A network diagram consisting of numerous light blue circular nodes of varying sizes, connected by thin white lines. The nodes are scattered across the blue background, with some forming small clusters and others being isolated.

KWRW 2026.033 | Juni 2026

**Voordat ze vertrekken:**

**Ondersteunen van  
kennisoverdracht met  
machinaal leren**



# Colofon

## Voordat ze vertrekken: Ondersteunen van kennisoverdracht met machinaal leren

KWRW 2026.033 | Juni 2026

**Projectnummer**

404300/141

**Projectmanager**

Ralph Beuken

**Opdrachtgever**

KWR Waterwijs – Verkennend Onderzoek

**Auteur(s)**

dr.ir. J. (Joeri) Willet, dr.ir. D. (Dirk) Vries, S. (Sharon) Clevers MSc, dr. X. (Xin) Tian

**Kwaliteitsborging**

dr.ir. E.J.M. (Mirjam) Blokker

**Verzonden naar**

Dit rapport is verspreid onder deelnemers aan het KWR Waterwijs-programma.

**Openbaarheid**

Dit rapport is openbaar.

**Keywords**

Machine Learning, implicit knowledge, anomaly detection, water system design, knowledge transfer, knowledge management

**AI-verantwoording**

Voor dit document zijn LLM's (ChatGPT, Claude, en Copilot) gebruikt om te sparren over de structuur en de argumentatie en om een Python script te schrijven. De tekst is door de auteurs geschreven.

**Year of publishing**  
2026

**More information**  
Joeri Willet

**T** +31 30 606 9752  
**E** joeri.willet@kwrwater.nl

PO Box 1072  
3430 BB Nieuwegein  
The Netherlands

**T** +31 (0)30 60 69 511  
**E** info@kwrwater.nl  
**I** www.kwrwater.nl

The logo for KWR (Knowledge and Water Research Institute) consists of the letters 'KWR' in a bold, blue, sans-serif font.

Juni 2026 ©

All rights reserved by KWR. No part of this publication may be reproduced, stored in an automatic database, or transmitted in any form or by any means, be it electronic, mechanical, by photocopying, recording, or otherwise, without the prior written permission of KWR.

# Samenvatting

## Inleiding

Effectieve kennisoverdracht tussen senior en junior is waardevol nu een significant deel van de beroepsbevolking die betrokken was bij het ontwerp en bouw van drinkwaterinfrastructuur de pensioenleeftijd nadert. Concepten en methodes om deze kennisoverdracht te faciliteren zijn echter nog onderontwikkeld. In dit project is onderzocht in hoeverre technieken voor kunstmatige of artificiële intelligentie (AI), en specifiek machinaal leren (ML), gebruikt kunnen worden voor kennisoverdracht over distributiesystemen.

## Doel

Het doel van dit project is om de mogelijkheden voor kennisoverdracht ondersteund door ML te verkennen en om een “proof of concept” te ontwikkelen specifiek voor kennisoverdracht op het gebied van drinkwaterdistributie op basis van makkelijk toegankelijke data. Dit wordt gevold door aanbevelingen voor vervolg en voor daadwerkelijke implementatie.

## Aanpak

Voor dit onderzoek is in eerste instantie een summiere literatuurstudie gedaan naar AI en ML in het algemeen, en naar hun toepassing in de drinkwatersector. In parallel is een denkkader voor kennisoverdracht uitgedacht. Vervolgens is een iteratief proces een ML concept ontwikkeld op basis van zes stappen.

## Resultaten

De literatuurstudie laat zien dat AI en ML voor verschillende doeleinden gebruikt worden in de drinkwatersector, maar dat kennisoverdracht hier niet of nauwelijks deel van uitmaakt. Het ontwikkelde denkkader voor kennisoverdracht is gebaseerd op een onderscheid tussen verschillende soorten kennis: expliciete kennis (te documenteren en dus overdraagbaar) en impliciete kennis (persoonsgebonden en contextgebonden en daardoor moeilijk(er) te documenteren en over te dragen). De combinatie van beide soorten kennis levert het uiteindelijke ontwerp van drinkwaterinfrastructuur. Het ontwikkelde ML concept zoekt naar de sporen van impliciete kennis op basis van een datagedreven model en anomaliedetectie. Het is de verwachting dat dit een overzicht van afwijkende ontwerpkeuzes oplevert en is bedoeld als concreet aanknopingspunt voor gesprekken met senior experts. Het is tijdens deze gesprekken, waarin gespeculeerd wordt over oorzaak van de ontwerpkeuzes, dat kennisoverdracht daadwerkelijk plaatsvindt.

## Implementatie

Voor het ontwikkelde ML concept gebruiken we vooralsnog alleen relatief makkelijk toegankelijke data en adviseren we een getrapte implementatie. Op dit moment is het effect van impliciete kennis op het ontwerp van distributiesystemen vooral hypothetisch en anekdotisch. Door implementatie in toenemende stappen van detail en nauwkeurigheid kan gedurende het proces nagegaan worden of de hypothese dat impliciete kennis ontwerpbeslissingen beïnvloedt standhoudt net als de toegevoegde waarde ten opzichte van de nodige input.

# Inhoud

<b>Colofon</b>	<b>2</b>
<b>Samenvatting</b>	<b>3</b>
<b>Inhoud</b>	<b>4</b>
<b>1    <b>Introductie</b></b>	<b>5</b>
1.1    Aanleiding en doel	5
1.2    Aanpak en leeswijzer	6
<b>2    <b>Artificiële intelligentie en Machine Learning</b></b>	<b>7</b>
<b>3    <b>AI/ML-toepassing in de drinkwatersector</b></b>	<b>10</b>
3.1    AI/ML voor analyse, beheer, planning en ontwerp van stedelijke waterinfrastructuur	10
3.2    Literatuurstudie	11
3.2.1    Selectie van artikelen	11
3.2.2    Inzichten uit geselecteerde artikelen	13
3.3    Rol van AI in WDSA/CCWI conferentie	14
3.4    Interview	15
<b>4    <b>Expliciete en impliciete kennis</b></b>	<b>17</b>
<b>5    <b>Concept voor kennisoverdracht ondersteund door machinaal leren</b></b>	<b>18</b>
5.1    Beschikbare data	18
5.2    Probleemdefinitie	22
5.3    Evaluatie	23
5.4    Kenmerken/features	24
5.5    Modelkeuze	25
5.6    Experimenten	26
<b>6    <b>Aanbevelingen voor implementatie</b></b>	<b>28</b>
6.1    Testen van de impliciete-kennis-aanname en use-case	28
6.2    Trapsgewijze implementatie	28
6.3    Alternatieve verklaringen voor ontwerpafwijkingen	29
<b>7    <b>Conclusies</b></b>	<b>30</b>
<b>8    <b>Referenties</b></b>	<b>31</b>
<b>I    <b>Keywords WDSA/CCWI 2026 programmaboekje</b></b>	<b>32</b>

# 1 Introductie

## 1.1 Aanleiding en doel

Vergrijzing kan binnen de Nederlandse drinkwatersector zorgen voor een verlies van kennis en expertise. De drinkwaterbedrijven hebben dus een groot belang bij het overdragen van decennia aan kennis aan de nieuwe generatie waterexperts. Deze kennisoverdracht wordt beschouwd als een grote en nog onopgeloste uitdaging. De aanleiding voor dit verkennende onderzoek zijn de razendsnelle innovaties op het gebied van artificiële intelligentie (AI) waaronder machinaal leren (ML) en de mogelijkheid dat deze technologie ingezet zou kunnen worden voor kennisoverdracht, specifiek binnen de drinkwatersector.

Dit onderzoek is verkennend van aard en is gebaseerd op de hypothese dat de kennis binnen drinkwaterbedrijven (specifiek met betrekking tot het ontwerp, aanleg, en werking van drinkwaterdistributiesystemen op basis van waterbehoefte, leveringszekerheid, enz.) is ingebed in de data rondom de fysieke waterinfrastructuur die in de loop van decennia is gebouwd, vervangen, gesloopt en/of opnieuw opgebouwd. De aanname dat een dataset een weerspiegeling is van menselijke kennis wordt uitgelegd aan de hand van een minder positief voorbeeld in Box 1. Tegelijkertijd veronderstellen we dat AI-toepassingen voldoende ontwikkeld zijn om patronen in het ontwerp en vervanging van distributiesystemen te identificeren op basis van historische gegevens (zoals momentopnamen van het systeem over meerdere jaren, gemeentelijke/stedelijke ontwikkelingen, enz.). Op basis van deze patronen kan nagegaan worden op welke locaties er mogelijk extra aandacht nodig is bij vervanging of uitbreiding. Deze kennis bestaat dan niet enkel meer in het geheugen van werknemers met jarenlange ervaring binnen een specifieke regio. Dit kan van waarde zijn wanneer de watersector tegen 2030 een slordige 900.000 huishoudens extra moet bedienen ([Nationale Woon- en Bouwagenda | Home | Volkshuisvesting Nederland](#)).

Het beoogde resultaat van dit project is een onderbouwde inschatting van de mogelijkheden om met ML kennisoverdracht in de drinkwatersector te faciliteren. Specifiek in het kader van het ontwerp van drinkwaterdistributie. Vervolgens kan besloten worden of het ontwikkelen van een ML-model in vervolgonderzoek gerechtvaardigd is. Deze inschatting is gebaseerd op een inventarisatie van de mogelijkheden van AI in relatie tot het doel van kennisoverdracht over distributiesystemen, en een inventarisatie van de beschikbare data in relatie tot de data behoefte voor het gebruiken van een dergelijk ML-model.

### Box 1: ML voor recruitment bij Amazon (aangepast uit Harari (2024))

Tussen 2014 en 2018 werd er bij Amazon gewerkt aan een ML applicatie die zou helpen bij het identificeren van geschikte kandidaten voor een baan bij het bedrijf. Op basis van meerdere jaren aan data (motivatiebrief, CV, uitslag sollicitatiegesprek, ontslag, etc.) werd het algoritme getraind om nieuwe potentiële kandidaten een score te geven tussen de een en vijf sterren. Waar de programmeurs geen rekening mee gehouden hadden was dat het overgrote deel van de ingestuurde sollicitaties, de **trainingsdata** voor het algoritme, afkomstig was van mannelijke kandidaten. Dit kwam simpelweg omdat er historisch gezien meer mannen in de techsector werken/werkten dan vrouwen. De consequentie hiervan was dat het algoritme, uiteraard foutief, concludeerde dat mannelijke kandidaten beter geschikt waren voor de posities dan vrouwelijke kandidaten, en deze hierdoor automatisch een betere score toegekend werden. Uiteindelijk is het project stopgezet omdat deze bias lastig te verwijderen bleek.

Wat we uit dit voorbeeld kunnen leren is dat historische ontwikkelingen, patronen en voorkeuren (voornamelijk mannen in posities bij techbedrijven) ingebed kunnen zijn in datasets zonder dat deze expliciet hieraan toegevoegd zijn. De kracht van ML is dat het deze patronen in data kan herkennen en aan het licht brengen, waardoor we meer inzicht kunnen krijgen in de keuzes die in het verleden gemaakt zijn.

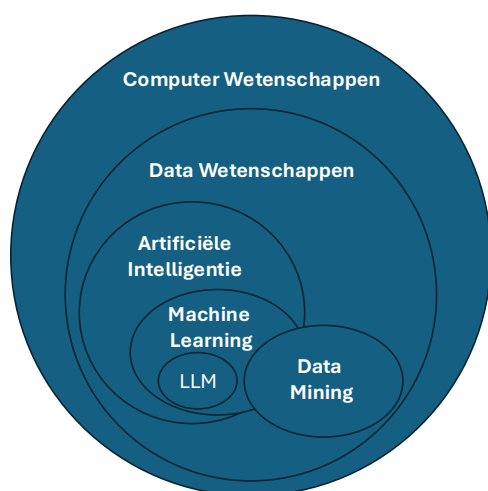
## 1.2 Aanpak en leeswijzer

In dit project zijn de volgende stappen gevolgd, die overeenkomen met de hoofdstukken in dit rapport.

