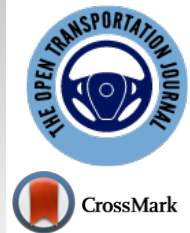




# The Open Transportation Journal

Content list available at: <https://opentransportationjournal.com>



## RESEARCH ARTICLE

### Bridge Frost Prediction Using K-nearest Neighbor Classifier

Jinhwan Jang<sup>1,\*</sup>

<sup>1</sup>Highway Research Division, Korea Institute of Civil Engineering and Building Technology, South Korea

#### Abstract:

#### Background:

Considering the frequent occurrence of accidents on icy bridges during winter nights, it would be advantageous to notify both road managers and drivers regarding the most perilous areas. This notification would allow road managers to address the icy conditions by applying de-icing substances, while drivers could be more adequately prepared for potential hazards.

#### Methods:

In this study, the focus was on investigating k-nearest neighbor algorithms to predict nighttime icing caused by frost on three distinct bridges located on the National Highways in Korea. The algorithms utilized atmospheric data as input, which was obtained from the weather agency's website through an open API. The input data included relative humidity, air temperature, and dew point temperature, as well as the disparities in air temperature and humidity between two consecutive days.

#### Results:

In order to assess the effectiveness of the prediction models, reference data were created using the fundamental principle that ice is formed when the temperature of the pavement is below freezing and lower than the dew point temperature. Consequently, the developed algorithm demonstrated favorable performance, achieving an accuracy of 95% when evaluated using a test dataset that occupies 30% of the entire data.

#### Conclusion:

Considering the increasing focus on preventive maintenance, these newly developed forecasting models can be employed proactively as a preventive measure against icing. This proactive approach will ultimately contribute to improving traffic safety on winter roads.

**Keywords:** Bride frost, Prediction, K-nearest neighbor, Atmospheric data, Black ice, Traffic safety.

#### Article History

Received: May 23, 2023

Revised: July 22, 2023

Accepted: July 26, 2023

## 1. INTRODUCTION

A significant number of traffic accidents occur on slippery roads. According to a study, these accidents lead to 1,300 fatalities and 13,735 injuries annually in the United States [1]. In Sweden, a report indicated that only 14% of drivers effectively adjust their speed on slippery surfaces [2]. Research conducted in Portugal suggests that the likelihood of a collision on icy roads is nine to ten times higher compared to dry surfaces [3]. In South Korea, data over a span of five years shows that 6,502 incidents on slippery roads resulted in 198 deaths. In 2019, a tragic accident took place on an icy bridge in South Korea, causing seven fatalities and numerous injuries. A similar incident occurred again in 2022.

Furthermore, the Korean government has instructed road maintenance agencies to incorporate pavement temperature measurements into their nighttime road maintenance patrols. These measurements serve as a foundation for implementing anti-icing initiatives [4]. Due to the fact that atmospheric sensors are typically situated far from the roads, reliable monitoring of road temperatures becomes challenging. Moreover, surface temperatures exhibit substantial fluctuations, even over short road sections, which restricts the practicality of using atmospheric temperature data for anti-icing purposes [5]. As a result, it becomes crucial to continuously measure pavement temperatures along the roadways to effectively combat black ice, particularly on bridges.

In general, snow removal operations are conducted according to weather forecasts, where snowplows are deployed

\* Address correspondence to this author at the Korea Institute of Civil Engineering and Building Technology, South Korea; E-mail: [jhjang@kict.re.kr](mailto:jhjang@kict.re.kr)

to clear snow and apply chemicals either during or shortly before (1-3 hours) the snowfall. However, in Korea, the absence of black ice forecasts puts a significant strain on maintenance personnel. As a result, they are required to carry out daily patrols along the roads, imposing a substantial workload on them.

In order to address this problem, the present research focused on creating forecasting models for black ice to enhance the effectiveness of anti-icing operations, specifically targeting frost-induced black ice on bridges. A k-nearest neighbor model, a widely recognized machine learning algorithm, was employed, utilizing atmospheric data as input. By utilizing this forecasting model, nighttime occurrences of black ice can be predicted using daytime weather forecasts. This enables maintenance personnel to perform anti-icing activities, such as patrolling and applying chemicals only when the occurrence of black ice is anticipated.

