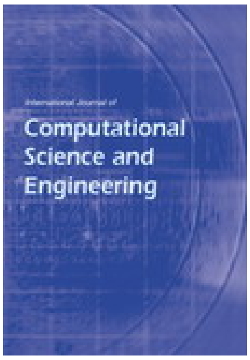


International Journal of Computational Science and Engineering

ISSN online: 1742-7193 - ISSN print: 1742-7185
https://www.inderscience.com/ijcse

A predictive model based on the LSTM technique for the maintenance of railway track system

Sharad Nigam, Divya Kumar

DOI: 10.1504/IJCSE.2024.10063100

Article History:

Received:	12 February 2023
Last revised:	15 November 2023
Accepted:	14 January 2024
Published online:	21 December 2024

A predictive model based on the LSTM technique for the maintenance of railway track system

Sharad Nigam* and Divya Kumar

Motilal Nehru National Institute of Technology,
Allahabad, Prayagraj, India
Email: sharad.2017rcs11@mnnit.ac.in
Email: divyak@mnnit.ac.in
*Corresponding author

Abstract: Maintenance is a substantial process to sustain the operations of transportation systems. Railway is a major way of transportation, defect and failure in track side equipment and track itself may causes major loss of human lives. So, an effective maintenance technique is needed which saves the passenger lives and maximises the utilisation of railway track equipment, just before it fails, that cause casualty. There are various types of track defects to be inspected, but this paper only deals with surface defects, cross level, and DIP. By examining the track condition and expecting the RUL, the railway industry can schedule maintenance of railway tracks. In this research paper, we have tried to conduct a relative analysis of various machine learning algorithms by comparing their performance for estimation of future failure points of railway tracks to inspect the reliability of LSTM technique. Dataset is taken from the trusted source RAS Track geometry analytics (2015).

Keywords: predictive maintenance; K-nearest neighbour; KNN; machine learning; long short-term memory; LSTM; railway track; support vector machine; SVM; remaining useful life; RUL.

Reference to this paper should be made as follows: Nigam, S. and Kumar, D. (2025) 'A predictive model based on the LSTM technique for the maintenance of railway track system', *Int. J. Computational Science and Engineering*, Vol. 28, No. 1, pp.10–20.

Biographical notes: Sharad Nigam is a senior research scholar in Computer Science and Engineering Department of Motilal Nehru National Institute of Technology, Prayagraj, India. His research interest includes railway transportation, predictive maintenance, machine learning, deep learning, internet of things, wireless sensor network, and data science.

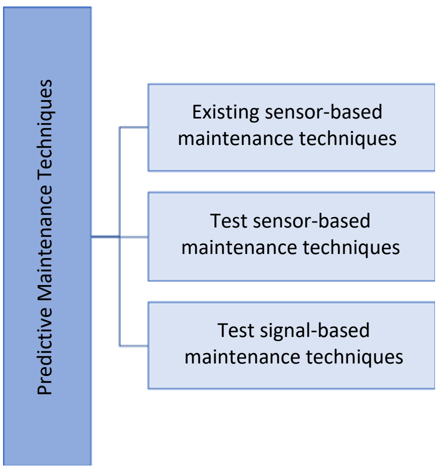
Divya Kumar is an Assistant Professor in Computer Science and Engineering Department of Motilal Nehru National Institute of Technology, Prayagraj, India. He has 12 years of teaching and research experience, and published numerous research articles in reputed SCI and Scopus journals. He has expertise in the field of optimisation algorithms, machine learning, block chain, reliability analysis and predictive maintenance. He has guided 30 post graduates and two PhD scholars.

1 Introduction

We depend on a large range of machines every day. But the fact is that, in the end, every system breaks down unless it is being maintained. The cost of maintenance is a substantial proportion of the overall cost of service in any industry. For industries, maintenance costs can vary from 15% to 60% of the total cost of the produced goods. For example, it is 15% for the food-associated industry, but for heavy industries like iron, steel, and paper it can be up to 60% (Mobley, 2002). Also, the unexpected failure of heavy machines can be a threat to human life. Dirk Claessen, Managing Director of IBM for its Royal Dutch Shell account, said that they conducted a study of China's coal-mining industry, where 3,000 people lose their lives every year and 80% of these deaths were caused by machine failure.

Predictive maintenance aims to predict the point of time at which the asset of the machine will fail and prevent the chance of failure by performing the required maintenance task (Wang et al., 2015). Earlier predictive maintenance was used in heavy machinery or very critical applications only but now it is used for simpler tasks also like in food storage to monitor the temperature, humidity and other related functions with progress in technology. Tremendous advancement has been made in the field of predictive maintenance over time (Kou et al., 2021). Nowadays, complex machinery and functions are included in most asset-intensive industries. The downtime cost is very high and the failure disrupts the manufacturing supply chain. Also, the competition in this field is very high and increases with time. The reliability and maintenance cost of their asset is one of the primary competitive means in this era of competition. To increase the reliability and decrease maintenance cost, actions such as application of machine learning on data to find a pattern of risk, big data analysis of real-time data apart from visual inspections, instrument readouts is needed and it is done by most of the critical applications-oriented companies and where the failure or downtime causes a critical effect (Yang et al., 2021). If keen eyes, nose, and ears are utilised to detect the cause of failure as equipment begins to fail, it can show signs that can be identified. Thankfully, sensors are there to act as keen eyes, nose, and ears to detect the beginning of equipment degradation and breakdown. As noted, in 89% of cases, predictive maintenance is the preferred form of maintenance, compared to time-based maintenance, which is prudent in just 11% of cases (Hashemian, 2010). Based on data sources, predictive maintenance can be divided into three basic techniques. Existing sensor-based maintenance techniques calculate variables such as temperature, pressure, level, and flow, using data from existing process sensors, like pressure sensors, temperature sensors, and resistance temperature detectors (RTDs). Test sensor-based maintenance techniques utilise data from test sensors such as vibration measuring accelerometers and leak detection acoustic sensors. In this method, wireless sensors will play a significant role. Test signal-based maintenance a technique relies on signals poured into the machinery and have them checked. Active measurements such as insulation resistance as well as inductance, capacitance, and measurements of resistance, also known as measuring for LCR are included in this category. This technology is used to detect defects such as cracks, corrosion, and wear for predictive maintenance of wires, generators, sensors and other services (Ayvaz and Alpay, 2021).

