

# RAPIDS

Pålitligt adaptivt prediktivt underhåll och intelligent beslutsstöd



Författare: Olof Steinert, Erik Frisk, Mattias Krysander, Anders Vesterberg,  
Henrik Boström, Tony Lindgren, Oskar Andersson Reyna

Datum: 2025-10-31

Projekt inom Elektronik, mjukvara och kommunikation - FFI

## Innehållsförteckning

|                                                     |           |
|-----------------------------------------------------|-----------|
| <b>1 Sammanfattning .....</b>                       | <b>3</b>  |
| <b>2 Executive summary in English.....</b>          | <b>3</b>  |
| <b>3 Bakgrund.....</b>                              | <b>4</b>  |
| <b>4 Syfte, forskningsfrågor och metod.....</b>     | <b>5</b>  |
| <b>5 Mål .....</b>                                  | <b>7</b>  |
| <b>6 Resultat och måluppfyllelse .....</b>          | <b>8</b>  |
| <b>7 Spridning och publicering .....</b>            | <b>17</b> |
| 7.1 Kunskaps- och resultatspridning.....            | 17        |
| 7.2 Publikationer.....                              | 18        |
| <b>8 Slutsatser och fortsatt forskning .....</b>    | <b>20</b> |
| <b>9 Deltagande parter och kontaktpersoner.....</b> | <b>22</b> |

### Kort om FFI

FFI, Fordonsstrategisk forskning och innovation, är ett samarbetsprogram mellan staten och fordonsindustrin som sedan 2009 finansierar forskning och innovation inom vägtransporter.

Läs mer på [www.ffisweden.se](http://www.ffisweden.se)

## 1 Sammanfattning

Projektets parter var Scania CV AB, Linköpings universitet (Institutionen för systemteknik), Stockholms universitet (Institutionen för data- och systemvetenskap) och Kungliga Tekniska högskolan (Skolan för elektroteknik och datavetenskap). Projektet pågick mellan januari 2022 och september 2025 med en total budget på 19,5 MSEK.

RAPIDS har flyttat fram forskningsfronten inom prediktivt underhåll genom att utveckla nya maskininlärningsmetoder för strömmade och obalanserade data, med tillämpningar demonstrerade på verkliga industriella fallstudier. Projektet har bidragit till framsteg inom överlevnadsbaserad prognostik och osäkerhetskvantifiering för datadrivna modeller, samt operationaliserat beslutsstödsverktyg som översätter sannolikhetsfördelningar till kostnadsoptimerade underhållsplaner.

Projektet har även tagit viktiga steg mot effektiv återkoppling av hälsoinformation, genom att utveckla prototyper för semi-online modelluppdatering baserad på aktiv inläring (eng. active learning). Grundläggande processer har etablerats för att ta in ny utfallsstatistik, träna om modeller och följa upp resultat. Den operativa återkopplingsloopen behöver dock formaliseras och implementeras fullt ut för att säkerställa långsiktig prestanda och robusthet i drift.

På infrastrukturen har RAPIDS etablerat skalbara CI/CD-pipelines (eng. Continuous Integration / Continuous Delivery) för maskininläring, genomfört A/B-tester och skuggutvärderingar (eng. shadow testing) för att verifiera produktionsduglighet. Genom tydliga styrningsmekanismer, inklusive GDPR-kompatibel datadelning, anonymisering och opt-out-funktioner (rätt att invända mot att personuppgifter används), har projektet lagt en solid grund för ansvarsfull och hållbar AI inom fordonsindustrin.

RAPIDS har genererat betydande forsknings- och innovationsvärde i form av vetenskapliga publikationer, konferensbidrag, examensarbeten, öppen källkod och en publik datamängd publicerad via Swedish National Data Service (SND). Projektet har även finansierat den första delen av två doktorandprojekt och bidragit till att stärka samarbetet mellan industri och akademi inom området prediktiv modellering och dataanalys för tunga fordon.

## 2 Executive summary in English

RAPIDS (Jan 2022–Sep 2025) was a joint project between Scania CV AB, Linköping University, KTH Royal Institute of Technology, and Stockholm University to advance predictive maintenance for heavy vehicles. The project targeted three core challenges: learning from streaming and imbalanced data, quantifying and using prediction uncertainty in decisions, and building sustainable feedback loops and data/Machine Learning infrastructure that meet governance and regulatory requirements.

### *Objectives & Approach*

RAPIDS aimed to develop machine learning methods that turn logged and streaming sensor data into reliable, vehicle-level health predictions, integrate uncertainty into maintenance planning, and establish processes for continuous improvement. Work was organized into five packages:

project management (WP1), predictive models for streaming data (WP2), decision-making under uncertainty (WP3), feedback for data driven models (WP4), and architecture & data management (WP5). Academia led theory and method development; Scania provided use cases, data access, and industrial validation.

#### *Key Results*

WP2: New survival-based neural models for censored, imbalanced time series; techniques for handling non-uniform sampling; a framework combining low-frequency aggregates with high-frequency signals; open-source package PySaRe.

WP3: Open-source Python package *crepes* for conformal prediction; semi-online calibration; test martingales for drift detection; example-based explanations for random forests (open-source Python package *xrf*); decision support translating probability distributions to cost-optimized maintenance actions.

WP4: Principles and prototypes for feedback and active learning that prioritize high-value inspections when full testing is costly, enabling continuous model improvement.

WP5: Scalable CI/CD pipelines (incl. shadow/A-B testing), GDPR-compliant data governance (anonymization, opt-out), and a public dataset published via SND with a Nature data descriptor; repeatable processes for responsible external data sharing.

#### *Execution*

Despite a nine-month extension due to PhD recruitment and external pressures, the plan was largely followed. Eleven workshops, regular pulse/steering meetings, a public webinar, keynotes, and industry events ensured strong dissemination and knowledge transfer.

#### *Impact and Outlook*

RAPIDS successfully met its objectives and significantly advanced the state of the art in predictive maintenance for heavy vehicles. All partners report strong collaboration and effective knowledge transfer between academia and industry. The project strengthened the scientific foundation of survival modeling, uncertainty quantification, and interpretable AI, while demonstrating direct industrial relevance through Scania case studies.

The public dataset and data-sharing framework have positioned Scania as a leader in open and responsible industrial AI, fostering collaboration, attracting talent, and generating numerous external research outputs. RAPIDS also laid the groundwork for self-learning predictive systems through active learning and feedback integration.

Long-term, RAPIDS contributes to increased vehicle uptime, sustainability, and competitiveness in Swedish transport technology. Future research directions include multimodal data fusion, online drift detection, adaptive recalibration, and transfer learning for new components and control systems. The project's success illustrates the value of tightly coupled academic-industrial research in achieving impactful, data-driven innovation.

## **3 Bakgrund**

För att effektivisera transport av gods och människor krävs att fordon kan utnyttjas enligt plan och att driftstopp undviks. Oplanerade avbrott i trafiken kan leda till kostsamma konsekvenser såsom stillestånd, köbildning, förstörd last och i värsta fall olyckor. Möjligheten att förutsäga ett

fordons hälsostatus inom en relevant underhållshorisont skapar förutsättningar för effektivare underhållsplanering, ökad fordonstillgänglighet samt minskad risk för haverier resulterande i ett mer hållbart transportsystem.

Datadrivna modeller baserade på maskininlärning erbjuder ett resurseffektivt sätt att prognostisera hälsostatus för kritiska komponenter i ett fordon, exempelvis startbatteriet, givet dess specifika användningshistorik och konfiguration. Genom att använda stora mängder loggad fordondata kan modeller identifiera mönster och trender som indikerar begynnande degradering, och därigenom möjliggöra proaktiva underhållsåtgärder.

Ett grundläggande problem är dock att de händelser som modellerna försöker förutse inträffar sällan. Det leder till obalanserade datamängder och begränsad information om komponenters verkliga hälsostatus, vilket gör att prediktionerna blir osäkra. Denna osäkerhet måste kunna kvantifieras och vägas in i beslutsprocesserna för att säkerställa att åtgärder grundas på tillförlitliga bedömningar. I takt med att prediktiva komponentbyten blir vanligare minskar dessutom mängden återkopplad feldata från verkstäder, vilket försvårar validering och vidareutveckling av modellerna.

