

# 기계학습을 통한 여름철 노면상태 추정 알고리즘 개발

## Estimation of Road Surface Condition during Summer Season Using Machine Learning

여지호\* · 이주영\*\* · 김강화\*\*\* · 장기태\*\*\*\*

\* 주저자 : 한국과학기술원 조천식녹색교통대학원 박사과정  
 \*\* 공저자 : 한국과학기술원 조천식녹색교통대학원 박사과정  
 \*\*\* 공저자 : 디토닉 주식회사 수석연구원  
 \*\*\*\* 교신저자 : 한국과학기술원 조천식녹색교통대학원 부교수

Jiho Yeo\* · Jooyoung Lee\*\* · Ganghwa Kim\*\*\* · Kitae Jang\*\*\*\*

\* The Cho Chun Shik Graduate School of Green Transportation, KAIST  
 \*\* The Cho Chun Shik Graduate School of Green Transportation, KAIST  
 \*\*\* Dtonic Corporation  
 \*\*\*\* The Cho Chun Shik Graduate School of Green Transportation, KAIST  
 † Corresponding author : Kitae Jang, kitae.jang@kaist.ac.kr

Vol.17 No.6(2018)

December, 2018

pp.121~132

pISSN 1738-0774

eISSN 2384-1729

<https://doi.org/10.12815/kits.2018.17.6.121>

2018.17.6.121

Received 21 September 2018

Revised 23 October 2018

Accepted 5 November 2018

© 2018. The Korea Institute of Intelligent Transport Systems. All rights reserved.

### 요약

기상은 교통흐름, 운전자의 주행패턴, 교통사고 등 여러 방면에서 도로교통에 영향을 미치는 중요한 요인이다. 본 연구는 기상상황과 노면상태 사이의 관계에 초점을 맞추어 기계학습을 통해 도로의 노면상태를 추정하는 모델을 개발하였다. 노면 상태의 수집을 위해 실험 차량에 노면센서를 부착하여 '건조', '습윤', '젖음', 3가지 범주로 구분된 노면상태 정보를 수집하였고, 이를 추정하기 위한 변수로 도로의 기하구조 정보(곡률, 구배), 교통정보(교통량), 기상정보(강우량, 습도, 온도, 풍속)를 활용하였다. 노면 상태를 예측하기 위한 알고리즘으로는 다양한 기계학습 알고리즘이 검토되었으며, 그 중 가장 높은 정확도를 보인 'Random forest'를 기반으로 한 2단계 분류모형을 구축하였다. 총 16일의 실측 데이터 중 14일의 데이터를 모델 학습하는 데 활용하였고, 2일의 데이터를 모형의 정확도를 검증하기 위해 사용하였다. 그 결과 81.74%의 검증 정확도를 가지는 노면상태 예측 모델을 구축하였다. 본 연구의 결과는 기상청에서 관측하는 기상정보로 도로의 노면상태를 추정할 수 있다는 가능성을 보여주며, 새로운 장비나 센서를 설치하지 않고도 기존의 기상 관측 정보와 교통정보 등을 활용하여 노면의 상태를 추정할 수 있음을 시사한다.

핵심어 : 교통안전, 기상, 노면상태, 기계학습

### ABSTRACT

Weather is an important factor affecting roadway transportation in many aspects such as traffic flow, driver's driving patterns, and crashes. This study focuses on the relationship between weather and road surface condition and develops a model to estimate the road surface condition using machine learning. A road surface sensor was attached to the probe vehicle to collect road surface condition classified into three categories as 'dry', 'moist' and 'wet'. Road geometry information (curvature, gradient), traffic information (link speed), weather information (rainfall, humidity, temperature, wind speed) are utilized as variables to estimate the road surface condition. A variety of machine learning algorithms examined for predicting the road surface condition, and a two-stage classification model based on 'Random forest' which has the highest accuracy was constructed. 14

days of data were used to train the model and 2 days of data were used to test the accuracy of the model. As a result, a road surface state prediction model with 81.74% accuracy was constructed. The result of this study shows the possibility of estimating the road surface condition using the existing weather and traffic information without installing new equipment or sensors.

Key words : Traffic safety, Weather, Road surface condition, Machine learning

## I. 서 론

기상의 변화는 여러 측면에서 도로교통에 영향을 미치고 있으며, 이를 반영하여 교통 흐름을 원활히 하고 교통사고를 예방하기 위해 교통상황과 기상 상황에 대한 다양한 연구가 수행되고 있다. 기존에 기상과 교통 사이의 관계를 분석한 연구는 크게 1) 기상에 따른 교통흐름의 변화(Kwon, 2013; Jeong, 2013; Keay et al., 2005; Knapp et al., 2000; Shi et al., 2011) 2) 기상이 교통사고에 미치는 영향(Morgan, 2011; Edwards, 1998; Broadsky, 1988; Choi et al., 2013) 그리고 3) 기상에 따른 운전자의 행동 분석(Rahman, 2012; Hoogendoorn, 2010)으로 요약해 볼 수 있다. 기존의 연구들은 기상이 교통흐름 및 안전에 매우 중요한 요인이고 따라서 기상 상황에 따른 도로의 상태를 정확히 모니터링 하고 관리해야 한다는 점을 시사한다.

위의 연구들은 대부분 기상관측소에서 제공하는 기상정보를 기반으로 하였으나, 실제로 특정 장소에 위치한 기상관측소에서 측정된 기상상황과 도로 위의 기상 상황은 차이가 존재한다. 이를 해결하기 위하여 정확한 노면상태를 판별하기 위한 다양한 연구가 수행되었다.

