Anomaly Detection in Railway Infrastructure

David Morandi and Stephan Jüngling

FHNW University of Applied Sciences Northwestern Switzerland, School of Business, Peter Merian-Strasse 86, 4052 Basel, Switzerland,

Abstract

In order to keep complex railway systems fail-safe, sophisticated maintenance of the rolling stock and infrastructure are most essential. Although AI-based predictive maintenance systems exist in many different industries, there is still quite large potential for different application scenarios. The current research shows such an example, where machine learning can be applied to detect anomalies in the pantograph-catenary system by using a simple convolutional neural network that is able to detect arc ignitions during train operation. The paper provides some insights into the process of the system development life cycle. Starting from the initial idea to use machine learning for anomaly detection, over the system design of a prototype and the training of the Keras-based machine-learning model, up until the evaluation of the conducted experiments. The arcVision system prototype provides valuable insights into how a predictive maintenance process could be established by combining the results from the machine-learning model with rules and insights from manual inspections.

Keywords [1]

Anomaly Detection, Railway Infrastructure, Predictive Maintenance, Arc Ignition, Machine Learning, CNN, Keras, System Development Life Cycle, Knowledge Engineering

1. Introduction

For many large technical systems, outages are either not tolerable or too costly. Therefore, many traditional processes exist to maintain complex systems to find anomalies and fix potential defects in advance. In most of these traditional cases, we use best practice rules, most of them implementing maintenance intervals, based on parameters such as time, frequency or intensity of use. However, lately and specifically in the context of the advances in machine learning, many novel types of predictive analytic solutions are used to further optimize the traditional, mostly time-based maintenance cycles. Predictive Maintenance (PdM) is one of the fields where AI-based prognostics and health management (PHM) could successfully be implemented in various industries such as construction-, automotive-, steel-, or in aeronautics and logistics. Furthermore, with the introduction of IoT, smart factories and smart cities in the context of Industry 4.0, standardized reference models such as RAMI 4.0 [1] (Reference Architecture Model Industry) are introduced, which provide guidelines for the implementation of predictive maintenance systems as described in the use case of aeronautics supply chains industries [2]. However, a recent systematic literature review of academic papers from the past five years about PdM by Dalzochio et al. [3] also collected existing challenges in applications of many different machine-learning techniques, while pointing out a high demand in further research investigations in the area of applying machine learning and reasoning in the context of Industry 4.0.

Despite the railway industry being an old and traditional industry, digitalization and the potential of predictive maintenance will not spare this industry. The complex railway system with its security

In A. Martin, K. Hinkelmann, H.-G. Fill, A. Gerber, D. Lenat, R. Stolle, F. van Harmelen (Eds.), Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021) - Stanford University, Palo Alto, California, USA, March 22-24, 2021.
E-mail: david@morandi.me, stephan.juengling@fhnw.ch

measures, which need to be fail-safe and the sophisticated maintenance of the rolling stock and infrastructure are most essential. For example, the European Union [4] has the aim to increase the rail competitiveness and pursues specific goals such as decreasing the life-cycle cost of railways by 50%, doubling the capacity and increasing the punctuality by 50% achieved by digitalization. It is important to prevent incidents as efficient as possible, since even a minor local interruption can quickly evolve to a major disturbance in the operation of the overall railway system. Therefore, the vision for the conducted master thesis [5] has been created to identify a potential use case for anomaly detection in the railway industry using artificial intelligence.

2. Current State of Applications of AI in the Railway Industry

Whenever safety plays an important role, and the resulting cost of manual maintenance effort is high, there is a potential to apply PdM using automatic anomaly detection. Nevertheless, the possible business cases, where AI could be used, to overall reduce the amount of manual inspection work are manifold, and therefore triggered the first two research questions:

- **RQ1:** Which area contains the biggest potential for anomaly detection in railway infrastructure?
- **RQ2:** What is the current state of applications of anomaly detection in railway infrastructure?

The answer to these two questions can be derived with a literature research. One of the main aspects, when talking about inspection of railway systems is the borderline of manual work done by humans versus inspections that can be automated with the help of AI-based systems. Al-Douri et al. [6] state that railway infrastructure is very complex, covers a large area and consists of many subcomponents, which are often difficult to maintain. Numerous stakeholders, weather conditions, physical components as well as administration, traffic situations and new investments need to be considered for the maintenance of the system. The infrastructure can be further split into technical subsystems, such as the substructure, track, electrical system, signaling system, and telecom system.

Gibert et al. [7] state that periodical inspection and monitoring is needed to ensure safe transportation. Safety can be improved by more frequent inspections, reducing human errors and automating line inspections using computer vision and pattern recognition methods. In contrast, full automation cannot yet be implemented due to the number of different possible error modes and the wide range of image variations that can potentially trigger false alarms. In addition, the number of faulty components is very small, so that to little training data is available for the machine to train a robust visual anomaly detector. Siemens Ltd. [8] has a system, which detects broken rails. They equip trains with sensors on one of the bogies. During operations, the sensors scan the tracks and send them to the control center. If a broken rail is detected, an alert is sent to the subsequent trains and the line can be blocked by the control center or the existing signaling solution.

Switches are as well critical sub-systems in the railway infrastructure, as they, if malfunctioning can lead to serious accidents. Either the train can derail or switch to the wrong line. Guzman et al. [9] developed an anomaly detection model and integrated it in an intelligent workflow, which is used in an operational environment. The model can calculate an anomaly score which is then aggregated with meta-data of the corresponding switch (e.g., last maintenance, number of failures) and can take context information such as different weather conditions or other environmental factors into account. However, they concluded that the model could be further improved by using more features.