Samtidigt ökar tillgången till strömmad data från fordon i drift, där sensorer kontinuerligt rapporterar driftparametrar i realtid eller nära realtid. Detta öppnar nya möjligheter för förbättrad prognostik, men innebär också nya tekniska utmaningar. De ökade datamängderna kräver utveckling av metoder som kan hantera och uppdatera modeller kontinuerligt utan att fullständig rådata behöver lagras eller bearbetas i sin helhet. Den ökade datatakten ställer också krav på effektiv datahantering, filtrering och prioritering av vilken information som är mest värdefull för modellernas förutsäggelseförmåga.

För att uppnå robusta och långsiktigt hållbara prediktiva modeller krävs även väl utformade återkopplingsmekanismer. När en komponent byts ut, eller när underhåll utförs, behöver information om komponentens verkliga status och livslängd kunna återföras till modellen. Detta är ofta förenat med kostnader, både tekniskt och organisatoriskt, och kräver balans mellan nytta och resursåtgång. Effektiv återkoppling är särskilt viktig när tillgången till feldata minskar som följd av mer planerade underhållsåtgärder.

Därtill måste datahantering och modellutveckling ske i enlighet med gällande lagkrav och regleringar, inklusive krav på spårbarhet och personuppgiftshantering. Den ökande mängden fordondata, särskilt när den är strömmad, kräver att organisationer etablerar strukturer för datastyrning (data governance) som uppfyller både legala och etiska krav. Detta omfattar bland annat efterlevnad av dataskyddsförordningen (GDPR), säker hantering av kunddata samt transparens i hur data används och delas.

Mot denna bakgrund har projektet RAPIDS adresserat behovet av ny kunskap och metodik kring hur prediktiva modeller kan utvecklas, uppdateras och användas i praktiken för tunga fordon. Projektet har fokuserat på att hantera osäkerhet i modellernas prediktioner, utveckla metoder för att använda strömmade data effektivt, samt skapa processer för återkoppling av information som möjliggör kontinuerlig förbättring av modellerna. Samtidigt har de krav på tillförlitlighet, spårbarhet och dataskydd som följer av storskalig användning av fordondata i en industriell kontext beaktats.

## 4 Syfte, forskningsfrågor och metod

Projektet RAPIDS syftade till att öka kunskapen om och praktiskt tillämpa prediktiva modeller för tunga fordon, med fokus på maskininlärningsmetoder baserade på ökad tillgänglighet av

strömmade och loggade fordonsdata. Genom att utveckla metoder för att kvantifiera osäkerhet, förbättra återkoppling av data och integrera dessa modeller i beslutsprocesser för underhåll, skulle projektet bidra till en mer robust, effektiv och hållbar fordonsdrift.

Projektet avsåg att skapa ny kunskap kring hur prediktiva modeller kan användas på ett tillförlitligt sätt i industriella miljöer, där kraven på tillgänglighet och säkerhet ökar i takt med introduktionen av autonoma och elektrifierade fordon. Ett särskilt fokus låg på att utveckla teoretiska och praktiska ramverk för att hantera obalanserade och strömmade datamängder, samt att möjliggöra individanpassade beslut om komponentunderhåll baserat på modellernas sannolikhetssuppskattningar.

Ett ytterligare syfte var att främja samverkan mellan akademi och industri genom att kombinera teoretisk metodutveckling med praktiska tillämpningar. Projektet förväntades därigenom stärka svensk konkurrenskraft inom området intelligenta och tillförlitliga system, bidra till ökad fordonstillgänglighet, samt minska resursförbrukning och miljöpåverkan genom mer träffsäkra och data-drivna beslut om underhåll.

#### Forskningsfrågor

Projektets forskningsfrågor formulerades för att adressera de huvudsakliga vetenskapliga och tekniska utmaningarna kopplade till prediktivt underhåll i en kontext av snabbt ökande datamängder och krav på transparens. De övergripande frågorna var:

#### Strömmade data

- Hur kan metodik för prediktiva prognostikmodeller utvecklas för att hantera en förväntad ökning i datatakten från fordon, där mer data finns tillgänglig men inte i full upplösning och takt?

#### Osäkerhet

- Hur kan osäkerheten i prediktiva modeller skattas och hanteras på ett tillförlitligt sätt, särskilt för datadrivna modeller baserade på strömmade data?
- Vilken potential har metoder som conformal prediction för att kvantifiera osäkerhet i prognostikmodeller för strömmade data?

#### Beslut under osäkerhet

- Hur kan beslut, till exempel inom underhållsplanering, fattas med hänsyn tagen till den estimerade osäkerheten i modellernas prediktioner?

#### Återkoppling

- Hur kan information återkopplas på ett effektivt sätt för att möjliggöra kontinuerlig uppdatering och förbättring av de prediktiva modellerna över tid?

Projektet byggde på en iterativ och inkrementell metodik där teoriutveckling och praktisk tillämpning löpte parallellt. Arbetet organiserades i fem arbetspaket, som tillsammans täckte hela kedjan från datainsamling till beslutsstöd.

Lärosätena (LiU, KTH och SU) fokuserade på teori och metodikutveckling inom maskininlärning, osäkerhetskvantifiering och återkoppling, medan Scania bidrog med affärs- och problemförståelse, datatillgång samt verifikation av koncept på verkliga industriella fallstudier. Metoder och modeller utvecklades och utvärderades både på syntetiska dataset och på verkliga data från Scantias fordon. Fallstudierna användes för att testa modellernas tillförlitlighet, tolkningsbarhet och praktiska nytta i beslutsprocesser för underhåll.

Projektet hade en stark betoning på kunskapsöverföring. Resultat spreds genom vetenskapliga publikationer, examensarbeten, seminarier och ett öppet webbseminarium. Kod, data och

metoder tillgängliggjordes där så var lämpligt för att främja fortsatt forskning och industriell tillämpning.

## 5 Mål

RAPIDS avsåg att flytta fram forskningsfronten inom prognostik och hälsostatusövervakning för tunga fordon. Projektet hade som mål att bidra till ett mer hållbart transportsystem genom en mer optimal planering av fordonsunderhåll genom utveckling av maskininlärningsmodeller som kan förutsäga komponenters hälsostatus med högre precision och tillförlitlighet. Genom de tunga fordonens komplexitet förväntades projektet ge upphov till nya vetenskapliga och industriella frågeställningar, särskilt inom områdena osäkerhetskvantifiering vid censurerad och strömmade data samt återföring av hälsostatus. Fokus låg på att utveckla och integrera prediktiva modeller i beslutsprocesserna för underhåll, så att beslut kan fattas mer robust, effektivt och individbaserat för varje fordon.

En central del av målsättningen var att dra nytta av den ökade tillgången till strömmade och loggade fordonsdata, samt att utforska hur moderna metoder för analys av sekventiella data och tidsserier kunde tillämpas för planerat underhåll. Projektet skulle utveckla metoder för att uppskatta och hantera osäkerhet i prediktioner, samt undersöka hur ny information om komponenters hälsa kan återkopplas till modellerna för att successivt förbättra deras prestanda.

Projektet var uppbyggt runt fem arbetspaket med följande mål:

### 1. Arbetspaket 1 – Projektstyrning

Målet med arbetspaketet var att säkerställa en effektiv och strukturerad genomförandeprocess för projektet. Det innefattade operativ ledning, rapportering till Vinnova, koordinering av projektgruppen samt spridning av resultat genom presentationer och externa aktiviteter såsom webbseminarier.

### 2. Arbetspaket 2 – Prediktiva modeller för prognostik baserat på strömmade data

Målet var att utveckla metoder för att bygga och uppdatera prediktiva modeller som hanterar kontinuerligt strömmade och obalanserade fordonsdata. Arbetspaketet skulle fokusera på att förbättra modellernas noggrannhet, hantera osäkerhet i prediktioner och kombinera information från individnivå med populationsdata för mer robust prognostik.

### 3. Arbetspaket 3 – Datadrivet beslutsfattande under osäkerhet

Målet med arbetspaketet var att implementera och utvärdera ramverk för att kvantifiera osäkerhet. Eftersom industriella data ofta är censurerade var ett av delmålen att anpassa ramverken för sådan förekomst och bedöma värdet av dem empiriskt. Ett annat delmål var att implementera och utvärdera ramverk för att detektera förändringar av underliggande fördelningar i strömmade data. En central fråga för detta arbetspaket var hur den uppskattade osäkerheten kan användas för att fatta bättre beslut, dvs. hur de uppskattade sannolikhetsfördelningarna på bästa sätt kan användas vid beräkningen av kostnaden för olika alternativ.