Kim et al.(2013)은 노면상태를 판별하는 것에 초점을 맞추어 영상정보를 기반으로 하는 SVM(Support Vector Machine)을 통해 노면상태를 판별하는 알고리즘을 개발하였다. 영상정보를 기반으로 하여 편광계수 및 웨이블릿 변환(Wavelet transform)을 통해 SVM 분류기법을 적용하여 실제 도로상의 4단계 노면상태(마름, 젖음, 적설, 결빙)에 대해 검지 모델을 구축하고 예측 정확도를 분석하였다. Kim et al.(2010)은 로봇에 부착된 카메라로부터 취득되는 영상정보를 통해 지면의 마찰계수를 추출하는 방법을 제안하였다. 총 6가지 물질(하늘, 흙, 작은 자갈, 자갈, 나무, 아스팔트)를 통해 물질의 조성비를 계산한 후 베이지안 분류기법을 활용하여 노면의 마찰계수를 추정하였다. 한편 영상정보를 이용한 모델링 외에도 광학센서를 통한 노면상태의 판별도 수행되었는데, 핀란드의 Teconer 사는 도로노면 상태 관측장비인 RCM411을 개발하여 광학센서 기반의 분광분석(Spectral analysis)를 통해 노면의 온도, 물의 양을 측정하였고, 이를 바탕으로 노면의 상태 및 마찰계수를 추정하였다(Haavasoia et al., 2011).

위의 연구들은 각각 다른 방식으로 노면의 상태를 수집하고 다른 판별 분석을 사용하였지만 모두 현장 실험을 통해 수집한 데이터를 기반으로 노면 상태를 추정하였다는 공통점이 있다. 이러한 접근 방법은 특정 지점의 정확한 노면의 상태를 파악할 수 있다는 장점이 있지만, 실험이 수행된 지역 외에서는 노면의 상태를 알 수 없다는 한계가 있다. 이를 보완하기 위해 보다 광역적으로 수집이 가능한 기상정보와 노면상태 사이의 관계를 규명하는 다양한 연구가 수행되었다. Chapman et al.(2001)은 다양한 공간지리번호(위도, 경도, 스카이뷰, 교통 밀도, 지형도 등)를 사용하여 노면 온도를 예측하였고, 72%의 분산을 설명하는 모델을 구축하였다. Shao(1996)은 7개 국가의 17곳의 장소에서 센서를 통해 도로 노면의 온도, 이슬점, 풍속, 운량을 측정하였고, 3시간 후까지의 노면 상태를 예보하는 모델을 개발하였다. 국내에서는 지능형 교통체계의 하나인 도로기상정보 시스템(Road Weather Information System, 이하 RWIS)을 통해 도로상에서의 강수유무, 강수량, 대기 온도 등에 대한 실시간 정보를 수집하여 기상청의 자동기상관측장비(Automatic Weather Station, 이하 AWS) 정보

를 보완하고 있다.

하지만 RWIS의 경우에도 도로 상 특정 지점에 설치하는 장비이기 때문에 해당 지점의 도로 노면 정보만을 취득할 수 있어 전체 도로구간에 대한 정보 수집을 할 수 없으며, 이를 기반으로 하는 기존의 연구들도 특정 지점에 국한해서 노면상태를 추정하였다는 한계가 존재한다.

따라서 본 연구는 보다 넓은 범위의 기상관측정보를 사용하여 도로의 노면상태를 추정하는 모델을 개발하는 것을 목표로 한다. 기존에 기상청, 지자체, 민간기업에서 설치한 기상관측센서를 활용하여 보다 넓은 범위에서의 기상정보를 수집하였고, 이를 활용하여 노면상태 예측 모형을 구축하였다. 노면 상태에 영향을 미치는 요인들을 추출하기 위해 기상정보 외에도 도로 기하구조 정보, 교통정보 등의 다양한 변수들을 수집하였으며, 정확한 도로 노면 상태 관측을 위해 RCM411(Teconer, 2015) 모델을 수집용 차량에 부착하여 초 단위로 도로의 노면상태를 측정하였다. 분석 기간은 2017년 7월부터 8월까지로 설정하여 강우 시의 노면상태 예측에 초점을 맞추었으며, 분석의 공간적 범위는 기상 상태 외의 교통상황에 대한 변수를 최소화 할 수 있고, 노면센서의 오차를 최소화 할 수 있는 연속류 도로인 내부순환로를 선정하였다.

본 논문의 구성은 다음과 같다. 2장에서는 노면상태, 기상정보, 기하구조 정보 등 데이터를 수집하는 방법과 수집된 데이터를 설명한다. 3장에서는 노면 상태를 예측하는데 활용한 분석 모델에 관해 설명하고 모형을 통해 예측된 결과와 예측 정확도를 제시한다. 마지막으로 4장에서는 결론 및 향후 연구 과제에 대해 논의한다.

## II. 자료 수집 및 개요

### 1. 분석의 시공간적 범위

본 연구에서는 서울의 내부순환로를 대상으로 노면상태 및 기상정보 데이터를 수집하였다. 기상 상태 외의 다른 교통상황에 대한 변수를 최소화 할 수 있고, 또 차량이 정지했을 시에 실험 차량에 부착하는 노면센서가 오작동을 일으킬 수 있는 위험이 있으므로 서울의 연속류 도로 중 하나인 내부순환로를 연구의 공간적 범위로 선정하였다. 연구의 시간적 범위로는 2017년 7월~8월 사이 중 비가 오거나 온 직후에 데이터를 수집하였으며, 총 16일의 데이터를 수집하였다. 맑은 날에 실험을 수행하여 노면이 건조할 때 건조하다는 정보를 얻는 것은 의미가 없으므로 비가 오는 도중이나 온 직후에 실험을 수행하여, 습윤한 노면이 서서히 마르는 과정 동안의 노면상태 정보를 수집하였다.

