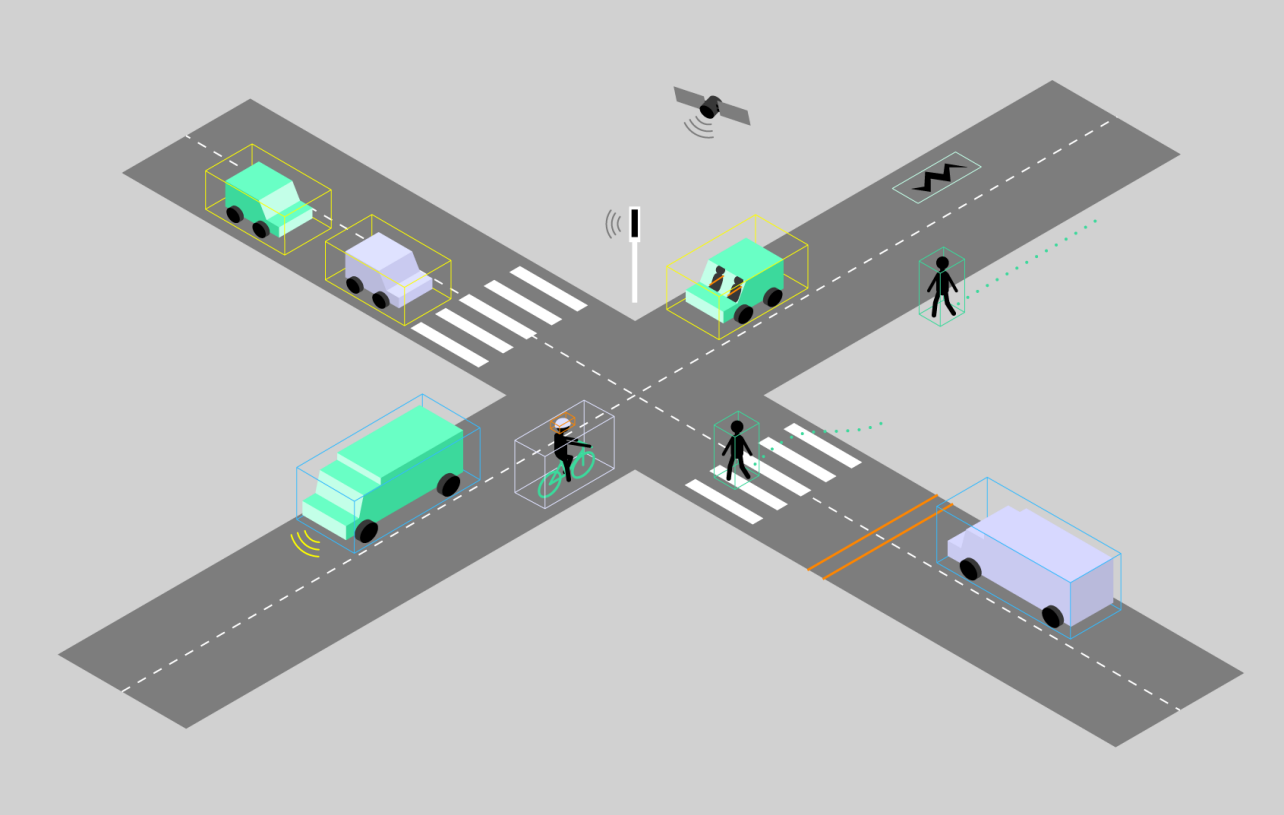
April 2021

Intelligenta beslutsstödsystem för utvecklad planering av transportsystemet

Sofia Löfstrand
Martin Modigh Karlsson



Illustrationer: Ella Frödin

TRAFIKVERKET
DRIVE SWEDEN
AI SWEDEN

Innehållsförteckning

Sammanfattning

Detta initiativ syftar till att bättre förstå hur ny teknik kan bidra till utvecklingen av Trafikverkets förmåga att planera ett effektivt, tillgängligt och hållbart transportsystem. Initiativet skall svara på om det finns förutsättningar för AI som nyttiggörare inom verksamheten, indikera var man kan få störst utväxling av AI för intelligenta beslutsstödsystem, samt vidare utveckling för att nå önskade effekter.

Projektet genomfördes under perioden juni 2020 – april 2021 under projektledning av Lindholmen Science Park, specifikt det strategiska innovationsprogrammet Drive Sweden i samarbete med AI Sweden och Trafikverket. Forskare knutna till *Data as Impact Lab* vid Högskolan i Borås har bidragit till analysen av vetenskapliga publikationer. Utöver dessa organisationer bidrog ett antal samarbetspartner genom intervjuer, deltagande i workshops, samtal och mejlväxlingar. Det materialet kompletterades med både nyhetssökningar/-bevakningar och nämnda sökningar av vetenskapliga publikationer för vidare sammanställning i en analys.

I projektet identifierades AI-applikationer för luftkvalitet och trafiksäkerhet som två för Trafikverket prioriterade områden, dels då de ses som grundläggande för ett mer hållbart transportsystem och dels då ett behov att utveckla nya arbetssätt identifierats. För att konkretisera frågeställningarna ytterligare lades fokus på vägtransportsystemet.

Både för luftkvalitetsanalyser och trafiksäkerhetsanalyser framträder behov av bättre kartdata i analyser. För luftkvalitet är det viktigt att förutom vägbanedata inkludera information om den omgivande miljön, till exempel vägsidesområden/-topografi och byggnationer i närhet av vägarna. För trafiksäkerhet finns behov att kartlägga till exempel mitträcken, sidoräcken, korsningar och korsningsutformning, räfflor i mitten och sidan av vägen, väglinjer, kurvatur av vägar och utformning av vägsidesområden.

Behoven av bättre data för luftkvalitet och trafiksäkerhet skulle troligtvis kunna lösas med AI-tekniker för bildanalys som idag är tillgängliga. Vägscanningen som görs inom verksamhetsområde Underhåll har potential att samordnas för flera syften. En utökad scanning för att täcka fler behov föreslås därför att utredas. För att gå vidare med ett sådant case är det till att börja med en samordning av behov från olika verksamheter som behöver göras innan en eventuell innovationsupphandling eller -utredning av vilka tekniska lösningar som är möjliga. Sannolikt finns det behov av mer detaljerad vägdata från andra delar av Trafikverkets organisation, synergier som inte undersökts i denna studie.

När det gäller trafikflödesdata finns även där överlappande behov för områdena luftkvalitet och trafiksäkerhet. För luftkvalitet behövs mer detaljerad information om hur trafiken varierar i tid (varje timme under ett dygn) när det gäller antal fordon, typ av fordon och hastigheter för dessa fordon. Detta för att kunna ta fram flödeskategorier. För trafiksäkerhet är information om vilka typer av fordon och om oskyddade trafikanter som rör sig längs och tvärs vägarna och vilka hastigheter de

har viktig. Man behöver även veta mer om avvikelser i hastighetsefterlevnad för att analysera problemen. För att täcka dessa behov behöver det utredas vilken teknik, eller kombination av tekniker som vägnära mätningar, mobildata etc., som kan täcka behoven. Vägnära kameror där trafikanter och trafiksituationer kan identifieras med bildanalys är en möjlig lösning som skulle kunna ersätta slangmätningar och slingor i marken. Tester för att utvärdera ny teknik för att göra trafikflödes- och hastighetsmätningar föreslås därför.

I fortsatt arbete är det av stor vikt att samordna insamling av data för olika syften och mellan olika verksamhetsområden på Trafikverket. Med dagens metoder görs skilda mätningar för olika syften, mätningar som med stor sannolikhet tekniskt sett kan kombineras i mycket högre grad. En stor risk är att dagens uppdelning är suboptimerad och ger en kostsam insamling, jämfört med samordnad insamling.

För att man inom Trafikverket ska kunna samordna sina insatser inom mer avancerad dataanalys och AI är det nödvändigt att ett tydligt ansvar och mandat för att samordna skapas på myndigheten. Den samordnande verksamheten bör samla in behov från myndighetens olika delar, koordinera insatser för att samla in data och bidra till analys.

För att på längre sikt arbeta med mer avancerad dataanalys och AI är det nödvändigt att dels använda data i flera olika syften, dels – och framför allt – att använda flera olika kompletterande datakällor för analys. För detta är samordning nödvändig.

Samordningen behöver göras internt men också med externa samarbetspartner både för att klarlägga behov och för att utveckla lösningar. För luftkvalitet bör arbetet göras i samarbete med främst kommuner, SMHI och Naturvårdsverket då de är viktiga parter. Vägscanning och trafikflödesmätningar skulle gynnas av samordning av metoder mellan Trafikverket och kommuner för att möjliggöra bättre analyser på ett nationellt plan. Precis som för frågan om intern samordning är den om ansvar i arbetet med att minska luftföroreningar från trafiken viktig för myndigheten att klargöra. Det är frågor som behöver lösas centralt på Trafikverket för att skapa förutsättningar för att utföra arbetet i organisationen.

Summary

This initiative aims to better understand how new technologies can contribute to the development of the Swedish Transport Administration's ability to plan an efficient, accessible and sustainable transport system. The initiative will answer how AI can be an enabler in that aim and where the use of AI for intelligent decision support systems, can be applied.

The project was carried out during the period June 2020 – April 2021 under project management by Lindholmen Science Park, specifically the strategic innovation program Drive Sweden in collaboration with AI Sweden and the Swedish Transport Administration. Researchers affiliated with *Data as impact lab* at the University of Borås have contributed to the analysis of scientific publications. In addition to these organizations, a number of partners contributed through interviews, participation in workshops, talks and email exchanges. This material was supplemented with both news searches and said searches of scientific publications for further compilation in an analysis.

In the project, AI applications for air quality and road safety were identified as two prioritized areas for the Swedish Transport Administration. They are seen as fundamental to a more sustainable transport system and there is a need to develop new ways of working to tackle these challenges. In order to further narrow down the topics, the focus was on put the road transport system.

Both for air quality analyses and road safety analyses there is a need for better map data. For air quality, it is important to include, in addition to roadway data, information about the surrounding environment, such as roadside areas/topography and constructions in the vicinity of the roads. For road safety, there is a need to map, for example, markings, grooves and railings in the middle and side of the road, intersections and intersection design, road curve and roadside area design.

The need for better data for air quality and road safety could probably be solved with AI imaging technologies that are currently available. Road scannings carried out today for maintenance purposes could also answer these needs. An extended scan to cover identified needs is therefore proposed to be investigated. In order to proceed with such a case, it is initially a coordination of needs from different ares of the organisation that is needed before innovation procurement or an investigation of technical solutions starts. There is probably a need for more detailed road data also from other areas within the Swedish Transport Administration, that have not been included in this study.

When it comes to traffic flows, there are also overlapping needs for the areas of air quality and road safety. For air quality, more detailed information is needed on how traffic varies in time (every hour) in terms of the number of vehicles, the type of vehicle and the speeds of those vehicles. This is to be able to develop different categories of traffic flow. For road safety, information on the types of vehicles and on vulnerable road users on the roads is important. The speeds and speed

compliance of the road users is also important to analyze the problems. In order to cover these needs, it is necessary to investigate which technology, or combination of technologies such as roadside measurements, mobile data, etc., can cover the needs. Roadside cameras where road users and traffic situations can be identified with image analysis is a possible solution that could replace pneumatic tubes and inductive loops. Tests to evaluate new technologies for making traffic flow and speed measurements are therefore proposed.

In continued work, it is of great importance to coordinate the collection of data for different purposes and between different areas of the Swedish Transport Administration. With today's methods, different measurements are made for different purposes, measurements that can most likely be technically combined to a much greater extent. A big risk is that today's division is sub-optimized and costly, compared to coordinated approach.

In order for the Swedish Transport Administration to coordinate its efforts in more advanced data analysis and AI, it is necessary that a clear responsibility and mandate to coordinate is created at the authority. The coordinating activities should collect needs from the different parts of the Authority, coordinate efforts to collect data and contribute to analysis.

In order to work with more advanced data analysis and AI in the longer term, it is necessary to use data for several different purposes and, above all, to use several different complementary data sources for analysis. Coordination is needed to achieve this.

Coordination needs to be done internally, but also with external partners both to clarify needs and to develop solutions. For air quality, the work should be done in cooperation mainly with municipalities, SMHI and the Swedish Environmental Protection Agency as they are important parties. Road scanning and traffic flow measurements would benefit from coordination of methods between the Swedish Transport Administration and municipalities to enable better analysis at a national level. As with the issue of internal coordination, the responsibility for reducing air pollution from traffic is important for the Authority to clarify. These are issues that need to be solved centrally at the Swedish Transport Administration in order to create the conditions for carrying out the work in the organization.

Bakgrund

Verksamhetsområde Planering har behov att analysera hur verksamheten kan utvinna värde ur AI som stöd i intelligenta beslutsstödsystem. Projektet syftar till att utveckla en ökad förståelse för hur organisationen kan dra nytta av AI för att utveckla den långsiktiga planeringen av och hållbarheten i transportsystemet. Rätt tillämpat och utvecklat bedöms AI i förlängningen kunna bli ett viktigt stöd och verktyg i arbetet med åtgärdsplaneringen och den nationella planen.

Initiativet skall svara på om det finns förutsättningar för nyttiggörande inom verksamheten, indikera var man kan få störst utväxling av AI för intelligenta beslutsstödsystem samt vad man kan behöva utveckla vidare för att nå önskade effekter.

Verksamheter i flera olika sektorer har satt upp kostsam IT-infrastruktur för att kunna jobba med exempelvis AI och andra typer av beslutsstödsystem i stor skala, men har haft svårt att hitta tillämpningar och arbetssätt som ger avkastning på investeringen. Flera andra verksamheter är också osäkra på hur de kan komma igång och jobba med AI samt vilket värde som finns i det. Därtill ställer sig flera frågor kring hur exempelvis AI påverkar organisationen och framtida kompetens-försörjning.

Inom Trafikverket pågår sedan en tid flera initiativ med kopplingar till Artificiell Intelligens (AI), Machine Learning (ML) eller Big Data Analytics, med syftet att utforska möjligheterna med de nya digitala analysstöd som växer fram på marknaden. Bland annat har ett initiativ till förbättrad analysförmåga initierats. TRV Labs är en annan satsning där Trafikverket tillsammans med Linköpings Universitet vill skapa förutsättningar för olika samverkansformer med flera aktörer. Av de projekt som hittills har genomförts är huvuddelen riktat mot järnväg men allt fler idéer och projekt börjar nu även riktas mot vägområdet.

Syfte och mål

Projektet syftar till att utifrån ett utmaningsdrivet perspektiv och i linje med målbild 2030 ta fram ett kunskapsunderlag som bidrar till en ökad förståelse för hur Trafikverket kan dra nytta av AI för att utveckla den långsiktiga planeringen av transportsystemet.

Målet är att identifiera relevanta och/eller mogna AI-tillämpningar och att ta fram ett ”State of the art”-kunskapsunderlag som kan användas inför en kommande innovationsupphandling. Resultaten i denna förstudie kan även nyttjas vägledande för framtida forskningsinriktade AI-initiativ.

Projekttid, parter och omfattning

Projektet genomfördes under perioden juni 2020 – april 2021 under projektledning av Lindholmen Science Park, specifikt det strategiska innovationsprogrammet Drive Sweden i samarbete med AI Sweden och Trafikverket. Forskare knutna till *Data as Impact Lab* vid Högskolan i Borås har bidragit till analysen av vetenskapliga publikationer. Utöver dessa organisationer bidrog ett antal samarbetspartner genom intervjuer, deltagande i workshops, samtal och mejlväxlingar. Det materialet kompletterades med både nyhetssökningar/-bevakningar och nämnda sökningar av vetenskapliga publikationer för vidare sammanställning i en analys. En mer utförlig beskrivning av metod och genomförande presenteras i nästa kapitel.

I projektet identifierades AI-applikationer för luftkvalitet och trafiksäkerhet som två prioriterade områden inom Trafikverket, då de är grundläggande att förbättra för ett mer hållbart transportsystem och ett behov finns att utveckla nya arbetssätt. För att konkretisera frågeställningarna ytterligare lades fokus på vägtransportsystemet. Därmed avgränsades arbetet till att utreda och sammanställa en ”State Of The Art” för AI tillämpningar för luftkvalitet och trafiksäkerhet inom vägtransportsystemet.

I rapporten Digitaliseringens bidrag till målbild 2030 (Magnusson, U., & Eriksson, P. (2020)) beskrivs målen för luftkvalitet och trafiksäkerhet med koppling till vägtrafiknätverket enligt följande.

> Luftkvalitet mål 2030:
>
> - Utsläppen från transportsektorn har minskat så att miljökvalitetsmålet Frisk luft för NO2 i urban bakgrund och PM10 i gaturum uppnås.
>
> Trafiksäkerhet mål 2030:
>
> - Minst 50 % färre dödas och minst 25 % färre skadas allvarligt i vägtransportsystemet jämfört med 2020.

Rapporten beskriver hur digitalisering kan bidra till dessa mål och för trafiksäkerhet exemplifierar nedan citat en av flera vägar framåt.

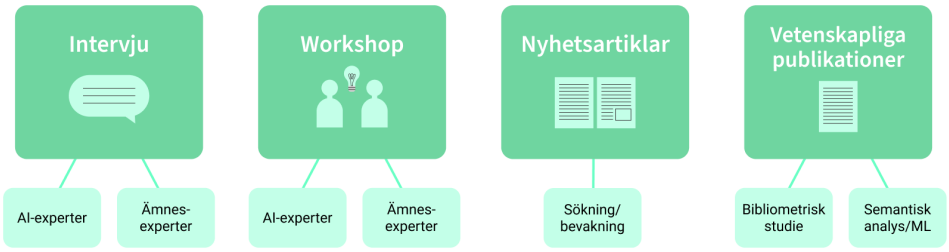
”Trafiksäkerheten kan ökas genom en trafikledning som har automatiserade beslutsprocesser i realtid med AI-stöd som baseras på data från och om transportsystemet. Automatiserad trafikledning kan leda till färre felbeslut än en helt manuell trafikledning.”

För luftkvalitet beskrivs bland annat följande.

”Luftkvaliteten i tätort och gaturum kan förbättras genom att data om luftkvalitet erhålls från olika mätsystem i realtid och nyttjas för att styra fordonens drivlina, transportvägar och tillträde till områden, så kallad geofencing. Ett första steg kan vara att bygga system som styr detta utifrån mer statisk information, till exempel fordons egenskaper och skyddsvärda områden. I ett nästa skede skulle geofencing kunna användas dynamiskt för att styra hastighet, drivlina och trafikflöde beroende på aktuella luftkvalitetsvärden.”

Rapporten beskriver en riktning för möjliga framtida lösningar som digitalisering, avancerad dataanalys och AI skulle kunna bidra till. Denna riktning och målen för 2030 har varit en ingång i arbetet under detta projekt.

Metod och genomförande

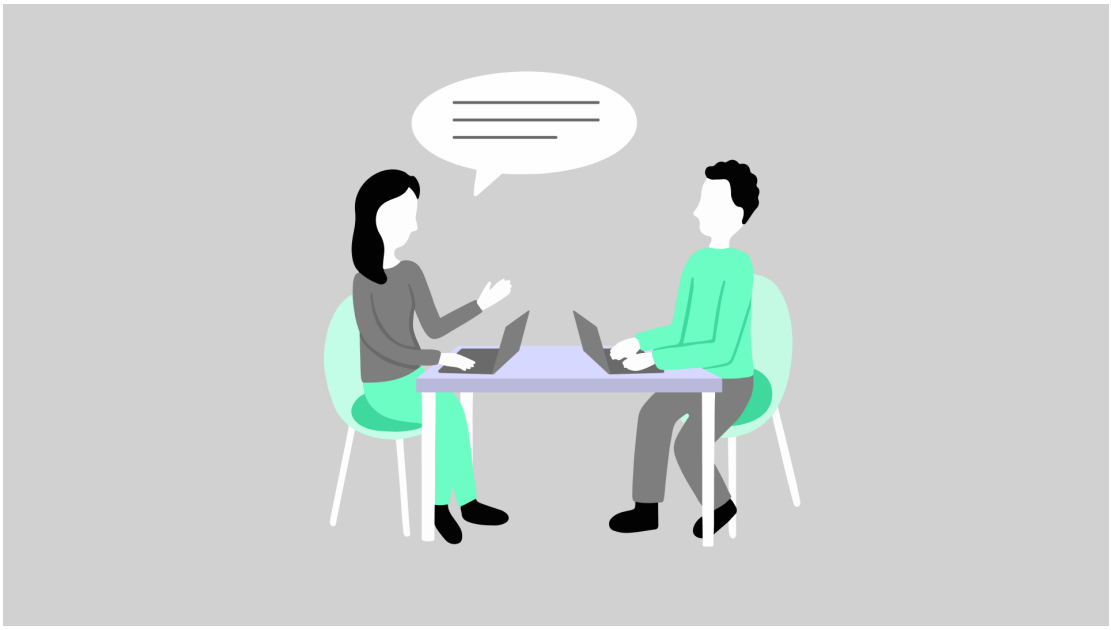


I genomförandet var det första steget att klargöra specifika utmaningar som finns i arbetet med att minska emissioner från och öka trafiksäkerheten i vägtransport-nätet. I ett andra steg har sedan lösningar på dessa behov analyserats där mer avancerad datainsamling, analys och AI-teknik kan bidra. Metoden för att samla in och analysera information har varit densamma för både luftkvalitet och trafik-säkerhet och arbetet med de båda har genomförts parallellt.

De datakällor som använts som underlag för analys består av:

- Intervjuer med ämnesexperter inom luftkvalitet, trafiksäkerhet och AI
- Workshopar inom luftkvalitet & AI respektive trafiksäkerhet & AI
- Sökning av nyhetsartiklar
- Sökning av vetenskapliga publikationer

Intervjuer



För luftkvalitetsområdet intervjuades totalt 16 personer från nedan listade organisationer. Dessa personer har erfarenhet och kunskap inom forskning, samt utvärdering, uppföljning och utveckling av åtgärder för att minska emissioner från vägtransportsystemet.

- Cowi
- Göteborgs universitet
- IVL Svenska Miljöinstitutet
- KTH
- Lunds universitet
- Naturvårdsverket
- NILU (Norsk institutt for luftforskning)
- Stockholms universitet
- Stockholms stad
- SMHI
- Trafikverket
- Umeå universitet
- Uppsala universitet
- VTI

För trafiksäkerhetsområdet intervjuades 13 personer totalt, varav 12 personer från Trafikverket och en representant för VTI.

Intervjupersonerna skiljer sig mellan ämnena luftkvalitet och trafiksäkerhet då majoriteten är aktörer utanför Trafikverket för luftkvalitet och för trafiksäkerhet är

majoriteten anställda på Trafikverket. En förklaring till detta är att Trafikverket historiskt sedan mitten av 90-talet och framåt har arbetat systematiskt med Nollvisionen (www.trafikverket.se/resa-och-trafik/trafiksakerhet/det-har-ar-nollvisionen/) och genomfört ett målinriktat arbete för att minska antal döda och skadade i trafiken. Det finns således många personer på Trafikverket som är involverade i att samla in och analysera data, samt vidta åtgärder för att förbättra trafiksäkerheten. Fokus under intervjuer inom trafiksäkerhetsområdet lades därför på att intervjua anställda på Trafikverket.

För AI-området har 10 personer från de nedan listade organisationer intervjuats. De har kompetens inom datainsamling, analys och forskning inom AI.

- Dynniq
- Halmstad Högskola
- KTH
- Microsoft
- MIT
- Peltarion
- RISE
- Univrses
- Viscando

Workshops

Två workshops hölls under projektet för att diskutera ett par prioriterade frågor som valts ut baserat på intervjuresultat.

För luftkvalitet var frågan:

”Hur kan vi, med stöd av AI, bättre förstå och utvärdera hur människor utsätts för utsläpp från vägtransporter?”

I denna workshop presenterades till att börja med hur exponeringsstudier görs idag av SMHI som ansvarar för verktyget SIMAIR och vilka datakällor som används för analysen. SIMAIR är ett webbaserat system som hjälper kommuner, luftvårdsförbund och andra regionala aktörer att enkelt utvärdera halterna av luftföroreningar. SIMAIR drivs av SMHI i samverkan med Trafikverket. Steg två var att diskutera hur datakällor och analyser behöver utvecklas för att ge bättre beslutsstöd för åtgärder. I diskussionen inkluderades både analys av källor, dvs. fordonens olika bidrag, till emissioner i vägtransportsystemet, och hur människor exponeras för dessa emissioner.

I workshopen om luftkvalitet deltog 12 representanter från SMHI, Peltarion, Göteborgs Universitet, COWI, Stockholms universitet, Stockholms Stad, RISE, Halmstad Högskola, Viscando, IVL Svenska Miljöinstitutet, Trafikverket och Naturvårdsverket. Moderatorer för workshopen var representanter från Drive Sweden och AI Sweden.

För den andra workshopen med fokus på trafiksäkerhet, var ämnet hastighetsefterlevnad och oskyddade trafikanter. Frågorna som prioriterades baserat på intervjuresultat var:

”Vad vore idealt att veta om fordonshastigheter och oskyddade trafikanter?”

”Vilka beslut ska informationen stödja?”

”Vad vore idealt att veta för att kunna prioritera trafiksäkerhetsåtgärder på vägnätet?”

”Vilka likheter och skillnader finns i databehoven från olika verksamhetsområden på Trafikverket?”

I workshopen deltog 7 representanter från Trafikverkets verksamhetsområden Planering och Underhåll. Moderatorer för workshopen var representanter från Drive Sweden och AI Sweden.

Nyhetssökningar

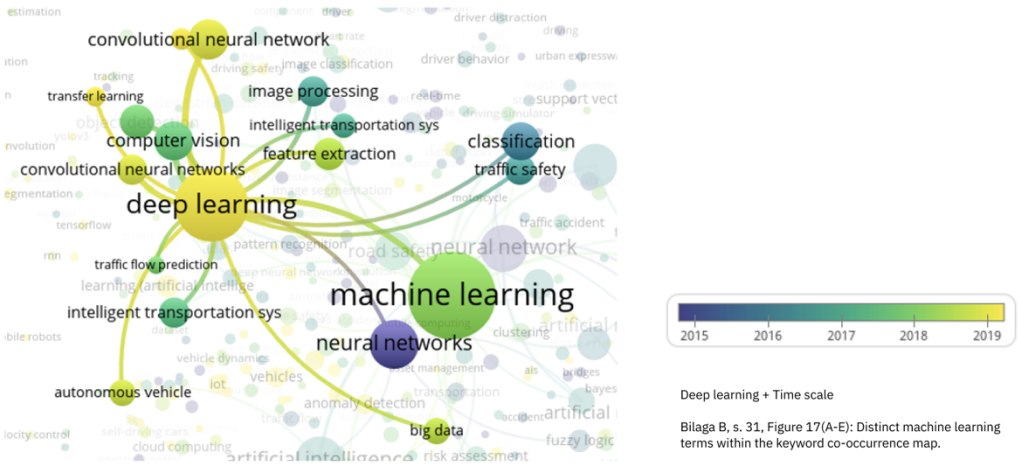
Nyhetssökningar gjordes med hjälp av tjänsten Retriever (retrievergroup.com). Sökningarna gjordes i följande medier:

- Svensk tryckt press (ex lokaltidningar, magasin, fackpress)
- Svensk webb, både webbsidor med redaktionellt innehåll samt webbsidor tillhörande myndigheter, organisationer, föreningar och liknande.
- Engelska, tyska, spanska & franska webbkällor. Indelning utifrån språk. Webbsidor med redaktionellt innehåll av olika karaktär, t ex stora internationella nyhetskällor, nyhetsbyråer, mindre lokala nyhetskällor, fackpress.

Sökningen sattes upp med sökord på engelska och svenska både för både områdena *luftkvalitet och AI* samt *trafiksäkerhet och AI*. Sökningen gjordes för material som publicerats från 1/9 2020 till 28/2 2021. Sökningen gjordes genom att kombinera ord som beskriver ämnesområdet, som *”Trafiksäkerhet”* och *”Olycksförebyggande”* i kombination med AI-termer som *”Maskininlärning”* och *”Djupinlärning”.*

För luftkvalitet och AI valdes 27 artiklar ut av totalt 2783 i sökningen. För trafiksäkerhet valdes 32 artiklar ut av totalt 1143 i sökningen. Några av de utvalda artiklarna refereras i kommande delar av denna rapport och alla 59 artiklar listas i Bilaga A.

Sökning av vetenskapliga publikationer – semantisk analys och scientometri

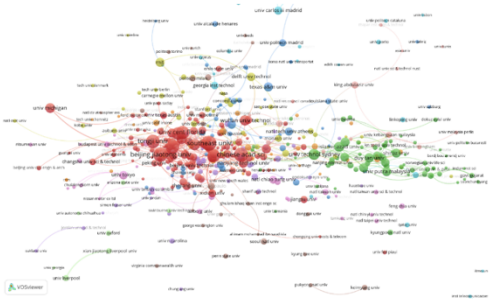


Bilaga B, s. 31, Figure 17(A-E): Distinct machine learning terms within the keyword co-occurrence map.

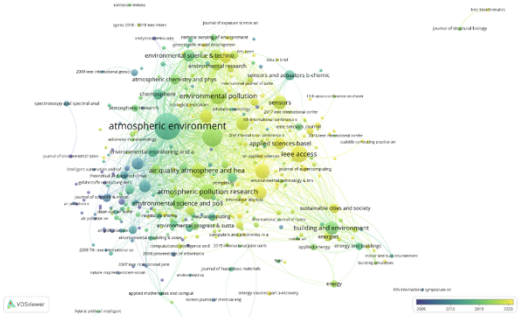
För att genomföra sökningar av vetenskapliga publikationer togs hjälp av Högskolan i Borås *Data as Impact Lab*. Deras uppdrag bestod i att studera aktuell forskning som tillämpar artificiell intelligens och maskininlärning inom två forsknings-områden: luftkvalitetsfrågor och trafiksäkerhet.

I ett första steg gjordes en scientometrisk[1] analys genom att söka efter relevant litteratur med specifika termer som täcker de aktuella ämnena och sedan använda olika bibliometriska metoder för att undersöka innehållet och metadata för den litteratur som hittades. Clarivate Web of Science valdes för uppgiften, vilken är en av de mest omfattande publikationsdatabaserna med omfattande kvalitetskontroll. Den innehåller också bibliografiska referenser till annan vetenskaplig forskning, vilket är en förutsättning för att göra citationsanalys.

Söktermerna utgjordes av ämnesord som ”air quality”, ”air pollution” och ”particle emission” för studier av luftkvalitetsforskning. För studier av trafiksäkerhet användes termer som ”traffic safety”, ”safety assessment” och ”traffic accidents”. Dessa termer kombinerades med metodologiska termer som används för att beskriva nuvarande AI-tillvägagångssätt inom forskning som ”artificial intelligence”, ”machine learning”, ”deep learning” och mer specifika termer som ”random forest” och ”support vector machines”.



Bilaga B, s. 24, Figure 12 Co-authorship – Organisations. Of 3,431 organisations, 643 (564) were found ≥3 times (clusters)



Bilaga B, s. 15, Figure 4: Bibliographic coupling – Sources. Of 1,409 sources, 334 (330) had ≥2 documents. Nodes clustered based on average publication year.

Sammantaget identifierades 3 166 artiklar inom luftkvalitetsforskning och 4 171 inom trafiksäkerhetsforskning. Ett öppet tidsfönster användes och i båda datamängderna finns de tidigaste identifierade publikationerna i tidsintervallet 1991–2021, men med en approximativt exponentiell ökning fram till dagens datum.

I ett andra steg gjordes semantiska analyser på de identifierade artiklarna i tidigare sökning. För att möjliggöra identifikation av specifika analysmetoder och analystekniker i vetenskapliga abstracts utvecklades en strategi för att extrahera fraser som potentiellt utgör svar på specifika frågor. Detta tillvägagångssätt, känt som *question answering* (QA), använder djupinlärning och språkmodellering för att

[1] Termerna scientometri och bibliometri används utbytbart i denna beskrivning, men med viss distinktion då bibliometrisk metod här behandlas som tillämpning av scientometrisk forskning.

identifiera avsnitt i texterna som motsvarar en maximal sannolikhet att utgöra svaren på givna frågor. Språkmodellen som har använts för denna QA-strategi är SciBERT (se Beltagy, Lo, & Cohan, 2019), som är tränad på vetenskapliga artiklar från webbplatsen Semantic Scholar.

För att förbättra möjligheterna att utforska de datamängder som erhållits från Web of Science utvecklades en probabilistisk sökmotor inom ramen för projektet. Denna sökmotor är en implementering av den probabilistiska best match-modellen BM25 (se Robertson & Zaragoza, 2009) och är som sådan baserad på sannolikhets-rankningsprincipen för dokument. Denna princip anger att dokument ska rankas som svar på användaren sökfråga utifrån sannolikheten för att dokumenten är relevanta för användaren (Robertson, 1977). BM25 tillämpar en binär egenskap som kallas ”eliteness” och som anger huruvida en specifik term beskriver dokumentets innehåll (Robertson & Zaragoza, 2009). Denna egenskap anses i sin tur vara statistiskt relaterad till de lokala frekvenserna för termerna, modellerade i form av diskreta Poisson-fördelningar. För att erhålla ett komplett viktningsschema för term-dokument-relationen normaliseras de lokala frekvenserna utifrån dokumentens längd och multipliceras med en vikt som konceptuellt liknar viktningsschemat inverse document frequency, introducerat av Spärck Jones (1972). I detta projekt användes modellen BM25 för att rangordna bibliografiska poster från Web of Science baserat på innehållet i motsvarande abstracts. Förutom att rangordna posterna med hjälp av BM25-modellen erbjuder sökmotorn också möjligheten att rangordna framsökta poster enligt citeringsfrekvens respektive citeringsfrekvens per år. I tillägg till den grundläggande bibliometriska informationen visas i anslutning till varje post en ikon (kallad ”munk”) som representerar Altmetrics Attention Score. Denna ikon innehåller information från Altmetric rörande i vilken utsträckning dokumentet har nämnts i källor som nyhets-artiklar, sociala medier, Wikipedia och patent. Resultaten av både den scientro-metriska och den semantiska analysen presenteras i Bilaga B.