Figure 1 Techniques of predictive maintenance (see online version for colours)



1.1 Related work

Some of the relevant and recent work done in the field of predictive maintenance is discussed in this section. Karimpour et al. (2018) have worked on neuro fuzzy model using artificial intelligence. The authors analysed the gauge values and found that straight sections were more responsible for serious faults. They predicted the gauge value for next year if past year data is present. Andrade and Teixeira (2012) have worked on Bayesian model due to different kind of defects by using the inspection data with help from Office of Rail Regulation (ORR). They analysed the risk associated cost in life cycle of rail. He et al. (2015) employed statistical model for analysing derailment risk associated with track related issues or defects. In the paper, they proposed a framework for capturing several kinds of defects and methods needed to be done. The problem and work resemble up to some extent a part of our work. Vale and Lurdes (2013) employed statistical model for analysing derailment risk associated with track related issues or defects. In the paper, they proposed a framework for capturing several kinds of defects and methods needed to be done. The problem and work resemble up to some extent part of our work. Shafahi et al. (2008) analysed states of track based on five-time interval. Markov chains, artificial neural network and neuro fuzzy network are used for the prediction and maintenance. Susto et al. (2014) designs predictive maintenance model for railway track, switches by applying tree-based classification machine learning technique (Bukhsh et al., 2019). To improve the predictive maintenance decision making process, this model collects local interpretable model agnostic explanation framework as well as global data explanation. Xie et al. (2020) has worked on predictive maintenance technique for railway tracks using unsupervised learning, deep learning, and ensemble method to detect rail geometry irregularity, missing rail components, rail head defects, and other critical issues in track. Li et al. (2014) introduced a model that automatically learns the track signalling rules and predicts build failure in railway tracks. This predictive model predicts the future failure by collectively analysing both historical and real time data. Overall data are explored by various analytical approaches including correlation analysis, time series analysis, and causal analysis. Gerum et al. (2019) presented an approach to predict rail track geometry defects by analysing inspection data and maintenance scheduling data. This paper concludes that to construct an accurate prediction model, detailed data may not be necessary if general yet accurate data are available. Faiz and Singh (2009) discussed time based predictive maintenance to predict issues in track geometry. This predictive model improves the accuracy in prediction in the maintenance management dispensation.

1.2 Main contribution

This paper concerns the predictive maintenance of railway tracks by performing the hypothesis testing using K-fold validation. There are various types of track defects to be inspected, but this paper only deals with surface defects, cross level, and DIP. Track defects are measured as red tag and yellow tag as per the level of severity. Railway track dataset is a sequential time series data that requires current as well as previous sequential data to predict the maintenance requirement in certain and uncertain condition. Various machine learning models like linear regression, logistic regression, random forest, naïve Bayes, decision tree, K-nearest neighbour (KNN) and support vector machine (SVM) are discussed. The paper discusses LSTM as a deep learning concept that can be used for predictive maintenance. As per the name of long short-term memory (LSTM), it does not process only for the single data points but also for entire sequence of data.

The main contribution of this paper is to compare other machine learning algorithms with LSTM and prove that LSTM performs better for the predictive maintenance of railway tracks. Section 2 will discuss dataset description of railway tracks geometry. Section 3 will discuss proposed schema with experimental setup and dataset pre-processing. Section 4 describes and compares various existing machine learning algorithms with LSTM for predictive maintenance. Section 5 will elaborate on the experimental observation and result with confusion matrix, ROC curve, accuracy, precision, recall, and F1 factor. Section 6 performs hypothesis testing of machine learning algorithms. Section 7 concludes the article and last section contains the included references.

Table 1 Attribute description of railway tracks dataset

Attribute	*Description*
MILEPOST	Point on the track
TRACK SDTK NBR	Track type
TEST DT	Inspection date
DEF NBR	Defect Sequence number
GEO CAR NME	Track geometry car name
DEF PRTY	Severity of the defect: yellow or red
DEF LGTH	Length of defect in feet
DEF AMPLTD	Maximum size of defect in inches
TSC CD	Track type (tangent, spiral and curve)
CLASS	Class of track
TEST FSPD	Operating speed of freight train
TEST PSPD	Operating speed of passenger train
DEFCT TYPE	Defect type (XLEVEL, SURFACE, DIP)
TOT CAR EAST	Total number of cars travelling to east
TOT CAR WEST	Total number of cars travelling to west
TOT TRN EAST	Total number of trains travelling to east
TOT TRN WEST	Total number of trains travelling to west
TOT DEFLT MGT	Sum of total gross tons

2 Dataset description

2.1 Railway tracks

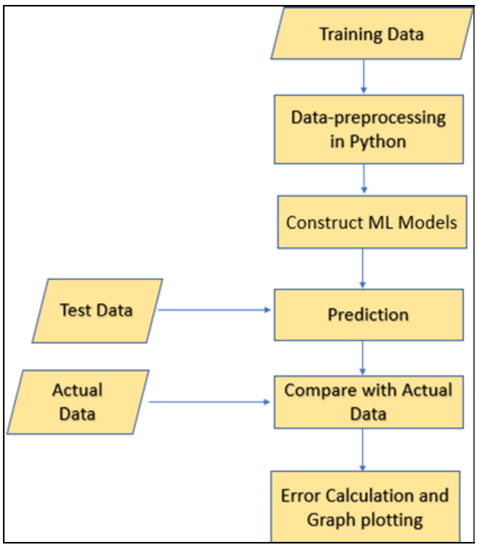
The dataset used for predictive maintenance of railway tracks is obtained from RAS Track Geometry Analytics (2015). RAS Problem Solving Competition 2015' (RAS 2015) is used to build a predictive model for the degradation of track geometry using information on track layout, its characteristics, traffic and a series of historical inspection records. Using visual inspection and technology such as induction and ultrasonic systems, the information was acquired by rail track vehicles operating on the rail switches and inspecting railway lines for rail defects. To reliably distinguish places where measurements have taken place, the vehicles are often fitted with the global positioning system (GPS). The car for examination can distinguish around 40 different variants of defects. But the dataset chosen can distinguish the following three defects, i.e., surface defects, cross level and DIP. Track defects are classified into two severity levels, i.e., red tags and yellow tags. Red tag defects do not fulfil the Federal Railroad Administration (FRA) track safety requirements and must be handled as soon as possible after they are detected while yellow tag defects meet FRA specifications, but do not comply with the basic requirements of railroad. If yellow tag defects are not corrected, they will gradually become red tag defects. Some of the important attributes are discussed in Table 1.

