MACHINE LEARNING APPLICATIONS IN SEWER SYSTEMS

BY JOSH MYRANS, FEIFEI ZHEN AND ZORAN KAPELAN

Given the growing scarcity of clean freshwater sources, the water industry as a whole has largely focused on the sustainable distribution and security of potable water. However, the less glamorous task of wastewater management is a constant pressure for all, requiring an equally significant investment into research and development. As if to further highlight this problem, the average age of sewer pipes in the UK is rapidly increasing, with many pipes still in service long past their intended lifespan. This article explores the advances in machine learning which are helping to better manage wastewater (or sewer) networks.

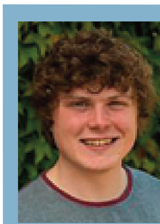
Since a wastewater network is often expected to collect sewers from all different water users in a particular urban region, its spatial scale as well as the structure complexity has typically substantially increased over the past few decades as a result of population growth and quick urbanization. These physical changes combined with system ageing result in a number of issues during the sewer network management or operation. Typical issues include (i) pipe blockages (e.g., sand sediments) that can directly affect flow capacity of the sewer pipes, causing manhole overflows and odour problems, (ii) illicit inflows (e.g., toxic discharges from local factories, rainwater, and groundwater) that may induce functional failures of wastewater treatment plants (WWTPs) and consequently result in significant contamination of the receiving water body, and (iii) leaks of the sewers that can directly induce serious contamination to the surrounding water environments. To solve these problems, deploying sensors in the sewer networks can be promising, aimed to detect or warn such events in an efficient manner.

Currently, sensors are often only placed at the end of a sewer system, monitoring treatment processes and discharges into the local environment. However, this is slowly changing with the introduction of low-cost robust sensors, providing the network visibility required to inform and improve pipe maintenance and rehabilitation. This constant stream of data can provide many insights into the status of a network, although many of these are hard to spot with only human eyes. Fortunately, machine learning thrives in the age of data, capable of interpreting patterns in vast quantities of data that no human being could ever hope to identify. These data driven techniques have been well demonstrated in many other professional sectors including telecommunications, gas/oil and finance, where inordinate quantities of data are produced every day.

Working with cutting edge AI technology provides the wastewater industry with a wealth of opportunities for more efficient means of practice. The strengths of machine learning include the ability to rapidly process and highlight trends and patterns in enormous volumes of data. From this skillset we can achieve the automation of tasks that would be extremely time consuming and tedious for a trained professional, real time analysis of sensor data and effective management of complex interrelated systems. This article will discuss a number of successful applications of machine learning within the wastewater sector, providing a number of examples, including one with more in-depth information.

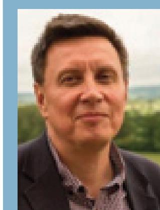
Machine Learning in sewer Management

Artificial Intelligence (AI) and Machine Learning (ML) in particular are playing an increasing role in the management of sewer systems, ranging from improved operation and maintenance of these systems to their long-term planning and asset management. Most of AI based solutions are built around smart processing of some data and extracting the useful information from it[5]. The data often comes from various sensors installed in these systems (e.g. level, flow and water quality sensors) but frequently from other sources too (e.g. inspection CCTV videos, digital maps, asset data, etc.). The current situation in most water and sewer utilities is often described as DRIP – Data Reach Information Poor. AI/ML enables to solve this problem by extracting useful information from large amounts of data and using it for improved management of sewer systems.



Dr Josh Myrans is a Data Scientist at the University of Exeter partnered with the UK water company: South West Water. He graduated from the University of Exeter in 2014 with a BSc in computer Science and Mathematics, before continuing to complete his PhD in Water Informatics Engineering in 2018 (also at Exeter University). Josh has been working within South West Water as a KTP associate since November 2018, where he continues to develop his postgraduate research for practical application within the water industry.

Professor Feifei Zheng, a PhD from the University of Adelaide, Australia. He is now a Professor in Zhejiang University, China, with research focus on design and operation of water infrastructures, hydroinformatics, as well as decision support systems for various water systems. He has led 12 research and consulting projects from national natural science foundation of China and various water utilities, and has published more than 50 journal papers in his research area.



