THE FUTURE OF TRANSPORT

PROJECT REPORT ACA108

The application of artificial intelligence to road pavement maintenance assessment

A proof of concept study

M Courtier, R Lee, P Sanders and R Dhammi

Report details

Report prepared for:	TRL Academy		
Project/customer reference:	SIP 21008 (Pathfinder - GLAM)		
Copyright:	© TRL Limited		
Report date:	26th April 2024		
Report status/version:	Final		
Quality approval:			
James Peeling (Project Manager)	**James Peeling**	Alex Wright (Technical Reviewer)	**Alex Wright**

Disclaimer

This report has been produced by TRL Limited (TRL) under a contract with the TRL Academy. Any views expressed in this report are not necessarily those of the TRL Academy.

The information contained herein is the property of TRL Limited and does not necessarily reflect the views or policies of the customer for whom this report was prepared. Whilst every effort has been made to ensure that the matter presented in this report is relevant, accurate and up-to-date, TRL Limited cannot accept any liability for any error or omission, or reliance on part or all of the content in another context.

Executive summary

Currently, national and local roads authorities use deterministic (rule based) methods within the tools they apply for the identification of lengths that should be considered for maintenance (schemes), and to determine their priority for maintenance. These tools utilise condition data (provided by automated and human assessors). The condition data is used to identify schemes, which are then further assessed to determine which will be treated. However, the current rules-based approaches do not produce outcomes that robustly reflect the decisions that engineers would themselves make. The availability of improved tools to support this process could assist engineers in making consistent decisions, improve efficiency and support better long-term planning.

This work investigates how an outcome-based approach could be developed to better identify lengths for treatment. The development of the model draws on network level data from the Strategic Road Network. This includes condition data such as visual condition, roughness and skid resistance, and contextualising information such as construction, traffic, material and age. These are collated and aligned with data on the actual treatments that were carried out on the lengths for which the condition and contextualising data were available. The dataset is split 80/20 to train/test a set of machine learning models. The best performing of these models deploys the Random Forest Classifier, which is referred to in this work as the Digital Engineer.

The locations identified for treatment by the Digital Engineer in the test dataset are compared with those that were actually treated. The locations are also compared with the locations identified using deterministic (rules-based) methods. The accuracy scores suggest that the Digital Engineer provides a significantly higher level of overall accuracy in the identification of lengths requiring treatment than the rules-based methods. The Digital Engineer also identifies the specific lengths that were treated to a much high level of accuracy and was more consistent in identifying the lengths that were not treated.

The results indicate that additional contextualising information is required to help digital tools deliver recommendations for treatment that better reflect the decisions made by engineers, and that Machine Learning techniques may be used to apply this additional contextualising information. However, there are complexities in the way that models apply the data to make decisions on treatments. The influences of contextual factors (e.g. location, type), and their balance with condition, may need to be better understood and explained as the development of the Digital Engineer matures. This will ensure that such an approach can be trusted and can be generally applied.

It is suggested that further development and testing of the Digital Engineer approach should aim to better understand the influence of features on tool decisions, including identifying any that could be included in the Digital Engineer that were not available as part of this study. A more comprehensive model verification could also be undertaken. There would also be benefit in considering the route to implementation, perhaps in collaboration with pavement engineers, and whether the approach could be implemented within TRL's iROADS system.

Table of Contents

Acknowledgement

The research presented in this report would not have been possible without the pavement condition and treatment data sourced from National Highways' databases. We would like to express our gratitude for the permission to use the data in this work.

1 Introduction

Currently, as part of their road maintenance management process, many national and local roads authorities use deterministic (rule based) methods for the identification of lengths that should be considered for maintenance (schemes), and to determine the priority for maintenance (treatments) on those lengths. For example, the condition of the English Strategic Road Network (SRN) is currently assessed through surveys carried out using specialist tools/devices and by human engineers. The condition data provided are used to identify schemes. These schemes are then further assessed by engineers to determine which schemes should be treated. Similar approaches are also applied on local roads in the UK.

Due to the human component, it is inevitable that a number of factors other than road condition are considered during the engineer's assessment. Experience on the SRN suggests that there is a disconnect between the lengths identified using current rules-based methods, and those lengths that actually receive treatment. The aim of the work presented in this report was to understand whether it would be possible to utilise the wider data available to highway engineers to improve the ability to identify the lengths for treatment that are actually treated on the network. The availability of improved tools to support this process could assist engineers in making consistent decisions, improve efficiency and support better long-term planning.

It was proposed that an outcome-based approach, in the form of a traditional machine learning algorithm, be investigated. Therefore, this work develops a machine learning model to identify lengths requiring treatment. The work focusses on the Strategic Road Network, for which data was made available by National Highways. Data is collated on road condition, along with wider contextualising information such as construction, traffic, material and age. By comparing these with data on actual treatments, a set of machine learning models are developed, trained and tested. The best performing model is selected, which is named the Digital Engineer, and its performance discussed.

2 Background - approach to maintenance management

On Strategic Roads National Highways use a seven-stage approach to the planning of pavement maintenance based on the 3D process (Develop, Design, Deliver), Table 1. Stages 0 and 1 (scheme identification, and options assessment) are the focus of this work.

Table 1: National Highways 3D process

Stage	Objective
Stage 0: Scheme identification	**Develop**
Stage 1: Options assessment	
Stage 2: Preliminary design	**Design**
Stage 3: Detailed design	
Stage 4: Commercial pricing	**Deliver**
Stage 5: Scheme construction	
Stage 6: Scheme close out	

National Highways' Decision Support Tool (DST) acts as the basis for scheme identification. The tool is used to identify and inform a programme of works, based on a range of data, including:

- Condition from TRAffic speed Condition Surveys (TRACS), TRAffic speed Structural Surveys (TRASS), and Sideway-force Coefficient Routine Investigation Machine (SCRIM) surveys
- Traffic from the Web TRaffic Information System (WebTRIS)
- Pavement construction information from Highways Agency Pavement Management System (HAPMS) and (more recently) Pavement-Asset Management System (P-AMS) pavement records

The rulesets used by the DST are based on those laid out in CS 228 (CS 228 Pavement inspection and assessment - Skidding resistance, 2021), and CS 230 (CS 230 Pavement maintenance assessment procedure, 2020). CS 230 mandates a series of condition surveys to assess the structural and surface properties of the road. TRACS surveys are carried out annually to assess the condition of the pavement surface by measuring parameters such as texture depth, rutting, cracking, and longitudinal profile. TRASS surveys are carried out to assess the structural integrity of the pavement by measuring pavement deflection and collecting ground penetrating radar information. CS 228 mandates an additional survey (SCRIM), which annually assesses the skid resistance properties of the pavements.

The results of these surveys are interpreted using the condition thresholds defined in CS 230, as well as the skid resistance investigatory levels provided in CS 228. The combination of these standards assists the DST, and highway engineers, to identify schemes where maintenance could be carried out.

In addition to the results of the DST, regional operators (which include private companies commissioned by National Highways to manage an Operational Area) will consider regional needs to satisfy the pavement condition metrics and performance indicators provided in the Design Manual for Roads and Bridges (DMRB) and the Asset Class Strategy (Highways England, 2020). These regional needs may incorporate network availability, regional major projects, accident data, or road user needs.

Once potential schemes have been identified, they are assessed for maintenance prioritisation. Table 2 presents National Highways guidance for initial prioritisation of potential schemes (National Highways, 2023) which aligns with National Highways' three imperatives, Safety, Customer Service and Delivery, with equal weighting given to each imperative.

Table 2: Maintenance prioritisation imperatives (National Highways, 2023)

	Priority approach	Specific considerations
Safety	The prioritisation of schemes under safety is based on skid resistance, texture depth and rutting, parameters that have been shown to correlate with collision risk.	**Skid resistance:** The priority assessment is based on skid deficiency i.e. the level of skid resistance relative to the Investigatory Level (IL). Higher priority is given to schemes where skid resistance is below the IL. **Texture depth and Rut depth:** The priority assessment for these parameters is based on the TRACS condition category as defined in CS 230 with higher priority given to the poorer condition categories.
Customer service	For customer service, priority is given to schemes that will improve ride quality, reduce noise or address customer complaints.	**Ride quality:** The assessment is based on the TRACS eLPV3, eLPV10 condition categories and the Bump Index as defined in CS 230, with higher priority given to schemes with poorer ride quality. **Noise:** The assessment is based on the noise sensitivity of the scheme (assessed in accordance with CD 236) and the existing surface type. **Customer complaints**: The assessment is based on schemes with high numbers of customer complaints that would be addressed by pavement renewal works (which could include potholes, bumps, noise etc.)
Delivery	For delivery, priority is given to schemes where there is a high risk of unplanned maintenance being required due to pavement condition. The three parameters considered are NSC (on TRASS) for flexible pavements, the number of RMMS/Confirm defects and the proportion of length that will improve the KPI3 score.	**TRASS structural condition:** The assessment is based on the TRASS NSC as defined in CS 230 with higher priority given to schemes in the poorer condition. **RMMS/confirm defects:** The assessment is based on the number of pavement-related defects, per carriageway-km, recorded over the most recent three-year period. **KPI3:** The assessment is based on the percentage of the scheme length that, when treated, will improve the KPI3.

Following the completion of the first stage (scheme identification), schemes move into the options assessment phase. During this phase, at least two options are created for each scheme: a ‘Do Minimum’ option to restore the pavement to a serviceable state; and (at least one) ‘Preferred Option’ to address a wider range of maintenance needs. National Highways select a single design from the options developed using the lifecycle costing software ‘Software for the Whole-life Economic Evaluation of Pavement Schemes’ (SWEEP.S). This software determines the option providing the best value for money over a 60-year period. SWEEP.S considers a wide variety of cost factors in its calculations, including: cost of works; costs due to accidents and delays; economic benefits; future traffic growth; current pavement condition and deterioration rates. Schemes selected based on a SWEEP.S analysis will then be taken forward to the design and deliver maintenance processes.

3 The potential for the application of Machine Learning to maintenance management

3.1 Machine Learning

Machine Learning (ML) is an umbrella term which refers to a wide range of statistical tools. These tools fall into a hierarchy of complexity as illustrated in Figure 1, where smaller circles indicate higher orders of complexity. It can be seen that Artificial Neural Networks (ANNs) are the most sophisticated methods shown. However, whilst ANNs may offer a high level of predictive power, they are also highly complex statistical models and require a considerable effort to setup and train.

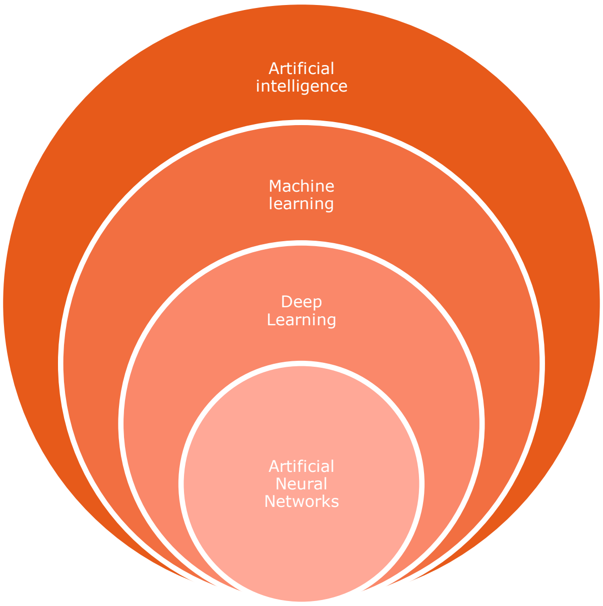


Figure 1: The hierarchy of artificial intelligence methods

In contrast to ANNs, machine learning and deep learning methods can provide advantages in terms of model training, the setting up of the models, computing power, storage, and implementation. Figure 2 presents the hierarchy of some specific machine learning and deep learning algorithms. With the exception of the neural network methods, the functional training and use of the algorithms is similar.

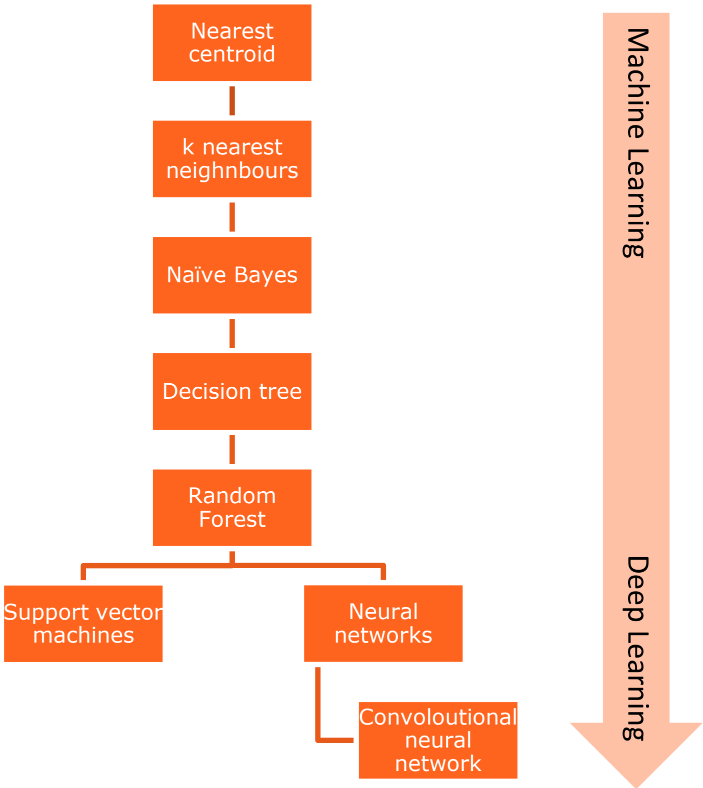


Figure 2: The hierarchy of machine learning and deep learning algorithms (adapted from (Kneusel, 2021))

This work aimed to investigate if the less complex ML tools could be used to predict locations for maintenance, and in particular whether these could provide a better level of agreement with human decisions than the current approach applying traditional rules-based methods. This would be achieved by creating a ML tool, called the 'Digital Engineer', which (if successful) could provide engineers with advice regarding locations for scheme selection. The guiding principle was therefore to supplement the deterministic (ruleset based) approaches for scheme identification laid out in CS 228 and CS 230 with an analytical outcome-based approach. This outcome-based approach would utilise the same data as CS 228 and CS 230. However, because the outcomes would utilise ML algorithm training drawn from real world schemes, they should indirectly accommodate the additional analyses applied by engineers (organisational preferences and other factors) that lead to the current schemes identified for maintenance on the SRN.

Note: It was not the intention of this work to develop an "automated engineer". Decisions would still require human intervention. Rather, the Digital Engineer aims to increase efficiency. The current rules-based approach for maintenance selection identifies many potential lengths, which must be sifted and prioritised. It was anticipated that the Digital Engineer would allow engineers to more robustly shortlist the schemes most likely to pass through to Stage 6. This would help engineers to focus on those schemes which are most likely to be implemented.

It is also anticipated that the Digital Engineer could support maintenance scheme design for local roads. As for the SRN, Local Highway Authorities (LHAs) undertake routine surveys to collect condition data. The Surface Condition Assessment of the National Network of Roads (SCANNER) survey is currently mandated for local classified roads. LHAs use the SCANNER data (in combination with other local data) to identify potential schemes. As SCANNER provides data that is similar to the TRACS surveys carried out on the SRN, the Digital Engineer could (subject to training) be used to identify potential schemes on local roads.

3.2 Previous work in this area

The following is a brief summary of other work that has been undertaken to apply machine learning methods to identify locations for maintenance. This is not an exhaustive list. It is provided as a contextualising overview.