3 Proposed schema

3.1 Experimental setup

The machine learning models were built and the results were tested using the Python programming language. The algorithms were applied to the data sets that were read from files using the Python programming language. Python has the advantage of being simple to import data, process it, and visualise the results. The outcomes were saved as variables. The visualisation methods available in the matplotlib library were used to plot these variables. Flow chart of proposed method is shown in Figure 2.

Figure 2 Flow-chart of proposed method (see online version for colours)



3.2 Dataset pre-processing

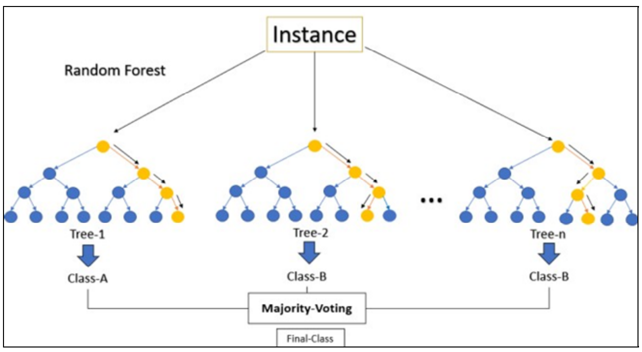
Dataset was required to be pre-processed. In this pre-processing, we have to remove the repeated inspections. We have to remove unnecessary defect parity changes over time. This means we are taking the records which are not downgraded from red to yellow. We have also removed some columns and depending upon previous research related to this, we have used certain attributes like defect parity, defect amplitude, milepost, defect magnitude, and the number of days after which we have to make a prediction (Kocbek and Gabrys, 2019).

4 Machine learning algorithm for prediction

We have explored many machine learning algorithms for prediction like random forest, logistic regression, decision tree, LSTM networks, etc. LSTM proved to be the best choice because it has memory to save long time series dependencies while extracting complex correlations between data, which has higher weightage in determining the condition of the system (Cai et al., 2022).

Figure 3 Machine learning for predictive maintenance (see online version for colours)

Figure 4 Random forest (see online version for colours)



4.1 Random forest

Random forest is a supervised learning algorithm. As its name suggests, it consists of many single decision trees that act as a group. Every single tree in the random forest draws out a class prediction, and the class with the majority votes becomes the prediction of the model, as shown in Figure 4. RF can be used for both classification and regression tasks. Also, there is a problem of overfitting in the decision tree. In contrast, in most cases, random forest does not suffer from overfitting as they operate with random feature subsets and construct smaller trees from such subsets (Biau and Scornet, 2016; Prytz et al., 2015).

4.2 Logistic regression

Regression analysis is a predictive modelling technique. It estimates the link between dependent (target) and independent variable (predictor). Logistic regression model is useful in probabilistic modelling. Logistic regression categorises the result in two categories – positive and negative. The categories can also be said as 0 or 1. Basically it generates results in binary format. So, the target label should be discrete (Kleinbaum et al., 2002).

4.3 Decision tree

Decision tree or choice tree is one of the most broadly utilised easy methods which is also used as a building block of several complex algorithm. It contains conditional control statements. Choice tree is basically a tree like structure which resembles a flow chart. A node within the tree corresponds to a feature in an instance to be classified. In a call tree, the category label is portrayed by terminal nodes. The inner node, representing nodes except leaf nodes, has some condition on some attributes and checkS the condition for attributes of data to tell apart them. Some of the decision tree induction algorithms are CART, ID3, and C4.5. From the same set of attributes, nearly exponential decision tree can be built. There are many optimal greedy algorithms for reasonably accurate decision tree construction (Navada et al., 2011).

4.4 Naive Bayes classifier

It is a parametric classifier that is often used when we need quick results, even if results are not the most accurate. It's called naive because the assumptions we make in our prior are not based on the contents of our data. We just assume that the data has a certain structure; we assume that every feature of our samples follows a Gaussian distribution (Rish et al., 2001).

4.5 K-nearest neighbour

It is a non-parametric algorithm called k-nearest neighbours, or KNN. Here, the letter k at the start refers to a number and we can pick any integer that is 1 or larger. KNN is fast to train because all it does, is save a copy of every incoming sample into a database. The interesting part comes when training is complete, and a new sample arrives to be categorised. KNN is a non-parametric (it means that it does not make any assumptions on the underlying data distribution) method used for both classification and regression, the input consists of k closest training examples in feature space. In other words, the category of the k nearest neighbour instances is utilised to observe the classification call of any instance (Zhang, 2016).

4.6 *Support vector machine*

Support vector machines or SVMs are one of the most popular supervised learning algorithms and are used for classification and regression problems. However, it is primarily used for machine learning classification problems. SVM selects extrema/vectors to help create hyperplanes (Susto et al., 2014). These extreme cases are called support vectors and the algorithm is called a SVM. Consider the figure, where two different categories are classified using decision boundaries or hyperplanes.

4.7 *Long short-term memory*

LSTM stands for long short-term memory. LSTM is an advanced version of recurrent neural network (RNN). A RNN is a type of neural network in which the previous step's output is fed as input for the current step. In some cases we require past data for better prediction, this is the case when RNN comes into the picture. In RNN there is a hidden layer. Hidden state, which recalls some details about a sequence, is the leading and most significant feature of RNN (Mathew et al., 2017).

