Image-based Road Surface Condition Detection Using Transfer Learning

Eleni Aloupogianni, Faiyaz Doctor, Charalampos Karyotis, Raymond Tang, and Rahat Iqbal

***Abstract*—Accurate prediction of road surface conditions can help authorities manage vehicular transportation effectively in large cities by helping to reduce congestion and the risk of accidents due to adverse weather. Image-based classification, using Convolutional Neural Networks (CNN) in combination with Transfer Learning (TL), can provide a real-time, data-driven and cost-effective solution for classifying road surfaces under varying weather, traffic and image recording conditions. This paper proposes an image classification approach leveraging TL with the ResNet50 architecture, enabling the efficient utilisation of pre-existing knowledge within deep neural networks to facilitate rapid model adaptation without extensive dataset collection. This study focuses on the challenging tropical conditions of Singapore as a use case where the performance of the proposed approach is evaluated on two distinct video footage/image datasets, namely fixed expressway cameras and dashcams. The strategy of coupling pre-trained models with TL consistently converges to better results more rapidly compared to training from scratch or with fine-tuning. The results provide valuable insights for traffic management authorities in selecting scalable architectures and training strategies for big data curation from traffic cameras, considering computational constraints for real-world deployment.**

***Index Terms*—road analysis, artificial intelligence, transfer learning, weather analysis, intelligent transport system.**

I. Introduction

THE effects of climate change, particularly in regions such as South East Asia, are causing an increasing frequency of adverse weather events such as typhoons and storms resulting in heavy rain and flooding that affect both rural and urban areas [1]. The state of roadways can significantly impact driving decisions and being able to accurately identify road surface conditions is vital for advancements in self-driving cars and monitoring systems supporting city authorities, for ensuring continuous traffic flow and road safety among diverse road users. The condition of the road surface, whether it is wet or dry, can influence vehicle speed, and is directly related to the risk of road accidents [2]. Moreover, accurate determination of weather and associated road conditions is a necessary feature for Digital Twin implementations of Intelligent Transportation Systems (ITS) [3]–[5]. The information obtained regarding road conditions serves both as a warning to drivers and a prompt for road maintenance personnel [6].

Meteorological stations and sensors are traditionally used to measure inclement weather conditions [7]. However, these technologies introduce complexities due to the need to install large sensor grids to cover a target area along with long-term hardware and data management requirements, resulting in high overall costs for their installation and maintenance. Implementing a road surface identification system using these data sources introduce several challenges. These relate to aligning and fusing together multiple data streams, extracted features that may be at a lower resolution and locationally uncorrelated for accurately classifying surface conditions in real-time at specific locations along a road. A low-cost, data-driven and quick-to-implement alternative for the identification of road surface conditions is image-based classification. The increased availability of image data from expressway cameras and dashcams can support this alternative. This enables a standalone, real-time classification based on a single frame, without the need for additional knowledge [8]. In this way, instead of constructing new weather monitoring stations, already existing grids of expressway cameras can be utilised together with using dashcams to increase resolution and coverage. However, manual and human-based image classification by monitoring personnel can be inconsistent and time-consuming. Challenges arise such as motion blur or low-brightness [9], difficulty in differentiating heavy rain from fog [7], or distinguishing a wet road surface from a shadow or a liquid spillage on the road, particularly with low-resolution input images. Additionally, there has been limited prior research investigating the issue of concurrent atmospheric phenomena, like rain within a storm or sunlight during rain, which cannot be adequately addressed by traditional image classification methods [10], [11].

Computer vision and Convolutional Neural Networks (CNN)-based algorithms have shown impressive abilities in image classification. However building a classification model from scratch requires substantial time until it is production-ready. These algorithms typically require a significant volume of labelled training data and computational resources to achieve a generalised solution [12], [13]. Although extensive traffic camera networks can amount to increasing data size, resulting in massive datasets [14], the available of such data is limited during the initial stages of data collection and model development. To speed up the process, model weights pretrained on giant datasets can be used as initial weights, instead of the weights being randomly initialised. The model

E. Aloupogianni is with Interactive Coventry Ltd, Coventry, UK (email: eleni@interactivecoventry.com).

F. Doctor is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK, and also with Interactive Coventry Ltd, Coventry, UK (email: fdocto@essex.ac.uk).

C. Karyotis is with Interactive Coventry Ltd, Coventry, UK (email: charalampos.karyotis@interactivecoventry.com).

R. Tang is with the Citymatrix Pte Ltd, Kuala Lumpur, Malaysia. (email: enquiry.thecitymatrix@gmail.com)

R. Iqbal is with Interactive Coventry Ltd, Coventry, UK (email: r.iqbal@interactivecoventry.com).

Manuscript submitted on March 19, 2024. (Corresponding author: E. Aloupogianni.)

can then be trained through several epochs, where all model weights are updated, in a process called Fine-Tuning (FT) [15]. However, large processing times and overfitting errors can still occur when training all layer weights, especially for deep networks. Transfer Learning (TL) [16] offers a solution to this issue, addressing challenges stemming from lengthy training times and limited data availability. TL, also known as domain adaptation, leverages knowledge gained from training a model on a sizable dataset for a related task, and transfers this knowledge to a new, potentially smaller dataset or different task [17]. This approach speeds up network convergence, enabling quicker and more efficient generalisation to new data. TL is particularly valuable in real-world applications where rapid model deployment is crucial. There is commonly an overlap between TL and FT definitions in the literature. Here, we define FT as re-training the entire weight network of a model on a new dataset for a couple more epochs. Similarly, we define TL as re-training only the final layers of the network, the ones that handle decision making. By combining pretrained CNNs with TL, this paper demonstrates this novel hybrid approach can provide a cost-effective and fast trainable alternative for classifying road conditions, where these models are evaluated on real road activity data elicited over varying weather conditions in Singapore.