- Beknopte introductie AI en ML (Hoofdstuk 2);
- Overzicht ML-mogelijkheden binnen de drinkwatersector (Hoofdstuk 3): er is een overzicht gemaakt van de mogelijkheden van ML op basis van literatuuronderzoek, een analyse van het programmaboekje van een conferentie, en een interview met onderzoekers met expertise op het gebied van AI/ML;
- Een conceptueel kader rondom kennis en kennisoverdracht is opgesteld op basis van dit overzicht (Hoofdstuk 4)
- De ontwikkeling van een ML concept voor kennisoverdracht binnen de drinkwatersector, specifiek op het gebied van distributie infrastructuur (Hoofdstuk 5) wordt beschreven;
- Aanbevelingen voor de implementatie van een dergelijke ML model (Hoofdstuk 6);
- Conclusies (Hoofdstuk 7).

## 2 Artificiële intelligentie en Machine Learning

In dit hoofdstuk introduceren we de benodigde basis om AI en ML te begrijpen in het kader van dit onderzoek. In het dagelijks leven lijkt de term artificiële intelligentie (AI) te staan voor Large Language Model (LLM) zoals ChatGPT, Copilot, of Claude. In werkelijkheid is het AI-veld veel breder en zijn LLM's eerder onderdeel van het sub veld machinaal leren. Het gebruik van data onderscheid ML van AI in het algemeen. Wanneer we uitzoomen naar het veld van computerwetenschappen dan kan daarbinnen de sub-discipline Data Wetenschappen geïdentificeerd worden, daarbinnen is het AI-veld te vinden, en weer daarbinnen het veld van Machine Learning (Theobald, 2017) (zie Figuur 1). Data mining is een aanverwante discipline waarin vergelijkbare methodes gebruikt worden, maar waarbij de focus ligt bij het ontwarren van grote datasets in plaats het creëren van modellen die het systeem beschrijven.

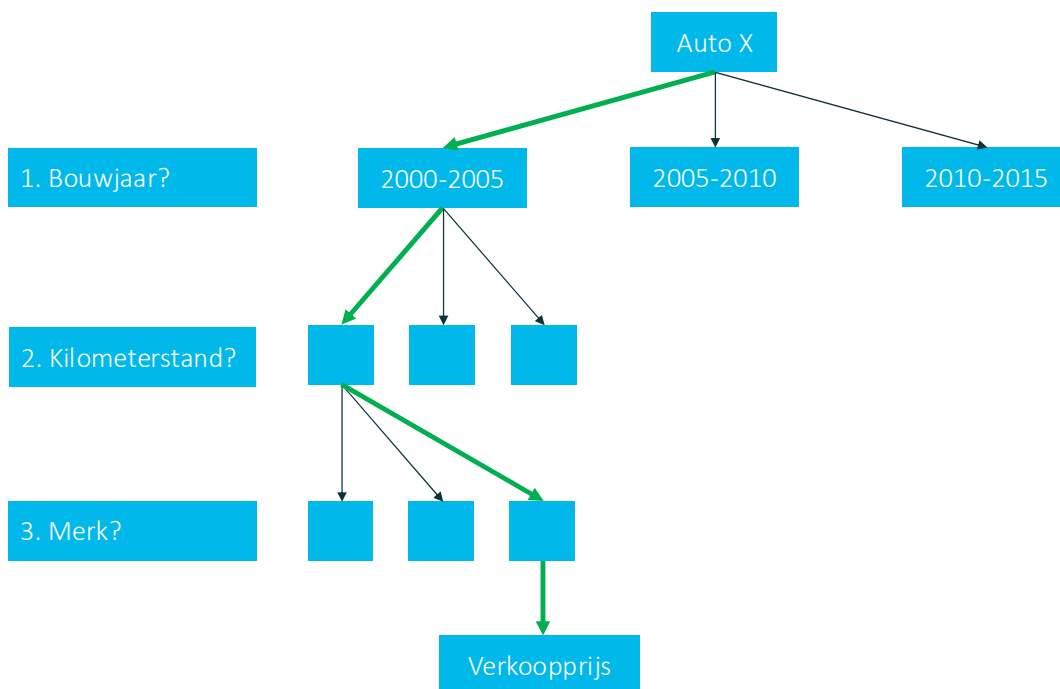


Figuur 1 Data-gerelateerde velden binnen computerwetenschappen, aangepast uit Theobald (2017)

In het kader van dit onderzoek is ML de juiste term voor de methodes die toegepast zouden kunnen worden op de historische infrastructuur data van drinkwaterbedrijven. In essentie is ML een verzameling methodes om patronen in data te herkennen zonder dat de onderliggende software hier specifiek voor geprogrammeerd is. Het is dus niet meer nodig om expliciete logica/regels te programmeren om een bepaalde output te genereren. Wanneer men in een programmeertaal (bijvoorbeeld Python) het commando `"2 == 3"` (wat neer komt op: 2 is gelijk aan 3) invoert zal deze als output `"False"` teruggeven. De achterliggende logica om deze output te genereren is expliciet geprogrammeerd. Binnen ML wordt deze expliciet geprogrammeerde logica echter vervangen door een model dat getraind is om patronen te herkennen in de input data om vervolgens een output te geven. We leggen dit verschil uit aan de hand van onderstaand voorbeeld.

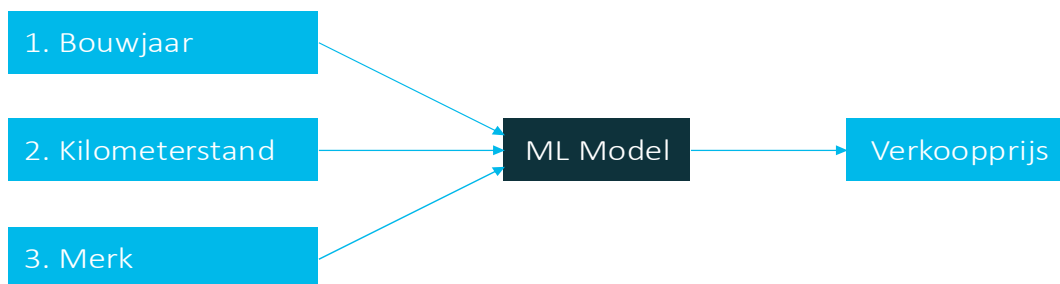
### Expliciet vs. impliciet geprogrammeerde logica: prijsbepaling auto's

Een autogarage wil het mogelijk maken voor nieuwe werknemers om snel een inschatting te maken over de verkoopwaarde van de tweedehandsauto's waarin zij handelt en overweegt hiervoor een software tool te ontwikkelen. Een optie is om collega's met veel ervaring te vragen naar de denkstappen die zij doorlopen bij het bepalen van de verkoopprijs en deze na te bootsen in een computerprogramma door deze expliciet te programmeren. Dit kan bijvoorbeeld door een beslisboom te maken waarbij er vragen beantwoord worden om tot een prijsinschatting te komen (zie Figuur 2). Het wordt al snel duidelijk dat er een grote hoeveelheid mogelijkheden expliciet doordacht en geprogrammeerd moeten worden om tot een uitkomst te komen.



Figuur 2 Voorbeeld van een expliciet geprogrammeerde beslisboom

Een alternatief is om op basis van historische data over de verkoopwaarde van tweedehandsauto’s in combinatie met hun eigenschappen (bouwjaar, merk, kleur, kilometerstand, etc.) een ML model te trainen dat zelf de verbanden tussen de kenmerken en de verkoopwaarde bepaalt (zie Figuur 3). Deze data heeft het bedrijf immers al ter beschikking omdat het in het verleden elke transactie geregistreerd heeft.



Figuur 3 Voorspelling van verkoopprijs op basis van een ML model

De achterliggende methoden binnen ML zijn uiteindelijk gebaseerd op statistiek en wiskunde om een model te maken waarin de input data gekoppeld is aan uitkomsten, en kan er dus eigenlijk niet echt gesproken worden over ‘leren’ (Andriy Burkov, 2019) of over intelligentie. Desondanks wordt de term ML breed toegepast in de praktijk. Over het algemeen is het echter onmogelijk voor mensen om de verbanden binnen een ML model te achterhalen en/of te begrijpen (Harari, 2024). Daarnaast zal een ML algoritme alleen (goed) werken binnen een nauwe bandbreedte van data, vergelijkbaar met de data waar het op getraind is.

Het trainen van een ML model wordt gedaan door verschillende statistische algoritmen toe te passen en/of te combineren, maar deze zijn over het algemeen onder te verdelen in drie hoofdcategorieën: **supervised learning** (begeleid leren), **unsupervised learning** (onbegeleid leren) en **reinforcement learning** (betrachtigend leren). Voor een introductie naar de verschillende algoritmen die toegepast kunnen worden binnen deze categorieën verwijzen we naar de teksten van Burkov en van Theobald (Andriy Burkov, 2019; Theobald, 2017).

**Supervised:** het identificeren van patronen in een dataset waarbij de correcte output (verkoopprijs van de auto) gekoppeld kan worden aan bekende eigenschappen van de input data (merk, bouwjaar, etc.). Een dataset waarin dit het geval is wordt omschreven als een ‘gelabelde’ dataset.

**Unsupervised:** het identificeren van verborgen patronen en labels in een dataset waarin niet alle relevante variabelen geïdentificeerd zijn. Een voorbeeld hiervan zou zijn om op basis van een dataset van kenmerken van auto's (motorinhoud, gewicht, etc.) een algoritme clusters te laten maken en achteraf na te gaan of deze overeenkomen met bepaalde merken. De hypothese hierbij zou zijn dat ieder merk voorkeuren heeft voor bepaalde eigenschappen en dat deze (zonder deze mee te geven aan het model) teruggevonden kunnen worden.

**Reinforcement:** voortdurende verbetering van het model door feedbackmechanismen te implementeren vanuit vorige iteraties. Over het algemeen wordt de output gecijferd op basis van wenselijkheid; positieve waarden voor wenselijke uitkomsten en negatieve waarden voor onwenselijke. Vervolgens leert het algoritme welke serie aan output de hoogste score opleveren onder verschillende omstandigheden. Een garage zou de prijs van hun wagens kunnen laten afhangen van veranderende omstandigheden zoals hun totale aanbod, de prijzen van hun naaste concurrenten, of zelfs op het uiterlijk van een klant. Het bepalen van de prijs op basis dit soort omstandigheden heeft niet alleen effect op de transacties van vandaag, maar ook op die van de toekomst. Het verkopen van een auto met korting vandaag zorgt ervoor dat je inventaris en onderhandelingspositie morgen anders is.

In Hoofdstuk 5 gebruiken we de volgende zes stappen van Bourke (2024) om een ML concept te ontwikkelen dat kan bijdragen aan kennisoverdracht, zie

Tabel 1 Algemene stappen voor het ontwikkelen van een ML concept inclusief voorbeelden

Stap	Beschrijving	Voorbeeld
1. Probleemdefinitie	Een zo duidelijk mogelijke definitie van het probleem dat met ML onderzocht zou kunnen worden	Voorspellen van de verkoopwaarde van tweedehandsauto's
2. Data	Een overzicht van de beschikbare data voor het probleem en een inschatting of een mens patronen in deze data zou kunnen ontdekken met genoeg tijd	De data over historische transacties van het autobedrijf, inclusief vraagprijs en uiteindelijke verkoopprijs. Deze weerspiegelen de keuzes die gemaakt zijn om tot een verkoopprijs te komen.
3. Evaluatie	Definitie van de criteria voor succes	Reële prijsindicaties met een foutmarge van minder dan 5%
4. Kenmerken / features	Kenmerken (categorische, numerieke, etc.) welke aanwezig in de data waarop getraind kan worden	Bouwjaar, merk, kilometerstand, kleur, accessoires, APK, etc.
5. Model	Keuze voor een model dat geschikt kan zijn voor het probleem dat we willen oplossen	Gelabelde data uit historische transacties maakt <b>supervised learning</b> mogelijk
6. Experimenten	Voorlopige implementatie om te onderzoeken wat wel/niet werkt om aan de succescriteria te voldoen	Iteratief proces van stap 1-5 waarin alternatieve modellen getest worden of door te experimenteren met de ratio tussen de training, validatie, en test data

## 3 AI/ML-toepassing in de drinkwatersector

De implementatie van AI/ML heeft de laatste jaren een vogelvlucht genomen binnen verschillende domeinen en zou kunnen bijdragen aan de digitalisatie van de watersector (Vekaria & Sinha, 2024). Vergeleken met andere sectoren (energie, gezondheid, transport) liep de implementatie van AI/ML achter in het waterdomein (Doorn, 2021). In dit hoofdstuk geven we een overzicht van het gebruik van AI/ML voor drinkwaterinfrastructuur.