### 4. Arbetspaket 4 – Återkoppling för datadrivna modeller

Målet för arbetspaketet var att undersöka hur man från utbytta komponenter kan erhålla information som maximerar kunskapen om komponenten för den datadrivna prediktiva modellen på ett så kostnadseffektivt sätt som möjligt. Detta innebar forskning kring hur och på vilket sätt man bör återkoppla information till en datadriven prediktiv modell, givet dess utsagor, i syfte att främja modellens prestanda. Som del i detta avsåg arbetspaketet

undersöka processer och organisation för att möjliggöra en sådan återkoppling, men även belysa tekniska frågeställningar kring återkoppling. Arbetspaketet fokuserade till stor del på de tekniska/algoritmiska delarna i hur återkoppling bör genomföras och mindre på process och organisation. Scania utvärderade nya tekniska mätverktyg för att förbättra återkopplingen men valde efter en utvärdering att inte gå vidare med mätverktyget, vilket gjorde att mindre fokus lades på process och organisationsfrågor.

#### 5. Arbetspaket 5 – Arkitektur och datahantering

Målet var att etablera en skalbar och hållbar dataarkitektur som möjliggör insamling, lagring och analys av strömmade fordonsdata i enlighet med lagkrav och god datastyrning.

Arbetspaketet syftade till att skapa en robust infrastruktur för modellträning, återkoppling och kontinuerlig förbättring av maskininlärningslösningar i produktionslik miljö.

Projektet hade dessutom som mål att bidra till forsknings- och kompetensutveckling genom anställning av två doktorander och en postdoktor. En ytterligare målsättning var att sprida forskningen och kunskapen från projektet genom vetenskapliga publikationer, konferenser, webbseminarium, koddelning och samverkan med andra aktörer.

Sammantaget var målet att skapa förutsättningar för mer träffsäkra, robusta och kostnadseffektiva prediktioner som möjliggjorde ett mer planerat, datadrivet och hållbart underhållsarbete för tunga fordon.

## 6 Resultat och måluppfyllelse

### 6.1 Arbetspaket 1 – Projektstyrning

Arbetspaketets huvudsyfte var att operativt leda projektet, samt ansvara för leveranser till Vinnova inklusive rapporter och presentationer. Totalt hölls elva workshops med projektgruppen, samt varannan vecka ett kortare pulsmöte. Varje workshop inleddes först med ett ledningsgruppsmöte och därefter presenterades och diskuterades delresultat och aktuella frågeställningar från varje arbetspaket. Vid workshoparna var det både industriella och akademiska presentationer. Mötena genomfördes huvudsakligen fysiskt i Södertälje.

Projekttiden förlängdes med nio månader eftersom rekryterings- och migrationsprocessen för de antagna doktoranderna tog nio månader längre än väntat samtidigt som Scania under samma period var under hård påfrestning på grund av införandet av nya lagkrav inom cybersäkerhet.

Utöver projektledning och koordinering omfattade arbetspaketet även ett betydande arbete med spridning av projektets resultat.

Presentationer och aktiviteter inkluderade:

- Presentation vid Data Innovation Summit (en av de mest inflytelserika årliga konferenser inom data och AI i Norden som samlar ledande experter och innovatörer inom området.) med fokus på tillämpningar av prediktivt underhåll.
- Inbjuden keynote vid SAFEPROCESS Symposium, en central akademisk konferens i området, där projektets metodik och resultat inom osäkerhetskvantifiering och modellering presenterades för en internationell publik.
- Monterutställning vid Scania Innovation Day, där projektets resultat visades upp för breda intressentgrupper inom Scania och Traton R&D.

- Inkorporering av projektets resultat i en sommarkurs inom maskininlärning och dataanalys, vilket bidrog till kunskapsspridning och kompetensuppbyggnad.
- Rundabordssamtal kring Combient-samarbete om industrins behov kopplade till prediktivt underhåll med särskilt fokus på data och datautläsning

Projektet avslutades med två större seminarier:

1. Ett publikt webbseminarium (2025-09-05) hölls med presentationer av projektresultaten, presentation samt paneldiskussion med experter inom utveckling tillgänglighetstjänster, datahantering samt prediktiv modellering inom fordonsbranschen från både akademien och industrin.
2. Ett slutseminarium på Scania (2025-09-30) med presentationer av projektresultaten (på en mer detaljerad nivå än vid det publika webbseminariet)

#### *Arbetspaketets bidrag till FFI:s mål*

Arbetspaketet har bidragit till FFI:s mål genom att säkerställa en effektiv projektstyrning, aktiv resultatspridning och bred kunskapsöverföring mellan industri och akademi. Projektkonstellationen har främjat samverkan mellan industri och akademi genom att lägga stort fokus på utbyte av kunskap och data. Genom presentationer vid internationella konferenser, publika seminarier och industrievenemang har projektets resultat gjorts tillgängliga för en större målgrupp och därmed stärkt svensk fordonsindustris position inom prediktivt underhåll, dataanalys och maskininlärning. Arbetspaketet har även bidragit till kompetensuppbyggnad och långsiktig kunskapsförankring genom utbildningsinsatser och forskarutbildning i linje med FFI:s övergripande mål om hållbar innovation och ökad konkurrenskraft.

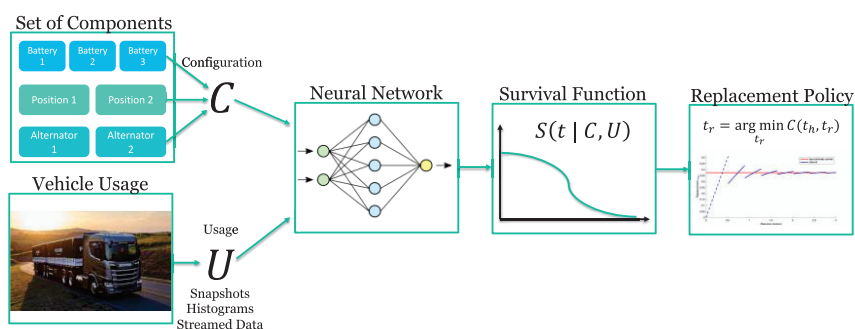
## **6.2 Arbetspaket 2 – Prediktiva modeller för prognostik baserat på strömmade data**

Grundfrågeställningen i detta arbetspaket har varit att utveckla metoder för effektiv inlärning av prediktiva modeller för prognostik baserat på tidsseriedata. Prediktivt underhåll ställer höga krav på modeller som inte bara kan förutsäga framtida komponenttillstånd, utan också hantera praktiska begränsningar i verkliga system. Två centrala egenskaper har studerats: att fel ofta inte inträffar under insamlingsperioden, vilket innebär att tidpunkten för en komponents livstid ej är känd, vilket resulterar i så kallad censurerad data, samt att insamlad data är inhomogent samplad och därmed svår att använda i standardiserade modeller och träningsmetoder.

För hantering av komplexa censurerade data har neurala nätverksmodeller som predikterar probabilistiska överlevnadsfunktioner utvecklats. Dessa modeller kräver inga fördelningsantaganden om komponenters livslängd, vilket gör dem attraktiva för komplexa dataset. Överlevnadsfunktioner beskriver sannolikheten för överlevnad och integrerar osäkerheter i prediktionerna samt möjliggör systematisk inkludering av censurerade data. Projektet har resulterat i energibaserade överlevnadsmodeller som hanterar censurering och visar konkurrenskraftiga resultat på både simulerade data och verkliga data från startbatterier i lastbilar (Holmer m.fl., 2023). Modellen har få hyperparametrar, vilket gör den lätt att applicera i industriella applikationer. Vidare har en familj av neurala nätverksmodeller med styckvis definierade risk- och täthetsfunktioner utvecklats, vilket utökar traditionella modeller. Simulerade dataset visar att dessa modeller uppnår liknande prediktiv prestanda som de energibaserade modellerna men med betydligt lägre beräkningstid (Holmer m.fl., 2024a). De utvecklade prognostikmodellerna är implementerade och tillgängliggjorda via Python-paketet PySaRe (PYthon package for neural network-based Survival Analysis and Reliability Engineering, <https://github.com/oholmer/PySaRe>). Alla dessa modeller kan utgöra

utmatningslager i neurala nätverk för statiska eller tidsseriebaserade data. En god egenskap med de här, mycket generella modellerna, är att de passar väl in i ett ramverk för optimal statistisk underhållsplanering och det finns även stöd för detta i programpaketet PySaRe. Se vidare beskrivning i arbetspaket 3.

