

Titel und Verantwortliche

- Titel des Projekts: DEMIS Signale 2.0
- Förderkennzeichen: DRM D81600, Aktenzeichen 1368-1600
- Leitung: Dr. Hermann Claus
- Autoren: Dr. Hermann Claus, Dr. Alexander Ullrich
- Kontaktdaten: ClausH@rki.de, UllrichA@rki.de
- Laufzeit: 01.12.2017 - 30.11.2020 (3 Jahre)
- Fördersumme: 1.209.631 €

Das Projekt wurde bis zum 31. Dezember 2021 kostenneutral verlängert, es wurde nicht die volle Fördersumme aufgebraucht.

Inhaltsverzeichnis

Titel und Verantwortliche	1
Zusammenfassung (max. 1 Seite).....	4
Einleitung.....	5
Ausgangslage des Projekts	5
Ziele des Projekts.....	5
Projektstruktur (Projektaufbau, Strukturen, Verantwortlichkeiten)	5
Erhebungs- und Auswertungsmethodik.....	5
Durchführung, Arbeits- und Zeitplan	6
Ergebnisse	6
Methodenentwicklung	7
Supervised Learning	7
Raum-zeitliches Modell	7
Evaluierung von Ausbruchserkennungsmethoden	8
Generative Adversarial Networks	9
Hierarchisches Forecasting mit Erklärungen.....	9
Fallbasierte Ausbruchserkennung durch Dimensionsreduktion	9
Kofaktoren und Integration.....	10
Wetterdaten.....	10
Kartenmaterial.....	10
Nachrichten	10
Überwachtes Lernen von Syndromen	11
Syndromische Surveillance durch "Social Media" Daten	12
Dashboard und Evaluation	12
Dashboards.....	12
Signale Dashboard.....	17
Influenza Dashboard	18
Vector-borne Dashboard.....	18
Vacmap.....	19
Kommunikation und Vernetzung	19
Vernetzung	20
Focus Group AI for Health	20
Hackathon	21
Gender Mainstreaming Aspekte	21
Diskussion der Ergebnisse, Gesamtbeurteilung.....	21

Verbreitung und Öffentlichkeitsarbeit der Projektergebnisse.....	22
Masterarbeiten.....	22
Veröffentlichungen in Fachzeitschriften	22
Präsentationen	23
Verwertung der Projektergebnisse (Nachhaltigkeit / Transferpotential	24

Zusammenfassung (max. 1 Seite)

Das bestehende Frühwarnsystem das im Vorgängerprojekt „DEMIS Signale“ entstanden ist und mit Hilfe von statistischen Methoden mögliche Ausbrüche in Meldedaten erkennt, wurde um neuartige Analysen aus dem Bereich des maschinellen Lernens und der Anreicherung durch weitere Datenquellen erweitert. Die Berichterstattung wurde dabei um mehrere Dashboards erweitert, welche in enger Zusammenarbeit mit den Nutzern und Nutzerinnen entstanden sind.

Es wurden neue Methoden des maschinellen Lernens zur Ausbruchserkennung und Vorhersage von Infektionszahlen entwickelt. Diese Methoden berücksichtigten Aspekte die in bisherigen Ansätzen nicht oder in der Ausprägung vorhanden waren. Eine der neuen Ausbruchserkennungsmethoden basierte auf Einzelfalldaten und deren vielfältigen Informationen und bediente sich dabei state-of-the-art Dimensionsreduktionsverfahren und semi-überwachten Clusterverfahren. Diese Art der Ausbruchserkennung unterscheidet sich stark von den üblicherweise Zeitreihen-basierten Ansätzen und entspricht eher der Arbeitsweise menschlicher Experten und Expertinnen. Die entwickelten Vorhersagemethoden adressierten ebenfalls Aspekte die selten in den Standardmethoden zu finden sind, aber doch wichtig für die Arbeit in der Epidemiologie. Dazu zählen die Bestimmung von Unsicherheiten und die Erklärbarkeit der Vorhersageergebnisse. Zur Vergleichbarkeit und Optimierung neuer und existierender Ausbruchserkennungsmethoden wurde ein Evaluierungs-Framework entwickelt.

Zur Gewinnung und Nutzbarmachung zusätzlicher Datenquellen wurden Methoden des maschinellen Lernens verwendet und Pipelines entwickelt, die automatisiert Daten extrahieren, verarbeiten und verfügbar machen. Für die Extraktion von relevanten Informationen aus epidemiologischen Texten wie ProMED Nachrichten und WHO Outbreak News Artikeln wurden Methoden aus dem Bereich der Verarbeitung natürlicher Sprache (NLP) angepasst und in einer Software verwendet. Die Software bestimmt zusätzlich die Relevanz eines Artikels auf Basis bisheriger Entscheidungen und ist daher ein nützliches Tool im Screening-Prozess einer event-based Surveillance. Um die aus einer Notaufnahmeh-Surveillance stammenden Daten für eine Surveillance von Infektionskrankheiten nutzen zu können wurde eine Methode des überwachten Lernens angewendet um Syndromdefinitionen zu generieren die mit fehlenden Einträgen in den Daten umgehen kann und trotzdem vergleichbare Ergebnisse liefern wie Syndromdefinitionen von Fachexperten und Fachexpertinnen.

Zur Darstellung der Analyseergebnisse und der Integration mehrere Datenquellen wurden verschiedene Berichte und Dashboards entwickelt die sowohl intern für Experten und Expertinnen, als auch für externe Fachgremien und während der Pandemie sogar die allgemeine Öffentlichkeit verfügbar gemacht wurden.

Die entwickelten Methoden wurden in Masterarbeiten oder Veröffentlichungen in Fachzeitschriften beschrieben und der Programmcode der entstandenen Softwaretools wenn möglich öffentlich zur Verfügung gestellt. Ebenfalls wurden einige der Ergebnisse und Erkenntnisse auf Fachkonferenzen präsentiert. Es hat eine starke Vernetzung mit internationalen Partnern auf dem Gebiet der Frühwarnung stattgefunden die in verschiedenen Formaten fortgeführt werden.

Einleitung

Ausgangslage des Projekts

Durch das Vorgängerprojekt gab es bereits ein existierendes Frühwarnsystem, welches mit Hilfe einfacher statistischer Methoden Signale generiert und diese in einem Berichtsformat zusammenfasst. Die Analyse geht dabei aber nicht über die Meldedaten und einer einfach Zeitreihenanalyse zur Erkennung von möglichen Ausbrüchen hinaus.

Eine Vielzahl von hilfreichen Datenquellen wurde bisher nicht nutzbar gemacht und die neuen Möglichkeiten durch Methoden des Maschinellen Lernens wurden noch nicht erschlossen. Ebenso wurde die Möglichkeit der interaktiven Exploration mit Hilfe von neuartigen Berichtsformaten und Dashboards noch nicht ausgeschöpft.

Ziele des Projekts

Diese Lücken sollen im Signale 2.0 Projekt geschlossen werden, in dem zum einen relevante Datenquellen zum Zwecke der Frühwarnung und Surveillance bereitgestellt werden sollen und neuartige Methoden des maschinellen Lernens erforscht und angewendet werden sollen mit denen weiterführende Analysen durchgeführt werden können. Die Daten aus verschiedenen Datenquellen sollen zusammen mit den Ergebnissen in interaktiven Dashboards zusammengeführt werden welche den Nutzern ein umfassenden Überblick auf die epidemiologische Lage erlaubt.

Projektstruktur (Projektaufbau, Strukturen, Verantwortlichkeiten)

Das Projekt wurde in drei größere Teile und neun Arbeitspakete aufgeteilt, die alle zum größten Teil eigenständig bearbeitet oder zumindest begonnen werden können. Ein Teil beschäftigt sich mit der Erforschung und Anwendung neuer Analysemethoden, während in einem weiteren Teil moderne Methoden des maschinellen Lernens angewendet werden um neue Datenquellen zu erschließen. Ein dritter Teil soll die Daten und Ergebnisse in geeigneter Form zusammen darstellen. Es gab übergreifende Arbeitspakete die sich mit der Integration, Evaluierung und Kommunikation beschäftigen.

Die inhaltliche Projektleitung wurde von einem Mitarbeiter übernommen, musste aber auf Grund von Personalwechsel einmal neu besetzt werden. Teilprojekte wurden, wenn möglich, von 2 Mitarbeitenden betreut die sich fachlich und methodisch ergänzen. Projekte die mit Studierenden durchgeführt wurden, wurden ebenfalls von jeweils 2 Mitarbeitenden betreut. Für die Evaluierung und den Entwurf des Konzepts des zusammenführenden Dashboards wurde allerdings im gesamten Projektteam zusammengearbeitet.

Erhebungs- und Auswertungsmethodik

Anforderungen an Auswertungen und Berichtsformate wurden durch einen mehrstufigen Feedback-Prozess mit Hilfe von Fragbögen und Interviews erhoben. Für alle entwickelten und angewandten Methoden wurden Evaluierungen durchgeführt. Für Ausbruchserkennungsalgorithmen im speziellen wurde sogar ein Evaluierungsframework entwickelt.

Eine Anforderung an alle Arbeitspakete war das Anfertigen von Dokumentationen der Methoden, dies wurde durch das Anfertigen von Manuskripten entweder in Form von Masterarbeiten oder zur Einreichung in einer Fachzeitschrift bzw. Fachkonferenz erfüllt. Viele Produkte wurden sowohl intern

als auch Kongressen und Seminaren für die Fachöffentlichkeit präsentiert und der Programmcode auf geeigneten Plattformen wie github und gitlab öffentlich zur Verfügung gestellt.

Durchführung, Arbeits- und Zeitplan

Die Laufzeit des Projekts wurde von 3 Jahren auf 4 Jahren kostenneutral verlängert und trotzdem konnte die Gesamtfördersumme nicht vollständig verwendet werden. Zu Beginn des Projekts konnten 3 der 4 Stellen sofort besetzt werden, allerdings konnte die verbleibende Stelle erst nach einem Jahr und 2 Bewerbungsrunden besetzt werden. Weiterhin kam es zu einem zwischenzeitlichem Personalwechsel für 6 Monate wegen einer Entsendung zur WHO. Außerdem wurden 3 der 4 Stellen vorzeitig beendet durch Wechsel in andere Projekte. Diese teilweise kurzfristigen Personalwechsel führten dazu das einige Teilprojekte nicht wie geplant zum Abschluss gebracht werden konnten.

Ein großer Teil des Projekts wurde weiterhin durch die veränderten Anforderungen während der Covid-19 Pandemie beeinflusst. Die Zusammenarbeit mit Fachkollegen an der EpiWebapp konnte so kaum fortgeführt werden. Außerdem waren die Projektmitarbeitenden in verschiedene Aufgaben zur Bekämpfung der Pandemie mit eingebunden weshalb einige Teilprojekte nicht abgeschlossen werden konnten. Allerdings führte die Pandemie auch zu neuen Kooperationen sowohl innerhalb mit Fachkollegen als auch extern mit Modellierungsgruppen. Dadurch sind spannende Produkte entstanden die so nicht geplant waren aber auch für die Arbeit über die Pandemie hinaus nützlich sein werden. Zudem waren die Entwicklungen während der Pandemie ein guter Hinweis das die Erfahrungen aus den bereits bearbeiteten Teilprojekten hilfreiche Erkenntnisse geliefert haben und viele Prozesse erleichtert haben.