### 3.1 AI/ML voor analyse, beheer, planning en ontwerp van stedelijke waterinfrastructuur

In de literatuur worden verschillende gebieden aangedragen waar AI/ML aan kan bijdragen. Volgens Doorn (2021) kunnen deze methoden in vier categorieën ingedeeld worden op basis hun inzet in de watersector: modellering, voorspellen en prognoses, beslisondersteuning en operationeel management, en optimalisatie. De verschillende typen analyse geïdentificeerd door Fu et al. (2023) kunnen hier vervolgens aan gekoppeld worden: Descriptieve Analytics, Diagnostische Analytics, Voorspellende Analytics, Prescriptieve Analytics. De gekozen methodes kunnen vervolgens bijdragen aan overkoepelende doelen als betrouwbaarheid, weerbaarheid en duurzaamheid (Fu et al., 2023).

#### Descriptieve Analytics

AI-gestuurde descriptieve analyses vatten historische gegevens samen om gedrag en prestaties van watersystemen in het verleden te begrijpen:

- AI-algoritmen analyseren grote datasets uit SCADA-systemen en bieden inzicht in systeemoperaties en prestatiepatronen.
- Dataminingstechnieken halen waardevolle informatie uit diverse bronnen, waaronder sensordata, klantfeedback en onderhoudsrapporten.
- Visualisatietools creëren intuïtieve weergaven van complexe watersysteemdata, wat patroonherkenning en besluitvorming ondersteunt.

#### Diagnostische Analytics

AI verbetert diagnostische mogelijkheden voor het identificeren van oorzaken van problemen binnen waterinfrastructuur:

- Anomaliedetectie-algoritmen signaleren ongebruikelijke patronen of gedragingen in waterkwaliteit, debieten of drukkiveaus.
- Graph Neural Networks (GNN's) modelleren complexe waternetwerken, waardoor potentiële storingspunten of inefficiënties snel kunnen worden geïdentificeerd.
- Machine learning-classificatoren categoriseren en prioriteren onderhoudsproblemen op basis van historische gegevens en deskundige kennis.

#### Voorspellende Analytics

AI-gestuurde voorspellende analyses voorspellen toekomstige gebeurtenissen in watersystemen:

- Deep learning-modellen voorspellen fluctuaties in watervraag, waardoor proactieve toewijzing van middelen mogelijk wordt.
- Tijdreeksanalyse-algoritmen voorspellen potentiële systeemstoringen, wat preventief onderhoud mogelijk maakt.
- Physics-Informed Neural Networks (PINN's) combineren data gestuurde benaderingen met fysische modellen om de voorspellingsnauwkeurigheid voor complex watersysteemgedrag te verbeteren.

## Prescriptieve Analytics

AI maakt prescriptieve analyses mogelijk om aanbevelingen te doen voor het optimaliseren van waterbeheer: Genetische algoritmen optimaliseren het ontwerp en de werking van watersystemen, rekening houdend met meerdere doelstellingen en beperkingen.

- Reinforcement learning-technieken ontwikkelen adaptieve regelstrategieën voor real-time beheer van waterdistributie.
- Beslissingsondersteunende systemen integreren door AI gegenereerde inzichten met deskundige kennis om investeringen in infrastructuur en beleidsbeslissingen te sturen.

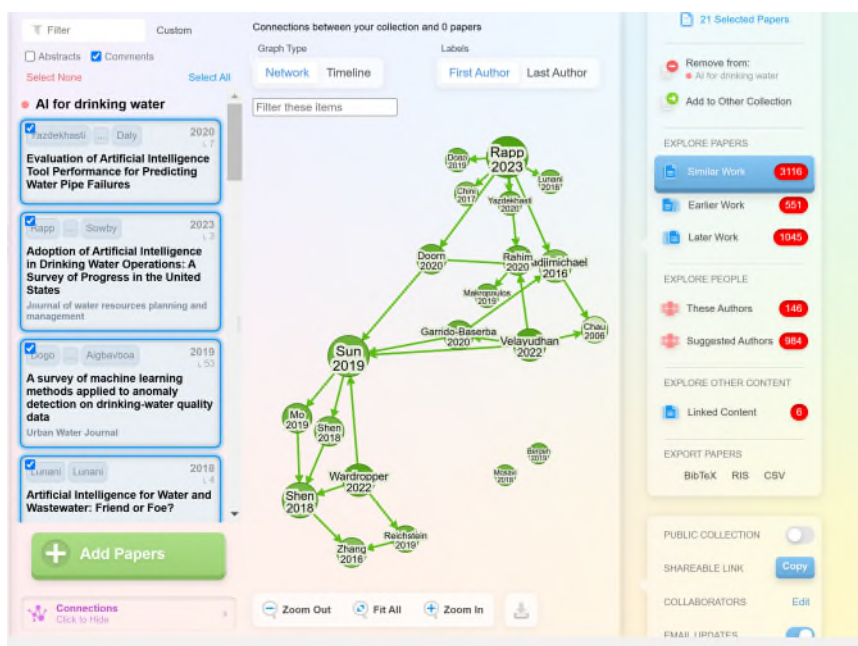
## 3.2 Literatuurstudie

### 3.2.1 Selectie van artikelen

Research Rabbit, een academisch platform om literatuur te doorzoeken, gebruikt om artikelen die zich richten op de toepassing van kunstmatige intelligentie (AI) in drinkwatersystemen te vinden en te analyseren. Dit platform biedt een intuïtieve manier om wetenschappelijke literatuur te verkennen via een interactieve, netwerk gebaseerde interface die verbanden tussen artikelen, auteurs en onderwerpen zichtbaar maakt. Bovenstaande is begin 2025 afgerond (en is dus niet compleet t/m verschijningsdatum) en was bedoeld om de onderzoekers te informeren m.b.t. de stappen in hoofdstuk 4 en 5.

Het startpunt van deze literatuurstudie was om de huidige rol van AI binnen drinkwatersystemen in kaart te brengen en van daaruit de zoektocht uit te breiden door gebruik te maken van de “Similar Work”-functie van Research Rabbit. Met deze functie kunnen nauw verwante artikelen ontdekt worden. Door te visualiseren hoe deze studies zich binnen het bredere onderzoekslandschap verhouden zijn de kernpublicaties geïdentificeerd en de voornaamste ontwikkelingen binnen het veld verkend.

Met behulp van Research Rabbit identificeerden we 22 relevante artikelen die zich richten op de toepassing van kunstmatige intelligentie (AI) in (drink)watersystemen (zie Figuur 4 en Tabel 2). De netwerkvisualisatie in de tool toont de relaties tussen deze studies en biedt inzicht in hoe verschillende werken verbonden zijn via gedeelde auteurs, citaties of vergelijkbare onderzoeksthema's.



Figuur 4 Overzicht van de literatuur binnen de Research Rabbit omgeving.

Tabel 2 Overzicht van geselecteerde artikelen

Titel	Auteurs	Tijdschrift
Artificial intelligence underpins urban water infrastructure of the future: a holistic perspective	Guangtao Fu, Siao Sun, Lan Hoang, Zhiguo Yuan, David Butler	Water
Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art	Liangpei Zhang, Lefei Zhang, Bo Du	IEEE Geoscience and Remote Sensing Magazine
IoT-Enabled Water Distribution Systems—A Comparative Technological Review	Nibi Kulangara Velayudhan, Preeja Pradeep, Sethuraman N. Rao, Aryadevi Remanidevi Devidas, Maneesha Vinodini Ramesh	IEEE Access
Decision-support systems for water management	Chloe B. Wardropper, Andrea E. Brookfield	Journal of Hydrology
A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists	Chaopeng Shen	Water Resources Research
HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community	Chaopeng Shen, Eric Laloy, Amin Elshorbagy, Adrian Albert, Jerad Bales, Fi-John Chang, Sangram Ganguly, Kuolin Hsu, Daniel Kifer, Zheng Fang, Kuai Fang, Dongfeng Li, Xiaodong Li, Wen-Ping Tsai	Hydrology and Earth System Sciences
Deep Convolutional Encoder-Decoder Networks for Uncertainty Quantification of Dynamic Multiphase Flow in Heterogeneous Media	Shaoxing Mo, Yin hao Zhu, Nicholas Zabaraz, Xiaoqing Shi, Jichun Wu	Water Resources Research
Machine learning for data-driven discovery in solid Earth geoscience	Karianne J. Bergen, Paul A. Johnson, Maarten V. de Hoop, Gregory C. Beroza	Science
Flood Prediction Using Machine Learning Models: Literature Review	Amir Mosavi, Pinar Öztürk, Kwok-wing Chau	arXiv: Learning
The fourth-revolution in the water sector encounters the digital revolution.	Manel Garrido-Baserba, Lluís Corominas, Ulises Cortés, Diego Rosso, Manel Poch	Environmental Science and Technology
Urban Hydroinformatics: Past, Present and Future	Christos Makropoulos, Dragan Savić	Water
Deep learning and process understanding for data-driven Earth system science	Markus Reichstein, Gustau Camps-Valls, Björn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, Prabhat	Nature
Artificial intelligence in the water domain: Opportunities for responsible use	Neelke Doorn	Science of the Total Environment
Where Are All the Data? The Case for a Comprehensive Water and Wastewater Utility Database	Christopher M. Chini, Ashlynn S. Stillwell	Journal of Water Resources Planning and Management

Do machine learning methods used in data mining enhance the potential of decision support systems? A review for the urban water sector	Antonia Hadjimichael, Joaquím Comas, Lluís Corominas	AI Communications
A review on integration of artificial intelligence into water quality modelling.	Kwok-wing Chau	Marine Pollution Bulletin
How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions	Alexander Y. Sun, Bridget R. Scanlon	Environmental Research Letters
Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review	Md Shamsur Rahim, Khoi Nguyen, Rodney A. Stewart, Damien Giurco, Michael Blumenstein	Water
Artificial Intelligence for Water and Wastewater: Friend or Foe?	Mahesh Lunani	Opflow
A survey of machine learning methods applied to anomaly detection on drinking-water quality data	Eustace M. Dogo, Nnamdi Nwulu, Bhekisipho Twala, Clinton Aigbavboa	Urban Water Journal
Adoption of Artificial Intelligence in Drinking Water Operations: A Survey of Progress in the United States	Alyson H. Rapp, Annelise M. Capener, Robert B. Sowby	Journal of water resources planning and management
Evaluation of Artificial Intelligence Tool Performance for Predicting Water Pipe Failures	Sepideh Yazdekhosti, Greta Vladeanu, Craig Daly	Pipelines 2020

### 3.2.2 Inzichten uit geselecteerde artikelen

De geselecteerde artikelen zijn gescand op (potentiële) toepassingen van AI in watersystemen, met als hoofddoel te begrijpen hoe AI-technologieën worden ingezet om uitdagingen in drinkwatersystemen aan te pakken. Denk hierbij aan uitdagingen als systeemoptimalisatie, anomaliedetectie, voorspelling van infrastructuurstoringen of monitoring van waterkwaliteit. We identificeren een aantal ontwikkelingen, methodologieën en kennislacunes in het huidige onderzoek. Er vallen een aantal thematische gebieden, zoals voorspellende modellering, Internet-of-Things (IoT)-integratie en beslissingsondersteunende systemen te onderscheiden. Daarnaast identificeren we ook dat de ethische en praktische uitdagingen van AI-implementatie aan bod komen in de literatuur; Garrido-Baserba et al. (2020) pleiten bijvoorbeeld voor ethische kaders en transparante praktijken. Onderstaande alinea's geven een globaal overzicht van terugkerende thema's uit de artikelen in Tabel 2, de thema's zijn tot stand gekomen op basis van een scan van de artikelen en een inschatting van de relevante thema's door de schrijver, maar pretenderen geen volledigheid.

#### *Voorspellende modellen voor storingen in waterinfrastructuur*

Yazdekhosti et al. (2020) geven een overzicht van het potentieel van AI-modellen aan om storingen in waterleidingen te voorspellen (voorheen onder de noemer van ML i.p.v. AI) door historische gegevens, omgevingscondities en leidingkenmerken te analyseren. Zij benadrukten de noodzaak van ontworpen kenmerken en robuuste evaluatiemaatstaven om precisie en betrouwbaarheid van het model te waarborgen. Deze aanpak helpt nutsbedrijven bij het prioriteren van onderhoud en het voorkomen van catastrofale storingen.