There is a wide range of potential areas for anomaly detection in railway infrastructure, and not only the rails and switches, but also defects in the catenary system should be considered. One of the key factors that influence the operation quality of high-speed trains is the electric contact between the pantograph and the catenary system, as stated by Wu et al. [10]. Combining this fact with the insights from conducted expert interviews that the manual inspections of rails or switches are much better accessible for humans than the inspection of the catenary system, the idea of narrowing down the scope to anomaly detection in the pantograph-catenary system gave the idea for the thesis statement for the conducted master thesis of Morandi [5]. The thesis hypothesis states that it is possible to design a prototype (*arcVision System*) which makes use of machine learning techniques, with the goal to help detecting arcing in the pantograph-catenary system. According to Wu et al. [10], arcs generate areas of high temperature on the strips, which causes the material to melt and evaporate, resulting in erosion of the wire and strip or high thermal temperature gradients, which cause thermal stress and potential breakage of the strip material. Predictive maintenance processes could be established with the help of digitally recorded sensor data that detect arcing during train operation. With the help of supervised learning, an arc detecting system could be trained, that helps to build up a predictive maintenance system for the pantograph catenary system.

3. Business Case of the arcVision System

With almost 29'000 kilometers, Switzerland has one of the densest public transport networks in Europe [11]. In consequence, a dense timetable, professional personnel and reliable rolling stock/infrastructure is needed to meet this high demand and to ensure fast and safe transportation services. Despite its dense regular interval timetables, Swiss railways are able to use their infrastructure in an optimized way. Faulty infrastructure or rolling stock can cause delays, and even minor disturbances could provoke major disruption to the entire timetable.

There are many causes for train delays. Some of them are predictable and some of them are yet unpredictable. Regular maintenance can minimize the risk of delays and accidents. However, periodic maintenance programs are time-consuming and cause unwanted idle times. Furthermore, it is important to identify the source and root cause of the defects in order to determine appropriate maintenance activities. Attrition or fatigue of material is one of the main causes for such maintenance.

During maintenance of the rolling stock, train parts, which have passed their operational lifetime and further parts, which show more attrition of material than expected prior to the expiry of the guarantee period are replaced. These procedures can be very time and resource intensive. To prevent additional defects and maintenance, it would be helpful to understand the source of such attrition and to localize the root causes. With the help of sensors, it is possible to collect plenty of data. However, manual extraction, evaluation and anomaly detection of the collected sensor data is a highly time-consuming process. In this case, artificial intelligence could help to minimize the time needed for anomaly detection. After evaluating the data, the source of attrition can be localized and possibly increase the operational lifetime of the affected parts on one hand and reduce time and effort for maintenance on the other.

In order to localize the ignition from the electric spark in the pantograph-catenary system during train operations, the time of the occurrence, the location (e.g. in a tunnel), the acceleration of the train (e.g. mechanical forces) or further context information (e.g. weather condition) could have an influence. As described by Wu et al. [10] the electrical contact between the pantograph and the catenary is rather complex and variable. The operating performance of the electrical contact system of pantograph and catenary depends on the contact resistance, contact surface heat, friction and wear. If arcing in the pantograph and the catenary contact occurs regularly, it can cause unusual wear and even breakage of the graphite strips, which leads to malfunction power supply respectively of the train. Therefore, it would be helpful to know all these parameters mentioned above in order to determine why arcs occur and where the catenary system should be inspected. Frequent arc ignition will accelerate attrition and faster loss of material results in more frequent maintenance or malfunctioning, which eventually causes a higher idle time or an interruption in operations.

The conducted literature review has revealed that until now, only limited research has been conducted on automatic detection of anomalies in the pantograph catenary system using machine learning techniques. Consequently, the feasibility of such an anomaly detection raised the following two additional research questions:

- **RQ3:** Is it possible to design a prototype (*arcVision System*) which makes use of machine learning techniques, with the goal to help detecting arcing in the pantograph-catenary system?
- **RQ4**: To what extent can *arcVision* assist a human inspector and where are its limitations?

4. Prototype Design and Architecture – arcVision System

For the envisioned prototype, a supervised learning approach was chosen, for which large amounts of test data needs to be recorded and labeled. The *arcVision System* prototype was developed in cooperation with Regionalverkehr Bern-Solothurn (RBS), a Swiss public transport company in the region of Bern and Solothurn. RBS operates a fleet of 49 trains, a train network of 53.9 kilometers and the corresponding infrastructure. The *arcVision System* contains two different parts. The *arcVision Scanner*, which collects the necessary sensor and context data. The *arcVision Model* consists of a Convolutional Neural Network (CNN) for the training and arc detection of the recorded pictures from *arcVision Scanner* to do a binary classification. Figure 1 shows the envisioned architecture of the system.