Trotz der Probleme mit Personalwechsel und Pandemie wurden die meisten Ziele erreicht und die verschiedenen Arbeitspakete zu einem zufriedenstellenden Abschluss gebracht. Alle methodischen Arbeitspakete wurden erfüllt und sogar mehr Teilprojekte und wissenschaftliche Arbeiten angefertigt als vorher angenommen. Die Entwicklung des Dashboards wurde zwar wie geplant konzipiert und entwickelt, d.h. in enger Zusammenarbeit mit den Fachkollegen und Fachkolleginnen, allerdings konnte die Veröffentlichung und endgültige Evaluierung aus oben genannten Gründen nicht mehr durchgeführt werden. Dafür wurde aber eine Vielzahl von zusätzlichen Dashboards für Teilaufgaben angefertigt die sowohl intern als auch teilweise extern in Verwendung sind. Das Arbeitspaket zur Personalisierung wurde schon im ersten Abschnitt des Projekts aus datenschutzrechtlichen Gründen

Ergebnisse

Im Rahmen des Signale 2.0 Projektes wurde ein Ausbruchsinformationssystem entwickelt, das Epidemiologinnen und Epidemiologen des ÖGD dabei unterstützt, Auffälligkeiten zu entdecken, zu analysieren und angemessen zu handeln. Dafür wurden zum einen Methoden entwickelt, erforscht und evaluiert, die Auffälligkeiten und Trends in Meldedaten erkennen, ggf. unter Verwendung externer Datenquellen. Zum anderen wurden neue Informationsquellen erschlossen, indem externe Datenquellen wie z.B. Wetter-Daten oder Berichte über Ausbrüche extrahiert, analysiert und bereitgestellt werden. Die Ergebnisse der entwickelten Methoden zusammen mit den Meldedaten und externen Daten werden dann in einem Dashboard so dargestellt werden, dass Epidemiologinnen und Epidemiologen einen umfassenden Eindruck der aktuellen epidemiologischen Lage erhalten und gezielt erforschen können. Sowohl beim Entwurf als auch bei der Evaluation der verschiedenen

Teilschritte wurden mögliche Nutzer, also Epidemiologinnen und Epidemiologen des ÖGD und andere Experten, eingebunden. Dafür wurden Fachkreise durchgeführt, Netzwerktreffen arrangiert und Publikationen angefertigt.

Methodenentwicklung

Im Vorgänger-Projekt wurde bereits ein System zur automatischen Ausbruchserkennung etabliert. Dieses System nutzte hauptsächlich klassische statistische Verfahren zur Ausbruchserkennung. Im ersten Projektabschnitt wurden neuartige Methoden erforscht, die zum einen andere Datenquellen integrieren als auch weitere räumliche und zeitliche Effekte berücksichtigen. Diese Methoden wurden in Tools implementiert und deren Performance evaluiert. Im zweiten Projektabschnitt wurden weitere Methoden erforscht, die neue kontextuelle Informationen erbringen und neue Datenquellen nutzbar machen.

Supervised Learning

Eine wichtige Informationsquelle wurde in den bisher verwendeten Methoden noch nicht berücksichtigt, neben den gemeldeten Fällen liefern die Gesundheitsämter über `SurvNet@RKI` nämlich auch Informationen zu Ausbrüchen. Ausbrüche sind hier Sammlungen von Fällen, die von den Gesundheitsämtern für ein bestimmtes Ereignis gemeinsam betrachtet werden. In `SurvNet@RKI` haben Fälle eines Ausbruchs eine gemeinsame Ausbruchs-ID. Diese Information wird nun genutzt um ein Hidden Markov Modell (HMM) zu trainieren, welches dadurch lernen soll, Ausbrüche in den Meldedaten besser zu erkennen. Dabei werden weiterhin wöchentlich aggregierte Zeitreihen generiert und dann die Wochen, welche Ausbruchsfälle enthalten, gekennzeichnet, womit der Algorithmus lernt, wie viele Fälle in einer Ausbruchswoche bzw. einer normalen Woche zu erwarten sind und sogenannte Übergangswahrscheinlichkeiten, also z.B. wie wahrscheinlich es ist in einer Ausbruchswoche zu sein, wenn die Vorwoche ebenfalls eine Ausbruchswoche war oder eben eine normale Woche. Aus diesen beiden Informationen berechnet das HMM eine Wahrscheinlichkeit für eine Woche, ob sie eine Ausbruchswoche ist. Dies ist für Epidemiologinnen und Epidemiologen besser verständlich, als die Ergebnisse der bisher verwendeten Algorithmen, mit p-Werten eines statistischen Tests mit der Nullhypothese dass kein Ausbruch vorliegt. Neben der besseren Interpretierbarkeit hat das HMM weitere Vorteile. So produziert die neue Methode bei gleicher Sensitivität weniger falsch-positive Signale. Weiterhin erlaubt es das Modell, weitere Formen von Experteninformationen zu integrieren, ein Beispiel dafür wäre das Feedback, dass direkt (welche Signale wurden als Ausbruch klassifiziert oder als irrelevant betrachtet) oder indirekt (welche Signale wurden weiter erforscht oder ignoriert) aus der Benutzung des Dashboards extrahiert werden kann.

Die Beschreibung der Ausbruchsdaten, der Methode und mögliche weitere Anwendungen wurden in einem Artikel im Detail beschrieben, auf Grund eines Personalwechsels wurde dieser aber nicht zum peer-review eingereicht. Die Ergebnisse wurden auf der ISDS-Konferenz im Februar 2019 vorgestellt. Eine vergleichbare Methode wurde von Kollegen und Kolleginnen am RKI in einer Fachzeitschrift veröffentlicht (Veröffentlichung 6).

Raum-zeitliches Modell

In den bisherigen Methoden zur Ausbruchserkennung wurde für jede generierte Zeitreihe bzw. für jede untersuchte Filterkombination (Fallzahlen aggregiert nach Kalenderwoche und gefiltert nach Meldeort, Altersgruppe, Krankheit, Subtyp, ...) ein Modell trainiert und eine Erwartung für eine bestimmte Kalenderwoche (z.B. in den Signalberichten die aktuellen 6 Wochen) berechnet. Dies vernachlässigt aber zeitliche und räumliche Abhängigkeiten, z.B. könnten hohe Fallzahlen in der

Vorwoche in benachbarten Landkreisen zu einer hohen Fallzahl in der aktuellen Woche in einem Landkreis führen. Es wurde eine neue Methode entwickelt, die genau diese Zusammenhänge modelliert und zusätzlich Faktoren wie Trend, Saisonalität und demographische Faktoren berücksichtigt. Die Güte der Vorhersagen sind vergleichbar mit den besten state-of-the-art Modellen und das neu entwickelte Modell hat den Vorteil, dass sich damit auch die Unsicherheit der Vorhersage bestimmen lässt. Weiterhin kann durch Visualisierung der gelernten Modellparameter interpretiert werden, wie sich die jeweilige Krankheit in Raum und Zeit ausbreitet, welche Saisonalität sie besitzt und welche demographischen Faktoren eine Rolle spielen. Es wird erwartet, dass dadurch zum einen ein besseres Verständnis der Krankheit selbst aber im speziellen auch über die auffällige Situation zu erhalten, d.h. es wird nicht nur berichtet, dass etwas auffällig ist, sondern auch was die Auffälligkeit ausgelöst hat. Die Methode ist flexibel genug um weitere Faktoren wie z.B. das Wetter zu integrieren.

Die Methode wurde in Zusammenarbeit mit Olivera Stojanovic und Johannes Leugering entwickelt. Frau Stojanovic hat in der Signale Gruppe einen zwei-monatigen wissenschaftlichen Besuch im Rahmen ihrer Doktorarbeit an der Universität Osnabrück absolviert. Die Projektidee entstand während des Hackathons an der Universität Osnabrück, woraufhin eine Zusammenarbeit vorgeschlagen wurde. Die Beschreibung der Methode und Resultate wurden in einem Manuskript zusammengefasst (Veröffentlichung 1) und in einem peer-reviewten Journal veröffentlicht, der Programmcode ist ebenfalls öffentlich verfügbar.

Evaluierung von Ausbruchserkennungsmethoden

Im Frühjahr 2019 hat Rüdiger Busche seine Masterarbeit „Systematic Evaluation and Optimization of Outbreak-Detection Algorithms Based on Labeled Epidemiological Surveillance Data“ abgeschlossen. Durch Hyperparameter-Optimierung, bei der gemeldete Ausbrüche als Labels verwendet wurden, konnten die optimalen Parameter für eine Reihe von gewöhnlichen Ausbruchserkennungsalgorithmen berechnet werden. Dabei wurde herausgefunden, dass diese Parameter in der Regel signifikant unterschiedlich zu den Standardeinstellungen der Algorithmen sind. Darüber hinaus konnte ein konsequenter Vergleich von Algorithmen erreicht werden. Der Standard „FarringtonFlexible“ Algorithmus erzielte die beste Leistung, welche insgesamt dennoch nicht sehr hoch ist im Vergleich zu anderen Anwendungsbeispielen des maschinellen Lernens. Siehe Abbildung 1.

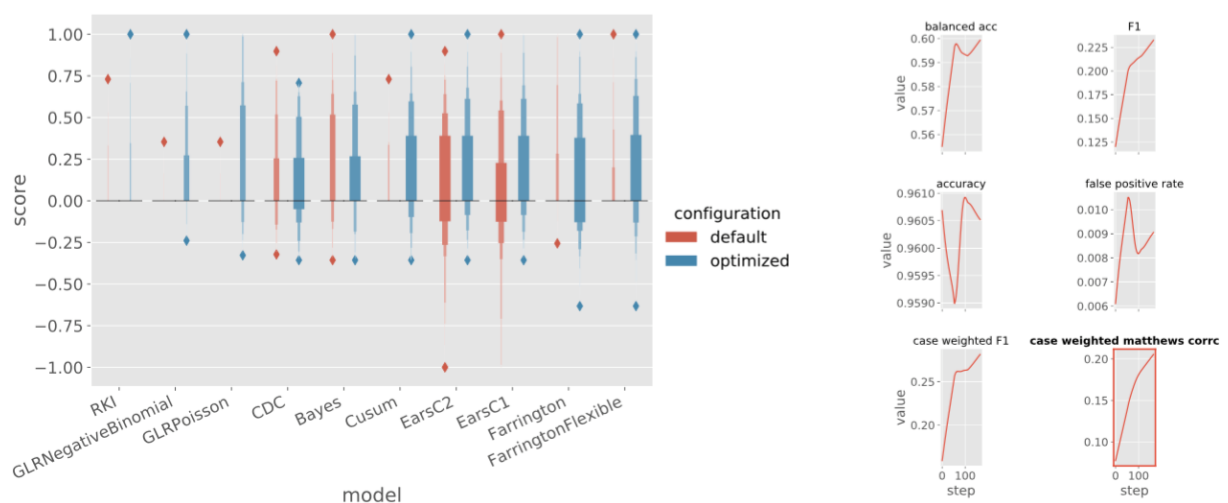


Abbildung 1: Optimierung von Ausbruchserkennungsalgorithmen. Links: Score (weighted Matthews

Correlation Coefficient) vor und nach Optimierung. Rechts: Evolution unterschiedlicher Scores während der Optimierung für den Algorithmus FarringtonFlexible.

