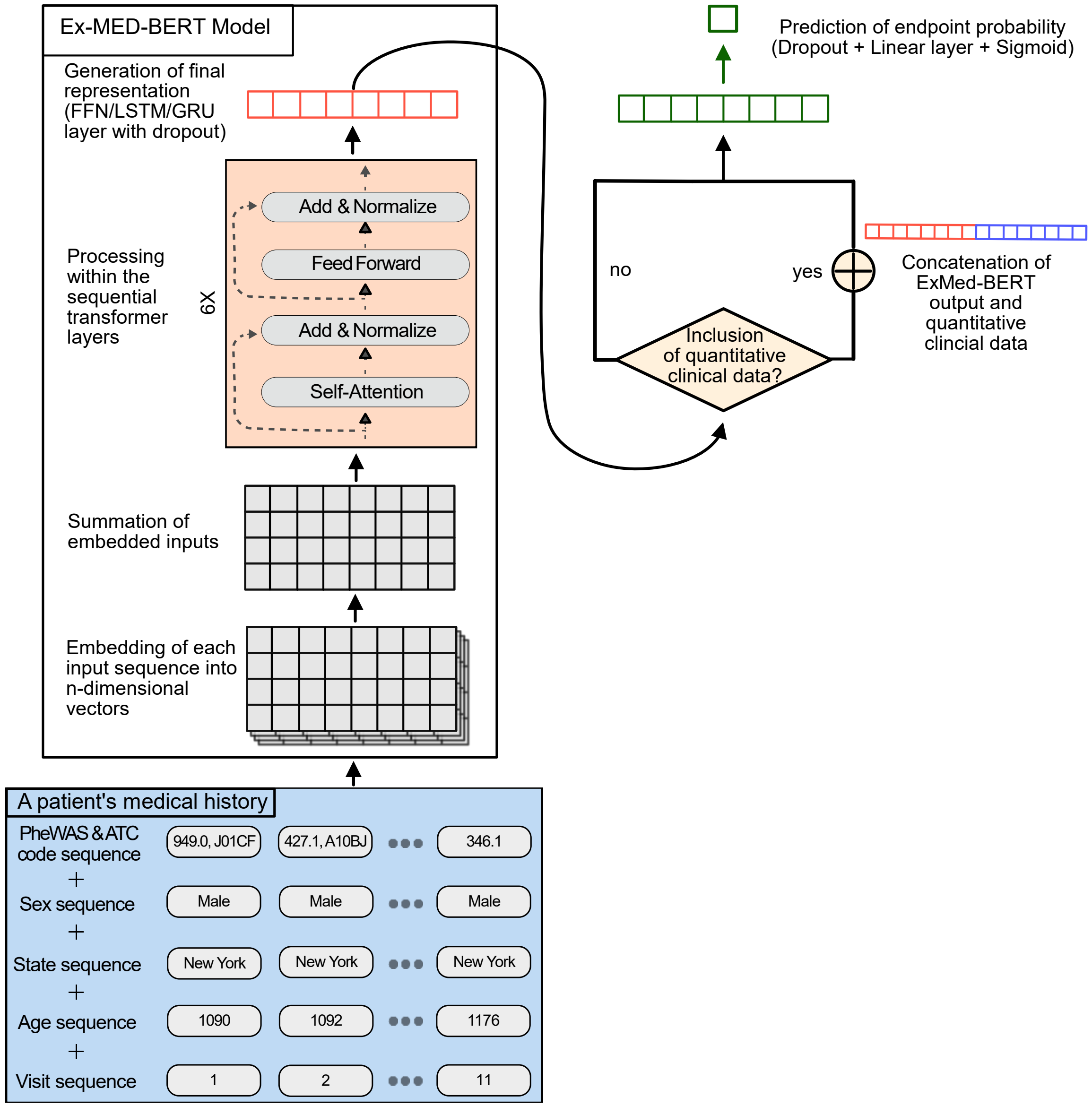


Slide credits: Manuel Lentzen

Fraunhofer SCAI

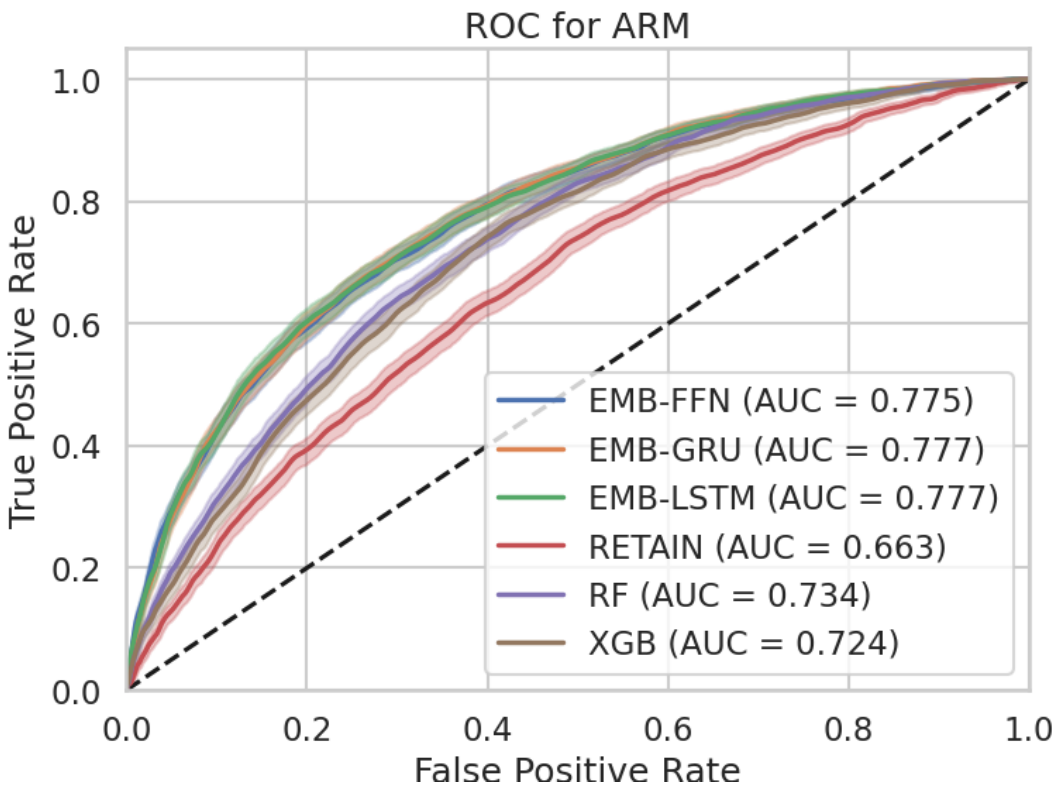
Development of a Transformer-based model for structured EHRs