### **Lekdetectie en operationeel beheer**

Lekdetectie en storingen hebben historisch gezien veel aandacht gekregen vanuit de wetenschap. Rapp et al. (2023) onderzochten 49 grote Amerikaanse waterbedrijven en ontdekten dat 24% heeft geëxperimenteerd met AI voor taken zoals lekdetectie en systeemautomatisering. Hoewel veelbelovend, vormen barrières zoals kennistekorten en onzeker rendement op investering een obstakel voor grootschalige adoptie. Zij pleitten voor partnerschappen en casestudies om verantwoord AI-gebruik te bevorderen en operationele voordelen aan te tonen.

### **AI voor anomaliedetectie in waterkwaliteit**

Dogo et al. (2019) geven een uitgebreid overzicht van AI-methoden voor het detecteren van anomalieën in drinkwaterkwaliteitsdata. Technieken zoals support vectormachines en neurale netwerken werden beoordeeld op robuustheid en effectiviteit. Deze studie benadrukt het belang van hybride modellen voor verbeterde nauwkeurigheid.

### **AI in digitale watermeters en data-analyse**

Recente reviews, zoals die van Sun et al. (2019), benadrukken de rol van ML bij het benutten van de enorme hoeveelheid data die door digitale watermeters wordt gegenereerd. Door gedetailleerde analyses van waterverbruik mogelijk te maken, bevorderen deze technologieën besparingsinspanningen en bieden ze bruikbare feedback aan consumenten en nutsbedrijven.

### **Integratie van AI in beslissingsondersteunende systemen**

Wardropper et al. (2022) onderzochten de evolutie van water gerelateerde systemen, die data en modellen integreren om besluitvorming te ondersteunen. Deze systemen zijn cruciaal geworden nu klimaatverandering de variabiliteit van watervoorraden vergroot. Toch blijven er uitdagingen bestaan op het gebied van bruikbaarheid en de wensen van belanghebbenden.

### **Toepassingen van Big Data en AI in milieu en waterbeheer**

Sun et al. (2019) identificeerden het transformerende potentieel van big data en AI voor toepassingen zoals rampenbeheer en de optimalisatie van de water-energie-nexus. Zij benadrukken de integratie van deep learning met domeinspecifieke kennis om de prestaties en nauwkeurigheid van AI-modellen te verbeteren.

### **IoT-geïntegreerde waterdistributiesystemen**

Velayudhan et al. (2022) beoordeelden de integratie van IoT-technologieën in waterdistributie en stelden een IoT-architectuur voor real-time monitoring en controle voor. Dit raamwerk benadrukt duurzaamheid en efficiënt middelenbeheer in steeds complexere waternetwerken.

### **Naar een nationale waterdatabank**

Chini et al. (2017) roepen op tot de oprichting van een gecentraliseerde waterdatabank om onderzoek en beleidsvorming te verbeteren. Een dergelijke gegevensbron zou nationale studies over waterverbruik, energiegebruik en duurzaamheid mogelijk maken, vergelijkbaar met modellen die in de energiesector worden gebruikt.

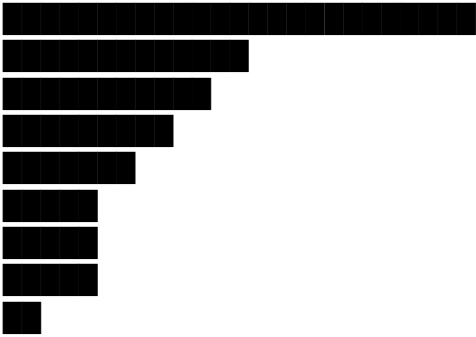
## **3.3 Rol van AI in WDSA/CCWI conferentie**

Naast de literatuurstudie is het programmaboekje van de WDSA/CCWI conferentie in 2026 ([WDSA-CCWI 2026](#)) een duidelijk voorbeeld van de mate waarin AI/ML aandacht begint te krijgen onder onderzoekers op het gebied van drinkwaterinfrastructuur. Voor dit onderdeel is gebruik gemaakt van claude.ai om:

1. Een overzicht te maken van de gebruikte termen in de 'keywords' van de abstracts (voor een overzicht zie Bijlage 0)
2. Een python script te schrijven dat elk abstract in het programmaboekje indeelt op basis van deze 'keywords' naar methodes, en toepassingsgebieden, en gebruik van AI-methodes.

Hieruit blijkt dat ruwweg 45% van de abstracts AI/ML methodes benoemt dan wel gebruikt. Een ander veelvoorkomend thema zijn digitale tweelingen (met of zonder AI) in 24% van de abstracts. Naast de inhoud van de abstracts was er een workshop gewijd aan 'AI voor waterprofessionals', waren er meerdere sessies waar AI centraal stond (bijvoorbeeld: Trustworthy Artificial Intelligence in Water Systems I&II), ging een keynote specifiek over ML, en wordt AI meerdere keren genoemd in de beschrijving van de afsluitende paneldiscussie.

De gebieden waar AI wordt toegepast op basis van de abstracts, in totaal 227, zijn voornamelijk (overlap tussen thema's is mogelijk):

Hydraulische simulatie / surrogaatmodellering	64	62.7%	
Waterkwaliteit	34	33.3%	
Digitale tweelingen	29	28.4%	
Lekdetectie	25	24.5%	
Waternvaagprognoses	19	18.6%	
Operationeel beheer	15	14.7%	
Wateroverlast / hemelwater / afwatering	14	13.7%	
Anomaliedetectie	14	13.7%	
Storingen / asset management	7	6.9%	

Kennisoverdracht, met of zonder AI, is terug te vinden in ongeveer 3% van de abstracts van WDSA/CCWI. Er lijkt, binnen deze onderzoeksgemeenschap, weinig aandacht voor dit onderwerp te zijn. Buiten deze gemeenschap wordt er wel expliciet aandacht besteed aan kennisoverdracht met AI/ML. Een voorbeeld hiervan is de 'Knowledge Summit' in Dublin, met de 2026 keynote 'Tacit Collective Knowledge – the New AI Frontier' (Tacite/Impliciete Collectieve Kennis – de nieuwe AI-grens) ([Speakers — Knowledge Management](#)). Er zijn dus mogelijke aanknopingspunten buiten de drinkwatersector. We komen terug op het concept van tacite/impliciete kennis in Hoofdstuk 4.

### 3.4 Interview

Naast de literatuurstudie zijn twee onderzoekers (Prof. Dragan Savic en Lydia Tsiami) met expertise op het gebied van AI/ML geïnterviewd. Doel van het interview was om een voorlopige inschatting te maken over de mogelijkheden voor het gebruik van AI bij het identificeren van ontwerpregels binnen stedelijke drinkwater infrastructuur welke vervolgens benut kunnen worden voor kennisoverdracht. Wat volgt zijn de kernpunten van het gesprek.

- AI en ML algoritmes kunnen gebruikt worden om een inschatting te maken van de historische data die de meeste invloed hebben op het ontwerp van een drinkdistributiewatersysteem.
- Wanneer ontwerpregels worden gedefinieerd in samenspraak met drinkwaterexperts dan is het mogelijk om na te gaan of deze weerspiegeld worden in werkelijke netwerken met behulp van AI en clusteringalgoritmes.
- Hoogstwaarschijnlijk zal het nodig zijn om in de datavoorbereiding de aansluitleidingen te verwijderen omdat er anders te veel details aanwezig zijn waar het model niet mee om kan gaan.
- Bij het implementeren van een AI-model dat getraind is op historische drinkwaternetten zal het van belang zijn om het juiste aggregatieniveau van de invoerdata te behalen. Bij te veel details zullen er wellicht geen patronen herkend kunnen worden en kan een algoritme niet getraind worden.

Naast deze kernpunten kwam ook duidelijk naar voren dat het definiëren van een duidelijk probleem om op te lossen op basis van een ML-model lastiger kan zijn dan het op eerste gezicht lijkt. Het gaat hierbij om de vraag: Wordt de keuze voor een ML-model en methode gebaseerd op de beschikbare data, of het doel dat we willen bereiken? Dit vraagstuk speelde ook een rol tijdens de rondvraag naar data bij de drinkwaterbedrijven, en is verder beschreven in Hoofdstuk 5. Los van de data of het doel is het voor kennisoverdracht belangrijk om eerst een onderscheid kunnen maken tussen de soorten kennis welke overgedragen kunnen worden (Hoofdstuk 4).

## 4 Expliciete en impliciete kennis

In het kader van kennisoverdracht rondom het ontwerp van drinkwaterdistributiesystemen is het van belang om onderscheid te maken tussen de verschillende soorten kennis die hierbij komt kijken. In dit project definiëren we dit onderscheid niet langs de lijnen van domeinen of expertisevelden maar langs de lijnen van persoonsgebondenheid — dat wil zeggen: de mate waarin kennis losgemaakt kan worden van de persoon, relatie of context waarin zij ontstaan is. Hiervoor gebruiken we de twee hoofdcategorieën die gedefinieerd worden in de literatuur rondom kennisoverdracht, expliciete en impliciete kennis (Falckenthal et al., 2025).

**Expliciete kennis:** is onafhankelijk van de situatie of context, waardoor het de status van ‘algemene kennis’ bereikt heeft en kan bestaan in normen, ontwerptekeningen of rekenmodellen, onafhankelijk van een specifieke medewerker. Expliciete kennis staat op zichzelf en kan relatief makkelijk gearticuleerd, gecodificeerd, opgeslagen en gedeeld worden doormiddel van verschillende media zoals documenten, boeken, of datasets. Een duidelijk voorbeeld van de overdracht van expliciete kennis binnen de drinkwatersector zijn de Praktijk Codes Drinkwater (de PCD's) (KWR, 2026).

**Impliciete kennis:** blijft verbonden met de persoonlijke interpretaties van een individu in een bepaalde context. Impliciete kennis is dus persoonlijk, is gebaseerd op ervaring en verandert continu naarmate expliciete kennis wordt toegepast in verschillende situaties. Het is de kennis van een medewerker die weet wanneer een hydraulisch model van een voorzieningsgebied “klopt” zonder dat volledig te kunnen formaliseren, of die de organisatorische en bestuurlijke context van een specifiek distributienetwerk begrijpt op basis van jaren van interactie met beheerders en andere stakeholders.

Het is lastig (dan wel onmogelijk) om impliciete kennis op te slaan (met of zonder AI/ML) en over te dragen omdat deze afhankelijk is van de specifieke situatie en/of de betrokken actoren. Tegelijkertijd is enige vorm van overdracht van impliciete kennis waardevol in een context waar besluiten niet enkel gebaseerd worden op expliciete kennis maar waar deze kennis wel moeilijker behouden blijft in een organisatie vanwege vergrijzing. Het is nog onduidelijk in hoeverre de verschillende soorten kennis een rol spelen/speelden binnen ontwerpkeuzes voor drinkwaterdistributie. In dit onderzoek werken we met de hypothese dat impliciete kennis dusdanig van belang is/was dat deze een effect heeft op de ontwerpkeuzes die gemaakt zijn in het verleden.

Onderzoek naar de overdracht van impliciete kennis met behulp van AI/ML wordt al wel uitgevoerd maar nog niet specifiek binnen de watersector (Falckenthal et al., 2025). In het volgende hoofdstuk beschrijven we een ML concept waarmee we verwachten dat het mogelijk wordt om de fysieke locaties/situaties in het drinkwaterdistributiesysteem waar impliciete kennis een significante rol speelde in het ontwerp/besluitvorming van te identificeren en het mechanisme om daarmee bij te dragen aan kennisoverdracht.

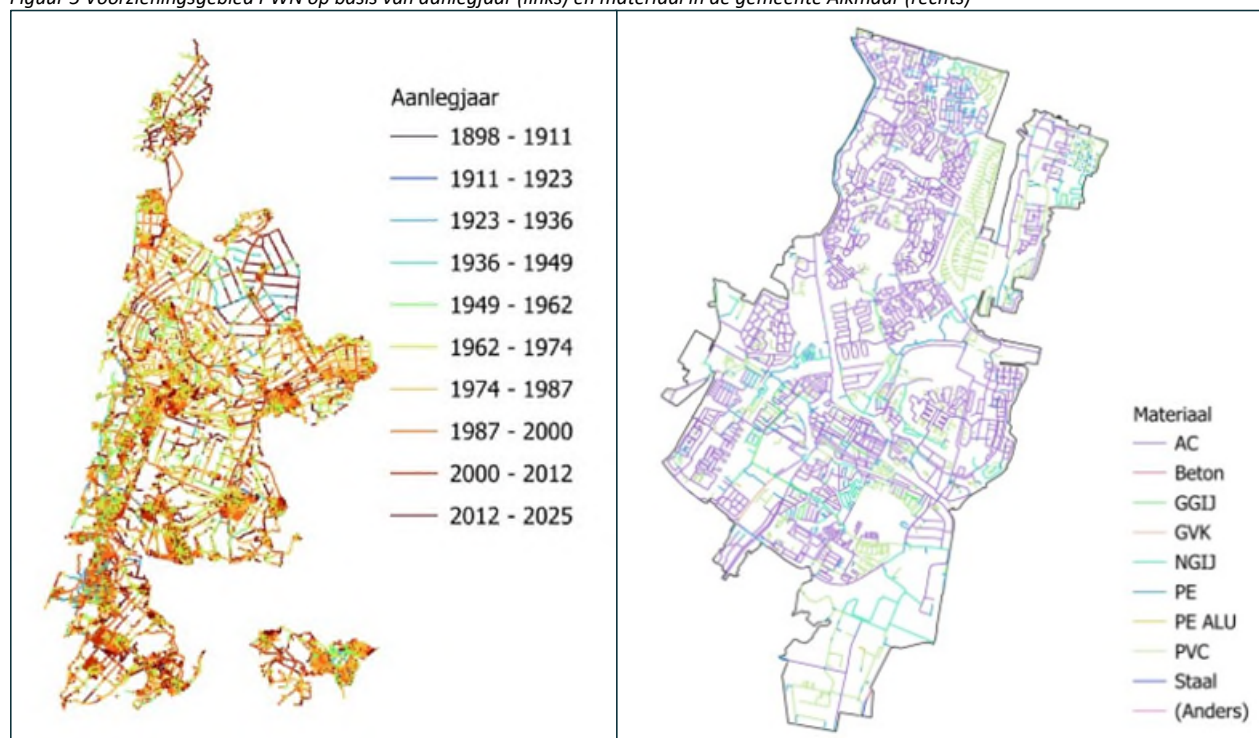
## 5 Concept voor kennisoverdracht ondersteund door machinaal leren

In dit hoofdstuk worden de stappen voor de ontwikkeling van een ML concept voor kennisoverdracht binnen de drinkwatersector op conceptueel niveau doorlopen (Probleemdefinitie, Data, Evaluatie, Kenmerken/features, Model, Experimenten). Zoals bleek uit het interview is het onmogelijk om op een lineaire manier deze stappen te doorlopen. Het opstellen van een probleemdefinitie vergt ook inzicht in de beschikbare data. Tegelijkertijd is het opstellen van een probleemdefinitie, om daar vervolgens de juiste data bij te zoeken of te genereren, op dit moment geen werkbare aanpak voor de drinkwaterbedrijven omdat de toegevoegde waarde van ML nog te onduidelijk is. In dit hoofdstuk beschrijven we eerst de (makkelijk) beschikbare data en op basis daarvan een probleemdefinitie die kan bijdragen aan de overdracht van (impliciete) kennis. Vervolgens worden de vier andere stappen in een ML concept doorlopen. Op basis van de conceptuele aanpak en een inschatting van de haalbaarheid die hieruit volgt kan besloten worden om het ML concept daadwerkelijke uit te voeren in een vervolgproject.

### 5.1 Beschikbare data

Voor dit onderzoek hebben we data over het voorzieningsgebied van PWN ontvangen (zie Figuur 5). Deze dataset geeft een globaal overzicht van de manier waarop het netwerk van PWN over de jaren ontwikkeld is doordat de eigenschappen van de verschillende segmenten gekoppeld kunnen worden aan het aanlegjaar. De mogelijkheid dat sommige segmenten niet meer origineel zijn en dat deze vervangen zijn wordt echter niet weergegeven in deze dataset, ook al zijn de aangrenzende segmenten vaak wel indicatief. Denk bijvoorbeeld aan de situatie waarbij een segment met een jongere leeftijd tussen segmenten met een oudere leeftijd ligt. In een dergelijke situatie valt te verwachten dat dit een vervanging vanwege een storing o.i.d. betreft maar valt niet te achterhalen of er opnieuw nagedacht is over ontwerpkeuzes zoals de diameter. Het verkrijgen van informatie over deze vervangingen is op dit moment niet haalbaar en kiezen we voor een aanpak welke hiermee kan omgaan.

Figuur 5 Voorzieningsgebied PWN op basis van aanlegjaar (links) en materiaal in de gemeente Alkmaar (rechts)



Binnen deze dataset zijn de eigenschappen van het distributiesysteem gekoppeld aan lijnsegmenten die staan voor de leidingen in de ondergrond. De geregistreerde eigenschappen voor de individuele lijnsegmenten zijn:

1. ID: een unieke code gekoppeld is aan de leiding.
2. Functie: Brandkraanleiding, Distributieleiding, Inwinningsleiding, Productieleiding, Ruwwaterleiding, Transportleiding.
3. Materiaal: AC, Beton, Beton/PE, GGIJ, GVK, Koper, NGIJ, PE, PE ALU, PVC, RVS, Staal.
4. Nominale diameter: 53 diameters tussen 15 mm en 1600 mm.
5. Datum aanleg
6. Wanddikte
7. Materiaalspecificatie
8. Lengte
9. Geometrie: de ligging en vorm van het lijnsegment

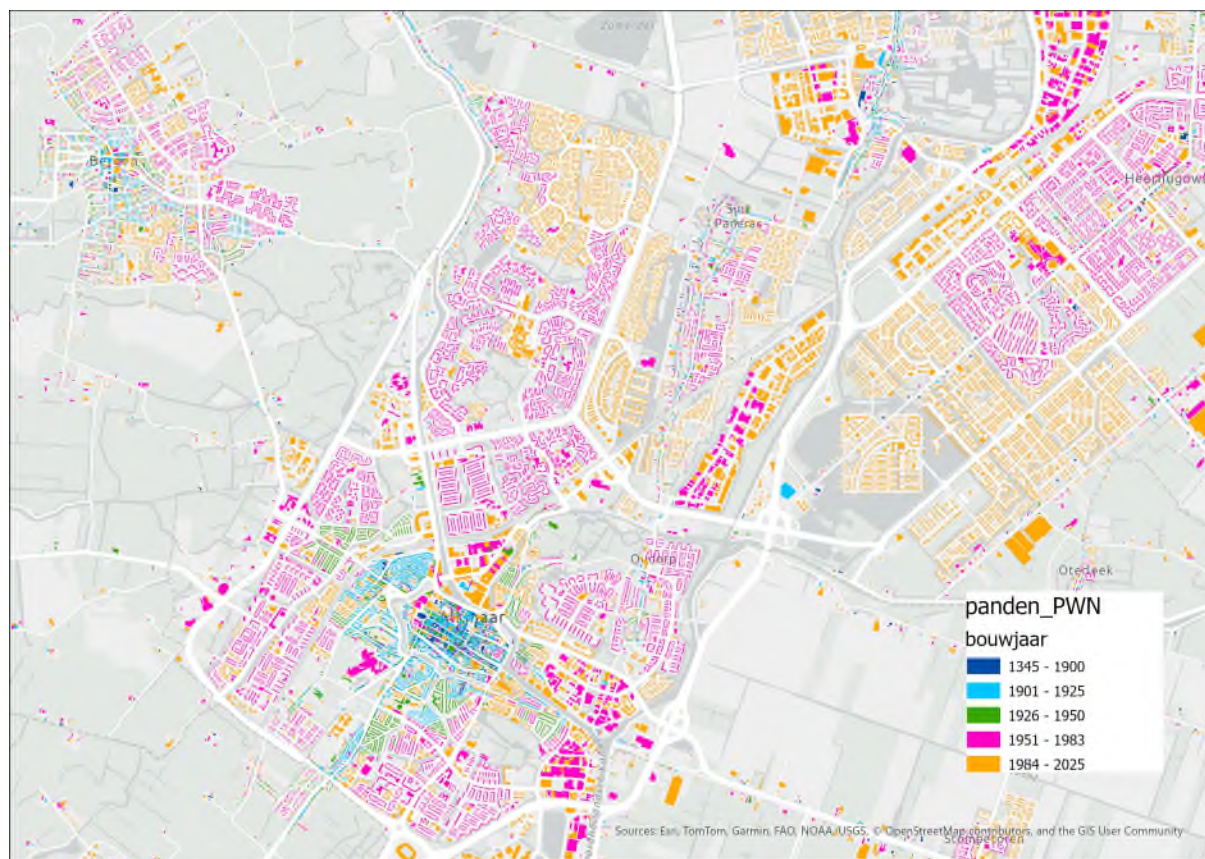
Naast de data over het netwerk van PWN zijn er publieke datasets die gebruikt kunnen worden voor een ML concept:

#### Bebouwing: Basisregistratie Adressen en Gebouwen (BAG)

De BAG is een database met alle adressen en gebouwen in Nederland, inclusief hun ligging, gebruik en status in Nederland (zie Figuur 6). Gezien leidingen altijd aangelegd worden om gebouwen van water te voorzien zijn data rondom bebouwing en leidingen onlosmakelijk met elkaar verbonden. Het eerste aanlegjaar van een leiding zal doorgaans overeenkomen met het bouwjaar van de eerste bebouwing op die locatie. De database wordt bijgehouden door gemeenten die verantwoordelijk zijn voor de kwaliteit ervan. Binnen de BAG wordt onderscheid gemaakt tussen verschillende objecten, de meest relevante voor dit onderzoek is het verblijfsobject:

*“Een verblijfsobject is een kleinste binnen één of meer panden gelegen en voor woon-, bedrijfsmatige, of recreatieve doeleinden geschikte eenheid van gebruik die ontsloten wordt via een eigen afsluitbare toegang*

*vanaf de openbare weg, een erf of een gedeelde verkeersruimte, onderwerp kan zijn van goederenrechtelijke rechtshandelingen en in functioneel opzicht zelfstandig is.”*  
 verblijfsobject | Basisregistratie Adressen en Gebouwen (BAG) | Kadaster



Figuur 6 Voorbeeld van BAG data voor Alkmaar binnen het PWN voorzieningsgebied

De relevante data binnen verblijfsobjecten voor dit onderzoek zijn:

1. Bouwjaar
2. Status: een aantal categorieën waaruit blijkt of het object wel of niet in gebruik is, of er een bouwvergunning verleend is, de bouw gestart is, en dergelijke.
3. Gebruiksdoel: bijeenkomstfunctie, celfunctie, gezondheidszorgfunctie, industrie functie, kantoorfunctie, logiesfunctie, onderwijsfunctie, sportfunctie, winkelfunctie, woonfunctie, en overig.
4. Aantal verblijfseenheden: de hoeveelheid verblijfseenheden binnen een enkel pand.
5. Oppervlakte: bebouwde grond en vloeroppervlak (inclusief verdiepingen)

### Nationaal Wegen Bestand (NWB)

Het nationaal wegenbestand is de database waarin informatie over verkeer en vervoer in Nederland verzameld wordt. Over het algemeen volgen leidingen het stratenpatroon en worden aangelegd onder de stoep, waardoor data over wegen relevant kan zijn voor drinkwaterdistributie. Nagenoeg alle openbare wegen in Nederland zijn opgenomen in het NWB. De relevante informatie zijn de geografische ligging van de individuele wegen en wegvakken en het type weg (Rijksweg, provinciaal, etc.). Het NWB bevat geen informatie over het moment waarop een weg aangelegd is. De bouwdatum kan opgevraagd worden bij de individuele wegbeheerders (Rijkswaterstaat, provincie, gemeenten).



Figuur 7 Voorbeeld van het Nationaal WegenBestand

### Satellietfoto's

Ontwikkelingen op het aardoppervlak kunnen gevolgd worden met lucht/satellietfoto's. De overkoepelende term voor analyse op basis van foto's is 'Remote Sensing' (RS). Deze technieken kunnen gebruikt worden voor analyses van ontbossing tot het smelten van gletsjers en verstedelijking. De eerste aaneengesloten reeks satellietbeelden is gegenereerd door het Landsat programma van de Verenigde staten dat in 1972 zijn eerste satelliet in werking stelde. De data van Landsat is sinds 2008 publiek en vrijelijk beschikbaar. Naast het Landsat programma bestaat in Europa het Copernicus programma dat sinds 2014 beelden verzamelt. Er zijn ook een aantal private partijen welke specialiseren in specifieke aspecten van RS. De data van deze programma's is toegankelijk via verschillende platforms:

1. Google Earth Pro
2. ESRI World Imagery Wayback (zie Figuur 8)
3. USGS EarthExplorer
4. Copernicus Browser



Figuur 8 Voorbeeld van ESRI World Imagery Wayback. Aan de linkerkant de situatie in 2014, rechts in 2025.

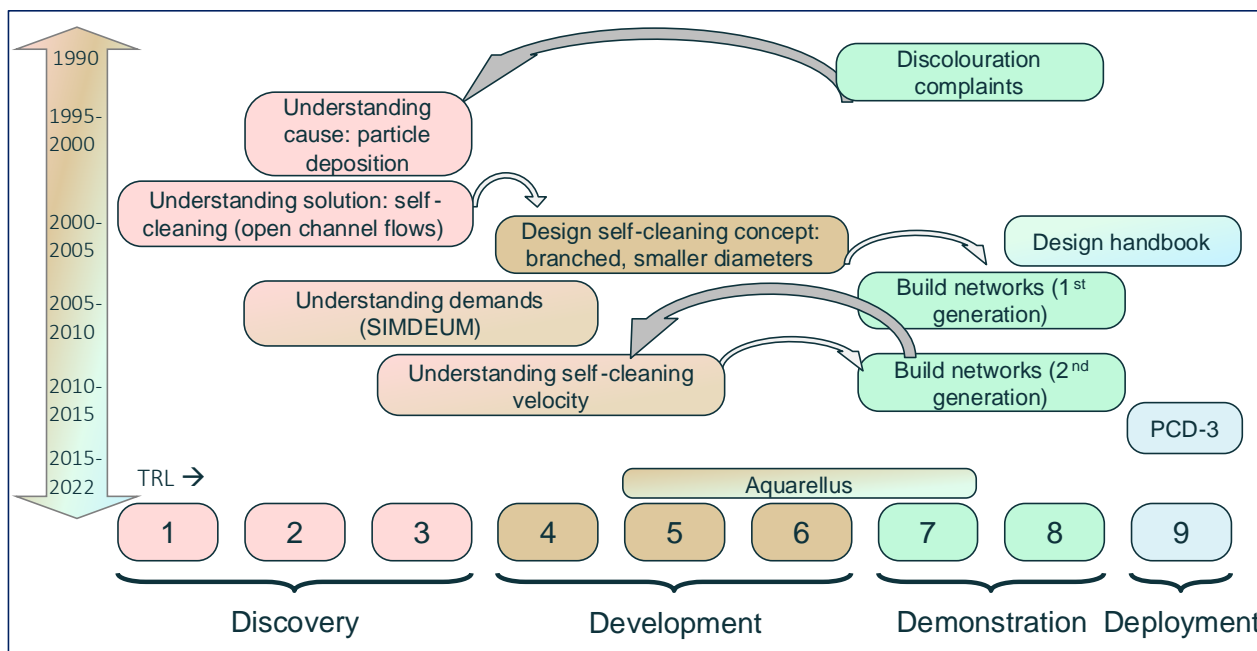
Met een tijdsreeks aan satellietbeelden kunnen veranderingen in een gebied gemonitord worden waarbij de mogelijkheden ook voorbij de zichtbare bandbreedtes voor mensen.

## 5.2 Probleemdefinitie

Als we aannemen dat het drinkwatersysteem ontworpen is op basis van expliciete ontwerpregels (hydraulische berekeningen, ontwerpboek, PCD's, etc.) dan kunnen lokale afwijkingen van deze ontwerpregels duiden op een doorslaggevende invloed van specifieke/afwijkende lokale omstandigheden. In de omgang met zulke omstandigheden wordt teruggevallen op de expertise (impliciete kennis) van professionals. Op basis van de beschikbare data kan anomaliedetectie met ML gebruikt worden om een overzicht te maken van locaties waar de infrastructuur afwijkt van de verwachting, wat kan duiden op de invloed van impliciete kennis op het ontwerpproces. De centrale vraag bij anomaliedetectie is: zien we iets wat op basis van patronen in de dataset onverwacht is. Het concept van anomaliedetectie wordt verder uitgelegd aan de hand van de transitie richting zelfreinigende netten. Het centrale probleem waar het ML concept voor ontworpen wordt is het identificeren van fysieke locaties in een drinkwaterdistributiesysteem waar de gerealiseerde ontwerpkeuzes afwijken van de verwachting.

### Anomaliedetectie: zelfreinigende netten

De aanleg van drinkwatersystemen is gebaseerd op een veranderlijke ontwerpfilosofie. Van oorsprong werden distributienetten ontworpen op basis van de drinkwatervraag, de bluswatervraag, en de wens naar leveringszekerheid en continuïteit (Mesman & Trietsch, 2000). Op basis van deze criteria werd in het verleden vaak gekozen voor een vermaasd ontwerp. In een vermaasd ontwerp is de kans op bruin water echter relatief hoog vanwege de kans op pendelend water, lange verblijftijden en lage snelheden (Mesman & Trietsch, 2000). Naarmate de oorzaken voor bruin water beter werden begrepen werd het mogelijk om een nieuwe ontwerpfilosofie te ontwikkelen en te implementeren, de zelfreinigende netten (zie Figuur 9).



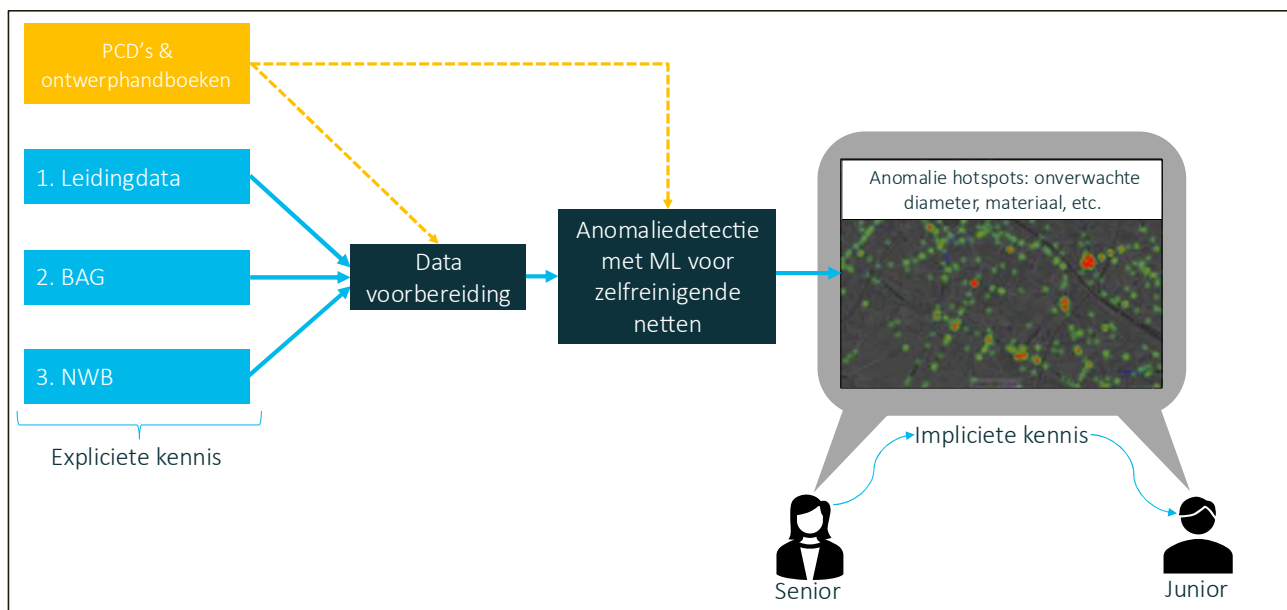
Figuur 9 Ontwikkeling van het concept voor zelfreinigende netten (aangepast uit Beuken & Blokker (2025))

De benodigde veranderingen in een drinkwaternetwerk voor het realiseren van zelfreinigende netten is samen te vatten als (Beuken & Blokker, 2025)

- Uni-directionele stroom: water moet altijd in dezelfde richting stromen, hiervoor is een vertakte structuur nodig.
- Snelheid: minstens 0.2 m/s per etmaal, hiervoor is een **reductie van diameters** richting het eind van het netwerk (110, 63, 40 mm) nodig en mogen er geen doodlopende uiteinden zijn.

- Corrosie: vermijdt corrosiegevoelige materialen, vervanging richting PVC of andere kunststoffen.
- Druk: zorg voor voldoende druk bij maximum watergebruik.

De transitie naar zelfreinigende netten gebeurt vooral bij nieuwbouwwijken (Horst, 2008) en vergt significante veranderingen aan de bestaande fysieke infrastructuur in bestaande wijken. Bij leidingvervanging in binnensteden waar men toch toewerkt naar zelfreinigende netten valt dus te verwachten dat de nieuwe diameter vaak lager moet zijn dan de oorspronkelijke diameter. Met anomaliedetectie is het mogelijk om te identificeren op welke plaatsen deze verandering wel/niet plaatsvindt terwijl deze juist niet/wel te verwachten zou zijn. Het is wel van belang om te benadrukken dat anomaliedetectie met ML niet kan achterhalen *waarom* de afwijking plaatsvindt, alleen *dat* er een afwijking van de norm bestaat. De reden voor de afwijking kan vervolgens achterhaald worden door na te gaan of specifieke omstandigheden gelden op de geïdentificeerde locatie. Bijvoorbeeld: vroegtijdige vervanging van een leiding kan ook gebeuren omdat er een grondverontreiniging vastgesteld is, in plaats van de transitie naar zelfreinigende netten. De manier waarop anomaliedetectie gebruikt kan worden in kennisoverdracht wordt uitgebeeld in Figuur 10. In eerste instantie stellen we een volledig datagedreven aanpak voor, hiermee kan achterhaald worden hoe de keuzes van individuele ontwerpers afwijken van die van andere ontwerpers in soortgelijke situaties. Deze aanpak vervolgens uitgebreid worden met de PCD's en ontwerprichtlijnen maar hiervoor moet een manier gevonden om het ML model de ontwerpregels te laten 'begrijpen' om vervolgens te testen of deze gevolgd worden, of ze worden gebruikt om domein kennis toe te voegen tijdens het voorbereiden van de data (zie paragraaf 5.4).



Figuur 10 Schematisch overzicht voor kennisoverdracht waarbij ML een rol speelt bij anomaliedetectie.

Een andere transitie in ontwerpfilosofie welke wellicht teruggevonden kan worden met bovenstaand concept is de manier waarop afsluiters geplaatst worden. In het verleden werden deze niet op de tekentafel bepaald, maar uit praktische overwegingen door de aannemer tijdens de realisatie van de werkzaamheden. De anekdote gaat dat afsluiters geplaatst werden op de locatie waar werkzaamheden aan het eind van de dag stopten. Anomaliedetectie op basis van ML, zoals hierboven beschreven, zou gebruikt kunnen worden om anomalieën in de transitie naar geplande afsluiter locaties te identificeren.

### 5.3 Evaluatie

Bij het formuleren van een ML aanpak is het nuttig om van tevoren de criteria voor succes vast te stellen. In het kader van kennisoverdracht op basis van anomaliedetectie is het van belang dat een algoritme voldoende

onderscheid kan maken tussen de ‘normale’ situatie en de ‘afwijkende’ situatie. Het is ook van belang dat de totale hoeveelheid gebieden/hotspots die aangemerkt worden niet dusdanig groot is dat het onmogelijk wordt om na te gaan welke lokale condities van toepassing zijn. Voorlopig formaliseren we deze criteria als:

1. Het verschil tussen een anomalie en de normale situatie kan achterhaald/uitgelegd worden door experts die bekend zijn met het specifieke gebied.
2. De hoeveelheid leidingsegmenten dat aangewezen wordt als ‘afwijkend’ is minder dan (bijvoorbeeld) 5% van het totaal.

## 5.4 Kenmerken/features

De beschikbare data bevatten voornamelijk categorische kenmerken en een aantal continue kenmerken. Hiermee kunnen er ook een aantal extra kenmerken afgeleid worden die van belang kunnen zijn voor de analyse. Bijvoorbeeld: de mate van stedelijke dichtheid (afgeleid op basis van de hoeveelheid adressen binnen een bepaald gebied) kan van invloed zijn op de diameter van de aangelegde drinkwaterinfrastructuur. Tabel 3 geeft een overzicht van de beschikbare data, en de mogelijke kenmerken/features welke hieruit afgeleid kunnen worden.

Tabel 3 Overzicht van relevante data voor anomaliedetectie

Bron/data	Type data	Opmerking/afgeleide features
<b>Leidingdata</b>		
Functie	Categorisch	Tertiair, secundair, primair (??)
Materiaal	Categorisch	Beton, PVC, etc.
Nominale diameter	Categorisch	
Datum aanleg	Continu	Mogelijk omzetten naar categorische data op basis van tijdsintervallen (2005-2010, 2010-2015, etc.).
Wanddikte	Categorisch	
Materiaalspecificatie	Categorisch	?
Lengte	Continu	
Geometrie	-	Bufferzone rondom lijnsegmenten voor het opstellen van afgeleide features, bijvoorbeeld: aantal verblijfseenheden binnen een straal van 25 m.
<b>BAG</b>		
Bouwjaar	Continu	Mogelijk omzetten naar categorische data op basis van tijdsintervallen.
Status	Categorisch	Namelijk ?
Gebruiksdoel	Categorisch	Mogelijk te koppelen aan verwacht waterverbruik → industrie gebuikt wellicht meer water per vierkante meter oppervlak dan huishoudens.
Aantal verblijfseenheden	Continu	Bepaling van de mate van verstedelijking op basis van verblijfseenheden, oppervlakte en geometrie.
Oppervlakte	Continu	
Geometrie	-	
<b>NWB</b>		
Weg type	Categorisch	Snelweg, etc.
Geometrie	-	Te combineren met de geometrie van de leidingdata om te achterhalen hoe vaak wegen en leidingen elkaar kruisen.

## 5.5 Modelkeuze

Het identificeren van anomalieën in een dataset vergt een model dat hiervoor geschikt is. De voorkeur is om een datagedreven aanpak te gebruiken omdat het dan niet nodig is om de ontwerprichtlijnen in de handboeken en PCD's mee te geven aan het model. Dit zou namelijk min of meer vergelijkbaar zijn met het maken van een model dat netwerken zelfstandig kan ontwerpen op basis van de ontwerpregels, om vervolgens te evalueren of deze overeenkomen met de werkelijkheid. In een datagedreven aanpak gaat men ervan uit dat de relevante verbanden aanwezig zijn in de data, en dat een ML model deze verbanden kan herkennen, zonder ze noodzakelijkerwijs te 'begrijpen'. Deze aanpak sluit ook aan bij het identificeren van locaties in het netwerk waar impliciete kennis doorslaggevend was in het ontwerpproces. Impliciete kennis is per definitie niet gecodificeerd, en kan dus niet meegegeven worden als input voor een ML model. Een datagedreven aanpak is achteraf dus de enige mogelijkheid om sporen van deze kennis in de data te herkennen, ook al levert dit nog geen verklaring voor de redenen achter afwijkingen in de patronen. Voor deze verklaringen zal het nog steeds nodig zijn om experts om naar hun oordeel te vragen. Hiervoor stellen we een getrapte, datagedreven aanpak voor.

### *1: Inzet van een combinatie van clustering en graaftechnieken*

Als eerste aanpak wordt een datagedreven, unsupervised ML-methode voorgesteld waarbij ook graaftechnieken worden ingezet. Een graaf is een modelrepresentatie van een netwerk en boomstructuur: het bestaat uit een verzameling punten, knopen ('nodes') genoemd, waarvan sommige of alle verbonden zijn door lijnen, de zijden ('edges'). Afhankelijk van de toepassing kunnen de lijnen gericht zijn, dan is sprake van een gerichte graaf. Met graaftheorie kunnen eigenschappen van een graaf worden bepaald. Een clusteringsmethode wordt toegepast, omdat deze techniek robuust is en het toepassen ervan relatief weinig training/'tuning' vergt waardoor snel resultaat geboekt kan worden. Met deze benadering kan (relatief eenvoudig) worden beoordeeld of ontwerpkenmerken überhaupt voldoende onderscheidend zijn in het identificeren van clusters. Met hulp van geoinformatie-gegevens en graaftheorie kan het leidingnetwerk topologisch worden vastgesteld om vervolgens metrieke te berekenen die een mate van vermaasdheid in het leidingnetwerk kunnen vastleggen (bijvoorbeeld hoeveel knooppunten in een netwerk met elkaar verbonden zijn: hoe lager dit getal, hoe meer vertakt het netwerk is). De clusteringtechnieken worden ingezet om diverse kenmerken met elkaar te relateren en te analyseren welke kenmerken doorslaggevend zijn om een ontwerpfilosofie te identificeren. Tenslotte zijn clusteringstechnieken ruim onderzocht en toegepast, ook voor waterdistributienetwerken in bijvoorbeeld het afbakenen van zones en verdere analyse en in combinatie met graaftheoretische concepten (Lifshitz & Ostfeld, 2018; Sela Perelman et al., 2015).

### *2: Inzet van graafneurale netwerken (GNN's)*

De eerste aanpak kan leiden tot het identificeren van clusters die te relateren zijn aan ontwerpfilosofieën, maar die wellicht niet goed of betrouwbaar te reproduceren zijn voor specifieke regio's of delen van een DWDN. De aanpak van deze evaluatie houdt in dat een expert met domeinkennis wordt betrokken en dat een bepaalde foutenmarge wordt gehanteerd, zie verder paragraaf 5.3. In dat geval is het de moeite waard om een alternatieve aanpak te volgen waarbij GNN's worden getraind en toegepast. Dit soort neurale netten kunnen worden toegepast om de bedrijfsvoering van waterdistributienetwerken te analyseren, afhankelijkheden te vatten en te optimaliseren (Vittori et al., 2025; Yazdani & Jeffrey, 2012). Deze neurale netwerkmodellen zijn bij uitstek geschikt om graaf-gestructureerde data te analyseren, zoals getoond in de analyse van kwetsbaarheden in drinkwaterdistributienetwerken. Topologische kenmerken en sensordata kan als inputdata dienen. Dezelfde techniek kan worden gebruikt om de vermaasdheid en andere ontwerpattributen te relateren en vast te leggen in een 'ontwerpfilosofiemodel', om vervolgens afwijkingen ten opzichte van wat in werkelijkheid is aangelegd te kunnen detecteren.

GNN's hebben veel potentieel voor waterdistributienetwerken vanwege hun gelijkenissen, maar er zijn ook aandachtspunten:

- Overfitting: als het model te veel ‘netwerklagen’ heeft dan kan overfitting (het trainen van te veel parameters betekent een grotere nauwkeurigheid, maar heeft ook tot gevolg dat bij nieuwe data het model er vaker naast kan zitten) te voorkomen.
- Over-smoothing kan optreden als de GNN het netwerk te veel versimpelt door knooppunten weg te laten in een neurale netwerklaag. Versimpeling is nodig om classificatie op weg te helpen, maar het kan ook belemmerend werken omdat de oorspronkelijke structuur van het leidingnet verloren gaat en er dus ook meer misclassificaties ontstaan.
- De uitlegbaarheid van een GNN is kleiner dan van andere, eenvoudigere ML-modellen.

## 5.6 Experimenten

Om te komen tot een getraind model voor de detectie van afwijkende ontwerpkeuzes en veranderende ontwerpfilosofieën, zijn een aantal stappen nodig. Het idee is dat het waterleidingnet via gelaagde kennis tot stand is gekomen — expliciete ontwerpstandaarden (ontwerpfilosofie van zelfreinigend netwerk: ZRN) en impliciete lokale aanpassingen vanwege bijvoorbeeld beperkingen in ruimtelijke inpassingen. De impliciete kennis zou ruimtelijk en topologisch zichtbaar moeten zijn in lay-out, diameters en materiaal. Met ML kan segmentering van een (deel van) een DWDN worden bewerkstelligd, waarbij er twee doelen worden nagestreefd:

- **Doel (A):** segmenten waar *alleen* expliciete kennis zichtbaar is versus segmenten met een duidelijke lokale afwijking (impliciete kennis).
- **Doel (B):** onderscheiden van vermaasde netten en ZRN op basis van bouwjaar-clusters.

De output van de clustering, en mogelijk ook de GNN die de clustering kan verfijnen, is input voor een interpretatie door experts:

- Bij **A** (impliciete kennis): de expert verklaart *waarom* een segment afwijkt, indien een verklaring voorhanden is. Bij (A) is het gesprek tussen een expert en een junior medewerker de manier om kennisoverdracht te faciliteren.
- Bij **B** (verloren expliciete kennis): de expert helpt de classificatie te ijken, en de ingedeelde ZRN gebieden kunnen vervolgens worden getoetst aan bijvoorbeeld bruinwaterklachten-data.

Tenslotte kan na evaluatie de classificatie worden getoetst aan operationele data zoals bruinwatermeldingen om de indeling praktisch te valideren. In Tabel 4 zijn de stappen van voorbereiding, training van de ML-technieken en expertinterpretatie en -validatie verder uitgewerkt.

Tabel 4: stappenplan om een categorisering van ZRN-generaties te realiseren waarbij de expert interpreteert en valideert.

Stap	Methodiek	Omschrijving	Output
<b>1: Voorbewerking</b>	Segmentatie	Verdeel netwerk in analyse-eenheden (DMA of een wijk)	Subnetten
	Graafopbouw	Bouw m.b.v. de GIS-data een graaf op, waarbij de leidingen de zijden zijn en de kruisingen en aansluitingen knopen (bijv. met de NetworkX Python module). De graafrepresentatie is nodig voor zowel de cyclusdetectie (vermazing/vertakking) met clustering als de GNN-aanpak.	Graaf per segment
<b>2: Featureconstructie en -selectie</b>	Bepalen topologische kenmerken	Aantal cycli, graadverdeling en padlengte. De graadverdeling geeft een verdeling aan van de knoopgraad. De knooppuntgraad is een getal die weergeeft hoeveel verbindingen (zijden) het heeft tot andere knooppunten.	Structuurfeatures

	Bepalen fysische kenmerken	Diameterstatistieken, materiaal, lengteverdeling	Asset features
	Bepalen tijdskenmerken	Aanlegjaar, vervangingen, bouwperiode	Historische features
<b>3: Training</b>	Clustering	Als eerste aanpak wordt clustering op de features of een selectie van features toegepast, zoals k-means, of DBSCAN. Dit levert een categorisering (clustering) op.	Clusterlabels
	(optioneel) GNN	De relatie tussen lokale topologie (buurleidingen, knooppuntgraad, diameter-overgang) en het type ontwerpfilosofie wordt door de GNN geleerd.	Latente representatie
<b>4: Expertvalidatiesessie</b>	Variatie-analyse	Spreiding binnen clusters wordt bepaald	Clusterkwaliteit
	Relatie met data	Distributies van bouwjaar, diameter, etc. binnen een cluster worden geanalyseerd	Clusterprofielen
	Interpretatie	Experts duiden clusters (ZRN-generaties) en labelen deze. De GNN kan vervolgens opnieuw getraind worden met deze labeling zodat impliciete en expliciete kennis ontvecht worden (zie stap 5).	Gelabelde clusters
	Kennisinbreng	Verklaring van afwijkingen met impliciete kennis (A) en typeert verloren expliciete kennis door middel van labeling (B).	Kennisregels
<b>5: Kalibratie en iteraties</b>	Verfijning clustering	Selectie van aantal clusters	Definitieve clustering
	Modelverbetering	Nieuwe features toevoegen aan GNN door toevoeging labeling ZRN-generaties (semi-supervised training) en/of verbetering van de selectie van features voor clustering	Verbeterd model
<b>6: Toepassing en validatie</b>	Anomaliedetectie	Afwijkingen t.o.v. clusterprofielen worden gevisualiseerd en geanalyseerd door een expert	Afwijkingsanalyse
	Validatie met praktijk	De classificatie wordt getoetst aan operationele data (zoals bruinwatermeldingen)	Operationeel inzicht

## 6 Aanbevelingen voor implementatie

In dit onderzoek wordt een conceptuele aanpak beschreven waarmee kennisoverdracht tussen senior en junior ondersteund kan worden, specifiek op het gebied van WDN-ontwerp. In dit hoofdstuk doen we een aantal aanbevelingen voor daadwerkelijke implementatie.

### 6.1 Testen van de impliciete-kennis-aanname en use-case

De opzet van het ML concept in dit onderzoek is gebaseerd op de aanname dat impliciete kennis doorslaggevend kan zijn in het ontwerp van een WDN, en is vooral anekdotisch van aard. Voordat een ML model daadwerkelijk gebouwd wordt is het aan te bevelen om deze aanname uitvoeriger te testen. Hiervoor beschrijven we twee opties:

1. Samen met de experts van een bepaald voorzieningsgebied nagaan of en hoe vaak het voorkomt dat de werkelijke huidige situatie in het distributienet afwijkt van hetgeen zij zouden verwachten of aanbevelen. De hoeveelheid en gemak waarmee dergelijke locaties aangewezen kunnen worden geven inzicht in de mate waarin het makkelijk of moeilijk zal zijn voor een model om ook dergelijke locaties te identificeren. Wanneer een eerste inschatting aangeeft dat er amper/weinig van dit soort locaties zijn betekent het dat een ML model erg precies/nauwkeurig zal moeten zijn om überhaupt een dergelijk patroon in de data te kunnen vinden. Wanneer de eerste reactie is dat een redelijk deel van het netwerk afwijkt van expliciete ontwerpregels dan is het aannemelijk dat dit teruggevonden kan worden in de data. Wanneer met aangeeft dat een groot deel van het netwerk afwijkt dan het voor het ML model ook lastig zijn om afwijkingen van normaal te onderscheiden. Dit komt neer op een inschatting van de signaalsterkte van het fenomeen.
2. Wat is de reactie wanneer een afwijkende situatie voorgelegd wordt aan een expert in het voorzieningsgebied? Als er een (overtuigende) verklaring gegeven wordt kan er vervolgens nagegaan worden of deze gebaseerd is op expliciete of impliciete kennis. Wanneer er geen verklaring voorhanden is dan laat dit ook zien dat er factoren meespeelden tijdens het beslisproces die moeilijk overdraagbaar blijken te zijn.

Deze sessies met experts kunnen een eerste beeld geven maar moeten niet gezien worden als vervanging voor het voorgestelde statistische model. Het is namelijk moeilijk voor mensen om hun besluitvormingsproces achteraf te doorgronden en men heeft de tendens om zijn besluit achteraf te rationaliseren op basis van uitlegbare redenen in plaats van eigen ervaringen en intuïtie (Nisbett & Wilson, 1977). Tegelijkertijd kunnen de sessies ook gebruikt worden om na te gaan in hoeverre een dergelijke aanpak van toegevoegde waarde zou voor kennisoverdracht, en daarmee te bepalen of er ook behoefte is om een dergelijk ML concept verder te ontwikkelen.

### 6.2 Trapsgewijze implementatie

De invloed van impliciete kennis op het ontwerpproces, en de mogelijkheid om deze via een datagedreven model te achterhalen, is voorlopig een hypothese. Daarom wordt een trapsgewijze implementatie van het bovenstaande ML concept aanbevolen. De eerste stap in dit proces is beschreven in de vorige paragraaf. De volgende stap is het ontwikkelen van een lichtgewicht versie van het ML model waarbij een minimum aan data, en daardoor een minimum aan voorbereiding, gebruikt wordt. Denk hierbij aan het achterwege laten van satellietfoto's. Het doel van deze eerste versie is om een eerste indicatie te krijgen of het mogelijk is om locaties waar afgeweken is van de expliciete ontwerprichtlijnen te identificeren. Wanneer blijkt dat een dergelijke versie van een ML model hier

succesvol in is, door dit te laten bevestigen door experts uit het gebied, is het de moeite waard om een completer model te bouwen voor nauwkeurigere uitkomsten.

### 6.3 Alternatieve verklaringen voor ontwerpafwijkingen

Naast impliciete kennis kunnen er ook andere redenen zijn dat het ontwerp afwijkt van de verwachtingen. Denk hierbij aan verwachtingen rondom de ontwikkelingen van een gebied die niet door bleken te gaan. Een dergelijke situatie kan wellicht teruggevonden worden in het WDN-ontwerp (grotere diameters), terwijl het nauwelijks/onmogelijk terug te vinden kan zijn in de beschikbare data om het model te trainen. Met de beschreven methode kan deze locatie alsnog aangewezen als een anomalie, maar reflecteert niet de implicite/explicite kennis aannames waarmee we werken. Op dat moment was het wellicht niet de correcte impliciete kennis van de toenmalige ontwerper, maar de foutieve expliciete aannames over een gebied welke tot deze ontwerpkeuze leidde. Van beiden valt iets te leren, maar ze moeten niet door elkaar gehaald worden. Het is dus nuttig om correctheid van kennis, en de vorm van kennis los van elkaar te zien:

- Expliciet en correct (hydraulische formules en simulaties)
- Expliciet en incorrect (achterhaalde ontwerprichtlijnen)
- Impliciet en correct (werkelijk begrip van een gebied of locatie)
- Impliciet en incorrect (foutieve aannames of een verouderd mentaal model)

De gebruiker van een ML model dat ontwikkeld is op basis van de datagedreven aanpak in de bovenstaande hoofdstukken moet zich ervan bewust zijn dat het moeilijk/onmogelijk zal zijn om enkel met het algoritme onderscheid te maken tussen correcte en incorrecte impliciete kennis.

## 7 Conclusies

In dit project is een conceptuele en methodologische basis gelegd voor kennisoverdracht tussen senior en junior, een onderwerp dat relevant is gezien de vergrijzing van de Nederlandse beroepsbevolking samenvalt met een significante vervangingsopgave van het drinkwaterdistributiesysteem. Door kennis te definiëren op basis van de mate waarin deze persoonsgebonden is, in plaats van langs kennisdomeinen, leveren we de argumentatie en raamwerk voor het soort kennis dat extra aandacht verdient tijdens kennisoverdracht. De beschikbare data over drinkwaterinfrastructuur, het denkkader over expliciete (te documenteren en overdraagbaar) en impliciete (persoonsgebonden en moeilijk te documenteren) kennis, en de beschikbare ML methodes komen samen in onderstaande hypothese:

*Impliciete kennis kan invloed hebben op het ontwerp en aanleg van (drinkwaterdistributie)systemen en statistisch relevante sporen hiervan verwachten we terug te vinden met datagedreven methodes voor anomaliedetectie op basis van machinaal leren. Het bespreken en verklaren van ontwerpkeuzes op locaties waar deze sporen aanwezig zijn is waardevol in het proces van kennisoverdracht.*

Een voorbeeld van een consequentie van bovenstaande hypothese binnen drinkwaterdistributie is een cluster van leidingen in een bepaald gebied/wijk met afwijkende diameters vergeleken met vergelijkbare gebieden elders in de stad. Deze afwijking/anomalie kan vervolgens samen met lokale experts onderzocht te worden om na te gaan wat de achterliggende oorzaak is. De gezamenlijke zoektocht naar deze verklaringen, waarbij men ook dieper moet zoeken dan enkel de mogelijke invloed van impliciete kennis, is waar kennisoverdracht daadwerkelijk zal plaatsvinden. Anomaliedetectie op basis van machinaal leren wordt dus gebruikt om aan te wijzen waar men moet zoeken, de kennisoverdracht zelf blijft afhankelijk van interacties tussen mensen.

## 8 Referenties

- Andriy Burkov, B. (2019). *The Hundred-Page Machine Learning Book*.
- Beuken, R., & Blokker, M. (2025). *KWRW 2025-049 Evaluatie zelfreinigende netten workshop 2-4-2025*.  
[www.kwrwater.nl](http://www.kwrwater.nl)
- Bourke, D. (2024, October 30). *A 6 Step Framework for Approaching Machine Learning Projects*.  
<https://dev.mrdbourke.com/zero-to-mastery-ml/a-6-step-framework-for-approaching-machine-learning-projects/>
- Doorn, N. (2021). Artificial intelligence in the water domain: Opportunities for responsible use. *Science of the Total Environment*, 755. <https://doi.org/10.1016/j.scitotenv.2020.142561>
- Falckenthal, B., Au-Yong-Oliveira, M., & Figueiredo, C. (2025). Intergenerational Tacit Knowledge Transfer: Leveraging AI. *Societies*, 15(8). <https://doi.org/10.3390/soc15080213>
- Fu, G., Sun, S., Hoang, L., Yuan, Z., & Butler, D. (2023). Artificial intelligence underpins urban water infrastructure of the future: A holistic perspective. *Cambridge Prisms: Water*, 1. <https://doi.org/10.1017/wat.2023.15>
- Harari, Y. N. (2024). *Nexus: van het stenen tijdperk tot AI, een beknopte geschiedenis van informatienetwerken*. Thomas Rap.
- Horst, P. (2008). *Gefaseerde aanleg van een zelfreinigend leidingnet in bestaande wijken (BTO 2008.047(s))*.  
[www.kwrwater.nl](http://www.kwrwater.nl)
- KWR. (2026). *Vigerende praktijkcodes transport & distributie - Praktijkcodes Drinkwater*.  
<https://www.praktijkcodesdrinkwater.nl/onderwerpen/vigerende-praktijkcodes-transport-distributie/>
- Lifshitz, R., & Ostfeld, A. (2018). Clustering for Analysis of Water Distribution Systems. *Journal of Water Resources Planning and Management*, 144(5). [https://doi.org/10.1061/\(asce\)wr.1943-5452.0000917](https://doi.org/10.1061/(asce)wr.1943-5452.0000917)
- Mesman, G. A. M., & Trietsch, E. A. (2000). *Ontwerprichtlijnen voor een vertakt leidingnet (BTO2000.03)*.  
[www.kiwa.nl](http://www.kiwa.nl)
- Nisbett, R. E., & Wilson, T. D. (1977). *Telling More Than We Can Know: Verbal Reports on Mental Processes* (Vol. 84).
- Sela Perelman, L., Allen, M., Preis, A., Iqbal, M., & Whittle, A. J. (2015). Automated sub-zoning of water distribution systems. *Environmental Modelling and Software*, 65, 1–14. <https://doi.org/10.1016/j.envsoft.2014.11.025>
- Theobald, O. (2017). *Machine Learning For Absolute Beginners*.
- Vekaria, D., & Sinha, S. (2024). aiWATERS: an artificial intelligence framework for the water sector. *AI in Civil Engineering*, 3(1). <https://doi.org/10.1007/s43503-024-00025-7>
- Vittori, G., Falkouskaya, Y., Jimenez-Gutierrez, D. M., Cattai, T., & Chatzigiannakis, I. (2025). Graph neural networks to model and optimize the operation of Water Distribution Networks: A review. In *Journal of Industrial Information Integration* (Vol. 47). Elsevier B.V. <https://doi.org/10.1016/j.jii.2025.100880>
- Yazdani, A., & Jeffrey, P. (2012). Water distribution system vulnerability analysis using weighted and directed network models. *Water Resources Research*, 48(6). <https://doi.org/10.1029/2012WR011897>

# I Keywords WDSA/CCWI 2026 programmaboekje

Programmaboekje: [https://www.wdsa-ccwi2026.ucy.ac.cy/wp-content/uploads/2026/05/wdsa\\_ccwi\\_2026\\_program\\_booklet\\_with\\_abstracts\\_2026-05-19.pdf](https://www.wdsa-ccwi2026.ucy.ac.cy/wp-content/uploads/2026/05/wdsa_ccwi_2026_program_booklet_with_abstracts_2026-05-19.pdf)

Termen voor AI en ML → elk abstract met deze term wordt gecategoriseerd als AI

- artificial intelligence
- machine learning
- deep learning
- neural network
- reinforcement learning
- graph neural
- large language model
- llm
- generative ai
- chatgpt
- gpt-N
- random forest
- gradient boost
- xgboost
- support vector machine
- svm
- convolutional neural
- lstm
- gnn
- autoencoder
- explainable ai
- xai
- attention mechanism

Termen voor digitale tweelingen

- digital twin
- digital-twin

Kennisoverdracht

- knowledge transfer
- knowledge sharing
- capacity building
- operator training
- workforce development
- tacit knowledge
- intergenerational
- mentoring
- skill transfer
- expertise transfer
- institutional knowledge

- implicit knowledge
- explicit knowledge