Generative Adversarial Networks

Die Anwendung von Generative Adversarial Networks (GANs) als Lösung für die Generierung von realistischen, anonymen Daten wurde erforscht. Hier geht es darum, zwei neuronale Netzwerke gleichzeitig zu trainieren: Das erste generiert Daten, das zweite vergleicht diese mit Realdaten. Der Ansatz ist ähnlich zu dem in Computersimulationen, außer, dass keine Annahmen über die Daten gemacht werden müssen und feine Korrelationen prinzipiell gelernt werden können. Realistische epidemiologische Zeitreihen konnten dadurch generiert werden. Siehe Abbildung 2.

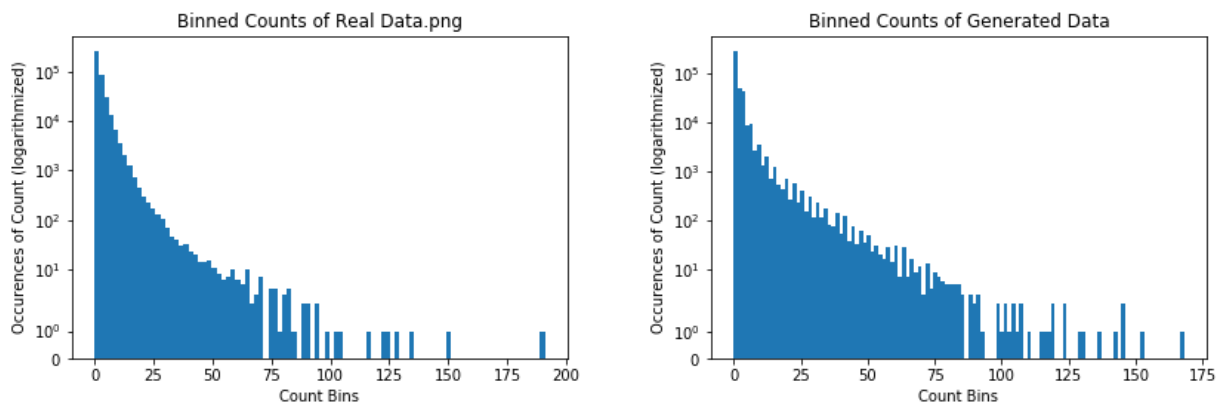


Abbildung 2: Generierung von epidemiologischen Daten durch ein GAN. Verteilung von wöchentlichen Fallzahlen in den realen Daten (links) und in den generierten Daten (rechts).

Hierarchisches Forecasting mit Erklärungen

Eine Vorhersage der weiteren Entwicklung der Fallzahlen ist eine wichtige Kontextinformation zum besseren Verständnis einer epidemiologisch relevanten Situation. Außerdem können solche Vorhersagen auch für sich genommen interessante Einsichten und Hinweise auf mögliche Veränderungen liefern. Allerdings ist dies nur möglich, wenn die Vorhersagen interpretierbar sind, d.h. die Gründe und Einflussfaktoren für Vorhersagen auch einsehbar sind. Ein weiterer Anspruch für Vorhersagen in unserem Kontext ist die hierarchische Konsistenz, d.h. Vorhersagen auf niedrigen Ebenen stimmen mit Vorhersagen auf höheren Ebenen überein, z. B. sollte die Summe der Vorhersagen aller Bundesländer gleich der Vorhersage für ganz Deutschland sein. Ebenso sollten auch die Erklärungen der Vorhersagen auf verschiedenen Ebenen konsistent sein. Dafür haben wir einen neuartigen Ansatz entwickelt der interpretierbares hierarchisches Vorhersagen erlaubt. Die Methode wurde so implementiert, dass sie im Routinebetrieb eingesetzt und in das Reporting eingebunden werden kann. Die Methodik und Ergebnisse sind in der Masterarbeit von Adrian Lison, die am RKI im Rahmen unseres Projekts entstanden ist, zusammengefasst.

Fallbasierte Ausbruchserkennung durch Dimensionsreduktion

Bisher werden Auffälligkeiten in Infektionszahlen durch Zeitreihenanalysen erkannt, d.h. die Meldedaten werden stratifiziert (z.B. Bundesland, Altersgruppe) und nach Meldewoche aggregiert, diese Zeitreihe wird dann mit einem klassischen Ausbruchserkennungsalgorithmus analysiert. In früheren Analysen konnten wir zeigen, dass auf univariaten Zeitreihen basierende Ansätze nicht alle Arten von Ausbrüchen erkennen können. Epidemiologen erkennen Zusammenhänge auf Basis von Daten auf Fallebene, d.h. Gemeinsamkeiten in den Eigenschaften von Fällen. Diesen Ansatz haben wir versucht auf eine computer-gestützte Clustererkennung zu übertragen. Dafür definieren wir Distanzen zwischen Fällen in dem wir Distanzmaße für alle vorhandenen und relevanten

Fallmerkmale definieren. Basierend auf diesen Distanzen und bekannten Zusammenhängen zwischen Fällen können wir zum einen eine semi-überwachte Clustererkennung durchführen sowie eine Visualisierung der Ähnlichkeit der Fälle mittels Dimensionsreduktion durchführen. Beide Ergebnisse können dann in das Reporting eingebunden werden. Die Methodik und Ergebnisse sind in einer Masterarbeit beschrieben, die von Kristina Kohn während ihres Aufenthalts am RKI unter unserer Leitung entstanden ist.

Kofaktoren und Integration

Ein weiteres Ziel des Signale 2.0 Projektes ist es, zusätzliche Datenquellen zu erschließen und für eine breite Anwenderschaft bereitzustellen. Zum einen sollen diese Datenquellen die Methoden zur Signalerkennung verbessern und zum anderen nützliche Kontextinformationen liefern, um die epidemiologische Lage besser einzuschätzen. Die Daten sollen dabei sowohl im Dashboard als auch teilweise in Form einer Schnittstelle zur Verfügung gestellt werden.

Wetterdaten

Die Nützlichkeit von Wetterdaten zur Vorhersage von Fallzahlen für verschiedene Infektionskrankheiten wurde im Vorprojekt erforscht, außerdem wurde ein R-Paket entwickelt, welches diese Daten extrahiert und für den internen Gebrauch aufbereitet. In diesem Abschnitt des Signale 2.0 Projektes wurde die Analyse dieser Daten konkretisiert. In Zusammenarbeit mit Wiebke Hellenbrand und Teresa Kreusch aus dem Fachgebiet 33 des RKI wurde ein Dashboard entwickelt, das helfen soll, den Zusammenhang zwischen Wettereinflüssen und vektorbasierten Krankheiten wie FSME, Hantavirus-Erkrankung oder Borreliose zu erforschen. Das Dashboard steht RKI-intern zur Verfügung und soll in einem nächsten Schritt für die Landesstellen der Bundesländer freigeschaltet werden.

Kartenmaterial

In vielen der im Projekt entwickelten Dashboards werden Karten angezeigt und auch für einige Methoden werden geographische Informationen benötigt (Nachbarschaften zwischen Landkreisen, oder Fläche für die Berechnung der Populationsdichte). Um nicht in jedem Teilprojekt diese Informationen ablegen zu müssen, wurde ein zentraler Dienst entwickelt, der diese Informationen über eine Webschnittstelle bereitstellt. Über HTTP-Requests kann jeder Rechner und Server innerhalb des RKI-Netzwerkes diese Geoinformationen im GeoJSON- und TopoJSON- Format abrufen. Neben der Wahl des Formats erlaubt der Dienst auch eine Auswahl über die Qualität (low oder middle für Webdienste, high oder raw für genauere Berechnungen), das Gebietslevel (Landkreis, Bundesländer, Deutschland) und den Datenstand (für jede Kreisgebietsreform wird eine neue Version erstellt). Die Schnittstelle wird bereits in verschiedenen Projekten im Institut verwendet.

Nachrichten

Auss Abbood hat im Frühjahr 2019 seine Masterarbeit „Automatic Information Extraction and Relevance Evaluation of Epidemiological Texts Using Natural Language Processing“ abgeschlossen. Er konnte anhand einer Reihe von Methoden der Verarbeitung natürlicher Sprache und maschinellen Lernverfahren Online-Nachrichtentexte analysieren, um daraus automatisch epidemiologische Information zu extrahieren und die epidemiologische Relevanz einzelner Nachrichten zu berechnen. So wird die Ereignis-basierte Surveillance am RKI unterstützt.

Die Analysen wurden als Analysepipeline (siehe Abbildung 3) in eine Webanwendung namens EventEpi integriert und RKI-intern zur Verfügung gestellt. Der Programmcode ist öffentlich verfügbar

und die Methoden und Ergebnisse sind in einer Veröffentlichung in einer Fachjournal zusammengefasst (Veröffentlichung 2).