### 2. 자료 수집 및 데이터 전처리

본 연구에서는 노면상태 정보 수집을 위해 핀란드의 Teconer사에서 제공하는 노면센서, RCM411을 수집용 차량에 부착하여 운영하였다(<Fig. 1> 참조). RCM411은 주행 중 노면상태 데이터 수집이 적합한 센서로 노면의 상태와 노면의 상태, 노면 마찰계수, 노면 온도 등을 실시간으로 측정한다(Teconer, 2018). 해당 센서는 핀란드 개발된 해당 센서는 2011-2012년에 성능의 테스트를 마쳤으며 향후 핀란드 기상청에서 계획하고 있는 Road Weather forecast model에 활용할 계획에 있다(Karsisto, 2016). 따라서 본 연구진은 해당 센서를 통해 구분되는 노면상태 정보가 신뢰성을 가진다고 판단하였고 본 연구에서 활용하였다.

이 중, 본 연구에서는 노면의 상태와 노면의 마찰계수를 중점적으로 살펴보았다. 노면의 마찰계수는 0.01

단위까지 측정할 수 있으며 노면의 상태는 건조(Dry), 습윤(Moist), 젖음(Wet), 살얼음(Slushy), 눈(Snowy), 아이스(Icy)의 5가지의 항목으로 분류된다. 노면센서로부터 수집되는 데이터는 스마트폰의 수집 앱을 통해 매 초 단위로 수집하여, 서버로 데이터를 전송하였으며, 시간정보와 공간 좌표를 추출하기 위해 스마트폰의 시간과 GPS 데이터를 함께 추출하였다(Kim et al., 2017).



<Fig. 1> Implementation of RCM411 in probe vehicle

7~8월 중 비가 오는 날의 총 16일의 데이터가 수집되었으며 비가 오는 도중과 비가 온 직후 노면이 마를 때까지 수집용 차량을 운용하여 노면이 젖음에서 습윤, 건조로 바뀌는 과정을 추적하였다. 본 연구의 공간적 범위는 여름철이므로 건조, 습윤, 젖음 3가지 노면상태만이 관측되었으며, 실험이 수행된 날짜와 해당 날에 수집한 노면 상태의 분포는 아래 <Table 1>과 같다. 비가 왔던 날에도 건조상태인 노면이 다수 관찰되는 것을 알 수 있으며, 상황에 따라 습윤이 많은 날과 젖음이 많은 날이 구분되는 것을 확인할 수 있다.

<Table 1> Experiment date and distributions of roadway surface condition

Number	Date	Surface condition		
		Dry	Moist	Wet
1	20170708	1829	203	34
2	20170710	70	38	12489
3	20170715	3141	783	3993
4	20170716	1924	62	0
5	20170717	1498	228	69
6	20170722	2556	229	324
7	20170723	11	11	3502
8	20170724	156	61	3310
9	20170728	3746	1399	2350
10	20170731	2637	588	3674
11	20170806	3325	108	15
14	20170815	0	0	2255
15	20170820	0	1	4159
16	20170821	1318	285	699

노면 상태를 예측하기 위한 설명변수로는 크게 기상관측자료, 도로 기하구조 자료, 교통정보 자료를 활용

하였다. 기상관측자료로는 자동기상관측장비(AWS)에서 분 단위로 온도, 습도, 풍속, 누적강수량을 수집하였다. 기상청, 서울시, 민간기업의 AWS를 모두 활용하였으며, 그중에서도 실험지역인 내부순환로 인근의 29개의 AWS를 사용하였다(<Fig. 2> 참조). 도로의 특정 지점에서의 기상정보는 직접적인 측정이 불가하므로, 29개의 AWS 중 가장 인접한 AWS의 기상정보를 해당 도로의 기상정보로 가정하였다.



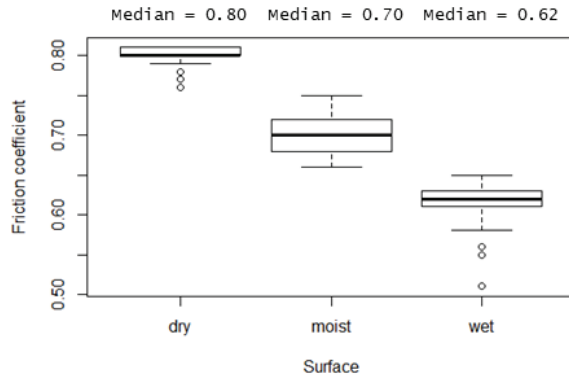
<Fig. 2> Location of AWS(Automatic Weather Station) sensors in the study site

도로의 기하구조 역시 노면의 상태를 결정하는 요인으로 과거 연구들에서 주요하게 다루어졌으며(Liu and Peng, 1996), 강수 시 배수 등에 영향 노면의 상태에 영향을 미치는 요인으로 판단하였다. 이에 도로의 곡선 반경, 종단경사, 횡단 경사를 변수로 활용하였으며, 내부순환로에서 이동형 차량을 이용해 기하구조 값을 측정하였으며, 성산방면, 성수방면 양방향에 대해 각각 추출하였다. 성산방향으로는 총 2,155개의 지점에서 기하구조 정보를 수집하였으며 성수방향으로는 2,336개의 지점에 대해 기하구조 정보를 수집한 후 노면센서를 통해 수집한 노면상태 측정 지점에 매칭하였다.