Med hjälp av den sökmotor (BM25) som skapades av Högskolan i Borås genomfördes i ett tredje steg sökningar av relevanta vetenskapliga publikationer i underlaget från den bibliometriska studien. Av de 3166 publikationer för området luftkvalitet och AI och de 4171 publikationer och för området trafiksäkerhet & AI valdes 41 publikationer av specifikt intresse ut. Några av publikationerna refereras i kommande delar av denna rapport och presenteras listade i Bilaga C.

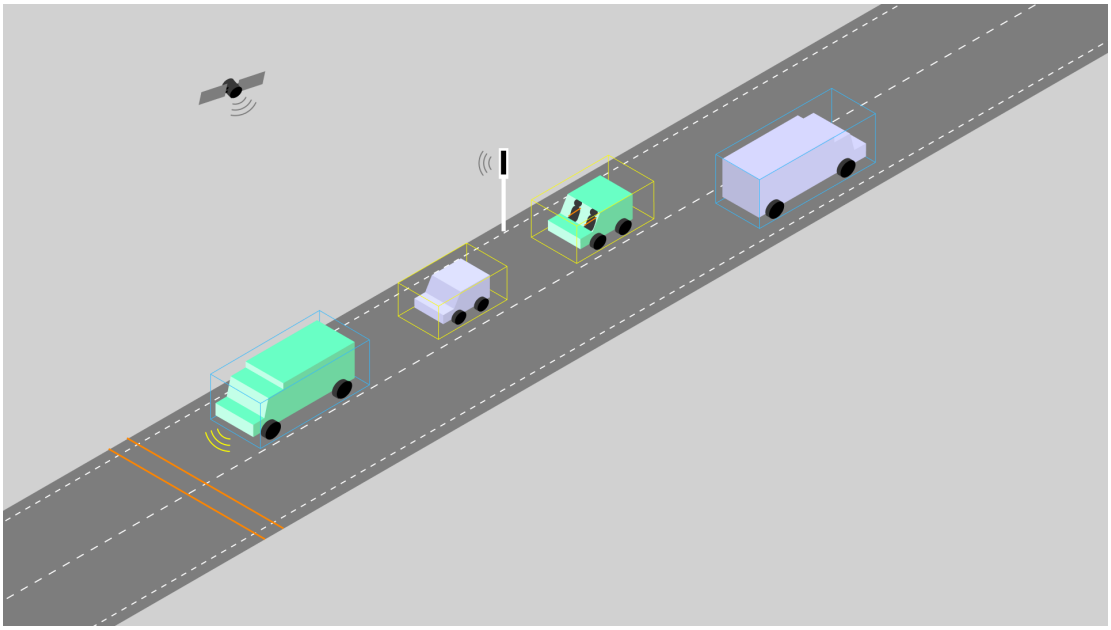
Resultat luftkvalitet och AI

I en studie från 2018 av Gustafsson et al. beräknades att 7 600 personer dör i förtid i Sverige varje år på grund av luftföroreningar. Trafikens luftföroreningar uppskattas på ett år totalt orsaka omkring 3 000 av dessa dödsfall. Från vägtrafiken är det till stor del kväveoxider och partiklar som bidrar till emissionerna. Dessa härstammar både från avgaser från fordonen och slitagepartiklar från däck, väg och bromsar.

Studien beskriver också att exponeringen av emissioner ökar med en ökande urbanisering, då fler människor utsätts för skadliga nivåer av emissioner i städer.

Idag ligger ansvaret på kommunerna att kontrollera och genomföra åtgärder för att möta luftkvalitetsdirektiv från EU (ec.europa.eu/environment/air/quality/) och de svenska miljökvalitetsnormerna för luft (www.naturvardsverket.se/mknluft). Mätningar av luftkvalitet görs idag i städer och andra områden, men för många områden saknas information för att få en komplett bild och för att kunna prioritera insatser på ett effektivt sätt. Det behövs mer detaljerade data för att förstå hur olika källor bidrar till emissioner och en försämrad luftkvalitet, för att kunna rikta åtgärderna där de har störst effekt.

Luftkvalitet och trafikflöden



Under intervjuer och workshop framkom att en bättre förståelse av trafikflöden är idag det viktigaste området att utveckla för att möjliggöra bättre analyser av vägtrafikens bidrag till emissioner och för att kunna prioritera åtgärder för att minska dessa. Det är grunden till att minska den exponering av emissioner som människor utsätts för från vägtrafiknätverket.

Forskning visar att trafikflödesmätningar är en tillförlitlig metod att utvärdera trafikkällornas bidrag till lokala luftföroreningar. Beckx et al. (2009) beskriver en aktivitetsbaserad modell för att förutsäga trafikkällornas bidrag till lokala luftföroreningar med tillräcklig noggrannhet och bekräftar användbarheten av aktivitetsbaserade transportmodeller för luftkvalitetsändamål. Författarna skriver att ALBATROSS – AURORA-modellen som utvecklats ger tillförlitliga koncentrationer av föroreningar per timme för hela Nederländerna istället för att bara använda dagliga medelvärden nära trafikstationer. Detta är en möjlighet för

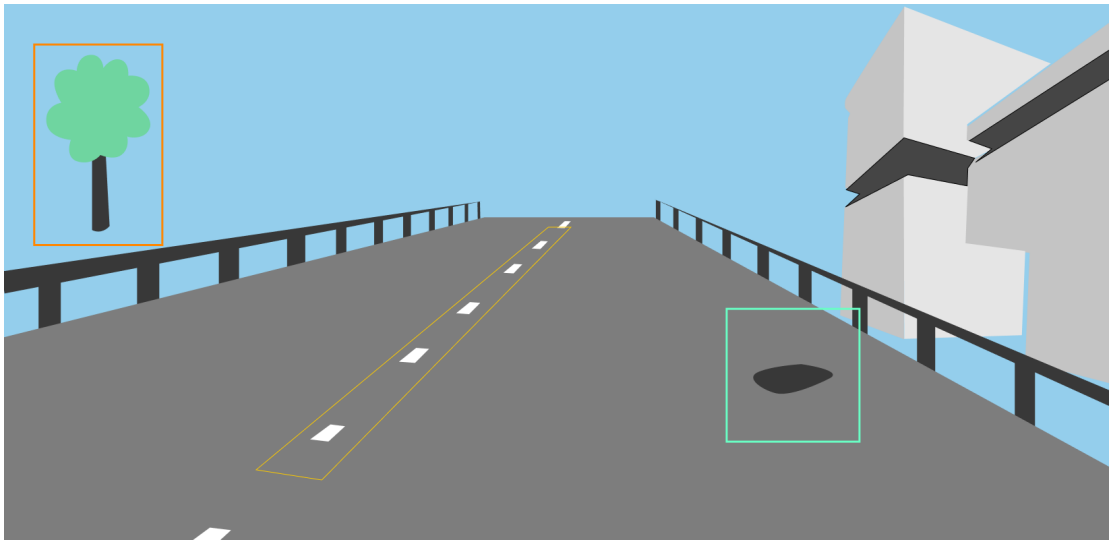
framtida exponeringsstudier som syftar till mer realistiska exponeringsanalyser och hälsoeffekter.

Idag används huvudsakligen måttet årsdygnstrafik (ÅDT) för att beskriva trafikflödet. ÅDT är det under ett år genomsnittliga trafikflödet per dygn mätt som fordon/dygn ,axelpar/dygn eller gående/cyklister/dygn. Medeltrafikflödet per dygn för ett visst år. För att beskriva hastighet används bland annat medelhastighet och andel fordon inom skyltad hastighet, skyltad hastighet + 5 km/tim. I arbetet med att beskriva fordonens bidrag på luftkvaliteten finns behovet att mer detaljerat beskriva variationer i trafikflödet eftersom dessa har stor påverkan på nivån av emissioner.

SIMAIR använder emissionsfaktorer från HBEFA (Handbook Emission Factors for Road Transport, hbefa.net) som för varje vägtyp och fordonstyp/fordonsflotta har 5 emissionsfaktorer för olika situationer från fritt flöde till kösituationer med krypfart. Idag finns ingen lösning för att välja flödeskategori systematiskt. För att komma framåt i arbetet behövs därför mer detaljerad information om hur trafiken varierar i tid (varje timme under ett dygn) när det gäller antal fordon, typ av fordon och hastigheter för dessa fordon.

Nästa nivå av information som är önskvärd är om trafiken beror på lokala, regionala, eller nationella resor. Detta skulle ge viktigt underlag för att förstå utmaningarna och vilka möjliga lösningar som behöver utvecklas för att minska emissionerna. Behovet är särskilt tydligt för kommuner som använder SIMAIR i sitt analysarbete. En kommuns lokala luftkvalitetsutmaningar varierar kraftigt beroende på hur trafiknätverket ser ut i kommunen och om trafiken beror på lokala, regionala resor eller nationella resor.

Kartdata för luftkvalitetssimuleringar



För simuleringar av luftföroreningar i SIMAIR bidrar Trafikverket med kartdata för det nationella nätverket och kommunerna för det kommunala nätverket. Kartdatan

för det kommunala vägnätverket är idag bristfällig. Det vore önskvärt att kartdata för det nationella vägnätverket och det kommunala nätverket presenterades på samma sätt och uppdaterades med samma frekvens för att öka tillförlitligheten i simuleringarna. Den skulle behöva uppdateras regelbundet och förutom vägkartdata inkludera information om den omgivande miljön, t.ex. vägsides-områden och byggnationer i närhet av vägarna då denna utformning påverkar spridningen av emissioner mycket.

Epidemiologiska studier

I frågan om hur människor utsätts för luftföroreningar så grundar sig idag epidemiologiska studier på samband mellan bostad (dygnsvila) och luftföroreningar. Människors verkliga exponering beror däremot till stor del på exponering i miljöer utanför hemmet och för att se effekten av åtgärder på befolkningens exponering behöver man därför kombinera mobilitet av människor med luftmiljö i både rum och tid. Denna andra aspekt behöver beaktas vid arbete med att minska exponeringen och bedömdes som viktig under workshopen.

Det finns exempel på studier som görs för att ta hänsyn till var människor befinner sig, förutom i bostaden. Ett pågående projekt på Umeå Universitet med finansiering från Trafikverket använder mobiltelefondata för att ta hänsyn till människors rörelsemönster och det är därmed möjligt att identifiera ”hotspots” där många människor befinner sig och kan utsättas. Mobildata har i flera andra exempel visat sig vara användbart för att identifiera rörelsemönster, till exempel i projekt inom Senseable Stockholm Lab (www.senseablestockholm.org). Metoder där rörelsemönster för befolkningen beaktas baserat på mobiltelefondata är ett intressant utvecklingsområde för att förbättra de epidemiologiska studierna.

Resultat trafiksäkerhet och AI

Trafikverket har under många år arbetat med att nå nollvisionen och minska antal döda och allvarligt skadade i trafiken. I uppföljningen av trafiksäkerheten i Sverige finns ett antal indikatorer som uppdateras ibland. Rapporten Analys av trafiksäkerhetsutvecklingen 2019, Trafikverket (2020) beskriver dessa indikatorer och om de är i linje med de mål som satts upp till 2020. I projektet har indikatorerna och uppföljningen av dessa varit en utgångspunkt under intervjuer.

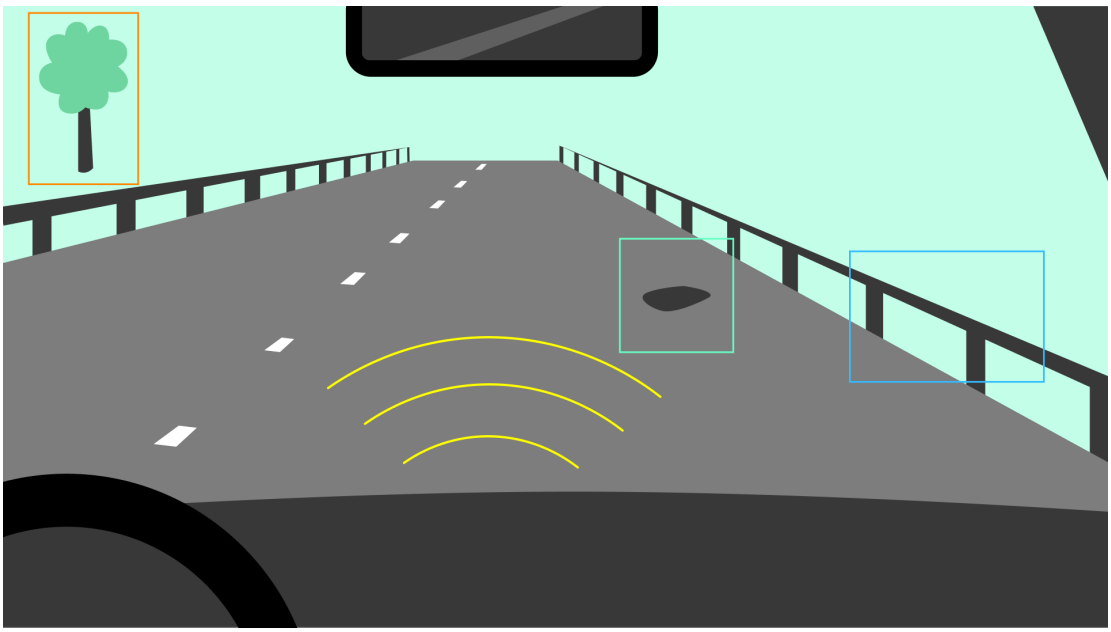
Indikatorerna är under uppdatering men följande användes för 2019:

1. Andel trafikarbete inom hastighetsgräns, statligt vägnät / kommunalt vägnät
2. Genomsnittlig resehastighet, statligt vägnät / kommunalt vägnät
3. Andel trafikarbete med nyktra förare
4. Andel bältade i framsätet i personbil

5. Andel cyklister med hjälm
6. Andel mopedister med rätt använd hjälm
7. Andel trafikarbete med högsta Euro NCAP-klass
8. Rätt användning av motorcykel (mäts inte)
9. Andel trafikarbete med mötesseparering på vägar över 80 km/tim, statligt vägnät
10. Andel säkra gång-, cykel- och mopedpassager
11. Andel kommuner med god kvalitet på underhåll av GC-vägar
12. Systematiskt trafiksäkerhetsarbete i linje med ISO 39001 (mäts inte)
13. Antal omkomna i trafiken
14. Antal allvarligt skadade i trafiken

Under intervjuer framkom att hastighetsefterlevnad och oskyddade trafikanter (cyklister, gående) är områden som för närvarande är prioriterade. För en bättre förståelse av trafiksäkerheten, för att göra analyser och för att prioritera åtgärder så finns ett behov av grundläggande data om väginfrastrukturen och mer detaljerade data om hur fordonen använder trafiknätverket.

Kartdata – utformning av vägar och vägsidesområden



Idag finns information om vägnätverket baserat på att Trafikverkets verksamhetsområde Underhåll samlar in data om vägarnas tillstånd som underlag för att prioritera underhåll. Denna data används idag även för andra syften, till exempel för att analysera vägarna ur ett trafiksäkerhetsperspektiv. Den data som samlas in idag är dock inte optimal i trafiksäkerhetssyfte då det behövs annan information än den som är viktig i underhållssyfte. Information som det finns behov att kartlägga är till exempel mitträcken, sidoräcken, korsningar och korsningsutformning, räfflor i

mitten och sidan av vägen, väglinjer, kurvatur av vägar och utformning av vägsidesområden. Detta för att mer effektivt kunna prioritera infrastrukturåtgärder från ett trafiksäkerhetsperspektiv.

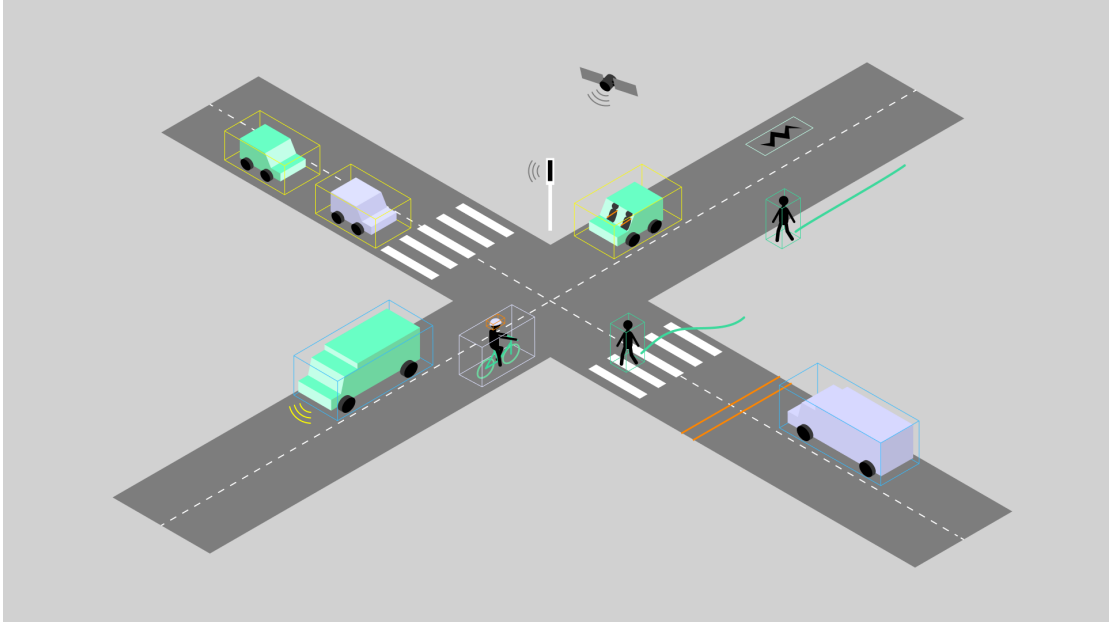
Behoven som finns från Trafikverkets verksamhetsområden Underhåll (i underhållssyfte) och från Planering (i trafiksäkerhetssyfte) skulle idag med AI-teknik, bland annat genom bildanalys under kameramätningar, högst troligt kunna samlas in med en och samma lösning. Det kritiska för att ta steg framåt är härmed inte de tekniska lösningarna, utan en samordning av behoven för datainsamling mellan Trafikverkets olika verksamhetsområden.

Trafiksäkerhet och trafikflöden

Den statiska informationen om vägarna som beskrevs i det tidigare kapitlet är viktig, men för att få en mer komplett bild av säkerheten på vägarna är det nödvändigt att veta mer om hur de används. Den statiska väginformationen behöver kombineras med dynamiska data om hur vägarna används. Dagens mätetal som till exempel årsdygnstrafik (ÅDT) och medelhastighet ger information om tillståndet i trafiken men anses för trubbiga för att användas i många analyser för att effektivt prioritera åtgärder. Det finns behov av att lättare kunna kombinera data från olika källor, till exempel statisk information om vägarnas utformning och dynamisk information om hur vägarna används, för att till exempel svara på frågor som *”Var finns kombinationen dåligt vägsidområde och hög hastighet?”* och därefter fokusera insatserna.

Dynamisk information som är viktig är vilka typer av fordon och oskyddade trafikanter rör sig längs och tvärs vägarna och vilka hastigheter de har. Dessa behov liknar de ovan nämnda behovet av att kategorisera trafikflödet på en mer detaljerad nivå i luftkvalitetssyfte. Både för luftkvalitetsanalyser och trafiksäkerhetsanalyser finns behov av att ha mer detaljerad information om hur trafiken varierar i tid när det gäller antal fordon, typ av fordon och hastigheter för dessa fordon. För trafiksäkerhet behöver man veta mer om avvikelser i hastighetsefterlevnad för att analysera problemen.

Med dagens tekniska lösningar finns stor potential att genomföra datainsamling flera syften samtidigt. Trafiksäkerhetsindikatorerna mäts idag med flera olika metoder. För bältesanvändning och hjälmanvändning utför personer observationer av passerande fordon och trafikanter, för trafikflödesmätningar och hastighets-mätningar görs mätningar i fasta stationer och genom slangmätningar, både i Trafikverkets regioner och nationellt. Hastighetmätningar görs även med fasta stationer i samband med trafikstyrning, till exempel på E4 utanför Stockholm (www.trafikverket.se/nara-dig/Stockholm/vi-bygger-och-forbattrar/e4e20-variabla-hastighetsgranser/). Förutom det finns även hastighetskameror som används för övervakning. Det är en stor variation av tekniker som används för sina specifika syften och det finns potential i att samordna så att fler mätningar kan genomföras med en teknisk lösning.



För att göra datainsamling och analys både mer kostnadseffektiv och ge bättre underlag till analys är det viktigt att utvärdera ny teknik. Med dagens sensorer, kameror och möjligheter till bildanalys, finns potential att samla in data till flera syften samtidigt.

Sådana lösningar utvärderas i flera projekt varav ett är inom Vivacity Labs i Manchester (highways.today/2020/11/09/vivacity-labs-ai-junctions-manchester/) där sensorer identifierar olika typer av trafikanter vid utvalda korsningar och trafiksignaler. Denna information används för att prioritera olika transportslag, alltså en typ av trafikstyrning, men information skulle även kunna användas för att till exempel analysera trafiksäkerheten i en viss korsning.

Ett par andra exempel på trafikmonitorering är Waycare Tech (www.itsinternational.com/its8/news/waycare-helps-manage-ohio-traffic) som utvecklat en plattform för att samla in realtidsinformation om trafiksituationen i Ohio för att kunna ta åtgärder för att förbättra trafiksäkerheten och minska restider.

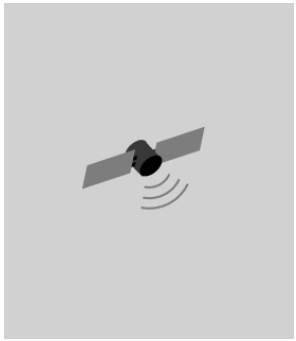
I England har Kent County Council och Amey installerat trafiksensorer (kccmediahub.net/kent-county-council-and-amey-install-traffic-insight-sensors745) som kan klassificera olika typer av trafikanter och deras hastigheter för att kunna styra trafiken effektivt. Denna information har potential att möjliggöra bättre trafikstyrning i realtid men även användas för planering på längre horisont.

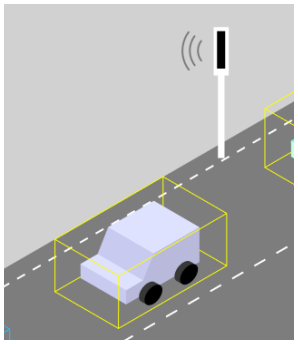
KCCmediahub.net skriver:

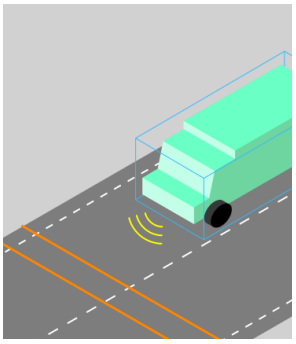
> *"A trial using state of the art traffic sensors is being carried out by Kent County Council and its partners at Amey to monitor traffic patterns to help make future transport decisions for the county.*
>
> *Across the county, 32 sensors have been placed which are able to classify what is using the highway, for instance pedestrians, cars, buses, bicycles, and count those users and record their speeds.*
>
> *Kent County Council Cabinet Member for Highways and Transport, Michael Page, said: "We have been working on several trials as part of the Live Labs programme which tests the very latest technology to see how it can help us save money, find and fix problems quicker, and make stronger evidence-based decisions about our road network."*

Mät- och analysmetoder

När det gäller att utreda emissioner från vägtrafiken så visar forskning att trafikflödesmätningar är en tillförlitlig metod för att utvärdera trafikkällornas bidrag till lokala luftföroreningar. Det finns även en förväntan på Trafikverket från externa samarbetspartner att bidra i luftkvalitetsarbetet just genom att minska trafikens påverkan på luftföroreningar. Även för trafiksäkerhetsanalyser så är trafikflödesdata där fordonens hasighet inkluderas viktig. Behov av mer detaljerad trafikflödes- och hastighetsdata finns både internt på Trafikverket och hos externa parter som intervjuats under projektet.







För att bidra med denna information så bör tekniker utredas både för hur de vägnära mätningar som upphandlas idag av Trafikverket kan utvecklas och hur andra källor kan bidra till värdefull kompletterande information. Zalakeviciute et al. (2020) beskriver en metod för att förutsäga urbana luftföroreningar från trafik genom att extrahera data från webbaserade applikationer (Google Traffic). Trafikdata som denna som samlas in på en mer övergripande nivå kan vara ett bra komplement till de mer detaljerade mätningar som trafiksensorer nära vägen kan ge. En annan möjlig källa som Trafikverket har utrett i syfte att samla in hastighetsdata är mobildata. Rapporten Big Data inom vägtrafikdata av Adolfsson, L. (2019). Mobildata beskriver att mobildata idag inte kan ersätta dagens vägnära mätningar genom fasta installationer och slangmätningar men kan möjligtvis komplettera dessa på en mer överskådlig nivå. I rapporten står:

> *”Utifrån erfarenheterna i detta projekt, att mobilbaserade trafikdata behöver kombineras med data från vägnära mätsystem, behöver samordning ske mellan ansvariga från samtliga verksamhetsgrenar för fasta mätsystemen i och kring fysisk väg. Detta är nödvändigt för att skapa synergier men också för att undvika dubbelinvesteringar och onödiga framtida driftkostnader.*
>
> *I fortsatt arbete är det av stor vikt att samordna insamling av data för olika syften och mellan olika verksamhetsområden på Trafikverket.”*

Detta resultat bekräftas även i denna studie. Med dagens metoder så görs mätningar för specifika syften, till exempel hastighetsmätningar, flödesmätningar, trafikmätningar och hastighetsövervakning, vilka med stor sannolikhet tekniskt sett kan kombineras. En stor risk med dagens uppdelning är att insamlingen är sub-optimerad och kostsam jämfört med om den skulle samordnas i högre grad. I inledningen av denna rapport nämns att flera verksamheter i olika sektorer har satt upp kostsam IT-infrastruktur för att kunna jobba med exempelvis AI och andra typer av beslutsstödsystem i stor skala, men har haft svårt att hitta tillämpningar och arbetssätt som ger avkastning på investeringen. Resultatet denna studie visar är att det antagligen kommer vara svårt att se avkastning på investeringar i data-insamling och analys på Trafikverket utan samordning av på en högre verksamhets-nivå görs. För att på längre sikt arbeta med mer avancerad dataanalys och AI är det nödvändigt att använda data i flera olika syften och framför allt att använda flera olika kompletterande datakällor i analysen. För detta är samordning nödvändig.

Utgår man från dagens mätningar som görs av Trafikverket skulle vissa flödes-mätningar möjligtvis kunna ge högre upplösning än dagens trafikflöden som beskrivs som ÅDT. Flödesmätningar som görs i 10 dagar kan visa medelhastigheten per timme. Även trafikflödesmätningarna på E4 utanför Stockholm som nämnts tidigare i rapporten kan potentiellt ha mer detaljerade data om trafikflöden och variationer i hastighet. Möjligtvis kan dessa interna källor vara en utgångspunkt för att beskriva trafikflöden på en mer detaljerad nivå.

För detektion av olika typer av trafikanter och en bättre förståelse skulle bildanalys med data från kameror vara en väg framåt.

Won (2020) presenterar en granskning av olika tekniker för trafikövervakning med fokus på fordonsindelning i olika klasser. Artikeln beskriver olika tekniker, forskningsfrågor och tekniska utmaningar som implementeringserfarenhet och systemprestanda. Denna genomgång är en intressant referens för utvärdering av möjliga framtida lösningar.

Utöver den trafikflödesdata som är prioriterad finns det ett generellt behov av en fördjupad kunskap om tillståndet för luftkvalitet. När det gäller luftkvalitet ligger i Sverige huvudansvaret för att mäta, enligt en tydligt definierade EU-regler, på kommunerna. Dessa mätningar innebär relativt dyra installationer på få platser i en stad. För att komplettera de fasta mätstationerna ger så finns olika potentiella tekniker.

I dagsläget utvecklas mätningar med mikrosensorer, som är billigare än de fasta mätstationerna. Fördelen med dessa är att de är små och kan placeras på många fasta positioner eller på mobila mätenheter för att täcka ett större mätområde. Nackdelen med mikrosensorerna är att de ännu har osäker kvalitet och en utvärdering behövs för att klargöra i vilka typer av mätningar de kan användas till. I flera olika projekt så har detta utvärderats genom jämförelser och kalibrering med de fasta stationerna. Ett exempel är luftkvalitetsmätningar som gjorts inom Senseable Stockholm Lab (www.senseablestockholm.org). Dessa tyder på att

mikrosensorerna kan, eller med ytterligare teknisk utveckling kommer att kunna, användas till vissa tillämpningar, till exempel för att mäta förändringar av luftkvalitet i ett område.

En annan källa för att mäta luftkvalitet är att använda satellitdata. Det är ett område under utveckling. Fördelarna med satellitdata är att de täcker en stor yta. Forskning pågår för att utveckla metoder för att använda satellitdata för att övervaka och predicera luftkvalitet. Det krävs dock ett arbete för att reda ut hur man kan applicera kunskapen på ett tillförlitligt sätt. Bland annat hävdar Sowden et al. 2018 att den rumsliga upplösningen i satellitdata är för grov för att göra lokala luftkvalitetsstudier. Ett flertal artiklar i Bilaga C diskuterar användningen av satellitdata för luftkvalitetsanalyser.

Eftersom ansvaret för monitorering av luftkvalitet ligger på kommunerna i Sverige så är det viktigt att samarbeta med dessa och andra parter som driver analysutvecklingen, till exempel SMHI och Naturvårdsverket. Arbetet bör samordnas så att olika datakällor och analyser kan komplettera varandra på ett effektivt sätt.

Samordning och ansvarsfördelning

Under denna studie har det visat sig vara komplicerat att sätta sig in i vilken typ av datainsamling och analys som genomförs på Trafikverket kopplat till arbetet med luftkvalitet och trafiksäkerhet. Data samlas in för många olika syften av många olika verksamhetsområden och regioner inom myndigheten och ansvaret för att definiera behov, genomföra mätningar och analyser sker inte på samordnat sätt. Denna brist på samordning medför med stor sannolikhet att budgeten som används inom detta område för de olika verksamheterna är stor i förhållande till värdet på resultaten. Det är även ett direkt hinder för att arbeta med avancerad dataanalys och AI, då data är svår att hitta, tolka och det skulle vara enormt resurskrävande att till en högre grad kombinera olika datakällor i analyser.

För att man inom Trafikverket ska kunna samordna sina insatser inom mer avancerad dataanalys och AI är det nödvändigt att ett tydligt ansvar och mandat för att samordna skapas på myndigheten. Den samordnande verksamheten bör samla in behoven från myndighetens olika delar, koordinera insatser för att samla in data och bidra till analysen.

I en artikel i realtid.se (www.realtid.se/debatt/statlig-datastrategi-kravs-framgangsrik-digitalisering) debatteras att en statlig datastrategi krävs för framgångsrik digitalisering. Följande citat är aktuellt även för Trafikverket:

> *”Förslag till företag och myndigheter är att knyta samman investeringar i datahantering med analys. Investera i tekniklösningar som förbättrar dataflödet, utbilda befintliga medarbetare och/eller anställ ett team av*

experter på data och analys, skapa en samarbetskultur mellan data- och analysteamen samt designa dataflöden som använder nya maskin-inlärnings-, kognitiva och artificiella intelligensfunktioner."

Genom att samordna behoven inom Trafikverket finns en grund för att göra innovationsupphandling på ett effektivt och utmaningsdrivet sätt. Det kan ge bättre svar att fokusera upphandlingen på vilka samlade utmaningar Trafikverket har och vilka specifika frågor man vill ha svar på och vara lösningsoberoende, i stället för att utvärdera en specifik produkt eller datakälla. Troligt är att mer kreativa lösningar presenteras om man i innovationsupphandlingen beskrev en utmaning som är lösningsoberoende, i stället för att utreda en viss teknik.

Slutligen är det även nödvändigt att Trafikverket på ett mer konkret sätt samordnar sina insatser med externa samarbetspartner. I arbetet med luftkvalitet är Trafikverkets roll inte tydlig. Det finns mål till 2030 att utsläppen från transport-sektorn har minskat så att miljökvalitetsmålet Frisk luft för NO2 i urban bakgrund och PM10 i gaturum uppnås. För att uppnå det målet är det rimligt att också klargöra vilka konkreta åtgärder som kan bidra till målet. För att komma fram till åtgärder är en utförlig analys av trafikens bidrag till luftföroreningar nödvändig. Den bör göras i samarbete med främst kommuner, SMHI och Naturvårdsverket då de är viktiga parter i frågan. Precis som frågan om intern samordning, så är den om ansvar i arbetet med att minska luftföroreningar från trafiken viktig för myndigheten att klargöra. Det är frågor som behöver lösas centralt på Trafikverket för att skapa förutsättningar för att utföra arbetet i organisationen.

Inspiration från Tysklands miljöminister Svenja Schulze finnes i en artikel i euractiv.com (www.euractiv.com/section/digital/interview/sustainability-should-be-underlying-principle-of-digital-policy-german-minister/)

"Digital innovations should be used as a tool for environmental protection and climate action, but we must act now, Germany's minister for the environment, Svenja Schulze, told EURACTIV in an interview.

Svenja Schulze currently serves as Germany's Minister for the Environment. She conducted this interview with EURACTIV's Samuel Stolton.

You've previously said you believe Europe's digital transition could help reach the bloc's climate goals for 2030 as part of the Green Deal. In what ways would you expect digital tools to be employed to help meet these targets?

Digital solutions are the key to bringing about the social and environmental transformation of our society. Digital management enables us to tap previously unimagined potential for efficiency, reduce resource consumption and achieve the climate goal of a CO2-free economy. Digital technology forms the basis for the energy transition and the mobility of tomorrow; just

think of how traffic in cities can be reduced through smart management and linking sharing services with a strong public transport system.

By collecting and analysing vast amounts of data, we are also harnessing the potential of AI for environmental protection and climate action.”

Slutsatser och rekommendationer

I denna studie var det första steget att klargöra vilka specifika utmaningar som finns i arbetet med att minska emissioner från och öka trafiksäkerheten i vägtransportnätverket. Fokus var att definiera vilken typ av data och analys som bättre skulle kunna ge underlag för åtgärder.

I arbetet framkom att behoven kopplat till analyser för luftkvalitet och för trafiksäkerhet delvis överlappar varandra.

Både för luftkvalitetsanalyser och trafiksäkerhetsanalyser så finns behov av bättre kartdata i analyser. För luftkvalitet är det viktigt att förutom vägkarta inkludera information om den omgivande miljön, till exempel vägsidesområden och byggnationer i närhet av vägarna. För trafiksäkerhet finns behov att kartlägga till exempel mitträcken, sidoräcken, korsningar och korsningsutformning, räfflor i mitten och sidan av vägen, väglinjer, kurvatur av vägar och utformning av vägsidesområden.