To address challenges in the urban monitoring domain efficiently and economically, this study advocates employing TL to swiftly train a deep learning model for accurate road surface classification from road images. The primary goal of this research is to develop a robust and efficient system for the real-time classification of weather conditions from road images by leveraging TL. Specifically, our contributions are as follows:

- Creating a deep learning model capable of accurately differentiating various conditions on road surfaces (wet and dry), while overcoming challenges such as motion blur, low-resolution images, and nighttime scenarios,
- Accurately classifying weather conditions on roads and expressways in real-time, thereby enhancing road safety and transportation efficiency,
- Exploring the unique and challenging tropical climate conditions of Singapore city-state on road surface condition detection,
- Contributing to the field of computer vision and deep learning by demonstrating the effectiveness of TL in rapidly developing models for specialised image classification tasks, with a focus on road conditions.

This research makes a novel contribution to the field by combining TL with advanced neural network models to address the specific challenges of weather condition classification on roads, a problem often overlooked in prior studies. While specialised previous works are scarce, they are primarily focused on traditional CNN and RNN networks trained from scratch, instead of more efficient pretrained CNN based architectures. Our approach focusing on using city road images targeted on South-East Asia leverages TL to overcome common model development limitations like lengthy training times, data scarcity, and the complexity of distinguishing intricate road features. This pioneering solution facilitates real-time road weather classification contributing to road safety and transportation efficiency. Two new datasets were collected and preprocessed for the investigation of this task. Another novel concept was the integration of the detection tool into an edge computing ecosystem, where devices like static cameras, dashcams, servers, and processors work together to monitor road surface and traffic conditions. The performance of the proposed models was evaluated through validation and runtime comparisons against alternative architectures. Ultimately, this research aims to offer valuable insights into the application of TL for specialised image classification tasks, supporting rapid model deployment in real-world contexts.

II. Related Works

The existing body of literature has highlighted several approaches for analysing road scenes through image analysis. Road condition detection can be thought of as a form of anomaly detection and in recent years image analysis approaches mostly based on CNNs, Recurrent Neural networks (RNNs) and masked auto-encoders have been proposed in [18]–[20] for video analysis. The majority of published work targets road analysis for driving assist systems, autonomous vehicles [21], [22], road maintenance [23], [24], flooding [25] and parking space [26] detection. Recent works applying deep learning frameworks for road traffic detection include models based on R-CNN for vehicle detection using segmentation [27], CNNs and Vision Transformers (ViTs) for weather condition prediction [28], and Generative Adversarial Networks (GANs) for data augmentation in pothole detection under adverse conditions [29]. CNN-based deep neural architectures have been proposed for weather classification for general images [10], [30]. Notably, a VGG19-based model trained from the ground up using road CCTV and dashcam images exhibited promising performance [31]. Similarly, a ResNet architecture, trained from scratch on a novel dataset of general images, displayed favorable outcomes [32]. In other works, YOLOv8 with visual transformers has been used for enhanced pothole detection [33], while a semi-supervised system using convolutional LSTM was developed for wet road event detection from surveillance cameras [34]. Such bottom-up approaches however necessitate substantial data, extended training periods, and significant fine-tuning to achieve optimal results.

The utilisation of TL for image-based weather classification has been sparsely explored. Prior work has employed TL in conjunction with pretrained Xception [35], [36], EfficientNet [37] or ResNet [38], [39] models. These models, whose weights are pretrained on ImageNet, coupled with data augmentation have shown impressive performance. During the Weather4cast competition, RNN and U-Net architectures have demonstrated superior capabilities in the prediction of up to 8 hours of weather from video frames with high temporal and spatial resolutions compared to other models [40]. The training datasets used in previous works predominantly consisted of general scenes, with images scrapped from the internet often emphasising the sky rather than on-the-road conditions. While

TL with conventional image classification models has been applied, comprehensive implementation details and validated results remain elusive. Furthermore, existing datasets lean towards generality, lacking a focus on road conditions and nighttime scenarios.

III. Road Surface Condition Modelling

A. Transfer Learning

TL is a machine learning technique where a pre-trained neural network model, typically trained on a large and diverse dataset, is leveraged as a starting point for solving a different but related task [16]. Instead of training a neural network from scratch for the new task, TL involves using the knowledge and features learned by the pre-trained model as a foundation. The model's weights and architecture are then adapted on a smaller dataset specific to the new task. For classification and prediction systems to be integrated into an ITS, they need to a) be able to receive training with unstructured, moderately annotated and limited numbers of data samples, and b) be easily updatable and integrate information quickly, in order to adapt to changing road conditions. The benefits of TL are multiple. By starting with a pre-trained model and adapting only a portion of its weights, TL significantly reduces the time required to train a model from scratch. TL can mitigate the impact of limited training data, as the model has already learned generic features from a large dataset. Moreover, TL helps the model generalise better to new data, as it has learned relevant features from a diverse source. This accelerated training process aligns with the demands of industry, where rapid model deployment is essential for practical applications.

A variety of TL implementations have been proposed [16] In this study, Inductive TL and Feature-based TL are used. Inductive TL refers to the scenario where the source and target tasks are different, but the source model's knowledge is leveraged to improve the performance of the target task. In this context, we adapt pretrained models on a new dataset, allowing the model to adjust to the specifics of the new task while benefiting from the general features it has already learned. During inductive TL, the minimisation to be approximated is the effective extraction of features that can reduce the divergence between the source domain and the target domain. Feature-based TL involves using the learned features from a pretrained model as a fixed feature extractor. Instead of retraining the entire model, only the final layers are retrained to adapt to the new task, while earlier layers retain the general features learnt from the original training. The data requirement is an extensive collection of labelled data from one or more base domains, along with a relatively small quantity of labelled data sourced from the target domain.