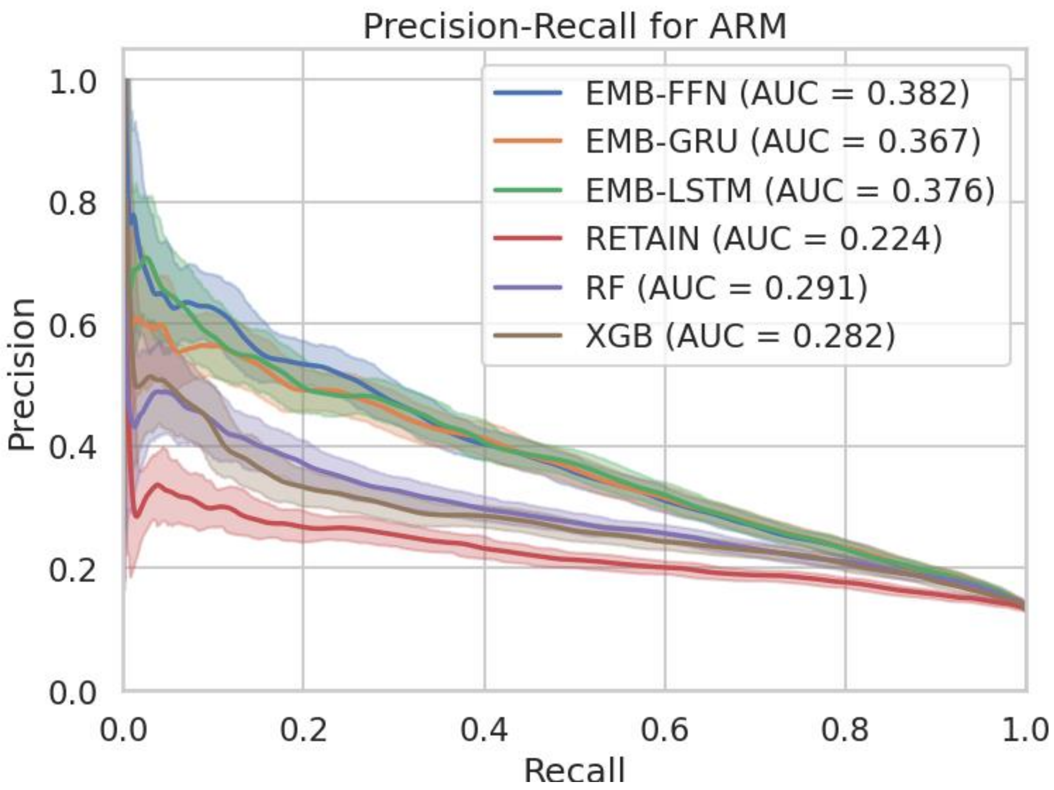
- Improved Med-BERT by incorporating further patient information (age, sex, residency / state, medication history)
- Pretrained the model on a large dataset of 3.5 million patients (988 million instances of drugs and diagnoses)
- Developed predictive risk models for forecasting Acute Respiratory Manifestations (ARM) post COVID-19 diagnosis using the refined model
- Validated model effectiveness and accuracy through benchmarks against Random Forest, RETAIN, and XGBoost



Slide credits: Manuel Lentzen

ExMed-BERT outperforms RF, RETAIN & XGBoost for predicting acute respiratory manifestation following COVID-19