또한, 도로의 교통량 역시 해당 노면의 상태에 영향을 미치는 요인이라는 과거의 연구 사례를 기반으로(Shi et al., 2011) 노면 상태에 영향을 미치는 변수로 활용하였다. 서울시 교통정보시스템(TOPIS, 2018)에서 제공하는 1시간 단위의 교통량을 사용하였다. 기상정보와 교통량의 경우 현재시점뿐만 아니라 시계열적인 변화도 노면상태에 영향을 미친다고 판단하여 노면상태를 수집한 시점으로부터 직전 12시간까지의 교통량 및 기상정보를 1시간 단위로 수집하였다. 교통량의 경우 표준노드링크 단위로 제공되기 때문에 내부순환로의 표준노드링크 단위로 정보를 수집하여 노면센서를 통해 1초단위로 수집되는 노면상태 자료에 매칭하였다.

### 3. 노면상태와 노면마찰계수 사이의 상관성 분석

본 장에서는 노면센서에서 취득 가능한 노면 온도, 수막 두께, 노면상태(건조, 습윤, 젖음), 마찰계수 중에서 노면상태와 노면의 마찰계수에 주목하여 두 변수 사이의 상관성과 노면상태에 따른 마찰계수의 차이를 분석하였다. 총 16일의 실험 기간 동안 수집된 데이터를 활용하여 노면상태에 따른 마찰계수의 분포를 도출하였으며, 결과는 <Fig. 3>과 같다. 노면 상태에 따라 마찰계수가 뚜렷한 차이를 보이는 것을 알 수 있으며, 건조상태에서 습윤 상태로 갈수록 마찰계수가 점점 감소하는 경향을 보이는 것을 알 수 있다. 또한, 습윤과 젖음 상태와 비교하면 건조상태에서는 마찰계수의 분산이 상대적으로 작은 값을 가지는 것을 알 수 있다.



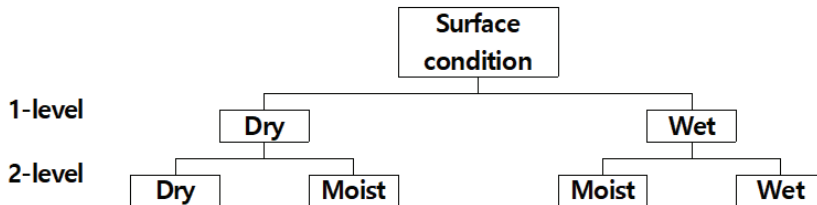
<Fig. 3> Coefficient of friction according to road surface condition

이와 같은 결과는 노면 센서를 통해 수집하는 노면의 상태정보는 마찰계수와 큰 상관성을 가지며, 그러므로 노면의 상태를 추정할 수 있다면 노면의 마찰계수를 추정할 수 있다는 것을 시사한다. 따라서 본 연구에서는 노면의 상태를 나타내는 변수로 3개의 범주로 분류된 노면상태(건조, 습윤, 젖음)를 활용하였으며, 정확한 노면의 상태를 추정하는 데 초점을 맞춰 연구를 진행하였다.

### Ⅲ. 분석방법론 및 결과

#### 1. 분석 방법론

노면상태를 예측하기 위해서 본 연구에서는 2단계 분류모형을 사용하였다. 2단계 분류모형(Two-level classification)은 샘플수가 부족한 class나 특징 차이가 적은 class를 구분하기 위한 기법으로 다양한 분야에서 활용되고 있다(Pouynfar and Sameti, 2014; Weidmann et al., 2003). 여름철 노면상태의 경우 ‘건조’, ‘습윤’, ‘젖음’, 3가지의 상태가 존재하는데 건조와 젖음에 비해 중간 단계인 습윤 상태는 적게 존재한다. 하지만 습윤의 경우 비가 온 후 노면이 언제 마르기 시작하는지를 알려주는 중요한 상태이므로 습윤 상태의 정확도를 향상시키기 위하여 <Fig. 4>와 같이 2단계 분류모형을 구축하였다.



<Fig. 4> 2-level classification model for estimating surface condition

노면 상태예측을 위한 모델로는 다양한 기계학습(Machine learning) 알고리즘을 검토하였다. 기계학습에는 의사결정나무, 판별분석, SVM(Support Vector Machine), Neural Network, KNN(K-Nearest Neighbors) 등 다양한 모델이 존재한다(Marsland, 2011). 본 연구에서는 학습 데이터의 이상치에 대한 대응력이 높고, 전체적인 분

산을 감소시켜 예측의 정확도를 높일 수 있는 앙상블트리 기법을 이용하였다. 앙상블 기법은 다수의 의사결정나무를 학습시키고, 이들의 분류결과를 종합하여 최종결과를 도출하는 방법론이다. 다수의 의사결정나무를 학습시키기 위한 방법들은 Random forest, Robust boosting, Random undersampling boosting, Random subspace 등이 존재하며, 이 중 가장 적합한 분류모델을 선정하기 위해 Matlab에서 제공하는 앙상블 매개변수 최적화 함수(OptimizeHyperparameters)를 이용하였다. 매개변수 최적화 함수는 각 앙상블 기법별로 학습률과 학습 트리 수를 조합하여 변수 최적화를 수행한다. 그 결과 본 연구에 사용된 데이터에 대해 정확도가 가장 높은 Random forest를 최종 모델로 결정하였다. Random forest는 기존 배깅(Bagging) 모형에 변수의 임의성을 추가한 기법이다(Breiman, 2001). 배깅 모형은 훈련 데이터로부터 크기가 같은 표본을 단순화를 반복추출하여 각각에 대해 의사결정나무를 생성하는 기법으로, 학습 데이터의 임의성을 부여한다. Random forest는 배깅 모형과 다르게 표본을 추출할 때 이용하는 변수의 종류도 임의 추출하여 의사결정나무를 학습시킨다. 이를 통해 학습된 모델에 변수의 임의성을 추가할 수 있어 모델의 다양성을 극대화하고 과적합 방지 및 이상치 대응력을 높일 수 있다. Random forest에서 추출 변수 수는 일반적으로 분류문제에 정확도가 높아 많이 사용하는 전체 샘플수의 제곱근을 이용하였다.