Figure 1 presents different pathways for training a model which are further elaborated. In the initial phase, an existing model architecture is selected. Weights can be randomly initialised using e.g. Uniform Xavier initialisation. Alternatively, pretrained weights can be used, where common base models have typically been trained on a large and diverse dataset, such as the ImageNet dataset. This enables them to learn a generic set of features that can be useful for a wide range of tasks. In

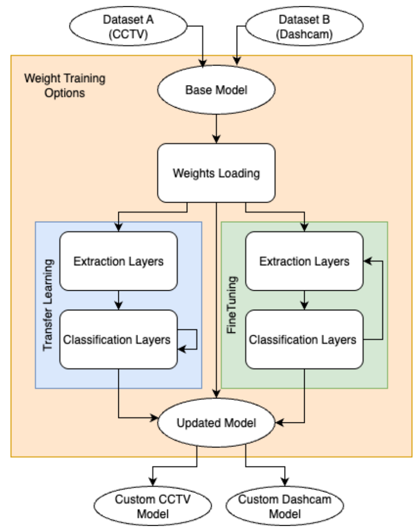


Fig. 1. A flowchart of the process of training a model with Transfer Learning and Fine Tuning.

the case of TL, the pre-trained model is utilised as a feature extractor. The earlier layers, which capture low-level features like edges and textures, are retained, while the later layers, which capture higher-level features specific to the original task, are discarded or updated. A new classification structure, often consisting of one or more fully connected layers, is added on top of the feature extraction layers. Similarly, FT can be performed, which is more resource-intensive involving re-training all or some of the earlier layers of the pre-trained model with the new data from the target domain. This allows the network to adapt its learned features to the task. In this sense, TL can be considered essentially a sub case of FT where only a specific subset of the weights are tuned.

B. Data Sources

Two independent datasets were used to evaluate the road surface detection problem using TL. Dataset A, the Expressway Image dataset, contains images captured by traffic monitoring cameras on Singapore's expressways. These images typically show less than 20% of the sky, depending on the camera angle. Dataset B - Dashcam Image dataset contains images sourced from dashcam footage on the internet with approximately 40% of the sky visible in the recorded scenes. The dashcam images although angled to capture more of the sky, have a closer proximity to the road surface for capturing finer grained details over a narrower field of view of the road directly in front of the vehicle. In contrast the expressway images are angled to capture more of the road however are positioned at a greater distance. This enables them to capture larger field of view of the road although at a lower resolution of detail. By utilising both datasets, we ensured that the evaluation captures a range of visibility conditions relevant to road surface detection. Further details of each of these datasets are described as follows.

1) Dataset A - Expressway Image dataset: Traffic images from expressway cameras are recovered from Datamall [41], a collection of dynamic datasets related to public transport and traffic metrics for Singapore. The expressway camera positions are static and the datasets are available through the official API of the Singapore Land Transport Authority (LTA) and are maintained, curated and monitored daily. Because the application takes into consideration weather data, OpenWeather's API [42] is used for weather data collection for Singapore city. Weather data, specifically weather descriptions, served as a ground truth label for the images. To match weather data with the images, only cameras located close to the weather station were selected, as shown in Figure 2. Data collection took place at irregular time periods between 2023-06-09 and 2023-09-10. The raw dataset consisted of 6081 weather description records (for label generation) and images (for classification). Class labels for moisture concentration levels on the road were identified as ['No', 'Low', 'High']. Images captured every 5 minutes at time points corresponding to weather descriptions 'Clear' or 'Clouds' were assigned the 'No' moisture concentration label. Images associated with weather descriptions 'Rain' and 'Thunderstorm' were assigned the 'Low' and 'High' labels, respectively. For the recovered dataset, weather conditions ranged as follows: $Temperature = [298.18, 305.76] \deg F$, $Pressure = [1007, 1012]Pa$, $Humidity = [57, 94]$, $WindSpeed = [0.51, 8.75]$ and $CloudCoverage = [20, 75]\%$.

In Dataset A, 'Low' and 'High' moisture concentration classes were severely underrepresented compared to 'No'. To address this the following Data Augmentation process was implemented to make the class distribution more balanced. 'Low' and 'High' classes were augmented to reach approximately 2000 sample images each. To preserve useful weather-related image features, the following augmentation transforms were used: Random Horizontal Flip, Random Rotation up to $[-10, +10]$ degrees and Random Perspective Adjustment. Brightness, color and blur were not augmented, because the differences between humidity and rain are minute, and such features might be irreparably obscured. On each image pending augmentation, all transforms were applied with a probability 50%. This resulted in 7998 samples, distributed as 4000 'No', 2000 'Low' and 1998 'High' moisture concentration labeled images, as shown in 3a. An example training batch is also shown in Figure 4.

2) Dataset B - Dashcam Image dataset: Free and available video footage from dashcams, which was recovered from the internet using the terms 'dashcam Singapore', was used to create a custom dashcam image dataset. Footage was classified according to the weather conditions on the road where the same process of using acquired weather descriptions (described in section III.B.1) was used to label the road condition states. Every tenth frame was recovered from the footage as an image sample. Footage that included large text, logos, etc was discarded. The resulting dataset consisted of 18686 labels and video frames, sampled at approximately 24 frames per second (FPS), with the descriptions shown in 3b. The dataset samples were distributed into 6926 'No', 8452 'Low' and 3308 'High' moisture concentration labeled images. An example training batch is shown in Figure 5.

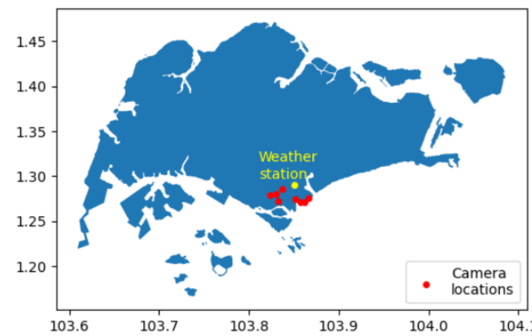


Fig. 2. Location of expressway cameras, used as data sources, relative to the weather station.

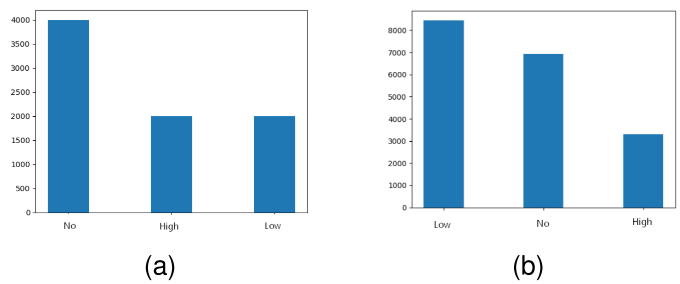


Fig. 3. The distribution of classes for a) Dataset A of expressway images and b) Dataset B of dashcam images.