Behoven för luftkvalitet och trafiksäkerhet skulle troligtvis kunna lösas med AI-tekniker för bildanalys som idag är tillgängliga. Vägscanningen som idag görs inom verksamhetsområde Underhåll har potential att samordnas i flera syften. En utökad scanning för att täcka fler behov föreslås därför att utredas. För att gå vidare med ett sådant case är det till att börja med en samordning av behov från olika verksamheter som behöver göras innan en eventuell innovationsupphandling eller utredning av vilka tekniska lösningar som är möjliga görs. Sannolikt finns det behov av mer detaljerad vägdata från andra delar av Trafikverkets organisation, som inte undersökts i denna studie.

När det gäller trafikflödesdata så finns även där överlappande behov för luftkvalitet och trafiksäkerhet. För luftkvalitet behövs mer detaljerad information om hur trafiken varierar i tid (varje timme under ett dygn) när det gäller antal fordon, typ av fordon och hastigheter för dessa fordon. Detta för att kunna ta fram flödeskategorier. För trafiksäkerhet är information om vilka typer av fordon och oskyddade trafikanter som rör sig längs och tvärs vägarna och vilka hastigheter de har viktig. Man behöver veta mer om avvikelser i hastighetsefterlevnad för att analysera problemen. För att täcka dessa behov behöver man utreda vilken teknik eller kombinationer av tekniker, vägnära mätningar, mobildata etc., som kan täcka behoven. Vägnära kameror där trafikanter och trafiksituationer kan identifieras med bildanalys är en teknik som skulle kunna ersätta slangmätningar och slingor i marken. Tester för att utvärdera ny teknik för att göra trafikflödes- och hastighetsmätningar föreslås därför.

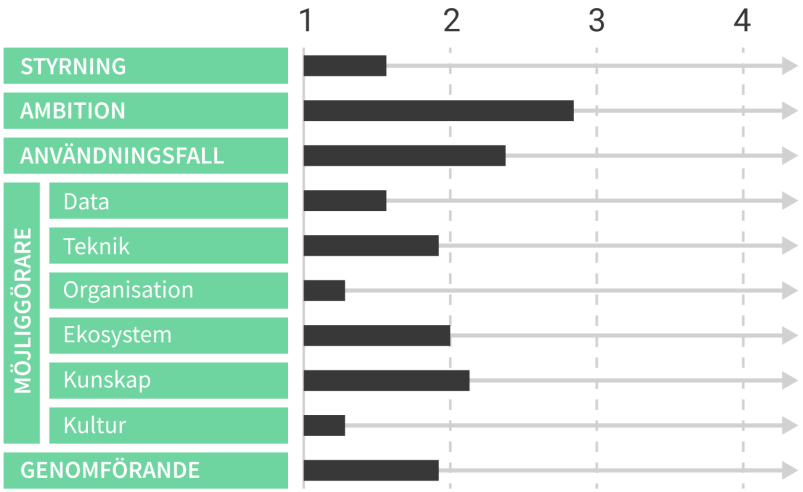
I fortsatt arbete är det av stor vikt att samordna insamling av data för olika syften och mellan olika verksamhetsområden på Trafikverket. Med dagens metoder görs mätningar för många olika syften, vilka med stor sannolikhet tekniskt sett skulle kunna kombineras i mycket högre grad. En stor risk med dagens uppdelning är att insamlingen är suboptimerad och kostsam jämfört med om den skulle samordnas.

För att man inom Trafikverket ska kunna samordna sina insatser inom mer avancerad dataanalys och AI är det nödvändigt att ett tydligt ansvar och mandat för att samordna skapas på myndigheten. Den samordnande verksamheten bör samla in behoven från myndighetens olika delar, koordinera insatser för att samla in data och bidra till analysen.

För att på längre sikt arbeta med mer avancerad dataanalys och AI är det nödvändigt att ha möjlighet att använda data i flera olika syften och framför allt att använda flera olika kompletterande datakällor i analysen. För detta är samordning nödvändig.

Samordningen behöver göras internt men också med externa samarbetspartner både för att klarlägga behoven och för att utveckla lösningar. För luftkvalitet bör arbetet göras i samarbete med främst kommuner, SMHI och Naturvårdsverket då de är viktiga parter i frågan. Vägscanning och trafikflödesmätningar skulle gynnas av att genomföras med samma metod för Trafikverket och kommuner för att möjliggöra bättre analyser på ett nationellt plan. Precis som frågan om intern samordning, så är den om ansvar i arbetet med att minska luftföroreningar från trafiken viktig för myndigheten att klargöra. Det är frågor som behöver lösas centralt på Trafikverket för att skapa förutsättningar för att utföra arbetet i organisationen.

En möjlig start för att ta nästa steg i arbetet i organisationen är att genomföra en skattning av organisationens AI-mognad med sikte på en AI-transformation. AI Sweden har ingått ett partnerskap med ett tyskt institut, UnternehmerTUM, som tagit fram en skattningsmodell som idag används av åtta olika länder. Skattningsmodellen består av flera påståenden som utvalda inom organisationen får ta ställning till. Man kan dela in organisationen i olika grupper för att se hur olika delar av organisationen ställer sig till olika områden. Detta är ett bra sätt för att se om det finns samsyn eller om det krävs ett arbete för att få samma bild av var man står. Resultatet sammanställs sedan enligt en modell och presenteras i en liknande figur som kan ses i 10 dimensioner. Detta är sedan ett underlag för vidare analys och arbete.



Skattningsmodell för AI-mognad

Referenser

Magnusson, U., & Eriksson, P. (2020). Digitaliseringens bidrag till målbild 2030 : Tillgänglighet i ett hållbart samhälle. Hämtad från Trafikverket website: http://urn.kb.se/resolve?urn=urn:nbn:se:trafikverket:diva-4197

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). https://doi.org/10.18653/v1/d19-1371

Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. Foundations and Trends in Information Retrieval, 3(4), 333–389. https://doi.org/10.1561/1500000019

Robertson, S. (1977). The probability ranking principle in IR. Journal of Documentation, 33(4), 294–304. https://doi.org/10.1108/eb026647.

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation, 28(1), 11–21. https://doi.org/10.1108/eb026526

Gustafsson, M., Lindén, J., Tang, L., Forsberg, B., Orru, H., Åström, S., & Sjöberg, K. (2018). Quantification of population exposure to NO2, PM2.5 and PM10 and estimated health impacts. Hämtad från http://urn.kb.se/resolve?urn=urn:nbn:se:naturvardsverket:diva-7870

Carolien Beckx, Luc Int Panis, Karen Van De Vel, Theo Arentze, Wouter Lefebvre, Davy Janssens, Geert Wets, The contribution of activity-based transport models to air quality modelling: A validation of the ALBATROSS–AURORA model chain, Science of The Total Environment, Volume 407, Issue 12, 2009, Pages 3814-3822, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2009.03.015.

Trafikverket. (2020). Analys av trafiksäkerhetsutvecklingen 2019 : Målstyrning av trafiksäkerhetsarbetet mot etappmålen 2020. Hämtad från http://urn.kb.se/resolve?urn=urn:nbn:se:trafikverket:diva-4327

Zalakeviciute, R.; Bastidas, M.; Buenaño, A.; Rybarczyk, Y. A Traffic-Based Method to Predict and Map Urban Air Quality.Appl. Sci. 2020, 10, 2035. https://doi.org/10.3390/app10062035

Adolfsson, L. (2019). Big Data inom vägtrafikdata : En förkommersiell upphandling för kreativa lösningar. Hämtad från https://fudinfo.trafikverket.se/fudinfoexternwebb/pages/ProjektVisaNy.aspx?ProjektId=3915

M. Won, Intelligent Traffic Monitoring Systems for Vehicle Classification: A Survey in IEEE Access, vol. 8, pp. 73340-73358, 2020, doi: 10.1109/ACCESS.2020.2987634.

M. Sowden, U. Mueller, D. Blake, Review of surface particulate monitoring of dust events using geostationary satellite remote sensing, Atmospheric Environment, Volume 183, 2018, Pages 154-164, ISSN 1352-2310, https://doi.org/10.1016/j.atmosenv.2018.04.020. (https://www.sciencedirect.com/science/article/pii/S1352231018302516)

Bilagor

Bilaga A Resultat av nyhetssökningar

Bilaga B Scientosemantics of applied research using AI and machine learning – a sequential approach

Bilaga C Utvalda vetenskapliga publikationer baserat på sökningar i BM25 sökmotor

Bilaga A – Resultat av nyhetssökningar

Nyhetssökningar gjordes med hjälp av tjänsten Retriever (retrievergroup.com). Sökningarna gjordes i följande medier:

- Svensk tryckt press (ex lokaltidningar, magasin, fackpress)
- Svensk webb, både webbsidor med redaktionellt innehåll samt webbsidor tillhörande myndigheter, organisationer, föreningar och liknande.
- Engelska, tyska, spanska & franska webbkällor. Indelning utifrån språk. Webbsidor med redaktionellt innehåll av olika karaktär, t ex stora internationella nyhetskällor, nyhetsbyråer, mindre lokala nyhetskällor, fackpress.

Sökningen sattes upp med sökord på engelska och svenska både för båda områdena luftkvalitet och AI samt trafiksäkerhet och AI. Sökningen gjordes för material som publicerats från 1/9 2020 till 28/2 2021.

Sökorden som användes för luftkvalitet och AI på engelska var Artificial intelligence (AI), Machine learning (ML), Deep learning (DL) i kombination med Particulate Matter (PM) PM1, PM2,5, PM10, NOx, CO2, Air quality + noise, Intelligent decision support systems.

Sökorden som användes för luftkvalitet och AI på svenska var artificiell intelligens (AI), maskininlärning (ML), djupinlärning (DL) i kombination med Luftburna partiklar (PM) PM1, PM2,5, PM10, NOx, CO2, Luftkvalitet + buller, Intelligenta beslutsstödsystem.

För sökningen om trafiksäkerhet och AI användes de engelska sökorden Artificial intelligence (AI), Machine learning (ML), Deep learning (DL) i kombination med Traffic safety, Safety assessment, Functional safety, Active safety, Passive safety, Advanced driving assistance systems (ADAS), Accident prevention, accident avoidance, Incident prevention, incident avoidance, Vision zero + traffic accidents, Road network, Road safety, Intelligent decision support systems.

För sökningen på svenska för trafiksäkerhet och AI användes sökorden *Artificiell intelligens (AI), Maskininlärning (ML), Djupinlärning (DL)* i kombination med *Trafiksäkerhet,*

Säkerhetsbedömning, Funktionell säkerhet, Aktiv säkerhet, Passiv säkerhet, Avancerade förarassistanssystem (ADAS), Olycksförebyggande, Olycksundvikande, Incidentförebyggande, incidentundvikande, Nollvisionen + trafikolyckor, Vägnät, Vägsäkerhet, Intelligenta beslutsstödsystem.

För luftkvalitet och AI valdes 27 artiklar ut av totalt 2783 i sökningen. För trafiksäkerhet valdes 32 artiklar ut av totalt 1143 i sökningen.

Resultat från nyhetssökning luftkvalitet och AI

Nedan listas 27 utvalda artiklar och citat av 2783 totalt i sökningen.

Machine learning reveals that prolonged exposure to air pollution is associated with SARS-CoV-2 mortality and infectivity in Italy - https://pubmed.ncbi.nlm.nih.gov/32882464/

“Air pollution can increase the risk of respiratory diseases, enhancing the susceptibility to viral and bacterial infections. Some studies suggest that small air particles facilitate the spread of viruses and also of the new coronavirus, besides the direct person-to-person contagion. However, the effects of the exposure to particulate matter and other contaminants on SARS-CoV-2 has been poorly explored. Here we examined the possible reasons why the new coronavirus differently impacted on Italian regional and provincial populations. With the help of artificial intelligence, we studied the importance of air pollution for mortality and positivity rates of the SARS-CoV-2 outbreak in Italy.”

A dataset of urban traffic flow for 13 Romanian cities amid lockdown and after ease of COVID19 related restrictions - https://pubmed.ncbi.nlm.nih.gov/32959019/

“This dataset comprises street-level traces of traffic flow as reported by Here Maps™ for 13 cities of Romania from 15th. of May 2020 and until 5th. of June 2020. This covers the time two days before lifting of the mobility restrictions imposed by the COVID19 nation-wide State of Emergency and until four days after the second wave of relaxation, announced for 1st. of June 2020. Data were sampled at a 15-min interval, consistent with the Here API update time. The data are annotated with relevant political decisions and religious events which might influence the traffic flow. Considering the relative scarcity of real-life traffic data, one can use this data set for micro-simulation during development and validation of Intelligent Transportation Solutions (ITS) algorithms while another facet would be in the area of social and political sciences when discussing the effectiveness and impact of statewide restriction during the COVID19 pandemic.”

Airly Raises $2m For Global Air Quality Platform - https://www.thestreet.com/press-releases/airly-raises-2m-for-global-air-quality-platform-15447950

“Airly provides accurate, ultra-local, predictive air quality data for governments, companies and individuals to understand the air their communities breathe.

Airly's sensors act as a global warning system for pollution at street level and in real time. Its proprietary breakthroughs in sensor miniaturization and machine learning allow it to monitor more pollutants with greater accuracy, at lower cost, and in more places.

The system involves dense networks of sensors that measure particulate matter (PM1, PM2.5 and PM10) and gaseous compounds (NO2, SO2, O3 and CO). It provides users live access to this data through an intuitive mobile app and online map.”

Statlig datastrategi krävs för framgångsrik digitalisering - https://www.realtid.se/debatt/statlig-datastrategi-kravs-framgangsrik-digitalisering

”Förslag till företag och myndigheter är att knyta samman investeringar i datahantering med analys. Investera i tekniklösningar som förbättrar dataflödet, utbilda befintliga medarbetare och/eller anställ ett team av experter på data och analys, skapa en samarbetskultur mellan

data- och analysteamen samt designa dataflöden som använder nya maskininlärnings-, kognitiva och artificiella intelligensfunktioner.”

Frost & Sullivan Presents Top Sensor Technologies Impacting the Future of Smart Cities - https://ww2.frost.com/news/frost-sullivan-presents-top-sensor-technologies-impacting-the-future-of-smart-cities/

Stadsutveckling kräver nya verktyg – och de finns - https://hallbartbyggande.com/stadsutveckling-kraver-nya-verktyg-och-de-finns/

Vivacity Labs to deploy AI-controlled road junctions in Greater Manchester - https://highways.today/2020/11/09/vivacity-labs-ai-junctions-manchester/

“Using sensors with inbuilt artificial intelligence, Vivacity enables TfGM to anonymously identify different types of road users at selected junctions and control traffic signals to allow different modes of transportation to be prioritised as and when required. With more cyclists on the road as people avoid public transport, these ‘smart junctions’ will be able to give priority to people on foot or bike where and when appropriate.

Vivacity Labs’ first-of-its-kind AI signal control system first went live early this year, before scaling to simultaneously control three, neighbouring junctions in the Blackfriars area of Salford in September 2020. This initiative also has the potential to reduce emissions and improve air quality in the Greater Manchester area. Congestion and queuing can be reduced by traffic signals that respond better and more quickly to changes in traffic conditions than existing systems.”

India uses artificial intelligence, geo testing to identify pollution hotspots - https://www.presstv.com/Detail/2020/11/12/638457/India-Artificial-Intelligence-Geo-Testing-Pollution-Hotspots

FRESH AIR AND QUIET: APP FINDS BEST HELSINKI ROUTES FOR BIKERS AND HIKERS - https://finland.fi/life-society/fresh-air-and-quiet-app-finds-best-helsinki-routes-for-bikers-and-hikers/

“See how Green Paths, a prototype app from the University of Helsinki’s Digital Geography Lab, helps cyclists and pedestrians in the Finnish capital area choose routes with the cleanest air and the least noise.

The app is part of Healthy Outdoor Premises for Everyone (HOPE), a project instigated by the City of Helsinki and funded by the European Regional Development Fund. Unlike other journey planning apps, Green Paths is for active travelling such as cycling and walking, not for public transport or car travel. It is optimised according to noise, environment and distance.

It calculates routes on the basis of real-time air-quality data provided by the Finnish Meteorological Institute and open data on noise levels from the municipalities of Helsinki, Espoo, Vantaa and Kauniainen, which comprise the capital metropolis.”

Mapping air pollution at the neighborhood level - https://www.axios.com/mapping-air-pollution-neighborhood-level-ccf4a764-3a73-4008-9527-3d97d41427d2.html

“How it works: Breezometer taps air quality sensors to create a microlocal forecast that changes every 15 ft. to 16 ft., in real time, says Ran Korber, the company's founder and CEO.

Users can see air pollution levels not just in their region or city, but along their block.

Details: Machine-learning algorithms can forecast air quality for dozens of pollutants with what Korber says is 90% confidence six hours into the future, though that accuracy declines over longer periods.”

Enabling data-driven governments - https://www.securitysa.com/11865r

“Across the globe, governments are exploring ways to use data to improve the life of citizens, increase efficiency of operations, and create more innovative business models. Gathering, accessing and analysing data can assist governments with insight to implement strategies, monitor the impact of those policies, and ensure a safe living and working environment, while data can also assist in delivering proactive services to the public.

Aruba is focused on creating intelligent network solutions that deliver the innovation and flexibility to power this next generation of data-driven, real-time applications.”

AI set to cut GHG emissions by 16% in the next 3-5 years - https://www.smart-energy.com/industry-sectors/new-technology/ai-set-to-cut-ghg-emissions-by-16-in-the-next-3-5-years/

Hela rapporten: https://www.capgemini.com/wp-content/uploads/2020/12/Report-Climate-AI.pdf

“The experts in our survey point to the absence of a data-driven culture (86%) as one of the major challenges in adopting AI for climate change. Organizations need to take a data-driven and scientific approach to tackling climate change with AI.”

“The coming decade will be critical when it comes to climate action and whether the world manages to avoid highly damaging climate change. Organizations across sectors need to take urgent actions to reduce their emissions. While many traditional methods are being mplemented, innovation is critical, and we have found that using AI remains largely unexplored. Leveraging AI’s full climate action potential sustainably is mission-critical for the world as a whole. To kick-start progress, organizations need to account for and take easures to combat the negative impact of AI on climate, educate sustainability teams on how AI can make a real difference and educate AI teams on the criticality of climate change, focus on building technological foundations, scale most impactful use cases, collaborate with the larger climate ecosystem and finally align scaling of AI use cases with emissions. Doing nothing is not an option.”

NASA model reveals how much COVID-related pollution levels deviated from the norm - https://www.eurekalert.org/pub_releases/2020-11/nsfc-nmr111720.php

“Using computer models to generate a COVID-free 2020 for comparison, NASA researchers found that since February, pandemic restrictions have reduced global nitrogen dioxide concentrations by nearly 20%. The results were presented at the 2020 International Conference for High Performance Computing, Networking, Storage, and Analysis.”

En genomgripande omvandling av transportsektorn: EU-kommissionen presenterar sin plan för miljövänlig och smart mobilitet till ett överkomligt pris - https://ec.europa.eu/commission/presscorner/detail/sv/IP_20_2329

”EU-kommissionen presenterade i dag sin strategi för hållbar och smart mobilitet tillsammans med en handlingsplan med 82 initiativ som kommer att vägleda arbetet under de kommande fyra åren. Strategin lägger grunden för hur EU:s transportsystem kan genomgå en grön och digital omvandling och bli mer motståndskraftigt mot framtida kriser. Enligt den europeiska gröna given kommer resultatet att bli en minskning av utsläppen med 90 % fram till 2050, med hjälp av ett smart, konkurrenskraftigt, säkert, tillgängligt och ekonomiskt överkomligt transportsystem.”

You Are What You Breathe: Polluted Air Tied to Plaques, Brain Atrophy - https://www.alzforum.org/news/research-news/you-are-what-you-breathe-polluted-air-tied-plaques-brain-atrophy

“Could air pollution even within Environmental Protection Agency limits—that is, not only in the world’s most polluted cities—hasten a person’s brain pathology and dementia? Yes indeed, according to two recent studies. In the November 30 JAMA Neurology, researchers led by Gil Rabinovici at the University of California, San Francisco, linked worse air pollution to higher brain amyloid burden in the Imaging Dementia–Evidence for Amyloid Scanning (IDEAS) cohort. In more than 18,000 cognitively impaired older adults from all over the U.S., the scientists compared brain amyloid plaque burden with concentrations of fine particulate matter (PM2.5) and of ground-level ozone. People who lived in an area with high PM2.5 levels had more plaques; ozone had no effect. A study in the November 18 Neurology led by Diana Younan, University of Southern California, Los Angeles, links air pollution to brain atrophy in 1,365 cognitively normal older participants in the Women’s Health Initiative Memory Study (WHIMS). The women who lived in areas of high PM2.5 had less gray matter in brain regions vulnerable to atrophy in Alzheimer’s disease. Taken together, these large cohorts add complementary evidence to a building consensus that toxic air may increase dementia risk.”

What Will the Workplace Look Like in 2025? - https://www.shrm.org/hr-today/news/all-things-work/pages/the-workplace-in-2025.aspx

“The shift to remote work will be among the biggest business trends in the coming years, though it won’t be the only lingering effect from the pandemic.”

What Are Smart Cities? - https://www.cbinsights.com/research/what-are-smart-cities/

“Technology is powering the rise of smart cities, transforming everything from traffic management to waste collection. We dig into the digital revolution giving rise to cities that are more connected, sustainable, and efficient — and what the future of urbanization might look like.”

Copernicus: studie bekräftar minskningar av utsläppen under den första vågen av COVID-19 i Europa - https://news.cision.com/se/copernicus-climate-change-services/r/copernicus--studie-bekraftar-minskningar-av-utslappen-under-den-forsta-vagen-av-covid-19-i-europa,c3256639

”Copernicus Atmosphere Monitoring Service bekräftar tillsammans med Spaniens Supercomputing Center i Barcelona minskningar av kväveoxider (NOx) och CO2-utsläpp av fossila bränslen i hela Europa under den första vågen av COVID-19-pandemin. Genom att använda artificiell intelligens ger studien en uppdelning av sektoriella och rumsliga utsläpp per dag samt information på landsnivå av olika föroreningar.”

Governments' Focus on Curbing Urban Air Pollution to Propel Adoption of Sensor and IoT Technologies in Air Quality Monitoring Equipment Market: TMR - https://www.prnewswire.com/news-releases/governments-focus-on-curbing-urban-air-pollution-to-propel-adoption-of-sensor-and-iot-technologies-in-air-quality-monitoring-equipment-market-tmr-301199260.html

“- Adoption of low-cost portable sensors by industry stakeholders and governments to advance monitoring of air quality, smart IoT-powered devices and drones seeing vast potential in making real-time air pollution response

- Strategic partnerships in developing and developed economies pave way to new technologies, global air quality monitoring equipment market to reach ~US$ 7 Bn by 2027”

You can now check your phone for hyperlocal air quality - https://www.israel21c.org/you-can-now-check-your-phone-for-hyperlocal-air-quality/

“iPhone users can now keep up to date on the air quality in their immediate surroundings with new features in the Apple Weather app courtesy of Israeli startup BreezoMeter.

The information from BreezoMeter, which utilizes machine learning to search through multiple sources of real-time air quality, will be accessible both through Apple Maps and Siri following the recent iOS 14.3 update.

The hyperlocal air quality data is available for locations in the United States, Britain, Germany, India and Mexico, allowing users to make immediate and informed decisions to reduce their exposure to air pollution (and giving them the perfect excuse not to go out for that run).”

Spatio-Temporal Representativeness of Air Quality Monitoring Stations in Mexico City: Implications for Public Health - https://www.frontiersin.org/articles/10.3389/fpubh.2020.536174/full#h4

“Assessment of the air quality in metropolitan areas is a major challenge in environmental sciences. Issues related include the distribution of monitoring stations, their spatial range, or missing information. In Mexico City, stations have been located spanning the entire Metropolitan zone for pollutants, such as CO, NO2, O3, SO2, PM2.5, PM10, NO, NOx, and PMCO. A fundamental question is whether the number and location of such stations are adequate to optimally cover the city. By analyzing spatio-temporal correlations for pollutant measurements, we evaluated the distribution and performance of monitoring stations in Mexico City from 2009 to 2018. Based on our analysis, air quality evaluation of those contaminants is adequate to cover the 16 boroughs of Mexico City, with the exception of SO2, since its spatial range is shorter than the one needed to cover the whole surface of the city. We observed that NO and NOx concentrations must be taken into account since their long-range dispersion may have relevant consequences for public health. With this approach, we may be able to propose policy based on systematic criteria to locate new monitoring stations.”

Association between exposure to airborne pollutants and COVID-19 in Los Angeles, United States with ensemble-based dynamic emission model - https://pubmed.ncbi.nlm.nih.gov/33417905/

“This study aims to find the association between short-term exposure to air pollutants, such as particulate matters and ground-level ozone, and SARS-CoV-2 confirmed cases.”

Luften blev inte så mycket bättre i nedstängda samhällen - https://www.dn.se/vetenskap/luften-blev-inte-sa-mycket-battre-i-nedstangda-samhallen/

Caligenix® Innovates Real-Time Way to Display 'Environmental Health Index™ ' Based on Zip Code - https://es-us.finanzas.yahoo.com/news/caligenix-innovates-real-time-way-150500474.html

"There's a constant dance between nature and nurture when it comes to the relationship between one's external and internal environments," said Eliad Josephson, Caligenix Co-Founder. "People need to have a clearer understanding of that relationship—especially during this pandemic—so that they're able to decide what actions they should take based on external factors in their area, like weather, pollution, the number of COVID cases, and so forth. Our mission is to provide wellness solutions personalized to one's lifestyle and genomic data, which is why we designed this tool to facilitate decisions based on real-time, location-specific data."

Air quality and health impact of 2019–20 Black Summer megafires and COVID-19 lockdown in Melbourne and Sydney, Australia - https://www.sciencedirect.com/science/article/abs/pii/S0269749121000762?dgcid=rss_sd_all

“Highlights

• Air pollution in Australia was modelled using meteorological normalisation.

• Accounting for meteorology is necessary for attributing air quality impacts.

• Air quality improvement during COVID-19 lockdown highly regionally dependent.

• PM2.5 and O3 from Black Summer smoke likely caused 187 excess deaths."

French court rules France not doing enough on climate change - https://www.carbonbrief.org/daily-brief/french-court-rules-france-not-doing-enough-on-climate-change

"A French court has ruled that the country's government must do more to combat climate change, reports Reuters, "in what environmental campaigners called a landmark ruling that could ramp up pressure on other countries to act on global warning". The case was brought by a group of non-governmental organisations (NGOs) – including the French branches of Oxfam and Greenpeace – who accused the French state of not living up to its own commitments, the newswire explains."

Sustainability should be 'underlying principle' of digital policy: German minister - https://www.euractiv.com/section/digital/interview/sustainability-should-be-underlying-principle-of-digital-policy-german-minister/

"Digital innovations should be used as a tool for environmental protection and climate action, but we must act now, Germany's minister for the environment, Svenja Schulze, told EURACTIV in an interview.

Svenja Schulze currently serves as Germany's Minister for the Environment. She conducted this interview with EURACTIV's Samuel Stolton.

You've previously said you believe Europe's digital transition could help reach the bloc's climate goals for 2030 as part of the Green Deal. In what ways would you expect digital tools to be employed to help meet these targets?

Digital solutions are the key to bringing about the social and environmental transformation of our society. Digital management enables us to tap previously unimagined potential for efficiency, reduce resource consumption and achieve the climate goal of a CO2-free economy. Digital technology forms the basis for the energy transition and the mobility of tomorrow; just think of how traffic in cities can be reduced through smart management and linking sharing services with a strong public transport system.

By collecting and analysing vast amounts of data, we are also harnessing the potential of AI for environmental protection and climate action."

Resultat från nyhetssökning trafiksäkerhet och AI

Nedan listas 32 utvalda artiklar och citat av totalt 1143 i sökningen.

Ford Is Working On New Technology To Predict Road Accidents Even Before They Happen - https://www.indiatimes.com/auto/alternative/ford-new-technology-predict-road-accidents-521757.html

“Auto maker Ford is working on a predictive road safety tool that will be able to foresee traffic incidents once complete. The project is being worked upon in collaboration with transport bodies, other local authorities as well as universities in the UK.

The consortium led by Ford will work upon how advanced analytics and data from connected vehicles can be used “to improve urban mobility and road safety,” the car major said in a release.”

Iteris Selected by Florida Department of Transportation to Implement Smart Mobility and Safety Initiative - https://www.businesswire.com/news/home/20200901005379/en/Iteris-Selected-Florida-Department-Transportation-Implement-Smart

“Iteris to Deploy Safety and Operational Improvement System at Key Intersections in the State of Florida

- Iteris will implement a near-miss identification safety system at key intersections across three Florida counties
- Groundbreaking program supports FDOT’s Driving Down Fatalities goals to proactively improve road safety by using smart mobility technologies to reduce crashes statewide
- Deal marks expansion of Iteris’ safety-focused software-as-a-service offering”

A complex junction recognition method based on GoogLeNet model - https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12681?af=R

“Complex junctions are typical microstructures in large-scale road networks with intricate structures and varied morphologies. It is a challenge to identify junctions in map generalization and car navigation tasks accurately. Generally, traditional recognition methods rely on low-level characteristics of manual design, such as parallelism and symmetry. In recent years, preliminary studies using deep learning-based recognition methods were conducted. However, only a few junction types can be recognized by existing methods, and these methods cannot effectively identify junctions with irregular shapes and numerous interference sections. Hence, this article proposes a complex junction recognition method based on the GoogLeNet model.”

Uber's self-driving operator charged over fatal crash - https://www.bbc.com/news/technology-54175359

“Elaine Herzberg, aged 49, was hit by the car as she wheeled a bicycle across the road in Tempe, Arizona, in 2018.

Investigators said the car's safety driver, Rafael Vasquez, had been streaming an episode of the television show The Voice at the time.

Ms Vasquez pleaded not guilty, and was released to await trial.”

StreetLight Data and Ford Mobility Collaborate to Provide Industry-Leading Traffic Safety Solutions for U.S. and Canadian Communities - https://www.benzinga.com/pressreleases/20/10/n17919937/streetlight-data-and-ford-mobility-collaborate-to-provide-industry-leading-traffic-safety-solution

“StreetLight Data, Inc. ("StreetLight"), the leader in Big Data analytics for mobility, and Ford Mobility, Ford Motor Company's subsidiary dedicated to solving the world's most pressing mobility issues, today announced their collaboration to provide industry-leading, safety solutions for cities, local government agencies, and departments of transportation. This unique combination of vast data and machine-learning resources offers users the most comprehensive mobility dataset and "smart" safety decision tools available today.”

When Self-Driving Cars Don’t Realize They’ve Been In A Car Crash - https://www.forbes.com/sites/lanceeliot/2020/10/18/when-self-driving-cars-dont-realize-theyve-been-in-a-car-crash/?sh=1d37e2f2102a

“Will self-driving cars realize when they've been in a car crash?”

Identifying safe intersection design through unsupervised feature extraction from satellite imagery - https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12623?af=R

“The World Health Organization has listed the design of safer intersections as a key intervention to reduce global road trauma. This article presents the first study to systematically analyze the design of all intersections in a large country, based on aerial imagery and deep learning. Approximately 900,000 satellite images were downloaded for all intersections in Australia and customized computer vision techniques emphasized the road infrastructure.”

Kerala to get 700 Automatic Number Plate Recognition cameras by next Feb. - https://www.thehindu.com/news/cities/kozhikode/kerala-to-get-700-automatic-number-plate-recognition-cameras-by-next-feb/article32933484.ece

“As part of improving its automated traffic enforcement activities, the Motor Vehicles Department (MVD) will introduce 700 Automatic Number Plate Recognition (ANPR) cameras in the State by February 2021. The portable devices with Artificial Intelligence (AI) capabilities will work with the support of 14 special control rooms.”

Drivers'' smartphones could be used for road safety: UK study - https://www.outlookindia.com/newsscroll/drivers-smartphones-could-be-used-for-road-safety-uk-study/1964491

“Motorists with smartphones could help highway chiefs maintain road quality by sending "crowdsourced" data from their mobiles that would allow engineers to assess when carriageway repairs are needed, according to a new UK study.”

Analysis of Machine Learning Techniques Applied to Sensory Detection of Vehicles in Intelligent Crosswalks - https://pubmed.ncbi.nlm.nih.gov/33114001/

“Improving road safety through artificial intelligence-based systems is now crucial turning smart cities into a reality. Under this highly relevant and extensive heading, an approach is proposed to improve vehicle detection in smart crosswalks using machine learning models. Contrarily to classic fuzzy classifiers, machine learning models do not require the readjustment of labels that depend on the location of the system and the road conditions. Several machine learning models were trained and tested using real traffic data taken from urban scenarios in both Portugal and Spain. These include random forest, time-series forecasting, multi-layer perceptron, support vector machine, and logistic regression models. A deep reinforcement learning agent, based on a state-of-the-art double-deep recurrent Q-network, is also designed and compared with the machine learning models just mentioned. Results show that the machine learning models can efficiently replace the classic fuzzy classifier.”

'I'm not drunk, it's my car:' Tesla's 'full self-driving' gets mixed reviews - https://edition.cnn.com/2020/10/30/cars/tesla-full-self-driving/index.html

“Tesla has released an early version of its "full self-driving" software to a small group of Tesla enthusiasts, who appear to be both delighted and alarmed by what they've experienced so far.”

Innovationspris till Svevia - https://www.entreprenad.com/article/view/748336/innovationspris_till_svevia

”Svevias digitala vägunderhållssatsning belönas med innovationspriset Quality Innovation Award. Svevia får pris för projektet Behovsstyrd hållbar väghållning.”

FLIR Systems Announces Artificial Intelligence Traffic Cameras for Predictive Traffic Management - https://www.tmcnet.com/usubmit/-flir-systems-announces-artificial-intelligence-traffic-cameras-predictive-/2020/11/09/9252924.htm

“FLIR Systems, Inc. (NASDAQ: FLIR) today announced two intelligent traffic system cameras, the FLIR TrafiSense™ AI* with thermal imaging and the FLIR TrafiCam™ AI visible camera, both with artificial intelligence (AI) to optimize traffic flow on roadways and at intersections. When combined with the FLIR Acyclica™ cloud platform, cities can apply the AI-camera data to predict traffic, prevent congestion and potential accidents, and create safer roads for drivers, cyclists, and pedestrians 24 hours a day.”