But there is a drawback of RNN, i.e., the vanishing gradient problem. The vanishing gradient problem is encountered in machine learning while training gradient-based techniques in neural networks (for example, back propagation). This problem makes it very difficult to learn the parameters of the initial layers in the network. One example of unstable behaviour that you may encounter while training a deep neural network is the vanishing gradient problem. It shows the condition where it is difficult for a deep multilayer feed-forward network or a RNN to transmit useful gradient information from the model's output back to the layers near the model's input (Thesia et al., 2022). The effect is the general failure to learn on a given dataset of models with models with many layers to converge to a weak solution prematurely (Hu et al., 2022). LSTM and GRU is a solution to this problem.

LSTM networks are a type of RNN that makes it simpler to include data from memory. The problem of vanishing gradients in RNNs is fixed here (Choudhary and Sharma, 2023). Time series data can be recognised, processed and predicted using LSTM. Back-propagation is used to train the model. In an LSTM cell, there are three gates, i.e., input gate, forget gate and output gate. Input gate figures out which input value should be used to change the memory (Zhang et al., 2021). Forget gate discovers what information from the block should be discarded. Output gate block's input and memory are used to determine the output. LSTM cell is shown in Figure 6. Because of its ability to keep long time series dependencies in memory, LSTM networks proved to be a winning choice for predictive maintenance. LSTM gates outputs are sigmoid function, i.e., output a value between 0 or 1. '0' means blocking and '1' means allowing to pass through gates (Figure 7).

Figure 5 Unrolled neural network (see online version for colours)

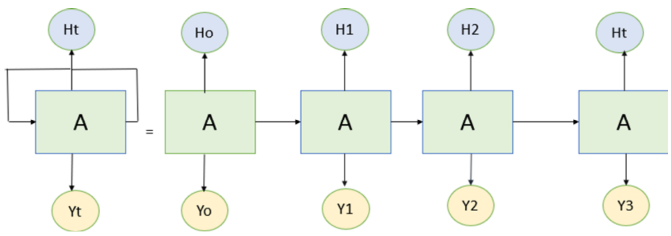


Figure 6 LSTM cell (see online version for colours)

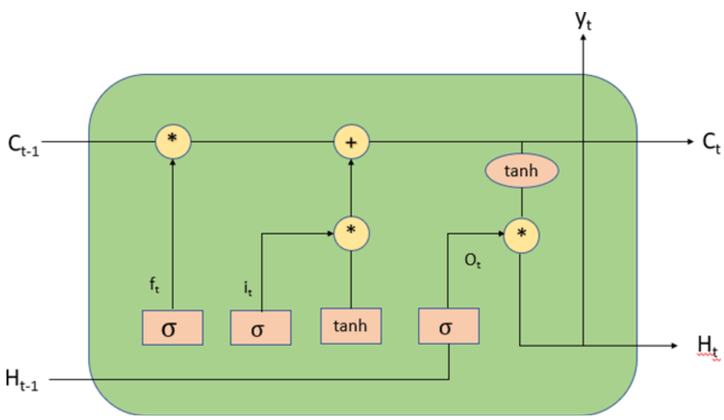
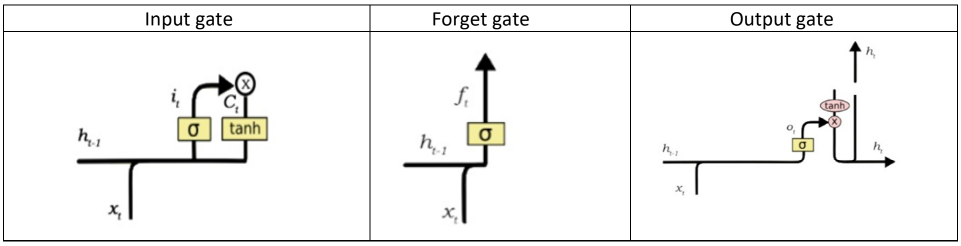


Figure 7 LSTM gates (see online version for colours)



These gates are ruled by the following equations:

$$i_t = \sigma\left(w_i\left[h_{t-1}, x_t\right] + b_i\right) \quad (1)$$

$$f_t = \sigma\left(w_f\left[h_{t-1}, x_t\right] + b_f\right) \quad (2)$$

$$o_t = \sigma\left(w_o\left[h_{t-1}, x_t\right] + b_o\right) \quad (3)$$

Here i_t, f_t, o_t, σ, w_x, h_{t-1}, x_t and b_x represent input gate, forget gate, output gate, sigmoid function, weight for respective gate(x) neurons, output of previous LSTM block at time $t - 1$, input at current time stamp and biases for the respective gates(x) respectively.

Equation (1) is for input gate that depicts how the new information will be stored in the cell state. Equation (2) is for forget gate and depicts the information that is going to be discarded from the cell state and equation (3) is for output gate that provides activation of the final output of LSTM block at timestamp ‘t’.

$$c_t' = \tanh\left(w_c\left[h_{t-1}, x_t\right] + b_c\right) \quad (4)$$

$$c_t = f_t \times c_{t-1} + t_t * c_t' \quad (5)$$

$$h_t = o_t * \tanh\left(c^t\right) \quad (6)$$

c'_t and c_t represents candidate for cell state and cell state at timestamp (t) respectively. The above equations, equation (4) and equation (5) tell what to remember from the current timestamp and what to forget from the last cell state.

Machine learning has a number of algorithms/models for pattern recognition and intrinsic structure in input dataset. Machine learning algorithms can be linear or nonlinear and also supervised and unsupervised learning. ML algorithms are majorly problem-based categories classification problem and regression problem. Some machine learning algorithms such as random forest, decision tree, naïve Bayes, logistic regression, SVM, KNN, and LSTM are discussed above for the designing of predictive maintenance model. The dataset used for predictive maintenance of railway track is obtained from RAS track geometry analytics (2015), which is a series of historical inspection and traffic records. This paper is only concerned about three defects, i.e., surface defects, cross level and DIP, these defects can be detected by the continuous data analysis of current point data and entire history sequence of data. For accurate and genuine maintenance schedule prediction, LSTM works better because it has a memory to store history data for long time. Along with a long-lasting memory as compared to the RNN, LSTM will accurately find the performance measure of prediction maintenance as compared to other machine learning algorithm.