Professor Zoran Kapelan is a Professor at the Delft University of Technology in the Netherlands where he is leading a research group on urban water infrastructure. He also holds a part-time professorial position at the University of Exeter in the UK. He is an IWA Fellow with 30 years of research and consulting experience in water engineering. His research interests cover a wide range of challenges related to water and wastewater infrastructure including development of various machine learning based technologies. Prof Kapelan pioneered the award winning burst/leak detection technology that is now used companywide in one of the largest UK water utilities resulting in large savings via reduced operational costs. He has published over 150 peer-reviewed journal papers.

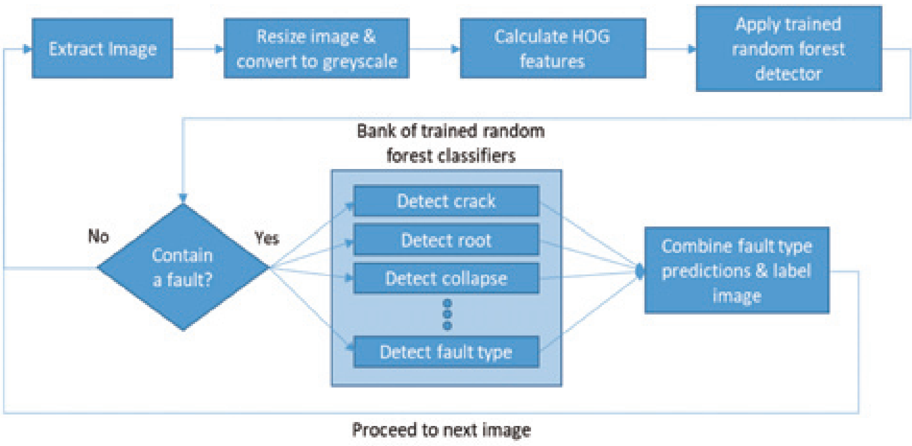


Figure 1. Flowchart depicting the process of applying automated labelling to raw images.

Some of the examples of ML methods developed for sewer systems include:

- ML for predicting sewer collapse/blockage rates and the remaining asset life.
  ML methods such as Artificial Neural Networks and Decision Trees have been used to predict sewer collapse / blockage rates that are critical for proactive asset management of sewer systems [1]. Most of these methods work by establishing a link between the above variables and potential explanatory factors such as sewer characteristics (e.g. pipe material, diameter, slope, condition), the environment (e.g. soil type, weather) and other factors (e.g. maintenance level). This data is used by the AI method to effectively learn under what combination of conditions sewer blockages or collapses occur.
- Early warning systems for blockages and other events in sewer systems. ML methods such as advanced Artificial Neural Networks and Fuzzy Theory have been combined with fault detection and isolation methods such as Statistical Process Control to detect or even predict blockages in sewer systems by raising alarms in near real-time [10]. Detection is typically done in the case of more instantaneous blockage events whereas prediction is usually more accurate for the gradually forming blockages (e.g. due to siltation or fat/oil/grease build up).
- Flood risk assessment and forecasting. The Cellular Automata based methodology has been used to predict the extent of flooding in the urban environment [6]. When compared to more conventional methods, these and similar ML-based methods tend to be computationally much faster yet accurate enough which enables their application over much larger geographical areas and/or in flood forecasting in the near real-time context.
- Augmented Reality (AR) for improved visualisation and inspection of sewer system assets. AR methods that combine Virtual Reality with conventional video feeds have been used to enable improved visualisation of sewers and other underground assets. This may involve presentation of other data of interest (e.g. asset characteristics, current or predicted water level at the location, etc.). These methods provide great help to technicians doing work in the field.
- Sewer self-cleansing. ML methods such as Random Forests have been used to develop models that can predict threshold flow conditions that lead to self-cleansing conditions in sewers [8]. This, in turn, can be used for the (re)design of these systems that ensures more effective sediment transport in sewer systems.
- Real-time (online) modelling of sewer system. Data is crucial to enable the applications of various ML methods. Unfortunately, in many cases system state observations (e.g., i.e. flows, water depth and other state variables) are scarce. Sensor data can be used to enable the estimation of sewer system state at different locations in the system, especially where sensors are not present. For example, a research group from Zhejiang University in China has successfully utilized the water supply data in a novel way to drive the real-time simulation of the wastewater network [11]. The key feature of this modelling approach is the novel use of smart demand metering sensors from the water supply systems to enable more accurate state estimation of sewer systems. This, in turn, enables to develop real-time sewer models in a more cost-effective manner.
- Real-time sewer sensor data validation. Bayesian type methods have been combined with Neural Networks and Interval Mathematics to validate sensor data on flows, depths, electro-conductivity) in near real-time [2].