모델의 학습을 위해서 총 16일의 데이터 중 14일의 데이터를 활용하였고, 구축된 모델의 검증을 위해서 16일 중, 7월 31일과 8월 20일의 2일 치 자료를 활용하였다. 최종적으로는 검증을 위한 데이터에서 출력된 추정 정확도를 모델의 판별 기준으로 사용하였다.

투입 변수로는 기상 관측 정보, 도로 기하구조 정보, 교통정보가 사용되었다. 1분 단위로 수집되는 기상정보를 10분 단위로 합산하였으며, 연속적인 기상 상황의 변화와 노면상태 사이의 관계를 반영하기 위해 직전 12시간까지의 기온, 강수량, 풍속, 습도 변수를 투입 변수로 활용하였다. 기하구조 정보로는 도로의 곡선반경, 횡경사, 종경사를 투입 변수로 설정하였으며 교통환경 변수로는 직전 1시간의 교통량을 변수로 설정하였다. 출력변수로는 노면센서로 측정된 노면상태 정보 값을 사용하였으며, 범주형 변수로 ‘젖음’, ‘습윤’, ‘건조’ 3가지의 노면상태 값이 존재하였다. 최종적인 모델구축에 활용한 데이터는 <Table 2>와 같다.

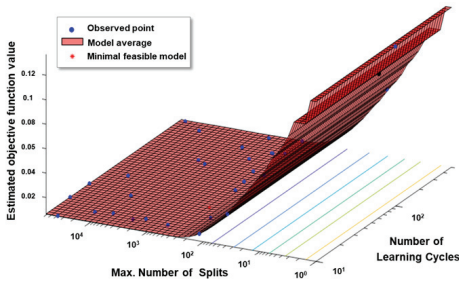
<Table 2> Input and output variable for classification

Type	Category	Unit	Variable name
Input variable	Weather information	°C	Temperature per 10 minutes (last 12 hours)
		mm	Precipitation per 10 minutes(last 12 hours)
		m/s	Wind speed per 10 minutes (last 12 hours)
		%	Humidity per 10 minutes (last 12 hours)
	Road geometry	%	Cross slope
		%	Vertical slope
		m	Curvature
Traffic information	km/h	Link speed per 1 hour (last 12 hours)	
Output variable	Road surface condition	Categorical	Road surface condition: Dry, Moist, Wet

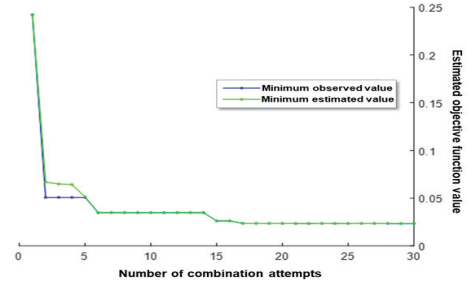
## 2. 분석 결과

Random forest에서 파라미터 최적화를 위해 조절 가능한 구성변수로는 앙상블에 활용하는 트리의 개수와 각 트리 구성 시 제약조건인 최대분기 수로, 각 값의 조합에 따라 모델의 정확도를 평가하여 가장 좋은 구성을 찾아 모델 파라미터를 최적화하였다. 일반적으로 트리의 개수가 증가할수록 모델의 정확도가 증가하는 경

향을 보이거나 과도하게 많을 시에는 학습 시간이 오래 걸리며 모델 자체의 크기도 증가하여 비효율성이 증가하므로 적합한 트리 개수의 선정이 중요하다. 개별 분류트리의 경우 분기가 많을수록 주어진 데이터 세트에 대한 분류 정확도가 증가하나, 지나치게 많은 분기가 만들어질 경우 과적합(Overfitting)이 발생하게 된다. 이에 각 트리의 일반성을 최대한으로 높이기 위해선 적합한 분기 수를 제약하여 학습시켜야 한다. 이에 각 값의 조합을 베이지안 최적화(Bayesian optimization) 기법을 이용하여 최적 조합을 찾았다. <Fig. 5>에서 각 조합 값에 따른 모델의 정확도를 나타내고 있다. 전체 30번의 조합을 이용하였으며, <Fig. 6>에서 볼 수 있듯이 약 20번 이후부터는 모델의 개선이 두드러지지 않아 30회 이상의 조합은 관측 필요성이 없다고 판단하였다.



<Fig. 5> Accuracy of model by combination



<Fig. 6> Model accuracy by combination attempts

2단계 분류모형 중 1단계 분류 모델의 학습에서는 습윤 자료를 제외한 건조/젖음 노면상태만을 학습시키고 검증에는 모든 노면상태를 포함시켜 분류 결과를 확인하였다. 최적의 조합을 이용한 1단계 분류모형의 검증 결과는 <Fig. 7>과 같다. 습윤 상태는 1단계 분류 모델에서는 학습하지 않아 검출이 불가능하나, 습윤 상태가 차지하는 비율이 낮으므로 전체 정확도는 93.62%를 보이는 것을 알 수 있다. 습윤 상태의 경우 표본 대부분이 건조상태로 분류된 것을 확인할 수 있으며, 이는 습윤 상태의 노면이 건조상태와 유사한 값을 가짐을 알 수 있다.