- In 1996 work was carried out to implement an Artificial Neural Network (ANN) to develop and implement an automatic procedure for screening and recommending roadway sections for pavement preservation on the Arizona Department for Transport network in the USA. The ANN correctly predicted the treatment recommendation in 76 percent of cases presented during model testing (an accuracy score of 0.76) (Flintsch *et al.*, 1996).
- (Salini *et al.*, 2015) carried out an investigation into the use of various AI tools to predict actual pavement maintenance actions based on pavement condition, budgetary constraints, and strategic importance. They concluded that genetic algorithms could be used to predict the most appropriate maintenance actions.
- A similar ANN approach to that applied by (Flintsch *et al.*, 1996) was used by (Mosa, 2017), who considered the use of pavement condition parameters and defects to identify optimum interventions. However, in the work of Mosa the optimum interventions were generated based on rulesets rather than using real world data. Because of this it was not possible to verify the real-world performance of the ANN.
- (Domitrovic *et al.*, 2018) analysed the possibility of using ANNs to evaluate pavement condition, and its application in the generation of maintenance schemes for the Croatian National Roads Authority. A backpropagation ANN was developed and tested on 481.3 km of national roads. The investigation indicated that ANNs could be used for optimisation of maintenance or rehabilitation strategies, and for the assessment of pavement condition at the project and network level.

It should be noted that with the exception of (Salini *et al.*, 2015), all the research presented above utilised ANNs. This work was seeking to investigate if the use of simpler ML tools can generate useful results.

4 Data Collection

Two key data sources were interrogated to provide data to develop the Digital Engineer model. The National Highways pavement management database 'HAPMS' provided treatment, inventory, and condition data. The National Highways Web-based TRaffic Information System (WebTRIS) provided traffic data. These data sources were processed to produce three main 'data types' as described in the following sections.

4.1 Treatment data

The treatment data reflect interventions actually carried out on the SRN. The Digital Engineer aimed to predict these interventions (i.e. to treat / not treat a length). A full description of the treatment dataset is provided in Appendix A.1. In summary, ~3 million rows of data were extracted from HAPMS, which separated the SRN into discrete sub-sections containing homogeneous constructions. The key information for these sub-sections was whether treatment had been carried out on the sub-section, and the date of the treatments. Whilst several categories of "treatment" were identified, some (such as 'Newly constructed') did not actually refer to a treatment. To simplify the analysis only the following categories were considered as treatments: Inlay, Re-surfacing, Re-construction, Surface Treatment, and Overlay.

4.2 Condition data

Condition data describe the condition of the SRN, as determined by annual surveys. In the light of the process discussed in Section 2, pavement condition data from the following annual surveys carried out by National Highways was used in this work:

- **TRAffic speed Condition Surveys (TRACS),** which provide information regarding the geometric properties of the road surface. Vehicles capture surface information though image, laser, accelerometer, inertial, and GPS measurements. For this work the following measured parameters were used from TRACS surveys; Rutting, Texture depth, Ride quality, Fretting, and Cracking. These parameters are described in the next chapter.
- **TRAffic speed Structural Surveys (TRASS),** which provide information regarding the structural condition of the pavement. TRASS survey vehicles measure the deflection of a pavement in response to loading from a HGV. This work has used the TRASS structural condition category data (NSC) which measures condition on a 1 (good) – 4 (poor) scale.
- **Sideway-force Coefficient Routine Investigation Machine (SCRIM) surveys**, which provide information on the skid resistance performance of the road surface. SCRIM devices use a test wheel to measure skid resistance under wet conditions. For this work the Characteristic Skid Coefficient (CSC) was used.

A full description of the condition dataset is provided in Appendix A.2. In summary, condition data were gathered for surveys carried out during the period 2005-2021 (inclusive), with the following parameters:

- Rutting (reported as values of 'maximum rut'), which represents the deepest rut measured in either wheelpath.
- Texture depth (reported as values of SMTD), which represent the texture depth of the road surface in the nearside wheel path.
- Ride quality; reported either as values of Moving Average Longitudinal Profile Variance MALPV, or enhanced Longitudinal Profile Variance (eLPV) at 3m, 10m, and 30m wavelengths. It is noted that the MALPV measure was changed to eLPV in 2014. As the data analysis extended before this date there was a need to accommodate this change. To maximise the amount of 3m and 10m MALPV / eLPV data available (and to enable them to be considered as comparable parameters), the ride quality parameter values were converted into Roughness Index (calculated using Equation 1). This provided Roughness Index values based on MALPV and eLPV. These Roughness Index values were later re-scaled so that they could be directly compared, as discussed in Section 5.3.

$$RI = \max\left(\frac{\sqrt{10 \cdot 3m\ eLPV}}{3} + \sqrt{10m\ eLPV} - 0.1, 0\right)$$

Equation 1: Calculating the roughness index

- Fretting, which characterises the amount of chip loss (pavement aggregate) over the lane width.
- Cracking (reported as whole carriageway cracking), which characterises the amount of cracking over the width of the whole carriageway.
- Deflection (reported as TSD / TRASS structural condition category), which characterises the amount of deflection of the pavement experienced in response to the loads exerted on it in the range 1 to 4 inclusive.
- Skid resistance (reported as Characteristic Skid Coefficient (CSC)), which describes the low speed, high slip (locked-wheel) skid resistance of the pavement in the nearside wheel path.
- Skid difference, which reports the difference between the measured skid resistance and the in-service requirement for skid resistance.

4.3 Contextualising data

Contextualising data may influence the selection of sites as schemes for intervention but are not data which describe the condition. The contextualising data used in this work (see also Appendix A.2) were:

- Traffic. The number of HGV passes per year

- Material age. The age of the surface layer, sub-surface (i.e. the layer immediately below the surface layer), and the base layer
- Thickness. The thickness of the surface layer in mm
- Material type. The type of the material used in the construction of the surface layer, sub-surface layer, and base layer
- The operational Area. The SRN is separated into 14 operational Areas. Some of these are managed by sub-contracting entities
- The operational environment. Reported as ‘urban’ or ‘rural’
- The carriageway type. Reported as ‘single carriageway’ or ‘dual carriageway’

With the exception of the traffic data, contextualising data were gathered at the same time as the treatment / condition data. Traffic data were collected from WebTRIS National Highways’ Web based Traffic Information System.

5 Data Processing

5.1 Data wrangling

Data wrangling is the process of converting raw data into a usable form, sometimes referred to as pre-processing. The data wrangling stage focussed on combining the datasets summarised in Section 4 so that they could be considered together (locationally aligned, comparable reporting intervals etc.) in the data exploration phase. A notable challenge for this work was the different methods of data delimitation / geo-referencing used. For example, condition data are nominally presented in 100m sub-sections, whereas the treatment data were not length delimited. This, along with other differences in data formats required complex joins and filters to be applied to the data. The data wrangling hence provided a single dataset of "features" for the development of a machine learning model, to be passed to the initial data exploration stage. A summary flowchart of the data wrangling process is presented in Figure 3. Further detail on the data wrangling is presented in Appendix B.

5.2 Initial exploration

Initial exploration was carried out to identify any further requirements for filtering or wrangling of the data before applying the machine learning. The following data analyses were carried out:

- The prevalence of each initial model feature was plotted as a bar charts
- The relationships between the initial model features were assessed for interdependence
- The distribution of the initial model features, and dependant variables were plotted as histograms or bar charts

The key findings are provided below. Further detail is provided in Appendix E.

Figure 4 shows the prevalence of each feature within the dataset, and Figure 5 presents the linear coefficients of determination between each of the features. The key observations from these figures are:

- The MALPV/eLPV ride quality data were incomplete. As there was a need to have all features present for the development of the machine learning model, the inclusion of the LPV data would have resulted in an unacceptable amount of data loss (i.e. lengths removed from the analysis because they had incomplete data sets). This was primarily a result of the move from LPV to eLPV in 2014, discussed above. As removal of ride quality as a maintenance factor was considered desirable, a need was identified for a process to address this issue.
- The data on cracking, deflection, trafficking, and base type were present in only approximately 30% of rows. This had the potential to lead to a large amount of data loss. Therefore methodologies for filling these data would need to be implemented.

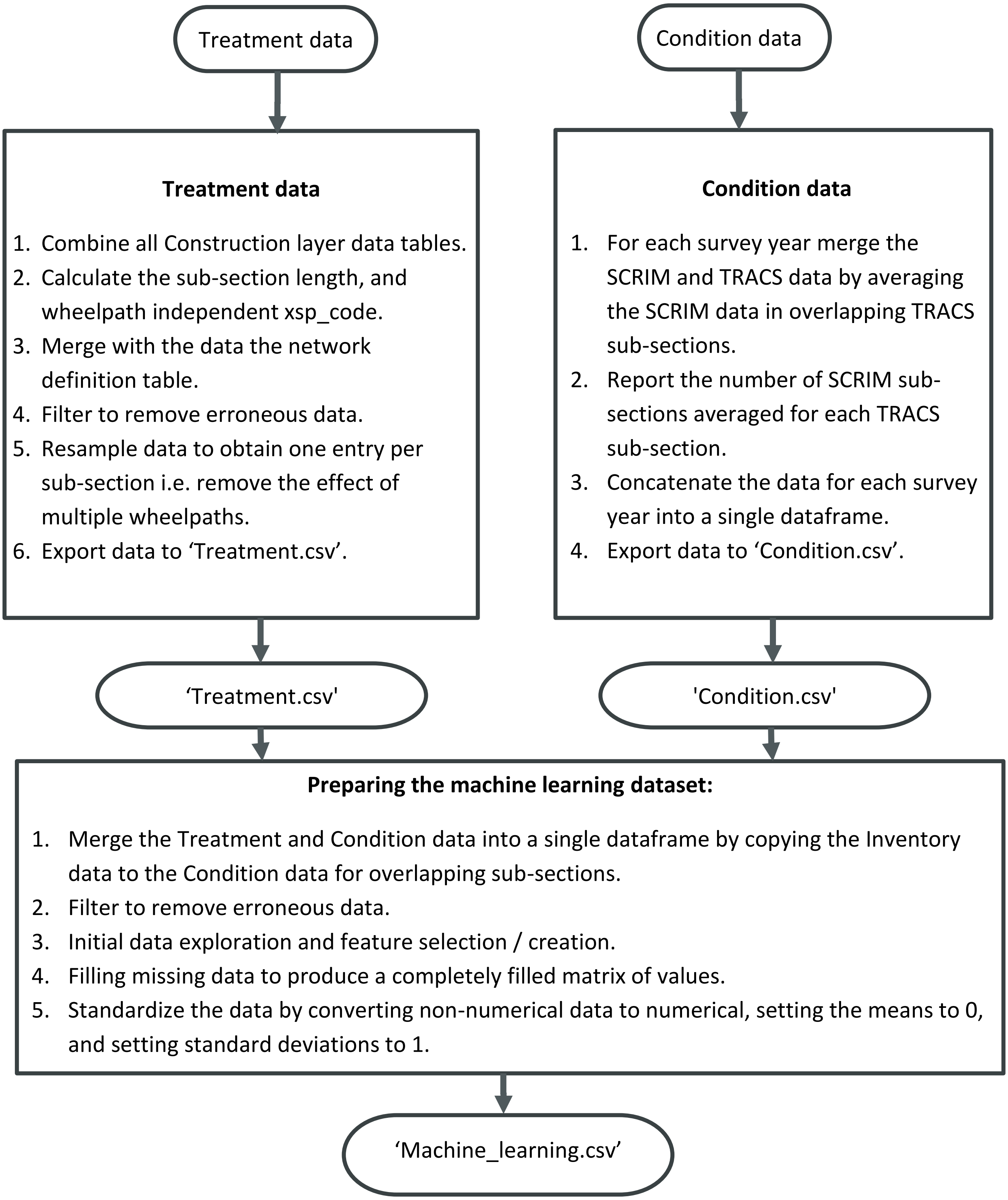


Figure 3: Summary data processing workflow

- Strong correlation was observed between skid resistance and skid difference. It was decided that scrim difference should be removed from the model, to reduce co-dependency in the data.
- The majority of data in the dataset was associated with pavements that had not been treated (which may be expected given only a few percent of the network is maintained each year) - Figure 6. If this dataset were provided to a machine learning algorithm it would be likely that the algorithm would develop a strong preference (bias) for not predicting a need for treatment. For training it would be necessary to balance the dataset such that the dataset was equally representative of treated and non-treated sections.