C. Models

Based on the literature review, the prominent ResNet and EfficientNet architectures, pre-trained on the ImageNet dataset were used as base models. ResNet is a deep neural network architecture known for its exceptional performance in image recognition tasks. It introduces the concept of residual blocks, which contain shortcut connections that allow gradients to flow more effectively during training. This design alleviates the vanishing gradient problem, enabling the training of ex-

(a) Without Data Augmentation

(b) With Data Augmentation

Fig. 4. Example of images in a batch for Dataset A.

Fig. 5. Example of images in a batch for Dataset B.

tremely deep networks with hundreds of layers. ResNet's skip connections ensure that the network can learn both low-level and high-level features, making it highly effective for various computer vision tasks. In contrast, EfficientNet is an architecture designed with a focus on efficiency and scaling. It uses a compound scaling method that optimizes model depth, width, and resolution simultaneously to achieve state-of-the-art performance while reducing computational demands. This approach allows EfficientNet models to strike a balance between model size and accuracy, making them efficient choices for resource-constrained environments. These specific architectures were chosen based on their proven track record in similar research, including works by [37] and [38].

IV. Model Evaluations

A. Experiments

For each dataset, 70% of the samples were used for training, 20% for validation and 10% for testing of the various models. Data were split in train-validation-split datasets in a stratified manner, so that a balanced number of all classes existed in all subsets. Each model and architecture was trained on the training data and evaluated on the validation data. During the validation stage, training parameters were minimally tuned. Performance was evaluated in terms of classification accuracy (defined as the percentage of correct predictions made divided by the total number of predictions made), F1-score, and confusion matrix. The models with the best performance were then trained on the union of the training and validation data and evaluated on the test data.

For each model architecture, some or all of the following training strategies were compared:

(a) Loading randomly initialised weights with Xavier initialisation and training the model from scratch,
(b) Loading ImageNet weights and training all layers with FT,
(c) Loading ImageNet weights and training the last layers with TL,
(d) Same as point (c), while applying on-the-fly data augmentation (OFDA) during each epoch of the training.

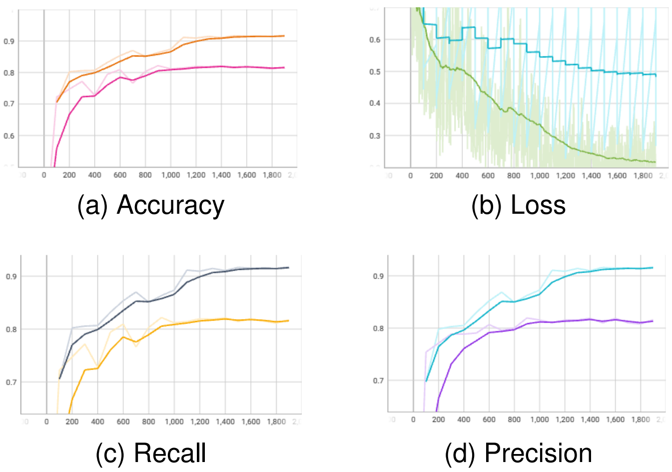
The reasoning behind OFDA is that training batches are not identical across epochs and, theoretically, this improves the model's generalisation properties. During validation, model architectures ResNet18, ResNet50, EfficientNet-b4 and EfficientNet-b7 were selected. As baseline, we also compared a simple randomly initialised CNN (with four convolutional layers, followed by an adaptive average pooling layer and two fully connected layers for classification), without any pretraining or TL, and a VGG. The SimpleCNN has far fewer parameters, computational cost, and training time compared to the powerful deeper NNs, which achieve higher accuracy at the cost of greater complexity and longer convergence. VGG (in this case VGG19) is one of the first NN that showed good performance in computer vision tasks.

B. Results

This study targets the task of image-based classification of the weather conditions on the road. During each experiment, a model was trained with a specific strategy and evaluated. A validation stage was included in the process to ascertain the effectiveness of training. Figure 6 illustrates the evolution of key metrics, namely Accuracy, Loss, Precision and Recall, across training and validation steps for ResNet50+TL on Dataset A. The model demonstrates a steady improvement in accuracy, converging to a stable value over time, while the loss consistently decreases during training. The close alignment between training and validation curves indicates effective model generalisation without much overfitting. Additionally, the relatively good performance in both Recall and Precision shows that the model is not skewed toward either metric.

The TL strategy transforms initial model weights so that the model can classify the 'No', 'Low' and 'High' surface moisture classes of our custom datasets instead of the classes from the ImageNet dataset. In order to demonstrate how FT and TL convert the model's output layers, the produced outputs when passing an image through the base classifier versus the customized classifier are displayed in Figure 7. Instead of predicting a standard ImageNet class (in this case 'truck'), the model is modified to produce the output classes as per the road condition classification task (in this case 'No'). The higher the output class score, the more certain the model is about its prediction.

During experimentation, the performance of different architectures and training strategies was thoroughly evaluated. For each architecture, various scenarios were explored, such as random initialisation or using weights pretrained on different versions of ImageNet. The architectures considered include EfficientNetb4, EfficientNetb7, ResNet18, and ResNet50 together with the baseline CNN and VGG19 models. Additionally, TL, FT and OFDA are applied in certain cases to the different base models. We chose not to apply OFDA to EfficientNetb4 and ResNet18, as their initial evaluations showed weaker performance compared to their larger counterparts, EfficientNetb7 and ResNet50. Given OFDA's mixed results in performance improvement, we determined it would likely yield unsatisfactory outcomes for these smaller models. The models that passed validation and the associated test performances for both image datasets are shown in Table I and Table II. The table provides insights into the impact of FT, TL and OFDA



(a) Accuracy (b) Loss

(c) Recall (d) Precision

Fig. 6. Comparison of key train and validation metrics across training steps (subunit of training epochs) for model ResNet50+TL and Dataset A.