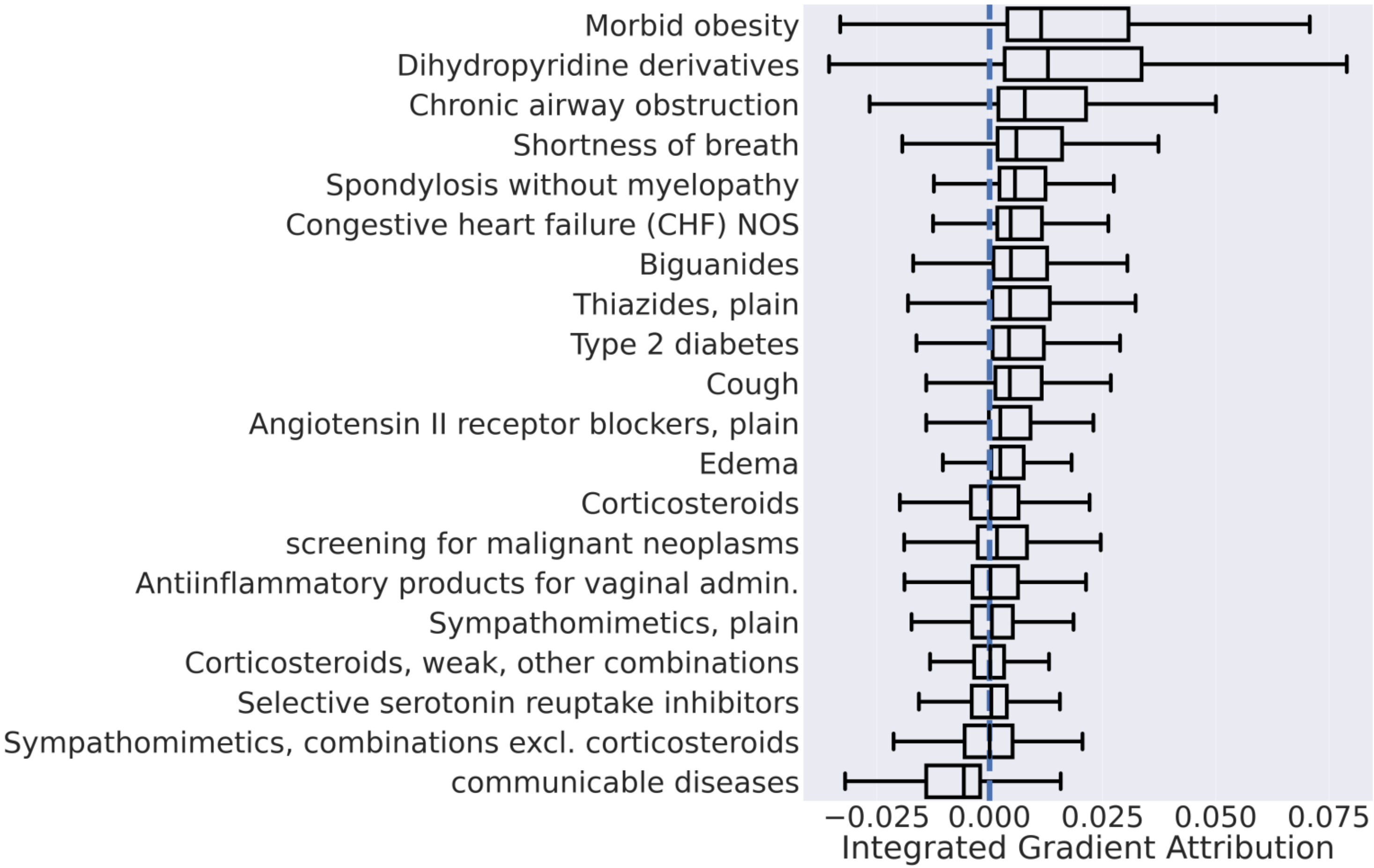




Slide credits: Manuel Lentzen

Fraunhofer SCAI

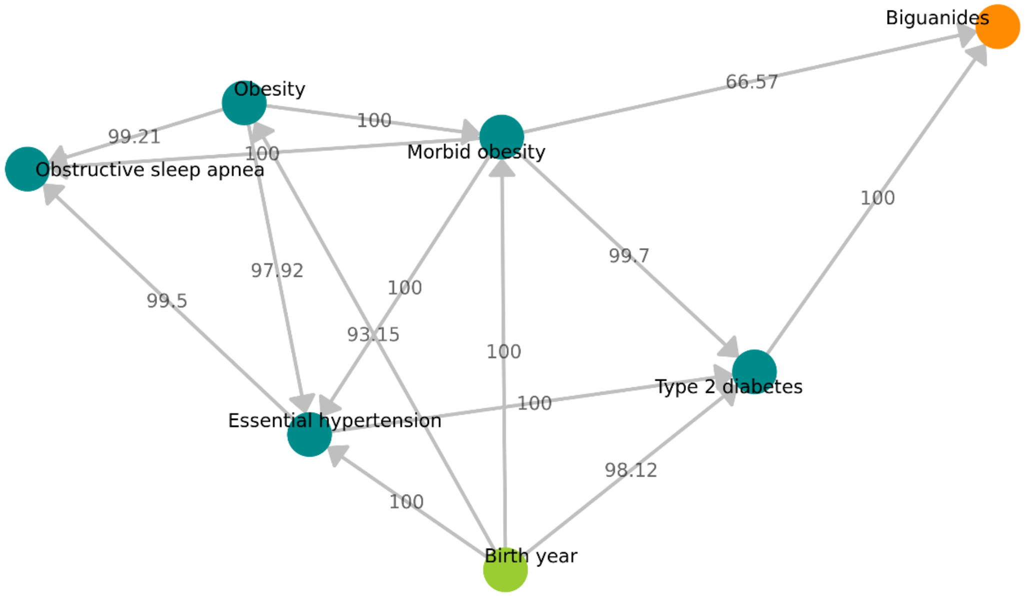
Assessing Global Feature Importance using Integrated Gradient Approach: Top 20 Diagnoses and Drugs for ExMed-BERT-GRU