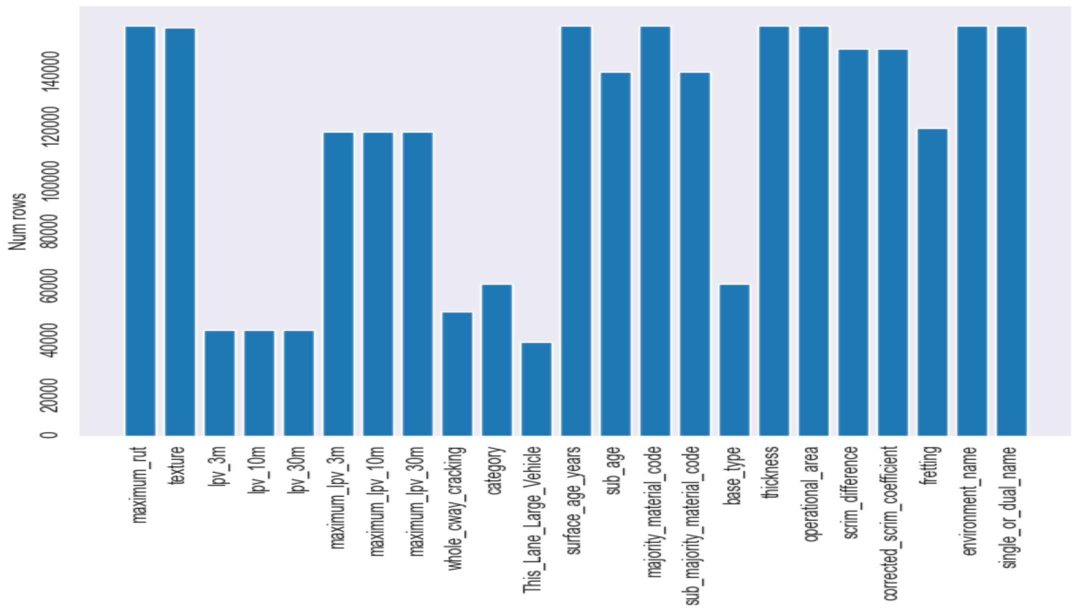


Figure 4: Distribution of initial model features

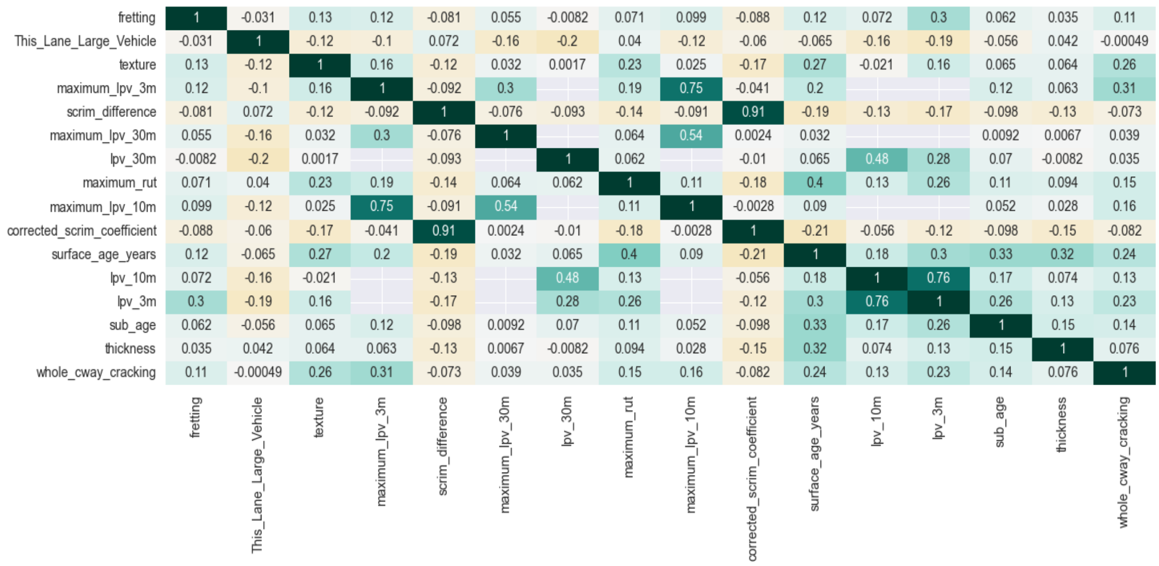


Figure 5: Correlations between initial model features

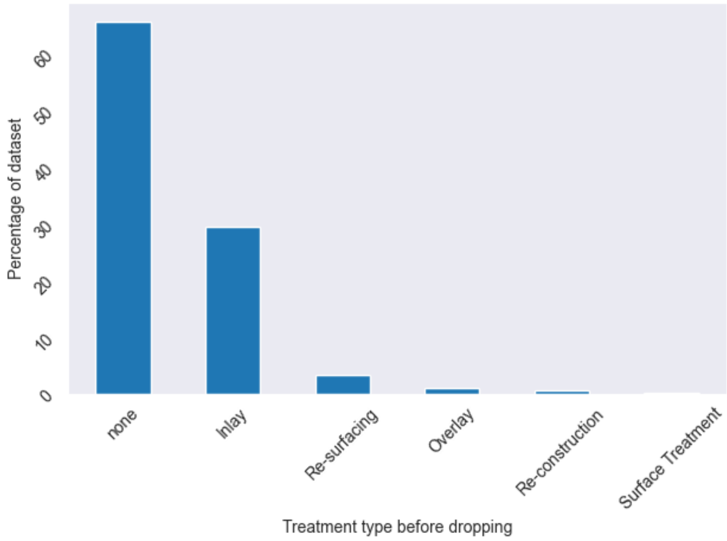


Figure 6: The prevalence of treatment types before balancing

5.3 Data filling

To accommodate the missing LPV data, the 3m and 10m MALPV / eLPV parameters were consolidated into a single Roughness Index (RI) using Equation 1. These were then standardised during the data standardisation process so that RI value obtained using eLPV and MALPV could be directly compared (see below). To accommodate the correlation between skid measurements, skid difference was removed from the list of model features.

A staged data filling process was used to maximise the availability of data. This process was applied to the cracking, deflection, trafficking, and base type data:

- For sub-sections/survey years not containing cracking data, the mean cracking value for the same sub-section from an adjacent survey year was accepted.
- For sub-sections/survey years not containing deflection category or base type data, the modal values for the same sub-section from an adjacent survey year was accepted.
- For sub-sections/survey years not containing trafficking data, the mean trafficking value for the same sub-section, from an adjacent survey year was accepted. As there were still significant gaps, for any remaining sub-sections/survey years not containing trafficking data the mean trafficking value for the whole network was accepted.

After the above had been carried out the data were filtered to remove any rows where there was an empty value in any row.

5.4 Data standardisation

After the above actions were completed, the data were standardised, this consisted of:

- Converting any non-numerical categorical data to numerical data. This was achieved using scikit-learn's LabelEncoder. (Scikit-learn, 2022a)
- Setting the means of each feature to zero. This was achieved by subtracting the feature mean from the feature value.
- Setting the standard deviations of each feature to 1. This was achieved by dividing the feature value by the feature standard deviation.

The processes resulted in approximately 14,725km (reported in nominally 100m delimited lengths) containing the following features:

- Treatment. A description of if the pavement sub-section was treated or not. This is the target that the Digital Engineer aimed to predict.
- Condition. Rutting, Texture depth, Roughness index, Cracking, Deflection category
- Contextualizing. Trafficking, Surface age, Sub-surface age, Base age, Surface material type, Sub-surface material type, Base material type, Surface thickness, Operational area, Operational environment, Carriageway type

5.5 Data balancing

The final stage of data wrangling was to balance the data such that the dataset equally represented lengths that had been treated and not treated. This was achieved by removing rows relating to treated sections until the data were balanced; the removal of rows was weighted by 'survey_year' so that data relating to more recent surveys would be biased over older surveys.

Following the completion of the filling, standardisation and balancing processes, the dataset was such that there was an equal number of rows for each model feature (i.e. the bars in Figure 4 became all of equal height), and the linear coefficients of determination between each of the features were as shown in Figure 7. It can be seen that, in the final dataset, instances of high between-feature correlations were significantly reduced.

Figure 7: Correlations between final model features

6 Machine learning model development and selection

6.1 Model selection

A number of machine learning and deep learning algorithms were considered for the Digital Engineer, as summarised below. Further descriptions of these are provided in Appendix C. All of these models seek to assign a category (treat or no-treat) to an input[1]:

- **Nearest Centroid (NC)**; Assigns a category to an input by determining the shortest distance to the category centroids (the average location of categories) within an n dimensional (nD) feature space[2].
- **k Nearest Neighbours (kNN)**; Assigns a category to an input by determining the nearest k category neighbours within an nD feature space and using these neighbours to vote on the category of the input.
- **Random Forest Classifier (RFC)**; Uses a 'forest' of decision trees to assign a category to an input variable. The decision trees are built to most 'cleanly' separate the modelled categories based on the model features.
- **Support Vector Machines (SVM)**; Separates an nD feature space into sectors which most 'cleanly' separate the modelled categories based on the model features. A category is assigned to an input based on the sector within which it falls.
- **Voting Classifier (VC)**; Utilises multiple algorithms to collect a group of votes regarding the category to be applied to an input.

Table 3 summarises the strengths and drawbacks of these models within the context of this work.

6.2 Implementation

The machine learning models used in this work were the implementations from the open source scikit-learn Python library. This provided a framework within which all the models listed in Table 3 could be applied, along with a common Application Programming Interface (API) for training and prediction. It also provided functions for assessing model performance. Scripts to prepare the data, prepare the models, train the models, perform predictions with the trained model, and assess the performance of the models were created in Python, also using other common data analysis libraries for data handling (Pandas) and visualisations (Matplotlib and Seaborn). This enabled each model to be trained and tested using the same input data to enable a fair comparison.

[1] An input is a set of data which the model has not 'seen' before for which a category is to be assigned.

[2] Where n = the number of model features

Table 3: Summary of the strengths and drawbacks of various machine learning models considered for the Digital Engineer

Model	Strengths	Drawbacks
NC	Models are very small in size. Models are computationally cheap to store and query. Does not require training.	Suffers from the curse of dimensionality. Does not cope well with diffuse groups of categories. Model is a 'black box'.
kNN	Does not require training.	Can be computationally expensive to carry out predictions. Model size scales with data size. Suffers from the curse of dimensionality. Does not cope well with diffuse groups of categories. Model is a 'black box'.
RFC	Can perform well with diffuse category groups. Allows trees to be interrogated to allow decision process to be understood.	Model training can be computationally expensive. Suffers from the curse of dimensionality. Resultant models can be very large. Can suffer from over fitting.
SVM	Can perform well with diffuse category groups.	Suffers from the curse of dimensionality. Can be computationally expensive to train. Model is a 'black box'.
VC	Allows a more precise determination of category.	Can be very computationally expensive. Can create very large models.

6.3 Training data

The training was conducted using 80% of the available dataset, with the remaining 20% being held for testing. When splitting data to train the machine learning models, care had to be taken to split the data such that pavement sections appearing in the training dataset did not appear in the testing dataset. This ensures that the test dataset only contains data that the model has not 'seen' during training. This was accomplished using scikit-learn's 'GroupShuffleSplit' tool (Scikit-learn, 2022b), which is a utility provided by scikit-learn to split data, whilst ensuring that no group appears in both the training and testing sets of the same split.

6.4 Model training and use

Machine Learning Models were initially trained using their default hyperparameters (model training settings). Each of the model types listed in Section 6.1 were investigated and trained. This was done using scikit-learn's toolbox of functions. First the model was trained using the 'fit' function. This feeds the training data to scikit-learn's implementation of each model and returns the trained model. The trained model is used for predicting treatments on the test dataset and saved for later use. To obtain predictions, the test data is fed to the trained model returned by the training function.

Later in the process, the hyperparameters of some of the models were varied to see if improved performance could be achieved. Hyperparameters are properties of the training process, whereas parameters are values in the model itself. The tuning of the hyperparameters was done using further tools provided in the scikit-learn library. This required multiple runs of training for the models being tuned.

7 Performance of the Digital Engineer

7.1 Comparison to the treatments in the test dataset

The performance of the trained models was investigated by applying the test data (the remaining 20% of the available dataset, with no overlap with the training dataset) within the trained model to obtain model predictions of the treatments required for each pavement sub-section. These predictions were then compared with actual treatments carried out on these sub-sections.

The performance was assessed using ‘accuracy scores’, which quantify the proportion of sub-sections cases in which the model made an accurate prediction of the treatment carried out (e.g. an accuracy score of 0.8 means that the model correctly predicted the treatments, including no treatment, for 80% of sub-sections). Four categories of accuracy scores were calculated:

- True positives for treated sub-sections (**True_pos_treat**): Where the model predicted that a sub-section received a treatment, and this was actually treated
- True positives for non-treated sub-sections (**True_pos_no_treat**): Where the model correctly predicted that a sub-section did not receive a treatment, and it was not actually treated
- False positives for treated sub-sections (**False_pos_treat**): where the model predicted that a sub-section received a treatment, and it was not actually treated
- False positives for non-treated sub-sections (**False_pos_no_treat**): where the model predicted that a sub-section did not receive a treatment, but it was actually treated

This was necessary because of the un-balanced nature of the network in terms of those lengths which are treated or not. The data suggested that approximately 6.3% of the National Highways network is treated annually (and therefore approximately 93.7% of the network is not treated). This adds importance to the *False_pos_treat* metric, because even a low tendency to report false positive predictions of treatment score (as a percentage) could be result in a large proportion of the network being falsely predicted to require maintenance in relation to the proportion of the network actually maintained.

Whilst all of the models listed in Section 6.1 were investigated, initial tests identified that a some approaches had very poor levels of performance. As a result, only the results for RFC, kNN and SVM models are presented here.

The solid- coloured bars in Figure 8 present the initial accuracy scores for the RFC, kNN and SVM models. It can be seen that, in the initial assessment, RFC provided the best overall accuracy (0.87), and the lowest level of false positives (*False_pos_treat* score, 0.09). Of these three models, SVM provided the lowest overall accuracy score (0.78). This was largely driven by the lower performance in correctly predicting the need for treatment (*True_pos_treat* metric). A fine-tuning phase was then undertaken in which the model hyperparameters were adjusted in an attempt to improve the accuracy scores for each model. This required re-training with new hyperparameters and repeat of the testing using the 20% test datasets. Model tuning was carried out using scikit-learn’s ‘GridSearchCV’ tool

(Scikit-learn, 2022c), which loops through predefined sets of hyperparameters to better match the model to the training set, enabling the optimum parameters to be selected from the listed hyperparameters.

The results of the tuning exercise are presented by the striped bars in Figure 8. It can be seen that the tuning process improved the accuracy for the kNN and SVM models, but the performance of these models was still below that of the RFC. The tuning process had little effect on the RFC. However, given that the RFC model provided the best overall performance, and provided the best *False_pos_treat* metric score, it is proposed that this model would be the most appropriate for application within a Digital Engineer tool.

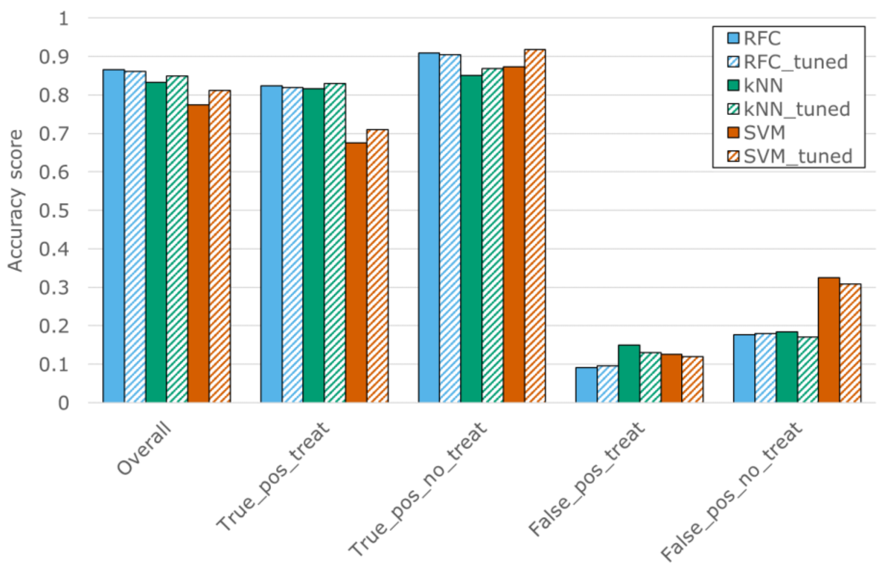


Figure 8: Initial (solid) and tuned (striped) machine learning accuracy scores for three models

7.2 Comparison to current methods

A further assessment of the Digital Engineer (RFC model) was undertaken by comparing its performance with that achieved using the current National Highways rules-based methods (Section 2). The objective of this assessment was to determine the performance of the current rules- based methodology in predicting the lengths that are maintained, and to use this to place the performance of the Digital Engineer into context. To undertake this assessment it was necessary to implement a version of the current rules-based process for maintenance assessment.

The advice in CS 228 (CS 228 Pavement inspection and assessment - Skidding resistance, 2021) can be applied to identify lengths which should be considered for maintenance based on skid resistance data (which hence are likely to be identified primarily as safety treatments). In addition, the current advice in CS 230 (CS 230 Pavement maintenance assessment procedure, 2020) can be used to identify lengths which should be considered for maintenance to meet functional/engineering needs. These are identified within two categories - technically simple or technically complex - referred to here as technical schemes for simplicity, and as a preventative maintenance category.

The technical schemes are identified based on the following decision points:

1. Are deflection categories available and all in Category 1 or 2?
   If no then technical scheme, if yes move to decision point 2.
2. Are rut depths less than 11mm?
   If no then technical scheme, if yes move to decision point 3.
3. Are there widespread visible surface defects[3] in any running lane in the Section?
   If yes then technical scheme, if no then no scheme.

The preventative maintenance schemes are based on the following criteria:

- The surface must be of a Thin Surface Course System (TSCS) construction,
- The surface must be greater than five years old,
- texture depths, rut depths and eLPVs must all be within condition category 1,
- lane fretting intensities must be <2,
- deflection categories must be within condition categories 1 or 2, and
- CSC must be above the investigatory level.

The CS 228 and CS 230 rulesets were applied, using the condition and contextual data available to this work (Section 4), to predict treatments for each sub-section. As was undertaken above to assessments the performance of the Digital Engineer, overall accuracy and four sub-accuracy scores were determined. These results are presented in Figure 9, where they are compared with the accuracy scores obtained by the Digital Engineer (RFC).