## 2. LITERATURE REVIEW

Black ice can develop for several reasons [6], such as the overnight freezing of melted snow, the freezing of rain on the pavement with sub-zero temperatures, and frost bonding with the road surface. The first two causes can be anticipated through atmospheric weather forecasting since they are associated with snow or rain. However, the last cause can only be identified through routine road maintenance patrols [7].

In order to forecast black ice, traditional approaches have relied on physical models and regression analysis. Physical models utilize a surface energy balance model that considers heat conduction, convection, radiation, and vapor movement to estimate pavement temperature [8 - 11]. On the other hand, regression models predict pavement temperature using various data sources, including atmospheric data, geometric factors, and air temperature collected from probe cars [12 - 14]. However, both types of models have their limitations. Physical models require extensive data, such as pavement thickness, heat transfer rate of pavement materials, and heat flux, which may not always be readily available. This indicates that relying solely on atmospheric data and pavement temperature is insufficient for accurate black ice prediction. While easier to comprehend, regression models struggle with predicting variables that exhibit nonlinear correlations. Hence, this study employed a machine learning model to overcome the limitations of previous methods.

## 3. METHODOLOGY

The k-nearest neighbor (k-NN) algorithm is a versatile machine learning model employed for classification and regression tasks. It is categorized as a non-parametric algorithm since it does not assume any specific data distribution.

The functioning of the k-NN algorithm involves comparing a new data point to its k-nearest neighbors within the training dataset. The value of k is determined by the user. In classification tasks, the majority class label among the k-nearest neighbors is assigned to the new data point. In regression tasks, the predicted value is obtained by calculating

the average or weighted average of the target values associated with the k-nearest neighbors.

The k-NN algorithm is straightforward and easy to understand, but its performance greatly depends on the choice of k and the distance metric utilized. Consequently, the quest for determining the optimal values of k and the distance metric becomes crucial in order to obtain a well-performing model.

## 4. RESULT

### 4.1. Data Collection

#### 4.1.1. Pavement Temperature and Atmospheric Data

In this research, road maintenance vehicles were fitted with a road surface temperature sensor (depicted in Fig. 1), which utilizes an infrared thermometer to gauge the temperature of three bridges, as illustrated in Fig. (2). While the vehicle is in motion, the sensor is capable of measuring the temperature every 0.2 seconds. Pavement temperature data were obtained from three bridges, each approximately 300 meters in length, from December 2022 to February 2023. Data collection took place on a total of 397 days, with measurements recorded once per day during nighttime. Its precision is greatly enhanced by a narrow half-angle of 5°, resulting in accurate surface temperature measurements [15 - 17]. At 0°C, the sensor exhibits an accuracy of  $\pm 0.3^\circ\text{C}$ . Furthermore, concurrent atmospheric data (air temperature and relative humidity) were acquired from the closest weather station during the same time period from an open API operated by the weather agency.



Fig. (1). Pavement temperature sensor.

#### 4.1.2. Pre-processing of Pavement Temperature Data

In order to efficiently utilize the gathered data for de-icing operations, it was crucial to consolidate the information according to specific roadway sections. To achieve this, the Korean government implemented a standardized node-link system (depicted in Fig. 3). This system divides the entire public highway network into nodes and links, taking into account various roadway characteristics, such as bridges, tunnels, overpasses, underpasses, intersections, and the number of lanes. Since the temperature of the pavement differs depending on the type of roadway, the aggregation of data using the node-link system is regarded as a logical and effective approach.



Fig. (2). Data collection sites.



Fig. (3). Standard node-link system of Korea.