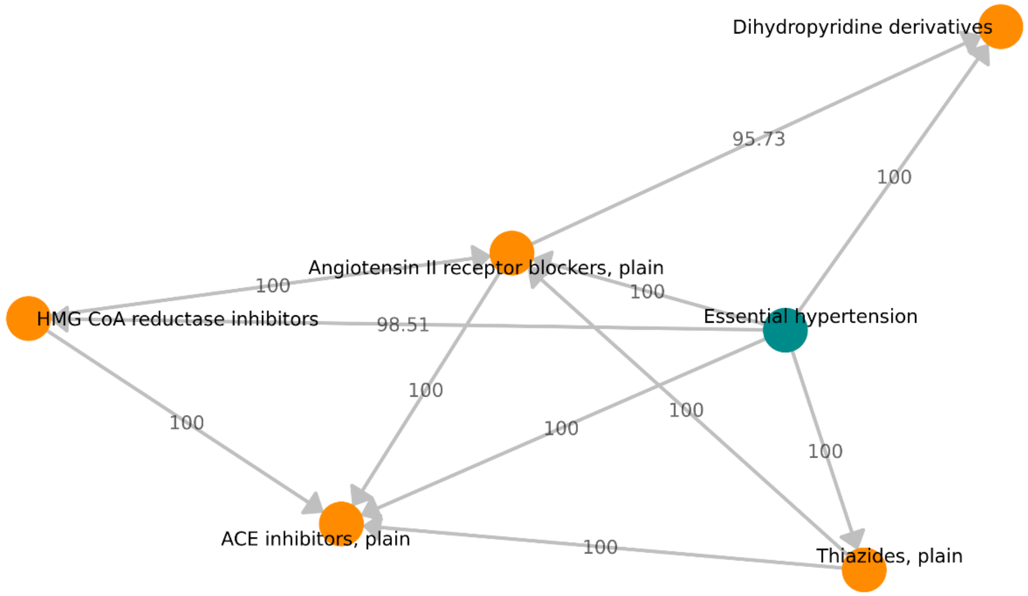
Slide credits: Manuel Lentzen

Fraunhofer SCAI

Understanding Relationships: Bayesian Network Analysis of Morbid Obesity / Angiotensin II Receptor Blockers and Other Diagnoses/Drugs



Morbid obesity



Angiotensin II receptor blockers

Slide credits: Manuel Lentzen

Fraunhofer SCAI

EXMEDBERT: CONCLUSION

- We generated customized transformer-based model for structured EHR data
- Risk models developed for ARM endpoint outperform RF, RETAIN and XGBoost models
- Performance enhancement through incorporation of quantitative clinical data
- Identification of several known risk factors as important features for our model
- Transfer learning enables utilization of pre-trained ExMed-BERT model for diverse clinical endpoint predictions

Slide credits: Manuel Lentzen

AI-DAS's Contributions

- Early alert model for pandemic situation

SCIENTIFIC REPORTS
natureresearch

Sci Rep. 2023; 13: 20780.
Published online 2023 Nov 27. doi: 10.1038/s41598-023-48096-3
PMCID: PMC10682010
PMID: 38012282

Development of an early alert model for pandemic situations in Germany

Danqi Wang,[1] Manuel Lentzen,[1,2] Jonas Botz,[1,2] Diego Valderrama,[1,2] Lucille Deplante,[3] Jules Perrio,[3] Marie Génin,[3] Edward Thommes,[4] Laurent Coudeville,[4] and Holger Fröhlich[1,2]

▸ Author information ▸ Article notes ▸ Copyright and License information PMC Disclaimer

- ExMedBERT: a Transformer-based model for structured EHRs

IEEE J Biomed Health Inform. 2023 Sep;27(9):4548-4558. doi: 10.1109/JBHI.2023.3288768.
Epub 2023 Sep 6.

A Transformer-Based Model Trained on Large Scale Claims Data for Prediction of Severe COVID-19 Disease Progression

Manuel Lentzen, Thomas Linden, Sai Veeranki, Sumit Madan, Diether Kramer, Werner Leodolter, Holger Frohlich

PMID: 37347632 DOI: 10.1109/JBHI.2023.3288768

- Drug repositioning for COVID-19

ELSEVIER
Artificial Intelligence in the Life Sciences
Volume 1, December 2021, 100020

Research Article

Machine Learning Based Prediction of COVID-19 Mortality Suggests Repositioning of Anticancer Drug for Treating Severe Cases

Thomas Linden [a b #], Frank Hanses [c q #], Daniel Domingo-Fernández [a], Lauren Nicole DeLong [a b], Alpha Tom Kodamullil [a], Jochen Schneider [d], Maria J.G.T. Vehreschild [e], Julia Lanznaster [f], Maria Madeleine Ruethrich [g], Stefan Borgmann [h], Martin Hower [i], Kai Wille [j], Torsten Feldt [k], Siegbert Rieg [l], Bernd Hertenstein [l], Christoph Wyen [m], Christoph Roemmele [n], Jörg Janne Vehreschild [e], Carolin E.M. Jakob [o], Melanie Stecher [p]...Holger Fröhlich [a b]