[3] Note that the condition dataset used in this work did not contain specific values for “widespread visible surface defects” but did contain TRACS cracking and fretting data. Therefore, it was assumed that widespread surface defects were present if TRACS Cracking >2 and TRACS Fretting >20.

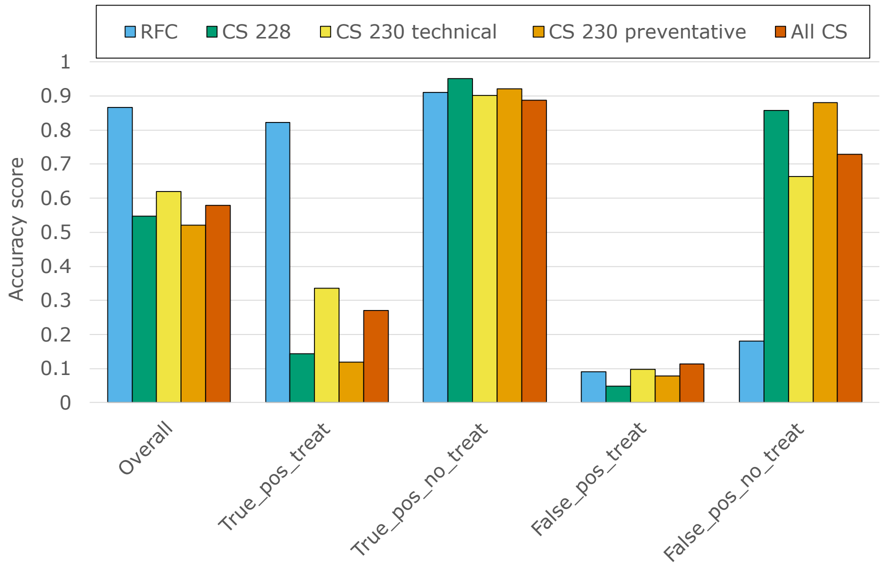


Figure 9: Digital Engineer (RFC) and current methodologies accuracy scores

Figure 9 Shows that the overall accuracy of the Digital Engineer exceeds that of the other approaches. The True_pos_treat performance has a strong influence here, with the rules-based methods not achieving an accuracy score better than 0.35. The performance of the methods is broadly comparable for the the False_pos_treat metric. However, the performance of the The False_pos_no_treat is surprising. The Digital Engineer significantly outperforms the rules-based methods, identifying a high number of treated lengths not identified using the rules-based approach. The rules-based methods are typically considered the starting point for the identification of lengths for treatment. In theory, engineers apply the rules and then filter down to a subset for maintenance. It might therefore be expected for only a few lengths to be treated that were not identified using the rules-based approach. This does not appear to be the case.

7.3 Discussion

The accuracy scores suggest there is a degree of uncertainty in the deterministic approach, as deterministic methods identify many schemes that will not be treated or would be treated with a low priority. In practice, the decisions recommended by these methods are augmented by several further stages of investigation and review. These further stages incorporate additional data sources, engineering assessments, financial assessments, scheduling, etc., to refine the selected sites. At the "model" stage the deterministic methods (as applied in this work) do not have access to these additional data sources that would be considered by engineers in further review stages. The Digital Engineer takes an outcome-based approach, predicting where treatments will be carried out based on observations of where treatments have been carried out on other sites. This approach draws on a wider range of information. The higher level of performance achieved using the Digital Engineer suggests that these additional data sources, combined with the Machine Learning approach, have the potential to significantly improve the ability to identify treatment lengths using digital tools.

However, as we expand the decision process to take account of a wider range of factors, we move away from a distinctly technical assessment, to one in which context is included, and this may become increasingly important. When developing a model such as the Digital Engineer care needs to be taken to ensure that the model is not adversely influenced by the content and distribution of the data contained in the training dataset, as they may not be balanced. An assessment of the features that have highest influence on the model (feature importance, Appendix F) suggests that carriageway type, location, construction and age might have had a strong influence on the model, with skid resistance having the largest influence for the condition data component of the features. However, whilst the training dataset was balanced with respect to the treat/no treat lengths, balancing did not consider other factors (for example the proportion of certain types of roads) which may have affected the model development/training process. Nevertheless, the model may also be reflecting the influence of current advice on engineering practice. For example, current guidance on the performance requirements for skid resistance (CS 228) specifies investigatory levels that increase with decreasing carriageway type, which may result in priority being given to some carriageway types for the treatment of skid resistance.

Furthermore, whilst the model development has considered many features, there may be features, which were not included in the model, that influence engineers' decisions on treatment. In addition, there is the risk of lurking features/variables, and lurking co-dependent variables[4]. For example, 'operational area' directly describes the geographical location of a section on the network. However, operational area may also be co-dependent with other features (not included in the Digital Engineer), such as sub-network length, budgetary information, local maintenance considerations, material types, etc. Another example may be trafficking. Whilst trafficking describes the number of HGVs travelling over a pavement section, it may also be co-dependent with operational priority[5], pavement design, traffic speeds and road user safety. Therefore, whilst the model developed in this work appears to have a high level of performance, further investigation and refinement may be necessary to ensure that development and training data do not have an adverse effect on the outcomes of the model - so that it can be robustly applied over the network.

[4] Variables (features) that have an influence, but which we are not including, and/or features not included in the model that are co-dependent with some of the model features.

[5] i.e. more heavily used roads may be prioritised for maintenance as they affect more users than roads with lower trafficking rates.

8 Conclusions and recommendations

Currently engineers utilise condition data and other contextual information to select lengths of the network that require treatment. The availability of digital tools to support this process could assist engineers in making consistent decisions, improve efficiency and support better long-term planning. However, current rules-based tools do not produce outcomes that robustly reflect the decisions that engineers would themselves make.

This work has developed an outcome-based approach to identify locations for treatment (using a machine learning model deploying the Random Forest Classifier model, which has been named the Digital Engineer) and compared the locations identified with those that were actually treated. A similar comparison was also undertaken using rule-based methods. The accuracy scores have shown that the Digital Engineer provides an overall accuracy score of approximately 0.88, whereas the rules-based methods (used by National Highways) provide overall accuracy scores of approximately 0.55. Hence the Digital Engineer identified the lengths that were treated to a much higher level of accuracy, and was also more consistent in identifying the lengths that were not treated.

The results suggest that measurements of pavement condition alone, as used by rules-based methods, are not sufficient for determining treatment needs as identified by engineers. Additional contextualising information is required to make an informed decision. It is apparent that Machine Learning techniques could be used to apply this additional contextualising information, in combination with the condition data, to predict locations where maintenance would be carried out.

However, there are complexities in the way that models apply the data to make decisions on treatments. The influences of contextual factors (e.g. location, type), and their balance with condition, may need to be better understood and explained as the development of the Digital Engineer matures. This will ensure that such an approach can be trusted can be generally applied.

Therefore, whilst this initial work shows the high potential of this approach, further work is recommended to better understand the influence and scope of the variables (factors) on the outcomes. This would include minimising the risk of over-fitting the model to some factors, and determining whether there are further factors that might affect real-world treatment decisions, but which have not been included in the current list of the Digital Engineer’s model features.

This work has restricted its scope to a binary ‘treat’ / ‘no treat’ classification. It may be possible to widen the scope of the Digital Engineer to suggest treatment types. It would also be of value to the development of the model, and its eventual implementation, to understand how the Digital Engineer would be provided to end users, for example within Asset Management Tools (AMTs) such as TRL’s IROADS system, and how the approach would be used by pavement engineers in the real world.

References

Domitrovic J, Dragovan H, Rukavina T and Dimter S (2018). *Application of an artificial neural network in pavement management system*. Technical Gazette 25.

Flintsch GW, Zaniewski P and Delton J (1996). *Artificial Neural Network for Selecting Pavement Rehabilitation Projects*. Transportation Research Record.

Highways England (2020). *Asset Class Strategy: Pavements*. Highways England: Birmingham.

Highways England, Transport Scotland, Welsh Government, Department for Infrastructure (2020). *CS 229 Data for pavement assessment*.

Highways England, Transport Scotland, Welsh Government, Department for Infrastructure (2020). *CS 230 Pavement maintenance assessment procedure*.

Highways England, Transport Scotland, Welsh Government, Department for Infrastructure (2021). *CS 228 Pavement inspection and assessment - Skidding resistance*. Highways England: London.

Kneusel RT (2021). *Practical deep learning: a Python based introduction*. No Starch Press: San Fransisco.

Mosa AM (2017). *Neural network for flexible pavement maintenance and rehabilitiation*. Applied Research Journal.

National Highways (2023). *Asset class handbook - pavements*. National Highways: Birmingham.

Salini R, Xu B and Lenngren CA (2015). *Application of artivicial intelligence for optimization in pavement management*. International Journal of Engineering and Technology Innovation.

Scikit-learn 2022a, *sklearn.preprocessing.LabelEncoder.*, <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html>.

Scikit-learn 2022b, *sklearn.model_selection.GroupShuffleSplit.*, <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GroupShuffleSplit.html>.

Scikit-learn 2022c, *sklearn.model_selection.GridSearchCV.*, <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html>.

Scikit-learn 2022d, *Permutation feature importance.*, <https://scikit-learn.org/stable/modules/permutation_importance>.

Scikit-learn 2022e, *Supervised learning.*, <https://scikit-learn.org/stable/supervised_learning.html>.

Appendix A Data descriptions

A.1 The treatment dataset

The treatment dataset was provided in several files which segmented the SRN into lengths containing homogeneous constructions and attributes, with a 'pavement_class_name' which described the treatment applied to that length at the time of last treatment. A separate column, 'constr_start_date' provided the date at which the last treatment was carried out on the pavement sub-section (Table 4).

Table 4: treatment (construction) data table

Column heading	Data type	Unit / categories	Description
road	integer	none	A numerical description of 'road_number'
road_number	text	none	The HAPMS road number, i.e. M4.
Section	integer	none	A numerical description of 'section_label
section_label	text	none	The HAPMS section label, i.e. 0100M4/5.
Section_sort	float	none	Unknown
section_length	float	meters	The length of the HAPMS section
section_function	integer	none	A numerical description of 'section_function_name
section_function_name	text	Main Carriageway, Ox Bow Lay-by, Roundabout, Slip road	The HAPMS section function.
Operational_area	integer	1, 2, 3, 4, 6, 7, 8, 9, 10, 13, 14, 125	The NH operational area
operational_area_name	text	not provided here for brevity	A text description of 'operational_area'
data_key	integer	7,8,1128	unknown
direction_key	integer	1,2,3,4,5,6	A numerical description of 'direction_name'
direction_code	text	CW, AC, NB, EB, WB, SB	A coded description of 'direction_name'
direction_name	text	Clockwise, Anticlockwise, North, East, West, South	The direction of the carriageway
constr_start_date	date	dd/mm/yyyy hh:ss	The date at which the pavement layer was last treated.

Column heading	Data type	Unit / categories	Description
Constr_end_date	date	dd/mm/yyyy hh:ss	The date at which the pavement layer was originally constructed
start_chainage	float	meters	The start chainage of the sub-section within the pavement section
end_chainage	float	meters	The end chainage of the sub-section within the pavement section
xsp	integer	none	A numerical description of 'xsp_name'
xsp_code	text	none	A coded description of 'xsp_name'
xsp_name	text	none	A description of the cross sectional position of the sub-section of the pavement over the carriageway width. This is split into left and right wheelpaths e.g. CL1L (left wheelpath) and CL1R (right wheelpath)
traf_acc_date	date	dd/mm/yyyy hh:ss	Unknown
layer	Integer	Not provided here for brevity	A coded description of 'layer_name'
layer_sequence	integer	1 – 15 inclusive	A numerical description of 'layer_name'
layer_name	text	Layer 1- 15 inclusive	A description of the pavement layer indexed from the bottom layer.
Material	integer	Not provided here for brevity	A numerical description of 'material _name'
material_code	text	Not provided here for brevity	A coded description of 'material_name'
material_name	text	Not provided here for brevity	The material comprising the pavement layer
date_laid	date	dd/mm/yyyy hh:ss	The date at which the pavement layer was originally constructed
thickness	integer	millimetres	The thickness of the pavement layer
condition_factor	integer	Not provided here for brevity	A numerical description of 'condition_factor_name'
condition_factor_name	text	Not provided here for brevity	A description of the overall pavement condition.
Xsp_left_offset	float	unknown	unknown
xsp_right_offset	float	unknown	unknown
road_sort	integer	unknown	unknown
section_function_sort	integer	unknown	unknown
operational_area_sort	integer	unknown	unknown
direction_sort	integer	unknown	unknown

Column heading	Data type	Unit / categories	Description
pavement_class	integer	Not provided here for brevity	A numerical description of 'pavement_class_name
pavement_class_name	text	Inlay, Provisional, Re-surfacing, GPR (Uncalibrated), Coring, GPR (Calibrated), Generated from Inventory, Re-construction, Newly Constructed, Extracted from PANDEF, Not Specified, Surface Treatment, Overlay, Re-setting Condition Factors.	A description of the last pavement treatment carried out.

A.2 The condition and contextualising datasets

The TRACS and TRASS datasets were provided as combined data files for each survey year containing data reported over 100m sub-section lengths. The construction data (surface, sub-surface, and base layers) were resampled and allocated to each 100m sub-section. The percentage of each 100m sub-section comprising the most prevalent material was also reported. Traffic data was reported in relation to specific traffic counting locations (geo-located using latitude and longitude co-ordinates). These locations were overlaid onto the National Highways network to determine the closest section. A description of the resulting data file content (TRACS, TRASS, construction) is presented in Table 5. Sideway-force Coefficient Routine Investigation Machine (SCRIM) surveys are carried out annually over the SRN and used to evaluate skid resistance of the pavement surface. SCRIM survey data were downloaded from the HAPMS system covering surveys from 2005 to 2021 in separate file. A description of the data in these files is presented in Table 6. Further contextualising data was provided through the network definition and additional information (Table 7).

Table 5: TRACS data table

Column heading	Data type	Unit / categories	Description
section	integer	none	A numerical description of 'section_label'
section_label	text	none	The HAPMS section
start_chainage	float	meters	The start chainage of the sub-section within the pavement section
end_chainage	float	meters	The end chainage of the sub-section within the pavement section
date laid	date	dd/mm/yyyy hh:ss	The date at which the material was laid.

Column heading	Data type	Unit / categories	Description
survey_start_date	date	dd/mm/yyyy hh:ss	The date the survey was carried out.
survey_year	year	yyyy	The year the survey was carried out
xsp_code	text	none	A coded description of 'xsp_name'
operational_area	integer	1-10, 13, 14, 25	The NH operational area
thickness	float	millimetres	The thickness of the surface layer
maximum_rut	float	millimetres	The maximum rut measured in either wheelpath
texture	float	millimetres	The Sensor Measured Texture Depth (SMTD)
(e)lpv_3m **(e)lpv_10m** **(e)lpv_30m**	float	none	Either the longitudinal profile variance (LPV) (for materials assessed before 2014), or the enhanced Longitudinal Profile Variance (eLPV) (for materials assessed during or after 2014).
fretting	float	none	The amount of fretting on the material
whole_cway_cracking	float	none	The amount of cracking on the material
category	integer	1, 2, 3, 4	The NSC category of the pavement.
majority_material_code	text	Not provided here for brevity	The material code of the most prevalent surface material in the 100m section.
majority_material_length	integer	meters	The length of the 100m section which is the 'majority_material_code'.
surface_age_years	integer	years	The age of the majority surface material
sub_majority_material_code	text	Not provided here for brevity	The material code of the most prevalent material immediately beneath the surface in the 100m section.
sub_majority_material_length	integer	meters	The length of the 100m section which is the 'sub_majority_material_length'.
sub_age			The age of the majority material immediately below the surface.
base_type	text	Not provided here for brevity	The subbase material code.
This_Lane_Large_Vehicle	integer	Vehicles per day	The daily flow rate of HGVs.

