# Machine Learning Based Bridge Condition Analysis

**UG Project Report submitted in the partial fulfillment for the award of**

**……………………….**

**By**

Aryan Tyagi

(20065025)

**Under the guidance of**

Dr. Sasankasekhar Mandal

**Department of Civil Engineering**

**INDIAN INSTITUTE OF TECHNOLOGY**

**(BANARAS HINDU UNIVERSITY)**

**VARANASI – 221 005**

November, 2023

# THESIS CERTIFICATE

This is to certify that the thesis titled **Machine Learning Based bridge Condition Analysis**, submitted by **Aryan Tyagi** (Roll No. 20065025), to the Indian Institute of Technology (Banaras Hindu University), Varanasi, in the partial fulfillment for the award of _________, is a bona fide record of work done by him/her under my/our supervision. It is certified that the statement made by the student in his/her declaration is correct to the best of my/our knowledge.

Date of Submission: 22nd November 2023

Dr. Sasankasekhar Mandal

Professor

Department of Civil Engineering

IIT (BHU), Varanasi – 221 005

# DECLARATION BY THE CANDIDATE

I, **Aryan Tyagi**, certify that the work embodied in this thesis is my own bona fide work carried out by me under the supervisions of **Dr. Sasanksekhar Mandal**, from July 2023 to November 2023 at the Department of Civil Engineering, Indian Institute of Technology (BHU), Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and/or cited to the researchers wherever their works have been utilized in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, etc., reported in journals, books, magazines, reports dissertation, thesis, etc., or available at websites.

.

Date: 22th November 2023 — Aryan Tyagi

Place: Varanasi — (Aryan Tyagi)

# ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who have supported me throughout the course of this research project.

First and foremost, I would like to thank my supervisor Dr. Sasankasekhar Mandal, for providing me with invaluable guidance, feedback, and encouragement throughout the research process. Their expertise and insight have been instrumental in shaping this thesis.

I would also like to thank Indian Institute of Technology BHU, Varanasi for their support in providing data, case studies, and other resources that have contributed to the research findings.

I am grateful to the participants who generously gave their time and expertise to participate in interviews and provide valuable insights into the challenges and opportunities associated with Machine Learning Based Bridge Condition Analysis.

# ABSTRACT

Bridges play a vital part in the transportation system by ensuring the connectedness of transportation systems, which is critical for a country's social and economic prosperity by offering daily mobility to the people. A comparison of the failure cases is made with those from available literature and observed some unique failure patterns of bridges.

More than 2130 bridges (excluding culverts and pedestrian bridges) have failed to provide intended service or got collapsed during various phases of construction in recent four decades.

Reasons for failure of bridges during their service life as well as during the construction phase has been analyzed by employing a mixed-methods approach, including FEA (Finite Element Analysis) for identifying potential areas of weakness. A comprehensive literature review is conducted to identify the current state of knowledge on Bailey Bridge failures. Case studies of Bailey Bridge failures were conducted to identify common patterns and causes of failure. The research findings indicated that the main causes of Bailey Bridge failures are

The results of this thesis can help bridge owners, engineers, and policymakers to better understand the challenges and opportunities associated with Bridges to take proactive measures to prevent failures.

# TABLE OF CONTENTS

## CONTENTS

# CHAPTER 1: INTRODUCTION

Bridges, as essential components of transportation infrastructure, are critical for economic growth, regional connectivity, and societal development. These engineering marvels span land and water, connecting people, goods, and services across vast landscapes. However, the aging and deterioration of bridges pose significant challenges for infrastructure authorities worldwide. Traditionally, bridge condition assessment in India, as in many other countries, has relied heavily on manual inspections and outdated data collection methods. These approaches are often costly, time-consuming, and have their limitations. To address these challenges and meet the demands of a rapidly growing nation, the exploration of machine learning and artificial intelligence (ML/AI) techniques for bridge condition analysis is becoming increasingly crucial.