Note that the above examples present only a small sample of AI/ML methods and applications for improved management of sewer systems. The next section presents another, more detailed example of a successful ML-based solution for solving a real-world challenge in these systems.

Automated sewer condition assessment using CCTV analysis

Background

Currently the most common method of inspection for sewers is through the use of CCTV cameras, which traverse the network recording footage of the pipe interiors for analysis by trained technicians. These surveys are performed regularly and are vital to the effective maintenance of the network. However, most networks contain tens, if not hundreds of thousands of kilometres of sewer pipe, resulting in a constant stream of CCTV footage which must be manually reviewed. The labour-intensive nature of this task, makes it both time consuming and expensive. Furthermore, surveys are commonly mislabelled due to subjective fault codes and pure human error. With some cameras footage can instead be labelled as it is collected, making the process more efficient. However, the accompanying analysis is often even worse, with technicians now performing multiple jobs at once, working in the elements and often next to a busy road.

Fortunately, AI can begin to improve upon this vital practice, automating elements of the analysis procedure in real time, so as to take the pressure off of the surveyor. Not only should this improve the speed and efficiency of a survey's collection, but dramatically reduce the cost and improve the uniformity of analysis. Removing the pressure of annotation from the surveyors enables them to concentrate on capturing high quality footage, only requiring additional input for the annotation of the most obscure faults.

AI-based methodology

To achieve automated fault detection and classification, a number of cutting edge machine learning and computer vision techniques are applied, namely random forests [3] and HOG (Histogram of Oriented Gradients) features [4]. In combination with a large dataset of labelled CCTV images these tools can first identify the presence of faults within an image, continuing to predict each individual fault type. This is done according to current industry standards, so as to produce a simplified report similar to that already used by the industry. Given the expedient and transportable nature of these techniques, the entire process can be performed in real time on site, in an office or on a server.

The procedure can effectively be broken down in to five stages: 'Frame Extraction & Pre-processing', 'Feature Extraction', 'Detection', 'Classification' and if applied to continuous footage 'Smoothing' [9]. The tasks associated with each stage are presented in the process diagram shown in Figure 1 require the collection of the image from the source video before re-sizing the image to match a uniform resolution and converting to greyscale. These two steps bring the data in line with the training set and eliminate unnecessary complexity from the image. This complexity is further reduced during the 'Feature Extraction' stage, where the image is reduced to a much smaller string of values representing its key components, this is done using HOG feature description. The next stage 'Detection' passes the feature descriptor to a single pre-trained random forest, which predicts the probability of the original frame containing a fault. If this is below a pre-determined threshold, the image is labelled as normal and the cycle restarts on a fresh image, otherwise a fault has been identified.

Once a frame is suspected to contain a fault the 'Classification' stage can occur, in which the feature descriptor is passed to a bank of random forests. Each of these random forests predicts the probability of the image containing a single fault type, i.e. that there is a single forest for cracks, a single forest for root intrusions etc. By combining and evaluating these predictions in a pairwise manner, a list of the most probable fault types can be identified for this image. Finally, if the image has been extracted from a continuous video source, additional information can be gained by comparing predictions to those of neighbouring frames. This is achieved during the 'Smoothing' stage, which applies a median filter among other techniques to process the entire sequence of predictions throughout a video. Amending predictions in this way massively reduces the impact of noise and eliminates many isolated misclassifications, producing a list of predictions much more in line with a surveyor's expectations.