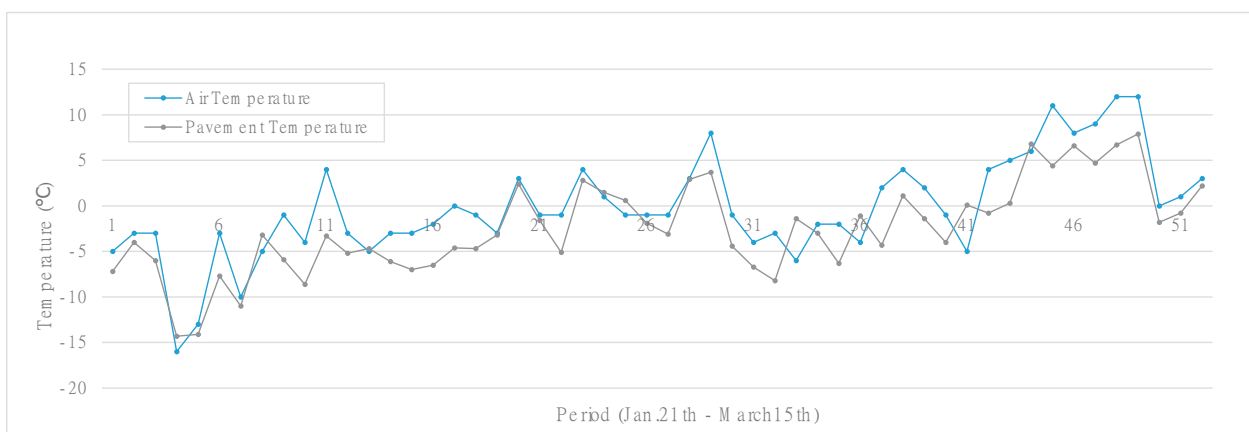


Fig. (4). Air temperature vs. pavement temperature.

For the analysis of pavement temperature, the standard node-link system was employed to aggregate the data into a single median value on a daily basis, which served as a representative measure of central tendency. Nevertheless, as

the measurements of pavement temperature obtained from moving vehicles may include outliers due to factors like debris on the road, driving on the shoulder lane, intermittent stops during patrolling, and so on, it was crucial to adopt a careful

aggregation method to address any unforeseen consequences caused by these outliers. Among the three commonly used measures of central tendency (mean, median, and mode), the median was chosen because it is less influenced by extreme observations [18, 19].

**4.1.3. Analysis of Pavement Temperature and Atmospheric Data**

Fig. (4) depicts a graph illustrating the pavement temperature and air temperature recorded at one of the aforementioned three bridges. Similar patterns were observed at the other two bridges. Overall, the bridge temperature was consistently lower than the air temperature, with the air temperature exhibiting greater variability. Tables 1 and 2 display that, on average, the bridge temperature was approximately 2 times lower than the air temperature, while the maximum temperature reached about 4 °C higher. In contrast, the air temperature had a minimum temperature of approximately 2 °C lower than the pavement temperature. Unlike the air temperature, which fluctuates considerably due to air flow, the pavement temperature is presumed to exhibit relatively low variability due to the geothermal and latent heat properties of the structure, as shown in Fig. (5).

**Table 1. Statistics of pavement (bridge) and air temperature.**

Statistics	Air Temp.	Pav. (Bridge) Temp.
Mean	-0.30	-2.47
Standard deviation	5.56	5.01
Minimum	-16.00	-14.30
25%	-3.00	-5.93
50%	-1.00	-3.15
75%	3.00	0.73
Maximum	12.00	7.90

**Table 2. Model performance.**

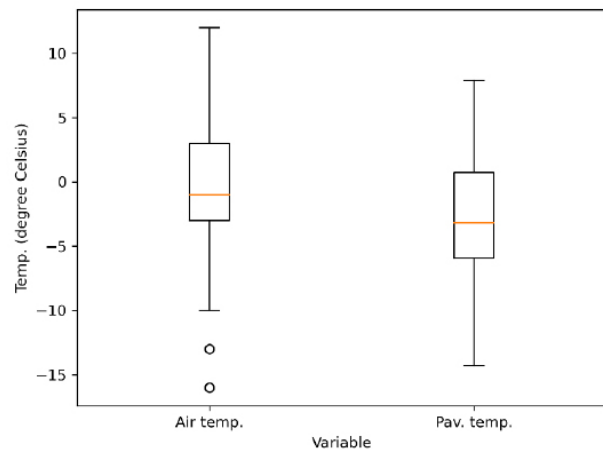
Index	Score
Accuracy	0.95
Precision	0.97
Recall	0.93
F1 score	0.95