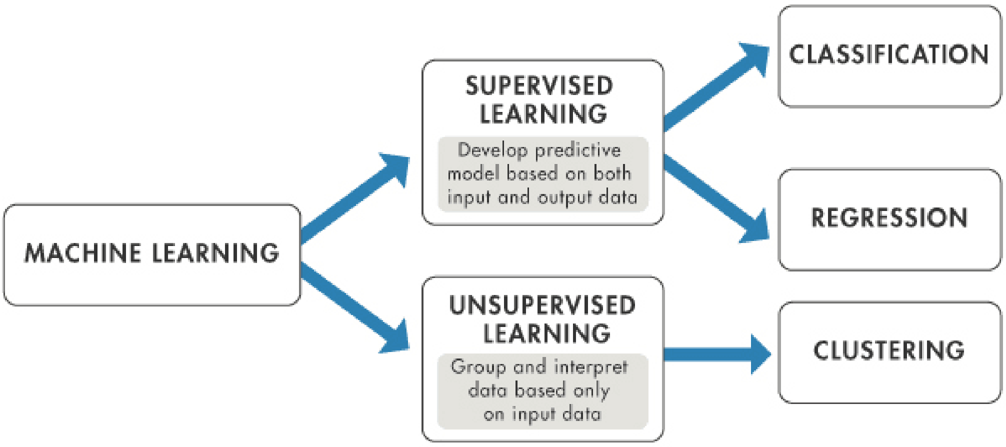
We highlight how machine learning can harness vast datasets generated from sensor networks, visual inspections, and other sources to optimize maintenance scheduling, detect early defects, and enable real-time monitoring. The application of machine learning techniques in bridge condition analysis promises to offer a cost-effective and data-driven approach that can help ensure the longevity and safety of India's vital bridge infrastructure. By harnessing data from an array of sources, including sensor networks, visual inspections, historical records, and environmental data, machine learning models can uncover hidden insights, patterns, and correlations.

The development of a computational data-driven, autonomous, and efficient bridge condition prediction machine learning model is crucial for improving bridge maintenance decision-making. This advanced model will enable proactive identification and mitigation of potential bridge issues, thereby enhancing bridge safety and reliability while minimizing resource expenditure and improving the overall transportation infrastructure system.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 INTRODUCTION TO MACHINE LEARNING

Machine learning is the science of empowering computers to learn from data and make intelligent decisions, all without explicit programming. In an era characterized by an explosion of data, machine learning is not just a technological innovation but a paradigm shift that has the potential to revolutionize numerous domains, from healthcare to finance, from transportation to entertainment. The driving force behind machine learning's ascent is its ability to unlock patterns, insights, and predictions that were previously hidden within vast and complex datasets. Rather than relying on explicit instructions, machine learning algorithms can generalize from historical data to make informed choices, recognize images, understand human language, and even predict future events.




In since we are classifying health of bridges, it makes it a supervised classification problem in machine learning.

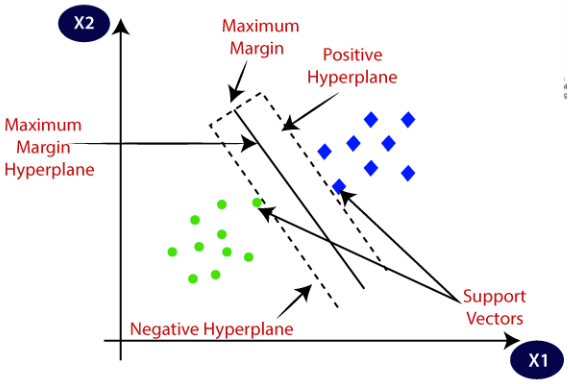
## 2.1 MACHINE LEARNING MODELS

The Machine Learning models that will be used in this study are mainly

- 2.1.1 Support Vector Machines (SVMs)
- 2.1.2 Artificial Neural Networks (ANNs)
- 2.1.3 Random Forests

### *2.1.1 Support Vector Machines (SVMs)*

Support Vector Machines (SVMs) are a type of machine learning algorithm that can be used for both classification and regression tasks. SVMs work by finding a hyperplane in feature space that separates the data into two classes. The hyperplane is chosen so that the margin between the two classes is as large as possible.

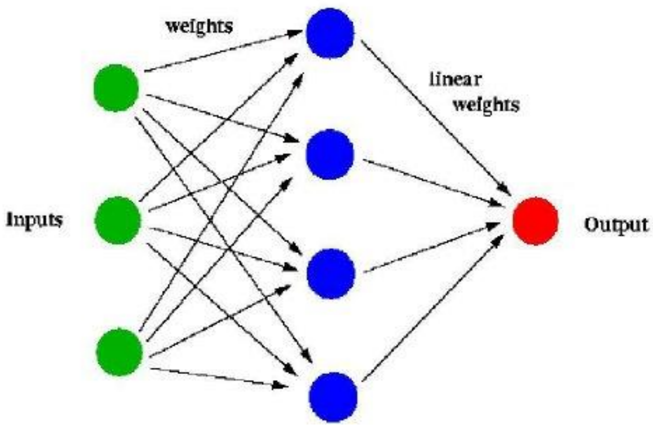



SVMs are a powerful machine learning algorithm that can be used for a variety of tasks, including:

- Image/Text/Data classification
- Fraud detection
- Medical diagnosis

### *2.1.2 Artificial Neural Networks (ANNs)*

Artificial Neural Networks (ANNs) are a type of machine learning algorithm that is inspired by the human brain. ANNs are made up of interconnected nodes, called neurons. Each neuron performs a simple mathematical operation, and the outputs of the neurons are connected to the inputs of other neurons.