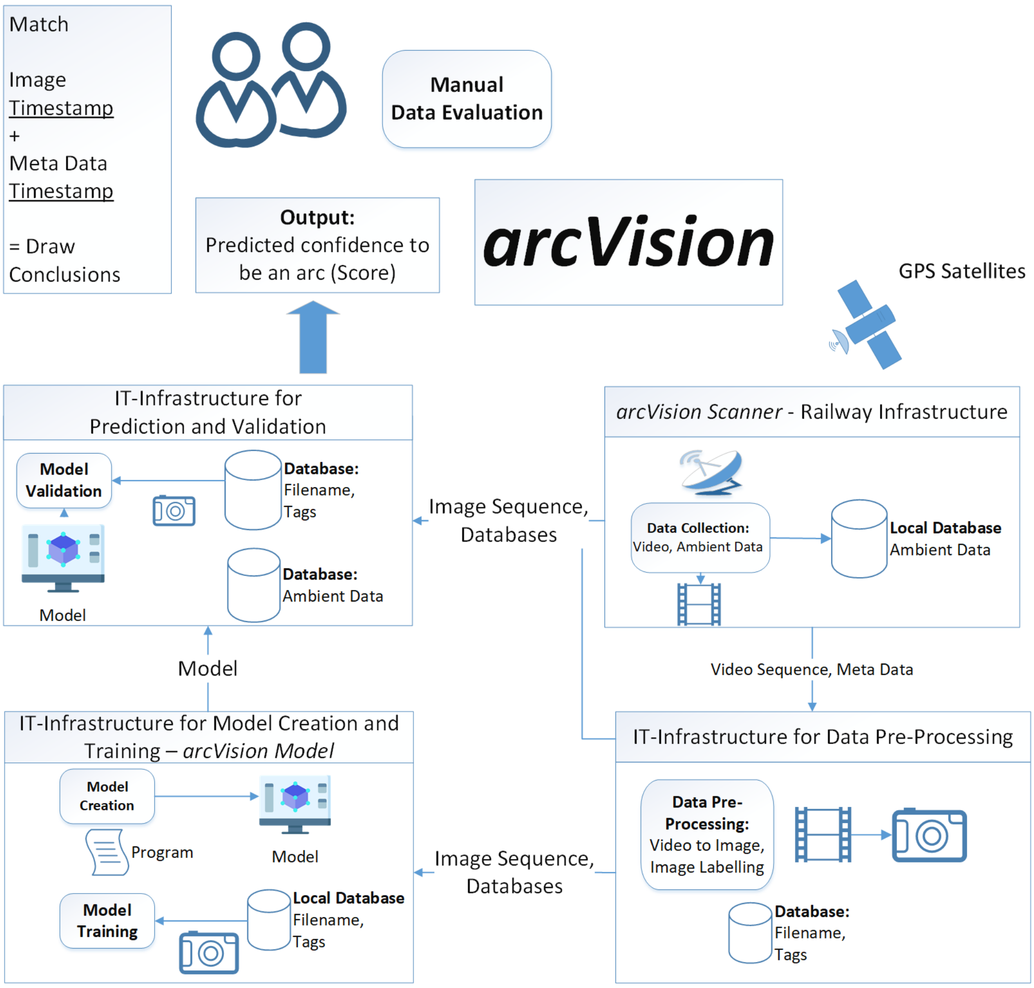


Figure 1: *arcVision System* Architecture

During train operation, data is collected with the *arcVision Scanner*. It consists of a single board computer, a camera and different sensors such as GPS, accelerator, barometer, thermometer and hygrometer and is mounted on the train's roof. The camera's field of view needs to be aligned with the touching point of the pantograph and the catenary. The collected video sequences could then be used by the *arcVision Model*, with an adequate pre-processing such as image extraction and the corresponding labeling of the images.

In a second step, the video sequence is divided into single frames. As no pre-trained model or pre-labelled data set can be used at the beginning, a screening of all video/image data is needed to search for potential pictures with arc ignitions to be labelled. The filename of the images and the corresponding label is then stored in a file-based database. The high effort for labeling the pictures is well known in literature [12], and different techniques such as data subset selection and active labeling [13] or self-supervised learning [14] have been developed, to reduce the labeling effort. The sparsity of the pictures with arcing turned out to be only 0.02% and the task resembled to the well-known situation of finding the needle in a haystack. However, due to an iterative approach, where the first arc picture could have been found, the initially badly trained model could never the less help to find further candidates of pictures with arc ignitions.

The *arcVision Model* is a CNN, based on Keras and TensorFlow, which allowed for a very fast experimentation with the before-mentioned iterative approach, where the entire dataset could be labeled, and the model be trained more effectively. The training was performed on a regular personal computer with a dedicated GPU, which turned out to have sufficient computing power for the machine learning task at hand.

During the fourth step, the pre-trained model was then used to predict new data sets that were collected in subsequent test-runs. The data collection is described in further detail in the next chapter.

In the very last step, the images, which showed arc ignition can be combined with the additional sensor data (e.g. temperature, GPS data) where additional context data can be recorded and analyzed for the anomaly detection process, which in the end, allows domain experts to draw conclusions for the PdM tasks.

4.1. Data Collection – arcVision Scanner

As mentioned before, to collect the necessary data (video and sensor), a corresponding device such as the *arcVision Scanner* was designed during the master thesis. It consists of a Single Board Computer (Raspberry Pi 4 Model B), a GPS receiver, a Raspberry Pi camera module and external temperature and humidity sensors.

As can be seen in figure 2, the prototype device was designed iteratively until the final packaging and mounting box was ready that resisted the various weather conditions. The sensitive electronic parts needed to be protected not only from rain, but also from fine dust. On the other hand, a continuous power supply is necessary, which can be taken most of the time from the train via the compressor control system. Since the train is in some situations decoupled from the catenary system, an additional power bank is used that can provide power for up to 40 hours. Furthermore, different adjustments were necessary to find the optimal angle and distance to the pantograph, in order to cover the entire contact range with the catenary system, where potential arcs can be detected. With the help of a Python script, all sensor data is stored in a file-based database.