2단계 분류에서는 젖음/습윤 상태와 건조/습윤 상태 각각에 대해 모델을 학습시켰다. 학습 데이터를 구성함에 있어서 습윤 상태의 검출 정확도 향상을 위해서 최적화된 구성변수 조합 학습 과정에서 습윤 상태에 대한 검지 가중치를 변형시켜 가장 최적의 조합을 찾는 과정을 통해 최종 모델을 도출하였다. 최종 모델 결과는 <Fig. 8>과 같다. 전체 모델의 정확도는 1단계 분류모형보다 10%정도 낮은 81.74%지만, 습윤 상태 검출률이 4.29%에서 52.97%로 크게 향상하였다.

**First-level Trees**  
-Test accuracy: 93.62%-

True class	Dry	2639	0	0
	Moist	536	0	70
	Wet	338	0	11214
		Dry	Moist	Wet
		Predicted class		

<Fig. 7> Confusion matrix of 1-level classification

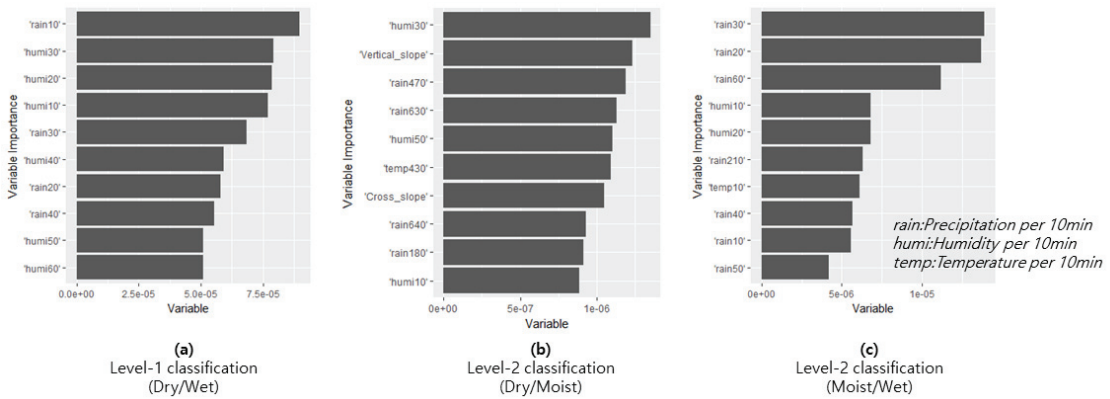
**Two-level Trees**  
-Test accuracy: 81.74%-

True class	Dry	1411	1228	0
	Moist	261	321	24
	Wet	222	967	10363
		Dry	Moist	Wet
		Predicted class		

<Fig. 8> Confusion matrix of 2-level classification

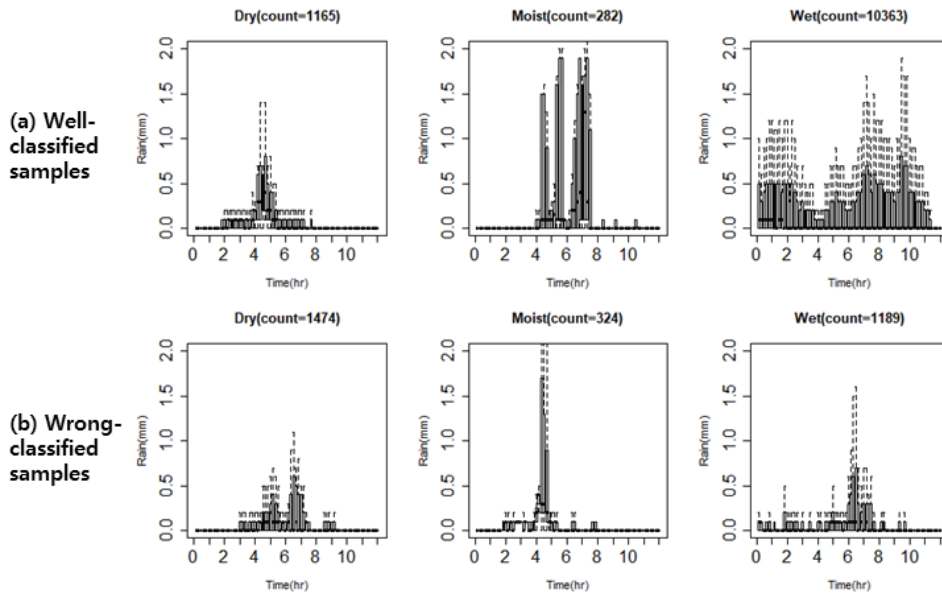


투입된 변수들의 중요성을 평가하기 위해 각각의 분류모델에서의 변수중요도(Variable Importance)를 산출하였다(<Fig. 9> 참조). 건조와 젖음을 분류하는 첫 번째 모델(First-level tree)에서는 직전 10분의 강수량이 가장 중요한 요인인 것을 알 수 있으며 1시간 이내의 습도와 강수량이 노면 상태를 결정하는 가장 큰 요인인 것을 알 수 있다. ‘습윤’ 상태를 분류하는 두 번째 모델(Two-level tree)에서는 건조/습윤을 분류하는 모델과 습윤/젖음을 분류하는 모델의 변수중요도가 상이한 것을 확인할 수 있다. 습윤/젖음을 분류하는 모델의 경우 첫 번째 모델과 유사한 기상 변수들을 중요한 분류변수로 가지지만(<Fig. 8-(c)> 참조), 건조/습윤을 분류하는 모델의 경우 습도를 제외하고는 유사성이 없는 것을 볼 수 있다. 이는 습윤의 상태가 건조와 유사한 기상 특성을 가지는 것을 시사한다. 기존의 1단계 분류 모형에는 없었던 도로의 횡단경사와 종단경사가 중요 변수인 것을 볼 수가 있는데 이는 건조와 습윤의 경우 기상요인보다는 도로의 기하구조 및 기타 다른 특성에 영향 받는다는 것을 의미한다.