Table 6: SCRIM data table

Column heading	Data type	Unit / categories	Description
survey_start_date	date	dd/mm/yyyy hh:ss	The date at which the survey was carried out.
survey_year	year	yyyy	The year during which the survey was carried out
section_label	text	none	The HAPMS section
start_chainage	float	meters	The start chainage of the sub-section within the pavement section
end_chainage	float	meters	The end chainage of the sub-section within the pavement section
xsp_code	text	none	A coded description of 'xsp_name'
lecf	float	none	The Local Equilibrium Correction Factor, is used to correct skid resistance data for seasonal, and between year variation and is used in the calculation of 'characteristic_skid_coefficient'.
characteristic_skid_coefficient	float	none	The pavement skid resistance corrected for seasonal, between year variation, and test speed.
site_definition_code	text	Not presented here for brevity	The skid resistance investigatory level as described in CS 228 (Highways England, Transport Scotland, Welsh Government, Department for Infrastructure, 2021)
investigatory_level	integer	Not presented here for brevity	A coded representation of 'site_definition_code'.
scrim_difference	float	none	The difference between the characteristic_skid_coefficient and in-service requirement.

Table 7: Network information/definition summary table

Column heading	Data type	Unit / categories	Description
road_class	Integer	1, 2, 3	A numerical description of 'road_class_name'
road_class_code	text	A, AM, M	A coded description of 'road_class_name
road_class_name	text	A, AM, M	The road class.
road_class_sort	integer	unknown	unknown
road	integer	none	A numerical description of 'road_name'
road_number	text	none	A coded description of 'road_name'
road_name	text	none	The road number relating to the section.
road_sort	integer	unknown	unknown
section	integer	none	A numerical description of 'section_label
section_label	text	none	The HAPMS section label, i.e. 0100M4/5.
section_sort	float	unknown	unknown
section_start_date	date	dd/mm/yyyy hh:ss	The date at which the section was last treated.
section_end_date	date	dd/mm/yyyy hh:ss	The date at which the section was originally constructed
section_length	float	meters	The length of the section
section_function	integer	1, 2, 3, 4	A numerical description of 'section_function_name'
section_function_code	text	MAIN, SLIP, OB, RBT	A coded description of 'section_function_name'
section_function_name	text	Main Carriageway, Slip Road, Ox Bow Lay-by, Roundabout	The function of the road section
section_function_sort	integer	unknown	unknown
core_status	Integer	unknown	unknown
core_status_name	text	unknown	unknown
tern_status	integer	792, 794, 795	unknown
tern_status_name	text	Yes, Not Specified, No	unknown
ha_region	integer	unknown	A numerical description of 'ha_region_name'
ha_region_name	text	unknown	The HA region in which the section occurs

Column heading	Data type	Unit / categories	Description
ha_regional_control_centre	integer	Not provided here for brevity	A numerical description of ‘ha_regional_control_centre_name’
ha_regional_control_centre_name	text	Not provided here for brevity	unknown
go	integer	unknown	unknown
go_name	text	unknown	unknown
round1_area	integer	unknown	unknown
round1_area_name	text	unknown	unknown
ha_category	integer	5516, 5517, 5518, 5519	A numerical description of ‘ha_category_name’
ha_category_name	text	Strategic, Not Specified, Regional, Toll Road or Route	The HA category related to the section
operational_area	integer	Not provided here for brevity	The NH operational area
operational_area_code	text	Not provided here for brevity	A coded description of ‘operational_area’
operational_area_name	text	Not provided here for brevity	A text description of ‘operational_area’
operational_area_sort	integer	unknown	unknown
data_key	integer	unknown	unknown
direction	integer	1,2,3,4,5,6	A numerical description of ‘direction_name’
direction_code	text	CW, AC, NB, EB, WB, SB	A coded description of ‘direction_name’
direction_name	text	Clockwise, Anticlockwise, North, East, West, South	The direction of the carriageway
direction_sort	integer	unknown	unknown
permanent_lanes	integer	1, 2, 3, 4, 5, 6	The number of permanent lanes present in the section
single_or_dual	integer	47, 49, 50, 51, 1612	A numerical description of ‘single_or_dual_name’
single_or_dual_code	text	DUALF, DUALT, S1W, S2W	A coded description of ‘single_or_dual_name’
single_or_dual_name	text	Dual Carriageway (Not Nominated), Dual Carriageway (Nominated), One Way Single Carriageway, Two Way Single Carriageway	A description if the pavement is a single or dual carriageway
single_or_dual_sort	integer	unknown	unknown

Column heading	Data type	Unit / categories	Description
environment	integer	8, 9, 10, 11	A numerical description of 'environment_name'
environment_code	text	NS, R, S, U	A coded description of 'environment_name'
environment_name	text	Not Specified, Rural, Suburban, Urban	The environment in which the section exists.
environment_sort	integer	unknown	unknown
local_authority	integer	Not provided here for brevity	A numerical description of 'local_authority_name'
local_authority_code	integer	Not provided here for brevity	A coded description of 'local_authority_name'
local_authority_name	text	Not provided here for brevity	The local authority within which the pavement exists.
local_authority_sort	integer	unknown	unknown
plan_reference	text	unknown	unknown
start_chainage	integer	meters	The start chainage of the section. This is always zero.
end_chainage	integer	meters	The end chainage of the section.

Appendix B Data Wrangling

B.1 Treatment data

The treatment data were prepared as follows:

1. The data from the construction data files were concatenated to a single dataframe.
2. Additional columns were created to describe the section lengths and the cross-sectional positions.
3. The data was merged with the Network Definition data.
4. Filters were applied to leave rows:
    a. Only including carriageway sections.
    b. Only including data from 2005.
    c. Having sub-section lengths >50m.
    d. Only having treatments of Inlay, Overlay, Re-construction, Re-surfacing, or Surface Treatment. This ensured that only sections subject to a pavement treatment were included in the dataset.

    This resulted in data containing a single row for each sub-section, describing when the sub-section was last treated and what that treatment was, and where the pavement construction and treatment was the same for each wheelpath.
5. Continuous sub-sections (connecting) within each section that were treated with the same treatment were joined together. This was carried out to improve the ease of joining the condition data during later data processing stages as well as overall data retention.

The resulting dataframe 'Treatment.csv' therefore contained data where each row represented a pavement sub-section where the same treatment was applied at (approximately) the same time. It is appreciated that within each pavement sub-section there is the potential for a mixture of material types but this was necessary for data retention purposes and is mitigated by the processing of the pavement condition data. A description of the data in Treatment.csv is presented in Table 8.

Table 8: Treatment.csv data description

Column heading	Data type	Unit / categories	Description
section	integer	none	A numerical description of 'section_label'
section_label	text	none	The HAPMS section
road_number	text	none	The HAPMS road number, i.e. M4.
single_or_dual_name	text	Dual Carriageway (Not Nominated), Dual Carriageway (Nominated), One Way Single Carriageway, Two Way Single Carriageway	A description if the pavement is a single or dual carriageway
environment_name	text	Not Specified, Rural, Suburban, Urban	The environment in which the section exists.
direction_code	text	CW, AC, NB, EB, WB, SB	A coded description of 'direction_name'
section	integer	none	A numerical description of 'section_label'
xsp_code	text	none	A coded description of 'xsp_name'
lane	integer	none	A description of the running lane derived from the xsp code.
section_function_name	text	Main Carriageway, Ox Bow Lay-by, Roundabout, Slip road	The HAPMS section function.
operational_area	integer	1, 2, 3, 4, 6, 7, 8, 9, 10, 13, 14, 125	The NH operational area
section_length_(m)	float	meters	The length of the HAPMS section.
start_chainage_(m)	float	meters	The start chainage of the sub-section within the pavement section
end_chainage_(m)	float	meters	The end chainage of the sub-section within the pavement section
subsection_length_(m)	float	meters	The length of the sub-section.
constr_start_date	date	dd/mm/yyyy hh:ss	The date at which the pavement layer was last treated.
pavement_class_name	text	Inlay, Overlay, Re-construction, Re-surfacing, Surface Treatment	A description of the last pavement treatment carried out.
group	integer	none	The group number.

B.2 Condition data

For each year, the SCRIM data were merged with the TRACS data. Because the SCRIM data are nominally 10m delimited and the TRACS data are nominally 100m delimited this merge was carried out by averaging the “characteristic_skid_coefficient” data occurring within each TRACS sub-section. The number of data points averaged were reported in the output data. These were then concatenated into a single dataframe and saved as a final ‘Condition.csv’.

The resulting ‘Condition.csv’ therefore contained data where each row represented a nominally 100m length of pavement and the condition, trafficking, and material data. A description of the data in Condition.csv is presented in Table 9.

Table 9: Condition.csv data description

Column heading	Data type	Unit / categories	Description
section	integer	none	A numerical description of ‘section_label’
section_label	text	none	The HAPMS section
start_chainage	float	meters	The start chainage of the sub-section within the pavement section
end_chainage	float	meters	The end chainage of the sub-section within the pavement section
sub_section_length	float	meters	The length of the sub-section
date laid	date	dd/mm/yyyy hh:ss	The date at which the material was laid.
survey_start_date	date	dd/mm/yyyy hh:ss	The date at which the survey was carried out.
survey_year	year	yyyy	The year during which the survey was carried out
xsp_code	text	none	A coded description of ‘xsp_name’
operational_area	integer	1, 2, 3, 4, 6, 7, 8, 9, 10, 13, 14, 125	The NH operational area
thickness	float	millimetres	The thickness of the surface layer
maximum_rut	float	millimetres	The maximum rut measured in either wheelpath
texture	float	millimetres	The Sensor Measured Texture Depth (SMTD)
(e)lpv_3m	float	none	Either the longitudinal profile variance (LPV) (for materials assessed before 2014), or the enhanced Longitudinal Profile
(e)lpv_10m			
(e)lpv_30m			

			Variance (eLPV) (for materials assessed during or after 2014).
fretting	float	none	The amount of fretting on the material
whole_cway_cracking	float	none	The amount of cracking on the material
category	integer	1, 2, 3, 4	The deflection category of the pavement.
characteristic_skid_coefficient	float	none	The characterisation of skid resistance of the pavement surface.
num_scrim_data	integer	none	The number of SCRIM data points included in the calculation of 'characteristic_skid_coefficient' for the sub-section length.
majority_material_code	text	Not provided here for brevity	The material code of the most prevalent surface material in the 100m section.
majority_material_length	integer	meters	The length of the 100m section which is the 'majority_material_code'.
surface_age_years	integer	years	The age of the majority surface material
sub_majority_material_code	text	Not provided here for brevity	The material code of the most prevalent material immediately beneath the surface in the 100m section.
sub_majority_material_length	integer	meters	The length of the 100m section which is the 'sub_majority_material_length'.
sub_age	integer	years	The age of the majority material immediately below the surface.
base_type	text	Not provided here for brevity	The subbase material code.
This_Lane_Large_Vehicle	integer	Vehicles per day	The daily flow rate of HGVs.

B.3 Preparing the data for machine learning

The preparation of data for machine learning was separated into the following stages:

1. Merging the Treatment and Condition data
2. Initial data filtering
3. Initial data exploration and feature selection / creation
4. Data filling
5. Data standardisation

6. Data balancing
7. Final data exploration

B.3.1 Merging the Treatment and Condition data

The Condition and Treatment data were merged in a similar manner to the SCRIM and TRACS data. Because the Condition data were 100m delimited and the Treatment data were not distance delimited, this merge was carried out by identifying the Treatment data which overlapped the Condition data and assigning the corresponding Treatment data to each nominal 100m sub section in the Condition data.

In addition to geographically merging the Treatment and Condition datasets, these datasets were also merged temporally. That is, the Condition data are available for a time period between 2005 and 2021 (inclusive). Condition data collected before treatment was carried out were assigned the 'pavement_class_name' (treatment carried out) for each sub-section. Condition data collected after treatment were likewise assigned a 'pavement_class_name' of 'none'. An example of this is provided in Table 10.

Table 10: Example of joining the Treatment and Condition data

Section_label	Start_chainage	End_chainage	Survey_year	Treatment_year	Years_to_maintenance	Pavement_class_name
0100A36/139	200	300	2008	2013	-5	Inlay
0100A36/139	200	300	2009	2013	-4	Inlay
0100A36/139	200	300	2010	2013	-3	Inlay
0100A36/139	200	300	2011	2013	-2	Inlay
0100A36/139	200	300	2012	2013	-1	Inlay
0100A36/139	200	300	2013	2013	0	treatment_year
0100A36/139	200	300	2014	2013	1	none
0100A36/139	200	300	2015	2013	2	none
0100A36/139	200	300	2016	2013	3	none
0100A36/139	200	300	2017	2013	4	none
0100A36/139	200	300	2018	2013	5	none
0100A36/139	200	300	2019	2013	6	none
0100A36/139	200	300	2020	2013	7	none
0100A36/139	200	300	2021	2013	8	none

B.3.2 Initial data filtering

Initial data filtering was carried out to obtain data for each sub-section, for which:

- The Treatment data overlapped the Condition data
- The 'treatment_year' and 'survey_year' is different. This was carried out to reject data where the Condition data could not have had an influence on a treatment being carried out or not
- There was at least three years between the most recent 'survey_year' and the actual 'survey_year'. This was carried out to improve the confidence in sections identified as not having been treated
- There was at least three years of condition data before the treatment. This was to account for the 'lag' between road surface condition monitoring and treatment

Appendix C Machine Learning models

C.1 Nearest centroid

Nearest centroid identifies the nearest average value (centroid) in an n-dimensional feature space and assigns a classification based on that assigned to the nearest average value An example of this for a 2 dimensional feature space with four classes is provided in Figure 10. In this example the centroid for class 3 is nearest to the new input value ('x') and so it is assigned a class of 3.

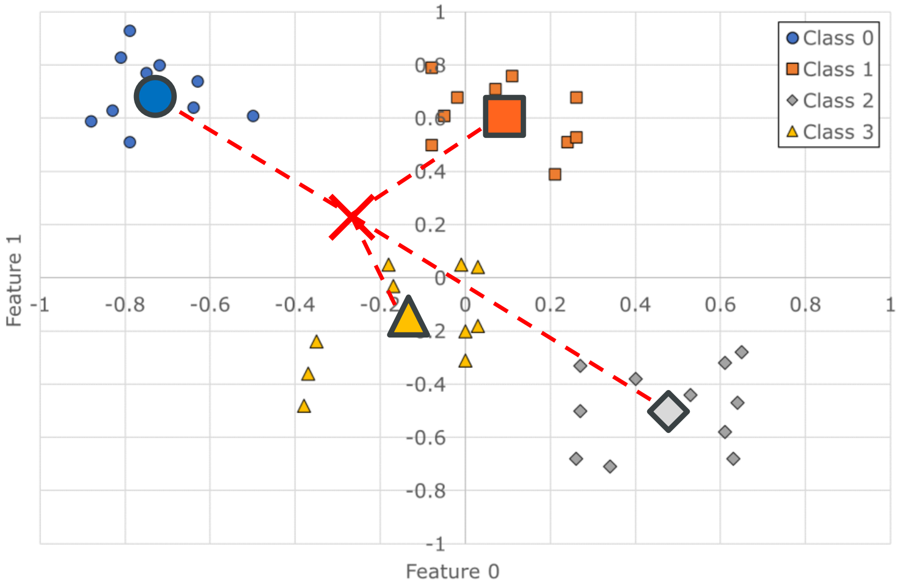


Figure 10: Example of nearest centroid (centroids are large grey series markers, the value to be classified is marked as a red 'X')

C.2 K Nearest neighbours (kNN)

A KNN model 'votes' on the class of a new input value based on the classes of the k number of neighbours to the new input value in an n dimensional feature space. In the case of a tie in voting the model will assign a class at random, or calculate the total distance to each class and assign that which is lowest. An example of this process is show in Figure 11. In this example the kNN model has received 3 'votes' for class 3 and 2 votes for class 2 so class 3 is assigned to the new input.