Traffic experts moot use of GPS to track potholes - https://www.thehindu.com/todays-paper/tp-national/tp-kerala/traffic-experts-for-use-of-gps-to-track-potholes/article33062681.ece

“With the Public Works Department (PWD) disbanding its team of non-muster roll (NMR) workers that used to be part of preventive maintenance of roads over two decades ago, GPS navigation system installed in public transport and other vehicles can well be relied on to track potholes that cause traffic hold-ups, damage vehicles and result in accidents, according to traffic and road safety experts.”

NSW Transport and Microsoft use machine learning and data to reduce road accidents - https://www.zdnet.com/article/nsw-transport-and-microsoft-use-machine-learning-and-data-to-reduce-road-accidents/#ftag=RSSbaffb68

“Transport for New South Wales and Microsoft have partnered to develop a proof of concept that uses data and machine learning to flag potentially dangerous intersections and reduce road accidents.

As part of the proof of concept, Transport for NSW ran a trial in Wollongong to uncover five potentially risky intersections. It involved 50 vehicles generating more than a billion rows of data over a 10-month period, before Databricks and Azure were used to curate, ingest, and interpret the data.”

Hazen.ai Wins the IRF Global Road Achievement Award for Best use of "AI in Traffic Safety" - https://www.prnewswire.com/news-releases/hazenai-wins-the-irf-global-road-achievement-award-for-best-use-of-ai-in-traffic-safety-301179808.html

“Today, Hazen.ai has won the Intelligent Traffic Systems (ITS) category of the 2020 IRF Global Road Achievement Awards (GRAA) for demonstrating significant achievements in making the most of existing infrastructure to accommodate present and future traffic, by using advanced or original traffic management concepts, organizations or systems.

Hazen.ai recently launched its proprietary Seatbelt and Mobile Phone detection technology, providing the first and highest accuracy of seatbelt and mobile phone violation detection at more than 90% accuracy for both.”

Henrik Berglund är Sveriges smartaste bilförare - https://news.cision.com/se/greater-than/r/henrik-berglund-ar-sveriges-smartaste-bilforare,c3243278

”Idag vann Henrik Berglund från Trollhättan tävlingen Sveriges smartaste bilförare. Tävlingen som har pågått i fyra veckor har belönat smarta, säkra och miljövänliga färdigheter bakom ratten och tusentals svenska bilister har deltagit.

Bakom den tekniska lösningen står det svenska bolaget Greater Than som med sin AI-baserade analys och app beräknar varje deltagares körmönster i realtid och poängsätter det i förhållande till C02 utsläpp och nivå av trafiksäkerhet.”

Autostrade Tech, IBM and Fincantieri NexTech Deploy a New Artificial Intelligence and IoT-based System for Monitoring Autostrade per l'Italia Civil Infrastructure - https://www.prnewswire.com/news-releases/autostrade-tech-ibm-and-fincantieri-nextech-deploy-a-new-artificial-intelligence-and-iot-based-system-for-monitoring-autostrade-per-litalia-civil-infrastructure-301180728.html

“- Autostrade per l'Italia has invested 60 million euros into improving its systems for monitoring and securing its civil infrastructure. IBM will provide advanced AI for data processing and Fincantieri NexTech supplies technologies and solutions for high-availability monitoring.

- The solution combines drones, 3D modeling tools and cognitive data analysis to help make controls more efficient, transparent, and traceable. The system, managed by Autostrade Tech, is already integrated with the Italian Ministry of Infrastructures and Transportation's Public Civil Infrastructure Database.

- The collaboration expects to make the new platform for the management of civil infrastructure, based on IBM Maximo, available to the market.”

Project to track traffic violators launched - https://timesofindia.indiatimes.com/city/madurai/project-to-track-traffic-violators-launched/articleshow/79574224.cms

“Automatic number plate recording cameras installed at 15 places in Sivaganga town will help track traffic violators who do not follow helmet rule and pillion rider safety rule.”

Ny avancerad teknik ska hjälpa polisen sänka hastigheten - https://sverigesradio.se/artikel/7620646

”Polismyndigheten har beställt 2 200 nya bilar utrustade med avancerad teknik som ska hjälpa polisen att stoppa fartdårar och andra brottslingar.

De nya fordonen är utrustade med kameror med artificiell intelligens som på egen hand läser av registreringsskyltar på bilar, en inbyggda radarsensorn som mäter hastigheten på andra fordon och poliserna kommer i och med det få en signal direkt om någon bryter mot hastighetsgränsen — samtidigt som fortkörningen fångas på film.

"Det här bidrar till att vi kommer kunna utvidga vår kontrollverksamhet, och det i sin tur hoppas vi ska kunna leda till att vi sänker medelhastigheten på vårt vägnät", säger Petter Wahllöf, gruppchef för trafikpolisen i Karlskrona.”

Currux Vision LLC Announces Industry Leading Accuracy Of Artificial Intelligence Smart City Traffic Platform Testing With The City Of San José - https://www.publicworks.com/doc/currux-vision-llc-announces-industry-smart-city-traffic-platform-testing-with-the-city-of-san-jose-0001

“Currux Vision Smart City Intelligent Traffic Management Platform (“SmartCity ITS”) uses Artificial Intelligence (“AI”) at the edge to rapidly optimize traffic management. Offering industry leading traffic monitoring the Currux Vision Platform provides the big data, autonomous traffic management, and faster response to incidents to help predict and improve traffic flow, safety, and reduce congestion and pollution.”

Waycare helps manage Ohio traffic - https://www.itsinternational.com/its8/news/waycare-helps-manage-ohio-traffic

“Waycare Tech is implementing its cloud-based mobility platform in collaboration with the Central Ohio Transit Authority (Cota) to increase traffic safety and reduce travel time for residents.

The project is part of the US Department of Transportation Integrated Mobility Innovation Demonstration Research grant of more than $1.7 million awarded to Cota and 13 partners, including the Ohio Department of Transportation (ODoT) and the city of Columbus.

Waycare says its Regional Cloud-Based Traffic Management Artificial Intelligence System allows participating transit agencies, counties and public safety organisations to prepare for and react to real-time traffic information across 13 central Ohio counties on a web-based platform.”

Belgium: are traffic cameras to spot use of mobile phones illegal? - https://newmobility.news/2021/01/06/belgium-are-traffic-cameras-to-spot-use-of-mobile-phones-illegal/

“The use of mobile phones behind the wheel has been prohibited for a long time, but it remains a persistent practice. To increase the chances of being caught, MP Joris Vandenbroucke (sp.a) wants the police to use traffic cameras for this purpose.

Vandenbroucke has submitted a bill to this effect. In doing so, he hopes to reduce the death toll in traffic in Belgium significantly. The reason for doing this is a new camera system capable of detecting the use of mobile phones while driving that was recently tested by the traffic institute Vias.”

Smart city: Windsor is first Canadian city to launch Ford Safety Insights Platform to reduce crashes - https://www.techrepublic.com/article/smart-city-windsor-is-first-canadian-city-to-launch-ford-safety-insights-platform-to-reduce-crashes/#ftag=RSS56d97e7

“Windsor's partnership with Ford will provide the city with data gathered from AI and machine learning to predict and prevent traffic accident hotspots.”

Toyota Mobility Foundation Announces Six Finalists for the City Architecture for Tomorrow Challenge - https://livenews.co.nz/2021/01/28/toyota-mobility-foundation-announces-six-finalists-for-the-city-architecture-for-tomorrow-challenge/

“Sixteen teams of innovators, from 9 countries, developed proofs-of-concept (PoC)s of their innovative data-driven solutions towards improving mobility and city planning in Kuala Lumpur

Six Finalists were selected by a panel of judges and will receive a US$125,000 grant each to develop a Minimum Viable Product (MVP) as a ready-to-implement solution”

Greater Than raises SEK 136 million to scale its AI technology - https://www.avanza.se/placera/pressmeddelanden/2021/01/29/greater-than-greater-than-raises-sek-136-million-to-scale-its-ai-technology.html

“Greater Than is a leading Insurtech providing a new actuarial pricing and risk predictions model, based on Artificial Intelligence.”

Just how dangerous are smart motorways? - https://www.prospectmagazine.co.uk/science-and-technology/smart-motorways-how-dangerous-vs-deaths-uk

“They have cut journey times, raised the capacity of Britain’s busiest roads, and saved money—but there's a reason why smart motorways are hated”

Hikvision ITS camera monitors traffic flow - https://www.iotm2mcouncil.org/iot-library/news/smart-cities-news/hikvision-its-camera-monitors-traffic-flow/

“Chinese video IoT company Hikvision has launched an ITS camera for improving road safety and traffic flow.

The All-Rounder ITS camera is engineered with an all-in-one structure, embedding video, radar and supplemental light in one module, helping traffic authorities ramp up the detection of violations.

Hangzhou-based Hikvision is an IoT provider with video as its core competency. The All-Rounder ITS camera is designed to improve road safety and optimise traffic flow. As the name implies, the camera encompasses different skills and abilities, boasting speed detection, traffic violation detection, automated plate recognition and vehicle attribute analysis in one housing.”

Traffic enforcement to go digital - https://www.thehindu.com/news/national/kerala/traffic-enforcement-to-go-digital/article33824033.ece

“Traffic enforcement activities will turn further digital with the round-the-clock State control rooms in the capital and six district-level control rooms set up under the ₹236-crore Safe Kerala project going live.

Through the new 700 Artificial Intelligence and Automatic Number Plate Recognition (ANPR) cameras in vantage points, enforcement officials in the control rooms can ensure road discipline and book erring motorists. Dedicated staff had been posted in the control rooms to issue challans to offenders.

The real time monitoring is part of improving Motor Vehicles Department’s (MVD) automated traffic enforcement activities and to gradually end the manual enforcement activities. Besides, the MVD can put an end to the complaints, ensure transparency and reduce manpower.”

Kent County Council and Amey install traffic insight sensors - https://kccmediahub.net/kent-county-council-and-amey-install-traffic-insight-sensors745

“A trial using state of the art traffic sensors is being carried out by Kent County Council and its partners at Amey to monitor traffic patterns to help make future transport decisions for the county.

Across the county, 32 sensors have been placed which are able to classify what is using the highway, for instance pedestrians, cars, buses, bicycles, and count those users and record their speeds.

Kent County Council Cabinet Member for Highways and Transport, Michael Page, said: “We have been working on several trials as part of the Live Labs programme which tests the very latest technology to see how it can help us save money, find and fix problems quicker, and make stronger evidence-based decisions about our road network.”

Intersections evolve from Inductive Loops to Artificial Intelligence - https://highways.today/2021/02/24/intersections-evolve-ai/

“Research and a good evidence-based survey of the area are vital to create an effective and safe intersection. City planners must also monitor the ever-changing conditions to improve and evolve the junction.

The methodologies used to monitor traffic fluctuations has changed radically over the years, from simple signal timings, to more advanced vehicle detection systems. Improvements in transportation technology and smart highways have driven intersections to become intelligent, safer, and more reliable.”

Smart City Traffic Safety Project Will Be Implemented In Samsun - https://raillynews.com/2021/02/Smart-City-Traffic-Safety-Project-will-be-implemented-in-Samsun/

“Samsun Metropolitan Municipality has started to apply smart city technologies in all investments from infrastructure to superstructure. In this context, the Smart City Traffic Safety Project is started. The municipality, which has completed its geometric design projects and preparations for artificial intelligence systems at 100 intersections, will hold the tender for the project in the near future. Mayor Mustafa Demir said, "We are excited to present this system, which is seen in the metropolitan cities of our country, to the service of our citizens".”

SCIENTOSEMANTICS OF APPLIED RESEARCH USING AI AND MACHINE LEARNING – A SEQUENTIAL APPROACH

Gustaf Nelhans and Johan Eklund,

Data as Impact Lab

2021-04-19

THE SWEDISH SCHOOL OF LIBRARY AND INFORMATION SCIENCE
UNIVERSITY OF BORÅS

Contents

Introduction - objectives and background

This report presents the results of a combined scientometric and machine learning exercise with the objective to help in the description of relevant research within the two topics of Air Quality research and Traffic safety research in relation to the application of artificial intelligence and machine learning techniques. It was performed by the Data as Impact Lab at the Swedish School of Library and information science, University of Borås.

The assignment consisted of investigating research where AI and advanced data analysis are used to study the respective subject areas of air quality issues and traffic safety. Specifically, we were asked to identify the development of research in the respective areas during recent years. Contexts that were requested related to which universities/areas/countries are active and identifying helpful review articles. There was also an interest in determining which data sources and methodological approaches have been used in the studies. For example, we were supplied with issues such as analysing air quality, distribution of emissions, and exposure to emissions in humans. The data sources could vary significantly from measuring emissions levels in measuring stations to satellite data and mobile phone data to measuring movement patterns.

The first stage of the work related to identifying meaningful terms to use in the searches for relevant research in the publication databases used. Clarivate Web of Science was chosen for the task, one of the most comprehensive publication databases with extensive quality control. It also contains bibliographical references to other scholarly research, which is a prerequisite for doing citation analysis.

Most publication databases employ a classic search interface. It means that data is first retrieved and is then stored for analysis in a sequential manner.[1] We first search for relevant literature using specific terms covering the subject in question and then employ various scientometric methods to investigate the content and metadata of the literature that we find. The client supplied these search terms. They consisted of subject terms such as “air quality”, “atmospheric pollution”, and “particle emission” for the study of Air Quality research. For Traffic Safety research, the terms included “traffic safety”, “road safety”, “safety assessment, and “traffic accidents”. These terms were combined with methodological terms used to describe current AI approaches within research such as “artificial intelligence”, “machine learning”, “deep learning”, and more specified terms as “random forest” and “support vector machine”. In order to transform these terms into Boolean search strings, several approaches need to be performed. It involves combining words into concepts, joining strings of terms within quotes (e.g. “air quality”), combining different concepts using the OR operator and combining them to identify the Union of two sets using the AND operator.

[1] Another approach, using live API access to publication data using the openly available CrossRef database as a source and Open Citations for linking references between the literature was investigated but rejected based on time constraints in the assignment and coverage. Such an exploratory approach would have yielded more emphasis on the acquisition data instead of a focus on analysis. Another issue is that not all publishers deliver citation data to CrossRef or do not allow its use, which means that novelty of data would be substituted for quality control. Though, it would be interesting to develop a live interface to live metadata for research, especially in a field such as AI/machine learning, where development is very quick.

In some cases, instead of using the less strict AND operator to combine terms that need to be close to each other, the NEAR/4 operator was used, meaning that if two terms are found within a four-word radius, they will be retrieved together (e.g. "*air* and water *quality*"). While it is important to yield a large enough data set to work with, no search string is perfect. Therefore, after an initial trial and error period, the effort most often ends with a reasonably wide string to identify as many relevant documents as possible while, at the same time, block off irrelevant terms based on the concept of precision and recall. At some point, introducing new terms does not yield a significant change in the number of hits. This coincides with the search string being "saturated". The individual search strings that were found meaningful for each topic are found at the beginning of each part of the study.

Scientometric analyses

An overview of each subject matter is presented in the following two chapters. The study uses a traditional bibliometric methodology to identify and aggregate bibliographic information from the Web of Science. Apart from Title, abstract and keywords found in each article that is retrieved, we are also able to extract metadata about authors, their affiliations and even funding data for the research (though, the latter is not used in the analysis. Lastly, citation data based on the bibliographic references of each article is retrieved from Web of Science).

Together, these different sources of information can be aggregated using scientometric methodology. Based on the concept of citation analysis, aggregation of data at different entity levels: article, author, source (e.g. the journal wherein the article is published in), or based on organisation data (university, department or country), three different scientometric methods are employed:

1. Co-authorship, meaning that authors or organisations are linked together if their respective entities are linked together based on authorship.
2. Bibliographic coupling, wherein two entities are linked together if they cite the same references as sources for the research, and
3. Co-citation analysis, where two cited sources (documents, authors or source journals) are linked together if the same entity cites them.

Additionally, a text mining approach, called co-word analysis, links relevant noun phrases to each other if found in titles and abstracts in the data set. The text-based analysis of keywords and key terms in the WoS dataset's titles and abstracts is used. Keywords are registered at the article level by the publisher, often chosen by researchers themselves but sometimes chosen from a list of pre-determined keywords. This algorithm considers pair-wise relationships between all keywords identified in the articles citing the institutes' publications based on how often the terms occur together in the "author generated" keyword list.[2] Another way of identifying key terms and phrases uses terms identified in the articles' titles and abstracts. This is a more "free form" of text, and while sometimes noisy, can provide insights in the actual terminology used instead of the more restricted set of keywords. Using VOSviewer, the co-word algorithm filters the text for meaningful noun phrases, including nouns and

[2] As opposed to Keywords PLUS™, which is a set of keywords that is added by Web of Science.

adjectives in front of nouns to identify semantic phrases of relevance, using linguistic techniques.

A note about "exploratory scientometrics"

As opposed to well-known scientometric uses for evaluative purposes, we are not interested in evaluating or ranking research per se but instead exploring the data generated and finding interesting patterns and aspects of the data to investigate further. In this study, most presented data is based on the notion of "exploratory scientometrics". This means that, instead of focusing on the ranking of entities for analysis, we try to convey the relational aspects of the scholarly papers found in the original searches. Be it citations or similar use of terminology, instead of top-10 lists; we try to show who collaborates with whom, the overlap between research interests at one organisation with another, as well as the similarities in terminology found between different levels of analysis. Therefore, the preferred means of exploring the results are from network visualisations of the scientometric data, which has the advantage of conveying much information in a condensed format. It also allows the user to explore the results themselves. Therefore, we only give some hints about interpreting the results and leaving it to the expert reader to convey meaning and conclusions about what is found.

We intend that the following analyses and illustrations of research publications in the field should give options for identifying the research's breadth and depth, as seen through the lens of scientometrics. Moreover, it affords "hypothesis generation" options to explore the vast set of data and find new insights into the work. Therefore, we intend to provide the reader with maps of the landscape and hints at interpreting the results. However, we intend that most of the actual analysis and conceptualisation will be done by the reader.

A final word about the visualisations shown in the report. Since the flat format of a report is somewhat limiting for a detailed analysis of the data, we also provide all visualisations in the report in an online appendix where the graphics are shown in a larger size and pdf format for vectorised versions that could be zoomed into.

Methodology

For the downloaded set of publications from WoS, the following report shows the most relevant data. All data is based on full counts at the document level. No fractionalisation or field normalisation is performed in tabular data. We perform contributor fractionalisation in the visualisations when relevant.

Tabular data is based on Web of Science data. Preparation of data was made using HistCite[3], a legacy software developed by Dr Eugene Garfield (1925-2017). To illustrate the bibliographic data and aggregate it so that more comprehensive information can be elucidated, tabular data is often accompanied by bibliographic visualisations. A software package, VOSviewer[4] (van Eck & Waltman, 2010), was used for most visualisations. It is created by

[3] A legacy version is available for Windows computers: https://support.clarivate.com/ScientificandAcademicResearch/s/article/HistCite-No-longer-in-active-development-or-officially-supported?language=en_US

[4] https://www.vosviewer.com/

researchers at the Centre for Science and Technology Studies at Leiden University. As opposed to generic visualisation software, it has been designed to reed output data files from citation databases such as Clarivate Web of Science and Elsevier Scopus, alleviating the often burdensome handling of these nested data frames. Additional handling of data was performed using R, Python and MS Excel.

Scientometric report, Air Quality Research

(All data were downloaded from Web of Science Core Collection on 2021-01-18)

Timespan: *open time window.* ***Indexes:*** *SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.*

Search string:

TS= ((air OR atmospher OR aerosol* OR particl*) NEAR/4 (qual* OR pollut* OR emiss* OR expos*))*

AND

TS=(("machine learning" OR "deep learning" OR "artific intel*" OR "neural network*" OR "support vector machine*" OR "reinforcement learning" OR "random forest*"))*

Identified documents: *3.166*

Results

Document type and publication year

For the publications identified in Web of Science, we show the number of documents published on a yearly basis (Figure 1). We find the number of articles per year is increasing, especially after 2014. Since no strict criteria for limiting the inclusion were performed, the citation database included documents published in 2021 and a few with an "unknown" date. These are generally preprints without a version of record that have yet to receive a publishing date.

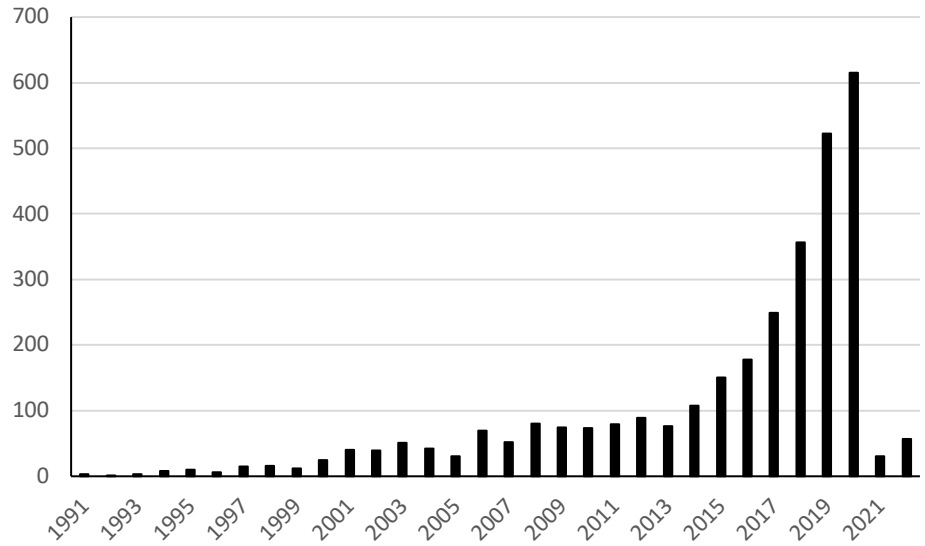


Figure 1. The number of yearly publications since 1991. The last column shows "unknown year, which primarily relates to ahead-of-print publications.

Most documents are of the category (peer reviewed) “article” (Table 1). Quite a large share of the documents are proceedings papers, which is quite common in engineering sciences. There is also a small number, but in terms of citations, quite significant publications of the review kind. Other document types were relatively few. Although some of these are not peer reviewed, it was deemed unnecessary to remove these publications since they might still be relevant for information purposes. As with self-citations, as will be seen later, when using bibliographic data for exploratory tasks, less strict inclusion criteria is often warranted. Apart from publication numbers, two additional columns are shown. The last one, GCS, stands for Global Citation Score and corresponds to the number of citations the documents at a particular year has received up to the time of retrieving data for the study. The middle one, LCS, stand for “Local Citation Score” and shows the number of citations the publications of a particular year has received *within* the data set.

Table 1: Document type

DOCUMENT TYPE	RECS	PERCENT	TLCS	TGCS
Article	2169	68.6	8907	34101
Proceedings Paper	750	23.7	478	2016
Review	92	2.9	188	2316
Article; Proceedings Paper	81	2.6	387	2125
Article; Early Access	55	1.7	0	14
Software Review	6	0.2	30	43
Article; Data Paper	4	0.1	0	15
Meeting Abstract	3	0.1	0	0
Review; Early Access	2	0.1	0	1
Editorial Material	1	0.0	0	10

Authorship

Although the focus is on the research content, it might still be relevant to show some data at the individual level in

Table 2 while then focusing on researchers' co-authorship with other researchers in the included publications identified in WoS (Figure 2).

Table 2: Author level data.

AUTHOR	RECS	TLCS	TGCS
Liu Y	31	131	470
Kumar A	23	180	366
Li Y	21	116	249
Lu WZ	21	261	692
Zhang L	21	20	224
Wang Y	20	16	83
Mlakar P	19	143	275
Perez P	17	340	629
Zhang Y	16	127	321
Li Q	15	177	337
Ma J	15	14	126
Oprea M	15	33	85
Kolehmainen M	14	519	1,088
Liu H	14	28	87
Schwartz J	14	115	391
Wang JZ	14	187	463
Li X	13	135	254
Nieto PJG	13	118	217
Wang ZY	13	76	230

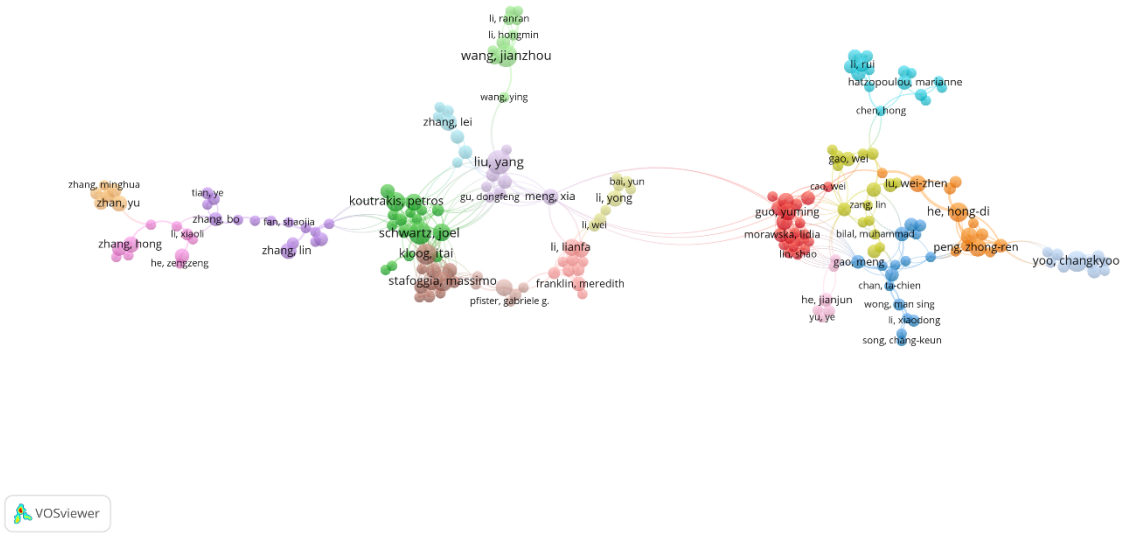


Figure 2: Co-authorship – Authors. Of 10,367 authors, 666 (321) were found ≥3 times. Visualisation: CoAuthAuth.png/pdf

Organisation level

A substantial share of the most prolific universities in the selection consists of Asian and Middle Eastern universities (Table 3). In Figure 3, instead, we see the full breadth of collaboration with organisations with at least three authorship contributions in the data set. Note that there are two visualisations, one using clusters to differentiate potential thematic clusters and one which uses the average publication year for each institution's contribution to the data set. Visualisations do not accompany the subdivision into smaller institutional units in Table 6 and the aggregation at the country level in Table 7.

Table 3 Organisation

#	INSTITUTION	RECS	PERCENT	TLCS	TGCS
1	Chinese Acad Sci	83	2.6	315	1,333
2	Tsinghua Univ	43	1.4	180	629
3	Nanjing Univ Informat Sci & Technol	41	1.3	51	227
4	Wuhan Univ	38	1.2	109	469
5	Peking Univ	36	1.1	244	534
6	City Univ Hong Kong	34	1.1	317	1,028
7	Shanghai Jiao Tong Univ	29	0.9	83	275
8	Univ Chinese Acad Sci	28	0.9	138	280
9	Aristotle Univ Thessaloniki	27	0.9	215	540
10	Sun Yat Sen Univ	26	0.8	92	302
11	Univ Tehran	26	0.8	60	326
12	North China Elect Power Univ	25	0.8	75	237
13	Islamic Azad Univ	24	0.8	60	413
14	Lanzhou Univ	24	0.8	246	537
15	NASA	24	0.8	101	447
16	Zhejiang Univ	24	0.8	126	351
17	Emory Univ	23	0.7	36	141
18	Beijing Univ Technol	22	0.7	28	87
19	Dongbei Univ Finance & Econ	22	0.7	233	610
20	CNR	21	0.7	49	203
21	Indian Inst Technol	21	0.7	119	381

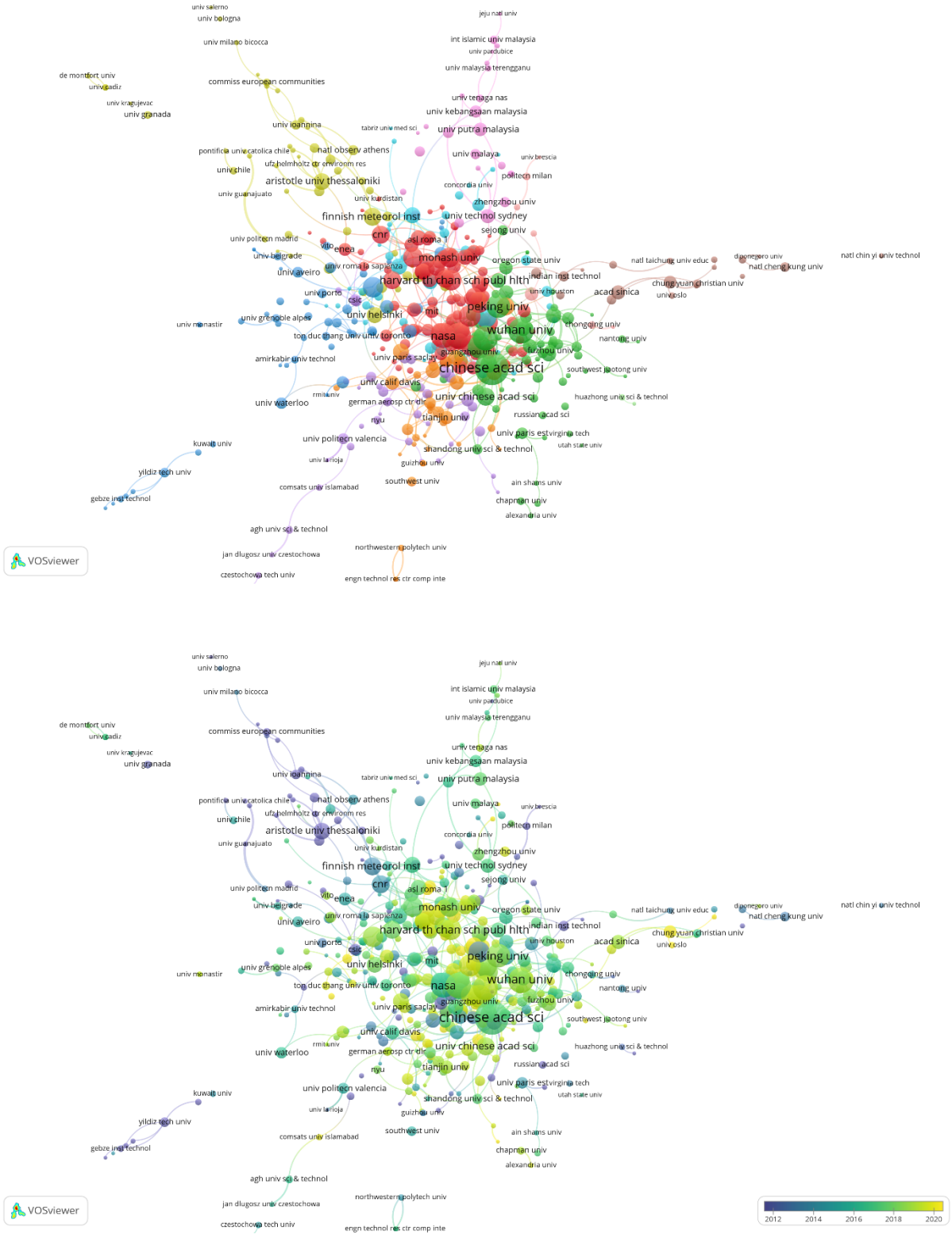


Figure 3: Co-authorship – Organisations. Of 2,952 organisations, 481 (570) were found ≥3 times. Top: Clusters, Bottom: Average Publication year. Visualisation: CoAuthOrg.png/pdf

Table 4: Institution with subdivision

INSTITUTION WITH SUBDIVISION	RECS	PERCENT	TLCS	TGCS
Dongbei Univ Finance & Econ, Sch Stat	22	0.7	233	610
Univ Chinese Acad Sci	21	0.7	138	261
City Univ Hong Kong, Dept Bldg & Construct	19	0.6	271	778
Chinese Acad Sci, Inst Geog Sci & Nat Resources Res	18	0.6	64	238
Emory Univ, Rollins Sch Publ Hlth	17	0.5	26	111
NASA, Goddard Space Flight Ctr	17	0.5	101	376
Univ Santiago Chile, Dept Fis	14	0.4	334	617
Tsinghua Univ, Sch Environm	13	0.4	79	244
Unknown	13	0.4	6	25
Aristotle Univ Thessaloniki, Dept Mech Engn	12	0.4	89	162
Beijing Univ Technol, Fac Informat Technol	12	0.4	11	44
Harvard TH Chan Sch Publ Hlth, Dept Environm Hlth	12	0.4	37	126
Univ Kuopio, Dept Environm Sci	12	0.4	449	966
Wuhan Univ, State Key Lab Informat Engn Surveying M…	12	0.4	69	230
Indian Inst Technol, Dept Civil Engn	11	0.3	70	205
Jozef Stefan Inst	11	0.3	6	55
Lanzhou Univ, Sch Math & Stat	11	0.3	167	342
Univ E Anglia, Sch Environm Sci	11	0.3	523	1,145
Univ Ioannina, Dept Phys	11	0.3	144	358
Wuhan Univ, Sch Resource & Environm Sci	11	0.3	54	182

Table 5: Country

#	COUNTRY	RECS	PERCENT	TLCS	TGCS
1	Peoples R China	899	28.4	2,865	10,872
2	USA	532	16.8	1,569	8,546
3	Italy	193	6.1	865	3,400
4	UK	193	6.1	815	3,714
5	India	179	5.7	509	1,560
6	Spain	165	5.2	430	2,337
7	Iran	126	4.0	236	1,300
8	South Korea	106	3.4	139	819
9	Germany	104	3.3	252	1,365
10	Taiwan	101	3.2	232	1,070
11	Australia	95	3.0	257	1,417
12	Turkey	90	2.8	386	956
13	Canada	89	2.8	228	1,355
14	Greece	87	2.8	733	2,302
15	France	83	2.6	269	1,381
16	Poland	77	2.4	93	625
17	Malaysia	71	2.2	81	602
18	Brazil	59	1.9	118	613
19	Unknown	51	1.6	227	645
20	Romania	46	1.5	43	238
30	*Sweden*	*29*	*0.9*	*41*	*248*

Journal level

When viewed as a top list of journals, we find a broad range of publication outlets in Table 6. Presenting the distribution as a network map of journals as in Figure 4, we can find patterns in the results within the total of 1,409 different publication outlets where the research was published. The bibliographic coupling algorithm is used here, meaning that two journals are closely connected based on the overlap of reference lists in their respective published articles. The node sizes are based on the relative number of published articles within each journal. The first one shows topics as different colours based on the layout algorithm used, while the second graph is colour coded based on the average publication year for each journal within the set.