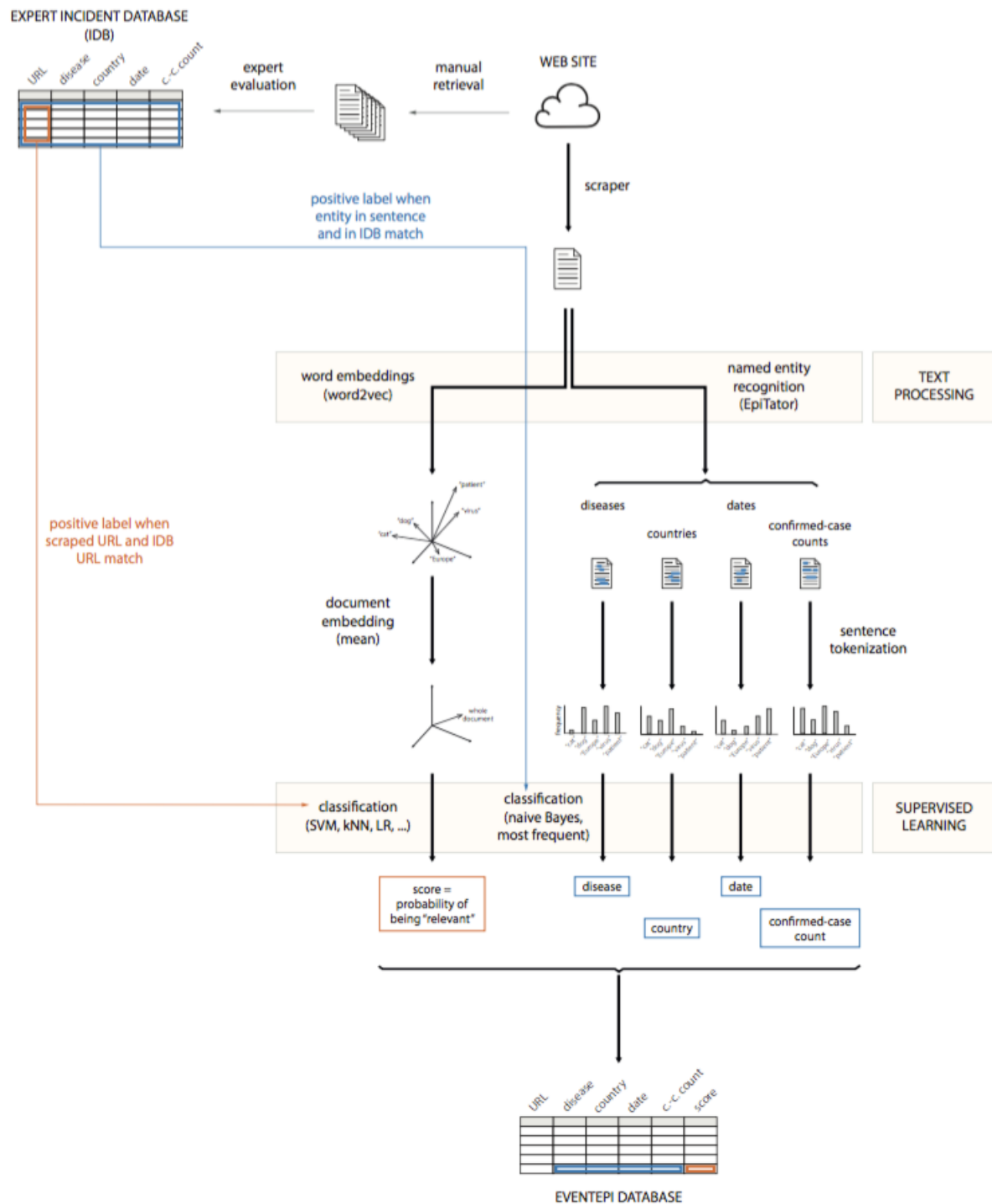


Abbildung 3: Analyse Pipeline von EventEpi.

Überwachtes Lernen von Syndromen

Im Rahmen des ESEG Projektes werden Notaufnahmedaten aus mehreren Krankenhäusern erhoben. Um diese Daten für eine syndromische Surveillance zu nutzen müssen allerdings passende Syndrome definiert werden. Üblicherweise bestehen solche Definitionen meist aus einer Kombination von ICD Codes, Vorstellungsgründen und anderen Parametern (z.B. auffällige Vitalparameter). Notaufnahmedaten sind aber meist nicht vollständig bzw. werden einige Angaben erst mit Verzug

nachgetragen. Wir haben daher ein Verfahren entwickelt welches basierend auf einer Expertendefinition eine angepasste Syndromdefinition lernt welches mit fehlenden Werten umgehen kann und trotzdem eine vergleichbare Performance, sowohl auf Fall- als auch auf Zeitreihenebene, liefert. Damit kann die Surveillance von Notaufnahmedaten als Kontext zur Surveillance der Meldedaten betrachtet werden. Die Methoden und erzielten Ergebnisse wurden 2020 in der Masterarbeit von Birte Wagner beschrieben, auf einer wissenschaftlichen Konferenz vorgetragen und ein wissenschaftlicher Artikel ist angefertigt der demnächst zum peer review eingereicht werden soll.

Syndromische Surveillance durch "Social Media" Daten

Suchergebnisse oder Mitteilungen in sozialen Medien können zeitnahe Hinweise auf epidemiologische Geschehen geben. Wir haben einen Ansatz entwickelt welcher Schnittstellen von Google Trends und Twitter nutzt um Häufigkeiten von bestimmten epidemiologisch relevanten Begriffen zu extrahieren und diese zu Zeitreihen von verschiedenen Syndromen zusammenzufassen. Wir konnten zeigen, dass diese Zeitreihen sehr gut mit Zeitreihen etablierter Surveillance Systeme übereinstimmen. Damit können Ergebnisse dieser syndromischen Surveillance auf "social media" Daten als Kontext zur Einordnung der epidemiologischen Lage genutzt werden. In der Masterarbeit von Florian Becker wurden die entwickelten Methoden und Ergebnisse zusammengefasst.

Dashboard und Evaluation

Das sichtbarste Produkt des Projekts sollte ein Dashboard zur Erkundung der Meldedaten und jeglicher Analyseergebnisse sein. Dabei wurden in jedem Schritt von der Konzeption, über Prototypen bis hin zur Evaluierung des finalen Produkts die Nutzer und Experten mit einbezogen. In der ersten Phase wurde eine Bestandsaufnahme und Evaluierung der bestehenden Produkte durchgeführt, sowie einige single-purpose Dashboards und Prototypen entwickelt, die später in das finale Dashboard eingefloßen sind. In der nächsten Phase wurden weitere Fachkreise durchgeführt und deren Erkenntnisse in ein Konzept für ein umfassendes Dashboard überführt und als Epi-WebApp umgesetzt.

Dashboards

Aus den erhobenen Anforderungen der Fachkreise und den Erkenntnissen mit den Dashboards der ersten Generation wurde ein Konzept für ein allgemeines Dashboard geschaffen. Dieses wurde für den internen Gebrauch entwickelt und bisher nur für einzelne Nutzer freigegeben. Durch die Belastungen und geänderten Prioritäten während der Pandemie wurde der weitere Veröffentlichungsprozess sowie deren Evaluation pausiert. Stattdessen wurden mehrere Dashboards für die Analyse der Covid Lage entwickelt und veröffentlicht.

Epi-WebApp:

Für Epidemiologinnen und Epidemiologen ist es schwierig die aktuelle Lage zu bewerten, weil in Abhängigkeit der Krankheit täglich viele Fälle hinzukommen, besonders wichtig ist hier:

- Wo sind die Hotspots des Krankheitsgeschehen?
- Ist das Krankheitsgeschehen bei den Landesstellen und Gesundheitsämtern bekannt
- Welche Fälle (Personen) sind konkret betroffen

Die *Epi-WebApp* erfüllt folgende Aufgaben:

- Zusammenfassen von verschiedenen Krankheiten in einer Analysegruppe, um auch Krankheitsübergreifende Analysen zu gewährleisten oder Analysen nur auf bestimmten Subtypen durchzuführen
- Wendet verschiedene relevante Analysen auf die gemeldeten Fällen an um **die aktuelle epidemiologische Lage** bezüglich einer Analysegruppe bewertbar zu machen
- Macht die aktuelle epidemiologische Lage für Analysegruppen anhand verschiedener Kennzahlen greifbar und somit bewert- und vergleichbar für Epidemiologen und Epidemiologinnen:
- **Fallzahl** differenziert nach einfacher Ausbruchszugehörigkeit
- Einschätzung der **Schwere** basierend auf bisherigen Beobachtungen
- Die gemeldeten **Cluster** mit räumlicher Ausdehnung; ergänzt um Cluster, die durch Analysen gefunden wurden
- **Vorhersage** der zu erwartenden Fallzahl
- Stellt die Fallzahlen aus verschiedenen Blickwinkeln dar um eine differenzierte Betrachtung zu ermöglichen:
- **Zeitlich** Fallzahlen nach Zeit => Steigen die Fallzahl oder sinken sie?
 - **Geografisch** Fallzahl nach Landkreis und/oder Bundesland => Gibt es einen Landkreis oder Bundesland, welche besonders betroffen sind?
 - **Demografisch** Fallzahl nach Alter und Geschlecht => Ist eine Altersgruppe oder ein Geschlecht besonders betroffen?
 - **Pathologisch** Fallzahl nach Subtyp => Macht ein konkreter Subtyp die Hauptlast aus?

Fallexplorer:

Epidemiologinnen und Epidemiologen haben die Möglichkeit Falldaten zu erkunden, indem sie mit Hilfe eines Filters bestimmen welche Datenmenge betrachtet wird und durch Hinzufügen von Visualisierungen definieren wie die Daten angezeigt werden

- Die betrachteten Fallzahlen können auf Einzelfälle abgebildet werden
- Die Untersuchung kann mit Kollegen und Kolleginnen geteilt werden um entdeckte Auffälligkeiten einfach zu kommunizieren und den Kollegen und Kolleginnen die Möglichkeit zu geben die Untersuchung weiter zu verfeinern

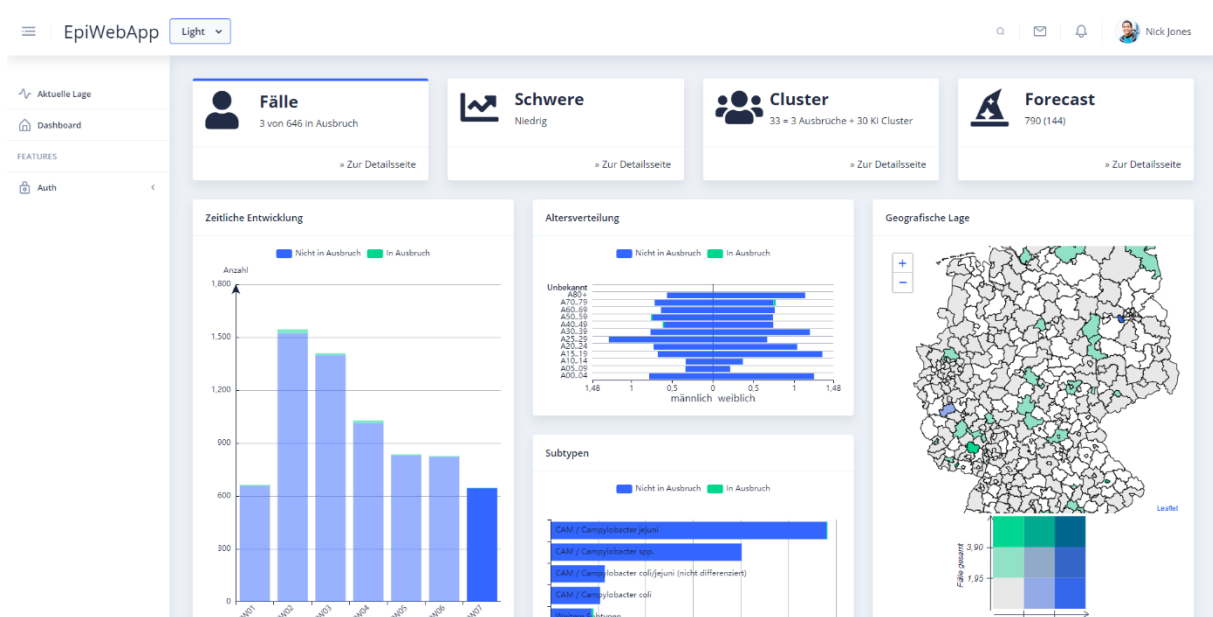


Abbildung 4: Screenshot der Epi-WebApp im Überichtsmodus in der Auswahl „Fälle“.