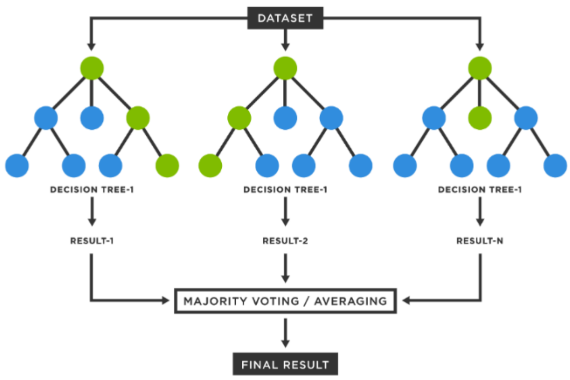



ANNs are trained by feeding them data and adjusting the weights of the connections between the neurons. The goal of training is to find a set of weights that allows the ANN to correctly classify or predict the output for a given input.

### *2.1.3 Random Forests*

In a Random Forest, multiple decision trees are constructed using subsets of the training data and features. The predictions from individual trees are then aggregated to make a final prediction.

This ensemble approach reduces overfitting and enhances the model's generalization capabilities. Random Forest is known for its ability to handle high-dimensional data and is widely used in applications such as disease prediction, image classification, and, in our case, bridge health assessment.

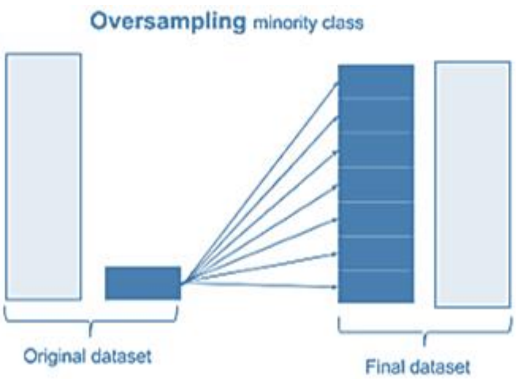



## 2.2 INTRODUCTION TO UNDER/OVER SAMPLING USING SMOTE

### *2.2.1 Under-Sampling:*

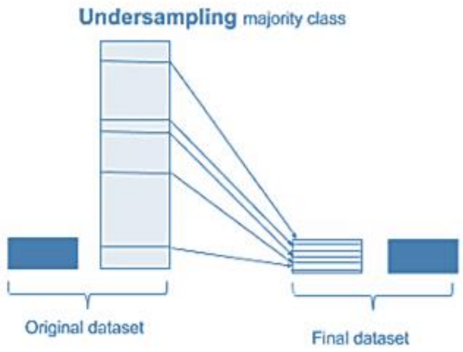
Under-sampling is a technique used to balance imbalanced datasets, typically in binary classification problems where one class has significantly fewer samples than the other. In under-sampling, we randomly remove a portion of the majority class samples to make the class distribution more balanced.

For example, if we have a dataset with 90% of samples belonging to Class A and only 10% to Class B, we can apply under-sampling by randomly selecting a subset of samples from Class A to match the number of samples in Class B. While under-sampling can balance the dataset, it also comes with a potential drawback.



### *2.2.2 Over-Sampling:*

Over-sampling, on the other hand, is a technique used to address class imbalance by increasing the number of samples in the minority class. This is typically done by duplicating or generating new samples in the minority class to match the number of samples in the majority class.




Continuing the same example, if you have a dataset with 90% of samples in Class A and 10% in Class B, you can apply over-sampling by replicating samples from Class B, creating a larger dataset with a balanced class distribution. While over-sampling helps address class imbalance, it may lead to overfitting if not applied carefully, as the model may become too sensitive to the minority class samples.