5 Experimental result

In this section, the predicted outputs of the railway tracks dataset are compared with their actual outputs and the comparison graph is shown.

5.1 Confusion matrix

A confusion matrix is a tabular structure that is used to explaining the performance of a classification model (or ‘classifier’) on a set of test data for which the result is known. A confusion matrix of size ‘m × m’ related to a classifier shows the expected and actual classification, where m is the number of various categories. Table 2 shows a confusion matrix for ‘m’ = two.

Table 2 Confusion matrix

	Predicted negative	*Predicted positive*
Actual negative	a	b
Actual positive	c	d

In the table, variable ‘a’ is known as ‘True negative’, variable ‘b’ is known as ‘False positive’, variable ‘c’ is known as ‘False negative’ and variable ‘d’ is known as ‘True positive’.

5.2 ROC curve

The ROC curve shows the performance of a classifier. It is known as receiver operating characteristic curve. This is basically a graph. Its x-axis denotes false positive rate (FAR) and y-axis denotes true positive rate (TPR). It explains the accuracy or performance of a graph at various threshold. It is found by the area under the curve (AUC). If the area of the plot is more in upper region where it cuts a line passing at 45 degrees from the coordinate (0, 0), then the model is better and if area is lesser than model is not good but bad one. At coordinate (1, 1), threshold is zero. The confusion matrix and roc curve for logistic regression, naive Bayes, random forest, KNN, decision tree and LSTM are shown in Figures 8, 9, 10, 11, 12, 13 and 14.

5.3 Accuracy

TP is true positive, *TN* is true negative, *FP* is false positive and *FN* is false negative.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

5.4 Precision

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Predictive Positive}} \tag{9}$$

5.5 Recall

$$\text{Recall} = \frac{TP}{TP + FN} \tag{10}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Actual Positive}} \tag{11}$$

5.6 F1 score

$$\text{F1 Score} = \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

Comparison among various models for total of 741 data points is shown in Table 3.

Table 3 Comparison between different models for railway track dataset

Model	*Correct output*	*Accuracy*	*Precision*	*Recall*	*F1 score*
Logistic regression	581	0.78	0.96	0.72	0.82
KNN	580	0.78	0.89	0.75	0.81
Naive Bayes classifier	561	0.75	0.99	0.69	0.81
Random forest	728	0.98	0.99	0.97	0.98
Decision tree	733	0.984	0.99	0.98	0.99
LSTM	737	0.987	1.0	0.99	1.0

From the comparison Table 3 of all models, it is evident that LSTM is best. The F1 score of LSTM model is 1. The F1 score is related to precision and recall. If F1 score is higher than it means to improve the performance if we want to increase precision then it will not demotivate recall and vice versa. It is basically harmonic mean of both of them. Here, the LSTM's F1 score is 1 which is best. The accuracy of LSTM is 0.987 which is more than others. LSTM will be useful in predictive maintenance of railway tracks. The decision tree is almost same to LSTM. The result can vary if we will have a large dataset and other factors or attributes of track is provided.

Figure 8 Naive Bayes (see online version for colours)

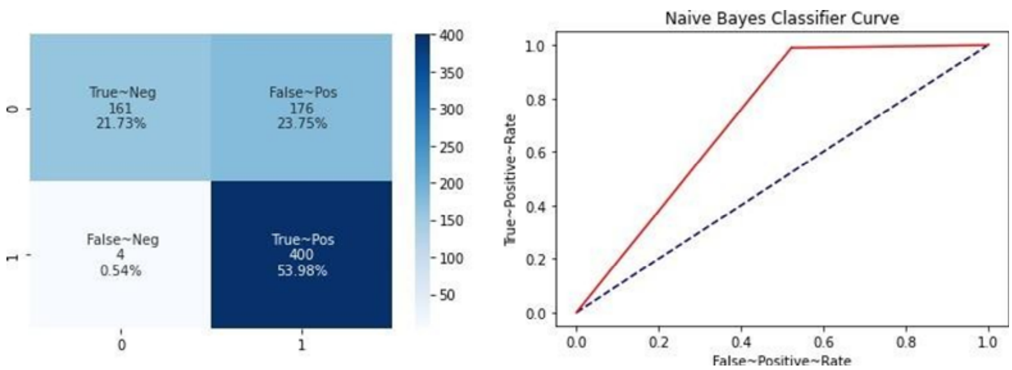


Figure 9 Logistic regression (see online version for colours)

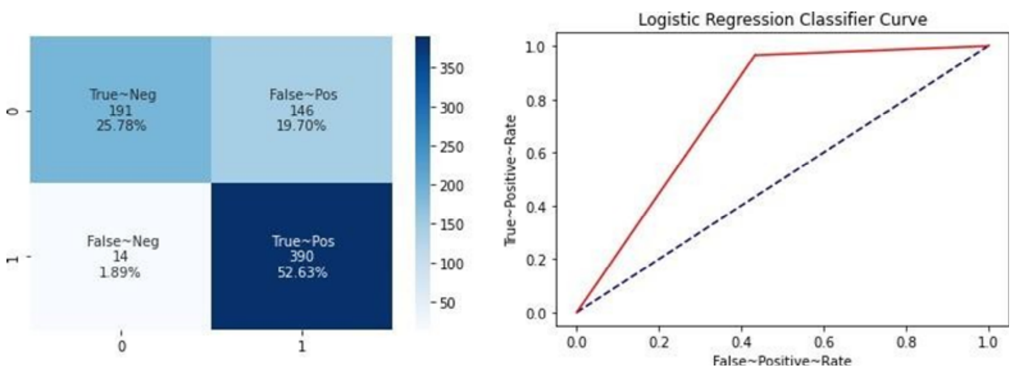


Figure 10 KNN (see online version for colours)