(a) Image with target label 'No'.

```
Predictions with efficientnet-b7-----
trailer truck, tractor trailer, trucking rig, rig, articulated lorry, semi   (9.17%)
traffic light, traffic signal, stoplight                                     (6.42%)
alp                                                                          (6.00%)
bicycle-built-for-two, tandem bicycle, tandem                                (5.83%)
street sign                                                                  (5.74%)
```

(b) EfficientNetb7 pretrained model

```
Predictions with v8_exp\efficientnet-b7-pretrained-v8_exp_TL-----
No        (59.99%)
Low       (38.81%)
High      (1.20%)
```

(c) EfficientNetb7+TL model

Fig. 7. Comparison of outputs for an EfficientNetb7 pretrained model without FT or TL against the outputs of EfficientNetb7+TL. The percentages indicate the class score.

on training time, testing accuracy and F1-score. The number of parameters for the various networks is given in thousands (K) or millions (M). The best testing accuracy for each architecture is highlighted in bold for easy reference. The performance on Dataset B is nearly perfect, thus the focus is shifted more towards Dataset A.

The results reveal that ResNet50+TL, leveraging ImageNetV2 pretrained weights, outperforms other configurations with a testing accuracy of 82.18% for Dataset A, consisting of expressway images, and 99.99% for Dataset B, consisting of dashcam images. This underscores the effectiveness of transfer learning in enhancing the model's ability to discern road surface conditions. It is also evident that the ResNet architecture is superior to the EfficientNet architecture for this specific task. TL demonstrated consistently improved performance compared to training from scratch and FT. Moreover, training with TL tends to require less training time compared to other training strategies. The only outlier is ResNet18+TL, which converges at a later epoch. Furthermore, larger and deeper architectures, namely ResNet50 and EfficientNetb7, performed better than their lighter predecessors, ResNet19 and EfficientNetb4, respectively. However, the performance improvement comes at the cost of an increased number of trainable parameters. Regardless, the number of trainable parameters of large models coupled with TL is still much smaller than the number of trainable parameters when training from scratch or with FT. When trained from scratch, models are shown to have performed poorly, indicating that a more detailed adjustment of the training parameters is needed, which requires substantially more time. The simple CNN evaluated on dataset A and B performed poorly as expected compared to the more complex model architectures. The FT and TL variants for VGG19 showed comparable performance gains. However, this came at a cost of training time, parameter size and convergence, compared to EfficientNet and ResNet. This justifies our selection of more efficient better performing ResNet and EfficientNet models for evaluating TL and FT architectures. Finally, the inclusion of OFDA in the training pipeline demonstrates mixed outcomes across architectures, indicating the nuanced influence of additional optimisation techniques.

Performances can be more clearly evaluated on a confusion matrix. Here the matrix provides insights into the ResNet50 model's performance in classifying different road surface conditions. Figure 8 depicts the confusion matrix for Dataset A, specifically for expressway images, while Figure 9 depicts the confusion matrix for Dataset B, for dashcam images. The values are given in parentheses and the color bar is normalised against the total size of the test data. Training with TL can be seen to clearly improve the model's performance for both datasets, compared to FT.

In Figure 10, we present an illustrative example of the ResNet50+TL model's application on various test images of Dataset A, offering further insights into its classification performance. The top row showcases images correctly classified by the model. Conversely, the bottom row of the figure illustrates instances where the model misclassified road surface conditions. The predicted and corresponding ground truth labels shown in parenthesis are provided for comparison. A limited area of the sky is visible in each test image, while scenes change from day to night. The model fails to predict the road conditions correctly in cases where there are strong street lights and increased humidity in the air causing light scattering effect. The model also fails when the expressway camera is located at the exit of a tunnel, such that its view captures both interior and exterior road conditions simultaneously. These results allow for a qualitative understanding of the model's limitations and the types of challenges it may encounter in real-world scenarios.

V. Discussion

This research developed an efficient real-time system for classifying road surface conditions using CNN deep learning

TABLE I
TESTING PERFORMANCE FOR DATASET A - EXPRESSWAY IMAGES

Architecture	Initialisation	Train Params	Train Time	Epoch	Train Acc. (%)	Test Acc. (%)	Test F1 (%)
SimpleCNN	Random	421 K	14m 15s	10	49.10	49.05	32.28
VGG19	Random	139 M	46m 23s	20	52.92	53.66	49.35
VGG19+FT	ImageNetV1	139 M	79m 6s	35	76.44	70.90	70.03
VGG19+TL	ImageNetV1	12.3 K	52m 28s	31	75.28	70.90	70.34
EfficientNetb4	Random	17.6 M	81m 17s	21	56.72	55.63	53.43
EfficientNetb4+TL	ImageNetV1	5.4 K	106m 45s	29	80.90	75.19	74.12
EfficientNetb7+FT	ImageNetV1	63.8 M	325m 1s	37	82.23	78.69	78.06
EfficientNetb7+TL	ImageNetV1	7.7 K	98m 25s	24	84.59	80.06	79.50
EfficientNetb7+TL+OFDA	ImageNetV1	7.7 K	116m 55s	13	64.25	63.94	63.40
ResNet18	Random	11.2 M	77m 44s	26	77.21	75.25	65.99
ResNet18+TL	ImageNetV1	1.5 K	97m 45s	33	83.09	77.00	76.62
ResNet50+FT	ImageNetV2	23.5 M	112m 37s	31	88.18	80.69	79.15
ResNet50+TL	ImageNetV2	**6.1 K**	43m 7s	12	92.24	**82.18**	80.90
ResNet50+TL+OFDA	ImageNetV2	6.1 K	57m 15s	19	62.57	60.50	58.49