Table 6: Journal sources

JOURNAL	RECS	PERCENT	TLCS	TGCS
ATMOSPHERIC ENVIRONMENT	147	4.6	2,593	5,999
SCIENCE OF THE TOTAL ENVIRONMENT	83	2.6	845	2,097
IEEE ACCESS	55	1.7	100	284
ENVIRONMENTAL POLLUTION	54	1.7	456	1,269
ATMOSPHERIC POLLUTION RESEARCH	48	1.5	415	791
AIR QUALITY ATMOSPHERE AND HEALTH	42	1.3	153	323
JOURNAL OF CLEANER PRODUCTION	42	1.3	82	415
SENSORS	39	1.2	2	336
BUILDING AND ENVIRONMENT	37	1.2	128	1,018
INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	34	1.1	31	283
ATMOSPHERE	32	1.0	5	161
ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH	32	1.0	168	442
SUSTAINABILITY	32	1.0	0	137
JOURNAL OF THE AIR & WASTE MANAGEMENT ASSOCIATION	30	0.9	385	803
APPLIED SCIENCES-BASEL	29	0.9	0	115
ENVIRONMENTAL SCIENCE & TECHNOLOGY	29	0.9	192	691
ENVIRONMENTAL MODELLING & SOFTWARE	28	0.9	422	1,669
REMOTE SENSING	26	0.8	0	135
ENVIRONMENT INTERNATIONAL	24	0.8	157	561
CHEMOSPHERE	23	0.7	182	853

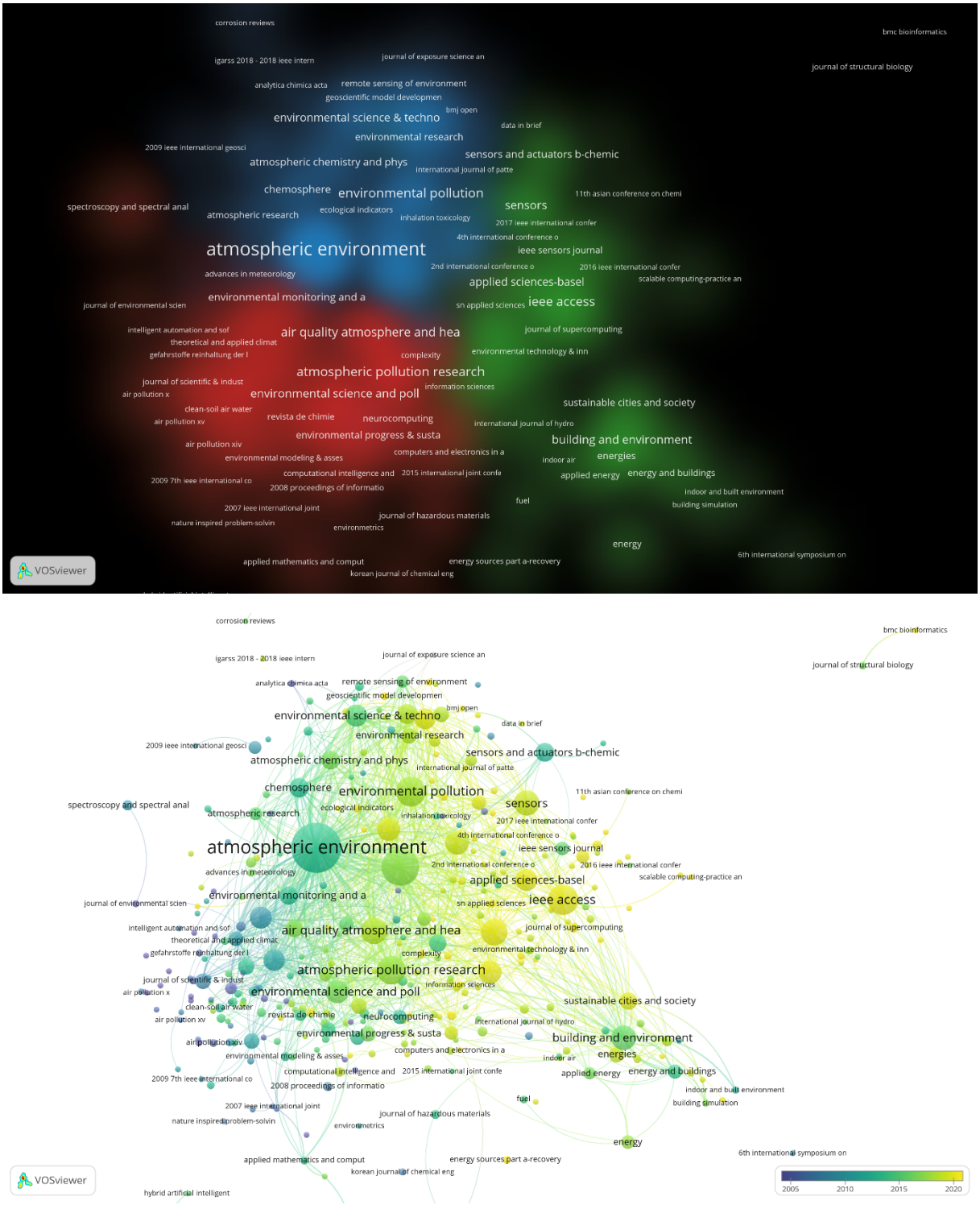


Figure 4: Bibliographic coupling – Sources. Of 1,409 sources, 334 (330) had ≥2 documents. Top: Density view, Bottom: Nodes clustered based on average publication year. Visualisation: BiblCoupSO.png/pdf; BiblCoupSOPY.png/pdf

Co-citation analysis

Lastly, we show the most cited documents (Figure 5) and all source outlets (Figure 6) that the researchers in the dataset cite within the publications. The Co-citation network shows the "intellectual basis" of the collective (Persson, 1994). Highly cited articles and source outlets contain the most frequently used research among the collected data. The articles cluster around specific topics, as can be seen through the titles of the sources. Both the green cluster on top and the red one to the right mainly consists of articles in the journal with the shortened title atmos environ. However, since these clusters are separated, these articles seem to cover different topics. Specifically, it can be seen that there is a temporal pattern, where the red cluster consists of articles published in the first decade of the 21st century.

In contrast, the green one consists of articles from the second decade. The purple cluster at the bottom consists of articles focusing on sensors, while the blue to the right seems to consist of interdisciplinary literature in engineering and environmental science. The yellow cluster, which divides the cited literature into two parts, seems to consist of a more fundamental kind of machine learning literature. This research is applied in the research found in the more distant clusters. Some of the details are lost when the data is aggregated at the journal level. It is because one journal stands out so much in terms of the numbers of citations. However, we find that a significant share of the cited literature is published in traditional disciplines, whether geophysics, remote sensing or the built environment.

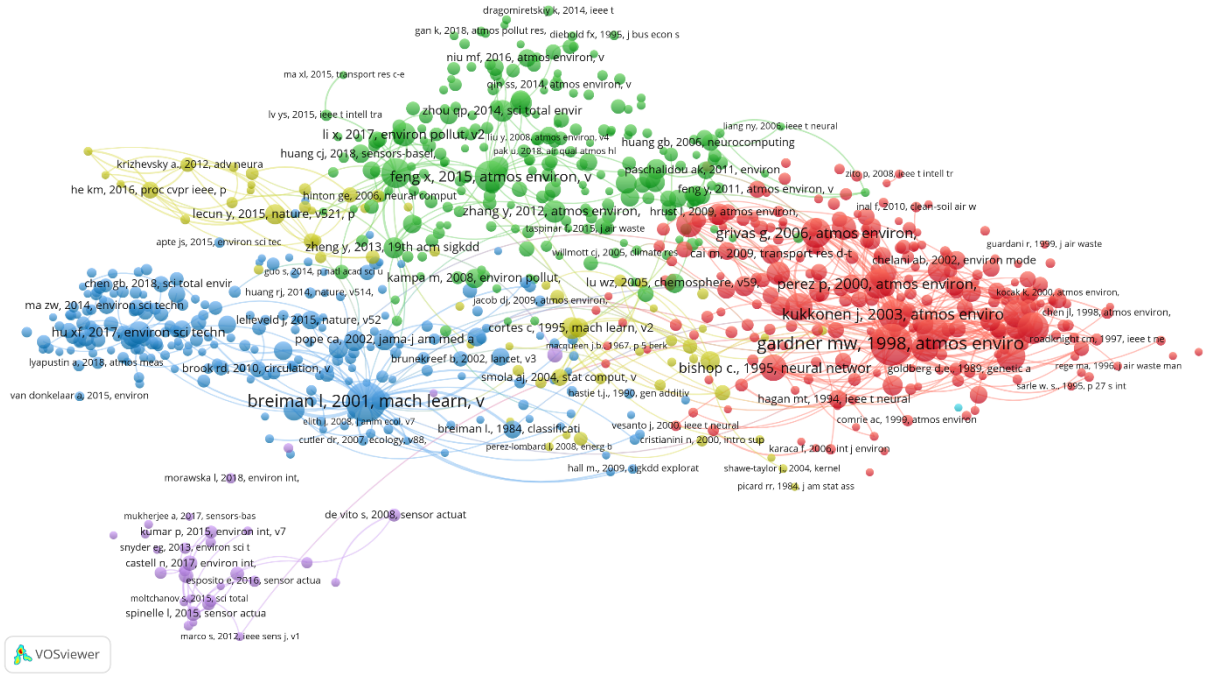


Figure 5: Co-citation – Documents. Of 81,062 sources, 818 had ≥10 citations. Visualization: CoCitDO.png/pdf

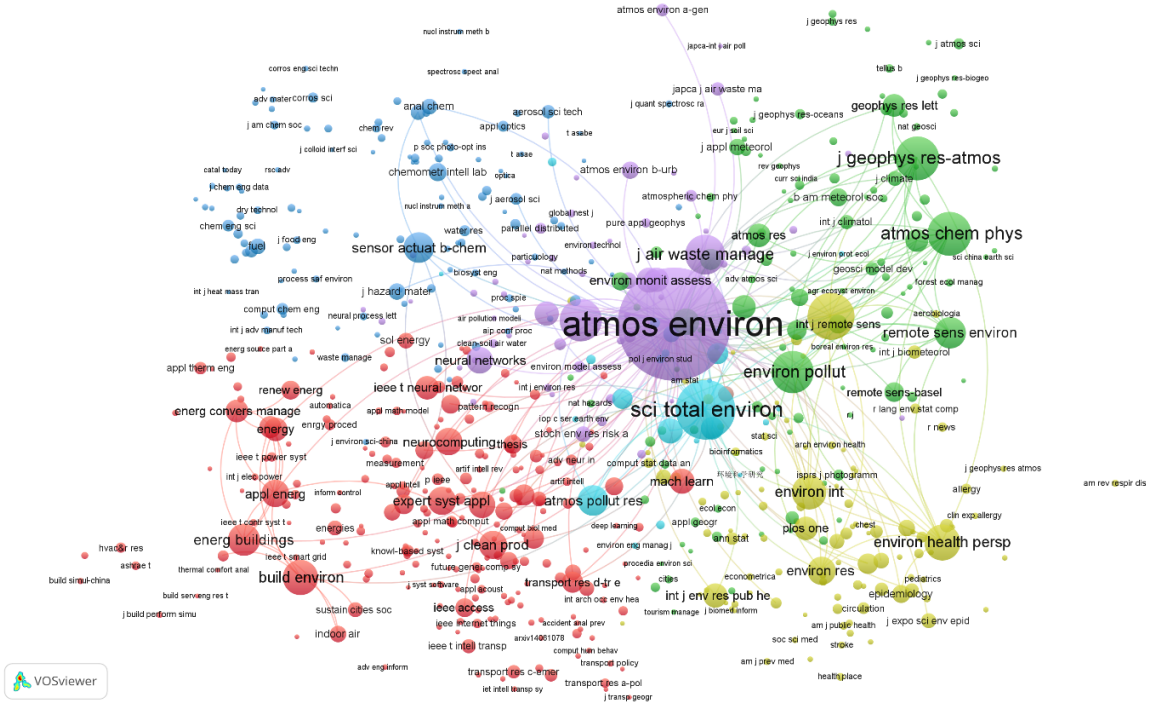


Figure 6: Co-citation – Sources. Of 25,644 sources, 644 had ≥20 citations. Visualization: CoCitSO.png/pdf

Keywords and co-word analysis

Lastly, we employ two techniques to identify thematic information about the research content covered by the articles identified in our searches. We select the most frequently used author keywords chosen for the publications (Figure 7, Figure 8). Another technique, co-word analysis, extract nouns and so-called noun phrases, using computer linguistic methods to identify phrases of text could sometimes elucidate more specific information (Figure 9).

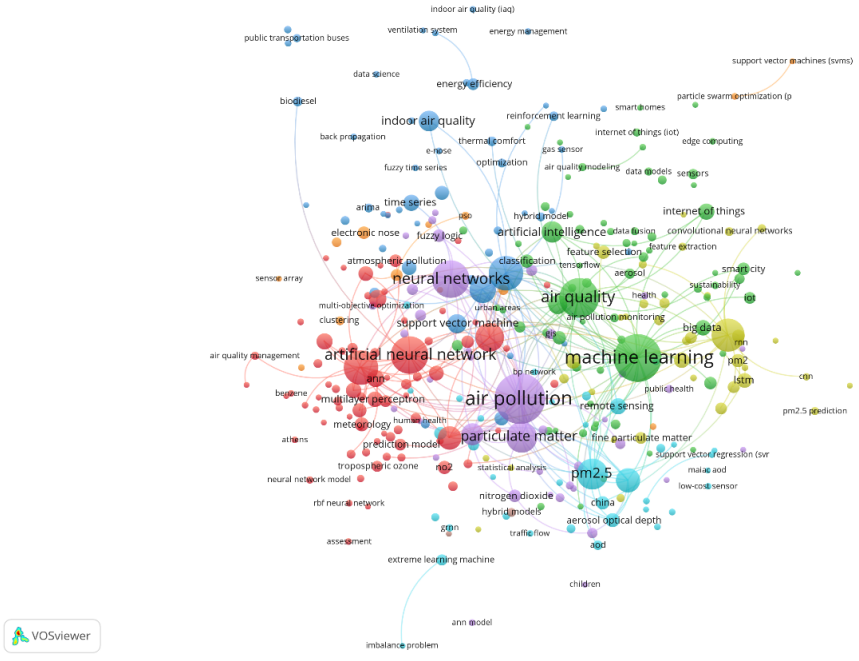


Figure 7. Author keywords. Of 6,517 keywords, 323 (323) were found ≥5 times. Visualisation: keywords.png/pdf

We add some details of the map to show that different machine learning algorithms seem to be related to specific research areas. Colours show which topic cluster each specific keyword belongs to:

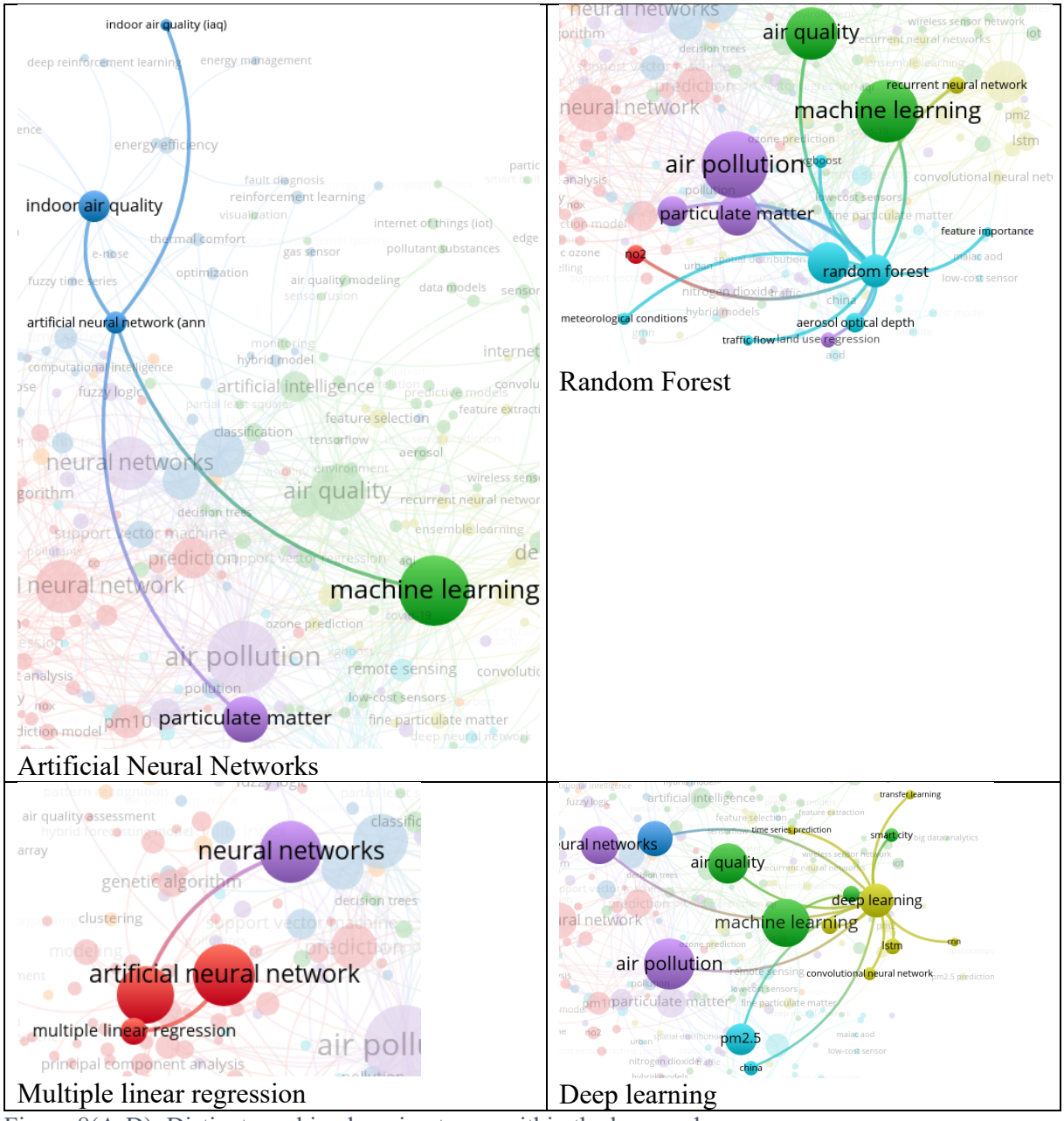


Figure 8(A-D): Distinct machine learning terms within the keyword co-occurrence map.

In the co-word analysis of relevant terms identified in the titles and abstracts of the retrieved documents, we identify terms and phrases relevant to different domains of analysis and specific topics. In a sense, it is possible to identify topics relating to research about air quality issues and AI methodology. The following description is tentative but shows one way of ascribing a "story" to the terms and phrases found in the co-word map. In the blue cluster, we find terms relating to forecasting and prediction technology. The green cluster covers particle matter, while the yellow cluster at the bottom depicts terms related to remote sensing approaches and epidemiological studies. The red cluster holds terms relating to instruments and identifies air quality issues in more general terms, while the purple cluster relates to indoor quality.

Figure 9: Co-word analysis. Of 64,048 noun phrases, 843 (506, TF-IDF=60%) where found ≥20 times. Visualization: cowords.png/pdf

Scientometric report, Traffic safety research

(All data were downloaded from Web of Science Core Collection on 2021-02-19)

Timespan: *open time window.* ***Indexes:*** *SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.*

Search string:

TS= (traffic OR transportation* OR road*)

AND

TS= (assessm* OR safety OR crash* OR accident* or collision*)

AND

TS=(("machine learning" OR "deep learning" OR "artific* intel*" OR "neural network*" OR "support vector machine*" OR "reinforcement learning" OR "random forest*"))

Identified documents: *4.171*

Results

Document type and publication year

For the publications identified in Web of Science, we show the number of documents published on a yearly basis (Figure 10). We find the number of articles per year is increasing, especially since 2014. Again, since no strict criteria for limiting the inclusion were performed, the citation database included some documents published in 2021 and a few with an "unknown" date. The latter are generally preprints that have yet to receive a publishing date.

Two additional columns are shown. The last one, GCS, stands for Global Citation Score and corresponds to the number of citations the documents at a specific year has received to the time of extraction of data. The middle one, LCS, stand for "Local Citation Score" and shows the number of citations the publications of a particular year has received within the data set.

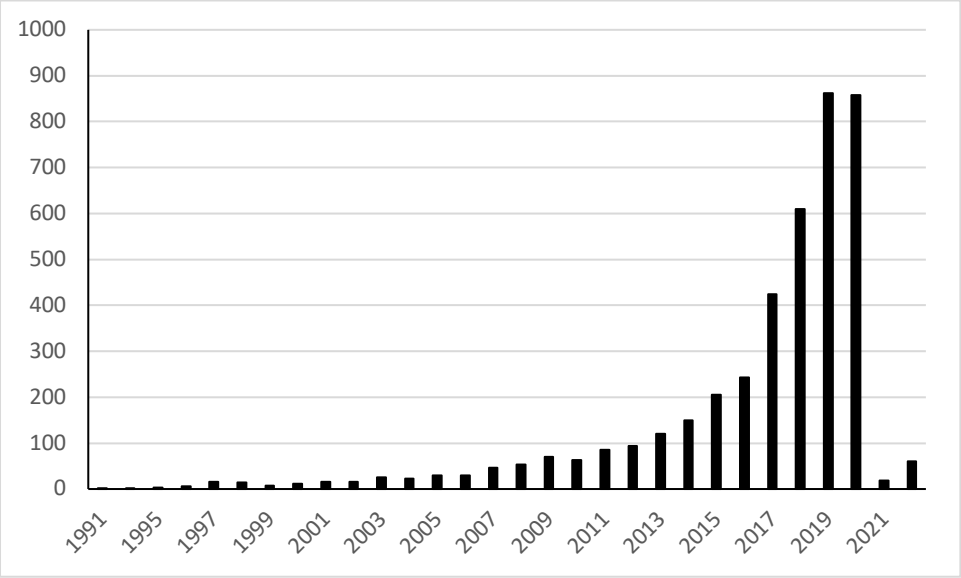


Figure 10. The number of yearly publications since 1991. The last column shows "unknown year, which primarily relates to ahead-of-print publications.

Most documents are of the category (peer reviewed) "article". A pretty large share of the documents is proceedings papers, which is quite common in engineering sciences. There is also a small number, but in terms of citations, quite significant publications of the review kind. Other document types were relatively few. Although some of these are not peer reviewed, it was deemed unnecessary to remove these publications since they might still be relevant for information purposes. As with self-citations, as will be seen later, when using bibliographic data for exploratory tasks, less strict inclusion criteria is often warranted.

Table 7: Document type

DOCUMENT TYPE	RECS	PERCENT	TLCS	TGCS
Article	2,377	57.1	3,533	37,683
Proceedings Paper	1,576	37.9	470	4,852
Review	81	1.9	68	1,920
Article; Proceedings Paper	54	1.3	142	851
Article; Early Access	53	1.3	0	22
Review; Early Access	7	0.2	0	6
Editorial Material	4	0.1	0	3
Article; Data Paper	2	0.0	0	0
Letter	2	0.0	0	0
Meeting Abstract	2	0.0	0	0
Article; Book Chapter	1	0.0	3	78
Proceedings Paper; Retracted Publication	1	0.0	0	2
Review; Book Chapter	1	0.0	0	12
Software Review	1	0.0	0	11

Authorship

Although the focus is on the research content, it might still be relevant to show some data at the individual level in

Table 2 while then focusing on the co-authorship of researchers with other researchers in the included publications identified in WoS (Figure 2).

Table 8: Author level data.

AUTHOR	RECS	TLCS	TGCS
Pradhan B	42	398	3,498
Abdel-Aty M	38	253	1,047
Wang C	33	29	248
Bui DT	29	135	1,875
Liu Y	26	3	96
Wang Y	26	24	170
Pourghasemi HR	25	251	2,151
Li J	23	6	146
Chen W	22	134	1,245
Li Y	18	5	47
Wang H	17	15	127
Hong HY	16	109	984
Liu J	16	3	87
Liu P	16	107	372
Wang W	16	89	260
Zhang Y	16	6	119

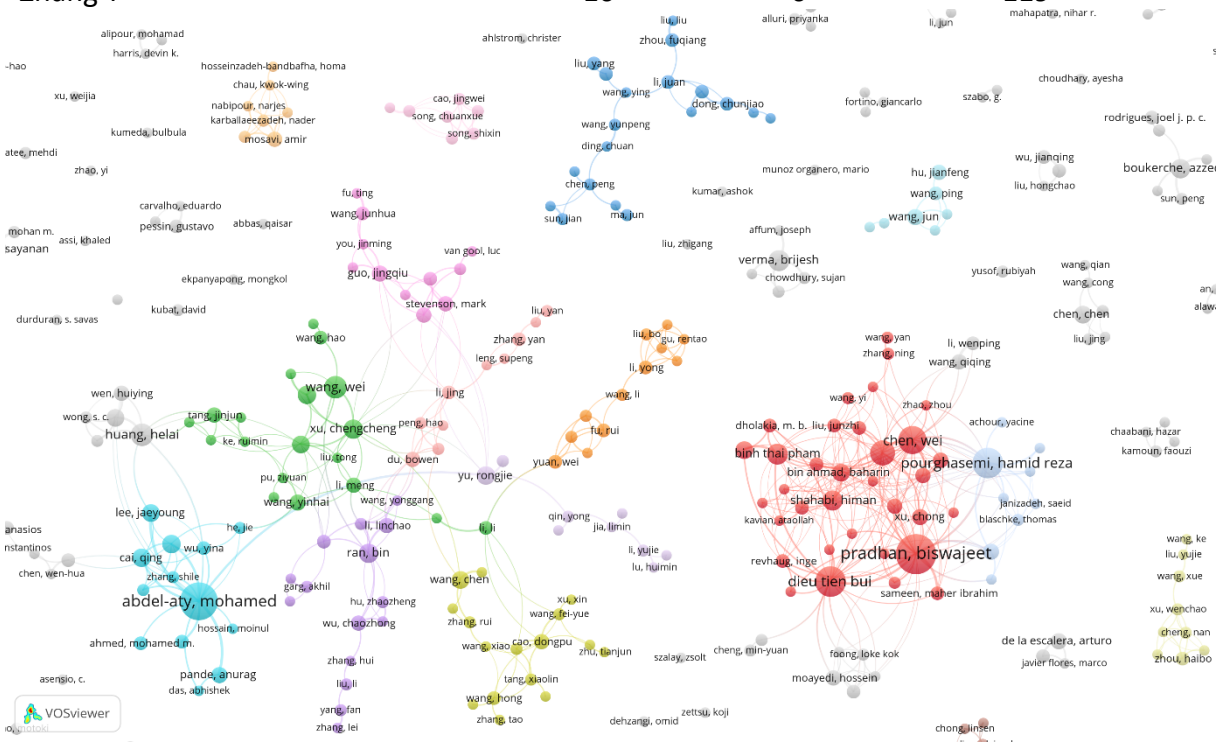


Figure 11: Co-authorship – Authors. Of 13,069 authors, 525 (321) were found ≥3 times. Visualization: CoAuthAuth.png/pdf

Organisation level

A large share of the most prolific universities in the selection consists of Asian and Middle Eastern universities. However, as opposed to the Air Quality research set, several US, British and Australian universities also occur. (Table 9). In Figure 12Figure 3, instead, we see the full breadth of collaboration with organisations with at least three authorship contributions in the data set. Note that there are two visualisations, one using clusters to differentiate potential thematic clusters and one which uses the average publication year for each institution's contribution to the data set. The distribution is very heterogeneous, and there are no obvious co-authorship patterns at any level of aggregation.

Regarding the average year of publication, there does not seem to develop new clusters of actors over time. Instead, the average year of publication seems to be in the middle of the range, indicating that the organisations have distributed output coverage over the years. The subdivision into smaller institutional units in Table 10 and the aggregation at the country level in Table 11 are not accompanied by visualisations.

Table 9: Organisation

INSTITUTION	RECS	PERCENT	TLCS	TGCS
Southeast Univ	69	1.7	160	573
Chinese Acad Sci	61	1.5	113	1,840
Beijing Jiaotong Univ	60	1.4	31	342
Tongji Univ	54	1.3	106	573
Tsinghua Univ	54	1.3	74	559
Univ Cent Florida	53	1.3	413	1,614
Beihang Univ	47	1.1	57	517
Changan Univ	47	1.1	33	243
Univ Technol Sydney	36	0.9	50	695
Univ Waterloo	35	0.8	22	543
Jilin Univ	32	0.8	5	231
Wuhan Univ Technol	32	0.8	13	111
Islamic Azad Univ	30	0.7	73	751
Univ Michigan	29	0.7	32	432
Duy Tan Univ	26	0.6	10	274
Southwest Jiaotong Univ	26	0.6	57	399
Hong Kong Polytech Univ	25	0.6	11	196
MIT	25	0.6	40	461
Nanyang Technol Univ	25	0.6	8	116
Univ Illinois	25	0.6	14	225
Univ Putra Malaysia	25	0.6	307	2,528
Univ Tehran	25	0.6	26	717

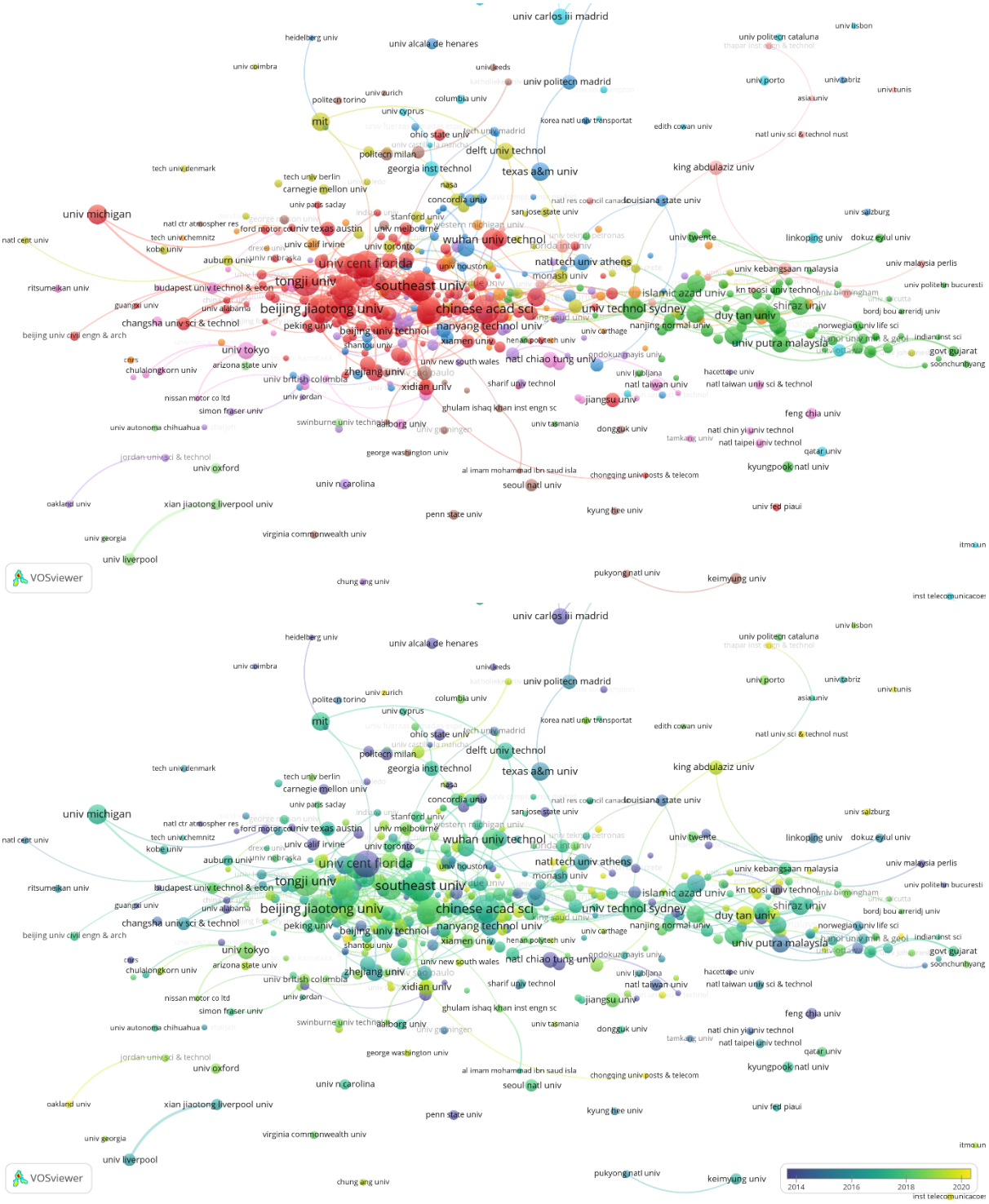


Figure 12 Co-authorship – Organisations. Of 3,431 organisations, 643 (564) were found ≥3 times. Top: Clusters, Bottom: Average Publication year. Visualisation: CoAuthOrg.png/pdf

Table 10: Institution with subdivision

INSTITUTION WITH SUBDIVISION	RECS	PRCN	TLCS	TGCS
Univ Cent Florida, Dept Civil Environm & Construct Engn	37	0.9	234	835
Southeast Univ, Sch Transportat	28	0.7	132	366
Duy Tan Univ, Inst Res & Dev	24	0.6	10	244
Unknown	19	0.5	9	112
Xian Univ Sci & Technol, Coll Geol & Environm	18	0.4	109	1,034
Beijing Jiaotong Univ, Sch Traff & Transportat	17	0.4	10	63
Sejong Univ, Dept Energy & Mineral Resources Engn	15	0.4	44	510
Univ Putra Malaysia, Fac Engn	15	0.4	232	1,671
South China Univ Technol, Sch Civil Engn & Transportat	14	0.3	18	168
Cent South Univ, Sch Traff & Transportat Engn	13	0.3	47	199
Shiraz Univ, Coll Agr	13	0.3	45	589
Beijing Jiaotong Univ, State Key Lab Rail Traff Control & Safety	12	0.3	3	163
Tongji Univ, Key Lab Rd & Traff Engn	12	0.3	26	94
Tsinghua Univ, State Key Lab Automot Safety & Energy	12	0.3	4	70
Univ Chinese Acad Sci	12	0.3	0	23
Southeast Univ, Jiangsu Key Lab Urban ITS	11	0.3	26	122
Univ Technol Sydney, Fac Engn & IT	11	0.3	41	373
Chinese Acad Sci, Inst Automat	10	0.2	43	1,002
Hong Kong Polytech Univ, Dept Comp	10	0.2	5	29
Jilin Univ, State Key Lab Automot Simulat & Control	10	0.2	0	45
Ton Duc Thang Univ, Dept Management Sci & Technol Dev	10	0.2	10	147
Univ Cent Florida, Dept Civil & Environm Engn	10	0.2	179	570
Univ Kurdistan, Fac Nat Resources	10	0.2	51	699
Univ Washington, Dept Civil & Environm Engn	10	0.2	37	138

Table 11: Country level

COUNTRY	RECS	PERCENT	TLCS	TGCS
Peoples R China	1,194	28.7	1,132	10,806
USA	898	21.6	1,361	12,565
India	256	6.2	189	2,027
South Korea	197	4.7	270	3,040
Canada	166	4.0	73	1,513
Australia	161	3.9	105	1,800
UK	159	3.8	133	3,853
Germany	158	3.8	114	1,654
Japan	158	3.8	149	1,443
Iran	150	3.6	445	4,564
Spain	138	3.3	78	1,371
Taiwan	120	2.9	78	1,202
Italy	119	2.9	117	1,599
France	104	2.5	56	1,125
Turkey	101	2.4	302	1,927
Malaysia	97	2.3	429	3,825
Saudi Arabia	66	1.6	105	588
Netherlands	65	1.6	43	1,419
Brazil	61	1.5	15	308
Singapore	58	1.4	66	652

Journal level

We find a broad range of publication outlets at the journal level in Table 6, while the breadth is drastically extended when we see all 2,023 journals where the research was published in Figure 4. The bibliographic coupling algorithm is used here, meaning that two journals are closely connected based on the overlap of reference lists in their respective published articles. There are three distinct clusters, containing research in transportation and accident analyses (top), environmental sciences and the geosciences (right), and intelligent transport systems research sensors and engineering, found to a large degree in conference publications (bottom). A less distinct cluster, focusing on infrastructures, structures and the built environment (yellow), is also found.