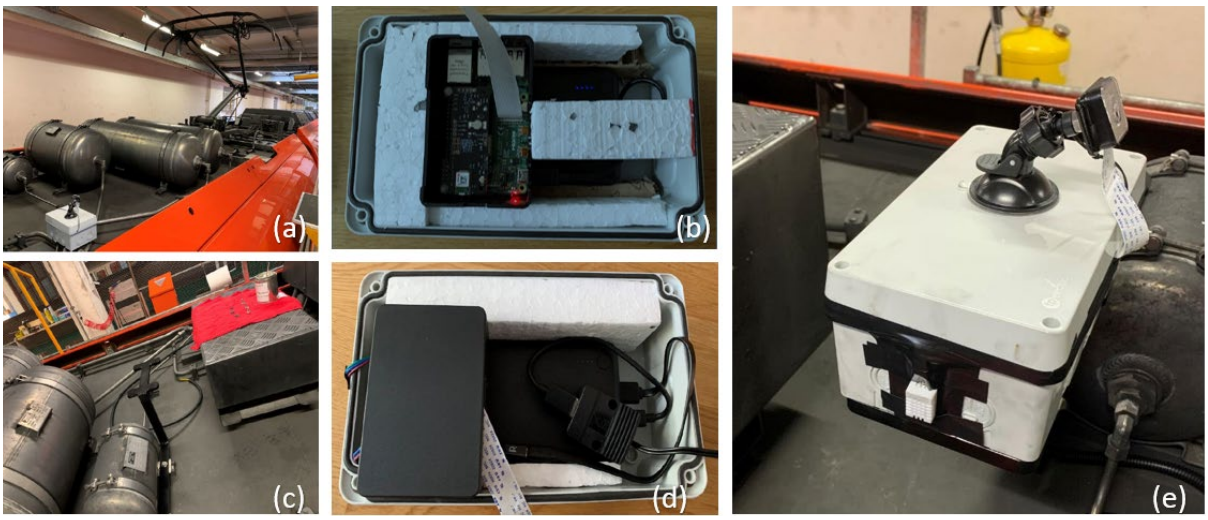


Figure 2: *arcVision Scanner (a) Temporary mount during first field test, (b) Packaging during first field test (without external power supply), (c) Final mount box, (d) Final packaging, (e) Final mounting*

4.2. Model Creation – arcVision Model

For the model creation and training, a CNN model, based on the frameworks TensorFlow, Keras and several other state-of-the-art tools such as numPy, scikit-learn, pandas, and opencv was used. Keras provides a deep learning API, running on top of the machine learning platform TensorFlow. When developing Keras, the focus was laid on allowing fast experimentation, being able to go from the idea to result as fast as possible, which is visualized and suggested by the Keras Special Interest Group [15] and shown in figure 3.

With increasing popularity of machine learning, many different pre-trained models exist. A recent survey pointing out the importance of transfer learning analysed the different approaches how existing pre-trained CNN networks can be re-used [16]. However, a pre-trained model (e.g., one for lightning detection), which would be potentially able to recognize arc ignition, could not be found. Furthermore, in our specific case of detecting arcs, the advantage of transfer learning, where lots of labelling effort and CPU hours from previous trainings could be reused to extend large sets of existing classes by this additional class, does not provide any advantage. Therefore, the CNN for the *arcVision Model* was built from scratch.

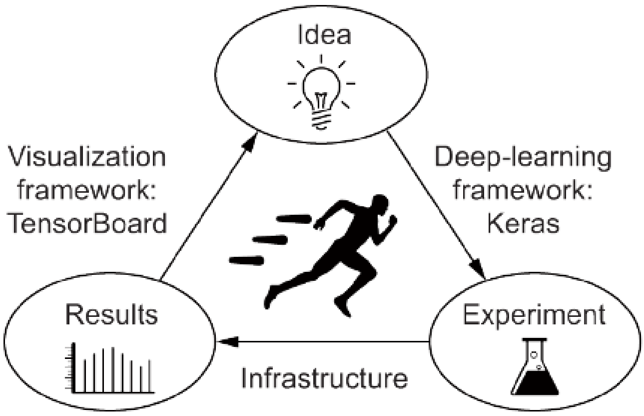


Figure 3: - The loop of progress, from Keras Special Interest Group [15]

A sequential Keras model was created and its layers are visualized using TensorBoard (see figure 4) as described in the loop of progress in figure 3. The model follows suggestions from O'Shea & Nash [17] and uses a plain stack of alternating convolutional and pooling layers followed by a flattening and dropout layer in the end before eventually feeding to the dense, fully connected layer.

The training and validation of the model were performed on PC with a decent graphics card (Nvidia GeForce GTX 980 Ti) and needed at least 500GB free space (SSD). Despite slightly outdated hardware, the task could be still performed successfully. During training, the model analysed 1'319 images with arc ignition and 4'194 images without arc ignition (total 5'513). As the numbers show, the classes were quite imbalanced. However, as stated by Kotsiantis et al. [18] such high imbalances occur in different real-world domains such as detecting oil spills in satellite radar images or detecting fraudulent telephone calls, where the aim is to detect an occasional but important case or event.