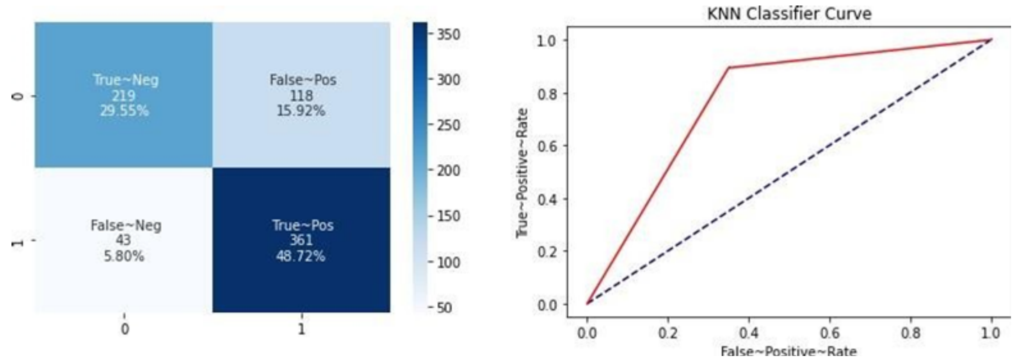


Figure 11 SVM (see online version for colours)

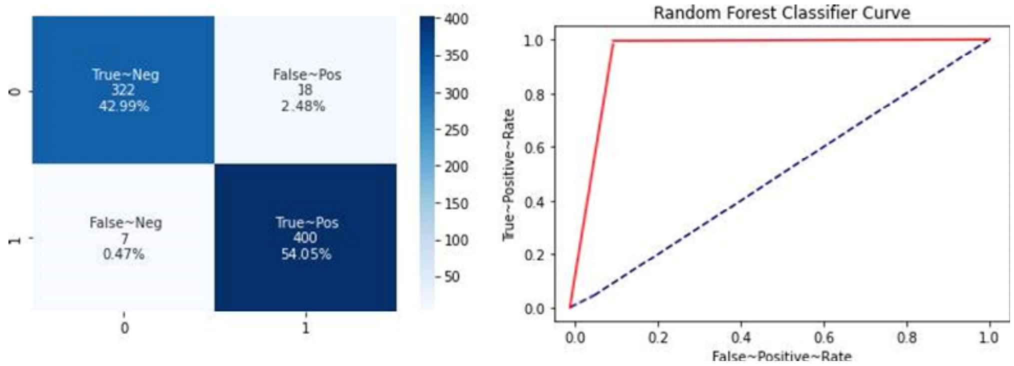


Figure 12 Random forest (see online version for colours)

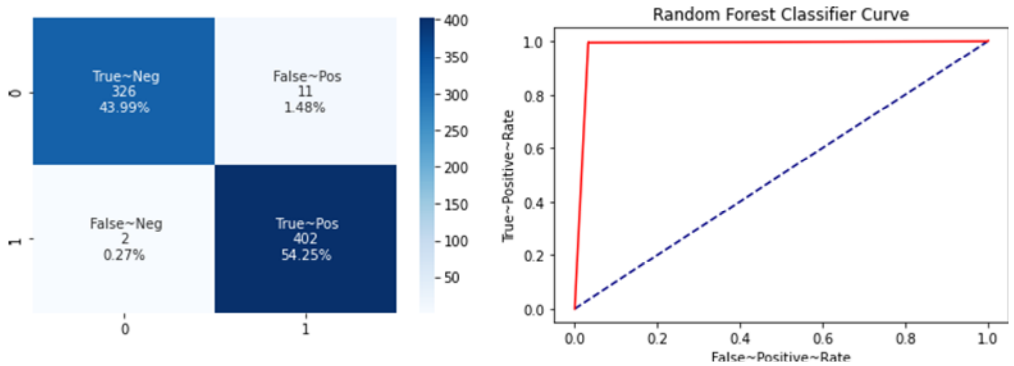


Figure 13 Decision tree (see online version for colours)

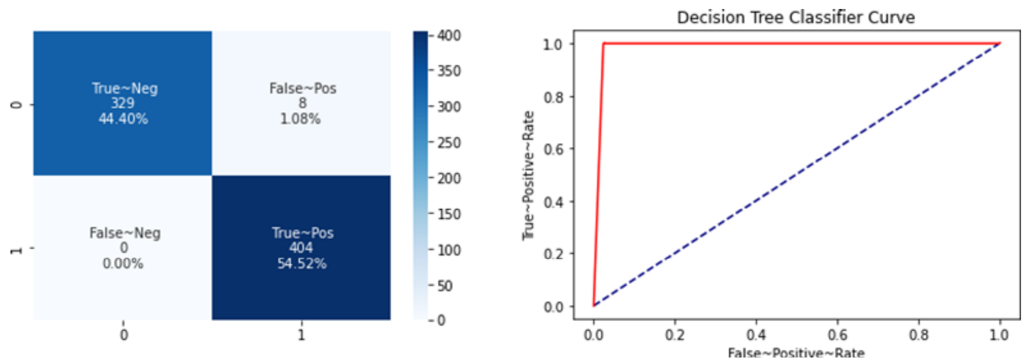


Figure 14 LSTM (see online version for colours)