**4.2. Bridge Frost Prediction**

**4.2.1. Building Blocks of K-Nearest Neighbor Models**

A predictive model for nighttime black ice using atmospheric data was established. The input data for a k-NN model included temperature, humidity, dew point temperature, temperature differences between consecutive days, and humidity differences between consecutive days (see Fig. 6). These data were collected over a span of two years at the three bridges mentioned earlier. To create the baseline data, we combined the input data with pavement temperature data obtained from patrol vehicles. Detailed information regarding the baseline data will be presented in the subsequent chapter. The configuration of the k-NN model is illustrated in Fig. (6). Since the scales of the input data varied, we applied standard

scaling to normalize the data prior to training the models. The training and test sets were divided into a 7:3 ratio, and the raw data were classified based on the ratio of dry/icing conditions to ensure a balanced distribution in both sets. The analysis encompassed a total of 397 days, with 193 days classified as icy conditions and 204 days as dry conditions.



**Fig. (5).** Boxplot of air and pavement temperature.

**4.2.2. k-NN Algorithm Procedure**

To implement the k-NN algorithm in Python, the following steps were followed:

1. Load the labeled training dataset, which contains input feature vectors and their corresponding class labels or target values.
2. Choose a value for k, which represents the number of nearest neighbors to consider during predictions.
3. Compute the distances between a new input sample and all the training samples using a distance metric. This metric quantifies the similarity or dissimilarity between two data points.
4. Identify the k-training samples with the smallest distances to the input sample. These samples will be the nearest neighbors.
5. For classification tasks, assign the class label that appears most frequently among the k-nearest neighbors to the input sample.

**4.2.3. Finding Optimal Parameters**

Finding the best value for k is crucial in the k-nearest neighbors (k-NN) algorithm. A small k can lead to overfitting, making the model overly sensitive to noisy data. Conversely, a large k can result in over-smoothing, causing the model to overlook local patterns in the data. The optimal value for k is typically determined through techniques like cross-validation or grid search. These methods involve evaluating different k values and selecting the one that performs best based on evaluation metrics. Fig. (7) illustrates the process of selecting the optimal k value, with a value of one being identified as the optimal choice.

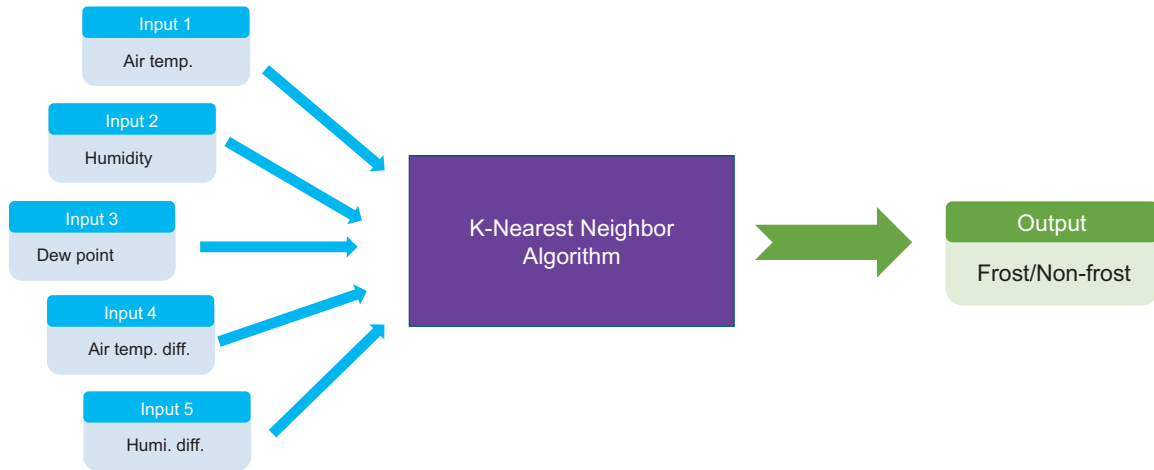


Fig. (6). Building block of k-NN model.