<Fig. 9> Variable Importance of 2-level classification model

어떠한 요인으로 인해 분류가 실패하고 오차가 발생하는지 보다 자세히 분석하기 위해 분류가 잘 된 표본 (True-Positive & False-Negative)과 그렇지 않은 표본(True-Negative & False-Positive)을 분리하여 각각에 대해 시간대별 강수량을 <Fig. 10>과 같이 도출하였다. <Fig. 10-a>는 모델에 의해 노면의 상태가 정확하게 분류된 표본들의 직전 12시간 강수량을 나타내며 <Fig. 10-b>는 모델이 추정에 실패한 표본들의 직전 12시간 강수량을 나타낸다. 그림에서 확인할 수 있듯이 정확히 분류된 샘플의 경우 젖음 상태일때와 건조/습윤 상태일 때의 강수량의 차이가 매우 큰 것을 알 수 있다. 반면 잘못 분류된 샘플의 경우 젖음 상태일 때의 강수량이 상대적으로 낮으며, 건조/습윤과 큰 차이를 보이지 않는 것을 확인할 수 있다. 이는 <Fig. 9>에서 볼 수 있듯이 강수량이 노면 상태를 결정하는데 매우 중요한 변수라는 것을 나타내며, 강우가 적으며 노면이 젖은 상태일 때 오차가 발생하는 것을 보여준다. 반면 건조와 습윤의 경우 강수량의 차이가 두드러지지 않는 것을 볼 수 있는데 이는 강우 변수만을 가지고는 건조/습윤 상태의 판별이 힘들다는 것을 시사한다.



<Fig. 10> Last 12 hours precipitation by surface condition by classification result

#### IV. 결 론

본 연구에서는 기계학습을 통해 도로의 노면상태를 추정하는 모델을 개발하였다. 노면 상태의 수집을 위해 실험 차량에 노면센서를 부착하여 ‘건조’, ‘습윤’, ‘젖음’, 3가지 범주로 구분된 노면상태 정보를 수집하였고, 이를 추정하기 위한 투입 변수로 도로의 기하구조 정보, 교통정보, AWS를 통해 수집 가능한 기상정보를 활용하였다. 노면 상태를 예측하기 위한 알고리즘으로는 다양한 기계학습 알고리즘이 검토되었으며 그 중, 가장 높은 정확도를 보인 ‘Random forest’를 기반으로 한 2단계 분류모형을 구축하였다. 총 16일의 실측 데이터 중 14일의 데이터를 모델을 학습하는 데 활용하였고, 2일의 데이터를 모형의 정확도를 검증하기 위해 사용하였다. 그 결과 81.74%의 검증 정확도를 가지는 노면상태 예측 모델을 구축하였다.

본 연구의 결과는 기상청에서 관측하는 기상정보로 도로의 노면상태 추정할 수 있다는 가능성을 보여준다. 새로운 장비나 센서를 설치하지 않고도 기존의 기상 관측 정보와 교통정보 등을 활용하여 노면의 상태를 80% 이상까지 추정할 수 있다는 것을 본 연구를 통해 증명하였다. 하지만 정확도의 향상에 있어서 한계점을 가지는 것을 알 수 있었다. 실제로 비슷한 기상 관측값을 가지지만 노면의 상태가 항상 동일하지 않은 것을 알 수 있었고, 20% 정도의 오차를 보였다. 이는 도로의 상태는 기상 상황뿐만 아니라 배수 시설, 교통량 등 다른 요인들에 의해서도 영향을 받을 수 있으며, 이러한 다양한 외부적 요인에 의해 동일한 기상 상황에서도 노면의 상태가 달라질 수 있다는 점을 시사한다.

본 연구는 산발적으로 설치되어 있는 기상관측 장비를 이용해서 도로 노면의 상태를 추정하였다는 점에서 그 의미가 있으며, 관측 차량에 의존하는 기존의 방법보다 더 넓은 공간적, 시간적 범위에서 노면 정보를 취득할 수 있다는 점에서 의의가 있다.

본 연구의 한계 및 향후 개선사항을 정리하면 아래와 같다. 첫째, 자료수집 과정에서 본 연구는 여름철 기

상 상황에만 초점을 맞추어 연구를 수행하였다. 향후 가을/겨울철 정보를 추가하여 모델을 고도화한다면, 결빙 상태의 노면 예측 또한 수행할 수 있을 것으로 기대된다. 둘째, 본 연구에서는 머신러닝 기법 중 Random forest 알고리즘을 사용하여 도시화 고속도로에 대해서 노면상태 추정 모형을 개발하였다. 향후 일반 국도로 까지 연구의 공간적 범위를 확대한다면 최적의 알고리즘과 파라미터가 달라질 수 있으며 이를 검토할 필요성이 있다. 셋째, 본 연구에서 사용한 입력 변수들은 10분단위 12시간동안의 측정값과 기하구조 정보를 이용하였다. 입력 정보량이 많아 모델의 과적합이 발생할 가능성이 있어 향후 연구의 시공간적 범위 확대시 입력 변수에 대한 차원감소 기법 적용을 통해 모델 고도화가 필요하다. 넷째, 본 연구의 노면상태 추정 정확도는 약 80%로 여전히 20%의 오차를 가지는 것을 확인할 수 있다. 이는 기상관측값이 넓은 범위에 걸쳐 동일한 값을 가지고, 따라서 보다 세밀한 단위에서의 기상정보를 수집한다면 더 높은 정확도로 노면의 상태를 추정할 수 있을 것으로 기대된다.