En arkitektur för hur de olika komponenterna kan kopplas ihop för ett generellt prediktivt underhållssystem ses i Figur 1. Där ses hur strömmade data, tillsammans med mer statiska konfigurationsdata används som indata till en neural nätverksmodell. Det neurala nätverket estimerar en individanpassad överlevnadsfunktion (eng. survival function) som sedan ligger till grund för en statistiskt optimal underhållsplan.



Figur 1: Arkitektur för ett prediktivt underhållssystem.

För homogent samplade tidsseriesdata fungerar metoderna väl, medan inhomogen sampling kräver omsampling. Omsamlingsfrekvensen är en viktig parameter: fler samplingar ger bättre resultat men ökar datamängden och träningstiden. För att minska datastorleken har slumpmässig omsampling vid varje träningsepok implementerats. Utvärdering på simulerade data och ett experimentellt dataset med startbatterifel visar att homogen sampling ger korrekta modeller och att den föreslagna tekniken effektivt reducerar datamängden (Holmer m.fl., 2024b). Metodiken har utvärderats och visat lovande resultat på data från faktisk användning av uptime-kritiska fordonskomponenter på Scania, såsom startmotor, generator, startbatteri, NOx-sensor och turboladdare.

För att kombinera lågfrekvent aggregerade data med högfrekvent strömmade signaler har ett ramverk utvecklats. Genom att använda tidsfönsterbaserade modeller för lågfrekvent data och GHMM/LSTM för avvikelsetektering i högfrekvent strömmade data har kombinationen visat sig ge både tidiga indikationer om fel och detaljerad information nära haveri. Trots begränsad tillgång till strömningsdata är metoden lovande och skalbar (Nogarin, 2023).

För att hantera begränsad tillgång till högfrekvent drift-till-fel-data har en testbänksmodell för Li-jon-batterier utvecklats, som genererar syntetiska data med realistiska egenskaper, inklusive brus och censurering. En fördel med syntetiska data är att feltidpunkterna är kända, vilket möjliggör mer omfattande utvärdering av modeller. Denna data har använts för att testa klassiska modeller och avancerade neurala nätverk, vilket bekräftar att den är tillräckligt komplex för utveckling och utvärdering av överlevnadsmodeller. Modellen och datasetet finns tillgängliga på [https://github.com/Arezou-Safdari/synthetic\\_sequential\\_data\\_generation\\_model](https://github.com/Arezou-Safdari/synthetic_sequential_data_generation_model) för att stödja vidare utveckling (Arezou m.fl., 2024).

En särskild typ av tidsseriesdata är strömmade felkoder. Moderna fordon genererar ett stort antal potentiella felkoder, vilket leder till hög dimensionalitet, särskilt i kombination med långa tidsserier där felkoder kan aktiveras och deaktiveras över tid. För att hantera detta har metodik

utvecklats för att ta fram vektorrepresentationer (embeddings) som bevarar mycket av informationen från felkoderna men samtidigt reducerar dimensionaliteten avsevärt.

Delprojektet ADEAMIT (Anomaly Detection in Autonomous Mining Trucks) var inriktat på att tidigt upptäcka minskat däcktryck för autonoma gruvlastbilar. Målet var snabb detektion ombord med begränsade beräkningsresurser. Supervised maskininlärning (Logistisk regression, Random Forest) gav överlag bättre resultat. Unsupervised maskininlärning (GMM, VAE) var tränat enbart på normalt data och kunde ändå upptäcka avvikelser. Med hänsyn till realtidskrav och hårdvarubegränsningar framstår GMM som mest lämpad för att användas på fordonet, medan logistisk regression är det mest relevanta alternativet med kalibrerat testdata.

#### *Arbetspaketets bidrag till FFI:s mål*

Arbetspaketet har utvecklat metoder för prediktivt underhåll som minskar oplanerade driftstopp och ökar fordonsutnyttjandet, vilket stödjer FFI:s övergripande mål om effektivitet och hållbarhet. Inom delprogrammet Elektronik, mjukvara och kommunikation (EMK) bidrar resultaten främst till området Intelligent och tillförlitliga system genom datadrivna modeller för prognostik och dynamiska underhållstjänster. De nya neurala nätverksmodellerna för överlevnadsanalys, ramverket för kombination av låg- och högfrekvent data samt syntetiska testdata stärker digitaliseringstrenden och möjliggör ett uppkopplat transportsystem med hög tillförlitlighet. Sammantaget skapar projektet förutsättningar för systemvinster och ökad konkurrenskraft i linje med FFI:s strategiska färdplan.

Totalt utlovades att följande leveranser skulle ske från arbetspaketet:

- Vetenskapliga artiklar
- Exjobbssarbeten
- Programkod

Alla dessa leveranser har blivit genomförda enligt plan, i kapitel 7.2 i avsnitt "Arbetspaket 2 – Prediktiva modeller för prognostik baserat på strömmade data" följer en lista över vetenskapliga artiklar och examensarbeten som skapats inom ramen för arbetspaketet. Vidare har programpaketet PySaRe och kod för generering av syntetisk prognostikdata publicerats.

### **6.3 Arbetspaket 3 - Datadrivet beslutsfattande under osäkerhet**

Olika ramverk för kvantifiering av osäkerhet har studerats genom implementation av konforma klassificerare, regressorer och prediktiva system i det publika Python-paketet *crepes* (conformal classifiers, regressors and predictive systems, se <https://github.com/henrikbostrom/crepes>). En tidig version (utan konforma klassificerare) beskrivs i (Boström, 2022) medan en utökad version beskrivs i (Boström, 2024a).

Ett flertal versioner av paketet har släppts under den öppna källkodslicensen BSD - 3 clause, inklusive:

- version 0.6.0 (2023-06-28), som utökar ramverket till konforma klassificerare
- version 0.7.0 (2024-06-26) som bland annat inkluderar en klass för att skapa så kallade Mondrianska kategorier
- version 0.8.0 (2025-03-31) som bland annat inkluderar konforma prediktorer (conformal predictors) som genereras semi-online, dvs. för vilka kalibreringsmängderna uppdateras sekventiellt
- version 0.9.0 (2025-10-08) som bland annat inkluderar testmartingaler för att undersöka om det underliggande antagandet om utbytbarhet (exchangeability) gäller

Ett ramverk som ursprungligen var planerat att utvärderas var Vennprediktorer. Dessa resulterar i en mängd av sannolikhetsfördelningar för varje enskild prediktion, till skillnad från konforma prediktorer som genererar p-värden. De senare kan användas för att med statistiskt garanterad konfidens förkasta möjliga klass- eller regressionsetiketter. För att Vennprediktorer ska kunna användas i praktiken behöver mängden sannolikhetsfördelningar transformeras till en enskild fördelning, något som inte kan göras med bibehållna garantier. Fortsatta studier av detta ramverk beslöts lyftas ur projektet.

Omfattande studier av anpassning av konform prediktion till censurerade data genomfördes. Resultat från dessa presenterades i (Boström et al, 2023). Med utgångspunkt från dessa togs en teoretiskt välgrundad ansats fram, som dock bygger på ett antagande utöver utbytbarhet (exchangeability) som inte kan testas utan kännedom om verkliga målvärden, varför den praktiska nyttan kan ifrågasättas. Den implementerade ansatsen utvärderades empiriskt på både syntetiska data (där det tillkommande antagandet kan testas) liksom på en datamängd gällande en komponent på Scania-fordon, där dock censureringsgraden var så hög att värdet av den nya metoden var begränsat.

Delmålet att implementera och utvärdera ramverk för att detektera förändringar av underliggande fördelningar i strömmade data uppfyllades genom studier av så kallade konforma testmartingaler. Dessa används för att testa antagandet om utbytbarhet (exchangeability) utifrån p-värden som genereras online av konforma prediktorer, där sannolikheten för ett falskt alarm kontrolleras av en användardefinierad signifikansnivå. En undersökning av hantering av multipla sekvenser av p-värden, genererade från flera konforma prediktorer, genomfördes och nya ansatser för att kontrollera felnivån presenterades i (Boström, 2025). En omfattande undersökning av existerande ansatser och nya kombinationer av dessa presenterades i (Boström, in press). Sammanställning av praktiska testfall på Scania, bland annat gällande uppskattning av kvarvarande batteriladdning, har initierats men ännu inte färdigställts.