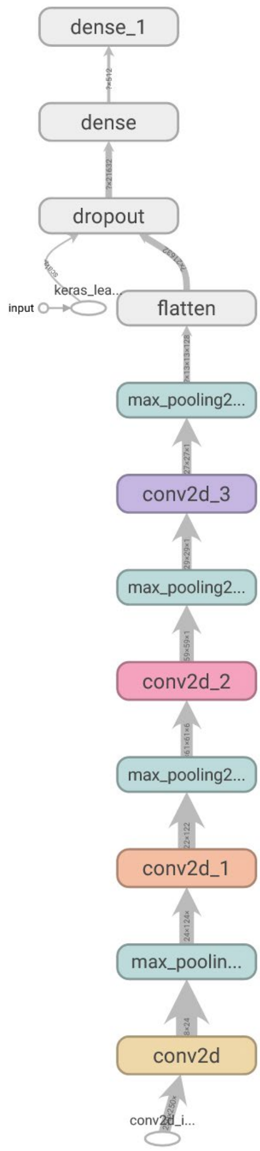


Figure 4: Model Structure

4.3. Experiments

As stated in [19], the characteristic system development life cycle (SDLC) methods from projects with pure classical software development tasks versus projects that include applications of machine learning might be different not only in terms of the skillset of the engineers, but also in terms of the suitability of the applied SDLC methodology. When talking about software engineering, test-driven development is considered as one of the “best practices” of software engineering. In the case of hardware related engineering tasks, where physical and environmental issues influence the experiments, testing is even more essential. In software development processes, APIs (application programming interfaces) can mimic real behavior and the code under development can be isolated from the environment. This is not as easy for the engineering and research task of designing the *arcVision Scanner* prototype. Nevertheless, partial testing can be implemented by breaking down experiments which are close to the reality into smaller experiments, which can at least provide partial insight into facts and constraints from the technical environment.

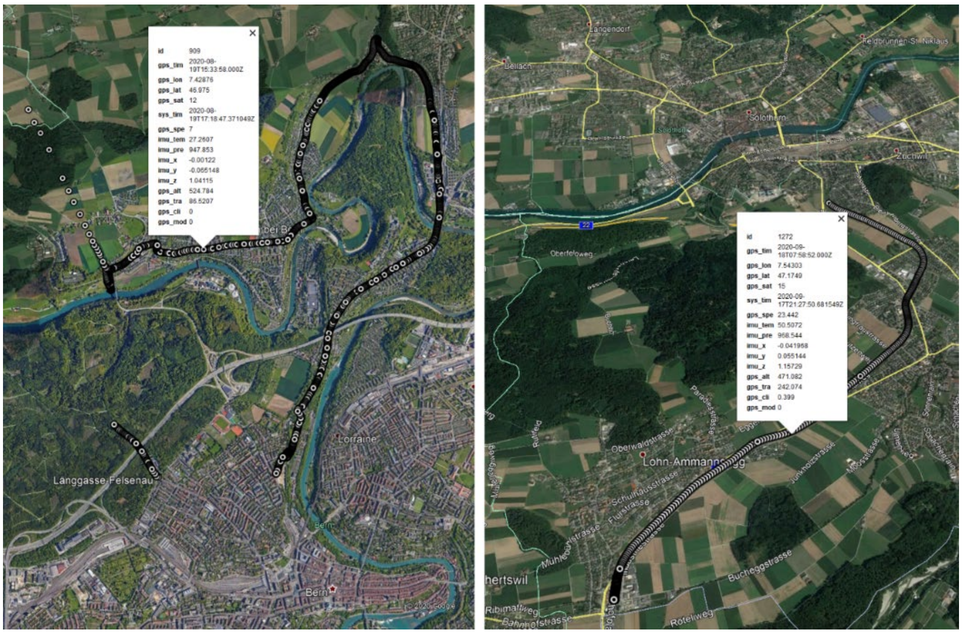


Figure 5: Field test maps (left: bike test, right: train)

To test the very basic functionality the first prototype iteration was equipped with a GPS receiver, different sensors (accelerator, barometer and thermometer) and a power supply in the form of a power bank. This test was conducted with a small bike tour, which started in the city of Bern, as can be seen on the left of figure 5. This test revealed the necessity of having a “time to first fix” (TTFF), which is necessary for the GPS receiver for an initial position fix. However, this first fix was achieved exactly at the entrance of the forest (northbound), which was approximately 5 minutes after the prototype was powered on and started. In the forest, the position could not be determined anymore, since the trees blocked the reception of GPS data from almost every angle under or over 90°. After leaving the forest, outlier positions were recorded. But shortly after about a minute the position could be adjusted thanks to the environment with an open sky. The rest of the track was correctly recorded until the small power bank ran out of power.

The second test iteration the *arcVision Scanner* was tested closer to real conditions and mounted on a train. In order to be optimally prepare for this setup, the capacity of the power bank was increased from approximately 1’000mAh to 21’400mAh. On the other hand, all devices were placed into an enclosure in order to ensure that the components are protected against spray water, strokes or dust. The ride was declared as a “not in service”-ride and therefore, no regular passengers were on the train. The test route started from the Solothurn train station to Lohn-Ammannsegg and went back to Solothurn with a total of approximately 30 minutes test time (see figure 5 – right side).

In order to collect all the necessary data in a final reality check, the *arcVision Scanner* was integrated into the regular daily operation time schedule for a few days on the regional express line Bern-Solothurn. It turned out, that the *arcVision Scanner* was always operational and even after rainfall, no damage could be found. Nevertheless, the roof of a train is very filthy, so the camera lens became quite dusty and needed to be cleaned. Nevertheless, the recorded data sets could be used for the validation and testing of the *arcVision Model*.

4.4. Model Validation

The validation data set contains 200 selected images with a balanced representation of the classes, where 92 had arcs and got the label "1" and 108 with no arc the label "0", whereas images with features and circumstances that rarely occurred were also taken into account. The accuracy of the model was optimized over several iterations by experiments with tuning the different training parameters such as batch size, number of epochs or different optimizers. At some point, the accuracy was considered as "good enough" and an accuracy of 74.5% could be achieved which is determined by the confusion matrix in table 1.

n=200	Predicted: Positive (arc)	Predicted: Negative (no arc)
Actual: Positive (arc)	76 (True Positives) (a)	16 (False Positives) (b)
Actual: Negative (no arc)	35 (False Negatives) (c)	73 (True Negatives) (d)

Table 1: Confusion Matrix

The data validation showed that the model is in general able to detect arc ignition with an adequate confidence as shown by some typical cases in figure 6.