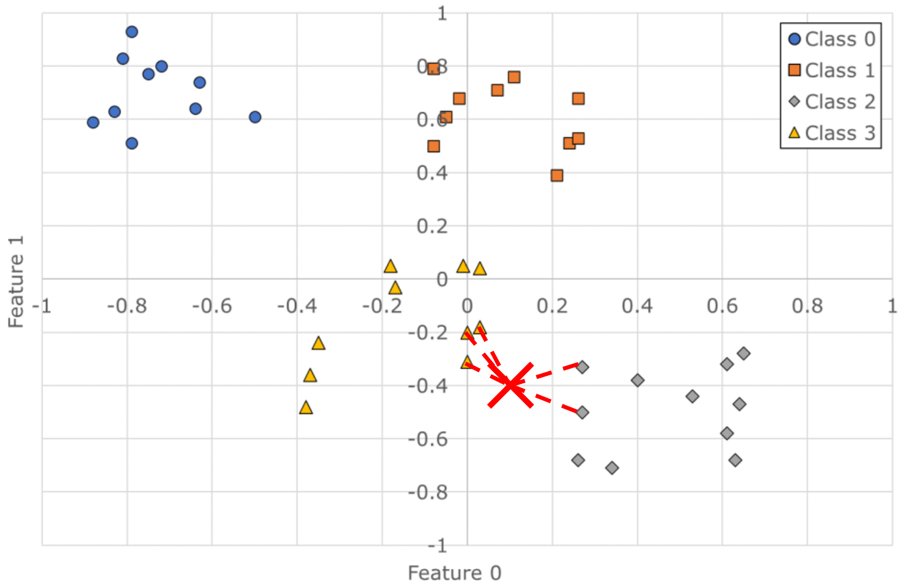


Figure 11: Example of k nearest neighbours with k=5

C.3 Random Forest Classifier (RFC)

A random forest classifier is based on the framework of a recursive decision tree. A decision tree functions by applying a set of binary rules to the model data in an iterative process which 'splits' the data into increasingly smaller parts until the decision tree terminates at a leaf node and produces a classification. The rules to be applied to the data are selected by a brute force method and the rule which maximises the 'purity' of the data as determined by the Gini index which is a measure of how well a rule has split the data by the defined classes.

A random forest classifier is a natural progression of recursive decision trees. In a random forest classifier a 'forest' of decision trees are trained on various sub-sets of the training data. This forest of trees allows for the classification of a new input value to be carried out multiple times (once for each tree in the forest) and a vote taken on the class that should be assigned to that input.

C.4 Support Vector Machines (SVM)

Support vector machines seek to segregate an n dimensional feature space into sectors which best separate the categories within that space. This can be thought of as drawing lines (vectors) through the space which maximise the distance between categories. A heavily simplified example of this process is show in Figure 12, which presents a two dimensional feature space with two classes. In reality, higher order geometries are used to segregate feature spaces but an example using linear vectors is provided here for simplicity.

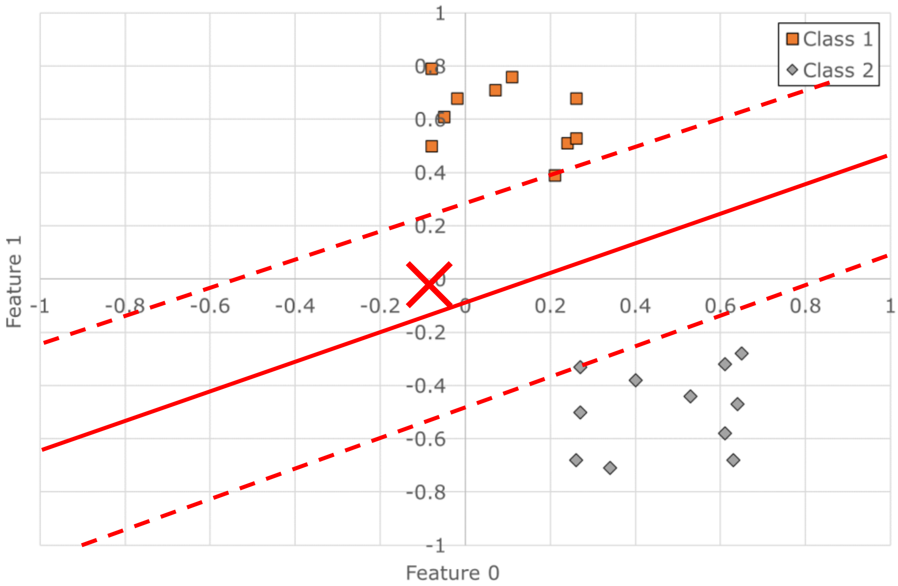


Figure 12: Example of support vector machines

Figure 12 shows the maximal margin separating the two classes (the solid line) between the two broken lines (the support vectors) which represent the maximum limits of the classes. In this example the new input (the red “x”) sits above the maximum margin separating Class 1 and Class 2 i.e. within the Class 1 category. The new input data would therefore be assigned Class 1.

C.5 Voting classifier

A voting classifier is an amalgamation of the previous methods mentioned in this section. A voting classifier will make predictions based on each of the models assigned to it and use the results of the predictions from individual models to ‘vote’ on the correct classification for a new input.

For example:

- The following models are assigned to a voting classifier:
  - RFC
  - kNN
  - SVC
- The models returned the following classifications:
  - RFC (Class 0)
  - kNN (Class 0)
  - SVC (Class 1)

If equal weight is applied to each vote then the voting classifier would assign Class 0 to the new input data.

Appendix D Final data exploration

To quality assure the data before assessing them with machine learning, the 'balanced' dataset was explored. The key findings from this data exploration, and the conclusions from it are provided below, a full suite of findings can be found in Appendix F. Figure 13 presents the linear coefficients of determination between each of the continuous model features. Here it can be observed that none of the final model features are well correlated with any other model feature. This has removed the co-dependency in the data originally observed.

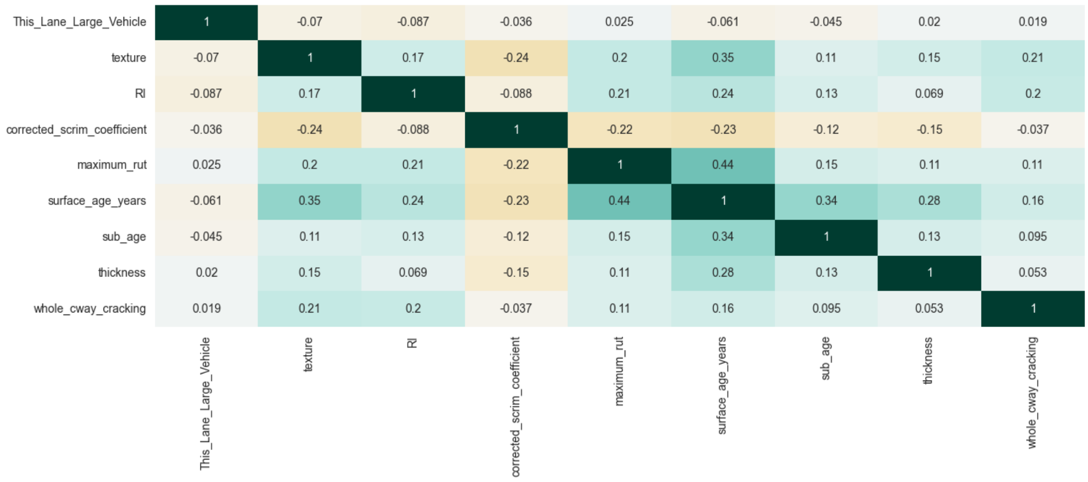


Figure 13: Correlations between final model features

Figure 14 confirms that the data have been correctly balanced as the '0' treated data (i.e. the data relating to lengths that were not treated.) is now at 50% of the total dataset.

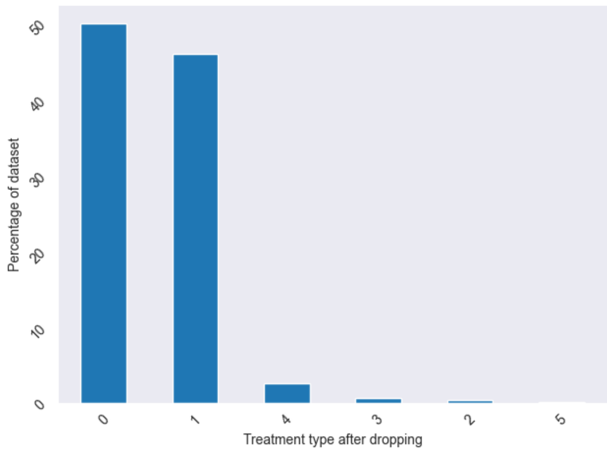


Figure 14: The prevalence of treatment types after balancing

D.1.1 Machine learning data description

The processes described above resulted in a dataset containing 147,254 rows representing a total length of approximately 14,725km). A description of the data in this dataset is presented in Table 11.

Table 11: Machine learning data description

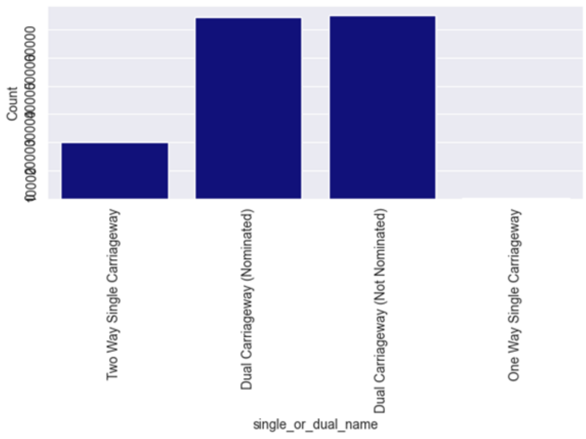
Column heading	Data type	Unit / categories	Description
treatment_binary	integer	0, 1	The sub-section was treated (1) or not treated (0)
treatment_multi	integer	0, 1, 2, 3, 4	A description of the treatment carried out.
maximum_rut	float	none	A standardised description of the amount of rutting present on the sub-section.
texture	float	none	A standardised description of the texture depth on the sub-section.
RI	float	none	A standardised description of the longitudinal profile on the sub-section.
whole_cway_cracking	float	none	A standardised description of the cracking on the sub-section.
category	integer	none	The deflection category on the sub-section.
This_Lane_Large_Vehicle	float	none	A standardised description of the trafficking on the sub-section.
surface_age_years	float	none	A standardised description of the surface age on the sub-section.
sub_age	float	none	A standardised description of the sub-surface age on the sub-section.
majority_material_code	integer	Not provided here for brevity	A numerical description of the material code of the most prevalent surface material in the sub-section.
sub_majority_material_code	integer	Not provided here for brevity	A numerical description of the material code of the most prevalent sub-surface material in the sub-section.
base_type	integer	Not provided here for brevity	A numerical description of the material code of the most prevalent base material in the sub-section.
thickness	float	none	A standardised description of the thickness of the surface layer on the sub-section.
operational_area	integer	1, 2, 3, 4, 6, 7, 8, 9, 10, 13, 14, 125	The NH operational area
characteristic_skid_coefficient	float	none	A standardised description of the corrected skid coefficient on the sub-section.
environment_name	integer	0, 1, 2	A numerical description of the environment name of the sub-section.

Column heading	Data type	Unit / categories	Description
single_or_dual_name	integer	0, 1, 2, 3	A numerical description of the single_or_dual_name name of the sub-section.
group	integer	none	The group assigned to the sub-section
survey_year	year	YYYY	The year that the condition surveys were carried out.

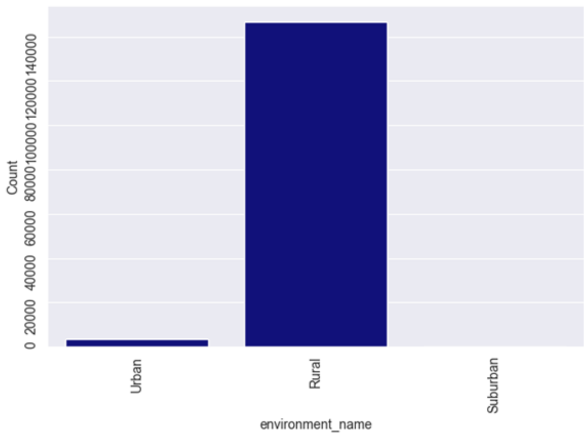
Appendix E Initial data exploration

E.1 The distribution of the initial model features

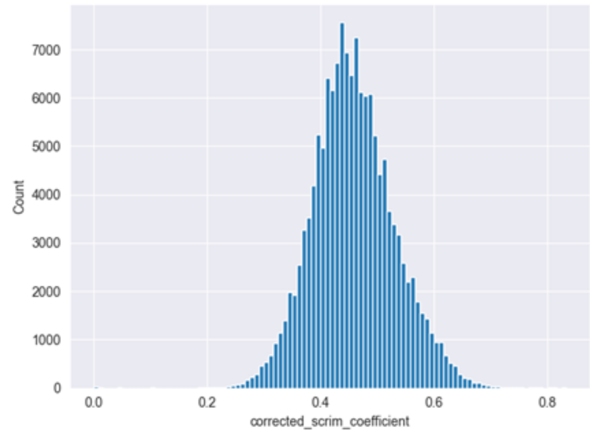
Table 12: The frequency distributions of the initial model features



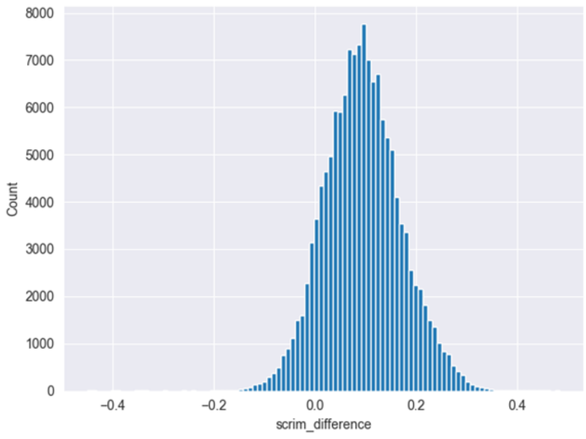
The distribution of carriageway types



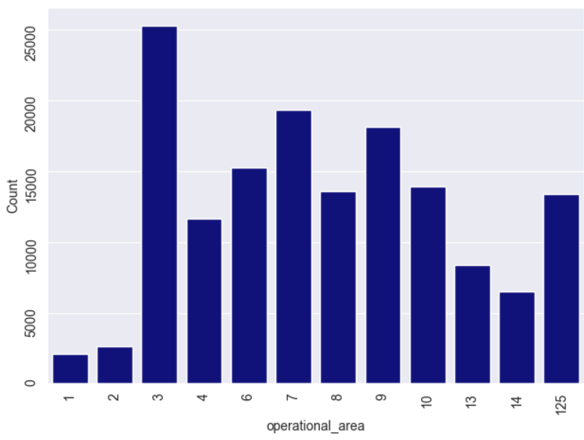
The distribution of environments



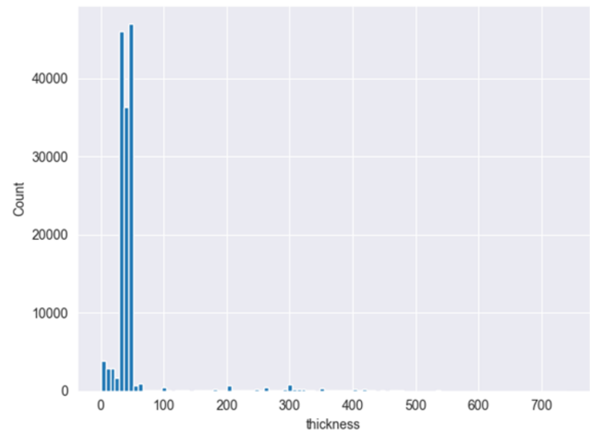
The distribution of characteristic skid coefficients