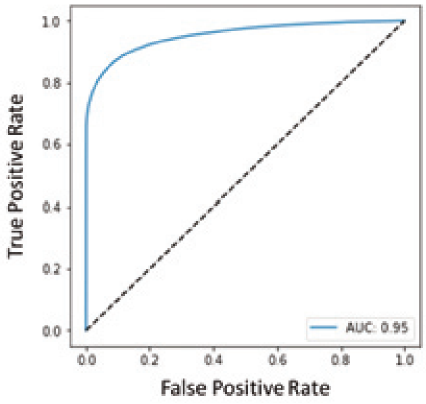


Figure 2. Receiver operator characteristic curve, demonstrating the range of achievable true (TPR) and false (FPR) positive rates. The dashed line represents the TPR and FPR for a 50:50 guess. Finally, the AUC (area under curve) is a measure of the methods overall performance.

It should be noted that all random forest classifiers will require training on a labelled dataset of images, processed using exactly the same 'Frame Extraction & Pre-processing' and 'Feature Extraction' stages as those intended for use on the video. This training sees each tree in a forest grown by randomly selecting features and splitting the training dataset according to their pre-assigned labels.

Results

This automated fault analysis has been performed in collaboration with the UK water company South West Water (SWW). This has granted access to a library of over 60,000 images, around half of which contained at least one labelled fault. In order to demonstrate the AI technology all these images are utilised via 25-fold cross validation [7]. This system ensures that training and testing datasets are not mixed, whilst making the most of the available data. Furthermore, the data has been arranged so as no images from the same pipe are present in both a training and testing fold.

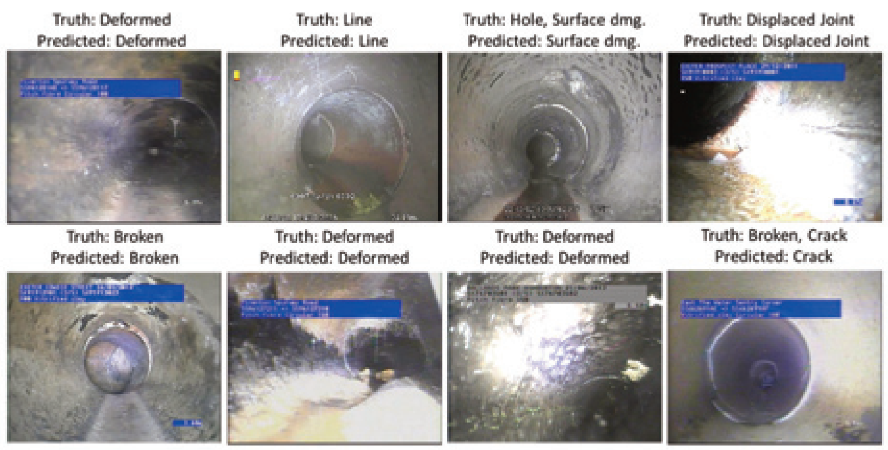
When the above approach was applied to the full dataset of labelled images an accuracy of 88% with a true positive rate (TPR) of 0.98 and a false positive rate of 0.24 was achieved. This means that the methodology correctly identified the status of the pipe 88% of the time, whether that be normal or faulty. Additionally, from the misidentifications, only 2% were missed defects and 24% were mislabelled normal pipe. By modifying the threshold on which an image is classified as faulty, the ratio between TPR and FPR can also be tweaked, as demonstrated by the receiver operating characteristic cure shown in Figure 2.

Applying the process of classification to detected faults, we must now acknowledge that a single image can contain multiple fault types. To do so, the methodology's results are evaluated using intersection over union (IoU), which measures the similarity of the predicted list of fault types with the true list of fault labels for a given image. This is a much more challenging task, assuming an image contains only a single fault, guesswork alone will only achieve an IoU of 6% (as we are using 18 different labels).

Although only a prototype, the methodology performs well, achieving an IoU of 35% and an accuracy on the primary fault of 70%. This performance is constantly improving, with the increased availability of high-quality labelled data. A handful of examples are shown below in Figure 3.