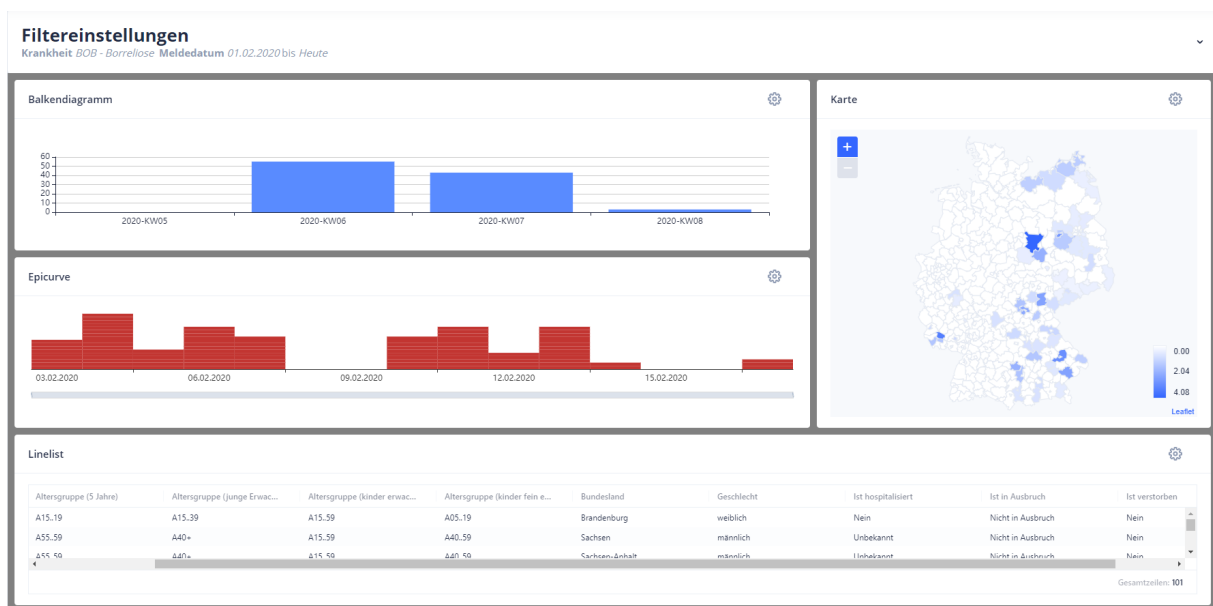


Abbildung 4: Screenshot des FallExplorer.

Covid-Signale-Dashboard:

Das COVID-Lagezentrum des RKI umfasst eine Position zur Ausbruchserkennung deren Aufgabe es unter anderem ist Landkreise zu identifizieren welche ein auffälliges Covid-Geschehen aufweisen. Zu diesem Zweck wurde ein Dashboard entwickelt welches tagesaktuell auf Landkreis-Ebene alle vorhandenen und relevanten Covid-Informationen zusammenträgt und nach einem Ampelsystem einordnet. Zu diesen Informationen gehört die 7-Tage-Inzidenz, die Veränderung zur Vorwoche, Positivenrate und ITS-Belegung. Die Einordnung in die Ampelstufen passiert dabei nach Experten-definierten Grenzwerten pro Indikator. Zusätzlich gibt es eine Gesamteinschätzung für jeden Landkreis. Zur Übersicht wird die Gesamteinschätzung in einer Deutschlandkarte dargestellt und die verschiedenen Indikatoren pro Landkreis in einer sortier- und filterbaren Tabelle aufgelistet. Für

jeden Indikator gibt es zusätzlich eine detailliertere Betrachtung. Die Erstellung dieses Dashboards in Zusammenarbeit mit dem Lagezentrum ergab wichtige Erkenntnisse welche Informationen und Darstellungen in einer Ausbruchssituation wirklich von Wichtigkeit sind.

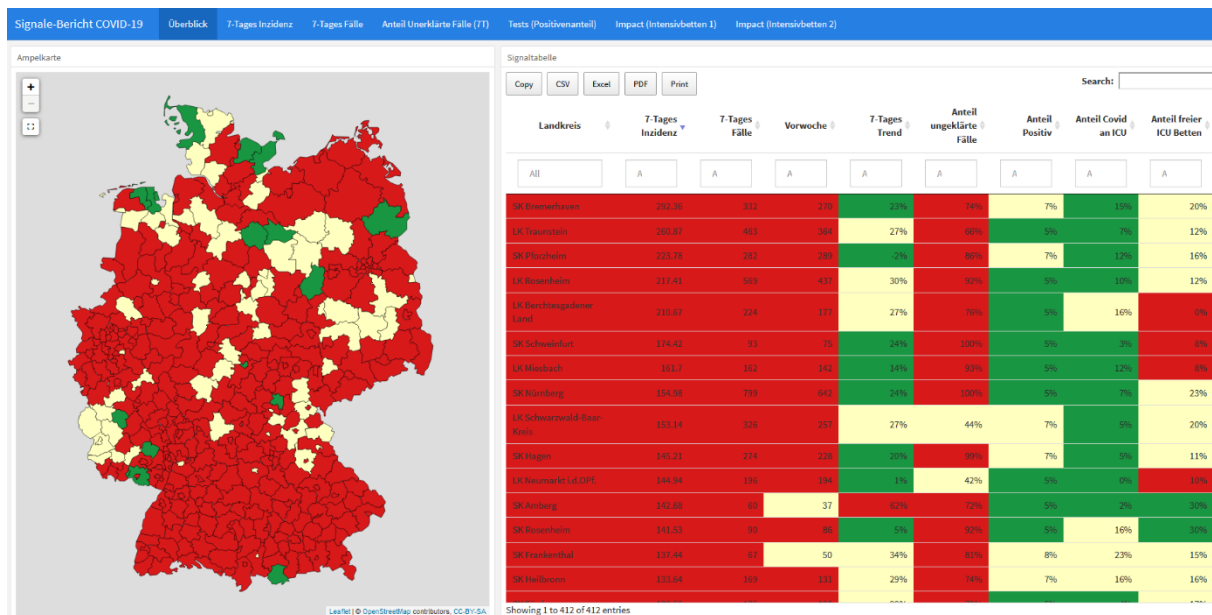


Abbildung 5: Screenshot des Covid-Signale-Dashboards.

Forecast-Hub-Dashboard:

In Zusammenarbeit mit der Forschungsgruppe von Johannes Bracher und Melanie Schienle am KIT, wurde eine Visualisierung der Ergebnisse des sogenannten Forecast Hubs (<https://kitmetricslab.github.io/forecasthub/forecast>) entwickelt. Der Forecast Hub selbst ist eine Plattform über den Modellierer Gruppen wöchentlich Vorhersagen bzgl. der Entwicklung der Covid-Fallzahlen und Todeszahlen übermitteln können (Veröffentlichung 5). Unser Dashboard erlaubt es nun diese Ergebnisse und deren Evaluierung in einer ansprechenden Visualisierung den Forschungsgruppen und der allgemeinen Öffentlichkeit zur Verfügung zu stellen. Diese Darstellung wurde ebenfalls für den ECDC-Forecast Hub (<https://covid19forecasthub.eu/>) von uns weiterentwickelt. Es gibt Planungen diese Art von Forecast-Hub auch über die Pandemie hinaus für andere Erreger zu nutzen.

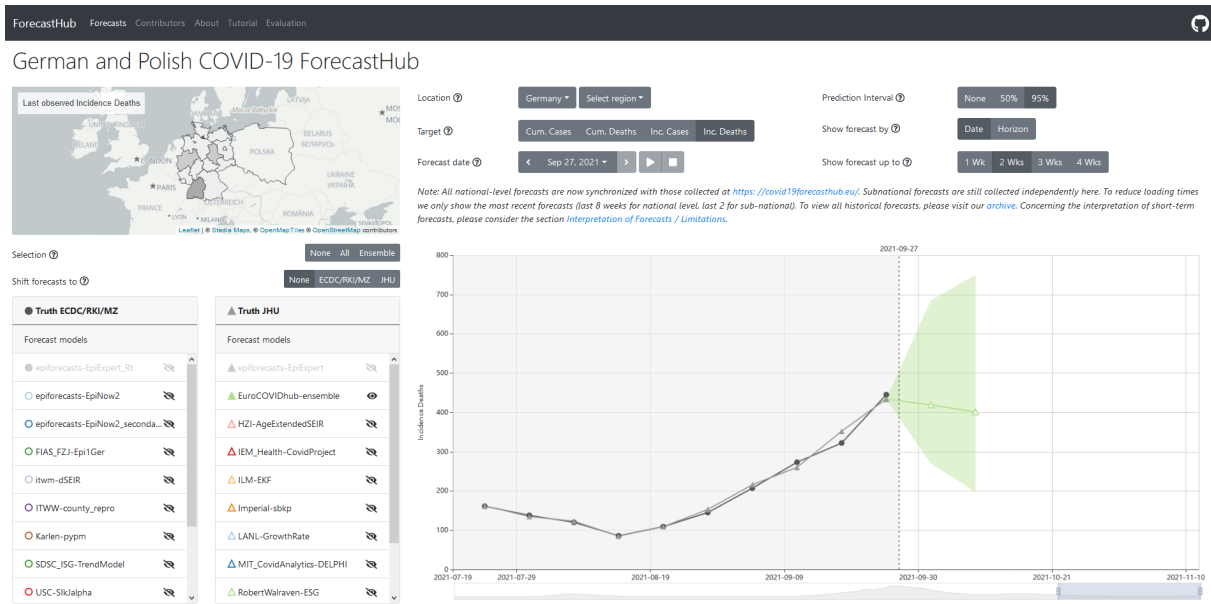


Abbildung 6: Screenshot des deutsch-polnischen Forecast-Hub-Dashboards.



Abbildung 7: Screenshot des ECDC Forecast-Hub-Dashboards.

Covid-Trend-Dashboard:

Neben dem Corona-Dashboard welches mit der Firma ESRI entwickelt wurde und die aktuelle Lage bzgl. der Covid-Fallzahlen abbildet, sollte auch ein Überblick über alle relevanten öffentlich verfügbaren COVID-Daten gegeben werden. Das von uns entwickelte Covid-Trend-Dashboard (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/COVID-19-Trends/COVID-19-Trends.html?_blob=publicationFile#/home) vereint 12 Indikatoren aus 6 Quellen zu einer Übersicht. Dabei sind alle Indikatoren einheitlich dargestellt mit einer Zeitreihe und Trendberechnung. Wenn möglich sind die Indikatoren auch nach Bundesland und Altersgruppe dargestellt. Auch bei der Erstellung dieses Dashboard wurden viele potentielle Nutzergruppen und

Stakeholder befragt, welche wichtige Erkenntnisse auch für die Erstellung anderer Dashboards von Nutzen war.