Ett bidrag till området stöd för beslutsfattande utgjordes av en ny ansats för att förklara en viss typ av prediktiv modell, så kallade random forests. Till skillnad från traditionella förklaringsansatser som analyserar enskilda variabels bidrag till en prediktion, så är den nya metoden exempelbaserad, där varje prediktion förklaras av en viktfordelning över samtliga träningsexempel (Boström, 2024b). Resultat från empiriska studier visar att antalet förklarande exempel kan begränsas i stor utsträckning utan att den prediktiva prestandan minskar. Ansatsen nytta är särskilt stor i de fall modellens variabler inte är enkelt tolkningsbara eller meningsfulla, vilket är fallet för t.ex. bildklassificering, och i de fall variablerna kan tolkas så kan den föreslagna ansatsen ge kompletterande information till de förklaringar som erhålls med traditionella förklaringstekniker. En implementation av metoden har gjorts publikt tillgänglig i form av Pythonpaketet xrf (Explainable random forests) under den öppna källkodslicensen BSD - 3 clause (se <https://github.com/henrikbostrom/xrf>).

De neurala nätverks-modeller som utvecklades i arbetspaket 2 möjliggör att moderna probabilistiska modeller kan användas inom etablerade ramverk för statistiskt beslutsfattande. I projektet har klassiska metoder inom åldersbaserat underhåll (eng. age-based replacement) från matematisk statistik utvärderats, och delvis integrerats i programpaketet PySaRe, på industriella data.

Ett praktiskt användningsfall inom projektet visade hur konform prediktion kan tillämpas för att hantera osäkerhet i diagnostik baserad på felkodsutläsningar. I vissa fall ger dessa utläsningar inte tillräcklig information för att en prediktionsmodell ska kunna göra säkra prediktioner. Genom att använda konform prediktion undersöktes därför hur säkerhetsnivån för varje prediktion kan uppskattas och om hur endast prediktioner med hög konfidens rapporteras.

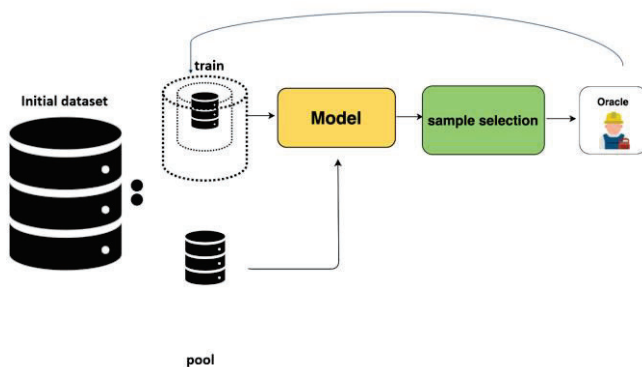
#### Arbetspaketets bidrag till FFI:s mål

Arbetspaketet har utvecklat metoder för att kvantifiera osäkerheten hos olika prediktiva modeller, t.ex. för att förutsäga komponenters kvarvarande livstid. Sådan kvantifiering är av vikt för att bättre kunna planera fordons underhåll, något som stödjer FFI:s övergripande mål om effektivitet och hållbarhet. Arbetspaketet har vidare bidragit till FFI:s mål gällande intelligenta och tillförlitliga system, där ansatserna för statistiskt beslutsfattande bidrar till de förra medan kvantifiering av osäkerhet bidrar till de senare. Utöver kvantifiering av osäkerhet så är även transparens en central aspekt för tillförlitlighet, t.ex. att det ska gå att förklara en modells prediktioner, något som har möjliggjorts via arbetspaketets bidrag inom exempelbaserade förklaringstekniker.

Arbetspaketets resultat i form av vetenskapliga publikationer och öppen källkod har i stort sett levererats i enlighet med den ursprungliga planen, se avsnitt 7.2 under rubriken "Arbetspaket 3 - Datadrivet beslutsfattande under osäkerhet" för en lista över vetenskapliga artiklar och konferensartiklar som arbetspaketet har bidragit med. Vidare har arbetspaketet producerat de två programpaketerna *crepes* och *xrf*, vilka har gjorts publikt tillgängliga.

#### 6.4 Arbetspaket 4 - Återkoppling för datadrivna-modeller

Detta arbetspaket har fokuserat på att undersöka hur man kan utforma återkopplingsloopar (eller aktivt lärande) för datadrivna prediktionsmodeller med syfte att förbättra modellernas prestanda över tid. En datadriven prediktionsmodell för underhåll kan misslyckas i sina förutsägelser på i huvudsak på två sätt 1) genom att förutsäga att en komponent bör bytas vid tidpunkt T men i verkligheten så fallerar komponenten innan tidpunkt T eller 2) genom att man byter komponenten vid tidpunkt T men att komponenten har lång livslängden kvar, dvs den blev ersatt för tidigt. I det första fallet så är det enkelt att återkoppla den faktiska tidpunkten för komponentens haveri till den datadrivna prediktionsmodellen. När det gäller det senare fallet så är det inte lika klart hur återkoppling skall göras till den datadrivna prediktionsmodellen och det är detta som arbetspaketet har undersökt.



Figur 2: Ett ramverk för att testa hur olika selektions metoder påverkar utfallet av de olika modellernas prestanda vid återkoppling av klassificerad data.

För att undersöka den återstående livslängd av en ersatt komponent så finns en kostnad associerad med undersökningen, detta gör att det typiskt inte är möjligt att undersöka alla ersatta komponenters kvarvarande livslängd. Här måste man alltså hitta metoder för att välja ut de fall som förbättrar den prediktiva modellen mest vid återträning med nya data. Detta är något

som vi undersökt i två vetenskapliga artiklar (Kharazian, Zahra, et al., 2023; Kharazian, Zahra, et al., 2024). I Figur 2, så visas ett ramverk för att simulera hur olika metoder för att välja ut vilka exempel som skall klassificeras av ett orakel för att sedan återkopplas till en ny mängd för träning. I början av simuleringen så undanhålls en del av datan i en Pool som selektions metoderna kan välja mellan (vilket simulerar att ett orakel klassificerar tidigare okända exempel), givet vilka exempel som väljs ut så kan nästa iteration av modellerna tränas, detta kan upprepas för att se hur de olika metoderna påverkar den slutgiltiga prestandan av de datadrivna modellerna. Även bilddata och datorseende har undersökts, med fokus på hur metoder som aktiv inlärning för objektigenkänning i videoströmmar och weakly supervised object detection (WSOD) kan minska behovet av manuella annoteringar och möjliggöra effektivare modellträning i industriella tillämpningar. Även nya sätt att återföra hälsostatusinformation från verkstadstexter har studerats. Det arbetet resulterade i ett nytt arbetssätt med mer strukturerad data definierad genom ett schema innehållande bl.a. version och diagnoskontext för hållbar och pålitlig modellförbättring. Genom att inkludera bl.a. bilddata från avvikelser och tidsserier av mätpunkter kan till större grad felmodeller identifieras och särskiljas mer effektivt. Däremot är den mängd sådan data som hittills hunnit samlas in begränsad, vilket innebär att den fulla effekten av denna återförda information på modellernas prestanda ännu inte kunnat utvärderas.

Vidare har arbetspaketet undersökt hur man skall modellera datadrivna prediktionsmodeller för att skapa modeller med bra prestanda. Här har vi undersökt hur man kan modifiera olika typer av traditionella prediktionsmodeller så de kan använda sig av censurerad information, dvs information om att vi observerat en komponent fram till en viss tid men att den inte har fallerat fram till en viss tidpunkt. Tre vetenskapliga artiklar har undersökt detta (Rahat, Mahmoud, et al., 2023; Sun, Kaiji, et al., 2023; Rahat, Mahmoud, and Zahra Kharazian, 2024). Vi har även i en vetenskaplig artikel (Kargar-Sharif-Abad, Monireh, et al., 2024) undersökt hur man kan göra prediktionerna från de datadrivna prediktionsmodellerna tolkningsbara för experter, detta genom användningen av SHapley Additive exPlanations (SHAP) som kan visualisera hur olika sensor värden kan öka eller minska en komponents skattade livslängd. Här kan dessa modeller också användas till att skatta nyttan undersöka olika typer av nya sensorer och därmed även möjliggöra beräkning av optimal sensorbestyckning.

#### *Arbetspaketets bidrag till FFI:s mål*

Höja den tekniska mognadsgraden. I arbetspaketet så har det utvecklats effektiva återkopplingsloopar för datadrivna prediktionsmodeller, vilket bidrar till FFI:s övergripande mål att höja den tekniska mognadsgraden inom området.