Fraunhofer SCAI

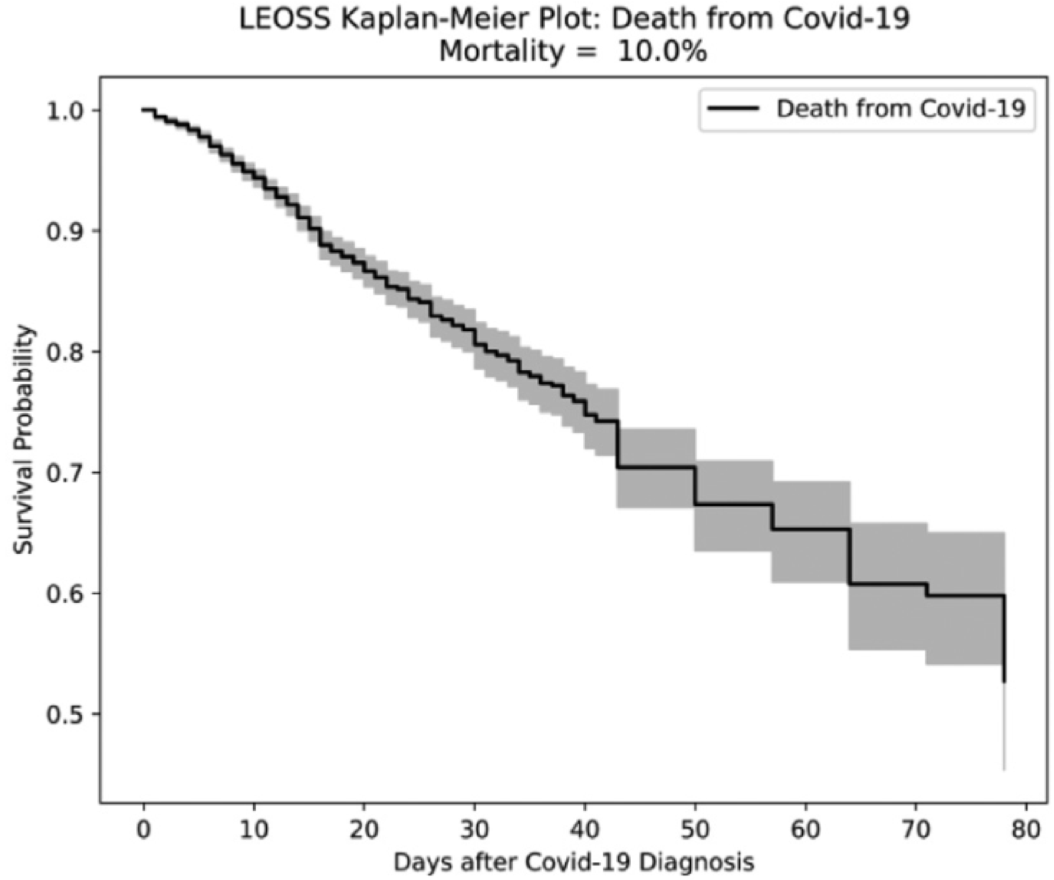
The Cohort: COVID-19 patients from the Lean European Open Survey on SARS-CoV-2 infected patients

(LEOSS - https://leoss.net)

Overview of patient demographics in LEOSS.

Age	
18 - 25 years	181
26 - 35 years	472
36 - 45 years	540
46 - 55 years	907
56 - 65 years	1125
66 - 75 years	981
76 - 85 years	1231
missing	242
Gender	
Male	3229
Female	2218
missing	232
Ethnicity	
Caucasian	4225
missing	1195
Asian & Pacific Islander	155
African & African American	98
Hispanic or Latino	6
Country	
Germany	5411
Turkey	65
Belgium	40
Czechia	33
Latvia	27
Other	26
GBR	23
Italy	19
Spain	15
France	11
Austria	9

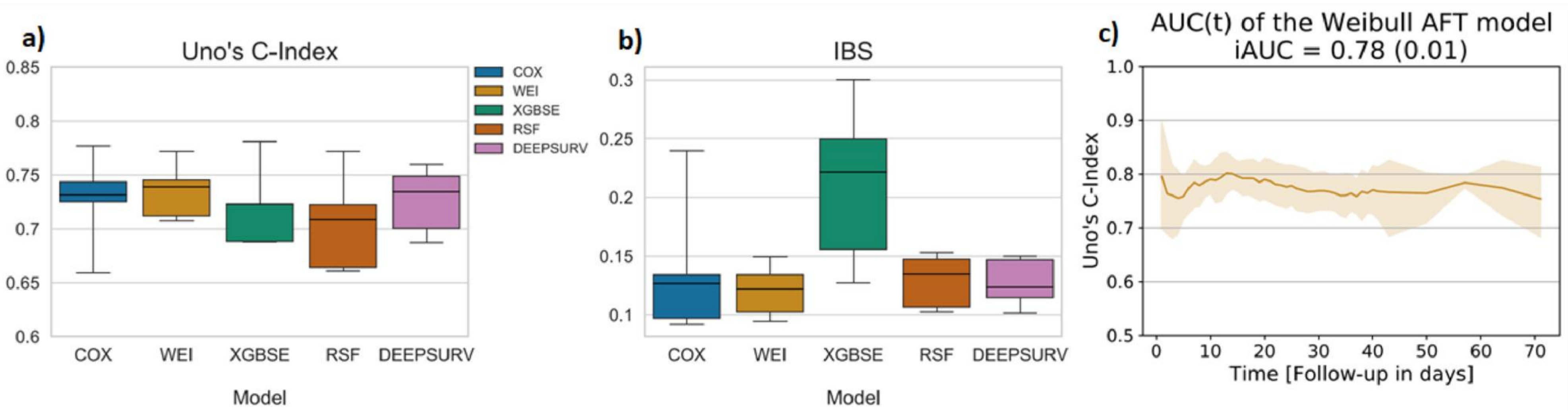
Data from ~5700 PCR or rapid test confirmed SARS-CoV-2 patients recruited in 100+ European active sites, primarily all over Germany



Death from Covid-19									
At risk	5677	3000	1301	584	294	129	87	62	0
Censored	0	2429	3946	4597	4855	4996	5034	5053	5108
Events	0	248	430	496	528	552	556	562	569