Table 12: Journal sources

JOURNAL	RECS	PERCENT	TLCS	TGCS
IEEE ACCESS	152	3.7	71	434
IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS	96	2.3	272	3,575
SENSORS	95	2.3	62	925
ACCIDENT ANALYSIS AND PREVENTION	92	2.2	788	2396
TRANSPORTATION RESEARCH RECORD	71	1.7	91	525
APPLIED SCIENCES-BASEL	49	1.2	0	249
TRANSPORTATION RESEARCH PART C-EMERGING TECHNOLOGIES	47	1.1	153	1,231
IET INTELLIGENT TRANSPORT SYSTEMS	40	1.0	57	287
JOURNAL OF ADVANCED TRANSPORTATION	34	0.8	8	166
EXPERT SYSTEMS WITH APPLICATIONS	32	0.8	88	1,085
SUSTAINABILITY	30	0.7	0	71
REMOTE SENSING	29	0.7	0	204
ENVIRONMENTAL EARTH SCIENCES	25	0.6	62	1,059
IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY	24	0.6	36	575
NEURAL COMPUTING & APPLICATIONS	24	0.6	21	212
2019 IEEE INTELLIGENT TRANSPORTATION SYSTEMS CONFERENCE (ITSC)	23	0.6	1	6
2018 21ST INTERNATIONAL CONFERENCE ON INTELLIGENT TRANSPORTATION SYSTEMS (ITSC)	22	0.5	12	59
INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS	20	0.5	2	11
COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING	19	0.5	38	909
ELECTRONICS	18	0.4	0	37
MATHEMATICAL PROBLEMS IN ENGINEERING	18	0.4	0	254
SAFETY SCIENCE	17	0.4	39	299
INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH	15	0.4	0	76
ISPRS INTERNATIONAL JOURNAL OF GEO-INFORMATION	15	0.4	0	79
MULTIMEDIA TOOLS AND APPLICATIONS	15	0.4	6	65

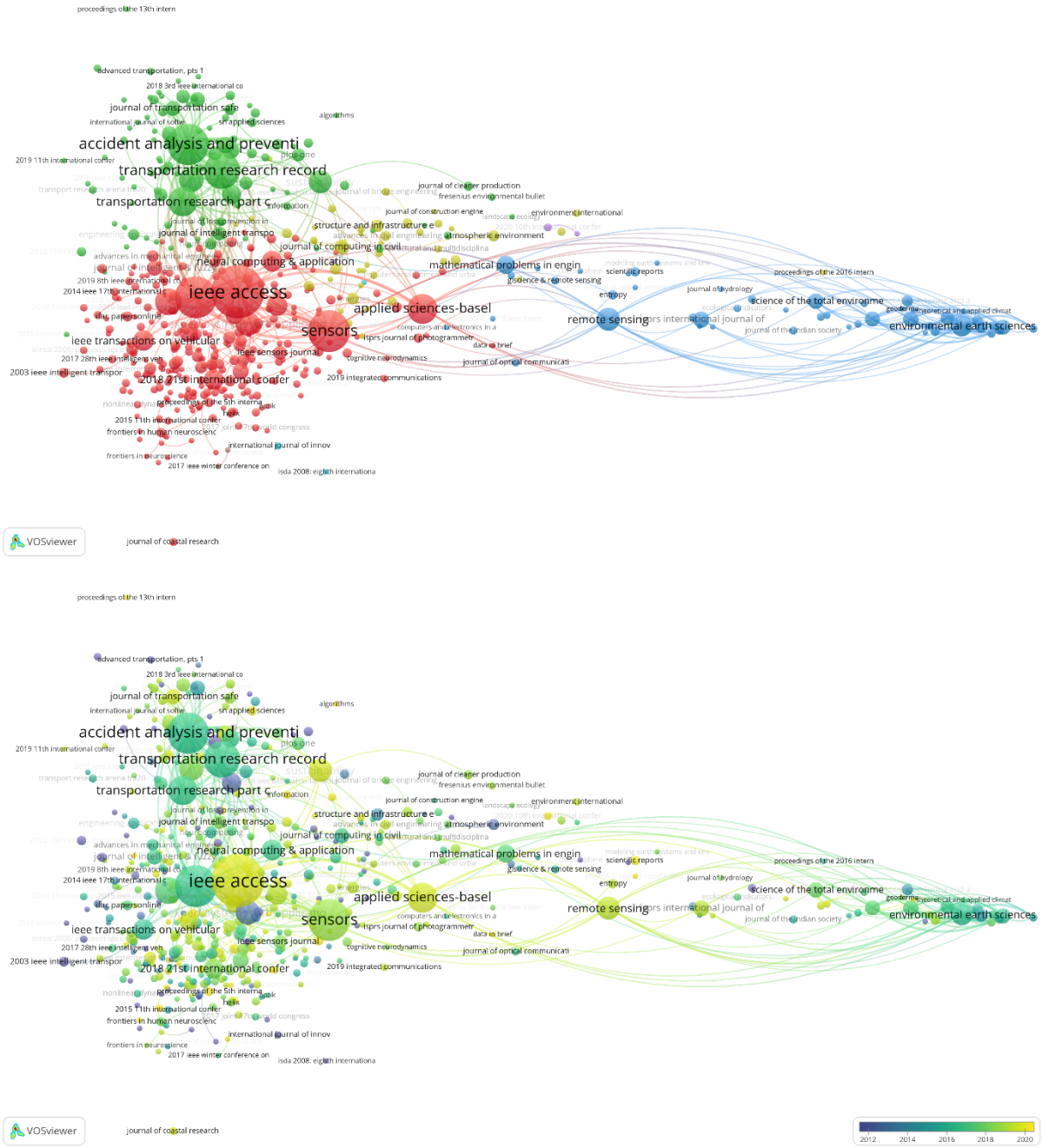


Figure 13: Bibliographic coupling – Sources. Of 2,023 sources, 504 (494) had ≥2 documents. Top: Density view, Bottom: Nodes clustered based on average publication year. Visualisation: BiblCoupSO.png/pdf; BiblCoupSOPY.png/pdf

Co-citation analysis

Lastly, we show the most cited documents (Figure 14Figure 5) and all source outlets (Figure 15) that the researchers in the dataset cite within the publications. The Co-citation network shows the "intellectual basis" of the collective (Persson, 1994). Highly cited articles and source outlets contain the most frequently used research among the collected data. The articles cluster around specific topics, as can be seen through the titles of the sources. Interestingly enough, some of the same aspects found in the bibliographic coupling analysis of the above are also seen here. However, at the centre, the blue cluster of important machine

learning literature used as the source of all research regardless of thematic focus has gained prominence. It is reasonably expected that there is a shared basic science cluster of machine learning technology and statistics literature shared among all research. However, instead of three distinct clusters as in the bibliographic coupling above, we find four distinct clusters surrounding the blue machine learning centre. Again, transportation and accident analyses are at the top, while environmental sciences and geosciences are found to the right. There is still an engineering cluster at the bottom, but there are many publications in medical engineering, neurosciences, and even "sleep research".

When aggregated at the journal level, we find the same issue as in the Air quality research report, that some of the details are lost since one journal stands out so much in terms of numbers of citations. However, we find that a significant share of the cited literature is published in traditional disciplines, whether geophysics, remote sensing or the built environment.

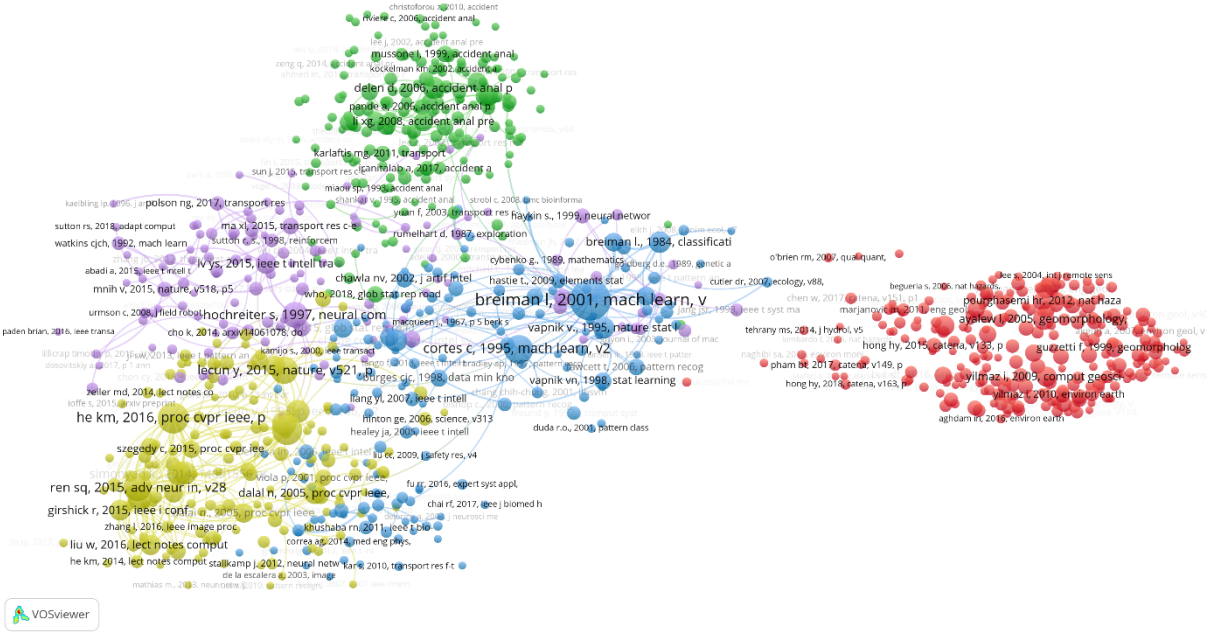


Figure 14: Co-citation – Documents. Of 104,582 sources, 887 had ≥10 citations. Visualization: CoCitDO.png/pdf

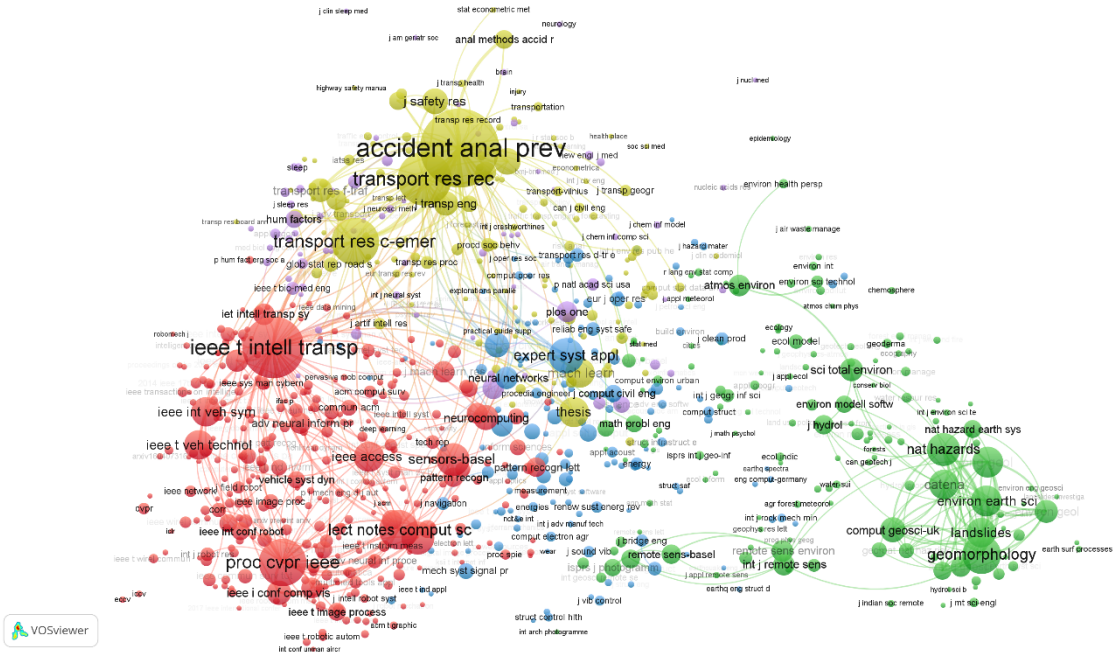


Figure 15: Co-citation – Sources. Of 38,867 sources, 806 had ≥20 citations. Visualization: CoCitSO.png/pdf

Keywords and co-word analysis

Lastly, we employ two techniques to identify thematic information about the research content covered by the articles identified in our searches. We select the most frequently used author keywords chosen for the publications (Figure 16, Figure 17). Another technique, co-word analysis, extracts nouns and so-called noun phrases, using computer linguistic methods to identify phrases of text could sometimes elucidate more specific information (Figure 18).

In the keyword analysis, specific machine learning technologies dominate the respective clusters. ‘Deep learning’ and ‘convolutional neural networks’ are associated with image analysis and computer vision in the blue cluster. In the green cluster, the generic term artificial intelligence and ‘reinforcement learning’, a much more specific term, are found in conjunction with autonomous vehicles. The red cluster, dominated by machine learning, is quite generic, although driver behaviour and physiological factors, including EEG, are connected to the support vector machines at the top. Lastly, although this is a cursory analysis, it seems like geographic information systems and land use concepts are found further away from some of the buzz words in machine learning. The most distinct concept is the more classic machine learning algorithm: ‘logistic regression’. Possibly, we have yet to see faster developments within this area in the future?

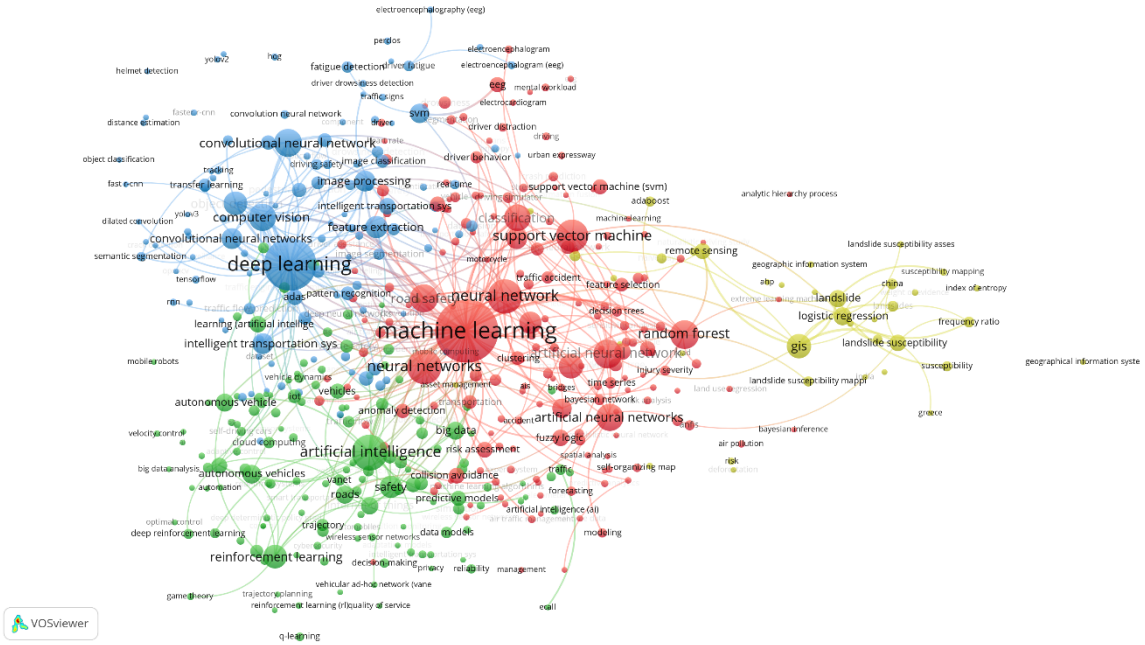
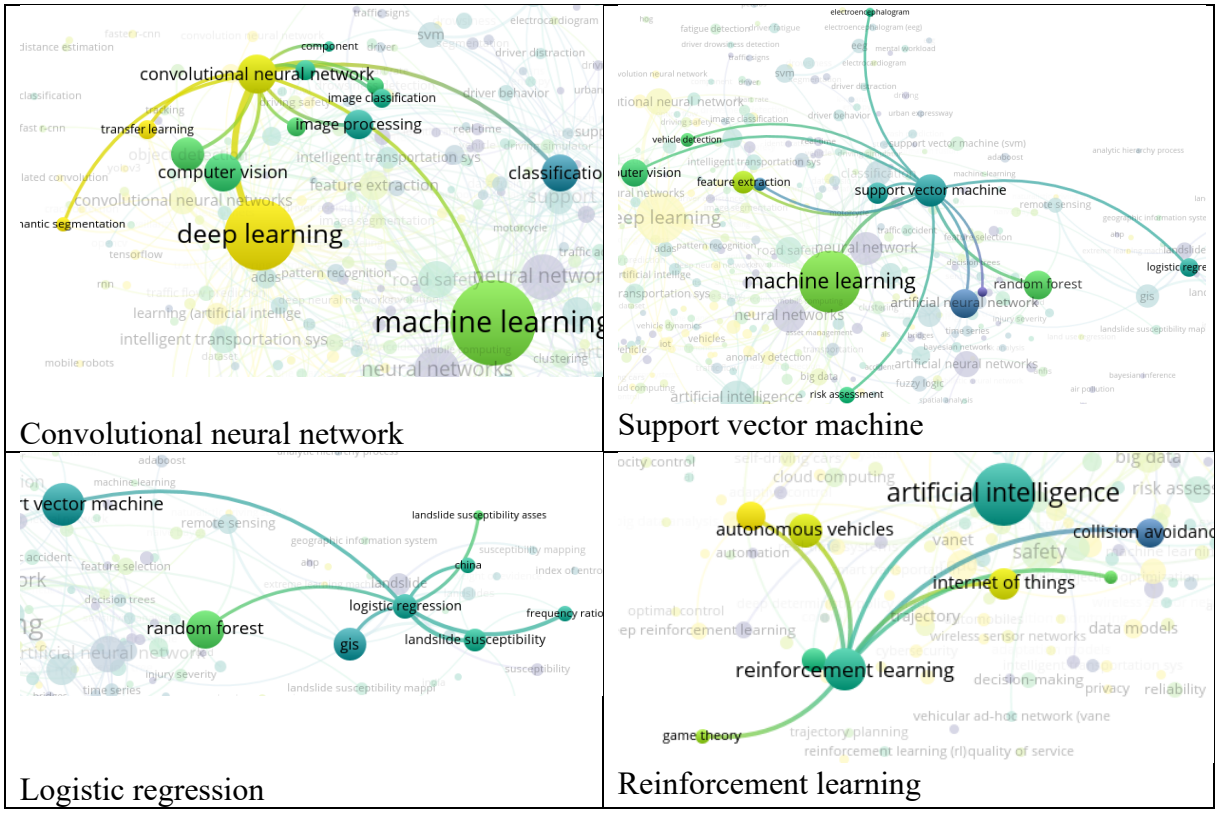
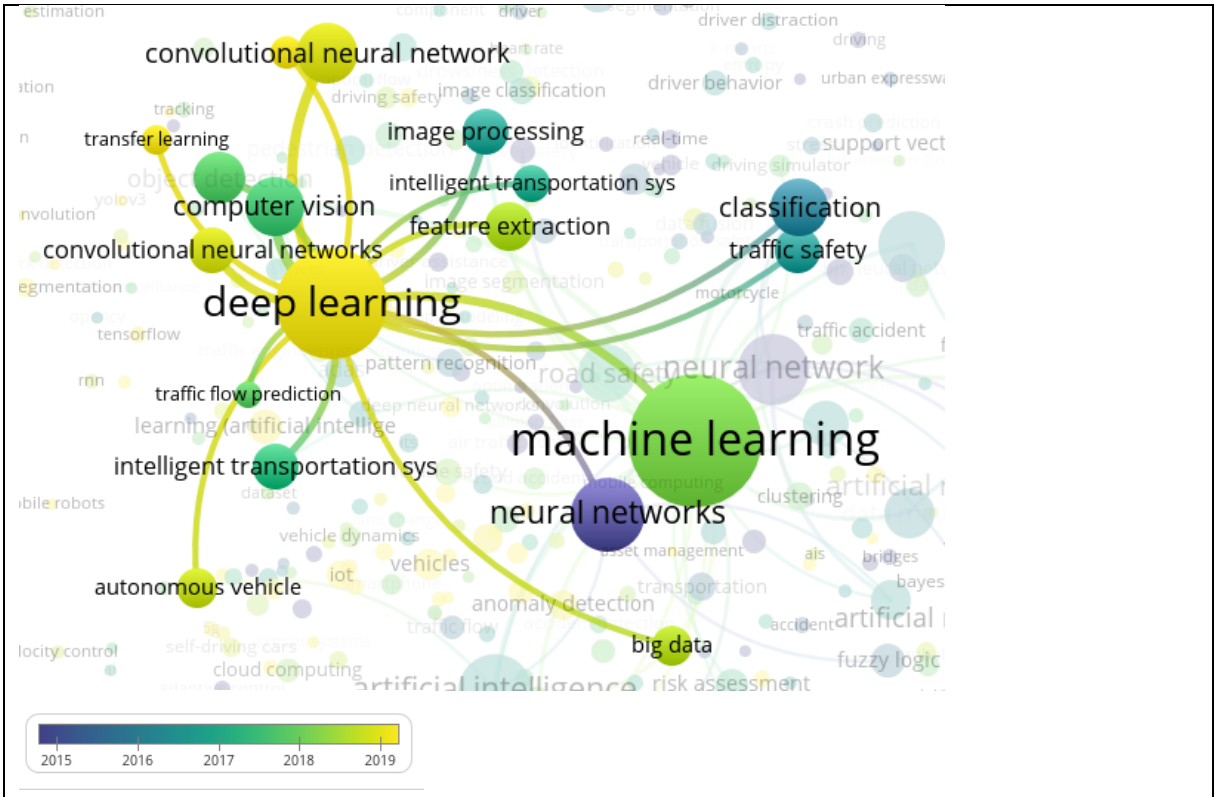


Figure 16. Author keywords. Of 9,979 keywords, 474 (473) were found ≥5 times. Visualisation: keywords.png/pdf

Below we add some details of the keyword map to show that different machine learning algorithms seem to be related to specific research areas, as noted above. Instead of Topic clustering, node colours show the average year of publication for articles using the specific keyword. Blue is earlier, yellow is more recent:





Deep learning + Time scaleScale

Figure 17(A-E): Distinct machine learning terms within the keyword co-occurrence map.

This study's last visualisation is based on co-word analysis of relevant terms identified in the retrieved documents' titles and abstracts. Here we identify terms and phrases relevant to different domains of analysis and specific topics. In a sense, it is possible to identify topics relating to research about air quality issues and AI methodology. The following description is tentative but shows one way of ascribing a "story" to the terms and phrases found in the co-word map. Here, four clusters are distinguished. The blue cluster contains phrases related to the driver's activity, including different detection systems, vision and psychological features like drowsiness. Instead, the green cluster focuses on the vehicle, autonomous systems and the technology of simulation and operation of the vehicle, and analyses of traffic flow, congestions, et cetera. The yellow cluster relates to features such as crashes and risk factors. In contrast, the green clusters cover the geographical settings of roads, GIS and mapping, and on the far right, the geological and geomorphological conditions.

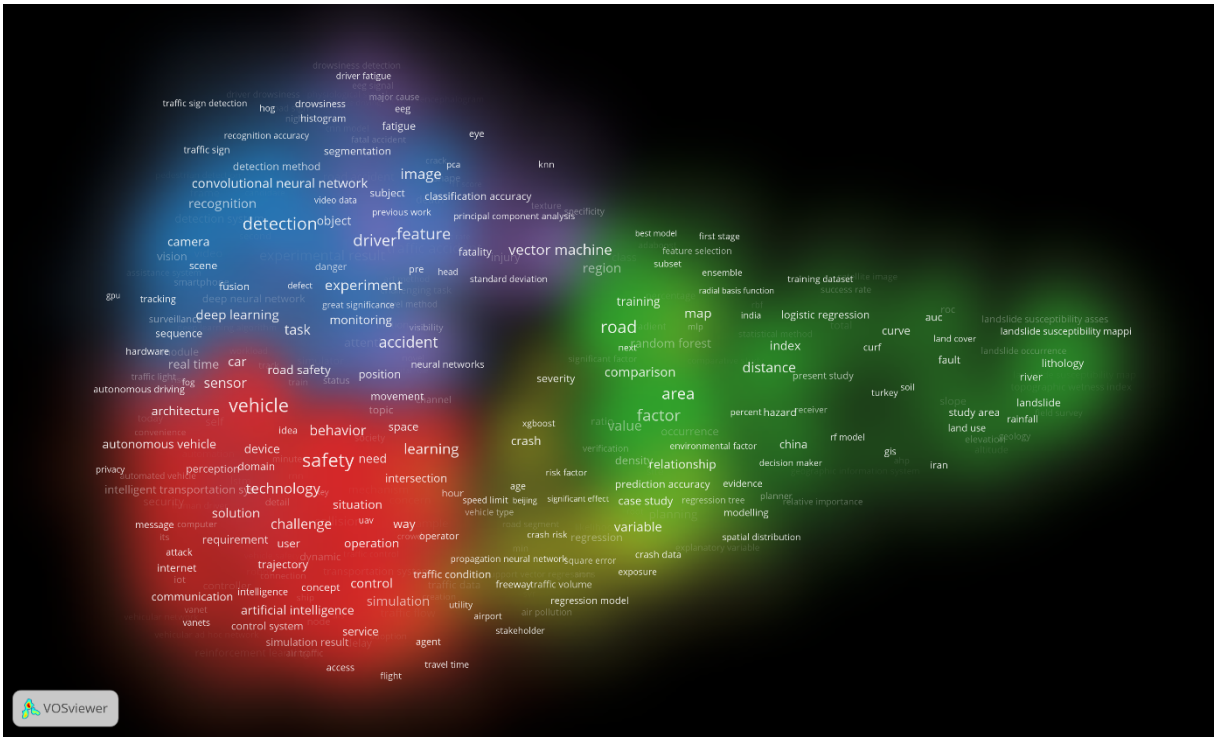


Figure 18. Co-word analysis. Of 74,165 noun phrases, 1,028 (617, TF-IDF=60%) where found ≥20 times. Visualization: keywords.png/pdf

Discussion

In the two scientometric chapters above, we have given an overview of the literature of the respective subject areas covered in the study. Using various methods, such as tabular data, graphs, and specifically scientometric visualisations of citation-based aggregated variables, together with co-occurrences of text, including keywords, and co-word analysis, we have made it possible to identify distinct features of the research. In a sense, this information provides a "distant reading approach" (Moretti, 2013) of the content of more than 3.000 articles on AI and air quality research and 4.000+ articles on traffic safety. From these analyses, we have gained insight into who is researching the area, where this research is published, and the outlets where the research is published.

Additionally, it has been possible to learn about the research content based on the co-occurrence text-based approaches above. However, scientometrics also has its limitations. These limitations include what is covered in the citation databases, substituting details for giving a more comprehensive picture, sometimes hiding the particulars of the content. Though, with the use of exploratory methodology and visualisations and the multitude of interpretations these afford, we try to bring tools to the reader to develop their own reading of the material. A way to overcome these limitations is described in the following chapter.

A scientosemantic approach to bibliographic data

This section describes an in-depth exploratory use of the bibliographic data retrieved from searches performed in Web of Science. We label this a *scientosemantic* approach to bibliographic data since we use probabilistic term weighting, computational linguistics, and machine learning to explore the content of titles and abstracts in bibliographic records. This approach complements the scientometric analyses described in the previous chapters by using results from these to investigate the content of the articles, here based on the text found in the abstracts of the articles.

To this end, we have developed tools for question answering and information searching within the retrieved data. These tools help the client identify specific methods and techniques used to measure air quality, the specific use of satellite data, but also what methods have been used to measure and analyse road traffic flow in the retrieved research. The information in the abstracts has also been used to achieve relevance ranking of the retrieved records to generate an elite set of documents for further reading.

These tools were provided to the client in different stages of completion during the project. While the question answering system lacks a graphical user interface and needs to be executed in a development environment, and consequently had to be run in our computer infrastructure, we made efforts to create a probabilistic search engine that the client could use to perform their own searches within the retrieved bibliographic material. Using openly available data provided by Crossref[5], we also made it possible for the client to retrieve the relevant articles and download the full-text PDF file from the publisher.

Lastly, the search engine utilises the Altmetric web service[6] to display the Altmetric donut icon, including usage metrics of the research. By clicking on the icon, the client can retrieve usage data from various outlets, including mentions of the articles in patents, policy documents and newspapers. It also identifies mentions of the articles on social media platforms like Facebook and Twitter, as well as additional scientometric data available through the Dimensions citation database[7], making it possible to explore citation data provided by Digital Science.

[5] https://www.crossref.org/
[6] https://www.altmetric.com/
[7] https://www.dimensions.ai/

Question answering in abstracts

In order to enhance the possibilities to explore the datasets obtain from Web of Science, we have developed a strategy for extracting phrases from abstracts that potentially constitute answers to specific questions. This approach, known as question answering (QA), utilises deep learning and language modelling to identify sections in the texts that correspond to a maximum probability of constituting the answers to given questions. The language model used for this QA strategy was SciBERT (see Beltagy, Lo, & Cohan, 2019), which is trained on scientific papers available at the Semantic Scholar website. Examples of questions and answers extract from the air quality dataset are given below.

Q. What measurement methods and types of data sources are used to map and analyse air quality?

A. CO2 sensors, low-cost air quality sensors, optical SPM observations and meteorological measurements, GOCI and Himawari-8, IoT-based portable air quality measuring devices, aerosol laser ablation mass spectrometry, gas chromatography coupled with mass spectrometry, low-cost miniaturised FTIR spectrometers, open-path FTIR spectroscopy, low-dimensional Linear Ventilation Models, land-use regression (LUR) models, electronic nose

Q. How is satellite data used for air quality analyses?

A. strengthen epidemiological studies investigating air pollution health effects, map and monitor surface air pollution, to reconstruct pollution concentrations at high spatio-temporal resolutions, spaceborne remote sensing, Land Surface Temperature (LST) and emissivity estimation, to investigate ground-level PM concentrations, estimate fine particle concentration on large spatial scale

With regard to the traffic safety dataset, the following question and examples of answers were produced:

Q. What methods are used to measure and analyse traffic flows in the road traffic network?

A. average daily traffic, induction loop sensors, inertial sensors, loop detectors and radar sensors, lidar, longitudinal velocity and lateral acceleration variance, vibration-counting, magnetometer and accelerometer sensor, acceleration data from vehicle-mounted sensors, using features of the generated acoustic signals, acoustic measurements.

Search engine for ranked retrieval of abstracts

In addition to the phrase extraction approach to answering natural language questions to material, a probabilistic search engine was developed within the project's scope. This search engine is an implementation of the BM25 best-match retrieval model (see Robertson & Zaragoza, 2009) and, as such, is based on the probability ranking principle of documents. This principle states that documents should be ranked in response to the user query according to the probability of the documents being relevant to the user (Robertson, 1977). The BM25 considers a binary property called *eliteness* of the relationship between document and terms, denoting the quality of a term being about the document content (Robertson & Zaragoza, 2009). This property is turn regarded to be statistically related to the local frequencies of the terms, as modelled by discrete Poisson distributions. To obtain a complete weighting scheme for the term-document relationship, the local frequencies are normalised by the length of the documents and multiplied by a weight conceptually similar to the inverse document frequency (IDF) weighting scheme introduced by Spärck Jones (1972).

In this project, the BM25 model was used to rank bibliographic records from Web of Science based on the content in the corresponding abstracts. Apart from ranking the records using the BM25 model, the search engine also offers the possibility to rank documents that are already selected by the BM25 document score according to citation frequency and citation frequency per year, respectively. To enhance the presentation of the records, an icon (called ”donut”) representing the Altmetric Attention Score is displayed for each document, containing information from Altmetric concerning the extent to which the document has been mentioned in sources like news articles, social media, Wikipedia, and patents. Screenshots of the search engine displaying results for the air quality and the traffic safety datasets, respectively, can be found in Figure 19 and Figure 20.