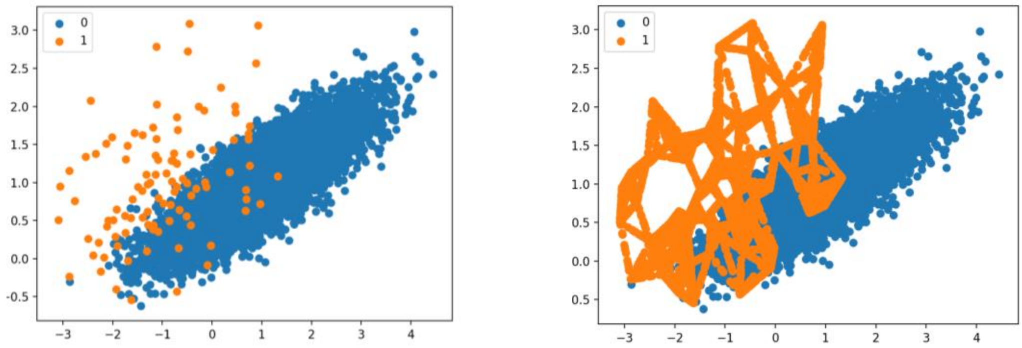
## 2.3 INTRODUCTION TO UNDER/OVER SAMPLING USING SMOTE

SMOTE: Synthetic Minority Oversampling Technique (SMOTE) is a method of data augmentation that creates new data points within the minority class of a dataset. This can be helpful when the dataset is imbalanced and the minority class is small. SMOTE works by creating new data points along the line segments between existing data points in the minority class. This creates new data points that are similar to the existing data points, but not identical. SMOTE can be used to improve the performance of machine learning models in a number of ways. For example, it can help to:

*Reduce bias*: SMOTE can help to reduce bias in machine learning models by creating a more balanced dataset.

*Reduce overfitting*: SMOTE can help to reduce overfitting by providing the model with more data to train on.

*Improve generalization*: SMOTE can help to improve generalization by providing the model with a wider range of data to train on.

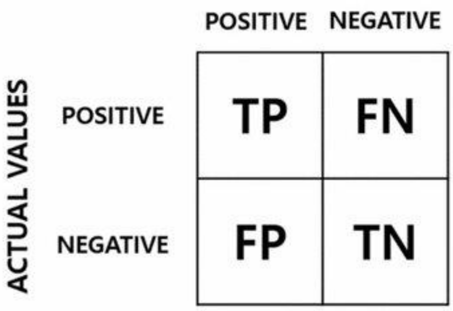


## 2.4 PERFORMANCE METRICS AND VALIDATION

To assess the effectiveness of the machine learning models, we employed accuracy, precision, recall, and F-score metrics. These metrics are derived from the confusion matrix, which summarizes the classification results based on the number of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). TP represents the correctly identified bridges in poor condition, while FN represents the bridges in poor condition that were misclassified. FP represents the bridges in good condition that were incorrectly classified as poor, and TN represents the correctly identified bridges in good condition.

### *2.4.1 Precision:*

Precision, also known as positive predictive value, measures the accuracy of the positive predictions made by a classifier. "Of all the instances that the model predicted as positive, how many were actually correct?"




### *2.4.2 Recall:*

Recall, also known as true positive rate or sensitivity, measures the model's ability to identify all relevant instances of the positive class. "Of all the actual positive instances, how many did the model correctly identify?"

***2.4.3 F1-Score:*** The F1-Score is a single metric that balances precision and recall. It is the harmonic mean of precision and recall and provides a more comprehensive assessment of a classifier's performance. The F1-Score is particularly useful when there is an imbalance between precision and recall, and you want to find a balance between avoiding false positives and false negatives.

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

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

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

***2.4.4 Support:*** The "support" in the context of classification metrics represents the number of actual occurrences of each class in the dataset. It can help provide context for the other metrics by indicating the distribution of the classes. For each class, support indicates how many instances belong to that class.

## 2.3 DATASET COLLECTION

This study utilizes the National Bridge Inventory (NBI) database, a comprehensive repository of information encompassing over 600,000 bridges across the United States, to generate bridge deterioration predictions. The NBI database, which is continuously expanding, provides detailed information on bridge location, classification, geometric characteristics, age, traffic data, structural characteristics, and constructional details for bridges on U.S. Highways, Interstate Highways, State and County roads, and publicly accessible bridges on Federal Lands. By employing this comprehensive and high-quality dataset, this study aims to generate accurate and reliable bridge deterioration predictions.