Fraunhofer SCAI

Machine learning can predict mortality with high accuracy



Model prediction performance measured via Uno's C-index on held out test sets (COX = elastic net penalized Cox proportional hazards regression; WEI = elastic net penalized Weibull accelerated failure time regression; XGBSE = XGBoost Survival Embeddings; RSF = Random Survival Forest; DEEPSURV = DeepSurv); (b) model calibration error measured via Integrated Brier Score (IBS) on held out test sets; (c) model prediction performance as function of time on held out test sets with 95% confidence interval, with integrated AUC (iAUC) denoting the mean (standard error) AUC over time. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Feature importances via SHAP values:

- lab measures were the most relevant type of features (23.5% cumulative importance)
- Disease symptoms ranked second (20.5%)
- Comorbidities third (13.2% cumulative importance).
- Comorbidity associated predictors included hypertension, an acute kidney injury, diabetes and dementia

a)

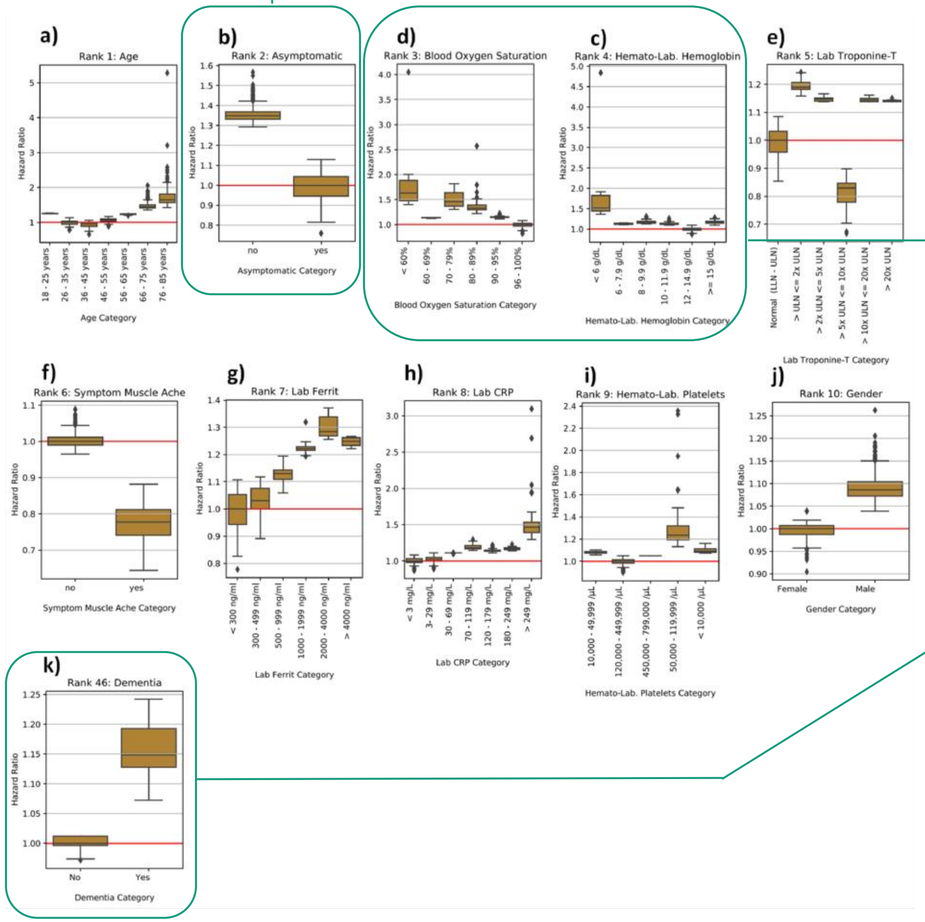
rank	domain	feature	importance
1	Demographic	Age	6.5%
2	Symptoms	Asymptomatic	2.9%
3	Vital	SO2 Oxygen Saturation	2.9%
4	Hemato. Lab	Hemoglobin	2.6%
5	Lab	Troponine T	2.6%
6	Symptoms	Muscle Aches	2.4%
7	Lab	Ferrit	2.4%
8	Lab	CRP	2.3%
9	Hemato. Lab	Platelets	2.2%
10	Demographic	Gender	1.9%
...	...	...	...
46	Dementia	Comorbidities	0.8%
		Total	29.4%

b)

Modality	Cumulative Importance	Number of Features
Lab	23.5%	25
Symptoms	20.5%	41
Comorbidities	13.2%	43
Vital	11.2%	11
Hemato. Lab	11.1%	8
Demographic	8.7%	3
Treatments	4.2%	5
CT_Xray	3.1%	12
Urine	2.0%	8
Other	1.6%	2
Smoking	1.0%	2
Total	100%	160