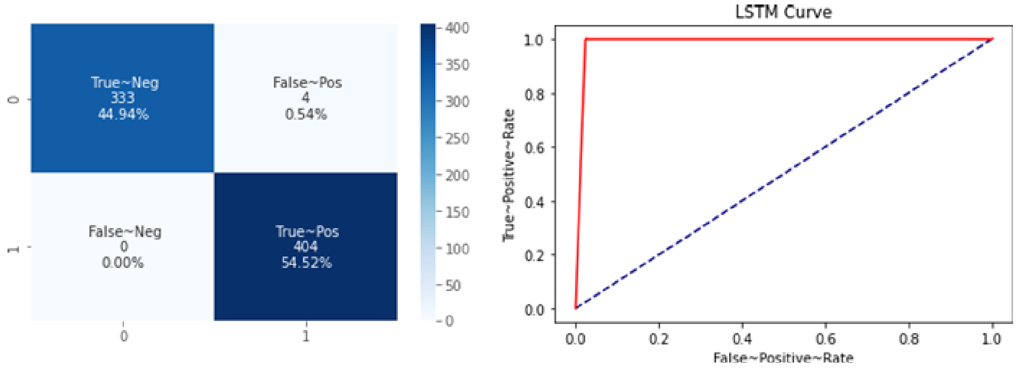


Table 4 Hypothesis data for different machine learning techniques

Model	*Validation*	*Log reg.*	*KNN*	*Naive Bayes*	*SVM*	*Random forest*	*Decision tree*	*LSTM*
1	25% data	78.0	77.2	75.5	91.1	98.0	98.4	98.7
2	30% data	77.5	77.6	76.9	95.0	98.2	96.9	98.8
3	35% data	79.2	79.2	70.6	92.3	97.1	98.3	97.8
4	40% data	74.3	81.1	72.45	89.7	96.5	98.1	99.0
5	45% data	80.0	76.2	78.9	92.5	94.8	97.8	98.5
6	50% data	76.8	73.5	73.5	90.2	98.2	98.2	98.8
8	5-folds cross validation	72.5	73.5	72.0	89.3	98.0	98.4	98.7
6	10-folds cross validation	78.6	74.8	78.9	91.0	98.1	98.7	98.9
9	15-folds cross validation	81.2	76.7	79.3	91.8	98.2	98.8	99.0

6 Hypothesis testing

In statistical analysis, sampling is used to derive inference about the dataset analysis from the sample. The sampling sometimes may lead to wrong conclusion of inferencing. So, inference derived from the sample goes through the statistical hypothesis testing process for ascertaining the confidence level of inference. Statistical hypothesis testing involves setting up a null hypothesis and an alternative hypothesis. Null hypothesis assumes there are significant performance differences between LSTM model and existing machine learning models, and an alternative hypothesis assumes existing machine learning performing is not better than LSTM. K-fold cross-validation is a evaluation technique for predictive models. The data set is divided into k subsets of equal size which are called folds. The model is trained and evaluated k times, each time using a different fold as the validation set. The performance metrics from each class are averaged to estimate the generalisation performance of the model. This method helps in model evaluation, meta-parameter selection and tuning and provides a more reliable measure of model effectiveness.

Here, we are using hypothesis testing to compare different machine learning algorithms. The hypothesis testing is done on railway dataset. Our hypothesis is that the LSTM is not better than the other models and will perform hypothesis testing based on data shown in Table 4.

The t-statistic is a measure of the difference between two sets expressed in units of standard error. A p-value is a measure of the probability of an observation lying at extreme t-values; therefore a low p-value also implies 'significance'.

Paired t-test is used to test our hypothesis. On the basis of t-value and p-value we will decide which hypothesis to accept or reject. Parameters for hypothesis testing are shown in Table 5.

Table 5 Parameters for hypothesis testing

Parameter	*Value*
Null hypothesis (H_o)	LSTM is not better than other models
Alternative hypothesis (H_1)	LSTM is better than other models
Level of significance	5% or 0.05

The formula for computing the t-value is:

$$t\text{-}value = \frac{mean1 - mean2}{\dfrac{(n1-1) * \operatorname{var}1^2 + (n2-1) * \operatorname{var}2^2}{n1+n2-2} * \sqrt{\dfrac{1}{n1} + \dfrac{1}{n2}}} \tag{13}$$

In equation (13), $n1$ and $n2$ represent the number of records in each sample, *mean*1 and *mean2* represent the mean of each model of selected pair, var1 and var2 represents the variance of each model of selected pair. $n1 + n2 - 2$ is the degrees of freedom.

In the results, it is clear that LSTM shows the best results among all other techniques so we are comparing LSTM with other models in this paired t-test.

Table 6 t-value and p-value of different models

Model	*t-value*	*p-value*
Logistic regression	35.2	1.2e-7
KNN	40.6	2.09311235e-8
Naive Bayes	30.3	4.1e-8
SVM	22.3	2.134e-7
Random forest	18.9	1.5e-9
Decision tree	12.7	1.12563e-10

The significance value is 0.05. It means that if the p-value is less than 0.05 then we reject the null hypothesis and go with the alternative hypothesis and vice-versa. We have used the MLxtend library to perform hypothesis testing. The results are shown in Table 6. All the p-values are smaller than the significance value (0.05) so we can reject the null hypothesis and go with the alternate hypothesis, i.e., LSTM is better than other models.

7 Conclusions

Railway transportation is the biggest infrastructure which mostly relies on railway tracks. This means that any defect in tracks may cause accidents. So, it is very prominent work to maintain rail tracks to avoid any such occurrence. By inspecting tracks to detect any defect or crack, one can schedule component remaining useful life (RUL), which improves operational reliability and cost effectiveness. Predictive maintenance applied on railway tracks using numerous machine learning algorithms such as KNN, decision tree, random forest, SVM, naïve Bayes, and LSTM can judge the condition of the track and schedule maintenance. As per Table 3 LSTM performs better as compared to other machine learning algorithms for criteria such as accuracy, precision, recall, and F1 score. Further, hypothesis testing uses K-fold cross validation to estimate the accuracy of the machine learning model on railway dataset. Hypothesis decision is based on t-value and p-value, with the significance value 0.05. As in Table 6, all p-values are smaller than the reliability and accuracy of LSTM track maintenance model. By predicting the actual condition and RUL of railway track, authorities can schedule required maintenance before breakdown of track, or broadcast a track fault warning to the nearest or responsible operators.

References

Andrade, A.R. and Teixeira, P.F. (2012) 'A Bayesian model to assess rail track geometry degradation through its life-cycle', *Research in Transportation Economics*, Vol. 36, No. 1, pp.1–8.

Ayvaz, S. and Alpay, K. (2021) 'Predictive maintenance system for production lines in manufacturing: a machine learning approach using IoT data in real-time', *Expert Systems with Applications*, July, Vol. 173, No. C, p.114598.