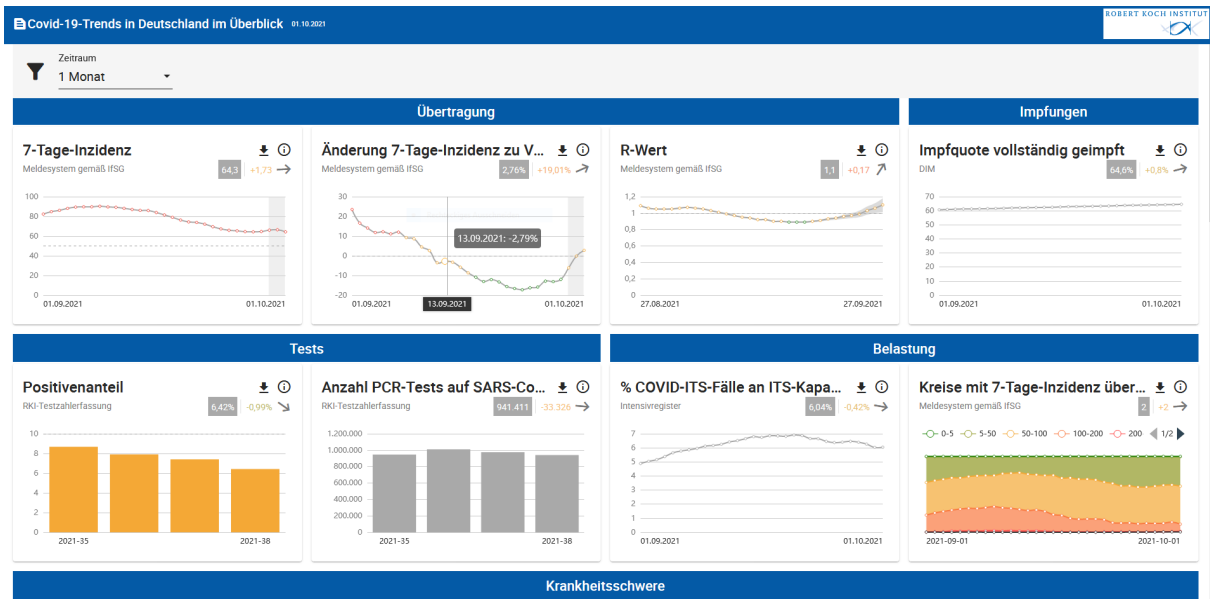


Abbildung 8: Screenshot des Covid-Trend-Dashboards.

Signale Dashboard

Das Signale Dashboard ist eine dynamische Variante der Signal Berichte die es erlauben soll die Ergebnisse der Signalerkennung sowie die IfSG Daten zu explorieren. Dies wird durch verschiedene Datenfilter (Zeitraum, Ort, Alter, Geschlecht, Subtyp) und die Verknüpfung der einzelnen Widgets ermöglicht.

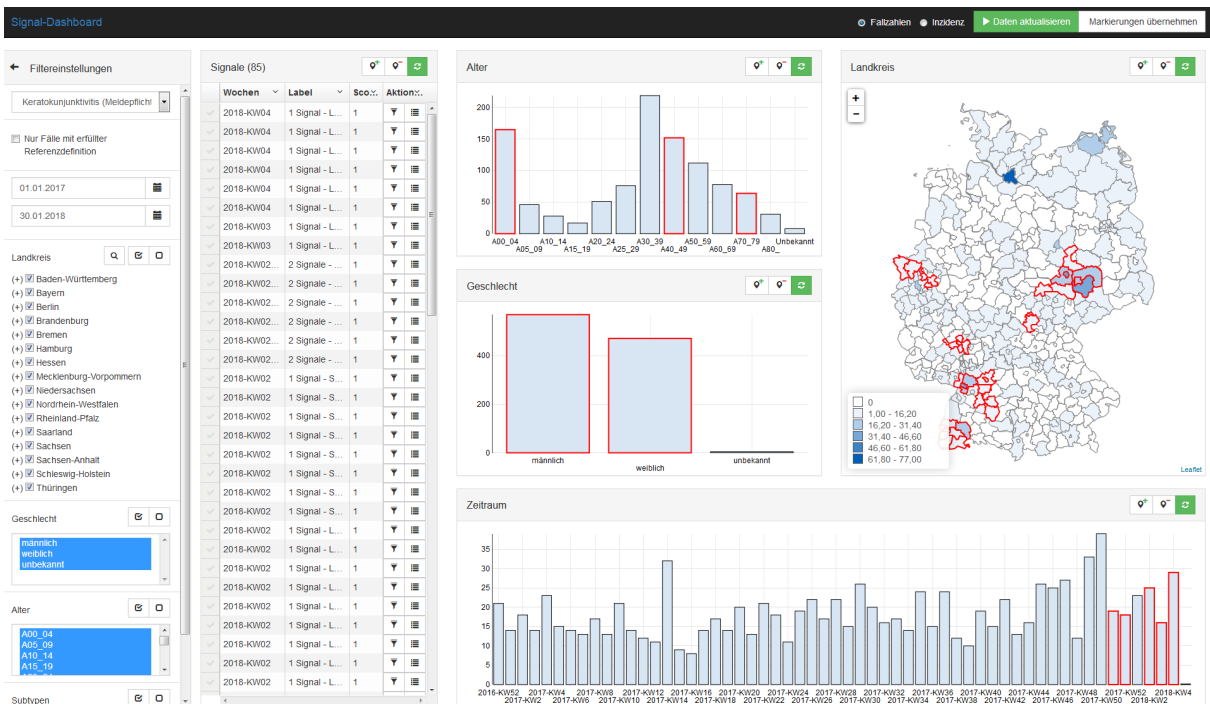


Abbildung 9: Screenshot des Signale Dashboards.

Influenza Dashboard

Die Mitgliedsländer der WHO sind dazu angehalten wöchentlich eine Einschätzung der Schwere der aktuellen Influenza-Saison abzugeben. Dazu gibt die WHO den sogenannte PISA-Leitfaden vor der die Bewertung bezüglich der Indikatoren „Transmissibility“, „Seriousness“ und „Impact“ vorsieht. Die Arbeitsgemeinschaft Influenza des FG-36 definiert diese Indikatoren mit Hilfe verschiedener Maßzahlen ihrer Surveillance Systeme. Dazu gehören z.B. Konsultationsinzidenz des AGI Systems oder Positivenrate der virologischen Surveillance. Das Influenza Dashboard stellt die genannten Indikatoren in geeigneten Zeitreihenvisualisierungen dar.

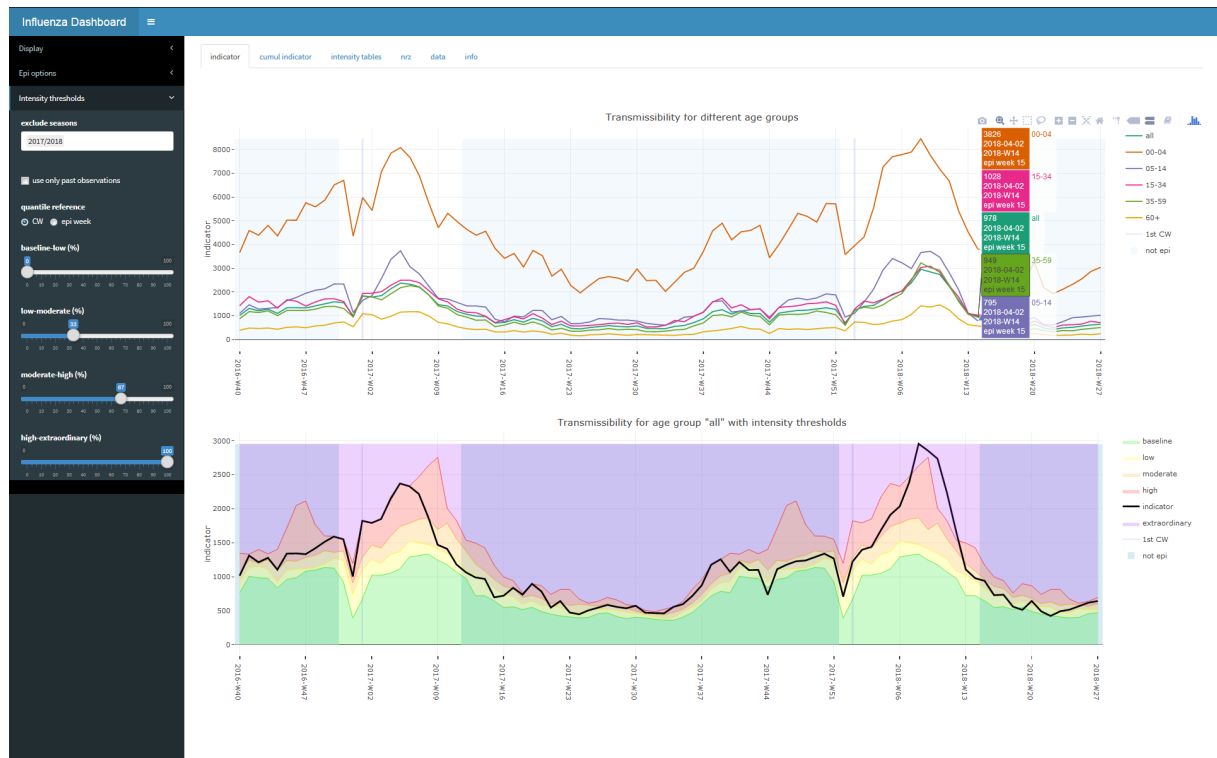


Abbildung 10: Screenshot des Influenza Dashboard in der Auswahl „Transmissibility“

Vector-borne Dashboard

Es wird davon ausgegangen, dass die Inzidenz von durch Zecken übertragbaren Krankheiten wie FSME und Borreliose stark durch externe Faktoren beeinflusst wird. Solche Einflüsse sind z.B. Wetter, Buchenmast, Bewaldung, Tierbestand. Das FSME Dashboard bietet Möglichkeiten zur Visualisierung und Modellierung des Zusammenhangs zwischen FSME bzw. Borreliose und verschiedenen Wettereinflüssen.

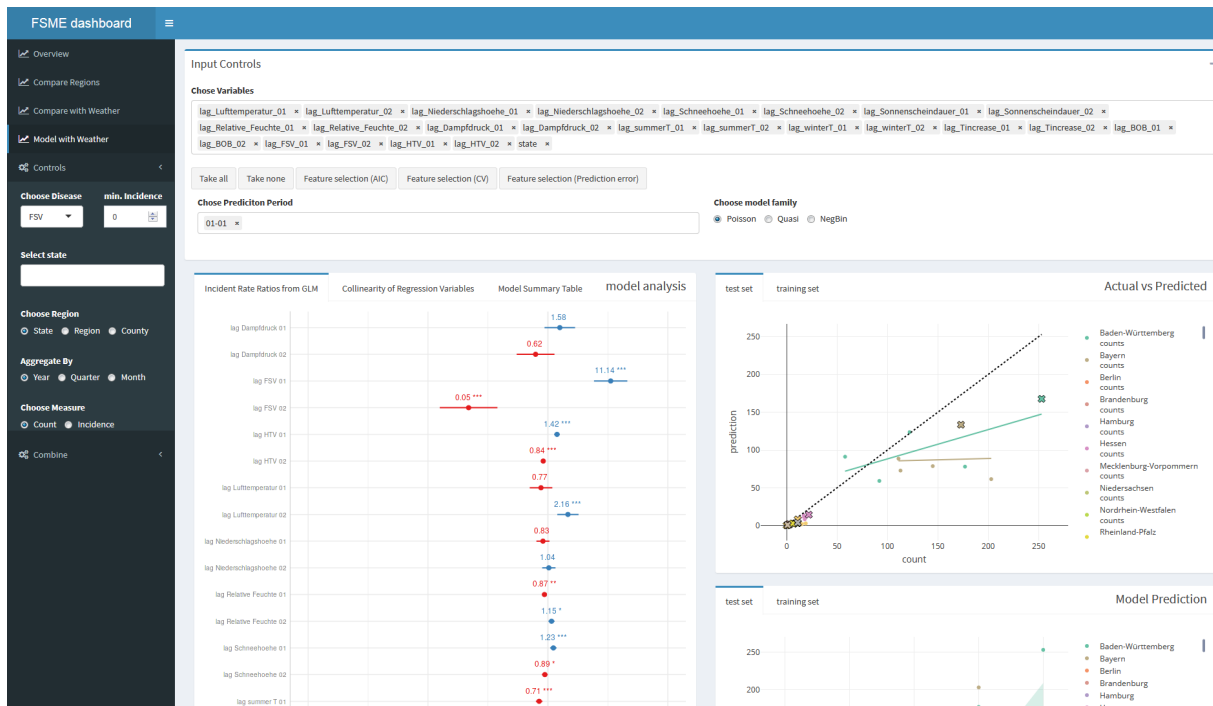


Abbildung 11: Screenshot des Vector-borne Dashboard im Model-Modus

Vacmap

In Anlehnung an die öffentliche Version der Vacmap (<http://vacmap.de/>) zur Darstellung der Impfquoten von Masern und Rotaviren erlaubt unsere Vacmap den Vergleich zwischen den Impfquoten und den Meldedaten durch zwei synchronisierte Karten.