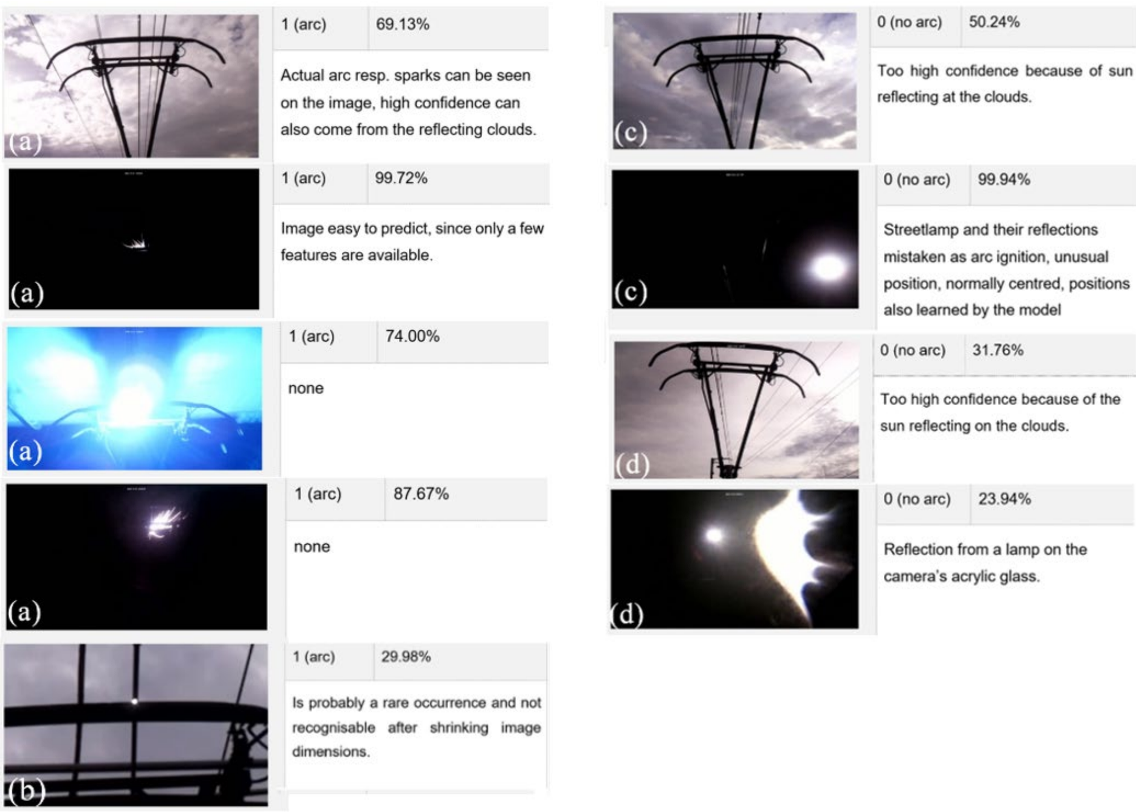


Figure 6: Typical classification samples – (a) true positives, (b) false positives, (c) false negatives, (d) true negatives

Nevertheless, the confusion matrix indicates that the model has problems with false positives and predicts an arc ignition where no arc ignition happened. Furthermore, the model shows a tendency of overfitting since it was trained with a small data set and possibly learned irrelevant features.

During the validation, the authors noticed some problematic cases where the prediction was wrong, as shown in figure 7. Especially rainy images, lamps, reflections, and small artefacts of arc ignition were problematic. In addition, the training data was captured in a short timeframe during a few days in autumn weather.

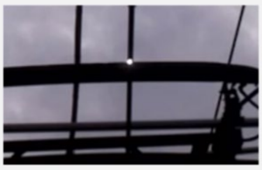
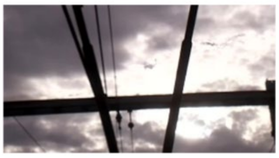
Image	Predicted confidence to be Class "1"	Remarks
	29.98% false positive	Is probably a rare occurrence and not recognisable after shrinking image dimensions
	76.02% false negative	Bright reflections in from the clouds in a position where also arcing could happened.

Image	Predicted confidence to be Class "1"	Remarks
	21.78% false positive	Only a few sparks recognisable and probably the same problem a mentioned in the first image, the bright clouds are in a position where normally no arcing happens.

Figure 7: Problematic cases

Based on the conducted experiments and the model validation, the **RQ3** can be answered as the follows: *Yes, it is possible to design a system, which makes use of machine learning techniques, and detects arcing in the pantograph-catenary system.*

However, the prototype needs further improvements in terms of data collection and optimization of the model in terms of accuracy. As we have seen in figure 6, the predictions differ heavily on the quite diverse situations and circumstances. Nevertheless, railway infrastructure domain experts evaluated and assessed the current prototype and there was a common understanding, that the *arcVision System* can provide very valuable insights in terms of finding potential defects early by analyzing the arc ignition pattern along the railway infrastructure system and the system seems to be a valid starting point for PdM. However, there was also a common understanding, that drawing conclusions out of such arc ignition patterns is still a task to be conducted by human experts. The potential overall improvement with a combination of ML and KE is addressed below. Reflecting these findings, the following chapter further elaborates on the combination of machine learning with knowledge engineering.