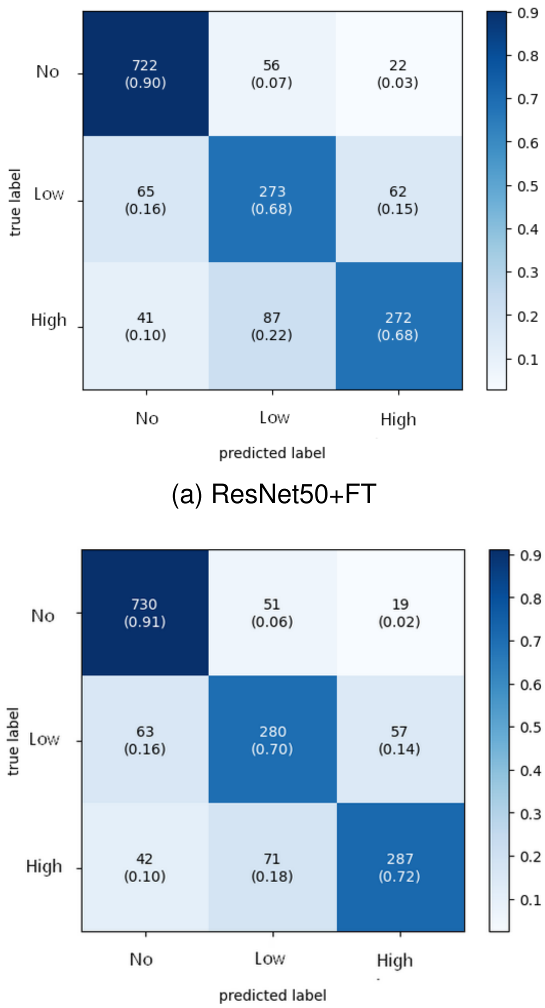
TABLE II
TESTING PERFORMANCE FOR DATASET B - DASHCAM IMAGES

Architecture	Initialisation	Train Params	Train Time	Epoch	Train Acc. (%)	Test Acc. (%)	Test F1 (%)
SimpleCNN	Random	421 K	22m 43s	10	49.01	49.29	38.82
VGG19	Random	139 M	43m 32s	10	66.30	64.48	61.8
VGG19+FT	ImageNetV1	139 M	43m 29s	10	99.61	99.52	98.77
VGG19+TL	ImageNetV1	12.3 K	133m 31s	10	99.75	99.49	98.85
EfficientNetb4	Random	17.6 M	32m 28s	5	45.32	43.54	43.05
EfficientNetb4+TL	ImageNetV1	5.4 K	24m 28s	5	98.77	98.31	98.08
EfficientNetb7+FT	ImageNetV1	63.8 M	120m 22s	9	97.76	97.21	97.21
EfficientNetb7+TL	ImageNetV1	7.7 K	43m 11s	7	99.17	98.37	97.07
ResNet18	Random	11.2 M	24m 20s	7	92.14	91.85	91.69
ResNet18+TL	ImageNetV1	1.5 K	20m 44s	7	99.87	99.58	98.86
ResNet50+FT	ImageNetV2	23.5 M	29m 13s	7	99.93	99.97	99.00
ResNet50+TL	ImageNetV2	**6.1** K	22m 6s	2	99.99	**99.99**	99.97

models with TL. While the training phase in these models can be time-consuming, the inference process for single image frames is remarkably fast, typically taking only a few milliseconds. This enables the system to operate in real-time, even for live video streams. Since most traffic data stream providers update video frames every few seconds at best (every few minutes in the case of Dataset A) and inference operations are much faster, the sufficient inference time demonstrates the system's capability for real-time deployment. This could be further elaborated in future work with a fully implemented system and measurement of the precise frame retrieval versus inference duration ratio. Nevertheless, the proposed system has several limitations that should be considered. Firstly, the dataset used for training and evaluation, even after OFDA, exhibits class imbalance, which may impact the model's ability to generalise across different road surface conditions. However, this imbalance should be mitigated to some extent by the influence of TL, which works as a generalised feature extractor for natural image scenes. Furthermore, the dataset labels were inferred from city-wide weather station values, which might not capture localised weather phenomena or lag in time. Additionally, due to the resource-intensive nature of training deep neural networks, especially with larger architectures like ResNet, the use of a small batch size is necessary, leading to increased memory requirements. This limitation can affect the model's overall performance, especially in scenarios where computational resources are constrained. Furthermore, attempts at OFDA, as described in the literature, have not yielded the expected improvements in model performance. Understanding and addressing these limitations are crucial for interpreting the study's findings and considering its applicability in real-world contexts.

Recent works have applied deep learning to road traffic detection and their results are mentioned for evaluation purposes. Weather prediction from road-facing images remains challenging due to varying conditions and data imbalances. In [28], ViTs outperformed CNNs, achieving 92% accuracy and 81.22% F1-score. However, ViTs require large amounts of training data to achieve high performances as compared to other approaches. In [29], pothole detection using YOLOv8 and transformers achieved 0.867 mAP and 35 FPS. In [33], wet roads were detected with a convolutional LSTM, achieving AUCs up to 0.92, while [38] reported accuracy 98.22% for weather detection on natural image scenes. In [34], transfer learning with VGG NET achieved 90% accuracy. In comparison, our proposed model with ResNet50+TL achieves an accuracy 82% and 99% and F1-Scores of 80.90% and 99.0%, for each dataset, respectively. This highlights the effectiveness of our approach, which is comparative and superior to other methodologies in similar tasks.

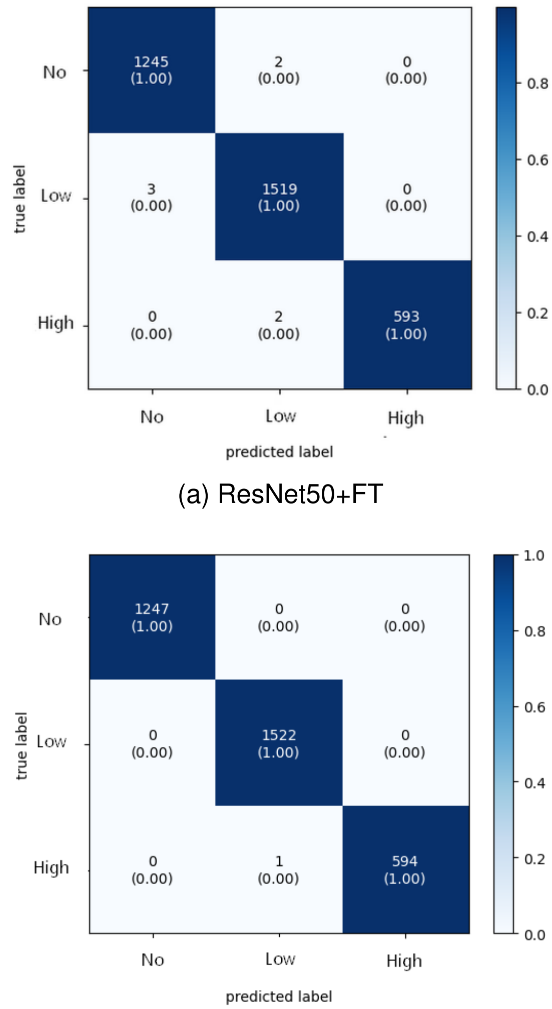
Performance was shown to differ for the two datasets, however, both are needed to provide relevant insights. Images captured from fixed expressway cameras and dashcams play distinct roles in an ITS, each offering unique advantages and applications. Fixed expressway cameras are strategically positioned along highways to provide a comprehensive and continuous view of major road segments. The fixed nature of these cameras facilitates long-term observation, making