## ACKNOWLEDGMENTS

본 연구는 기상청 「기상·지진 See-At기술개발연구(KMIPA-2017-5020)」의 지원으로 수행하였습니다.

## REFERENCES

- Breiman L.(2001), "Random forests," *Machine learning*, vol. 45 no. 1, pp.5-32.
- Brodsky H. and Hakkert A. S.(1988), "Risk of a road accident in rainy weather," *Accident Analysis & Prevention*, vol. 20, no. 3, pp.161-176.
- Chapman L., Thornes J. E. and Bradley A. V.(2001), "Modelling of road surface temperature from a geographical parameter database," *Part 2: Numerical. Meteorological Applications*, vol. 8, no. 4, pp.421-436.
- Choi S., Kim M., Oh C. and Lee K.(2013), "Effects of weather and traffic conditions on truck accident severity on freeways," *Journal of the Korean Society of Civil Engineers*, vol. 33, no. 3, pp.1105-1113.
- Edwards J. B.(1998), "The relationship between road accident severity and recorded weather," *Journal of Safety Research*, vol. 29, no. 4, pp.249-262.
- Hoogendoorn R. G., Tamminga G., Hoogendoorn S. P. and Daamen W.(2010), "Longitudinal driving behavior under adverse weather conditions: Adaptation effects, model performance and freeway capacity in case of fog. In Intelligent transportation systems," *2010 13th international ieee conference on IEEE*, pp.450-455.
- Jeong E., Oh C. and Hong S.(2013), "Prediction of speed by rain intensity using road weather information system and vehicle detection system data," *The Journal of the Korea Institute of Intelligent Transport Systems*, vol. 12, no. 4, pp.44-55.
- Karsisto V. and Nurmi P.(2016), "Using car observations in road weather forecasting," In *International Road Weather Conference*.
- Keay K. and Simmonds I.(2005), "The association of rainfall and other weather variables with road

- traffic volume in Melbourne, Australia,” *Accident Analysis and Prevention*, vol. 37, no. 1, pp.109-124.
- Kim D. G., Kim J. Y., Lee J. H., Choi D. G. and Kweon I. S.(2010), “Utilizing Visual Information for Non-contact Predicting Method of Friction Coefficient,” *Journal of the Institute of Electronics Engineers of Korea SP*, vol. 47, no. 4, pp.28-34.
- Kim J. H. and Won J. M.(2013), “A Development of The Road Surface Decision Algorithm Using SVM (Support Vector Machine) Clustering Methods,” *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 12, no. 5, pp.1-12.
- Kim S., Kim K., Lee M., Cho Y. and Hahm Y.(2017), “Improving Vehicle Safety Through Road Surface Condition and Weather Data Analysis,” *Journal of Information Technology and Architecture*, vol. 14, no. 4, pp.375-386.
- Knapp K. K. and Smithson L. D.(2000), “Winter storm event volume impact analysis using multiple-source archived monitoring data,” *Transportation Research Record*, 1700, pp.10-16.
- Kwon T., Fu L. and Jiang C.(2013), “Effect of winter weather and road surface conditions on macroscopic traffic parameters,” *Transportation Research Record: Journal of the Transportation Research Board*, no. 2329, pp.54-62.
- Liu C. S. and Peng H.(1996), “Road friction coefficient estimation for vehicle path prediction,” *Vehicle System Dynamics*, vol. 25, no. S1, pp.413-425.
- Malmivuo M.(2011), “Friction Meter Comparison Study 2011,” Abstract. Original report Kitkamittareiden vertailututkimus.
- Marsland S.(2011), *Machine learning: an algorithmic perspective*, Chapman and Hall/CRC.
- Morgan A. and Mannering F. L.(2011), “The effects of road-surface conditions, age, and gender on driver-injury severities,” *Accident Analysis & Prevention*, vol. 43, no. 5, pp.1852-1863.
- Pouyanfar S. and Sameti H.(2014), “Music emotion recognition using two level classification,” In *Intelligent Systems (ICIS), 2014 Iranian Conference on, IEEE*, pp.1-6.
- Rahman A. and Lownes N. E.(2012), “Analysis of rainfall impacts on platooned vehicle spacing and speed,” *Transportation Research Part F: Traffic psychology and behaviour*, vol. 15, no. 4, pp.395-403.
- Shao J. and Lister P. J.(1996), “An automated nowcasting model of road surface temperature and state for winter road maintenance,” *Journal of Applied Meteorology*, vol. 35, no. 8, pp.1352-1361.
- Shi L., Cheng Y., Jin J., Ran B. and Chen X.(2011), “Effects of rainfall and environmental factors on traffic flow characteristics on urban freeway,” *Transportation Research Board 90th Annual Meeting*, at Washington, D.C.
- Teconer, [http://www.teconer.fi/downloads/RCM\\_Datasheet\\_en.pdf](http://www.teconer.fi/downloads/RCM_Datasheet_en.pdf), 2018.08.27.
- Weidmann N., Frank E. and Pfahringer B.(2003), “A two-level learning method for generalized multi-instance problems,” In *European Conference on Machine Learning*, Springer, Berlin, Heidelberg, pp.468-479.