Figure 20 can serve as a use case for how the approaches presented in this report could be used together. The terms ‘driver’, ‘drowsiness’ were identified in the keyword map in Figure 16 and the co-word map in Figure 18. However, identifying specific terms is not enough. Finding the actual articles that discuss driver drowsiness makes it possible for the user to include this research in their work. The highlight feature helps the user evaluate if the results are relevant, while the possibility of switching to sorting the results based on citation data and the information provided by Altmetric further help evaluate if the article is relevant to use. Lastly, by clicking on the document's title, the user could find the article and download its PDF file.

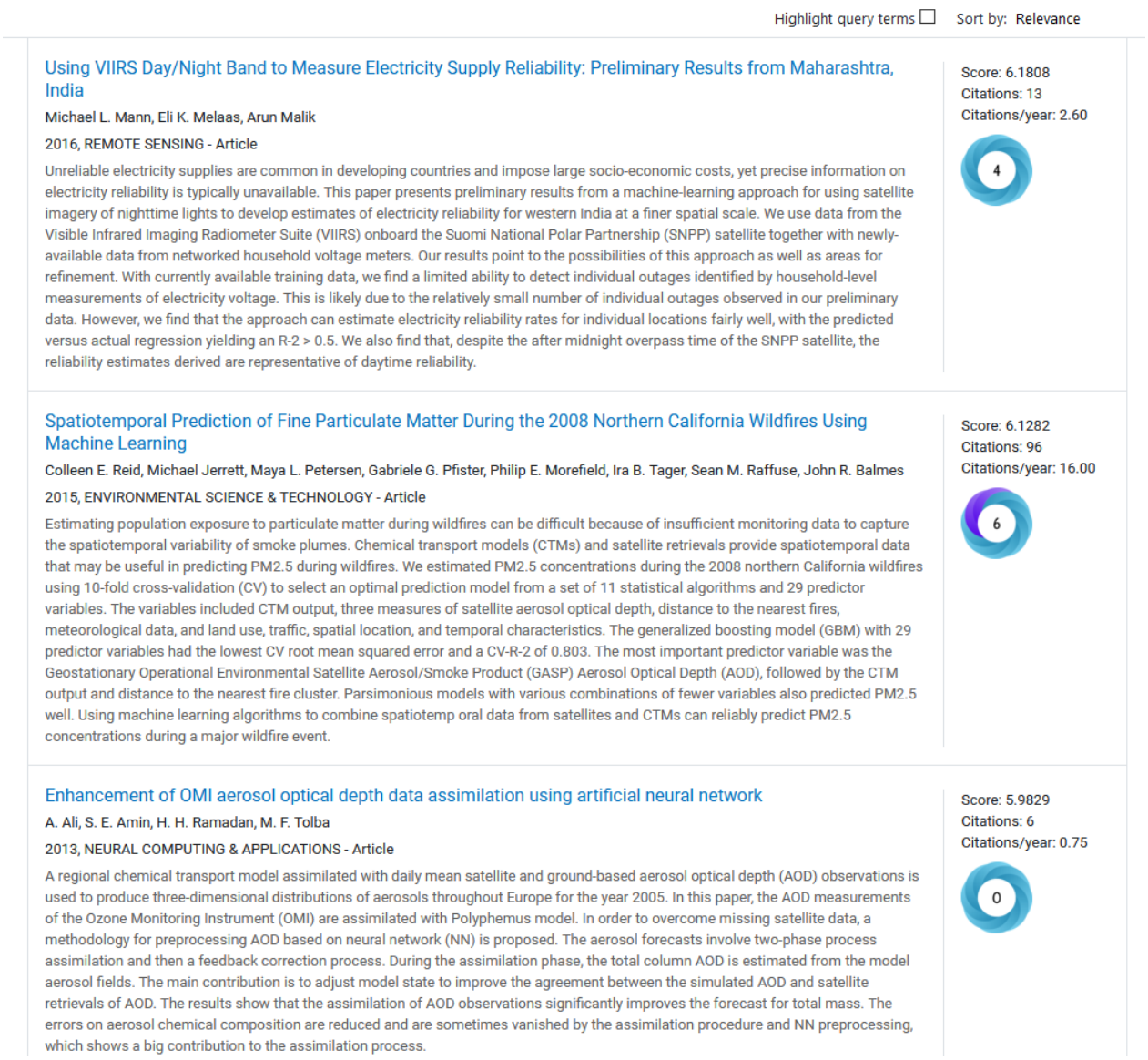


Figure 19. Search result based on the air quality dataset.

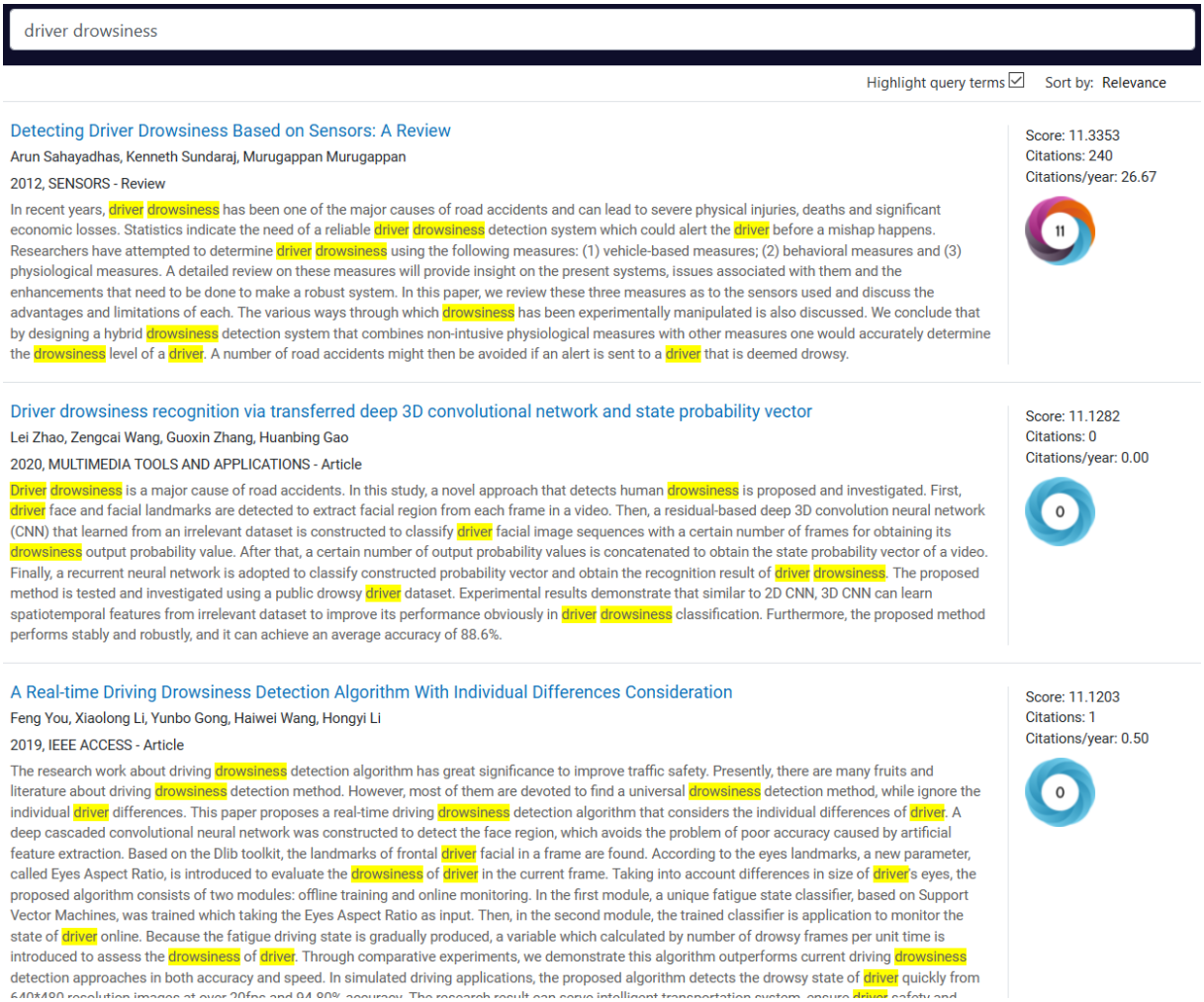


Figure 20. Search result based on the traffic safety dataset. Here, the highlight feature is turned on to show the search terms used.

Conclusions

This report has served to describe the means of providing tools and opportunities to investigate and evaluate research on the subject matters in focus for the project. It has also allowed us to develop approaches for creating a framework for doing scientometric and machine learning-based text analysis on research articles and their content. Furthermore, it has served as a stepping stone for developing combined scientometric and semantic machine learning approaches. In research terminology, this entails methods development that can yield new insights into bibliographic data that will be generalised in further publications from the Data as Impact Lab.

A second purpose of the development of these tools was to showcase the use of machine learning and, in a sense, AI technology to analyse text in large bibliographic data sets. Time did not permit us to develop any language comprehension models of our own. Nevertheless, using pre-trained language models and developing a probabilistic search engine made it possible to identify specific techniques and methodologies used within the research.

While exploratory, we show that in conjunction with the insights gained from the scientometric approach, which provides the opportunity for the user to identify topics and possibly to formulise hypotheses about the content of the retrieved research, using terms and phrases identified in the visualisations, the user can ask questions to the material in the Question Answer system. Using insights from the visualisations and the answers to the questions posed, the user can use the search engine to identify the particular documents wherein the methodologies and techniques have been applied.

Taken together, these approaches provide a complete ecosystem for the user who wants to explore the research about a subject matter and to identify relevant research covering specific aspects of interest within the large set of results retrieved through the Boolean searches described at the beginning of the report.

References

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A Pretrained Language Model for Scientific Text. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. https://doi.org/10.18653/v1/d19-1371

Moretti, F. (2013). *Distant Reading.* London: Verso.

Persson, O. (1994). The intellectual base and research fronts of JASIS 1986–1990. *Journal of the American Society for Information Science, 45*(1), 31-38.

Robertson, S. (1977). The probability ranking principle in IR. *Journal of Documentation, 33*(4), 294–304. https://doi.org/10.1108/eb026647.

Robertson, S., & Zaragoza, H. (2009). The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval, 3*(4), 333–389. https://doi.org/10.1561/1500000019

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation, 28*(1), 11–21. https://doi.org/10.1108/eb026526

Van Eck, N.J., & Waltman, L. (2010). Software survey: *VOSviewer*, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538.

Bilaga C. Utvalda vetenskapliga publikationer baserat på sökningar i BM25 sökmotor

Med hjälp av den sökmotor (BM25) som skapades av Borås högskola genomfördes sökningar av relevanta vetenskapliga publikationer i underlaget från den bibliometriska studien. För området luftkvalitet och AI identifierades baserades sökningen på 3166 publikationer och för området trafiksäkerhet & AI identifierades baserades sökningen på 4171 publikationer.

Söksträngen i sökningen (Web of Science Core Collection den 2021-01-18) för området luftkvalitet och AI, vilken resulterade i 3166 svar:

Timespan: open time window. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.

Search string:

TS= ((air OR atmospher* OR aerosol* OR particl*) NEAR/4 (qual* OR pollut* OR emiss* OR expos*) OR “smart cit*”)

AND

TS=(("machine learning" OR "deep learning" OR "artific* intel*" OR "neural network*" OR "support vector machine*" OR "reinforcement learning" OR "random forest*"))

Söksträngen i sökningen (Web of Science Core Collection den 2021-01-19) för området trafiksäkerhet och AI, vilken resulterade i 4171 svar:

Timespan: open time window. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI.

Search string:

TS= (traffic OR transportation* OR road*)

AND

TS= (assessm* OR safety OR crash* OR accident* or collision*)

AND

TS=(("machine learning" OR "deep learning" OR "artific* intel*" OR "neural network*" OR "support vector machine*" OR "reinforcement learning" OR "random forest*"))

Publikationerna är listade under det publikationsurval sökningarna gjordes (AQ = air quality, TS = traffic safety) samt under använt sökord.

För urvalet AQ gjordes sökningar för termerna review article, data sources, satellite emissions, satellite particles, traffic flow och traffic emissions.

För urvalet TS gjordes sökningar för termerna traffic flow, traffic emissions, speed compliance, vehicle speed och review article.

Nedan listas 41 utvalda publikationer med abstrakt.

Sökord: AQ review article

M. Sowden, U. Mueller, D. Blake,

Review of surface particulate monitoring of dust events using geostationary satellite remote sensing

Atmospheric Environment, Volume 183, 2018, Pages 154-164,

ISSN 1352-2310,

https://doi.org/10.1016/j.atmosenv.2018.04.020.

(https://www.sciencedirect.com/science/article/pii/S1352231018302516)

“The accurate measurements of natural and anthropogenic aerosol particulate matter (PM) is important in managing both environmental and health risks; however, limited monitoring in regional areas hinders accurate quantification. This article provides an overview of the ability of recently launched geostationary earth orbit (GEO) satellites, such as GOES-R (North America) and HIMAWARI (Asia and Oceania), to provide near real-time ground-level PM concentrations (GLCs). The review examines the literature relating to the spatial and temporal resolution required by air quality studies, the removal of cloud and surface effects, the aerosol inversion problem, and the computation of ground-level concentrations rather than columnar aerosol optical depth (AOD). Determining surface PM concentrations using remote sensing is complicated by differentiating intrinsic aerosol properties (size, shape, composition, and quantity) from extrinsic signal intensities, particularly as the number of unknown intrinsic parameters exceeds the number of known extrinsic measurements. The review confirms that development of GEO satellite products has led to improvements in the use of coupled products such as GEOS-CHEM, aerosol types have consolidated on model species rather than prior descriptive classifications, and forward radiative transfer models have led to a better understanding of predictive spectra interdependencies across different aerosol types, despite fewer wavelength bands. However, it is apparent that the aerosol inversion problem remains challenging because there are limited wavelength bands for characterising localised mineralogy. The review finds that the frequency of GEO satellite data exceeds the temporal resolution required for air quality studies, but the spatial resolution is too coarse for localised air quality studies. Continual monitoring necessitates using the less sensitive thermal infra-red bands, which also reduce surface absorption effects. However, given the challenges of the aerosol inversion problem and difficulties in converting columnar AOD to surface concentrations, the review identifies coupled GEO-neural networks as potentially the most viable option for improving quantification.”

Molina-Gómez, N.I., Díaz-Arévalo, J.L. & López-Jiménez, P.A.

Air quality and urban sustainable development: the application of machine learning tools

Int. J. Environ. Sci. Technol. 18, 1029–1046 (2021). https://doi.org/10.1007/s13762-020-02896-6

“Air quality has an effect on a population’s quality of life. As a dimension of sustainable urban development, governments have been concerned about this indicator. This is reflected in the references consulted that have demonstrated progress in forecasting pollution events to issue early warnings using conventional tools which, as a result of the new era of big data, are becoming obsolete. There are a limited number of studies with applications of machine learning tools to characterize and forecast behavior of the environmental, social and economic dimensions of sustainable development as they pertain to air quality. This article presents an analysis of studies that developed machine learning models to forecast sustainable development and air quality. Additionally, this paper sets out to present research that studied the relationship between air quality and urban sustainable development to identify the reliability and possible applications in different urban contexts of these machine learning tools. To that end, a systematic review was carried out, revealing that machine learning tools have been primarily used for clustering and classifying variables and indicators according to the problem analyzed, while tools such as artificial neural networks and support vector machines are the most widely used to predict different types of events. The nonlinear nature and synergy of the dimensions of sustainable development are of great interest for the application of machine learning tools.”

Soomro, Kamran & Bhutta, Muhammad Nasir Mumtaz & Khan, Zaheer & Tahir, Muhammad. (2019).

Smart city big data analytics: An advanced review

Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 9. 10.1002/widm.1319. https://onlinelibrary.wiley.com/doi/full/10.1002/widm.1319

“With the increasing role of ICT in enabling and supporting smart cities, the demand for big data analytics solutions is increasing. Various artificial intelligence, data mining, machine learning and statistical analysis-based solutions have been successfully applied in thematic domains like climate science, energy management, transport, air quality management and weather pattern analysis. In this paper, we present a systematic review of the literature on smart city big data analytics. We have searched a number of different repositories using specific keywords and followed a structured data mining methodology for selecting material for the review. We have also performed a technological and thematic analysis of the shortlisted literature, identified various data mining/machine learning techniques and presented the results. Based on this analysis we also present a classification model that studies four aspects of research in this domain. These include data models, computing models, security and privacy aspects and major market drivers in the smart cities domain. Moreover, we present a gap analysis and identify future directions for research. For the thematic analysis we identified the themes smart city governance, economy, environment, transport and energy. We present the major challenges in these themes, the major research work done in the field of data analytics to address these challenges and future research directions.”

Ullo, S.L.; Sinha, G.R.

Advances in Smart Environment Monitoring Systems Using IoT and Sensors

Sensors 2020, 20, 3113. https://doi.org/10.3390/s20113113

“Air quality, water pollution, and radiation pollution are major factors that pose genuine challenges in the environment. Suitable monitoring is necessary so that the world can achieve sustainable growth, by maintaining a healthy society. In recent years, the environment monitoring has turned into a smart environment monitoring (SEM) system, with the advances in the internet of things (IoT) and the development of modern sensors. Under this scenario, the present manuscript aims to accomplish a critical review of noteworthy contributions and research studies on SEM, that involve monitoring of air quality, water quality, radiation pollution, and agriculture systems. The review is divided on the basis of the purposes where SEM methods are applied, and then each purpose is further analyzed in terms of the sensors used, machine learning techniques involved, and classification methods used. The detailed analysis follows the extensive review which has suggested major recommendations and impacts of SEM research on the basis of discussion results and research trends analyzed. The authors have critically studied how the advances in sensor technology, IoT and machine learning methods make environment monitoring a truly smart monitoring system. Finally, the framework of robust methods of machine learning; denoising methods and development of suitable standards for wireless sensor networks (WSNs), has been suggested.”

Bellinger, C., Mohomed Jabbar, M., Zaïane, O. et al.

A systematic review of data mining and machine learning for air pollution epidemiology

BMC Public Health 17, 907 (2017). https://doi.org/10.1186/s12889-017-4914-3

“Background

Data measuring airborne pollutants, public health and environmental factors are increasingly being stored and merged. These big datasets offer great potential, but also challenge traditional epidemiological methods. This has motivated the exploration of alternative methods to make predictions, find patterns and extract information. To this end, data mining and machine learning algorithms are increasingly being applied to air pollution epidemiology.

Methods

We conducted a systematic literature review on the application of data mining and machine learning methods in air pollution epidemiology. We carried out our search process in PubMed, the MEDLINE database and Google Scholar. Research articles applying data mining and machine learning methods to air pollution epidemiology were queried and reviewed.

Results

Our search queries resulted in 400 research articles. Our fine-grained analysis employed our inclusion/exclusion criteria to reduce the results to 47 articles, which we separate into three primary areas of interest: 1) source apportionment; 2) forecasting/prediction of air pollution/quality or exposure; and 3) generating hypotheses. Early applications had a preference for artificial neural networks. In more recent work, decision trees, support vector machines, k-means clustering and the APRIORI algorithm have been widely applied. Our survey shows that the majority of the research has been conducted in Europe, China and the USA, and that data mining is becoming an increasingly common tool in environmental health. For potential new directions, we have identified that deep learning and geo-spacial pattern mining are two burgeoning areas of data mining that have good potential for future applications in air pollution epidemiology.

Conclusions

We carried out a systematic review identifying the current trends, challenges and new directions to explore in the application of data mining methods to air pollution epidemiology. This work shows that data mining is increasingly being applied in air pollution epidemiology.

The potential to support air pollution epidemiology continues to grow with advancements in data mining related to temporal and geo-spacial mining, and deep learning. This is further supported by new sensors and storage mediums that enable larger, better quality data. This suggests that many more fruitful applications can be expected in the future."

Sökord: AQ data sources

Chen, Y.-C.; Lei, T.-C.; Yao, S.; Wang, H.-P.

PM2.5 Prediction Model Based on Combinational Hammerstein Recurrent Neural Networks

Mathematics 2020, 8, 2178. https://doi.org/10.3390/math8122178

"Airborne particulate matter 2.5 (PM2.5) can have a profound effect on the health of the population. Many researchers have been reporting highly accurate numerical predictions based on raw PM2.5 data imported directly into deep learning models; however, there is still considerable room for improvement in terms of implementation costs due to heavy computational overhead. From the perspective of environmental science, PM2.5 values in a given location can be attributed to local sources as well as external sources. Local sources tend to have a dramatic short-term impact on PM2.5 values, whereas external sources tend to have more subtle but longer-lasting effects. In the presence of PM2.5 from both sources at the same time, this combination of effects can undermine the predictive accuracy of the model. This paper presents a novel combinational Hammerstein recurrent neural network (CHRNN) to enhance predictive accuracy and overcome the heavy computational and monetary burden imposed by deep learning models. The CHRNN comprises a based-neural

network tasked with learning gradual (long-term) fluctuations in conjunction with add-on neural networks to deal with dramatic (short-term) fluctuations. The CHRNN can be coupled with a random forest model to determine the degree to which short-term effects influence long-term outcomes. We also developed novel feature selection and normalization methods to enhance prediction accuracy. Using real-world measurement data of air quality and PM2.5 datasets from Taiwan, the precision of the proposed system in the numerical prediction of PM2.5 levels was comparable to that of state-of-the-art deep learning models, such as deep recurrent neural networks and long short-term memory, despite far lower implementation costs and computational overhead."

Sökord: AQ satellite emissions

Mohammad K. Younes, Ghassan Sulaiman, Ali Al-Mashni.

Integration of Traffic Management and an Artificial Intelligence to Evaluate Urban Air Quality

Asian Journal of Atmospheric Environment Vol. 14, No. 3. pp. 225-235, September 2020

https://doi.org/10.5572/ajae.2020.14.3.225

"Emissions from motor vehicles are the primary source of air pollution, especially in congested urban centres. However, through effective traffic management, it has been found that the level of pollution can be significantly reduced, facilitating the mobility of urban arterials. This study aims to quantify the extent of traffic emissions and to identify the influence of traffic management to improve air quality and reducing traffic emissions. An Adaptive Neuro-Fuzzy Inference System (ANFIS) model was developed to estimate the extent of traffic emissions (NO2 and PM10) at certain intersections. Then, a traffic management simulation software was also used to simulate traffic and to build a traffic improvement scenario at these intersections. This was followed by measuring the improvement in air quality due to traffic management modification, analysed using the developed ANFIS model. The results showed that reducing the delay at certain intersections may reduce NO2 and PM10 significantly. The proposed hybrid model increased the forecasting accuracy and improved the perception between the relationship between traffic characteristics and pollutant emissions. Additionally, it facilitates the work of city planners and helps decision making regarding urban air quality."

Schneider, R.; Vicedo-Cabrera, A.M.; Sera, F.; Masselot, P.; Stafoggia, M.; de Hoogh, K.; Kloog, I.; Reis, S.; Vieno, M.; Gasparrini, A.

A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain

Remote Sens. 2020, 12, 3803. https://doi.org/10.3390/rs12223803

"Epidemiological studies on the health effects of air pollution usually rely on measurements from fixed ground monitors, which provide limited spatio-temporal coverage. Data from

satellites, reanalysis, and chemical transport models offer additional information used to reconstruct pollution concentrations at high spatio-temporal resolutions. This study aims to develop a multi-stage satellite-based machine learning model to estimate daily fine particulate matter (PM2.5) levels across Great Britain between 2008–2018. This high-resolution model consists of random forest (RF) algorithms applied in four stages. Stage-1 augments monitor-PM2.5 series using co-located PM10 measures. Stage-2 imputes missing satellite aerosol optical depth observations using atmospheric reanalysis models. Stage-3 integrates the output from previous stages with spatial and spatio-temporal variables to build a prediction model for PM2.5. Stage-4 applies Stage-3 models to estimate daily PM2.5 concentrations over a 1 km grid. The RF architecture performed well in all stages, with results from Stage-3 showing an average cross-validated R2 of 0.767 and minimal bias. The model performed better over the temporal scale when compared to the spatial component, but both presented good accuracy with an R2 of 0.795 and 0.658, respectively. These findings indicate that direct satellite observations must be integrated with other satellite-based products and geospatial variables to derive reliable estimates of air pollution exposure. The high spatio-temporal resolution and the relatively high precision allow these estimates (approximately 950 million points) to be used in epidemiological analyses to assess health risks associated with both short- and long-term exposure to PM2.5."

Sökord: AQ satellite particles

Massimo Stafoggia, Tom Bellander, Simone Bucci, Marina Davoli, Kees de Hoogh, Francesca de' Donato, Claudio Gariazzo, Alexei Lyapustin, Paola Michelozzi, Matteo Renzi, Matteo Scortichini, Alexandra Shtein, Giovanni Viegi, Itai Kloog, Joel Schwartz,

Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model

Environment International, Volume 124,

2019, Pages 170-179, ISSN 0160-4120, https://doi.org/10.1016/j.envint.2019.01.016.

"Particulate matter (PM) air pollution is one of the major causes of death worldwide, with demonstrated adverse effects from both short-term and long-term exposure. Most of the epidemiological studies have been conducted in cities because of the lack of reliable spatiotemporal estimates of particles exposure in nonurban settings. The objective of this study is to estimate daily PM10 (PM < 10 μm), fine (PM < 2.5 μm, PM2.5) and coarse particles (PM between 2.5 and 10 μm, PM2.5–10) at 1-km2 grid for 2013–2015 using a machine learning approach, the Random Forest (RF). Separate RF models were defined to: predict PM2.5 and PM2.5–10 concentrations in monitors where only PM10 data were available (stage 1); impute missing satellite Aerosol Optical Depth (AOD) data using estimates from atmospheric ensemble models (stage 2); establish a relationship between measured PM and satellite, land use and meteorological parameters (stage 3); predict stage 3 model over each 1-km2 grid cell of Italy (stage 4); and improve stage 3 predictions by

using small-scale predictors computed at the monitor locations or within a small buffer (stage 5). Our models were able to capture most of PM variability, with mean cross-validation (CV) R2 of 0.75 and 0.80 (stage 3) and 0.84 and 0.86 (stage 5) for PM10 and PM2.5, respectively. Model fitting was less optimal for PM2.5–10, in summer months and in southern Italy. Finally, predictions were equally good in capturing annual and daily PM variability, therefore they can be used as reliable exposure estimates for investigating long-term and short-term health effects."

Stirnberg, R., Cermak, J., Fuchs, J., & Andersen, H. (2020).

Mapping and understanding patterns of air quality using satellite data and machine learning

Journal of Geophysical Research: Atmospheres, 125, e2019JD031380. https://doi.org/10.1029/2019JD031380

"The quantification of factors leading to harmfully high levels of particulate matter (PM) remains challenging. This study presents a novel approach using a statistical model that is trained to predict hourly concentrations of particles smaller than 10 µm (PM10) by combining satellite-borne aerosol optical depth (AOD) with meteorological and land-use parameters. The model is shown to accurately predict PM10 (overall R2= 0.77, RMSE = 7.44 µg/m3) for measurement sites in Germany. The capability of satellite observations to map and monitor surface air pollution is assessed by investigating the relationship between AOD and PM10 in the same modeling setup. Sensitivity analyses show that important drivers of modeled PM10 include multiday mean wind flow, boundary layer height (BLH), day of year (DOY), and temperature. Different mechanisms associated with elevated PM10 concentrations are identified in winter and summer. In winter, mean predictions of PM10 concentrations >35 µg/m3occur when BLH is below ~500 m. Paired with multiday easterly wind flow, mean model predictions surpass 40 µg/m3 of PM10. In summer, PM10 concentrations seemingly are less driven by meteorology, but by emission or chemical particle formation processes, which are not included in the model. The relationship between AOD and predicted PM10 concentrations depends to a large extent on ambient meteorological conditions. Results suggest that AOD can be used to assess air quality at ground level in a machine learning approach linking it with meteorological conditions."

Sökord: AQ traffic flow

Zalakeviciute, R.; Bastidas, M.; Buenaño, A.; Rybarczyk, Y.

A Traffic-Based Method to Predict and Map Urban Air Quality

Appl. Sci. 2020, 10, 2035. https://doi.org/10.3390/app10062035

"As global urbanization, industrialization, and motorization keep worsening air quality, a continuous rise in health problems is projected. Limited spatial resolution of the information on air quality inhibits full comprehension of urban population exposure. Therefore, we propose a method to predict urban air pollution from traffic by extracting data from Web-

based applications (Google Traffic). We apply a machine learning approach by training a decision tree algorithm (C4.8) to predict the concentration of PM2.5 during the morning pollution peak from: (i) an interpolation (inverse distance weighting) of the value registered at the monitoring stations, (ii) traffic flow, and (iii) traffic flow + time of the day. The results show that the prediction from traffic outperforms the one provided by the monitoring network (average of 65.5% for the former vs. 57% for the latter). Adding the time of day increases the accuracy by an average of 6.5%. Considering the good accuracy on different days, the proposed method seems to be robust enough to create general models able to predict air pollution from traffic conditions. This affordable method, although beneficial for any city, is particularly relevant for low-income countries, because it offers an economically sustainable technique to address air quality issues faced by the developing world."

W. Alajali, W. Zhou and S. Wen,

Traffic Flow Prediction for Road Intersection Safety

2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guangzhou, China, 2018, pp. 812-820, doi: 10.1109/SmartWorld.2018.00151.

"Road safety is a significant issue in any intelligent transportation system (ITS). Intersections are the most complex part of the road network, as they involve various participants, such as vehicles and pedestrians. Therefore, providing an accurate traffic flow prediction model will enhance traffic efficiency and reduce common problems, such as accidents, congestion and air pollution. However, there are two challenges in the traffic flow prediction problem: first, traffic is a dynamic nonlinear problem due to nonrecurrent events, such as accidents and roadworks, that occur near intersections as an unexpected event will impact the accuracy of the prediction method. The second challenge is that there is a large amount of data which needs a scalable model to efficiently handle big data. To overcome the first issue, in this study, accidents and roadworks data are used, in addition to sensor data that are updated in real time. The datasets are published by VicRoads for the state of Victoria, Australia. Moreover, ensemble decision trees for regression, namely the gradient boosting regression trees (GBRT) and random forest (RF), are proposed. To address the second challenge, the extreme gradient boosting Tree (XGBoost) algorithm, which is a scalable system, is examined to explore its ability to handle traffic data. Finally, a comparative analysis of the proposed methods in terms of time and accuracy is presented."

Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013.

U-Air: when urban air quality inference meets big data

In Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '13). Association for Computing Machinery, New York, NY, USA, 1436–1444. DOI: https://doi.org/10.1145/2487575.2488188

“Information about urban air quality, e.g., the concentration of PM2.5, is of great importance to protect human health and control air pollution. While there are limited air-quality-monitor-stations in a city, air quality varies in urban spaces non-linearly and depends on multiple factors, such as meteorology, traffic volume, and land uses. In this paper, we infer the real-time and fine-grained air quality information throughout a city, based on the (historical and real-time) air quality data reported by existing monitor stations and a variety of data sources we observed in the city, such as meteorology, traffic flow, human mobility, structure of road networks, and point of interests (POIs). We propose a semi-supervised learning approach based on a co-training framework that consists of two separated classifiers. One is a spatial classifier based on an artificial neural network (ANN), which takes spatially-related features (e.g., the density of POIs and length of highways) as input to model the spatial correlation between air qualities of different locations. The other is a temporal classifier based on a linear-chain conditional random field (CRF), involving temporally-related features (e.g., traffic and meteorology) to model the temporal dependency of air quality in a location. We evaluated our approach with extensive experiments based on five real data sources obtained in Beijing and Shanghai. The results show the advantages of our method over four categories of baselines, including linear/Gaussian interpolations, classical dispersion models, well-known classification models like decision tree and CRF, and ANN.”

Sökord: AQ traffic emissions

Carolien Beckx, Luc Int Panis, Karen Van De Vel, Theo Arentze, Wouter Lefebvre, Davy Janssens, Geert Wets,

The contribution of activity-based transport models to air quality modelling: A validation of the ALBATROSS–AURORA model chain

Science of The Total Environment, Volume 407, Issue 12, 2009, Pages 3814-3822, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2009.03.015.

“The potential advantages of using activity-based transport models for air quality purposes have been recognized for a long time but models that have been developed along these lines are still scarce. In this paper we demonstrate that an activity-based model provides useful information for predicting hourly ambient pollutant concentrations. For this purpose, the traffic emissions obtained in a previous application of the activity-based model ALBATROSS were used as input for the AURORA air quality model to predict hourly concentrations of NO2, PM10 and O3 in the Netherlands. Predicted concentrations were compared with measured concentrations at 37 monitoring stations from the Dutch air quality monitoring network. A statistical analysis was performed to evaluate model performance for different pollutants, locations and time periods. Results confirm that modelled and measured concentrations present the same geographical and temporal variation. The overall index of agreement for the prediction of hourly pollutant concentrations amounted to 0.64, 0.75 and 0.57 for NO2, O3 and PM10 respectively. Concerning the predictions for NO2, a major traffic pollutant, a more thorough analysis

revealed that the ALBATROSS–AURORA model chain yielded better predictions near traffic locations than near background stations. Further, the model performed better in urban areas, on weekdays and during the day, consistent with the emission results obtained in a previous study. The results in this paper demonstrate the ability of the activity-based model to predict the contribution of traffic sources to local air pollution with sufficient accuracy and confirms the usefulness of activity-based transport models for air quality purposes. The fact that the ALBATROSS–AURORA chain provides reliable pollutant concentrations on hourly basis for the whole Netherlands instead of using only daily averages near traffic stations is a plus for future exposure studies aiming at more realistic exposure analyses and health impact assessments."

Sadiq, A., El Fazziki, A., Ouarzazi, J. et al.

Towards an agent based traffic regulation and recommendation system for the on-road air quality control

SpringerPlus 5, 1604 (2016). https://doi.org/10.1186/s40064-016-3282-2

"This paper presents an integrated and adaptive problem-solving approach to control the on-road air quality by modeling the road infrastructure, managing traffic based on pollution level and generating recommendations for road users. The aim is to reduce vehicle emissions in the most polluted road segments and optimizing the pollution levels. For this we propose the use of historical and real time pollution records and contextual data to calculate the air quality index on road networks and generate recommendations for reassigning traffic flow in order to improve the on-road air quality. The resulting air quality indexes are used in the system's traffic network generation, which the cartography is represented by a weighted graph. The weights evolve according to the pollution indexes and path properties and the graph is therefore dynamic. Furthermore, the systems use the available pollution data and meteorological records in order to predict the on-road pollutant levels by using an artificial neural network based prediction model. The proposed approach combines the benefits of multi-agent systems, Big data technology, machine learning tools and the available data sources. For the shortest path searching in the road network, we use the Dijkstra algorithm over Hadoop MapReduce framework. The use Hadoop framework in the data retrieve and analysis process has significantly improved the performance of the proposed system. Also, the agent technology allowed proposing a suitable solution in terms of robustness and agility."