Arbetspaketets bidrag till följande delområden:

Vår forskning bidrar till intelligenta och tillförlitliga system genom att utveckla metoder för hur datadrivna prediktionsmodeller kontinuerligt kan förbättras över tid via återkopplingsloopar. Genom aktivt lärande säkerställs att modellerna blir mer robusta och kan fatta bättre beslut även under osäkerhet, vilket ökar systemens tillförlitlighet.

Genom användningen av tolkningsbara modeller (t.ex. SHAP) möjliggörs en effektiv människa-maskin-interaktion, där experter kan förstå och utvärdera modellernas beslut. Detta stärker förtroendet för de datadrivna systemen och gör det lättare för människor att samarbeta med och förbättra dessa modeller.

Slutligen bidrar forskningen till verifiering och validering genom att skapa metoder för att utvärdera modellernas prestanda under realistiska förhållanden och hantera censurerad data. På så sätt säkerställs att de prediktiva modellerna inte bara är träffsäkra utan även verifierade mot verkliga driftdata och tillförlitliga i praktisk användning.

Stora delar av kunskapsbidraget som arbetspaketet bidragit med sammanfattas väl i Zahra Khrazian halvtidsrapport, som hon presenterad, måndagen 24 mars 2025. Vid seminariet fanns åhörare från akademien och industrin och tack vare opponentens föredömliga arbete och Zahra bidrag, så ledde diskussionerna antagligen till att flera deltagare fick nya kunskaper och möjligtvis nya idéer för framtida forskning inom området.

Totalt utlovades att följande leveranser skulle ske från arbetspaketet:

- Vetenskapliga artiklar.
- Examensarbeten.
- Programkod.

## 6.5 Arbetspaket 5 - Arkitektur och datahantering

Arbetspaketet 5 har fokuserat på de tekniska och system-arkitekturella förutsättningarna för datadriven forskning och utveckling inom prediktivt underhåll. Fokus har legat på att skapa en skalbar och hållbar dataarkitektur för strömmade fordonsdata, säkerställa att lösningen är kompatibel med GDPR och arbeta med förutsättningar för kontinuerlig förbättring av maskininlärningsmodeller i en produktionsmiljö. Ett viktigt delmål har varit att förse övriga arbetspaket med relevant fordonsdata.

I arbetspaketet har man analyserat hur Scantias befintliga elsystem påverkar förutsättningarna för datainsamling som är lämplig för prognostikmodeller på en komponentnivå. Eftersom många tunga fordon använder CAN som standard för transportmedium mellan styrenheter i elsystemet så är bandbredden begränsad, vilket leder till ett krav på selektiv signalinsamling och downsampling.

Utöver arbetet med datainsamling har arbetspaketet utvecklat struktur och metodik för regelbunden träning och omträning av överlevnadsmodeller. Detta inkluderar framtagande av utvärderingsmått för överlevnadsmodeller på censurerad data, där traditionella metriker inte är tillräckliga, samt simulering av modellutfall för att bedöma den ekonomiska effekten av modellförbättringar innan produktionssättning. Den etablerade dataarkitekturen och utvärderingsmått som togs fram under detta arbete används nu som en grund i Scantias interna MLOps-plattformar och stöder fortsatt utveckling av datadrivna underhållslösningar i produktion.

Som en vidareutveckling av detta arbete har även LLMOps-principer etablerats för tjänster baserade på stora språkmodeller (LLM). Metodik har utvecklats för att systematiskt utvärdera modellkvalitet och övervaka känslighet mot variationer i indata. Detta möjliggör en kontrollerad och iterativ förbättringsprocess, motsvarande regressionstester i traditionell mjukvaruutveckling, vilket säkerställer stabilitet och långsiktig kvalitet i modellernas beteende.

En viktig leverans har varit utvecklingen av processer och verktyg för datasparbarhet, riskbedömning och anonymisering. Inom projektet togs en publik datamängd fram tillsammans med tillhörande processer och riktlinjer för ansvarsfull datadelning. Datamängden har anonymiserats, dokumenterats och tillgängliggjorts via Swedish National Data Service (SND), där både datamängd och beskrivning av datan finns publicerad. En data descriptor som beskriver innehåll, struktur och användningsområden har publicerats i Nature Journal, vilket bidragit till bred spridning och vetenskaplig synlighet. Datamängden är utformad för att stödja ett brett spektrum av forsknings- och utvecklingsaktiviteter, inklusive maskininläring, djupinläring, klassificering, regression, överlevnadsanalys och avvikelседetektering. Den har använts som grund för flera vetenskapliga publikationer och som underlag för en industriell tävling (industrial challenge) där deltagare fick utveckla och utvärdera nya metoder baserade på datan. Under

arbetet att ta fram den publika datamängden utvecklades processer och riktlinjer för hur Scania och projektets parter kan dela data externt på ett ansvarsfullt och repeterbart sätt. Dessa omfattar rutiner för anonymisering, juridisk och etisk granskning, kvalitetskontroll samt teknisk tillgängliggörande. Processerna har etablerats för att säkerställa efterlevnad av GDPR och dataspårbarhet och kommer att underlätta framtida publicering av ytterligare datamängder.

I projektet har ett flertal dataset tagits fram för bruk i övriga arbetspaket, dessa dataset har tagits fram på ett sätt som är både spårbart och säkert då vi har använt mekanismer från molnleverantörer som inneburit att inga filer behövt skickats mellan användare, utan datat har kunnat delas "at-rest". Man har även kunnat säkerställa att endast data från fordon som har tecknat avtal med Scania som tillåter dataanvändning för forskning och produktutveckling varit inkluderat genom ett kodbaserat filtersystem som exkluderar kunder som har "opt-out" från att dela data med Scania för syftet forskning och utveckling.

Arbetspaketet har även vidareutvecklat processer för återkoppling av hälsoinformation från verkstäder. Genom utökad datainsamling under felsökningssessioner har ny data gjorts tillgänglig för analys i molnet. Detta stödjer framtida modelluppdatering och felprediktion och kopplar till projektets övergripande mål om att etablera en kontinuerlig återkopplingsloop.

De flesta mål för arbetspaketet har uppnåtts enligt plan. En avvikelse var att tillgången till strömmade tidsseriedata från elektrifierade fordon försenades, vilket påverkade planerad användning för prognostikmetoder i övriga arbetspaket. Istället användes andra datakällor i snapshot-format, vilket visade sig ge värdefulla insikter och möjliggjorde parallell utveckling av pipeline och arkitektur.

Målet att göra tidsserie-data från elektrifierade fordon tillgänglig för analys kunde ändå uppnås mot slutet av projektet, vilket utgör en viktig grund för fortsatta aktiviteter inom både Scania och akademiska parter.

#### *Arbetspaketets bidrag till FFI:s mål*

Arbetet med att bygga en hållbar struktur för öppen och ansvarsfull datadelning mellan industri och akademi för utveckling och forskning inom prediktivt underhåll bidrar till FFI:s mål om att stärka Sveriges forsknings- och innovationskapacitet inom digitalisering och AI samt till FFI:s mål om ansvarsfull och hållbar digitalisering genom att kombinera tekniska lösningar med datastyrning i linje med GDPR och etiska krav.

Sammanfattningsvis har arbetspaketet uppfyllt sitt mål att etablera en robust, skalbar och hållbar dataarkitektur för prediktivt underhåll, samtidigt som det har bidragit till ökad forskningssamverkan, öppenhet och datadriven innovation inom svensk fordonsindustri.

# 7 Spridning och publicering

## 7.1 Kunskaps- och resultatspridning

Projektet har bedrivit en omfattande spridnings- och samverkansverksamhet mot både akademi och industri. Genom presentationer, rundabordssamtal, utbildningsinsatser och publicering av öppen källkod har resultaten nått breda målgrupper inom området för prediktivt underhåll, tillförlitlighet och maskininläring.

På industrisidan har projektet deltagit vid flera större evenemang, såsom Scania Innovation Day, Scania–KTH Strategic Partnership Day, samt i en rundabords-session arrangerad av Combient om utmaningar och behov inom prediktivt underhåll. Resultaten har också presenterats vid Data Innovation Summit och i ett gemensamt slutseminarium mellan Scania och TRATON R&D.

Projektet har haft en stark akademisk närvaro, bland annat genom keynote-presentationer vid den akademiska konferensen SAFEPROCESS'2024 i Italien, inbjudna föredrag vid internationella konferenser som Robotics: Science and Systems och COPA 2024, ICCDE 2024, SafeComp2024 samt genom universitetskurser i prognostik vid Linköpings universitet (<https://isy.gitlab-pages.liu.se/fs/en/phd-courses/prognostics/>). Projektet har även varit representerat vid workshops t.ex. vid Joint workshop SEC and BASE: System Control in Battery Production and Management.