5. Combining Machine Learning with Knowledge Engineering

Based on the data of the *arcVision System*, a map with all arc ignitions can be generated. For all these data points, additional sensor data can complement the recorded data. Based on the current primary data of the sensors, such as time, x/y-coordinates, and acceleration in x/y-direction, humidity, air pressure, temperature and humidity, secondary data such as real speed and acceleration/deceleration of the train can be calculated.

Furthermore, the perceived intensity of the arc ignition is heavily dependent on static as well as dynamic context, as shown in figure 8. With static context, environmental situations of the train such as entering a tunnel, distance from the tunnel entry, entering or exiting stations, at the location to a switch/catenary split etc. can be determined and added to the dataset. Dynamic context information such as the brightness of the sky (e.g. on cloudy or sunny days) could be added, in order to determine the absolute intensity of the arc rather than the relative perception of the intensity, which depends on the background of the picture. Based on a hypothesis, that the intensity of the arc caused by a defect at a particular location in a tunnel has the same intensity for all trains could lead to the conclusion that the intensity of an arc outside of tunnels is equal for different trains as well. Given this reproducibility, the effect of different weather conditions (e.g. sunny/cloudy sky) on the perceived relative intensities on the recorded pictures could be learned and the absolute intensities be calculated from the perceived, relative intensities.

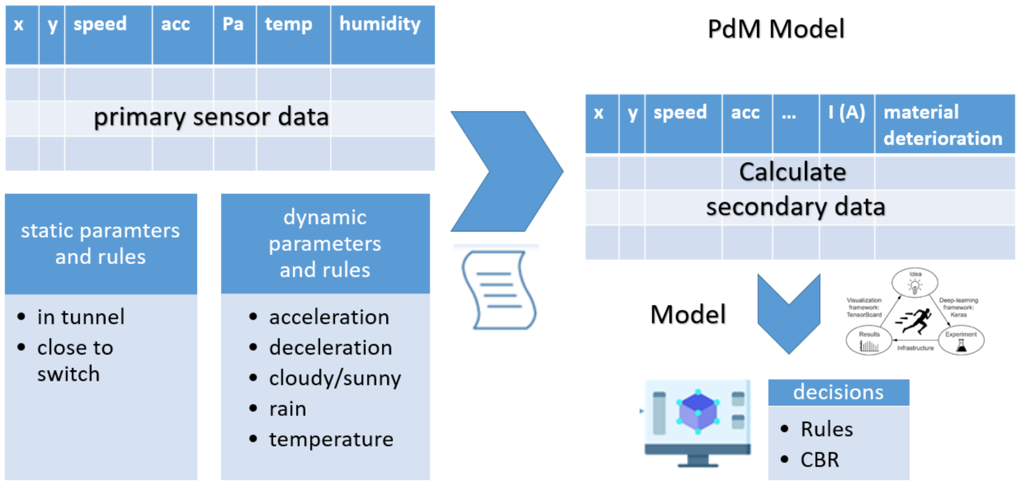


Figure 8: PdM Model

This new set of calculated secondary data could serve as new features to train a final PdM Model, which directly calculates a measure for presumed material deterioration based on the insights from the *arcVision System*.

Another option could be to combine the predictive PdM model with Case Based Reasoning (CBR), as demonstrated by Zhao et. al. [20] for PdM of railway turnout systems, which are most essential for the current high speed railway infrastructure in China. Suggestions from the machine-learning model could be compared to the results from existing cases in the case database. It is left to the humans to decide which suggestion they will follow in case of diverse results from the ML and the CBR model.

Based on the current state of insights from the ML model, which certainly needs more training data to be improved over time, some facts, rules can already be stated:

- The most intense arc ignition events are in tunnels (e.g. in the current set of the 22 most intense arcs in terms of brightness were 15 in a tunnel).
- Arc intensities seem to correlate with acceleration and deceleration of the train.
- The phenomenon of sparking can be observed throughout the line.
- Arc ignition happens more frequently while passing a switch, which is reasonable, since this always corresponds with a split of the wire in the catenary system, which causes the arc due to an air gap between the pantograph and the catenary.

However, this list of facts are currently assumption, which could not yet be proven based on the current pictures from the experiments. However, when more data is collected, these patterns can be validated, and rules for maintenance activities can be derived.

During the evaluation with domain experts the statement was made that the *arcVision System* offers an increased reliability in detecting existing arc ignitions, which were until now, only occasionally observed and reported by the train crew in cases when, they considered it to be important. So far, humans were not even able to detect smaller ignition events. Furthermore, the accuracy of the location of the arc ignitions reported by humans can be substantially improved. In cases of many arc ignitions on specific route sections, the depots have more time to plan the maintenance intervals and get more up to date information on wear conditions of the train engines. Up until now, the visual inspections were on a monthly base only.

The *arcVision System* can detect arc ignition patterns and provides some additional data about static as well as dynamic context information. The combination and aggregation of the available data may result in PdM tasks and activities that over time will provide more insights for human inspectors to draw conclusions and initiate appropriate measures. As such, human knowledge and expertise is still needed to properly categorize the cases and optimize the necessary maintenance activities. The *arcVision System* can assist humans by collecting enormous amount of data, detect anomaly patterns, visualize and analyze them in order to reduce the manual inspection work and increase the quality and reliability of the railway infrastructure. In consequence, the **RQ4** can be answered in terms of that the *arcVision System* can assist a human inspector in terms of providing and processing data on a large scale.

6. Conclusion and Outlook

It is important to detect potential incidents as fast as possible and prevent defects in the railway infrastructure, where even a minor interruption can evolve to a major disturbance. Predictive maintenance is possible to some extent with artificial intelligence, especially when using results from machine learning techniques to derive appropriate rules for the maintenance tasks with the help of knowledge engineering.