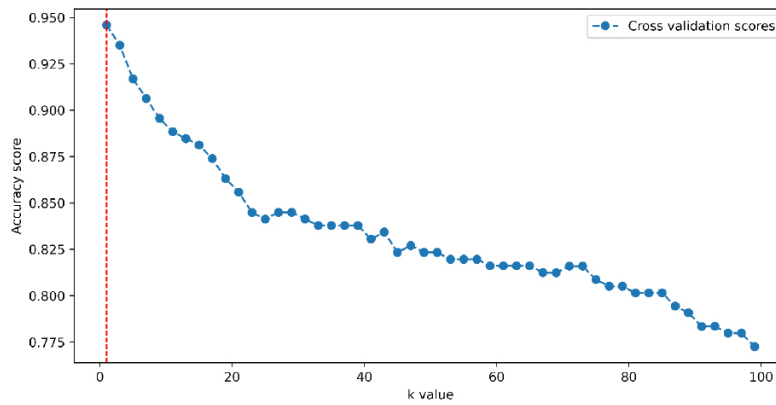


Fig. (7). Model performance by k values.

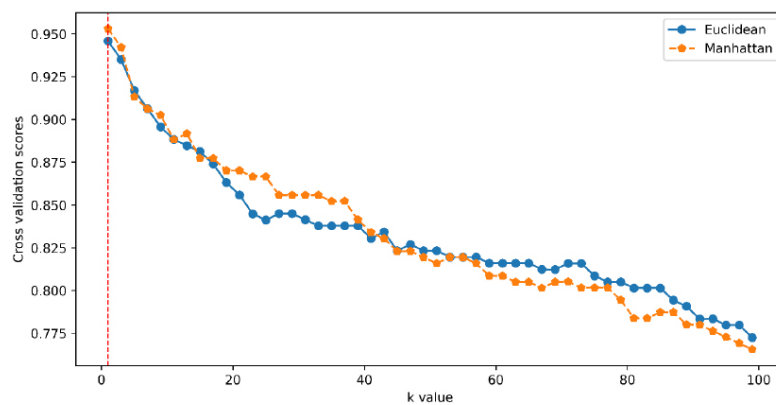


Fig. (8). Model performance by distance measures.

4.2.4. Optimal Metric Selection

The choice of distance metric in the k-nearest neighbors (k-NN) algorithm depends on the characteristics of the data and the specific problem being addressed. Euclidean and Manhattan distances (Eqs. 1 and 2) are commonly used

options. Therefore, the selection of the optimal metric was performed for these two metrics, as shown in Fig. (8). The analysis revealed that for a value of k equal to one, the optimal metric was found to be the Euclidean distance. Typically, k-NN algorithms can be computationally expensive when dealing with large datasets, as they involve calculating distances for

each new sample against all the training samples. However, in this study, the data size was relatively small (only a few bytes), which did not pose a hindrance to the extensive computations.

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$d_{(x,y)} = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

where  $d_{(x,y)}$  = distance between points x and y.

### 4.3. Assessment

#### 4.3.1. Baseline Data Generation

To assess the precision of the predicted black ice data, it is crucial to establish a baseline. While employing a reference device to directly measure the road surface conditions would be the most reliable and straightforward approach, it was not feasible within the scope of this study due to labor-intensive requirements, costs, and the need for device acquisition. Instead, we relied on a physical principle described in Eq. (3), which indicates that frost develops on the pavement when the pavement temperature is not only below freezing but also lower than the dew point temperature.

$$T_p \leq 0^\circ\text{C} \text{ and } T_p \leq T_d \quad (3)$$

where  $T_p$  = pavement temperature and  $T_d$  = dew point temperature.