Forskare inom projektet har även bidragit till den öppna forskningsgemenskapen genom att publicera flera Python-paket – bland annat crepes (<https://crepes.readthedocs.io/>), xrf (<https://xrf.readthedocs.io/>), PySaRe (<https://github.com/oholmer/PySaRe>) och SyntheticDataGeneration ([https://github.com/arezou-safdari/synthetic\\_sequential\\_data\\_generation\\_model](https://github.com/arezou-safdari/synthetic_sequential_data_generation_model)) – samt syntetiska data för prediktivt underhåll, vilket möjliggör fortsatt forskning och tillämpning. Projektet har dessutom haft populärvetenskaplig genomslagskraft genom en intervju i MotorMagasinet ([https://www.motormagasinet.se/article/view/1062141/spakulan\\_som\\_ska\\_fa\\_fordonen\\_att\\_leva\\_langre?token=vnayc5gbhjrse8c35190oavqjv2lqya](https://www.motormagasinet.se/article/view/1062141/spakulan_som_ska_fa_fordonen_att_leva_langre?token=vnayc5gbhjrse8c35190oavqjv2lqya)) och en artikel i IEEE Reliability Magazine om industriell dataåtkomst och dess utmaningar (<https://ieeexplore.ieee.org/document/11151816>).

Tillsammans visar dessa aktiviteter på ett brett och aktivt engagemang för att sprida kunskap, stimulera samverkan och stärka förståelsen för prediktivt underhåll och tillförlitlig maskininläring inom både akademi och industri

| Hur har/planeras projektresultatet att användas och spridas?   | Markera med X | Kommentar                                                                                                                                                                                                            |
|----------------------------------------------------------------|---------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Öka kunskapen inom området                                     | X             | Resultaten har under projekttiden spridits inom projektgruppen genom kontinuerliga arbetsmöten och externt genom en stor mängd akademiska publikationer och ett antal genomförda presentationer, kurser och webinar. |
| Föras vidare till andra avancerade tekniska utvecklingsprojekt | X             | Lärdomar från projektet har till viss del nyttjats inom avancerade tekniska utvecklingsprojekt och                                                                                                                   |

|                                                                      |   |                                                                                                                                                                  |
|----------------------------------------------------------------------|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                      |   | forskningsprojektet inom prediktivt underhåll, cirkulära affärsmodeller, diagnostik och hälsostatusmonitorering.                                                 |
| Föras vidare till produktutvecklingsprojekt                          | X | Resultaten används redan idag inom ett produktutvecklingsprojekt på Scania och Traton R&D och fler projekt är under konfigurering.                               |
| Introduceras på marknaden                                            | X | Scania kommer introducera tjänster som till viss del baseras på resultat från projektet. Resultat kommer också vidareutveckla och förbättra befintliga tjänster. |
| Användas i utredningar/regelverk/tillståndsärenden/ politiska beslut |   |                                                                                                                                                                  |

7.2 Publikationer

Projektets publikationer listats under respektive arbetspaket nedan:

Arbetspaket 2 – Prediktiva modeller för prognostik baserat på strömmade data

Vetenskapliga artiklar

Beikmohammadi, Ali, et al. "A cost-sensitive transformer model for prognostics under highly imbalanced industrial data." arXiv preprint arXiv:2402.08611 (2024).

Holmer, Olov; Frisk , Erik; Krysander, Mattias. "Energy-Based Survival Models for Predictive Maintenance". *IFAC-PapersOnLine*, 22nd IFAC World Congress, vol. 56, 2023, s. 10862–67. *ScienceDirect*, <https://doi.org/10.1016/j.ifacol.2023.10.762>.

[Holmer], Olov; Frisk, Erik; Krysander, Mattias. "Neural Network-Based Piecewise Survival Models". arXiv:2403.18664, arXiv, 27 mars 2024a. *arXiv.org*, <https://doi.org/10.48550/arXiv.2403.18664>.

Holmer, Olov; Krysander, Mattias; Frisk, Erik. "Usage-Specific Survival Modeling Based on Operational Data and Neural Networks". arXiv:2403.18739, arXiv, 27 mars 2024b. *arXiv.org*, <https://doi.org/10.48550/arXiv.2403.18739>.

Safdari, Arezou, Frisk , Erik; Holmer, Olov; Krysander, Mattias. "Synthetic Generation of Streamed and Snapshot Data for Predictive Maintenance". *IFAC-PapersOnLine*, 12th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2024, vol. 58, 2024, s. 270–75. *ScienceDirect*, <https://doi.org/10.1016/j.ifacol.2024.07.229>.

Examensarbeten

Fredriksson, Gabriel. *Prognostics for Condition-Based Maintenance of Electrical Control Units Using On-Board Sensors and Machine Learning*. 2022. [www.diva-portal.org](http://www.diva-portal.org), <https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-184235>.

Jayaraman, Vijay. "Enhancing failure prediction from timeseries histogram data: through fine-tuned lower-dimensional representations." (2023).

Nogarin, Leonardo. *Prognostics Models for Multi-Frequency Data Collection*. 2023.

Commented [MK1]: Otydligt att ref. denna och nästa artikel med (Holmer m.fl., 2024)

Commented [EF2R1]: Fixat

Sandberg, Elis. Second Life Applications for Degraded EV Batteries : Evaluating Benefits Based on Remaining Useful Life and Battery Configurations. 2023. [www.diva-portal.org](http://www.diva-portal.org), <https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-196013>.

Sun, Kaiji. "Robust Contrastive Learning and Multi-shot Voting for High-dimensional Multivariate Data-driven Prognostics." (2023).

### Arbetspaket 3 - Datadrivet beslutsfattande under osäkerhet

#### Vetenskapliga artiklar

Boström, H. (in press). "An Investigation of Conformal Test Martingales". *Enhancing the Learnability and Reliability of Machine Learning Algorithms. Essays Dedicated to Alexander Gammerman on his 80th Birthday*. Springer.

Boström, Henrik. "Crepes: A python package for generating conformal regressors and predictive systems." *Conformal and Probabilistic Prediction with Applications*. PMLR (2022): 24-41.

Boström, Henrik, Henrik Linusson, and Anders Vesterberg. "Mondrian Predictive Systems for Censored Data." *Conformal and Probabilistic Prediction with Applications*. PMLR (2023): 399-412.

Boström, Henrik. "Example-based explanations of random forest predictions." *International Symposium on Intelligent Data Analysis*. Cham: Springer Nature Switzerland (2024a): 185-196.  
Boström, Henrik. "Conformal prediction in Python with crepes." *Proceedings of Machine Learning Research* 230 (2024b): 236-249.

Boström, Henrik. "Testing Exchangeability for Multiple Sequences of P-values." *Proceedings of Machine Learning Research* 266 (2025): 1-18.

### Arbetspaket 4 - Återkoppling för datadrivna-modeller

#### Vetenskapliga artiklar

Kharazian, Zahra, et al. "Aid4hai: Automatic idea detection for healthcare-associated infections from twitter, a framework based on active learning and transfer learning." *International Symposium on Intelligent Data Analysis*. Springer Nature Switzerland, 2023.

Rahat, Mahmoud, et al. "Bridging the gap: A comparative analysis of regressive remaining useful life prediction and survival analysis methods for predictive maintenance." *Phm society asia-pacific conference*. Vol. 4. No. 1. 2023.

Rahat, Mahmoud, and Zahra Kharazian. "Survloss: A new survival loss function for neural networks to process censored data." *PHM Society European Conference*. Vol. 8. No. 1. 2024.

Kharazian, Zahra, et al. "CoPAL: Conformal Prediction in Active Learning An Algorithm for Enhancing Remaining Useful Life Estimation in Predictive Maintenance." *The 13th Symposium on Conformal and Probabilistic Prediction with Applications*. PMLR, 2024.

Sun, Kaiji, et al. "Robust contrastive learning and multi-shot voting for high-dimensional multivariate data-driven prognostics." *2023 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, 2023.

Kargar-Sharif-Abad, Monireh, et al. "SHAP-Driven Explainability in Survival Analysis for Predictive Maintenance Applications." *ECAI: European Conference on Artificial Intelligence, HAI15. 0: Embracing Human-Aware AI in Industry*. Vol. 5. 2024.

#### Examensarbeten

Lindfors, Martin. "Computer Vision for Internal Damage Assessment: Enhancing Reporting in Heavy-Duty Trucks: A Study on Weakly Supervised Localization and Hierarchical Classification in Real-World Troubleshooting Pipelines." (2025).