Biau, G. and Scornet, E. (2016) 'A random forest guided tour', *Test*, Vol. 25, No. 2, pp.197–227.

Bukhsh, Z.A. et al. (2019) 'Predictive maintenance using tree-based classification techniques: a case of railway switches', *Transportation Research Part C: Emerging Technologies*, Vol. 101, pp.35–54.

Cai, S. et al. (2022) 'A hybrid parallel deep learning model for efficient intrusion detection based on metric learning', *Connection Science*, Vol. 34, No. 1, pp.551–577.

Choudhary, S. and Sharma, R. (2023) 'CNN-based battlefield classification and camouflage texture generation for real environment', *International Journal of Computational Science and Engineering*, Vol. 26, No. 3, pp.231–242.

Faiz, R.B. and Singh, S. (2009) 'Time based predictive maintenance management of UK rail track', *2009 International Conference on Computing, Engineering and Information*, IEEE.

Gerum, P.C.L., Altay, A. and Baykal-Gürsoy, M. (2019) 'Data-driven predictive maintenance scheduling policies for railways', *Transportation Research Part C: Emerging Technologies*, Vol. 107, pp.137–154.

Hashemian, H.M. (2010) 'State-of-the-art predictive maintenance techniques', *IEEE Transactions on Instrumentation and Measurement*, Vol. 60, No. 1, pp.226–236.

He, Q., Li, H., Bhattacharjya, D., Parikh, D.P. and Hampapur, A. (2015) 'Track geometry defect rectification based on track deterioration modelling and derailment risk assessment', *Journal of the Operational Research Society*, Vol. 66, No. 3, pp.392–404.

Hu, N. et al. (2022) 'Graph learning-based spatial-temporal graph convolutional neural networks for traffic forecasting', *Connection Science*, Vol. 34, No. 1, pp.429–448.

Karimpour, M., Hitihamillage, L., Elkhoury, N., Moridpour, S. and Hesami, R. (2018) 'Fuzzy approach in rail track degradation prediction', *Journal of Advanced Transportation*, Vol. 2018, Article ID 3096190, 7pp.

Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M. and Klein, M. (2002) 'Logistic regression', *A Self-Learning Tekst*, Springer.

Kocbek, S. and Gabrys, B. (2019) 'Automated machine learning techniques in prognostics of railway track defects, in *2019 International Conference on Data Mining Workshops (ICDMW)*, IEEE, pp.777–784.

Kou, K. et al. (2021) 'A location privacy protection algorithm based on differential privacy in sensor network', *International Journal of Embedded Systems*, Vol. 14, No. 5, pp.432–442.

Li, H. et al. (2014) 'Improving rail network velocity: a machine learning approach to predictive maintenance', *Transportation Research Part C: Emerging Technologies*, Vol. 45, pp.17–26.

Liu, W. et al. (2021) 'Using machine learning techniques for DSP software performance prediction at source code level', *Connection Science*, Vol. 33, No. 1, pp.26–41.

Mathew, V., Toby, T., Singh, V., Rao, B.M. and Kumar, M.G. (2017) 'Prediction of remaining useful lifetime (RUL) of turbofan engine using machine learning', *2017 IEEE International Conference on Circuits and Systems (ICCS)*, IEEE.

Mobley, R.K. (2002) *An Introduction to Predictive Maintenance*, 2nd ed., Elsevier Science, USA.

Navada, A., Ansari, A.N., Patil, S. and Sonkamble, B.A. (2011) 'Overview of use of decision tree algorithms in machine learning, in *2011 IEEE Control and System Graduate Research Colloquium*, IEEE, pp.37–42.

Prytz, R., Kowalczyk, S., Rögnvaldsson, T. and Byttner, S. (2015) 'Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data', *Engineering Applications of Artificial Intelligence*, Vol. 41, pp.139–150.

RAS Track Geometry Analytics (2015) [online] https://www.informs.org/content/download/302130/2890480/RAS (accessed 12 June 2021).

Rish, I. et al. (2001) 'An empirical study of the naive Bayes classifier', in: *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, Vol. 3, pp.41–46.

Shafahi, Y., Masoudi, P. and Hakhamaneshi, R. (2008) 'Track degradation prediction models, using markov chain, artificial neural and neuro-fuzzy network', in *8th World Congress on Railway Research*, pp. 1–9.

Susto, G.A. et al. (2014) 'Machine learning for predictive maintenance: a multiple classifier approach', *IEEE Transactions on Industrial Informatics*, Vol. 11, No. 3, pp.812–820.

Thesia, Y., Vidhey, O. and Priyank, T. (2022) 'Predicting stock price movement using a stack of multi-sized filter maps and convolutional neural networks', *International Journal of Computational Science and Engineering*, Vol. 25, No. 1, pp.22–33.

Vale, C. and Lurdes, S.M. (2013) 'Stochastic model for the geometrical rail track degradation process in the Portuguese railway northern line', *Reliability Engineering & System Safety*, Vol. 116, pp.91–98.

Wang, K-S., Li, Z., Braaten, J. and Yu, Q. (2015) 'Interpretation and compensation of backlash error data in machine centers for intelligent predictive maintenance using ANNs', *Advances in Manufacturing*, Vol. 3, No. 2, pp.97–104.

Xie, J. et al. (2020) 'Systematic literature review on data-driven models for predictive maintenance of railway track: implications in geotechnical engineering', *Geosciences*, Vol. 10, No. 11, p.425.

Yang, X. et al. (2021) 'Service component recommendation based on LSTM', *International Journal of Embedded Systems*, Vol. 14, No. 2, pp.201–209.

Zhang, Z. (2016) 'Introduction to machine learning: k-nearest neighbors', *Annals of Translational Medicine*, Vol. 4, No. 11, p.218.

Zhang, S. et al. (2021) 'A novel method of mental fatigue detection based on CNN and LSTM', *International Journal of Computational Science and Engineering*, Vol. 24, No. 3, pp.290–300.