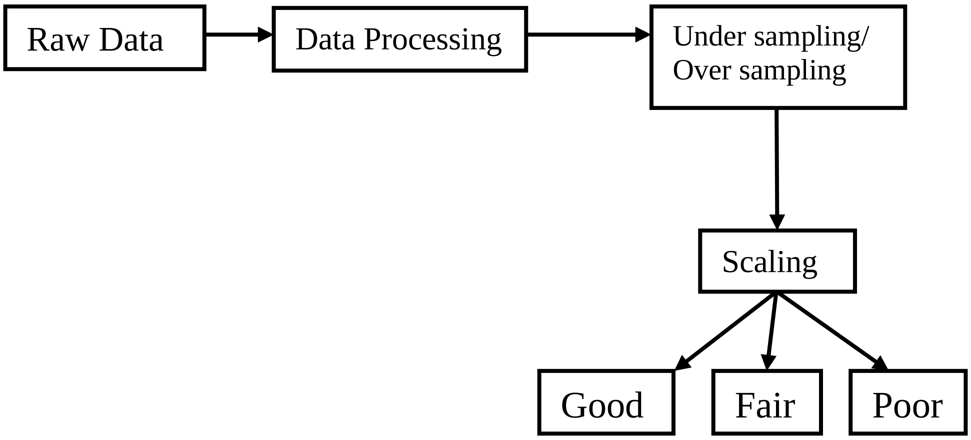
Due to limited computation power, only the bridges in the state of Georgia have been taken. The dataset contains features of bridges for years 1990 to 2022 with each year having around 10,500 datapoints and in total around 4,00,000 entries.

Also, since bridge health doesn’t change much in 1 to 2 years, data of intermediate years, for example, in between 1990 to 1995 only data of year 1992 has been taken. After applying these selections out total dataset has 1,32,000 datapoints in total till the year 2022.

As this dataset is highly imbalanced, we applied oversampling/under sampling techniques on the dataset as described above. Details of data-preprocessing is further discussed in the methodology section.

# CHAPTER 3: METHODLOGY

The figure shows the flowchart of the proposed framework.




`

The raw bridge data is prepared by performing data cleaning tasks such as dropping unnecessary features manually, and dropping features and data based on analyzing missing values.

## 3.1 DATA PRE-PROCESSING

### *3.1.1 Data Specification*

The initial NBI dataset contains 1,32,895 datapoints including the missing/NaN values which needed to be removed for the ML model. After removing the null values, remaining dataset contained 1,17,104 datapoints whose numeric and non-numeric features are listed below. The number of numeric features count to 30 and non-numeric features count to 11.

| NUMERIC COLUMNS |
|---|
| 'YEAR', '1 - STATE CODE', '3 - COUNTY CODE', '27 - YEAR BUILT', '29 - AVERAGE DAILY TRAFFIC', '45 - NUMBER OF SPANS IN MAIN UNIT', '49 - STRUCTURE LENGTH (FT.)', 'BRIDGE AGE (YR)', 'CAT29 - DECK AREA (SQ. FT.)', '17 - LONGITUDE (DECIMAL)', '16 - LATITUDE (DECIMAL)', '106 - YEAR RECONSTRUCTED', '34 - SKEW ANGLE (DEGREES)', '48 - LENGTH OF MAXIMUM SPAN (FT.)', '51 - BRIDGE ROADWAY WIDTH CURB TO CURB (FT.)', '91 - DESIGNATED INSPECTION FREQUENCY', '64 - OPERATING RATING (US TONS)', '66 - INVENTORY RATING (US TONS)', '30 - YEAR OF AVERAGE DAILY TRAFFIC', '109 - AVERAGE DAILY TRUCK TRAFFIC (PERCENT ADT)', '114 - FUTURE AVERAGE DAILY TRAFFIC', '115 - YEAR OF FUTURE AVERAGE DAILY TRAFFIC', '96 - TOTAL PROJECT COST', 'COMPUTED - AVERAGE DAILY TRUCK TRAFFIC (VOLUME)', 'AVERAGE RELATIVE HUMIDITY', 'AVERAGE TEMPERATURE', 'MAXIMUM TEMPERATURE', 'MINIMUM TEMPERATURE', 'MEAN WIND SPEED', 'CITY - INFOBRIDGE PLACE CODE' |

| NON-NUMERIC COLUMNS |
| --- |
| '1 - STATE NAME', '8 - STRUCTURE NUMBER', '2022 NBI STRUCTURE NUMBER', '22 - OWNER AGENCY', '3 - COUNTY NAME', '43A - MAIN SPAN MATERIAL', '43B - MAIN SPAN DESIGN', '58 - DECK CONDITION RATING', '59 - SUPERSTRUCTURE CONDITION RATING', '60 - SUBSTRUCTURE CONDITION RATING' |

For converting these non-numeric features into numeric ones, we use Label-Encoder from ScikitLearn library which encodes these non-numeric features into sparse categorical numerical features. Also, after removing some garbage features like structure-number, county-name, owner-agency, state-code we end up with 20 usable features relevant to this study.