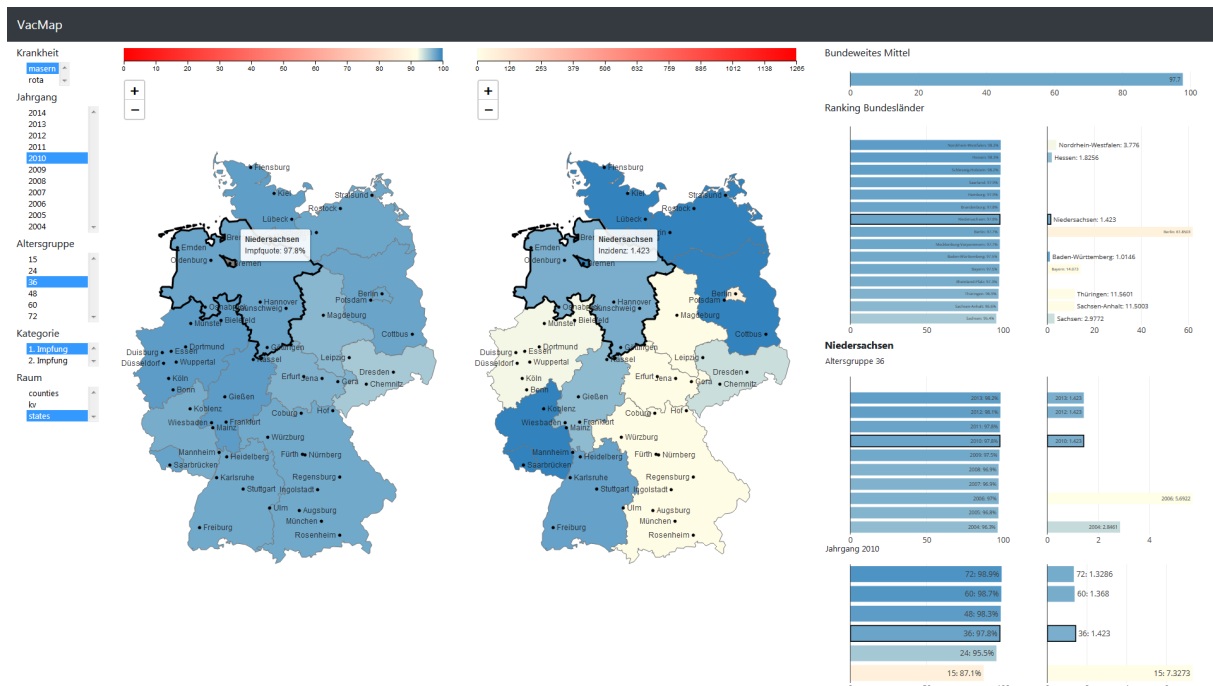


Abbildung 12: Screenshot des Vacmap Dashboard.

Kommunikation und Vernetzung

Neben der Veröffentlichung der entwickelten Methoden und weiterer Forschungsergebnisse soll eine Vernetzung mit Experten aus Epidemiologie sowie Statistik und Machine Learning erreicht werden. Im Vorgängerprojekt wurde dazu bereits ein internationaler Workshop organisiert, im ersten Teil des

Signale 2.0 Projektes wurde ein Hackathon organisiert und es wurden Konferenz- und Forschungsbesuche durchgeführt. Dabei wurden einige der im Projekt entstandenen Arbeiten bereits in mehreren Präsentationen vorgestellt und die Einreichung in wissenschaftlichen Journals wird vorbereitet. Im weiteren Teil des Projektes wurden neue Kontakte mit verschiedenen Modellierergruppen und Institutionen die sich mit neuartigen Machine-Learning Methoden beschäftigen geknüpft.

Vernetzung

Es gibt bereits einen aktiven Austausch mit einigen Instituten und Forschungsgruppen im In- und Ausland. Im Rahmen des Forecast Hubs gab es einen regen Austausch mit der Forschungsgruppe von Melanie Schienle am KIT und Sebastian Funk am LHSTM in England. Während der Pandemie haben wir zudem an der Modellierungsinitiative teilgenommen bei dem sich mehrere Modellierer-Gruppen aus Deutschland ausgetauscht haben. Mit DAKI-FWS (Daten- und KI-gestütztes Frühwarnsystem zur Stabilisierung der deutschen Wirtschaft) wurde ein neues Projekt beantragt welches in Kooperation mit dem Fraunhofer HHI, dem HPI und vielen weiteren Instituten und Firmen der Logistik Branche ein Frühwarnsystem etablieren soll welches auf verschiedene Risiken für die deutsche Wirtschaft hinweist.

- Im Rahmen des ESEG-Projekts hat die Projektgruppe an mehreren Workshops über den Stand und künftige Entwicklungen des Projekts teilgenommen.
- Im März 2019 wurde das RESST-Team (syndromische Surveillance) von Public Health England in Birmingham besucht. Im Dezember 2019 haben zwei Kollegen der Knowledge-Engineering-Gruppe der Technischen Universität Darmstadt die Projektgruppe besucht.
- Am 26. Juni 2019 hat sich die Projektgruppe mit Richard White, Statistiker des norwegischen Public-Health-Instituts FHI per Videokonferenz über automatisierte Ausbruchserkennung, Computer-Simulationen von Krankheitsverläufen und Deployment von Anwendungen ausgetauscht.
- Vom 11. bis zum 21. Juni 2019 war Savina Stoitsova vom National Center of Infectious and Parasitic Diseases (NCIPD) in Bulgarien im Rahmen eines durch das ECDC finanzierten "Senior Exchange"-Programms am RKI. Themen der Kooperation waren automatisierte Ausbruchserkennung und Berichterstattung in R.

Focus Group AI for Health

Die „Focus Group AI for Health“ (<https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx>) ist eine Initiative der Internationalen Fernmeldeunion (ITU) und der Weltgesundheitsorganisation (WHO), die darauf zielt, Standards für die Evaluierung von Ansätzen der künstlichen Intelligenz (KI) im Gesundheitsbereich zu entwickeln. In diesem Rahmen wird von uns die Topic Group Outbreaks koordiniert, die sich dem Benchmarking von Ausbruchserkennungsalgorithmen widmet. Regelmäßige Videokonferenzen wurden organisiert an denen etwa 30 Experten und Expertinnen aus der ganzen Welt teilgenommen haben. Ein Topic Driver Document, das das Vorhaben beschreibt, wird kontinuierlich ergänzt. Dazu gehört die Zusammenfassung des Wissenstands und Zusammenstellung der relevanten Literatur. Fortschritte der Topic Group werden regelmäßig auf Versammlungen der Focus Group vorgestellt. Eine Veröffentlichung des Evaluierungskonzepts in Form eines wissenschaftlichen Artikels und die Bereitstellung des Programmcodes zur Evaluierung als Python Bibliothek wird angestrebt.

Darüber hinaus wird in Working Groups an allgemeinen Leitlinien für das Benchmarking von KI-Ansätzen im Gesundheitsbereich mitgearbeitet.

Hackathon

Mit dem Ziel moderne Methoden des maschinellen Lernens auf klassische Probleme der Infektionsepidemiologie anzuwenden, wurde der Hackathon "hack4health - Cognitive Computing Hackathon on Epidemiology" veranstaltet. Der Hackathon wurde vom 12. bis 16. Februar 2018 an der Universität Osnabrück ausgetragen. Die Veranstaltung wurde gemeinschaftlich von der Signale Gruppe, dem Institut für Kognitionswissenschaften (IKW) der Universität Osnabrück und dem Gesundheitscampus Osnabrück organisiert.

Im Mai 2018 wurden in einer 2-tägigen Veranstaltung am RKI die Ergebnisse der Gewinner Projekte des Hackathons den Fach-Kollegen präsentiert. Dabei fand ein konstruktiver Austausch zwischen den Teilnehmenden des Hackathons und Epidemiologinnen und Epidemiologen statt. Während des Hackathons gab es Medienpräsenz in Form von Interviews und Berichterstattung in Fernsehen und Radio (NDR). Aus dem Hackathon resultierten ein zwei-monatiges Praktikum und zwei von uns betreute Masterarbeiten, deren Ergebnisse jeweils in einer wissenschaftlichen Veröffentlichung mündeten.

Gender Mainstreaming Aspekte

Bei der Analyse von Meldedaten wird neben anderen Parametern (z.B. Alter, Subtyp des nachgewiesenen Erregers) auch Angaben zum Geschlecht (männlich, weiblich, divers, unbekannt) berücksichtigt. In den Dashboards zu den Meldedaten werden geeignete Darstellungen verwendet die die Häufigkeiten nach allen Angaben zu Geschlecht stratifiziert darstellen. Allerdings erfassen nicht alle Meldesysteme das Geschlecht bzw. kann nicht bei allen Datenquellen eine getrennte Betrachtung nach Geschlecht durchgeführt werden.

Diskussion der Ergebnisse, Gesamtbeurteilung

Im Rahmen des Projekts wurden wie geplant verschiedene Aspekte der Analyse von Surveillance-Daten zum Zwecke der Ausbruchserkennung untersucht. Für den Bereich des Forecastings wurden wichtige praktische Erkenntnisse durch die Durchführung des Forecast Hubs erlangt. Die Vorhersage von Infektionsfällen ist für einen Zeitraum von 1-2 Wochen im Voraus möglich allerdings sehr stark von der Situation abhängig. Gerade in Situationen einer Trendumkehr durch Faktoren die nur schwer oder verzögert berücksichtigt werden können (z.B. neue Varianten, Maßnahmen) sind Vorhersagen eher unzuverlässig. Durch die Entwicklung eigener Methoden wurde klar das die Berücksichtigung möglichst vieler Ebenen (räumlich, demografisch) notwendig sind. Im Bereich Ausbruchserkennung ist durch die Evaluierung der bestehenden Algorithmen klar geworden dass die üblichen univariaten Zeitreihen Ansätze nicht ausreichend sind um Ausbrüche zu erkennen, hier werden multivariate oder fallbasierte Ansätze nötig werden. Die systematische Evaluierung selbst ist einer der größten Lücken im Bereich der Ausbruchserkennung. Insgesamt haben wir besonders durch die Betreuung von Masterarbeiten deutlich mehr Methodenentwicklung durchführen können als vorher geplant.