The concept of dew point temperature pertains to the specific temperature at which moisture in the air reaches saturation, resulting in the formation of fog or frost and causing the relative humidity to reach 100%. To determine the dew point temperature, the commonly employed approach is to utilize the Magnus formula, which is a widely accepted method for this purpose [15]. This formula calculates the saturation vapor pressure over liquid water at a given temperature (T) using Eq. (4).

$$EW = \alpha \cdot e^{\left(\frac{\beta \cdot T}{\lambda + T}\right)} \quad (4)$$

The Eq. (4) incorporates the parameters  $\alpha$  (6.112 hPa),  $\beta$  (17.62), and  $\lambda$  (243.12°C). By rearranging the terms in Eq. (4), we can derive Eq. (5), which provides an expression for the dew point ( $T_d$ ) based on the vapor pressure.

$$T_d = \frac{\lambda \cdot \ln\left(\frac{E}{\alpha}\right)}{\beta - \ln\left(\frac{E}{\alpha}\right)} \quad (5)$$

By substituting the definition of relative humidity ( $E = RH \cdot EW / 100$ ) into Eq. (5), we can derive Eq. (6), which enables us to calculate the dew point ( $T_d$ ) using both the atmospheric temperature (T) and relative humidity (RH).

$$T_d(T, RH) = \frac{\lambda \cdot \left[ \ln\left(\frac{RH}{100}\right) + \frac{\beta \cdot T}{\lambda + T} \right]}{\beta - \left[ \ln\left(\frac{RH}{100}\right) + \frac{\beta \cdot T}{\lambda + T} \right]} \quad (6)$$

Eqs. (4-6) employ the well-established Magnus formula, which is widely recognized as a reliable method for estimating the dew point temperature using the atmospheric temperature

and relative humidity. This formula exhibits an error rate of approximately 0.35°C [16].

To validate the precision of the Magnus formula, a winter road segment was surveyed, specifically focusing on frost-induced black ice occurrences. It was observed that black ice formed (on the left) when the criteria for frost formation were satisfied, whereas no black ice was observed (on the right) when those criteria were not met, as depicted in Fig. (9).

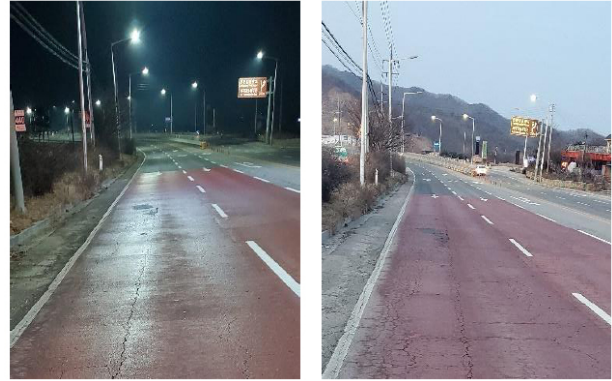


Fig. (9). Pavement with frost (left) and without frost (right).

Table 3. Confusion matrix.

Classification		Prediction	
		Dry	Icy
Baseline	Dry	0.935	0.065
	Icy	0.034	0.966

#### 4.3.2. Performance of k-NN Model

The developed model was evaluated using a test dataset consisting of 119 samples, which accounted for 30% of the entire data, and the performance is presented in Table 1. The evaluation metrics used, including accuracy, precision, recall, and F1 score, are commonly employed to assess the effectiveness of machine learning models. These metrics, as defined in Eqs. (7-10), indicate that the model yielded satisfactory outcomes with accuracy, precision, recall, and F1 score values of 0.95, 0.97, 0.93, and 0.95, respectively. Additionally, Table 3 showcases the confusion matrix, which further demonstrates the model's strong performance across all cases. Particularly noteworthy is its capability to accurately identify "icy" conditions, which is crucial for winter road maintenance.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (7)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (8)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (9)$$