(a) ResNet50+FT

(b) ResNet50+TL

Fig. 8. Confusion Matrix for Dataset A - Expressway Images



(a) ResNet50+FT

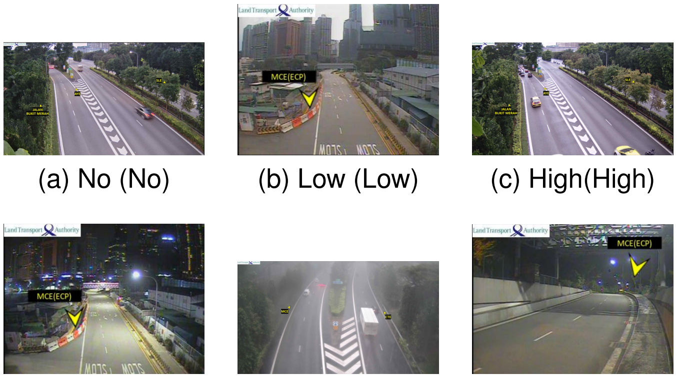
(b) ResNet50+TL

Fig. 9. Confusion Matrix for Dataset B - Dashcam Images

them particularly valuable for tasks such as traffic surveillance, incident detection, and overall system monitoring. The images from fixed expressway cameras serve as a foundational source of data for understanding macroscopic traffic patterns, infrastructure health, and environmental conditions. On the other hand, dashcams, mounted on vehicles, offer a dynamic and mobile perspective of the road environment. Dashcams are instrumental in providing fine-grained details of individual driving experiences and can be particularly useful for capturing transient events, such as sudden weather changes, accidents, or road hazards. The combination of fixed expressway cameras and dashcams in an ITS creates a synergistic data-driven approach, where the strengths of each type of camera contribute to a comprehensive understanding of the transportation network. Due to the differences in angle, motion blur, image size and frame rate for data availability, a difference in performance is expected. Road surface detection from expressway cameras, where the sky or a close-up view of the road is not visible, proved to be a significantly harder task to solve compared to the analysis of dashcam images.

The accurate classification of weather conditions on the road holds significance in numerous aspects of modern society. One crucial domain where this research finds utility is in the realm of road and expressway management. As transportation systems continue to evolve towards automation and enhanced efficiency, the ability to provide real-time and precise weather and road condition detection becomes a critical component for ensuring passenger safety and optimising traffic flow. By utilising advanced image-based weather classification techniques, this research aims to enhance the way we perceive and react to changing weather dynamics on roads and expressways, e.g. flooded road surfaces or oil spillage after an accident. The integration of such technology into transportation infrastructure promises to empower authorities and commuters alike with timely and accurate information, enabling proactive responses to adverse weather events, reduced accident rates, and improved overall road safety.

As a direction of future work, in the context of an ITS, the proposed TL-based road condition classification system can contribute significantly to the development of a digital twin framework. This framework, encompassing the integration of advanced technologies for real-time data acquisition and anal-



(a) No (No) (b) Low (Low) (c) High(High)

(d) Low (No) (e) High (Low) (f) No (High)

Fig. 10. Example of ResNet50+TL model's application on various test images. The label is the prediction label and the label in parenthesis is the ground truth label.

ysis, aims to enhance the overall efficiency of ITS. The application of TL in road condition classification, as demonstrated by the results discussed earlier, offers a crucial element for the digital twin's predictive capabilities. In the context of big data, storing extracted analytics values instead of the entire image can assist database and storage optimisation. Furthermore, the integration of this road condition classification approach within a broader weather monitoring framework enhances its applicability and impact. The combination of fixed and mobile dashcams forms a robust data acquisition system that captures real-time traffic data, including road surface conditions. This dynamic data collection process, complemented by the proposed TL model, enables the creation of a comprehensive digital representation of a city's transportation network. Such a digital twin framework not only facilitates accurate and timely road condition predictions but also contributes to traffic management, route optimisation, and overall system resilience. The ability to quickly update the model to varying environmental conditions makes this AI driven digital twin a valuable tool for optimising transportation systems in real time.

This research leverages cutting-edge computer vision technologies to enable real-time image-based detection of road surface conditions. Beyond individual advantages, the study holds the potential for optimising traffic management processes. By dynamically adjusting traffic flow and implementing precautions during adverse weather conditions, road authorities can enhance overall transportation efficiency and safety. Furthermore, this research has the potential to streamline traffic management processes, enabling road authorities to dynamically adjust traffic flow and implement necessary precautions during adverse weather conditions. Additionally, the analysis of training times and model sizes addresses practical considerations for deployment in real-world applications, particularly in scenarios with limited computational resources. These findings advance the understanding of deep learning approaches for image-based road surface condition detection, opening avenues for further research in this domain.

VI. Conclusion

In conclusion, this research presents a novel and efficient real-time and data-driven system for classifying road surface conditions, employing CNN deep learning models with TL. The developed model demonstrates the capability to differentiate various road surface conditions, addressing challenges such as motion blur and low-resolution images. The exploration of Singapore's tropical climate adds a unique dimension to the study, and the application of TL, particularly with the ResNet50 architecture, proves to be a powerful tool for fast model adaptation without extensive data collection. The research contributes to the field of image-based road surface condition detection using big data from diverse traffic camera networks, paving the way for advancements in data-driven ITS. The integration of this approach within a broader traffic monitoring framework, forming a digital twin, holds promise for enhancing transportation efficiency and safety in the face of changing environmental conditions.