It is also worth noting that these results are achieved using the labels assigned by the human observers which we know can be

Figure 3. Example images and the classifications generated by the machine learning algorithm.



inconsistent. A recent quality survey of 5% of the dataset found more than 30% of the labels to be incorrect, and 10% of them to be uninterpretable. Anecdotally this is good for the industry in general, however this does not bode well for the performance of data driven methodologies such as this.
This first step in the application of AI to the problem offers a great option for screening vast amounts of CCTV footage. It is much quicker than human analysis and can be performed outside of work hours in a massively parallel manner. Given its current role as a decision support tool, it can assist with operational efficiency, but continued development and increased data quality provide great prospects.

Conclusion

This article addresses the use of Artificial Intelligence and machine learning in particular in the daily management of sewer systems. Several examples of such applications are provided including the technology for automated detection of faults in sewers.

This technology is a good example of how machine learning and AI can be influencing the wastewater sector. Current practices rely on the slow and expensive, human based coding of CCTV sewer surveys that is not always fully reliable. The machine learning based technology enables the automation of some of that process, accurately and more consistently identifying the presence of faults whilst providing a good estimate of potential fault types. Therefore, the AI-based solution has a great potential to help technicians do their job more effectively in the future whilst reducing related costs.

Based on the above and other examples presented in the paper it is clear that the future of AI and machine learning in the wastewater sector is bright and that the full potential of these methods is yet to be fully explored. ■

References

[1] Bailey, J., Harris, E., Keedwell, E., Djordjevic, S. and Kapelan, Z., (2016), "Developing Decision Tree Models to Create a Predictive Blockage Likelihood Model for Real-World Wastewater Networks", Procedia Engineering, 154, 1209–1216, https://doi.org/10.1016/j.proeng.2016.07.433.

[2] Branisavljevic, N., Kapelan, Z. and Prodanovic, D., (2011), "Improved Real-time Data Anomaly Detection using Context Classification", Journal of Hydroinformatics, 13(3), 307-323, https://doi.org/10.2166/hydro.2011.042.

[3] Breiman, L., (2001), "Random forests", Machine learning, 45(1), 5-32, https://doi.org/10.1023/a:1017934522171.

[4] Dalal, N., & Triggs, B. (2005), "Histograms of oriented gradients for human detection", 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), 886-893, https://doi.org/10.1109/cvpr.2005.177

[5] Draude, S., Keedwell, E., Harris, E., Hiscock, R. and Kapelan, Z., (2019), "A statistical analysis on the effect of preceding dry weather on sewer blockages in South Wales", Water Science and Technology, 80 (12), 2381–2391, https://doi.org/10.2166/wst.2020.063.

[6] Guidolin, M., Chen, A., Ghimire, B., Keedwell, E.C. Djordjevic, S and Savic, D., (2016), "A weighted cellular automata 2D inundation model for rapid flood analysis", Environmental Modelling & Software, vol. 84, 378-394, https://doi.org/10.1016/j.envsoft.2016.07.008.

[7] Kohavi, R. (1995), "A study of cross-validation and bootstrap for accuracy estimation and model selection", International Joint Conference on Artificial Intelligence (IJCAI), 44(2), 1137-1145, https://www.ijcai.org/Proceedings/95-2/Papers/016.pdf.

[8] Montes, C., Vanegas, S., Kapelan, Z., Luigi, B., and Saldarriaga, J., (2020), "Non-deposition self-cleansing models for large sewer pipes", Water Science and Technology, wst2020154, https://doi.org/10.2166/wst.2020.154.

[9] Myrans, J., Kapelan, Z. and Everson, R., (2018), "Automated detection of faults in sewers using CCTV image sequences", Automation in Construction, vol. 95, 64-71, https://doi.org/10.1016/j.autcon.2018.08.005.

[10] Rosin, T, Romano, M., Keedwell, E. and Kapelan, Z., (2019), "Near Real-Time Detection of Blockages in Wastewater Systems using Evolutionary Artificial Neural Networks and Statistical Process Control", Proc. 10th International Novatech Conference, 1-5 Jul 2019, Lyon, France.

[11] Zhang, Q., Zheng, F., Jia, Y., Savic, D. and Kapelan, Z. (2021), "Real-time foul sewer hydraulic modelling driven by water consumption data from water distribution systems". Water Research, vol. 188, 116544, https://doi.org/10.1016/j.watres.2020.116544.