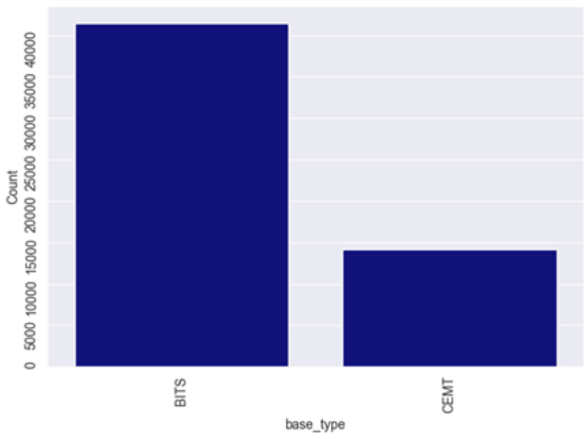
The distribution of skid differences



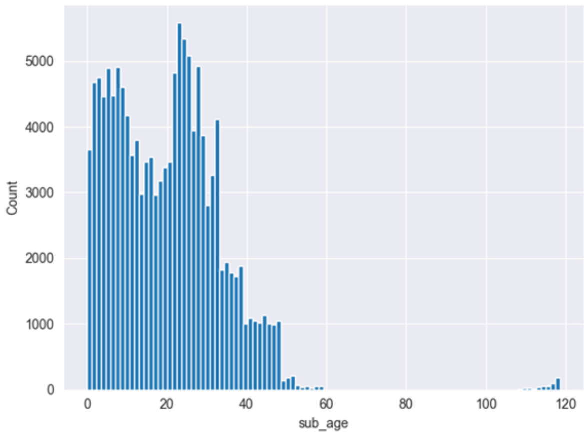
The distribution of operational areas



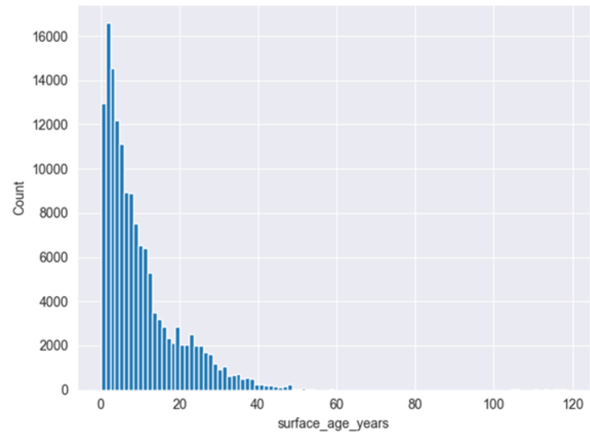
The distribution of surface layer thicknesses



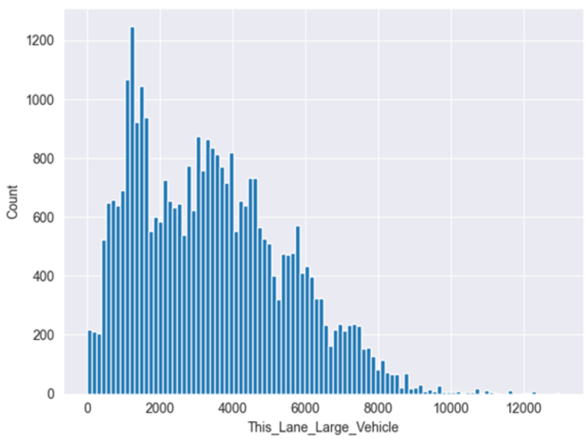
The distribution of foundation (base) types



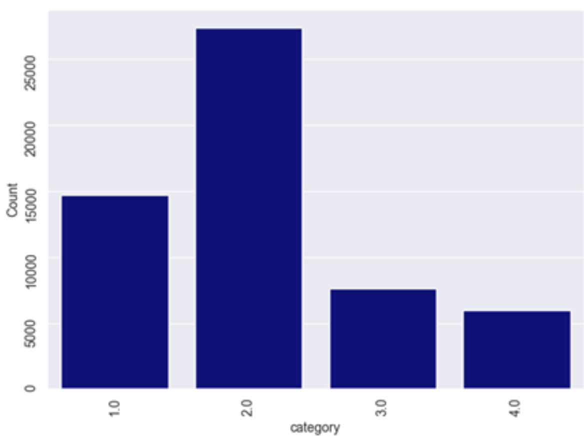
The distribution of sub surface material ages



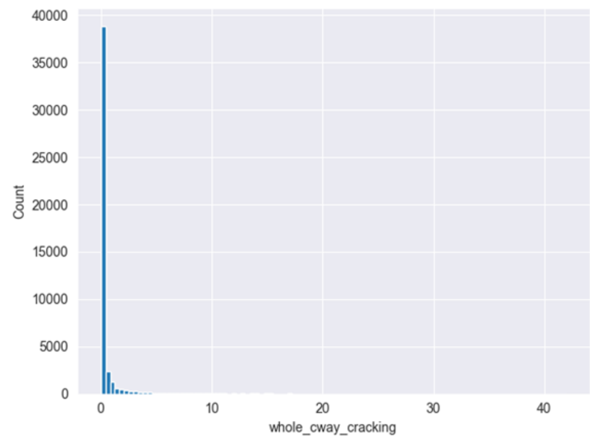
The distribution of surface material ages



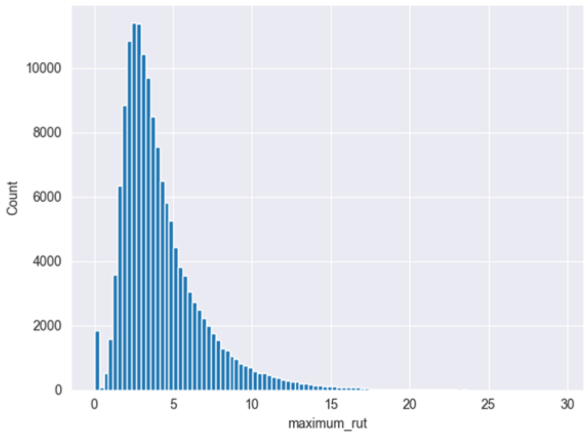
The distribution of daily HGV trafficking rates



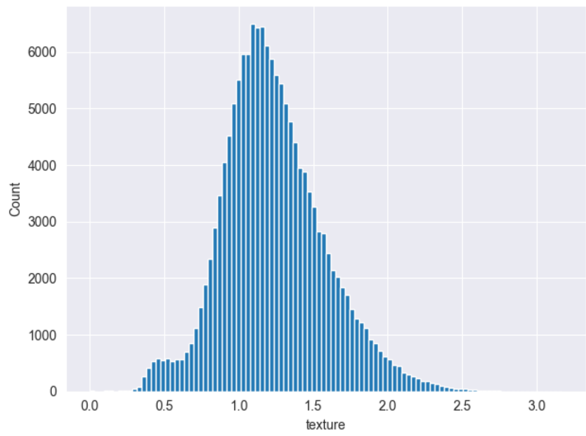
The distribution of deflection categories



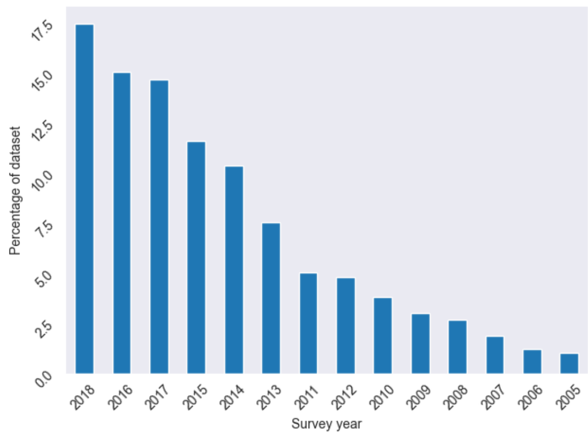
The distribution of cracking values



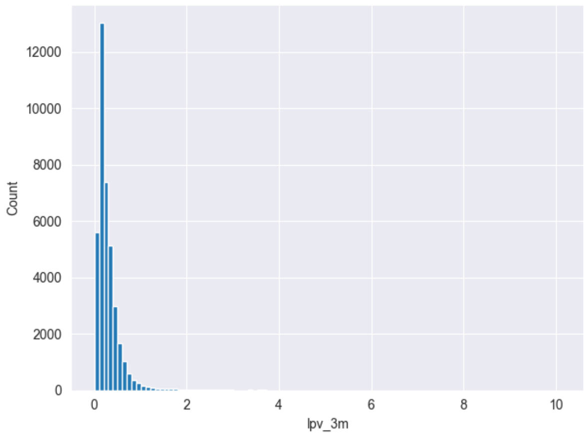
The distribution of rutting values



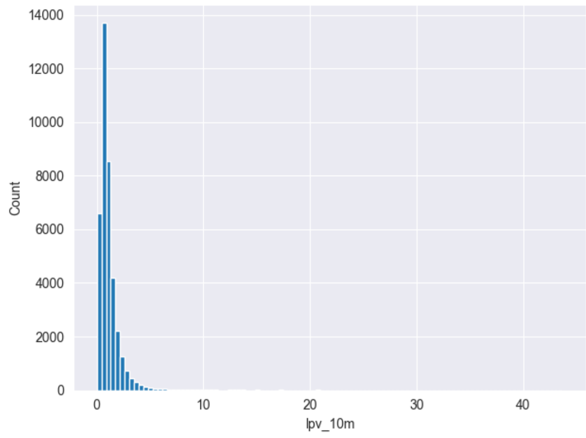
The distribution of texture depth values



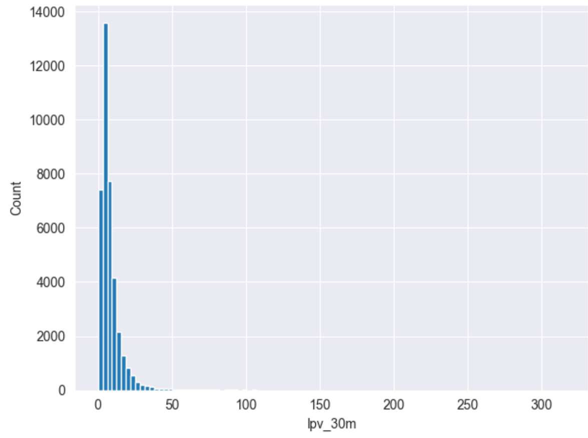
The distribution of survey years



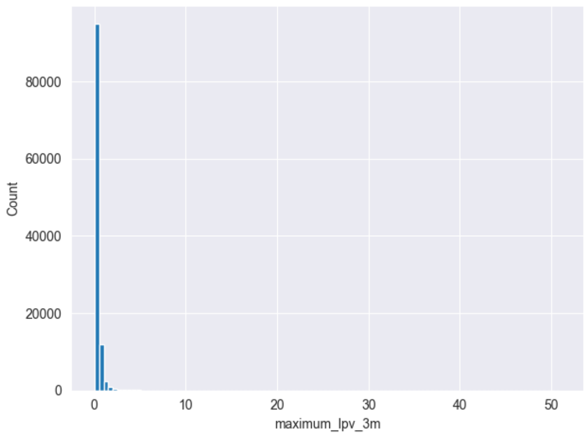
The distribution of 3m LPV values



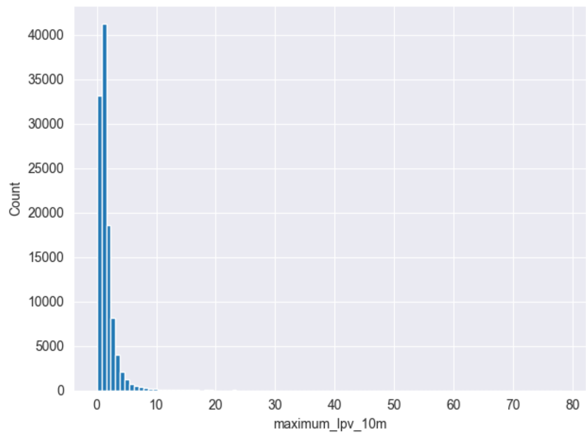
The distribution of 10m LPV values



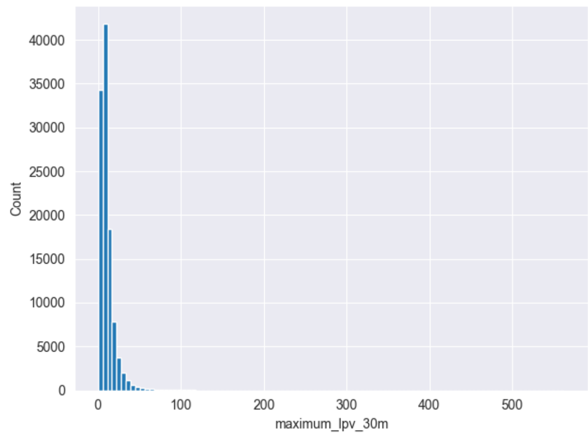
The distribution of 30m LPV values



The distribution of 3m eLPV values



The distribution of 10m eLPV values

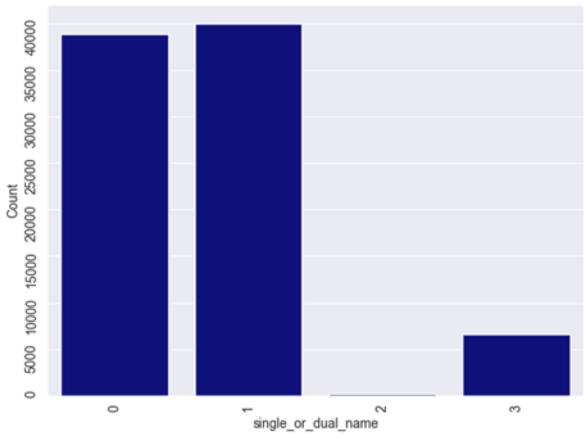


The distribution of 30m eLPV values

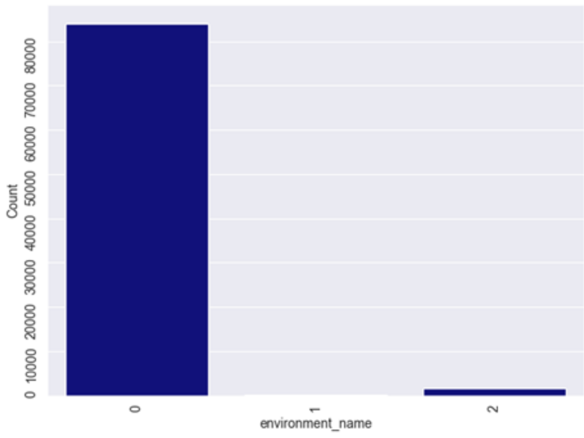
Appendix F Final data exploration

F.1 The distribution of final model features

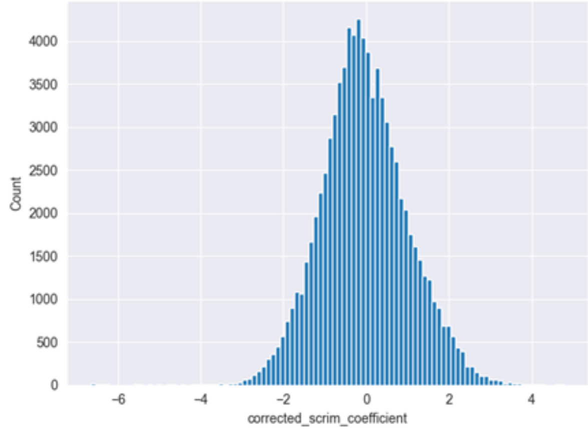
Table 13: The frequency distributions of the final model features