Fraunhofer SCAI

Partial dependence plots for most influential predictors

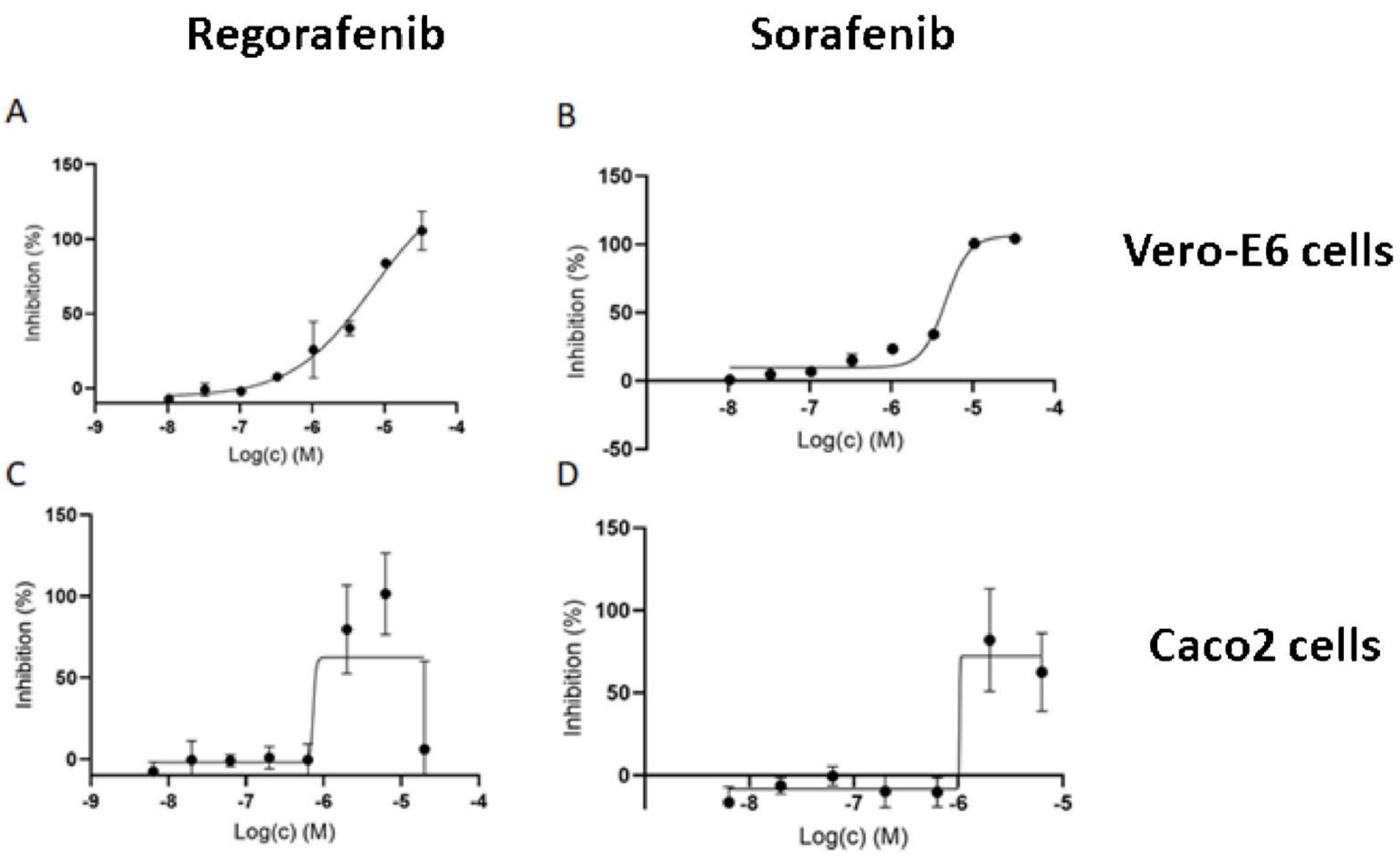


- An asymptomatic Covid-19 infection resulted into a 35% lower mortality risk compared to more severe disease symptoms
- For patients with low hemoglobin level or low oxygen saturation mortality risk was even increased by 50%
- Prior diagnosis of dementia results into an 15% increased mortality risk after SARS-CoV-2-19 infection (hazard ratio dementia vs. non-dementia: 1.15; 95% CI: [1.08, 1.24])

Fraunhofer SCAI

Anti-cancer drugs Regorafenib and Sorafenib help treat COVID-19 patients

- By calculating the overlap of multiple knowledge graphs and subsequent research in GWAS* studies:
- Identified Regorafenib and Sorafenib as potential COVID-19 drugs targeting TYK2 through cellular screening assay for anti-cytopathic effect.
- Regorafenib and Sorafenib were tested in Vero-E6 and Caco2 cell lines against SARS-CoV-2.
- Results shown: percentage inhibition of viral cytopathic effect normalized to Remdesivir as positive control (100%).
- Drug administration occurred 48 or 96 hours post-infection, with stained, washed, and counted cells indicating some toxicity, particularly in Caco2 cells at higher concentrations.c effect compared to Remdesivir
- The slightly negative relative inhibition shown in panel D is caused by plate control differences within plates.



*GENOME-WIDE ASSOCIATION STUDIES (HTTPS://WWW.GENOME.GOV/GENETICS-GLOSSARY/GENOME-WIDE-ASSOCIATION-STUDIES)

Fraunhofer SCAI

GREAT NEWS: ONGOING CLINICAL TRIAL!

An Observational Study, Called ROCURS, to Learn About COVID-19 Related Outcomes in People With Cancer Who Are Treated With Tyrosine Kinase Inhibitors (TKIs) Including Regorafenib or Sorafenib (ROCURS)

ClinicalTrials.gov ID NCT05594147

Sponsor Bayer

Information provided by Bayer (Responsible Party)

Last Update Posted 2023-11-07

Fraunhofer SCAI

DRUG REPOSITIONING FOR COVID-19: CONCLUSION

- Despite vaccination efforts, COVID-19 cases continued to rise globally until 2021, highlighting the need for effective medications against severe cases.
- A machine learning (ML) model was developed to predict COVID-19 mortality using extensive observational data from LEOSS, primarily from German inpatient settings.
- The ML model performs similarly to the established 4C mortality score (from UK) but differs in formulation, focusing on time-dependent mortality risk after COVID-19 diagnosis.
- The model identifies dementia as a predictor, suggesting TYK2 as a potential drug target for COVID-19.
- Regorafenib and Sorafenib, anti-cancer drugs, show promise as potential COVID-19 treatments based on cellular screening assays.
- The study demonstrates the utility of ML-based risk models in identifying potential drug targets and treatment options for COVID-19.
- Clinical trials already ongoing to confirm the efficacy of Regorafenib and Sorafenib in treating COVID-19.