$$F1\ Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (10)$$

## 5. DISCUSSION

Conventionally, the Road Weather Information System

(RWIS) has been used to predict black ice. To utilize RWIS, Environmental Sensor Stations (ESSs) are essentially deployed on the road, implying a substantial budget need to be allocated. Unfortunately, no available RWIS ESSs are installed in Korea. In this regard, predicting frost-induced nighttime black ice using readily available atmospheric data can be regarded as a cost-efficient countermeasure. Additionally, weather phenomena are known to be recurrent, indicating that the k-NN model developed in this study could be effectively applied. The suggested model can be practically applied to nighttime road maintenance patrolling, which would make winter road maintenance resources more efficient.

It should also be noted that the baseline data employed in this analysis do not directly reflect the actual surface conditions observed in the field. Instead, these values are calculated using road surface temperature and atmospheric data. Additionally, it should be noted that even when the conditions for frost formation are met based on physical laws, road icing may not occur in situations with high traffic volume [20, 21]. Nonetheless, the findings of this study hold significant value from various perspectives. From the standpoint of road administrators, if the presence of icing due to frost is anticipated, they can proactively engage in anti-icing measures. In this regard, the outcomes of this study can be practically applied to real-world winter road maintenance activities.

## CONCLUSION AND FUTURE STUDIES

In Korea, the occurrence of accidents caused by black ice, which led to seven fatalities in 2019 and numerous injuries in 2023, has prompted significant changes in winter road maintenance practices. As a response, field workers have been assigned daily maintenance patrols throughout the winter season. However, there has been a growing demand among field personnel for an improved patrolling schedule, particularly concerning the development of nighttime ice predictions.

Currently, there are no available predictive models specifically for bridge frost treatment using atmospheric data. As a result, maintenance patrols are often conducted on bridges even when the occurrence of black ice is not anticipated, leading to inefficiencies in terms of labor and equipment. To address this issue, a reliable machine learning model using k-NN was explored to predict the potential for nighttime black ice, yielding satisfactory results with an accuracy rate of 95%. Notably, the strength of this model lies in its ability to forecast bridge frost solely based on atmospheric data, making it highly applicable to real-world scenarios. Utilizing the findings of this study, winter road maintenance personnel can streamline their patrols by focusing solely on predicted occurrences of black ice. This approach makes winter road patrol activities more efficient and scientifically informed.

However, it is essential to note that the evaluation of the k-NN model was solely based on baseline data derived from a physical principle. In the future, if feasible, it is recommended to gather baseline data through field observations using a device that accurately measures road slipperiness to ensure a more robust evaluation process.

## LIST OF ABBREVIATIONS

<b>k-NN</b>	=	k-Nearest Neighbor
<b>RWIS</b>	=	Road Weather Information System
<b>ESSs</b>	=	Environmental Sensor Stations

## CONSENT FOR PUBLICATION

Not applicable.

## AVAILABILITY OF DATA AND MATERIALS

The data and supportive information are available within the article.

## FUNDING

This work was financially supported by a grant (No. KMI 2002-00110) from the Korea Meteorological Institute (KMI).

## CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

## ACKNOWLEDGEMENTS

Declared none.