Acknowledgments

We would like to express our sincere gratitude to the Singapore Land Authority for their invaluable collaboration in providing access to the Datamall dataset. This work was carried out during Project "EMIA - Epi-terrestrial Multimodal Input Assimilation" (Reference: 10036013), supported by Innovate UK and Enterprise Singapore as part of the UK-Singapore Collaborative R&D competition. For the purpose of Open Access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript (AAM) version arising from this submission.

References

[1] W. M. Organization, "Climate change and extreme weather impacts hit asia hard," 2024. [Online]. Available: https://wmo.int/news/media-centre/climate-change-and-extreme-weather-impacts-hit-asia-hard

[2] H. M. Hammad, M. Ashraf, F. Abbas, H. F. Bakhat, S. A. Qaisrani, M. Mubeen, S. Fahad, and M. Awais, "Environmental factors affecting the frequency of road traffic accidents: a case study of sub-urban area of pakistan," *Environmental Science and Pollution Research*, vol. 26, pp. 11 674–11 685, 4 2019.

[3] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," *IEEE Access*, vol. 8, pp. 108 952–108 971, 2020.

[4] G.-L. Huang, A. Zaslavsky, S. W. Loke, A. Abkenar, A. Medvedev, and A. Hassani, "Context-aware machine learning for intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, pp. 17–36, 1 2023.

[5] E. Aloupogianni, F. Doctor, C. Karyotis, T. Maniak, R. Tang, and R. Iqbal, "An ai-based digital twin framework for intelligent traffic management in singapore," in *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*, 2024. DOI: 10.1109/ICECET61485.2024.10698642 pp. 1–6.

[6] P. Jonsson, "Classification of road conditions: From camera images and weather data." IEEE, 9 2011. ISBN 978-1-61284-924-9 pp. 1–6.

[7] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," 2020, pp. 11 682–11 692.

[8] K. Dahmane, P. Duthon, F. Bernardin, M. Colomb, F. Chausse, and C. Blanc, "Weathereye-proposal of an algorithm able to classify weather conditions from traffic camera images," 2021.

[9] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," 4 2016.

[10] Y. Zhang, A. Carballo, H. Yang, and K. Takeda, "Perception and sensing for autonomous vehicles under adverse weather conditions: A survey," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 196, pp. 146–177, 2 2023.

[11] M. R. Ibrahim, J. Haworth, and T. Cheng, "Weathernet: Recognising weather and visual conditions from street-level images using deep residual learning," *ISPRS International Journal of Geo-Information*, vol. 8, 11 2019.

[12] A. Bailly, C. Blanc, Élie Francis, T. Guillotin, F. Jamal, B. Wakim, and P. Roy, "Effects of dataset size and interactions on the prediction performance of logistic regression and deep learning models," *Computer Methods and Programs in Biomedicine*, vol. 213, 1 2022.

[13] I. H. Sarker, "Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions," 11 2021.

[14] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: Computational intelligence techniques and application areas," *Technological Forecasting and Social Change*, vol. 153, p. 119253, 2020.

[15] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1299–1312, 5 2016.

[16] S. Niu, Y. Liu, J. Wang, and H. Song, "A decade survey of transfer learning (2010–2020)," *IEEE Transactions on Artificial Intelligence*, vol. 1, pp. 151–166, 10 2020.

[17] J. G. A. Barbedo, "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification," *Computers and Electronics in Agriculture*, vol. 153, pp. 46–53, 10 2018.

[18] M. Qasim and E. Verdu, "Video anomaly detection system using deep convolutional and recurrent models," *Results in Engineering*, vol. 18, p. 101026, 2023. DOI: https://doi.org/10.1016/j.rineng.2023.101026

[19] J. Gao, H. Yang, M. Gong, and X. Li, "Audio–visual representation learning for anomaly events detection in crowds," *Neurocomputing*, vol. 582, p. 127489, 2024. DOI: https://doi.org/10.1016/j.neucom.2024.127489

[20] N.-C. Ristea, F.-A. Croitoru, R. T. Ionescu, M. Popescu, F. S. Khan, and M. Shah, "Self-distilled masked auto-encoders are efficient video anomaly detectors," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. DOI: 10.1109/CVPR52733.2024.01513 pp. 15 984–15 995.

[21] Q. Chu, S. Liang, and M. Yang, "Visibility analysis in fog weather based on convolution neural network and transfer learning." IEEE, 8 2022. ISBN 978-1-6654-6287-7 pp. 294–298.

[22] M. L. Ahmed, R. Iqbal, C. Karyotis, V. Palade, and S. A. Amin, "Predicting the public adoption of connected and autonomous vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, pp. 1680–1688, 2 2022.

[23] A. K. Pandey, R. Iqbal, S. Amin, T. Maniak, V. Palade, and C. Karyotis, "Deep neural networks based approach for pothole detection." IEEE, 11 2021. ISBN 978-1-6654-3796-7 pp. 1–4.

[24] M. Hijji, R. Iqbal, A. K. Pandey, F. Doctor, C. Karyotis, W. Rajeh, A. Alshehri, and F. Aradah, "6g connected vehicle framework to support intelligent road maintenance using deep learning data fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, pp. 7726–7735, 7 2023.

[25] C. Karyotis, T. Maniak, F. Doctor, R. Iqbal, V. Palade, and R. Tang, "Deep learning for flood forecasting and monitoring in urban environments." IEEE, 12 2019. ISBN 978-1-7281-4550-1 pp. 1392–1397.

[26] R. Iqbal, T. Maniak, and C. Karyotis, "Intelligent remote monitoring of parking spaces using licensed and unlicensed wireless technologies," *IEEE Network*, vol. 33, pp. 23–29, 7 2019.

[27] A. Chaudhuri, "Smart traffic management of vehicles using faster r-cnn based deep learning method," *Scientific Reports*, vol. 14, no. 1, p. 10357, 2024.

[28] M. Samo, J. M. Mafeni Mase, and G. Figueredo, "Deep learning with attention mechanisms for road weather detection," *Sensors*, vol. 23, no. 2, p. 798, 2023.

[29] M. