Verslag workshop AI en Dijkmonitoring

26 maart 2024

Op dinsdag 26 maart organiseerde het Netwerk Dijkmonitoring (NDM) in samenwerking met DigiShape een workshop over AI en Dijkmonitoring. De workshop vond plaats bij de TU Delft. Het onderwerp was AI en Dijkmonitoring. Hierover vertelden Jeroen Baars vanuit Hoogheemraadschap Hollands Noorderkwartier, Joost Stenfert vanuit HKV Lijn in water, Muriel Serrurier Schepper vanuit het AI Annotatielab, Juan Pablo Aguilar Lopez vanuit de TU Delft en Kin Sun Lam vanuit Deltares. Aan het einde was er nog ruimte voor discussie met een borrel. De sessie was Engelstalig.

Introduction by Wouter Zomer (NDM) and Chris Karman (DigiShape)

In the introduction by Wouter Zomer about the Network Dike Monitoring, he discussed the benefits of early deployment of monitoring systems and how sharing user experiences can be advantageous. This practice accelerates development and enhances value. Quality is achieved through usage, and efficiency increases the quality over time. The Network Dike Monitoring has the following plans for 2024: technical-oriented workshops will be conducted, an article about the "Spoorboekje Implementatie Dijkmonitoring" will be published in Land & Water magazine, and the website dijkmonitoring.nl will be updated to a Wiki environment focused on dike monitoring techniques.

Following this, Chris Karman spoke to the audience about DigiShape. DigiShape is an open innovation platform comprising companies, knowledge institutions, and government bodies that collaboratively explore and exploit the potential of digitalization in the water sector. We work together on concrete projects, experimenting as an open community with available data, new data, and advanced techniques. The strength of DigiShape lies in the interaction between partners and the community. During DigiShape events such as this workshop, the community can contribute ideas and participate actively. Additionally, initiating ideas and sharing outcomes are always encouraged. The thematic areas of focus include Water & Logistics, Water Management, and North Sea Spatial Planning. Key elements include precompetitive propositions, sharing models and data, and strategic network expansion.

Jeroen Baars (Hoogheemraadschap Hollands Noorderkwartier)

There has been a significant shift in Artificial Intelligence (AI), evident in evolving parameters, models, and the ever-changing landscape of large language models and foundation models. This shift has been particularly noticeable since 2006, with the development of increasingly complex models. Concurrently, the water sector faces its own challenges, such as increased precipitation, more frequent peaks, and greater deficits. These changes present difficulties not only in AI but also in water management. Together, they necessitate a shift in water security tasks, focusing on:

- Extreme peak situations
- Shifts in endurance and extremities
- Shifts in possibilities → AI and new data sources

The HHNK (Hoogheemraadschap Hollands Noorderkwartier) aims to ensure water security through a data-driven approach. Centralized data storage is essential, as is ensuring the accuracy of data to understand the current state of a dike on its degradation curve. This approach is part of HHNK's vision in the Waterveiligheid 2030 project, which seeks to provide continuous insight into water security.

The water board focuses on the dominant failure mechanisms, which for HHNK are: height/overflow, erosion of the outer slope, and micro-instability. Their approach emphasizes the "data story" of dikes by providing continuous insight, risk-based management, accurate estimates, detection of changes, understanding the effects of these changes, and measuring effective parameters.

To test various measurement techniques, HHNK initiated the Case Study Purmer project. The project began on a small scale, aiming to prove new methods and scale up effective techniques. AI was tested for several purposes within this context.

Initially, measurements were integrated into a machine learning model to gain continuous insight into the phreatic line in the dike. The time series of groundwater measurements were analyzed using Pastas, an open-source software for groundwater time series analysis. This allowed for the estimation of gaps in the data.

Next, insights into drought conditions were gathered by implementing an indicator in the dike. The report *Een nieuwe grondwater droogte-indicator voor boezemkaden* provides further details on this project.

Additionally, AI was utilized in conjunction with drone time series to detect changes such as armour movement, vegetation changes, digging, and cracks. Multiple measurements were taken annually, allowing for differential analysis. Joost Stenfert will provide more information about this AI Toolbox. The insights gained from the data enable the water board to decide on appropriate actions.

Based on the findings from the above measurements, detailed inspections can be conducted on request using the Lidar feature on iPhones. This method allows for detailed field inspections to be performed remotely and enables the monitoring of changes over time.

In conclusion, a robust combination of data and information is crucial for continuous insight and on-demand operations. Water boards must decide the level of control they desire, and HHNK has used a phased approach with small pilots. Although the vision for 2030 is still a work in progress, significant steps have been taken.

One of HHNK's goals is to work with open-source solutions, making approaches and insights shareable. Collaboration is key: how can we assist each other further? For example, in 2023, HHNK collaborated with Wetterskip Fryslân by sharing drone images of dike cracks, creating a larger database for model development.

Water boards cannot keep up with all changes alone. Implementing and sharing these changes is crucial. Within the deep learning field, the rise of foundation models can be leveraged, which can probably be more effectively fine-tuned instead of training new deep learning models from the start.

We are currently in a shifting world with significant ambitions and progress, but also challenges. Every water board faces these challenges. HHNK seeks to avoid reinventing the wheel and instead leverages collaborations. Sharing developments is efficient. The question is how to share major shifts, data, trained models, and standards? How can we find mutual interests and collaborate effectively?

Joost Stenfert (HKV Lijn in water)

Artificial Intelligence (AI) is everywhere, and we can no longer ignore it. We all use it daily, whether we realize it or not. The task now is to see how we can use it in a proper and effective way.

Joost was inspired by the use of AI to detect veins in the human eye. He thought: How can we apply this AI method to other areas? This led to the idea of using AI to detect cracks in dikes. Testing soon began.

At Flood Proof Holland, experiments were conducted on real dikes. Cracks were artificially created, and drones captured these cracks in photographs. These images were then taken to the office, where the algorithm designed for detecting veins in human eyes was tested on the dataset of dike cracks. This marked the beginning of the crack detection part of the AI Toolbox, focusing on improving the algorithm specifically for dikes. It was challenging due to the limited number of available images. However, data from other experiments over the years helped. The Water Innovation Prize awarded to HHNK last year demonstrated significant progress in their work.

Why the AI Toolbox? Each year, water boards inspect thousands of kilometers of levees. This is a substantial amount of work, and we are increasingly interested in the varying conditions of levees and the differences between inspection years. Inspectors face a high workload, and their work is labor-intensive. They need assistance. Manually recording and processing the damage found by inspectors is time-consuming and laborious. Meanwhile, the development of measurement techniques and AI is advancing rapidly. We must learn to use these new tools responsibly. Once we do, they can greatly aid in the inspection and monitoring of levees.

The current concept of use is a risk-based approach, which includes the following steps:

1. Inspectors and personnel working on the dikes are tasked with identifying holes, damage, and the quality of vegetation. For example, if a young tree is found, it is removed from the dike.
2. Drones or photographs, combined with algorithms, assist these personnel. This can involve RGB (color) images, infrared (heat), or Lidar (structure, height, 3D) from various devices:
   - Satellites
   - Airplanes
   - Drones
   - Handheld photo cameras

   The balance between time and quality determines the desired precision from these measurement devices. Satellites are not suitable for detecting cracks in dikes. Handheld cameras produce inconsistent results as everyone captures images differently. Drones offer flexibility in timing, resolution choice, and access to interesting locations.

3. Algorithms perform repetitive tasks and record information. They work well for comparing data across years and provide more precise field information, reducing monotonous work.
4. The work of inspectors and dike personnel becomes more efficient and effective with the use of drones and algorithms. The focus is on critical locations, making inspection and monitoring more effective. The systems support the inspector: due to the risk-based approach, the inspector can prioritize visiting locations flagged by the techniques, and only visit less critical areas if time allows.

The AI Toolbox started with HKV and HHNK, followed by STOWA. Currently, Wetterskip Fryslân and Waterschap Brabantse Delta, among others, are interested in joining.

The question is: how do we proceed?

HKV aims to expand the AI Toolbox beyond crack detection. For example, work is already underway on detecting young trees, with more developments in progress. Additionally, working open-source and collaboratively is crucial to avoid falling behind in AI advancements in a few years. Establishing cloud storage for training data would be beneficial. Learning to collaborate effectively will ensure practical use of these technologies. Encouraging each other to pioneer AI development for dikes is not risky but necessary.

Muriel Serrurier Schepper (AI Annotatielab)

The AI Annotatielab demonstrates the importance of cooperation. It is a reintegration program company that aims to:

1. Improve job opportunities for people with barriers to employment through socially responsible data annotation.
2. Support clients in the ethical training of AI models.
3. Help employers comply with job agreement laws and meet their diversity and inclusion standards.

People come to the company and get personal coaching there. After about three months, a job hunter assists them, and over the next six months, they gain work experience. This experience involves annotating data: transforming unstructured data into structured data suitable for AI development. Human annotations make unstructured data comprehensible to machines. The success of the program is evident, as 80% of participants find suitable jobs afterward. It provides a low-stimulation environment where they can develop their skills.

At AI Annotatielab, data is annotated in a socially responsible way for companies. The reality of annotating data is often less ideal, with annotation work abroad done by people earning less than $2 per hour.

Several projects AI Annotatielab has worked on include selecting thumbnail images for Videoland and helping HHNK detect cracks in dikes. HHNK challenged AI Annotatielab to create training data for detecting dike cracks, as they lacked the time and personnel to annotate the data themselves. This project was challenging, involving multiple meetings with HHNK to define what constitutes a crack and what does not. The cracks had to be of a certain size to be relevant.

This type of challenge is common in AI Annotatielab projects: clients provide documentation on annotation requirements, but unforeseen situations always arise. Should an object in the picture be annotated or not? What if a sheep is on the crack? The team aligns with the client to decide the course of action. Muriel was pleased to see HHNK win the Water Innovation Prize, recognizing the precise work done by AI Annotatielab. Their efforts allowed HHNK to evaluate the quality of their algorithm.

Juan Palbo Aguilar Lopez (TU Delft)

Juan will discuss the role of sensors in AI systems for dike monitoring. He collaborated on this research with three students. First, he will explain the differences between Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL).

- AI is the use of technologies to build machines and computers capable of mimicking human cognitive functions seeing, feeling, smelling, listening, reasoning, responding, etcetera.
- ML is a set of statistical (or not) databased learning algorithms for predicting, classifying, deciding, optimizing, augmenting, reducing and generalizing.
- DL is one subset of ML methods based on ANN's with specific model architectures for predicting, classifying, deciding, optimizing, augmenting, reducing and generalizing.

Sensors in AI are used during and after events (predicting) while sensors in ML/DL are used before events (training and validation). When monitoring a system, it is crucial to determine what you want the system to learn. Sometimes training occurs during system operation, as with ChatGPT.

With 3,500 primary dikes and 14,000 regional dikes, there is a significant amount of work to be done. Dikes are deteriorating more rapidly due to climate change cycles, including flooding and drought. Therefore, we need to monitor them more intensively and frequently. Relying solely on humans as monitoring devices is impractical due to the sheer volume of work. Sensors are essential but they generate vast amounts of data. AI can assist in managing and analyzing this data. Juan will present three case studies to illustrate these concepts.

Case a: Where to place and how many pressure head sensors to predict/monitor Backward Erosion Piping?

In this research, simple machine learning techniques are used to determine optimal sensor placements to predict the progression of piping over time. Two scenarios were considered: placing sensors throughout the entire aquifer and placing sensors at the sides of the dike. The latter is preferred since placing sensors under the dike is less practical in real-world locations.

A time-dependent, physically-based Backward Erosion Piping model was proposed, comprising three parts:

1. The IJkdijk Experiment 2009
2. A Finite Element Method (FEM) model and laminar sediment transport equations
3. Piping erosion progression over time (including pore pressure estimations)

Machine learning was implemented using Principal Component Analysis (PCA). Three principal components were identified, two of which were effective in predicting piping. One component correlated strongly with water loading, while the other correlated with the progression of the pipe.

The sensors that best correlated with these components were linked to the principal components, enabling the prediction of pipe growth. Sensors placed closer to the water loading area best indicated the loading component, while sensors placed near the pipe best indicated pipe growth. However, as previously mentioned, placing sensors under the dike is not desirable. Therefore, the focus shifted to the toe locations of the dike and determining the optimal depth for sensor placement.

Case b: Which dikes are more prone to develop drought induced cracks?
To answer this question, proxy variables were used to try to predict observations. The first step involved analyzing the inspections and crack locations between 2018 and 2021 in a case study conducted by the Hoogheemraadschap van Delfland. The available data pertained to the cracks themselves rather than the drought process. Cracks are a dynamic phenomenon; what is visible today may disappear tomorrow. Because the dike locations were so close to TU Delft, it was possible to model, program, and directly verify the locations in real life.

To predict the cracks, data augmentation was necessary. In a GIS environment, all the cracks were mapped. For these locations, time series data on soil subsidence, precipitation, evaporation, soil subsidence, and NDVI (Normalized Difference Vegetation Index) "green-ness" were extracted. Additionally, local soil characteristics were included. Some dikes are more prone to cracking due to their orientation to wind and sun. Building models by combining this data is challenging due to the different spatial and temporal resolutions of the sensors.

An association analysis was used to build the models. It was decided to create two models: one for cracks larger than 2 meters and one for all cracks. Principal Component Analysis (PCA) was implemented as a machine learning method. These models operated on a decision tree framework and were trained using a Random Forest model for hazard mapping.

The decision trees revealed that:

- Long cracks are not observed on levees with flexibility smaller than 0.355 m/kPa.
- Long cracks (longer than 2 meters) are more often found on levees with slopes oriented towards the southern side.
- Both model trees indicate that a peat thickness of at least 31 cm in the upper layer makes levees susceptible to crack formation.
- Levees composed of soils with peat layers thinner than 31 cm do not seem to crack when precipitation deficit values are lower than 311 mm.

Case c: How to detect cracks with distributed temperature fiber optic sensors and deep learning algorithms?
A fiber optic cable works on the principle of light pulses. By measuring the time it takes for light pulses to travel through the cable, information can be transmitted. The cable is highly sensitive to bending, and changes in the properties of the light occur due to deformation or heat. This makes fiber optic cables excellent for measuring changes.

In this research, a peat dike prone to cracking was monitored using fiber optic cables. Thirty-two lines along the dike were installed, along with a webcam. After monitoring for three months, interpolation between the measured temperatures provided a comprehensive view of the situation, allowing for the creation of heat distribution maps along the dike.

As a second step, a convolutional high-resolution image-to-image encoder-decoder was built and trained. Data augmentation was necessary for this process because algorithms designed for big data require large datasets. The initial dataset of 922 images was enlarged to 31,744 images by cropping at different locations, flipping, and mirroring the images. The resulting trained model shows promise, but more training and testing of the architecture are needed. It is already capturing patterns, indicating its potential.

In conclusion: Machine learning can be used to design a sensor system based on dimensionality reduction capacity. AI requires sensors before being operational (ML) and after becoming operational. Sensor combination can pose a challenge in AI and ML performance due to its variability in frequency and spatial resolution. AI, ML and DL are tools for knowledge augmentation. However, their capacity is completely determined by the sensor with highest spatial resolution and frequency.

Kin Sun Lam (Deltares)

Kin Sun's presentation is about the DigiTwin and proven dike strength. A Digital Twin is a dynamic, virtual representation of a physical asset, product, process, or system. It represents something in the field and can predict what might happen under certain circumstances.

Monitoring and AI play a crucial role in Digital Twins. Monitoring, observations, and sensing technology provide input for existing models, including indirect data like remote sensing. Data science and AI help experts link this data with the models. Expertise is still essential for creativity, domain knowledge, and critical thinking. Updating the models with monitoring, observations, and sensing technology is a key part of a Digital Twin. AI can support decision-making in the future.

Example: Subsoil Modelling
As a pilot project at HHNK, the subsoil was modelled using CPTs, boreholes, and EM measurements with Machine Learning algorithms. This generated a subsoil model that is continuously updated with new CPTs and boreholes.

Case Study: Proven Dike Strength with Groundwater Monitoring
In the Proeftuin Purmer project, which Jeroen Baars also discussed, the goal was to create a complete Digital Twin with all its elements. One objective was to update the behavior model, which assesses dike stability based on groundwater levels at any given moment.

Additionally, Pastas software was used to forecast future groundwater levels in the dikes. The input data for this forecast included precipitation data, evaporation data, and measured groundwater levels. The output provided the best fit with a reliability interval for the groundwater levels at various points within the dikes, enabling both forecasting and hindcasting.

Using measured meteorological data, observed groundwater levels, and generated groundwater levels from Pastas, the failure probability of the dike could be updated.

Conclusion
In conclusion, real-time stability based on groundwater modelling is already achievable. Deltares enhanced the utility of measured data by using Pastas for predictions. Forecasts and hindcasts of dike stability are possible with these groundwater level predictions. The probability of failure can be updated with groundwater monitoring. Lastly, there are opportunities to improve models and predictions with more and better data (monitoring), as well as other AI algorithms and data science techniques.

Discussion

During the discussion, it became clear that both attendees and speakers are interested in sharing data in the future. This is the best way to improve the models fast. How to achieve this is something we need to consider as a community.