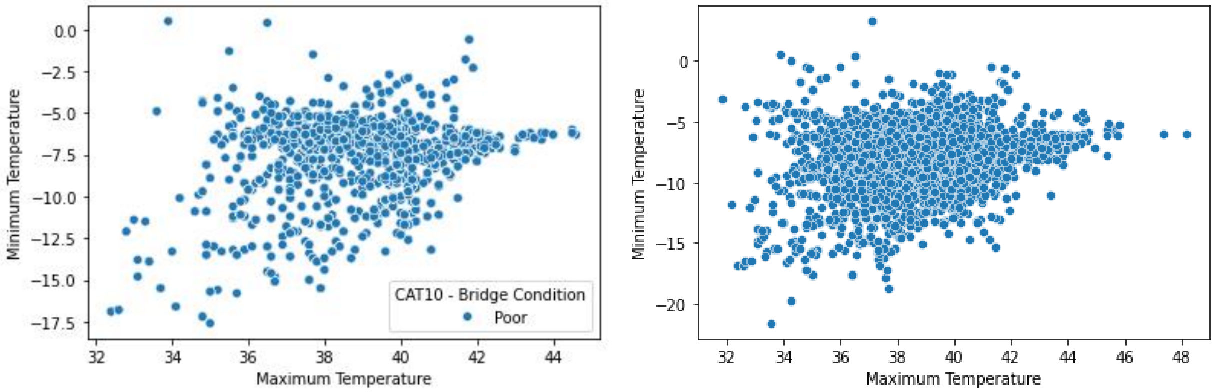
### *3.1.2 Training Data Over/Under sampling & Test Data*

The NBI dataset is highly imbalanced data in terms of bridge status distributions., since, the number of faulty bridges will be much lower compared to good/fair health bridges in a real-world scenario. ML-based methods encounter difficulties learning from imbalanced class distribution data and display bias results towards majority class and minority class samples are more likely to be misclassified.
The distribution of the dataset is given below in the table:

| Bridge Condition | Number of observations | Adjusted number of observations |
| --- | --- | --- |
| F(Fair) | 63,822 | 20,000 |
| G(Good) | 45,514 | 20,000 |
| P(Poor) | 7,768 | 10,000 |

Directly applying oversampling techniques to the minority class of the dataset to oversample up to the number of observations of F(Fair) or G(Good) classes may result in absurd over-estimations which may not help the model perform good on training and testing dataset.

s
After applying oversampling to the majority classes (Good, Fair), Oversampling techniques such as SVM-SMOTE is applied on the minority class. SVM-SMOTE focuses on generating new minority class instances near borderlines with SVM so as to help establish boundary between classes. The final dataset contains 10,000 “poor” health class samples.




(Figure shows oversampled datapoints of minority class (Poor) after SMOTE)

For the testing dataset, 20% train-test split was used with the final test dataset comprising of 23,421 samples with 54.27% as "good", 38.93% as "fair" and 6.54% as "poor" health samples which is similar to the distribution of original dataset.