The development of the *arcVision System* has shown that it is possible to detect arc ignitions in the pantograph-catenary system, which can assist human inspectors to analyze the state of the health of the current railway infrastructure and draw the right conclusions when predictive maintenance is necessary. Although the current *arcVision System* prototype demonstrated that in principle, the model is able to detect arc ignitions with an adequate accuracy. Nevertheless, the confusion matrix shows that the model has still some problems with false positives and predicts arc ignitions, which do not exist. The model also shows signs of overfitting since it was trained with a small data set and possibly learned irrelevant features. By using Keras, the design of the CNN is easily possibly and the application of AI and machine learning is feasible within the context of a typical system development lifecycle. The amount of time for the design, implementation and testing of the prototype was quite reasonable and the return on investment of an AI-based solution development is feasible even with little prior knowledge about machine learning. Some valuable insights about the arc ignition patterns could be derived, which were out of reach before.

During the evaluation of the *arcVision System,* several railway operators were interviewed in a qualitative way. Some of their key findings were that today, only little effort was raised to address anomaly detection with machine learning techniques. Many hurdles need to be considered but the main message was the importance of the cost-benefit-analysis. Some concerns were mentioned about technology replacing humans. When AI-based systems such as *arcVision* can assist humans to reach a higher quality and better reliability in collecting and processing data, they are very valuable. When drawing conclusions, human knowledge and human wisdom were still considered superior.

7. References

[1] Reference Architecture Model Industrie 4.0 (RAMI4.0), https://www.beuth.de/en/technical-rule/din-spec-91345/250940128

[2] Hribernik, K., von Stietencron, M., Bousdekis, A., Bredehorst, B., Mentzas, G., Thoben, K.D.: Towards a unified predictive maintenance system - A use case in production logistics in aeronautics. Procedia Manuf. 16, 131–138, 2018. doi:10.1016/j.promfg.2018.10.168

[3] Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., Barbosa, J.: Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. Comput. Ind. 123, 103298, 2020. doi:10.1016/j.compind.2020.103298

[4] European Parliament: Digitalisation in railway transport A lever to improve rail competitiveness, 2019

[5] Morandi, D.: Anomaly Detection in Railway Infrastructure, Master Thesis University of Applied Sciences and Arts Northwestern Switzerland, School of Business, 2020

[6] Al-Douri, Y.K., Tretten, P., Karim, R.: Improvement of railway performance: a study of Swedish railway infrastructure. J. Mod. Transp. 24, 22–37 (2016). doi:10.1007/s40534-015-0092-0

[7] Gibert, X., Patel, V.M., Chellappa, R.: Deep Multitask Learning for Railway Track Inspection. IEEE Trans. Intell. Transp. Syst. 18, 153–164 (2017). doi:10.1109/TITS.2016.2568758
[8] Siemes Ltd.: Broken Rail Detection, https://assets.new.siemens.com/siemens/assets/api/uuid:246cb11523d1a7b06f5423baa17f4479cf93f951/version:1525252841/datasheet-mrx-brd-audisclaimer.pdf
[9] Guzman, D.N., Hadzic, E., Schuil, R., Baars, E., Groos, J.C.: Turning data driven condition now- and forecasting for railway switches into maintenance actions. (2018)
[10] Wu, G., Gao, G., Wei, W., Yang, Z.: The Electrical Contact of the Pantograph- Catenary System. Springer Nature Singapore Pte Ltd (2019)
[11] VÖV: Facts & Figures Swiss Public Transport 2016/2017. (2017)
[12] Culotta, A., McCallum, A.: Reducing labeling effort for structured prediction tasks. Proc. Natl. Conf. Artif. Intell. 2, 746–751 (2005)
[13] Kaushal, V., Iyer, R., Kothawade, S., Mahadev, R., Doctor, K., Ramakrishnan, G.: Learning from less data: A unified data subset selection and active learning framework for computer vision. Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vision, WACV 2019. 1289–1299 (2019). doi:10.1109/WACV.2019.00142
[14] Zhao, H., Chen, H., Dong, W., Sun, X., Ji, Y.: Fault diagnosis of rail turnout system based on case-based reasoning with compound distance methods. Proc. 29th Chinese Control Decis. Conf. CCDC 2017. 4205–4210 (2017). doi:10.1109/CCDC.2017.7979237
[15] Chollet, F.: Keras API Special Interest Group (SIG)
[16] Ribani, R., Marengoni, M.: A Survey of Transfer Learning for Convolutional Neural Networks. Proc. - 32nd Conf. Graph. Patterns Images Tutorials, SIBGRAPI-T 2019. 47–57 (2019). doi:10.1109/SIBGRAPI-T.2019.00010
[17] O'Shea, K., Nash, R.: An Introduction to Convolutional Neural Networks. 1–11 (2015), ArXiv:1511.08458 [Cs]. http://arxiv.org/abs/1511.08458
[18] Kotsiantis, S., Kanellopoulos, D., Pintelas, P.: Handling imbalanced datasets : A review. Science (80-.). 30, 25–36 (2006)
[19] Jüngling, S., Peraic, M., Martin, A.: Towards AI-based Solutions in the System Development Lifecycle. Proc. AAAI 2020 Spring Symp. Comb. Mach. Learn. Knowl. Eng. Pract. (AAAI-MAKE 2020) - Vol. I. 2600, (2020)
[20] Zhao, H., Chen, H., Dong, W., Sun, X., Ji, Y.: Fault diagnosis of rail turnout system based on case-based reasoning with compound distance methods. Proc. 29th Chinese Control Decis. Conf. CCDC 2017. 4205–4210 (2017). doi:10.1109/CCDC.2017.7979237