Fraunhofer SCAI

Appropriate Solutions for Different Phases of COVID-19 Pandemic

Early Alert Model

- Evaluation of Google Trends and Twitter for symptom-based time-series data
- Models: Probabilistic log linear regression and Random Forest for trend analysis and forecasting
- Comparison with Kogan et al. in the U.S., showcasing F1 scores for tracking RKI confirmed cases
- Google Trends predated RKI confirmed cases and hospitalizations by a median of 16 and 19 days

Clinical Endpoint Prediction

- Customized transformer-based model for structured EHR data and ARM endpoint risk modeling
- Transfer learning enables utilization of pre-trained ExMed-BERT model for diverse clinical endpoint predictions
- Performance enhancement through incorporation of quantitative clinical data
- Identification of several known risk factors as important features for our model

Drug Repositioning

- Identification of TYK2 as a potential drug target for COVID-19 using ML predictions
- Potential COVID-19 treatments: Regorafenib and Sorafenib, based on cellular screening assays
- Ongoing clinical trialds for Regorafenib and Sorafenib efficacy
- Demonstration of ML-based risk models' utility in identifying drug targets and treatment options for COVID

Fraunhofer SCAI

ACKNOWLEDGEMENTS

AI & Data Science Team @ Fraunhofer SCAI

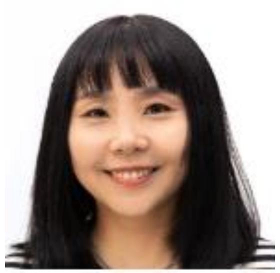
Holger Fröhlich

Mohamed Aborageh
Jonas Botz
Jannis Guski
Sophia Krix
Manuel Lentzen
Sumit Madan
Tamara Raschka
Diego Valderrama
Sobhan Moazemi

Tim Adams
Sabyasachi Patyoshi
Tom Hähnel
Kiril Klein (guest)
Johanna Driever
Lennart Carsten-Behrens
Konrad Gerischer
Georg v. Arnim
Sebastian Schwick

Former Associates

Danqi Wang



Thomas Linden

sobhan.moazemi@scai.fraunhofer.de holger.froehlich@scai.fraunhofer.de

innovative health initiative
ERAPerMed
JPND research
Bundesministerium für Wirtschaft und Klimaschutz
DFG Deutsche Forschungsgemeinschaft

References

[1] https://upload.wikimedia.org/wikipedia/commons/b/b3/COVID-19-outbreak-timeline.gif (access on 17.09.2022)

[2] https://ourworldindata.org/explorers/coronavirus-data-explorer (access on 07.02.2023)

[3] N.E. Kogan, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. Sci. Adv. 7, eabd6989 (2021).

[4] V. Madhur, et al. Google Search Trends predicting disease outbreaks: an analysis from India. Healthc Inform Res. 24(4):300-308 (2018).

[5] F. Dean. Short-term forecasting of the COVID-19 pandemic using Google Trends data: Evidence from 158 countries. Applied Econometrics. 1-20 (2020).

[6] Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SRN. Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill. (2020) 6:e18828. doi: 10.2196/18828.

[7] Ortiz-Martínez Y, Garcia-Robled JE, Vásquez-Castañeda DL, Bonilla-Aldana DK, Rodriguez-Morales AJ. Can Google R© trends predict COVID-19 incidence and help preparedness? The situation in Colombia. Travel Med Infect Dis. (2020) 37:101703. doi: 10.1016/j.tmaid.2020.101703.

[8] M. J. Pau, et al. Twitter improves influenza forecasting. PLOS Curr. 6, ecurrents.outbreaks.90b9ed0f59bae4ccaa683a39865d9117 (2014).

[9] R. Nagar, et al., A case study of the New York City 2012-2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J. Med. Internet Res. 16, e236 (2014).

[10] A. Z. Klein, et al. Toward using Twitter for tracking COVID-19: A natural language processing pipeline and exploratory data set. J. Med. Internet Res. 23, e25314 (2021).

[11] Vu Tran, et al. Tweet analysis for enhancement of COVID-19 epidemic simulation: a case study in Japan. Front. Public Health. (2022).

[12] R. Lamsal, et al. Twitter conversations predict the daily confirmed COVID-19 cases. Applied Soft Computing. 2206.10471v2 (2022).

[13] S. Yousefinaghani, et al. The assessment of Twitter's potential for outbreak detection: avian influenza case study. Sci. Rep. 9:18147(2019).

[14] A. Mavragani. Tracking COVID-19 in Europe: Infodemiology Approach. JMIR Public Health Surveill. 6(2):e18941(2020).

[15] Y.M. Zhang, et al. An intelligent early warning system of analyzing Twitter data using machine learning on COVID-19 surveillance in the US. Expert Systems with Applications 198(2022).

[16] Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). STL: A seasonal-Trend Decomposition Procedure Based on Loess. Journal of Official Statistics, 6(1):3–73.

[17] R.J. Hyndman, G. Athanasopoulos. Forecasting: principles and practice, 3rd edition, Otexts:Melbourne, Australia. Otexts:com/fpp2.

© Fraunhofer

Fraunhofer SCAI