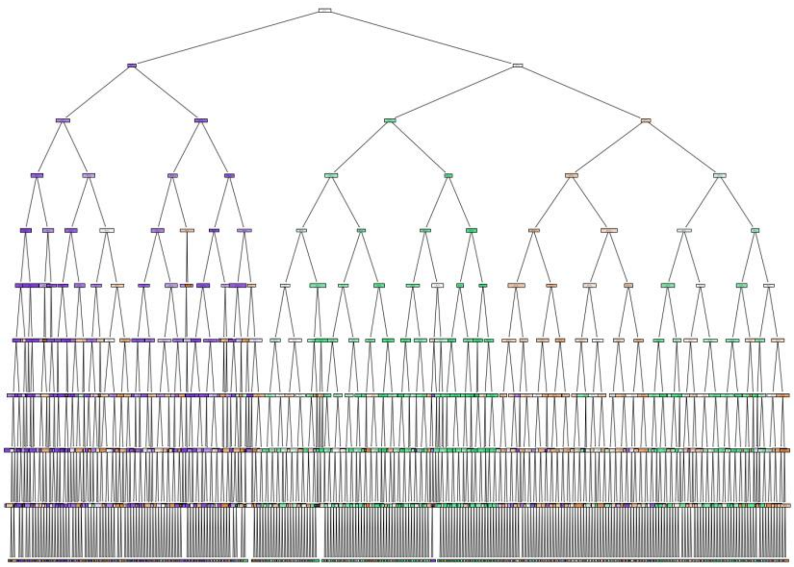
### *3.1.3 Experimental Design*

We applied 4 different Machine Learning classifiers for this task namely: Random Forest (RF), Support vector Machines (SVM), Artificial Neural Networks (ANN) and Logistic Regression.

For training these models, training data was used and testing data was used for evaluation purposes. Hyperparameter tuning was a trial-error method for finding the best set of hyperparameters.

### *3.1.4 Random Forest (RF)*

For the Random Forest model, 80 estimators with each having maximum depth of 500 nodes was chosen.

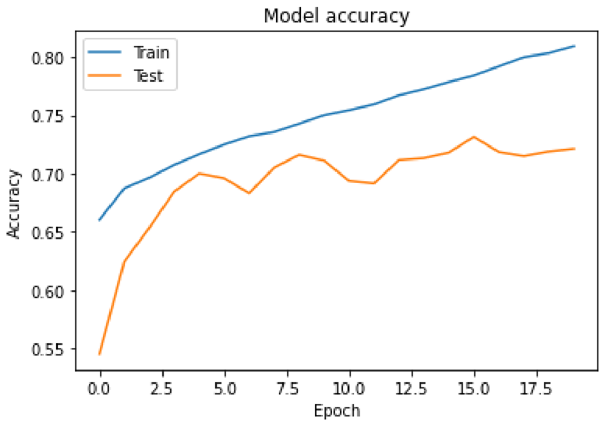


### *3.1.5 ANN*

For the Deep Neural Network, number of hidden layers were explored and best fit was total of 8 *dense layers* (including last output layer) and 1 Batch Normalize layer. The number of neurons followed an inverted pyramid structure with numbers ranging from 2056 to 128 to 3 (output layer). The best activation function was found to be *ReLU* for the hidden layer and since this is a multi-classification task output activation layer used is *Softmax*. Without standardizing the neural network struggles to converge to the optima thus, first layer used is the Batch Norm layer for standardizing the inputs.

```
Layer (type)                 Output Shape              Param #
=================================================================
batch_normalization (BatchN  (None, 19)                76
ormalization)

dense (Dense)                (None, 2056)              41120

dense_1 (Dense)              (None, 2056)              4229192

dense_2 (Dense)              (None, 1024)              2106368

dense_3 (Dense)              (None, 1024)              1049600

dense_4 (Dense)              (None, 512)               524800

dense_5 (Dense)              (None, 256)               131328

dense_6 (Dense)              (None, 128)               32896

dense_7 (Dense)              (None, 3)                 387

=================================================================
Total params: 8,115,767
Trainable params: 8,115,729
Non-trainable params: 38
```

Number of epochs were set to 30 with batch-size being 256. The optimizer used is *Adam optimizer* with learning rate as 0.0009 and decay as 0.0005. Particularly large learning rates tend to overshoot the optimum which fails to converge the optimization.




### *3.1.6 SVM & Logistic Regression (LR)*

For the SVM, no such adjustments had to be made and the gamma and c value were taken as 0.5 and 0.1.

# CHAPTER 4: RESULTS AND DISCUSSION

## 4.1 Random Forest

The final performance of the model on the test dataset having 23,421 samples was 78% overall accuracy with "poor" health class recall being 82%.

| Class | PRECISION | RECALL | F1-SCORE | SUPPORT |
|---|---|---|---|---|
| **0** *(Fair)* | 0.70 | 0.76 | 0.73 | 9118 |
| **1** *(Good)* | 0.85 | 0.79 | 0.82 | 12771 |
| **2** *(Poor)* | 0.76 | 0.82 | 0.79 | 1532 |
| **Accuracy** | | | 0.78 | 23421 |

## 4.2 ANN

Accuracies for ANN model is listed below in the table.

| Class | PRECISION | RECALL | F1-SCORE | SUPPORT |
|---|---|---|---|---|
| **0** (*Fair)* | 0.64 | 0.69 | 0.66 | 9118 |
| **1** *(Good)* | 0.80 | 0.74 | 0.77 | 12771 |
| **2** *(Poor)* | 0.67 | 0.77 | 0.72 | 1532 |
| **Accuracy** | | | 0.72 | 23421 |

## 4.3 SVM & Logistic Regression (LR)

The SVM model performed very poorly with it getting only 64% overall accuracy. Similarly, Logistic regression also performs very poorly with overall accuracy being only 62% on the testing dataset.