## REFERENCES

- [1] U.S. FHWA, "21st Century Operations using 21st Century Technologies", Available From: [https://ops.fhwa.dot.gov/weather/weather\\_events/snow\\_ice.htm](https://ops.fhwa.dot.gov/weather/weather_events/snow_ice.htm)
- [2] J. Bogren, and P.E. Caran, "SRIS-Slippery Road Information System", Available From: <https://trid.trb.org/view/908698>
- [3] P. Luque, J. Wideberg, and D. Mantars, *ITS to Improve Safety and Efficiency OBD-II and Smartphone Apps.*, CreateSpace Independent Publishing Platform: California, 2013.
- [4] Korean Government, "Winter Road-Traffic Safety Enhancement Measures (in Korean)", <https://www.oecd.org/regional/regional-policy/Traffic-Safety-Korea-Vulnerability-of-Elderly-Pedestrians.pdf>
- [5] R. Mats, G. Torbj, B. Jorgen, and J. Per-Erik, "Ice Formation Detection on Road Surfaces using Infrared Thermometry, Cold Regions Science and Technology", *Cold Regions Science and Technology*, vol. 83–84, pp. 71–76, 2012.
- [6] American Association of State Highway and Transportation Officials, "Update of the AASHTO Guide for Snow and Ice Control", Available From: <https://trid.trb.org/view/1246275>
- [7] PIARC Technical Committee, "Snow and Ice Databook 2018", Available From: <https://www.piarc.org/en/order-library/37033-en-Snow%20And%20Ice%20Databook%202022%20-%20Technical%20Report>
- [8] C. Louis-Philippe, and D. Yves, "METRO: A New Model for Road-Condition Forecasting in Canada", *J Appl Meteorol*, vol. 40, no. 11, pp. 2026–2037, 2001.
- [9] M. G. Tina, and S. T. Eugene, "Bridge Frost Prediction by Heat and Mass Transfer Methods", *J Appl Meteorol Climatol*, vol. 45, no. 11, pp. 517–525, 2006.
- [10] D.N. Claudia, C. Roberto, A. Gianluca, and B. Guido, "Thermal mapping as a valuable tool for road weather forecast and winter road maintenance: An example from the Italian Alps", *Proceedings of the Fourth International Conference on Remote Sensing and Geo-information of the Environment*, 2016 [<http://dx.doi.org/10.1117/12.2240484>]
- [11] K. Virve, "Observing and Forecasting Road Surface Temperatures", Available From: <https://helda.helsinki.fi/items/35b6e037-8cf5-42d3-8014-dfbc542f9f7e>
- [12] L. Chapman, J.E. Thornes, and A.V. Bradley, "Modelling of road surface temperature from a geographical parameter database. Part 2: Numerical", *Meteorol. Appl.*, vol. 8, no. 4, pp. 421–436, 2001. [<http://dx.doi.org/10.1017/S1350482701004042>]

- [13] V.J. Berrocal, A.E. Raftery, T. Gneiting, and R.C. Steed, "Probabilistic Weather Forecasting for Winter Road Maintenance", *J. Am. Stat. Assoc.*, vol. 105, no. 490, pp. 522-537, 2010. [http://dx.doi.org/10.1198/jasa.2009.ap07184]
- [14] H. Yumei, A. Esben, G. Torbjorn, and B. Jrgen, "Modeling Road Surface Temperature from Air Temperature and Geographical Parameters—Implication for the Application of Floating Car Data in a Road Weather Forecast Model", *American Meteorological Society*, vol. 58, pp. 517-525, 2019.
- [15] D. Sonntag, "Vapour Pressure Formulations based on the IST-90 and Psychrometer Formulae", *Z. Meteorol.*, vol. 70, no. 5, pp. 340-344, 1990.
- [16] R. Paul, and J.D. Paul, "Night Icing Potential – Demonstration Project", *Proc. of the 2008 Annual conference of the Transportation Association of*, 2008 Toronto, Ontario
- [17] Apogee Instrument, "Field of View", Available From: <https://www.apogeeinstruments.com/field-of-view/>
- [18] J. Shao, and P.J. Lister, "Data filtering for thermal mapping of road surface temperatures", *Meteorol. Appl.*, vol. 2, no. 2, pp. 131-135, 1995. [http://dx.doi.org/10.1002/met.5060020206]
- [19] ASTM International, "Standard Practice for Dealing with Outlying Observations", Available From: <https://www.document-center.com/standards/show/ASTM-E178>
- [20] B.R. Temeyer, "Use of an artificial neural network to predict air temperature, surface temperature, dew point and wind speed for the prediction of frost",
- [21] T. Gustavsson, "Variation in road surface temperature due to topography and wind", *Theor. Appl. Climatol.*, vol. 41, no. 4, pp. 227-236, 1990. [http://dx.doi.org/10.1007/BF00866454]

© 2023 The Author(s). Published by Bentham Science Publisher.



This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.