Srinivas Akshath, "Active learning for object detection of vehicle components" (2024)

Kargar-Sharif-Abad, Monireh, "Explainable Predictive Maintenance using Survival Analysis: A Remaining Useful Life prediction of heavy vehicle components" (2024)

San, Wenshi. "Explainable Anomaly Detection in Predictive Maintenance Using Shapelet Transform—Application in heavy vehicle components failure prediction." (2024).

Kargar Sharif Abad, Monireh. "Explainable Predictive Maintenance using Survival Analysis." (2024).

Montilla Tabares, Oscar. "Neural Networks for Predictive Maintenance on Highly Imbalanced Industrial Data." (2023).

#### Arbetspaket 5 - Arkitektur och datahantering

##### Vetenskapliga artiklar

Kharazian, Zahra, et al. "SCANIA component X dataset: a real-world multivariate time series dataset for predictive maintenance." *Scientific Data* 12.1 (2025): 493.

## 8 Slutsatser och fortsatt forskning

Alla parter i projektet är mycket nöjda med resultaten, samarbetet och den omfattande kunskapsöverföring som har skett mellan industri och akademi. Projektet har haft en tydlig förankring i faktiska behov och verkliga fordonsdata, vilket har gjort det möjligt att kombinera vetenskaplig forskning med industriella tillämpningar på ett djupgående och meningsfullt sätt. De väsentliga projektmålen, som beskrivs i kapitel 5, har helt eller delvis uppfyllts. RAPIDS har därigenom avsevärt stärkt forskningsområdet inom maskininlärning och prognostik, särskilt inom områdena konceptdrift, osäkerhetskvantifiering och tolkningsbarhet av modeller.

Projektet byggde vidare på de vetenskapliga och tekniska resultat som togs fram i det tidigare forskningsprojektet CODA, vilket möjliggjorde en snabb och effektiv start trots en initial utdragen rekryteringsperiod. Erfarenheterna visar att framtagning av nya datamängder är tidskrävande och måste påbörjas tidigt i projektets livscykel för att möjliggöra effektiv exekvering. Samarbetsformen mellan de deltagande parterna visade sig mycket framgångsrik. En viktig framgångsfaktor var att den teoretiska forskningen hade en tydlig industriell mottagare. Det möjliggjorde samutveckling av metoder som inte bara testades utan även anpassades till verkliga

användningsfall. Denna nära koppling mellan teori och tillämpning resulterade i högre relevans, snabbare kunskapsöverföring och större genomslag i industrin.

Sammantaget har RAPIDS flyttat fram forskningsfronten inom prediktivt underhåll för tunga fordon och skapat nya möjligheter att tillämpa maskininlärningsbaserade metoder i industriell drift.

Framtagna neurala nätverksmodellerna för överlevnadsanalys gör det möjligt att uppskatta sannolikheter för komponenters livslängd utan antaganden om underliggande fördelningar, vilket ger robusta prediktioner även för komplexa fordonsdata. Genom kombinationen av låg- och högfrekvent data i ett gemensamt ramverk har grunden lagts för skalbara lösningar som möjliggör tidiga felindikeringar. Resultaten, inklusive modeller och programvara publicerade i det öppna paketet PySaRe, har stärkt den vetenskapliga basen för datadriven prognostik inom fordonsindustrin.

Bidragen inom konform prediktion gör att prediktiva modeller kan förse med statistiska garantier för den tolererade felnivån. Ansatserna bygger på ett antagande om utbytbarhet (exchangeability), vilket kan testas med konforma testmartingaler. Nivån av falska alarm för signalering av att den underliggande fördelningen har förändrats kontrolleras med en användarspecifierad signifikansnivå.

Bidraget inom exempelbaserad förklaringsbar maskininläring tillåter att en modells prediktioner kan förklaras av de träningsexempel som modellen är genererad ifrån, något som kan vara av betydelse för att ge insikter och stödja beslutsfattande i de fall de variabler som modellen använder inte är direkt tolkningsbara. Förklaringarna som ges av den presenterade metoden kompletterar, och kan kombineras med, förklaringar som ges av traditionella variabelbaserade ansatser.

Uttecklade metoder för aktivt urval av återkopplingsdata och tolkningsbara modeller baserade på SHAP bidrar till ökad transparens och tillförlitlighet i beslutsstöd för underhåll. Framtagna teoretiska och praktiska principer för hur återkoppling och aktivt lärande kan användas för att förbättra datadrivna prediktionsmodeller över tid skapar en grund för självlärande och verifierbara prediktiva modeller i framtida driftmiljöer.

Ett av projektets mest framgångsrika resultat var framtagningen av en publik datamängd och etableringen av tillhörande processer och riktlinjer för ansvarsfull datadelning med externa användare. Datamängden skapades för att möjliggöra forskning och innovation inom prediktivt underhåll och maskininläring, samtidigt som högsta krav på dataskydd, anonymisering och regelefterlevnad uppfylldes. Den publika datamängden har fått stort genomslag inom data science-communityn och har redan resulterat i ett flertal vetenskapliga publikationer och nya tillämpningar utanför projektet. Den har etablerat Scania som en föregångare inom öppen och ansvarsfull industriell datadelning, vilket stärkt varumärket både inom forskningsvärlden, gentemot samarbetspartners och för att attrahera ny talang till både industri och akademi. Utöver själva datamängden har RAPIDS etablerat tydliga processer för framtida publicering av data. Dessa processer gör det nu betydligt enklare att på ett säkert och strukturerat sätt göra nya datamängder publika, vilket skapar långsiktiga förutsättningar för öppen innovation och fortsatt samverkan mellan industri och forskning.

Projektet har lagt en stabil grund för fortsatt forskning. Forskningsfrågor som uppkommit under projektets gång, men som inte har kunnat behandlas fullständigt, är exempelvis:

- Multimodal modellering: Hur flera datakällor – såsom loggad fordonsdata, tidsupplöst sensordata och bilder – kan kombineras i gemensamma modeller för att förbättra noggrannhet och robusthet i prediktiv prognostik.

- Online drift- och modelluppdatering: Hur datadrift kan upptäckas i realtid och modeller uppdateras adaptivt för att bibehålla prestanda när nya data kontinuerligt tillkommer.
- Hantering av nya komponenter och systemvarianter: Hur metoder kan utvecklas för att hantera situationer där nya komponenter introduceras, eller när kontroll- och loggningssystem förändras och därmed skapar nya datadistributioner. I sådana fall finns ofta begränsat eller inget historiskt data tillgängligt.
- Kontinuerlig återkoppling i drift: Även om initiala processer för att ta in ny utfallsstatistik, reträna modeller och följa upp effekter har etablerats inom RAPIDS, återstår behov av att undersöka hur dessa återkopplingsflöden kan formaliseras och utökas. Vidare krävs fördjupade studier av både organisatoriska och tekniska processer som möjliggör en effektiv och hållbar återkoppling i drift. Beslutssystem för när fysiska, potentiellt kostsamma tester bör genomföras för att facilitera återkoppling har hittills endast undersökts översiktligt och bör utvecklas vidare.

Alla dessa forskningsfrågor skulle kunna vara lämpliga för fortsatt forskning.

## 9 Deltagande parter och kontaktpersoner

Kontakta följande personer vid frågor gällande projektet:

- Arbetspaket 1 / Projektledare: Olof Steinert, Scania CV, [olof.steinert@scania.com](mailto:olof.steinert@scania.com)
- Arbetspaket 2: Prediktiva modeller för prognostik baserat på strömmade data, Erik Frisk, LiU, [erik.frisk@liu.se](mailto:erik.frisk@liu.se), Mattias Krysander, LiU, [mattias.krysander@liu.se](mailto:mattias.krysander@liu.se)
- Arbetspaket 3: Datadrivet beslutsfattande under osäkerhet, Anders Vesterberg, Scania CV, [anders.vesterberg@scania.com](mailto:anders.vesterberg@scania.com), Henrik Boström, KTH, [bostromh@kth.se](mailto:bostromh@kth.se)
- Arbetspaket 4: Återkoppling för datadrivna-modeller, Tony Lindgren, SU, [tony@dsv.su.se](mailto:tony@dsv.su.se)
- Arbetspaket 5: Arkitektur och datahantering, Oskar Andersson Reyna, Scania CV, [oskar.x.andersson@scania.com](mailto:oskar.x.andersson@scania.com)