Final accuracies are listed in the table below: -

| Model | Accuracy (%) |
|---|---|
| *Random Forest* | 78 |
| *Deep Neural Network / ANN* | 72 |
| *SVM & Logistic Regression* | 62-64 |

# CHAPTER 5: CONCLUSION AND FUTURE WORK

In this thesis, we have explored the application of machine learning techniques to bridge condition analysis. We have presented a novel machine learning-based framework for bridge condition prediction using advanced machine learning algorithms on the National Bridge Inventory (NBI) dataset. Our results demonstrate that machine learning can be used to accurately predict bridge condition ratings, with an accuracy of over 78%+. We have also shown that machine learning can be used to identify critical bridges that require maintenance or repair.

Our work has several important implications for bridge management. First, our findings suggest that machine learning can be used to develop more efficient and cost-effective bridge inspection programs. Second, our work can be used to prioritize maintenance and repair activities, ensuring that resources are allocated to the bridges that need them most. Third, our work can be used to develop predictive models that can be used to anticipate future bridge failures.

Overall, our work demonstrates that machine learning has the potential to revolutionize bridge condition analysis. By using machine learning, we can improve the accuracy and efficiency of bridge inspections, prioritize maintenance and repair activities, and develop predictive models that can be used to anticipate future bridge failures.

Future research directions include:

- Developing more sophisticated machine learning models that can account for a wider range of factors that affect bridge condition.
- Collecting and analyzing larger datasets of bridge inspection data to improve the accuracy of machine learning models.
- Developing real-time monitoring systems that can use machine learning to detect bridge damage in real time.
- Deploying machine learning-based bridge condition analysis tools in bridge management systems.

We believe that machine learning has the potential to make a significant impact on bridge management. By continuing to develop and apply machine learning techniques, we can improve the safety and reliability of our nation's bridges.

# CHAPTER 6: REFERENCES

Fielding, E., Lo, A. W., & Yang, J. H. (2010). The National Transportation Safety Board: A model for systemic risk management. Available at SSRN 1695781.

Lo Assaad, R., & El-adaway, I. H. (2020). Bridge infrastructure asset management system: Comparative computational machine learning approach for evaluating and predicting deck deterioration conditions. Journal of Infrastructure Systems, 26(3), 04020032.

Wu, N. C., & Chase, S. (2010). An exploratory data analysis of national bridge inventory (No. UVA-2009-03). Mid-Atlantic Universities Transportation Center.

Trans, G. I. S. (1999). An in-depth analysis of the national bridge inventory database utilizing data mining, GIS and advanced statistical methods.

Cook, W. (2014). Bridge failure rates, consequences, and predictive trends. Utah State University.

Morcous, G. (2005). Modeling Bridge Deck Deterioration by Using Decision Tree Algorithms. In Transportation Research Record: Journal of the Transportation Research Board.

Moriyoshi Kushida, B., Miyamoto, A., & Kinoshita, K. (n.d.). DEVELOPMENT OF CONCRETE BRIDGE RATING PROTOTYPE EXPERT SYSTEM WITH MACHINE LEARNING.

Okazaki, Y., Okazaki, S., Asamoto, S., & Chun, P. jo. (2020). Applicability of machine learning to a crack model in concrete bridges. Computer-Aided Civil and Infrastructure Engineering, 35(8), 775–792.

Qian, Y., Aghaabbasi, M., Ali, M., Alqurashi, M., Salah, B., Zainol, R., Moeinaddini, M., & Hussein, E. E. (2021). Classification of imbalanced travel mode choice to work data using adjustable svm model. Applied Sciences (Switzerland), 11(24). https://doi.org/10.3390/app112411916

Tao, W., Aghaabbasi, M., Ali, M., Almaliki, A. H., Zainol, R., Almaliki, A. A., & Hussein, E. E. (2022). An Advanced Machine Learning Approach to Predicting Pedestrian Fatality Caused by Road Crashes: A Step toward Sustainable Pedestrian Safety. Sustainability (Switzerland),