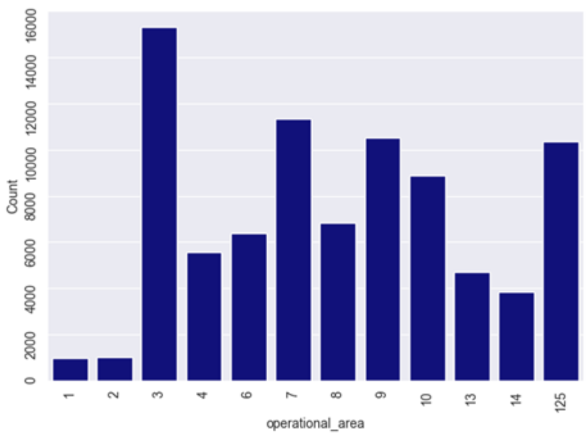
The distribution of carriageway types



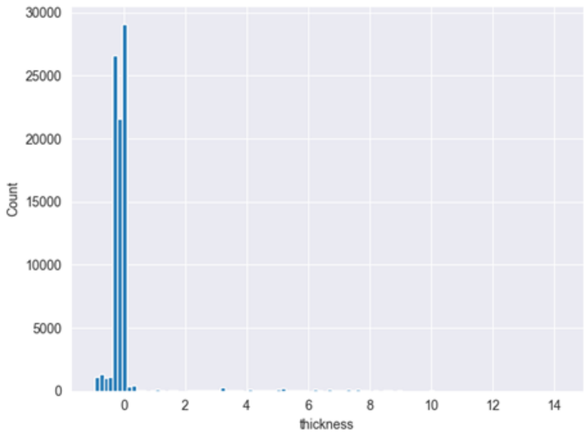
The distribution of environments



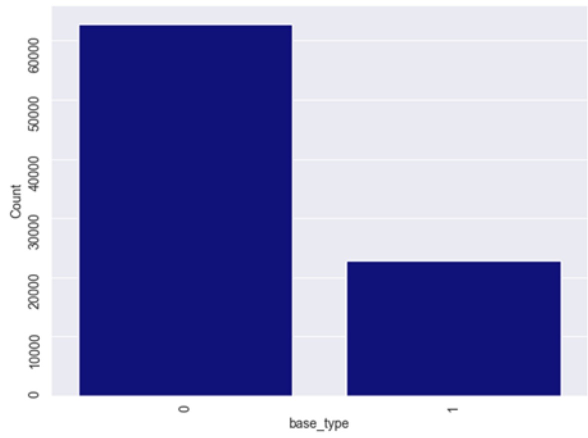
The distribution of standardised characteristic skid coefficients



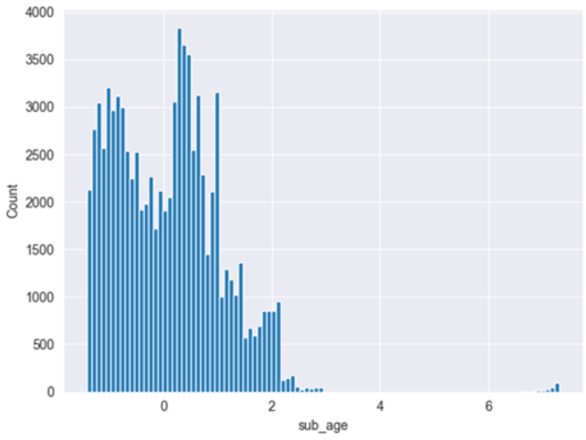
The distribution of operational areas



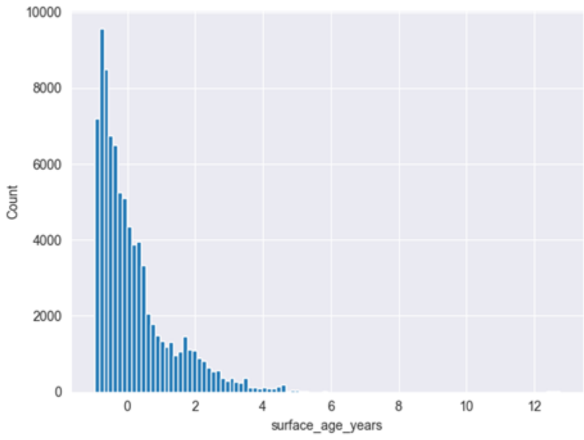
The distribution of standardised surface layer thicknesses



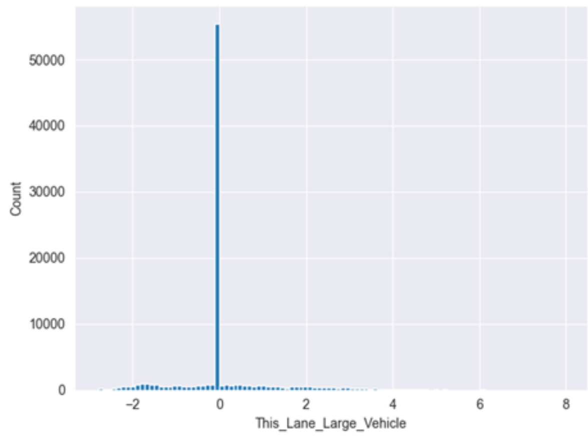
The distribution of foundation (base) types



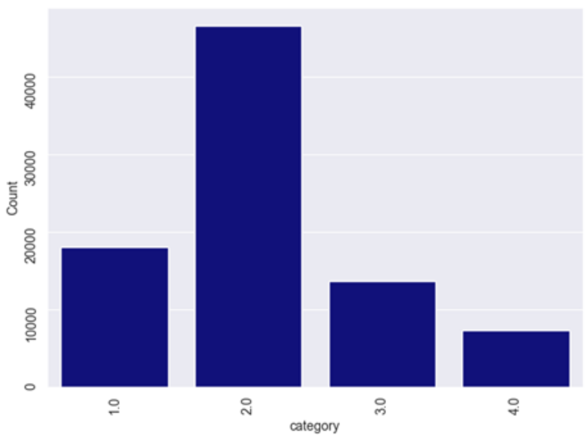
The distribution of standardised sub surface material ages



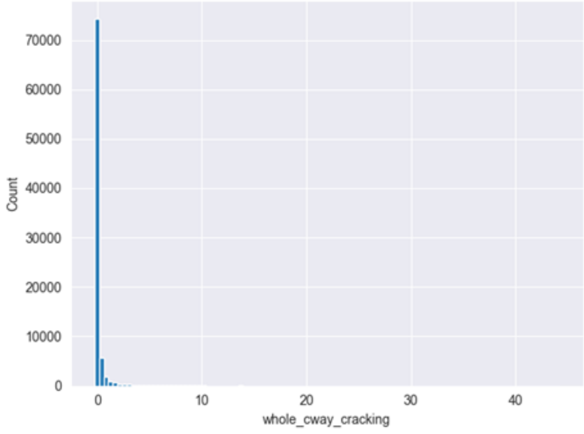
The distribution of standardised surface material ages



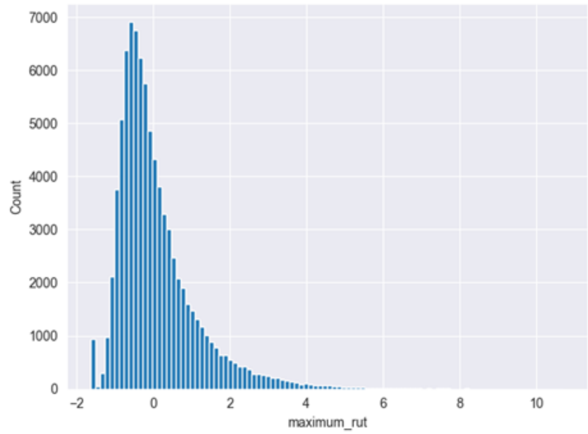
The distribution of standardised daily HGV trafficking rates



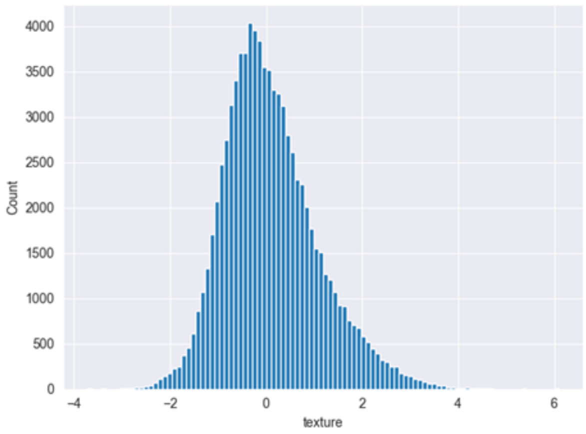
The distribution of deflection categories



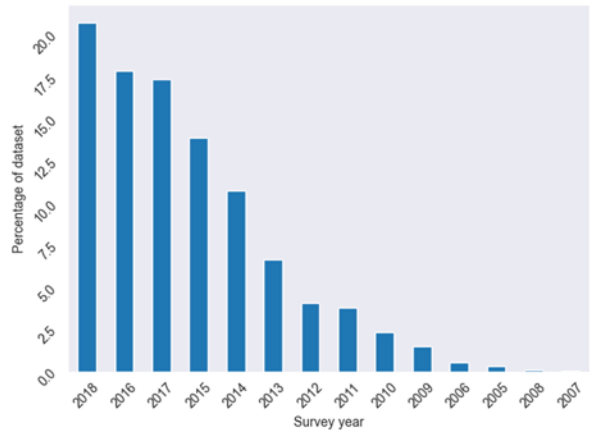
The distribution of standardised cracking values



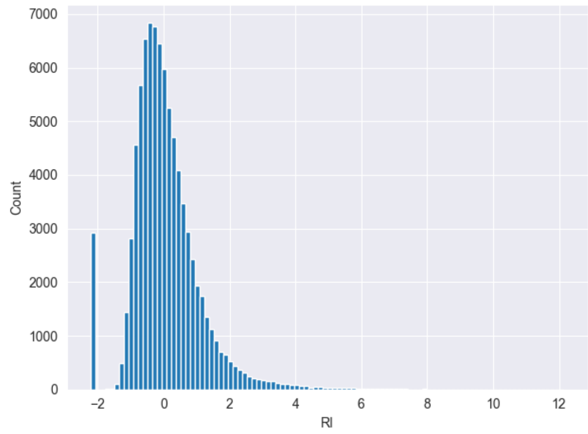
The distribution of standardised rutting values



The distribution of standardised texture depth values



The distribution of survey years



The distribution of standardised rutting index

F.2 Feature Importance

The toolkit provided functions to characterise the influence of each model feature in determining the outcome of the model (treat / do not treat).The assessment of feature importance is presented in Figure 15 as a box plot, where the magnitude of the feature importance is shown on the x-axis and the features are presented on the y-axis.

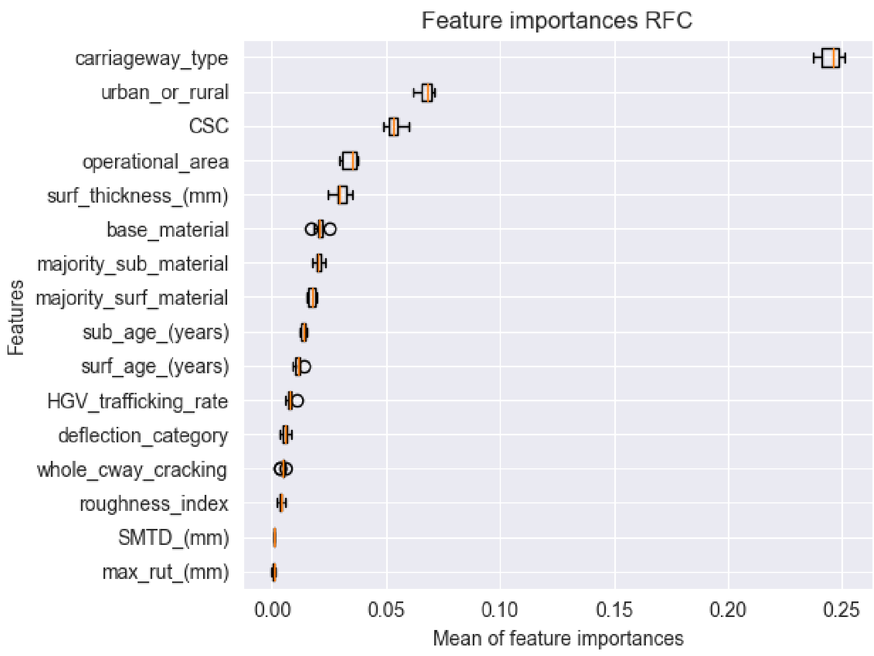


Figure 15: Ranking of importance of features to the outcome of the model

The application of artificial intelligence to road pavement maintenance assessment

Currently, roads authorities apply deterministic (rule based) digital tools to recommend lengths that should be considered for maintenance (using condition data provided by automated and human assessors). However, these recommendations do not robustly reflect the decisions that engineers would themselves make. This work investigates how an outcome-based model could be developed to better identify lengths for treatment. The development of the model draws on network level condition data (from the Strategic Road Network) that includes visual condition, roughness and skid resistance, and contextualising information such as construction, traffic, material and age. These are collated and aligned with data on the actual treatments that were carried out on the network, in order to train and test a set of machine learning models. The best performing of these models deploys the Random Forest Classifier, which is referred to in this work as the Digital Engineer.

A comparison between the locations identified for treatment by the Digital Engineer, the locations identified by rules-based tools, and the locations that were actually treated, shows that the Digital Engineer provides a significantly higher level of overall accuracy in the identification of these lengths. The results suggest that additional contextualising information assists in achieving outcomes from digital tools that better agree with the decisions made by engineers, and that Machine Learning techniques may be used to apply this additional information. It is recommended that further development and testing of the Digital Engineer approach should: better understand the influence of features on tool decisions; more comprehensively verify the model; and determine the route to implementation.

Other titles from this subject area

PPR863	Applications of machine learning in transport. K Nesnas, R Khatry, A Smirnov, D Peeling, S Mistry, M Crabtree and S Reeves. 2018
XPR015	A new approach to asset management of unpaved roads. R Workman, Z Wang and K Nesnas. 2022
PPR2042	Road Condition Monitoring Data - Network Study. A Wright and S Brittain. 2024
PPR2001	Investigation of the impacts of climatic conditions on skid resistance variation. T Andriejauskas and Z Wang. 2022

TRL
Crowthorne House, Nine Mile Ride,
Wokingham, Berkshire, RG40 3GA,
United Kingdom
T: +44 (0) 1344 773131
F: +44 (0) 1344 770356
E: enquiries@trl.co.uk
W: www.trl.co.uk

ISSN 2514-9695

DOI 10.58446/lnxe9255

ACA108