Sökord: TS traffic flow

A. Abadi, T. Rajabioun and P. A. Ioannou,

Traffic Flow Prediction for Road Transportation Networks With Limited Traffic Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 653-662, April 2015, doi: 10.1109/TITS.2014.2337238.

“Obtaining accurate information about current and near-term future traffic flows of all links in a traffic network has a wide range of applications, including traffic forecasting, vehicle navigation devices, vehicle routing, and congestion management. A major problem in getting traffic flow information in real time is that the vast majority of links is not equipped with traffic sensors. Another problem is that factors affecting traffic flows, such as accidents, public events, and road closures, are often unforeseen, suggesting that traffic flow forecasting is a challenging task. In this paper, we first use a dynamic traffic simulator to generate flows in all links using available traffic information, estimated demand, and historical traffic data available from links equipped with sensors. We implement an optimization methodology to adjust the origin-to-destination matrices driving the simulator. We then use the real-time and estimated traffic data to predict the traffic flows on each link up to 30 min ahead. The prediction algorithm is based on an autoregressive model that adapts itself to unpredictable events. As a case study, we predict the flows of a traffic network in San Francisco, CA, USA, using a macroscopic traffic flow simulator. We use Monte Carlo simulations to evaluate our methodology. Our simulations demonstrate the accuracy of the proposed approach. The traffic flow prediction errors vary from an average of 2% for 5-min prediction windows to 12% for 30-min windows even in the presence of unpredictable events.”

Azzedine Boukerche, Yanjie Tao, Peng Sun,

Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems

Computer Networks, Volume 182, 2020, 107484, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2020.107484.

“ In recent years, the Intelligent transportations system (ITS) has received considerable attention, due to higher demands for road safety and efficiency in highly interconnected road networks. As an essential part of ITS, traffic prediction can provide support in many aspects, such as road routing, traffic congestion control, etc. To provide a more comprehensive overview of the role of traffic forecasting in ITS systems, we will first introduce the corresponding ITS applications and discuss how traffic forecasting can improve the performance of these applications. Next, we will introduce the general prediction procedure as well as some basic concepts of traffic flow prediction, followed by a description of a general framework for implementing the traffic flow prediction. In this survey, mainly two sorts of prediction methods are focused, statistics-based and machine learning (ML)-based. These two types of approaches are more used in ITS traffic flow predictions these years, and service for different contexts. Generally speaking, the statistics-based models have better model interpretability, but the rigorous model structure limits the adaptability, while ML-based models are more flexible. Accordingly, we will introduce the characteristics of these two types of methods through specific examples of state-of-the-art approaches. Last but not least, some potential and meaningful development directions corresponding to this domain are introduced to do a great favor for future research.”

Sun, S.; Wu, H.; Xiang, L.

City-Wide Traffic Flow Forecasting Using a Deep Convolutional Neural Network

Sensors 2020, 20, 421. https://doi.org/10.3390/s20020421

“City-wide traffic flow forecasting is a significant function of the Intelligent Transport System (ITS), which plays an important role in city traffic management and public travel safety. However, this remains a very challenging task that is affected by many complex factors, such as road network distribution and external factors (e.g., weather, accidents, and holidays). In this paper, we propose a deep-learning-based multi-branch model called TFFNet (Traffic Flow Forecasting Network) to forecast the short-term traffic status (flow) throughout a city. The model uses spatiotemporal traffic flow matrices and external factors as its input and then infers and outputs the future short-term traffic status (flow) of the whole road network. For modelling the spatial correlations of the traffic flows between current and adjacent road segments, we employ a multi-layer fully convolutional framework to perform cross-correlation calculation and extract the hierarchical spatial dependencies from local to global scales. Also, we extract the temporal closeness and periodicity of traffic flow from historical observations by constructing a high-dimensional tensor comprised of traffic flow matrices from three fragments of the time axis: recent time, near history, and distant history. External factors are also considered and trained with a fully connected neural network and then fused with the output of the main component of TFFNet. The multi-branch model is automatically trained to fit complex patterns hidden in the traffic flow matrices until reaching pre-defined convergent criteria via the back-propagation method. By constructing a rational model input and network architecture, TFFNet can capture spatial and temporal dependencies simultaneously from traffic flow matrices during model training and outperforms other typical traffic flow forecasting methods in the experimental dataset.”

Sökord: TS traffic emissions

Ho-Wen Chen, Hsi-Hsien Yang, Yu-Sheng Wang,

Automobile gross emitter screening with remote sensing data using objective-oriented neural network

Science of The Total Environment, Volume 407, Issue 22, 2009, Pages 5811-5817, ISSN 0048-9697, https://doi.org/10.1016/j.scitotenv.2009.07.016.

“One of the costs of Taiwan's massive economic development has been severe air pollution problems in many parts of the island. Since vehicle emissions are the major source of air pollution in most of Taiwan's urban areas, Taiwan's government has implemented policies to rectify the degrading air quality, especially in areas with high population density. To reduce vehicle pollution emissions an on-road remote sensing and monitoring system is used to check the exhaust emissions from gasoline engine automobiles. By identifying individual vehicles with excessive emissions for follow-up inspection and testing, air quality in the urban environment is expected to improve greatly. Because remote sensing is

capable of measuring a large number of moving vehicles in a short period, it has been considered as an assessment technique in place of the stationary emission-sampling techniques. However, inherent measurement uncertainty of remote sensing instrumentation, compounded by the indeterminacy of monitoring site selection, plus the vagaries of weather, causes large errors in pollution discrimination and limits the application of the remote sensing. Many governments are still waiting for a novel data analysis methodology to clamp down on heavily emitting vehicles by using remote sensing data. This paper proposes an artificial neural network (ANN), with vehicle attributes embedded, that can be trained by genetic algorithm (GA) based on different strategies to predict vehicle emission violation. Results show that the accuracy of predicting emission violation is as high as 92%. False determinations tend to occur for vehicles aged 7–13 years, peaking at 10 years of age."

Viaene P. et al. (2018)

Application of a Comprehensive Integrated Assessment Tool for the Brussels Capital Region

In: Mensink C., Kallos G. (eds) Air Pollution Modeling and its Application XXV. ITM 2016. Springer Proceedings in Complexity. Springer, Cham. https://doi.org/10.1007/978-3-319-57645-9_43

"While in general air quality has improved in Europe over the past decades, there are still problems with exceedances of ambient air quality limit values in many urban areas. To design efficient Air Quality Plans to face these problems, methodologies and tools are required to assess the effects of possible abatement measures on local air quality. One such tool is RIAT+ (http://www.riatplus.eu) which was designed to help regional decision makers select air pollution reduction policies that will improve the air quality at minimal costs. In this contribution to ITM we present the results obtained as well as the lessons learned for an application of the RIAT+ tool to the Brussels Capital Region. RIAT+ has been previously applied successfully to regions in the Po Valley in Italy and to the Alsace region in France. The application to the BCR however poses specific challenges due to the fact that both the area on which the abatement measures can be applied as the emissions are more limited than in previous cases. Inside the BCR, emissions of nitrogen oxide and particulate matter are mainly from non-industrial combustion and traffic. For these two source categories a list of possible air quality abatement measures was provided by the Brussels Environmental agency. To allow RIAT+ to determine the optimal combination of abatement measures with minimal cost, information was collected on both the emission reduction efficiency and the costs of each of these measures. RIAT+ efficiently calculates concentration changes from emission changes using a receptor model based on an artificial neural network. Input for this receptor model was obtained from the results of a validated AURORA chemical transport model setup for the BCR. Once the receptor model was validated, RIAT+ was then used to calculate the effect of the different proposed abatement measures on air quality."

Jason G. Su, Ying-Ying Meng, Xiao Chen, John Molitor, Dahai Yue, Michael Jerrett,

Predicting differential improvements in annual pollutant concentrations and exposures for regulatory policy assessment

Environment International, Volume 143, 2020, 105942, ISSN 0160-4120,

https://doi.org/10.1016/j.envint.2020.105942.

“Over the past decade, researchers and policy-makers have become increasingly interested in regulatory and policy interventions to reduce air pollution concentrations and improve human health. Studies have typically relied on relatively sparse environmental monitoring data that lack the spatial resolution to assess small-area improvements in air quality and health. Few studies have integrated multiple types of measures of an air pollutant into one single modeling framework that combines spatially- and temporally-rich monitoring data. In this paper, we investigated the differential effects of California emissions reduction plan on reducing air pollution between those living in the goods movement corridors (GMC) that are within 500 m of major highways that serve as truck routes to those farther away or adjacent to routes that prohibit trucks. A mixed effects Deletion/Substitution/Addition (D/S/A) machine learning algorithm was developed to model annual pollutant concentrations of nitrogen dioxide (NO2) by taking repeated measures into consideration and by integrating multiple types of NO2 measurements, including those through government regulatory and research-oriented saturation monitoring into a single modeling framework. Difference-in-difference analysis was conducted to identify whether those living in GMC demonstrated statistically larger reductions in air pollution exposure. The mixed effects D/S/A machine learning modeling result indicated that GMC had 2 ppb greater reductions in NO2 concentrations from pre- to post-policy period than far away areas. The difference-in-difference analysis demonstrated that the subjects living in GMC experienced statistically significant greater reductions in NO2 exposure than those living in the far away areas. This study contributes to scientific knowledge by providing empirical evidence that improvements in air quality via the emissions reductions plan policies impacted traffic-related air pollutant concentrations and associated exposures most among low-income Californians with chronic conditions living in GMC. The identified differences in pollutant reductions across different location domains may be applicable to other states or other countries if similar policies are enacted.”

Abduljabbar, R.; Dia, H.; Liyanage, S.; Bagloee, S.A.

Applications of Artificial Intelligence in Transport: An Overview

Sustainability 2019, 11, 189. https://doi.org/10.3390/su11010189

“The rapid pace of developments in Artificial Intelligence (AI) is providing unprecedented opportunities to enhance the performance of different industries and businesses, including the transport sector. The innovations introduced by AI include highly advanced computational methods that mimic the way the human brain works. The application of AI in the transport field is aimed at overcoming the challenges of an increasing travel demand, CO2 emissions, safety concerns, and environmental degradation. In light of the availability

of a huge amount of quantitative and qualitative data and AI in this digital age, addressing these concerns in a more efficient and effective fashion has become more plausible. Examples of AI methods that are finding their way to the transport field include Artificial Neural Networks (ANN), Genetic algorithms (GA), Simulated Annealing (SA), Artificial Immune system (AIS), Ant Colony Optimiser (ACO) and Bee Colony Optimization (BCO) and Fuzzy Logic Model (FLM) The successful application of AI requires a good understanding of the relationships between AI and data on one hand, and transportation system characteristics and variables on the other hand. Moreover, it is promising for transport authorities to determine the way to use these technologies to create a rapid improvement in relieving congestion, making travel time more reliable to their customers and improve the economics and productivity of their vital assets. This paper provides an overview of the AI techniques applied worldwide to address transportation problems mainly in traffic management, traffic safety, public transportation, and urban mobility. The overview concludes by addressing the challenges and limitations of AI applications in transport."

Yisheng Lv, S. Tang, Hongxia Zhao and Shuang Li,

Real-time highway accident prediction based on support vector machines

2009 Chinese Control and Decision Conference, Guilin, China, 2009, pp. 4403-4407, doi: 10.1109/CCDC.2009.5192409.

"Traditional traffic accident prediction uses long-term traffic data such as annual average daily traffic and hourly volume. In contrast to traditional traffic accident prediction, real-time traffic accident prediction uses real-time traffic data, obtained from inductive loop detectors and usually collected every 20 or 30 seconds, to identify hazardous traffic conditions to potentially prevent the traffic accident occurrence. We aim at identifying traffic patterns leading to traffic accidents and not leading to traffic accidents in this study. Support vector machines (SVM) are used to classify traffic conditions into those two patterns with real-time traffic data. Traffic accident data and its corresponding real-time traffic data are collected from the traffic simulation software TSIS, which is a microscopic traffic simulation software. This is the first time the SVM method is applied for real-time traffic accident prediction. The experimental results show that it is promising for real-time traffic accident prediction by using the support vector machine method."

T. Pamula,

Road Traffic Conditions Classification Based on Multilevel Filtering of Image Content Using Convolutional Neural Networks

in IEEE Intelligent Transportation Systems Magazine, vol. 10, no. 3, pp. 11-21, Fall 2018, doi: 10.1109/MITS.2018.2842040.

"Classification of traffic conditions is a vital task for determining traffic control strategies in ITS. Systematic assessment of the volume of traffic enables appropriate changes of control measures for directing traffic streams to reach set goals of performance. Video traffic monitoring is a suitable and convenient source of traffic data. The paper presents a method

of classification of road traffic conditions based on video surveillance data. Convolutional neural network is used to classify the video content and establish measures of congestion of the observed traffic. Four levels of traffic conditions are distinguished which correspond to LOS categories. The network is validated using video data from several traffic observation sites. The trained CNN is capable of processing video data for systematic use by subsystems of ITS responsible for traffic management. The results of classification are compared with neural network based classifiers: a MLP (multi layer perceptron) and a DLN (deep learning network) with autoencoders. The proposed method is more accurate and less sensitive to the quality of video data."

Sökord: TS speed compliance

Weijia Xu, Natalia Ruiz-Juri, Ruizhu Huang, Jennifer Duthie, and John Clary. 2018.

Automated pedestrian safety analysis using data from traffic monitoring cameras

In Proceedings of the 1st ACM/EIGSCC Symposium on Smart Cities and Communities (SCC '18). Association for Computing Machinery, New York, NY, USA, Article 3, 1–8. DOI: https://doi.org/10.1145/3236461.3241972

"Transportation agencies often own extensive networks of monocular traffic cameras, which are typically used for traffic monitoring. However, the information captured by such cameras can also be of great value for transportation planning and operations applications, particularly when large data sets may be systematically analyzed. In this paper, we propose an approach to use data collected by existing monitoring cameras to automatically identify locations where pedestrian safety may be a concern. Our methodology utilizes a convolutional-neural-network-based method to recognize pedestrians in traffic camera feeds. Results are stored and aggregated, and may be queried for further analyses. The proposed computational approach may leverage hardware such as GPUs and distributed computing clusters to enable the analysis of large volumes of data. The post recognition analysis utilizes unsupervised learning methods to identify the spatial and temporal patterns of pedestrian positions, which are then correlated to specific scenarios such as usage of crosswalk, compliance with traffic signals, and pedestrian-vehicle interactions. Applications include the identification of potential safety concerns, measuring the effectiveness of proposed safety strategies, and identifying the need for improvements. This work provides preliminary results based on data from cameras owned by the City of Austin. We also discuss outputs such as pedestrian volume estimation and crossing hot-zones identification in the context of Smart Cities, and identify potential challenges and limitations."

Cheng Xu, Qiangwei Li, Zhaowei Qu, Sheng Jin,

Predicting Free Flow Speed and Crash Risk of Bicycle Traffic Flow Using Artificial Neural Network Models

Mathematical Problems in Engineering, vol. 2015, Article ID 212050, 11 pages, 2015. https://doi.org/10.1155/2015/212050

“Free flow speed is a fundamental measure of traffic performance and has been found to affect the severity of crash risk. However, the previous studies lack analysis and modelling of impact factors on bicycles’ free flow speed. The main focus of this study is to develop multilayer back propagation artificial neural network (BPANN) models for the prediction of free flow speed and crash risk on the separated bicycle path. Four different models with considering different combinations of input variables (e.g., path width, traffic condition, bicycle type, and cyclists’ characteristics) were developed. 459 field data samples were collected from eleven bicycle paths in Hangzhou, China, and 70% of total samples were used for training, 15% for validation, and 15% for testing. The results show that considering the input variables of bicycle types and characteristics of cyclists will effectively improve the accuracy of the prediction models. Meanwhile, the parameters of bicycle types have more significant effect on predicting free flow speed of bicycle compared to those of cyclists’ characteristics. The findings could contribute for evaluation, planning, and management of bicycle safety.”

S. Hua, M. Kapoor and D. C. Anastasiu,

Vehicle Tracking and Speed Estimation from Traffic Videos

2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 2018, pp. 153-1537, doi: 10.1109/CVPRW.2018.00028.

“The rapid recent advancements in the computation ability of everyday computers have made it possible to widely apply deep learning methods to the analysis of traffic surveillance videos. Traffic flow prediction, anomaly detection, vehicle re-identification, and vehicle tracking are basic components in traffic analysis. Among these applications, traffic flow prediction, or vehicle speed estimation, is one of the most important research topics of recent years. Good solutions to this problem could prevent traffic collisions and help improve road planning by better estimating transit demand. In the 2018 NVIDIA AI City Challenge, we combine modern deep learning models with classic computer vision approaches to propose an efficient way to predict vehicle speed. In this paper, we introduce some state-of-the-art approaches in vehicle speed estimation, vehicle detection, and object tracking, as well as our solution for Track 1 of the Challenge.”

J. Wang, Q. Gu, J. Wu, G. Liu and Z. Xiong,

Traffic Speed Prediction and Congestion Source Exploration: A Deep Learning Method

2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 2016, pp. 499-508, doi: 10.1109/ICDM.2016.0061.

“Traffic speed prediction is a long-standing and critically important topic in the area of Intelligent Transportation Systems (ITS). Recent years have witnessed the encouraging potentials of deep neural networks for real-life applications of various domains. Traffic

speed prediction, however, is still in its initial stage without making full use of spatio-temporal traffic information. In light of this, in this paper, we propose a deep learning method with an Error-feedback Recurrent Convolutional Neural Network structure (eRCNN) for continuous traffic speed prediction. By integrating the spatio-temporal traffic speeds of contiguous road segments as an input matrix, eRCNN explicitly leverages the implicit correlations among nearby segments to improve the predictive accuracy. By further introducing separate error feedback neurons to the recurrent layer, eRCNN learns from prediction errors so as to meet predictive challenges rising from abrupt traffic events such as morning peaks and traffic accidents. Extensive experiments on real-life speed data of taxis running on the 2nd and 3rd ring roads of Beijing city demonstrate the strong predictive power of eRCNN in comparison to some state-of-the-art competitors. The necessity of weight pre-training using a transfer learning notion has also been testified. More interestingly, we design a novel influence function based on the deep learning model, and showcase how to leverage it to recognize the congestion sources of the ring roads in Beijing."

Sökord: TS vehicle speed

António Lobo, Marco Amorim, Carlos Rodrigues, António Couto,

Modelling the Operating Speed in Segments of Two-Lane Highways from Probe Vehicle Data: A Stochastic Frontier Approach

Journal of Advanced Transportation, vol. 2018, Article ID 3540785, 10 pages, 2018. https://doi.org/10.1155/2018/3540785

"Most of the existing operating speed statistical models are applicable to individual design elements, particularly horizontal curves and tangents. A segment approach to operating speed has rarely been followed, with a few exceptions mainly related to the performance assessment of urban and freeway corridors, or design consistency studies using speed profiles built from successive design elements. This study introduces a new model to predict operating speeds in segments of two-lane highways. The maximum operating speed is given by a stochastic frontier function of the average daily traffic and road geometrics; the asymmetric disturbance accounts for the diversity in drivers' behaviour and vehicle characteristics, allowing estimating any percentile speed. The model was calibrated using probe vehicle data from noncongested roads. The accuracy of the average daily traffic in representing the actual driving conditions was further validated using simultaneous speed-traffic measurements. The new model aims to assist practitioners in the evaluation of design consistency from a macroscopic perspective since the early stages of road planning and design, as well as to support the definition of speed limits at new or existing infrastructures."

T. Xu, X. Sun, Y. Wu, Y. He and C. XIE,

Artificial neural network and wavelet analysis application in speed forecast of Beijing urban freeway

2009 IEEE Intelligent Vehicles Symposium, Xi'an, China, 2009, pp. 1004-1008, doi: 10.1109/IVS.2009.5164418.

“Speed is an important factor for traffic safety evolution. The use of ITS technology in speed management with real-time information of urban freeway is one of strategies to enhance road safety. This paper presents methods for prediction short-term traffic flow speed on Beijing urban freeway with real time information from inductive loops. The source data sets including traffic volumes, speed and occupancy which are collected 24 h/day over several years on Beijing ring road. Traffic flow is divided into three statuses, including free flow, transition status and congestion according to occupancy. To achieve objective prediction results, wavelet technology is applied to de-noising process. The artificial neural network, which does not require any pre-defined underlying relationship between dependent and independent variables, is a powerful tool in dealing with prediction problems. In this paper, RBF network is designed for predicting speed for future five minutes. Results show that the proposed RBF network model produces reliable estimates of vehicle speed for three various traffic conditions, especially congestion condition.”

Wang, L., Shi, Q., & Abdel-Aty, M. (2015).

Predicting Crashes on Expressway Ramps with Real-Time Traffic and Weather Data

Transportation Research Record, 2514(1), 32–38. https://doi.org/10.3141/2514-04

“Limited research has been conducted on real-time crash analysis of expressway ramps, although there have been many studies in recent years on estimating real-time crash prediction models for main lines. This study presents Bayesian logistic regression models for single-vehicle (SV) and multivehicle (MV) crashes on expressway ramps by using real-time microwave vehicle detection system data, real-time weather data, and ramp geometric information. The results find that the logarithm of the vehicle count, average speed in a 5-min interval, and visibility are significant factors for the occurrence of SV and MV crashes. The Bayesian logistic regression models show that curved ramps and wet road surfaces would increase the possibility of an SV crash, and off-ramps would result in high risk of MV crashes. The high standard deviation of speed in a 5-min interval would significantly increase MV crash likelihood. Random Forests software was applied in variable importance analysis, and the results revealed that the most important factors influencing crashes on ramps were traffic variables, the second most important factors are weather variables, and the least important but still significant factor was the ramp geometry.”

M. M. Ahmed and M. A. Abdel-Aty,

The Viability of Using Automatic Vehicle Identification Data for Real-Time Crash Prediction

in IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 2, pp. 459-468, June 2012, doi: 10.1109/TITS.2011.2171052.

“Real-time crash prediction research attempted the use of data from inductive loop detectors; however, no safety analysis has been carried out using traffic data from one of

the most growing nonintrusive surveillance systems, i.e., the tag readers on toll roads known as automatic vehicle identification (AVI) systems. In this paper, for the first time, the identification of freeway locations with high crash potential has been examined using real-time speed data collected from AVI. Travel time and space mean speed data collected by AVI systems and crash data of a total of 78 mi on the expressway network in Orlando in 2008 were collected. Utilizing a random forest technique for significant variable selection and stratified matched case-control to account for the confounding effects of location, time, and season, the log odds of crash occurrence were calculated. The length of the AVI segment was found to be a crucial factor that affects the usefulness of the AVI traffic data. While the results showed that the likelihood of a crash is statistically related to speed data obtained from AVI segments within an average length of 1.5 mi and crashes can be classified with about 70% accuracy, all speed parameters obtained from AVI systems spaced at 3 mi or more apart were found to be statistically insignificant to identify crash-prone conditions. The findings of this study illustrate a promising real-time safety application for one of the most widely used and already present intelligent transportation systems, with many possible advances in the context of advanced traffic management."

Sökord: TS review article

Jeevith Hegde, Børge Rokseth,

Applications of machine learning methods for engineering risk assessment – A review

Safety Science, Volume 122, 2020, 104492, ISSN 0925-7535,

https://doi.org/10.1016/j.ssci.2019.09.015.

"The purpose of this article is to present a structured review of publications utilizing machine learning methods to aid in engineering risk assessment. A keyword search is performed to retrieve relevant articles from the databases of Scopus and Engineering Village. The search results are filtered according to seven selection criteria. The filtering process resulted in the retrieval of one hundred and twenty-four relevant research articles. Statistics based on different categories from the citation database is presented. By reviewing the articles, additional categories, such as the type of machine learning algorithm used, the type of input source used, the type of industry targeted, the type of implementation, and the intended risk assessment phase are also determined. The findings show that the automotive industry is leading the adoption of machine learning algorithms for risk assessment. Artificial neural networks are the most applied machine learning method to aid in engineering risk assessment. Additional findings from the review process are also presented in this article."

M. Won,

Intelligent Traffic Monitoring Systems for Vehicle Classification: A Survey

in IEEE Access, vol. 8, pp. 73340-73358, 2020, doi: 10.1109/ACCESS.2020.2987634.

“A traffic monitoring system is an integral part of Intelligent Transportation Systems (ITS). It is one of the critical transportation infrastructures that transportation agencies invest a huge amount of money to collect and analyze the traffic data to better utilize the roadway systems, improve the safety of transportation, and establish future transportation plans. With recent advances in MEMS, machine learning, and wireless communication technologies, numerous innovative traffic monitoring systems have been developed. In this article, we present a review of state-of-the-art traffic monitoring systems focusing on the major functionality-vehicle classification. We organize various vehicle classification systems, examine research issues and technical challenges, and discuss hardware/software design, deployment experience, and system performance of vehicle classification systems. Finally, we discuss a number of critical open problems and future research directions in an aim to provide valuable resources to academia, industry, and government agencies for selecting appropriate technologies for their traffic monitoring applications.”

Zantalis, F.; Koulouras, G.; Karabetsos, S.; Kandris, D.

A Review of Machine Learning and IoT in Smart Transportation.

Future Internet 2019, 11, 94. https://doi.org/10.3390/fi11040094

“With the rise of the Internet of Things (IoT), applications have become smarter and connected devices give rise to their exploitation in all aspects of a modern city. As the volume of the collected data increases, Machine Learning (ML) techniques are applied to further enhance the intelligence and the capabilities of an application. The field of smart transportation has attracted many researchers and it has been approached with both ML and IoT techniques. In this review, smart transportation is considered to be an umbrella term that covers route optimization, parking, street lights, accident prevention/detection, road anomalies, and infrastructure applications. The purpose of this paper is to make a self-contained review of ML techniques and IoT applications in Intelligent Transportation Systems (ITS) and obtain a clear view of the trends in the aforementioned fields and spot possible coverage needs. From the reviewed articles it becomes profound that there is a possible lack of ML coverage for the Smart Lighting Systems and Smart Parking applications. Additionally, route optimization, parking, and accident/detection tend to be the most popular ITS applications among researchers.”

Mehdizadeh, A.; Cai, M.; Hu, Q.; Alamdar Yazdi, M.A.; Mohabbati-Kalejahi, N.; Vinel, A.; Rigdon, S.E.; Davis, K.C.; Megahed, F.M.

A Review of Data Analytic Applications in Road Traffic Safety. Part 1: Descriptive and Predictive Modeling

Sensors 2020, 20, 1107. https://doi.org/10.3390/s20041107

“This part of the review aims to reduce the start-up burden of data collection and descriptive analytics for statistical modeling and route optimization of risk associated with motor vehicles. From a data-driven bibliometric analysis, we show that the literature is divided into two disparate research streams: (a) predictive or explanatory models that

attempt to understand and quantify crash risk based on different driving conditions, and (b) optimization techniques that focus on minimizing crash risk through route/path-selection and rest-break scheduling. Translation of research outcomes between these two streams is limited. To overcome this issue, we present publicly available high-quality data sources (different study designs, outcome variables, and predictor variables) and descriptive analytic techniques (data summarization, visualization, and dimension reduction) that can be used to achieve safer-routing and provide code to facilitate data collection/exploration by practitioners/researchers. Then, we review the statistical and machine learning models used for crash risk modeling. We show that (near) real-time crash risk is rarely considered, which might explain why the optimization models (reviewed in Part 2) have not capitalized on the research outcomes from the first stream."

Hu, Q.; Cai, M.; Mohabbati-Kalejahi, N.; Mehdizadeh, A.; Alamdar Yazdi, M.A.; Vinel, A.; Rigdon, S.E.; Davis, K.C.; Megahed, F.M.

A Review of Data Analytic Applications in Road Traffic Safety. Part 2: Prescriptive Modeling

Sensors 2020, 20, 1096. https://doi.org/10.3390/s20041096

"In the first part of the review, we observed that there exists a significant gap between the predictive and prescriptive models pertaining to crash risk prediction and minimization, respectively. In this part, we review and categorize the optimization/ prescriptive analytic models that focus on minimizing crash risk. Although the majority of works in this segment of the literature are related to the hazardous materials (hazmat) trucking problems, we show that (with some exceptions) many can also be utilized in non-hazmat scenarios. In an effort to highlight the effect of crash risk prediction model on the accumulated risk obtained from the prescriptive model, we present a simulated example where we utilize four risk indicators (obtained from logistic regression, Poisson regression, XGBoost, and neural network) in the k-shortest path algorithm. From our example, we demonstrate two major designed takeaways: (a) the shortest path may not always result in the lowest crash risk, and (b) a similarity in overall predictive performance may not always translate to similar outcomes from the prescriptive models. Based on the review and example, we highlight several avenues for future research."

N. O. Alsrehin, A. F. Klaib and A. Magableh,

Intelligent Transportation and Control Systems Using Data Mining and Machine Learning Techniques: A Comprehensive Study

in IEEE Access, vol. 7, pp. 49830-49857, 2019, doi: 10.1109/ACCESS.2019.2909114.

"Traffic congestion is becoming the issues of the entire globe. This study aims to explore and review the data mining and machine learning technologies adopted in research and industry to attempt to overcome the direct and indirect traffic issues on humanity and societies. The study's methodology is to comprehensively review around 165 studies, criticize, and categorize all these studies into a chronological and understandable category. The study is focusing on the traffic management approaches that were depended on data

mining and machine learning technologies to detect and predict the traffic only. This study has found that there is no standard traffic management approach that the community of traffic management has agreed on. This study is important to the traffic research communities, traffic software companies, and traffic government officials. It has a direct impact on drawing a clear path for new traffic management propositions. This study is one of the largest studies with respect to the size of its reviewed articles that were focused on data mining and machine learning. Additionally, this study will draw general attention to a new traffic management proposition approach."

Moinul Hossain, Mohamed Abdel-Aty, Mohammed A. Quddus, Yasunori Muromachi, Soumik Nafis Sadeek,

Real-time crash prediction models: State-of-the-art, design pathways and ubiquitous requirements

Accident Analysis & Prevention, Volume 124, 2019, Pages 66-84,

ISSN 0001-4575, https://doi.org/10.1016/j.aap.2018.12.022.

"Proactive traffic safety management systems can monitor traffic conditions in real-time, identify the formation of unsafe traffic dynamics, and implement suitable interventions to bring unsafe conditions back to normal traffic situations. Recent advancements in artificial intelligence, sensor fusion and algorithms have brought about the introduction of a proactive safety management system closer to reality. The basic prerequisite for developing such a system is to have a reliable crash prediction model that takes real-time traffic data as input and evaluates their association with crash risk. Since the early 21st century, several studies have focused on developing such models. Although the idea has considerably matured over time, the endeavours have been quite discrete and fragmented at best because the fundamental aspects of the overall modelling approach substantially vary. Therefore, a number of transitional challenges have to be identified and subsequently addressed before a ubiquitous proactive safety management system can be formulated, designed and implemented in real-world scenarios. This manuscript conducts a comprehensive review of existing real-time crash prediction models with the aim of illustrating the state-of-the-art and systematically synthesizing the thoughts presented in existing studies in order to facilitate its translation from an idea into a ready to use technology. Towards that journey, it conducts a systematic review by applying various text mining methods and topic modelling. Based on the findings, this paper ascertains the development pathways followed in various studies, formulates the ubiquitous design requirements of such models from existing studies and knowledge of similar systems. Finally, this study evaluates the universality and design compatibility of existing models. This paper is, therefore, expected to serve as a one stop knowledge source for facilitating a faster transition from the idea of real-time crash prediction models to a real-world operational proactive traffic safety management system."

Sanjana, S., Sanjana, S., Shriya, V.R. et al.

A review on various methodologies used for vehicle classification, helmet detection and number plate recognition

Evol. Intel. (2020). https://doi.org/10.1007/s12065-020-00493-7

“Vehicle detection and classification has been an area of application of image processing and machine learning which is being researched extensively in accordance with its importance due to increasing number of vehicles, traffic rule defaulters and accidents. This paper aims to review various methodologies used and how it has evolved to give better results in the past years, closely moving towards usage of machine learning. This has resulted in advancing the problem statement towards helmet detection followed by number plate detection of defaulters. Object detection and Text recognition that are available in various frameworks offer built-in models which are easy to use or offer easy methods to build and train customized models.”

Philippe Barbosa Silva, Michelle Andrade, Sara Ferreira,

Machine learning applied to road safety modeling: A systematic literature review

Journal of Traffic and Transportation Engineering (English Edition), Volume 7, Issue 6, 2020, Pages 775-790, ISSN 2095-7564,

https://doi.org/10.1016/j.jtte.2020.07.004.

“Road safety modeling is a valuable strategy for promoting safe mobility, enabling the development of crash prediction models (CPM) and the investigation of factors contributing to crash occurrence. This modeling has traditionally used statistical techniques despite acknowledging the limitations of this kind of approach (specific assumptions and prior definition of the link functions), which provides an opportunity to explore alternatives such as the use of machine learning (ML) techniques. This study reviews papers that used ML techniques for the development of CPM. A systematic literature review protocol was conducted, that resulted in the analysis of papers and their systematization. Three types of models were identified: crash frequency, crash classification by severity, and crash frequency and severity. The first is a regression problem, the second, a classificatory one and the third can be approached either as a combination of the preceding two or as a regression model for the expected number of crashes by severity levels. The main groups of techniques used for these purposes are nearest neighbor classification, decision trees, evolutionary algorithms, support-vector machine, and artificial neural networks. The last one is used in many kinds of approaches given the ability to deal with both regression and classification problems, and also multivariate response models. This paper also presents the main performance metrics used to evaluate the models and compares the results, showing the clear superiority of the ML-based models over the statistical ones. In addition, it identifies the main explanatory variables used in the models, which shows the predominance of road-environmental aspects as the most important factors contributing to crash occurrence. The review fulfilled its objective, identifying the various approaches and

the main research characteristics, limitations, and opportunities, and also highlighting the potential of the usage of ML in crash analyses."