Die Bemühungen bzgl. der Extraktion und Bereitstellung von neuen Datenquellen haben gezeigt dass hier noch großes Potential liegt bisher ungenutzte Datenquellen für Surveillance-Zwecke aufzubereiten. Besonders public-health-nahe Daten wie ProMed oder Notaufnahmedaten haben

einen direkten Mehrwert der für die Beachtung in Routine Surveillance akzeptiert wird. Der praktische Nutzen von Daten aus sozialen Netzwerken muss erst noch gezeigt werden und die Akzeptanz ist ungewiss. In diesem Bereich haben wir unsere Ziele voll erfüllt und mehrere Datenquellen neu erschlossen bzw. bereitgestellt.

Durch die Erstellung von zahlreichen Berichten und Dashboards haben wir wichtige Erkenntnisse gewonnen die auch schon bei der Erstellung von Produkten in der Covid-19 Pandemie angewendet werden konnten. Unser Ziel eines einheitlichen Dashboards welches RKI-weit verwendet wird haben wir nicht vollständig erreicht. Das Dashboard wurde entwickelt konnte aber auf Grund anderer Aufgaben während der Pandemie und Personalwechsel nicht bis zur (internen) Veröffentlichung gebracht werden und damit konnte auch die Evaluierung durch Nutzer und Nutzerinnen nicht durchgeführt werden. Allerdings sind mehrere Dashboards entstanden die sowohl intern als auch extern in Nutzung sind, was eigentlich nur als Zwischenschritt geplant war.

Insgesamt haben wir unsere Ziele erfüllt auch wenn wir durch die Covid-19 Pandemie und Personalwechsel für einen großen Teil des Projekts stark eingeschränkt waren. Im Bereich der Datenbereitstellung und Dashboards wurden Produkte entwickelt die direkte Verwendung am RKI gefunden haben. Im Bereich der Methodenentwicklung wird eine intensivere Integration in bestehende Systeme erst noch angestrebt.

Verbreitung und Öffentlichkeitsarbeit der Projektergebnisse

Alle erwähnten Methoden und Ergebnisse wurden in Form von Manuskripten, Masterarbeiten und Artikeln in Fachzeitschriften dokumentiert und sind somit für die Öffentlichkeit verfügbar. Außerdem wurden einige der Teilprojekte auf Konferenzen und Netzwerktreffen vorgestellt.

Masterarbeiten

- a) Kohn, K.J. Individual-based detection of infectious disease outbreaks using machine learning techniques. 2020. Christian-Albrechts-Universität Kiel
- b) Wagner, B. Classifying Emergency Department Data to Improve Syndromic Surveillance: From Mixed Data Types to ICD Codes and Syndromes. 2020. Universität Bielefeld
- c) Lison, A. Interpretable Hierarchical Forecasting of Infectious Diseases. 2020, WWU Münster, Universität Münster
- d) Becker, F. Use of Social-Media Data in an Epidemiological Context. 2020, Beuth Hochschule für Technik Berlin
- e) Abbood, A. Automatic Information Extraction and Relevance Evaluation of Epidemiological Texts Using Natural Language Processing. 2019. Universität Osnabrück
- f) Busche, R. Systematic Evaluation and Optimization of Outbreak-Detection Algorithms Based on Labeled Epidemiological Surveillance Data. 2019. Universität Osnabrück

Veröffentlichungen in Fachzeitschriften

- 1) Stojanović O, Leugering J, Pipa G, Ghazzi S, Ullrich A. A Bayesian Monte Carlo approach for predicting the spread of infectious diseases. PLoS ONE 14 (12): e0225838. Epub Dec 18. <https://doi.org/10.1371/journal.pone.0225838>
- 2) Abbood A, Ullrich A, Busche R, Ghazzi S (2020) EventEpi—A natural language processing framework for event-based surveillance. PLOS Computational Biology 16(11): e1008277. <https://doi.org/10.1371/journal.pcbi.1008277>

- 3) Sarma N, Ullrich A, Wilking H, Ghazzi S, Lindner AK, Weber C, Holzer A, Jansen A, Stark K, Vygen-Bonnet S. Surveillance on speed: Being aware of infectious diseases in migrants mass accommodations – an easy and flexible toolkit for field application of syndromic surveillance, Germany, 2016 to 2017. Euro Surveill. 23 (40): pii=1700430. <https://doi.org/10.2807/1560-7917.ES.2018.23.40.1700430>
- 4) Ullrich, Alexander et al. Impact of the COVID-19 pandemic and associated non-pharmaceutical interventions on other notifiable infectious diseases in Germany: An analysis of national surveillance data during week 1–2016 – week 32–2020. The Lancet Regional Health – Europe, Volume 6, 100103. <https://doi.org/10.1016/j.lanepe.2021.100103>
- 5) Bracher, J., Wolfram, D., Deuschel, J., Ullrich A. et al. A pre-registered short-term forecasting study of COVID-19 in Germany and Poland during the second wave. Nat Commun 12, 5173 (2021). <https://doi.org/10.1038/s41467-021-25207-0>
- 6) Zacher B, Czogiel I (2022) Supervised learning using routine surveillance data improves outbreak detection of Salmonella and Campylobacter infections in Germany. PLOS ONE 17(5): e0267510. <https://doi.org/10.1371/journal.pone.0267510>

Präsentationen

Um die Methoden und Strategien, die im Projekt entwickelt wurden, vorzustellen, wurde an verschiedenen Kongressen teilgenommen und Präsentationen gehalten.

- Am 21. Januar 2019 wurde im Rahmen der Vorlesungsreihe "Infectious Disease Epidemiology" am RKI ein Vortrag zur automatischen Ausbruchserkennung gehalten.
- Auf der 2019 ISDS Annual Conference am 29.01.-01.02.2019 in San Diego (Kalifornien) mit dem Thema "Harnessing Data Science to Improve Population Health and Public Health Surveillance" wurden zwei Präsentationen gehalten. Die Titel der Präsentationen lauteten "Dashboards as strategy to integrate multiple data streams for real time surveillance" und "Supervised Learning for Automated Infectious-Disease-Outbreak Detection".
- Am 14. März 2019 wurden im Rahmen des PAE-Treffens die Signale-Berichte und weitere intern verfügbare Tools vorgestellt.
- Am 1. Juni 2019 wurde der Vorschlag für das Benchmarking von Ausbruchserkennung-Algorithmen der Focus Group AI for Health während des Meetings E am ITU in Genf vorgestellt.
- Die Ergebnisse zur Unterstützung der Ereignis-basierten Surveillance durch Verarbeitung natürlicher Sprache und maschinelle Lernverfahren wurde auf zwei Konferenzen vorgestellt: auf der 14. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi) in Ulm (11. bis 13. September 2019) und auf der ESCAIDE 2019 in Stockholm (27. bis 29. November 2019).
- Am 21. November 2019 wurden die Arbeiten im Projekt und das deutsche Meldewesen am National Institute of Infectious Diseases (NIID) in Tokyo vorgestellt.
- Das Projekt wurde eingeladen, beim Workshop "Infectious Disease Surveillance" der Universität Bern am 25. November 2019 die Arbeiten vorzustellen. Dort wurde über zwei Ansätze des maschinellen Lernens in der Infektionsepidemiologie berichtet.
- 8. und 9. Januar 2020: Workshop "AI4H standardized assessment framework & handling and assessment methods" am Fraunhofer HHI in Berlin, wo wir Generative Adversarial Networks (GANs) als Lösung für die Generierung von realistischen, anonymen Daten vorstellen.

- 16. bis 17. Januar 2020: Kick-Off des EJP One Health Projekts "BeONE" am RIVM in Bilthoven, Niederlande: One-Health-Ansatz zur Ausbruchserkennung und Visualisierung mit epidemiologischen und genetischen Daten für Lebensmittelbedingte Erkrankungen.
- 22. bis 24. Januar 2020: Wir stellen Fortschritte der Topic Group Outbreaks beim Meeting H der Focus Group AI for Health in Brasilia vor (vor Ort und remote).
- 25. bis 29. Januar 2020: Wir sind bei den Applied Machine Learning Days an der EPFL in Lausanne, wo wir unser Framework "EventEpi" (Verarbeitung natürlicher Sprache für die ereignisbasierte Surveillance) vorstellen.
- 30. September bis 2. Oktober 2020: Wir stellen eine neue Metrik für Ausbruchserkennungsalgorithmen für die Topic Group Outbreaks beim Meeting der Focus Group AI for Health vor.
- 2. Oktober 2020: Wir stellen unsere Ergebnisse zum Syndrom Lernen auf der 15. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi) vor (remote).

Verwertung der Projektergebnisse (Nachhaltigkeit / Transferpotential)

Viele der Forschungsergebnisse sind als Manuskripte niedergeschrieben und der dahinter liegende Programmcode auf öffentlichen Plattformen geteilt und damit für die wissenschaftliche Community verfügbar gemacht. Außerdem wurden mehrere Produkte entwickelt die innerhalb des RKI aber auch teilweise extern genutzt werden können. Neben den gewonnenen Erkenntnissen und entwickelten Produkten sind es besonders die Netzwerke mit Forschungsgruppen sowohl nationaler als auch internationaler Institute in den Bereichen der Public Health, Statistik und Machine Learning die einen Mehrwert sowohl für das RKI als auch die Forschungsgemeinschaft erbringen.

Generell wurde die Kompetenz bzgl. Ausbruchserkennung, Datenanalyse und Datenvisualisierung des RKI gestärkt. Im Rahmen von DEMIS und der ÖGD-Kontaktstelle werden dauerhaft eine der Aufgaben und Produkte die innerhalb von Signale 2.0 entstanden sind fortgeführt. Weiterhin wurden neue Projekte angeworben, die gewonnene Erkenntnisse in vielfältigen Anwendungen umsetzen werden.

Die Verbesserungen im Bereich der Ausbruchserkennung führen hoffentlich zu verbesserter Entscheidungsfindung und Arbeitserleichterung bei den Fachexperten am RKI und den Landesstellen.

Die große Aufgabe wird nun darin bestehen die erstellten Tools und entwickelten Methoden auf ihre Praxistauglichkeit noch weiter zu evaluieren und noch mehr in die vorhandenen Surveillance Systeme zu integrieren. Dies soll ebenfalls eine leichte Integration und Überprüfbarkeit von neuen Methoden ermöglichen welche dann ggf. in externen Forschungsgruppen oder dem ZKI-PH entwickelt werden.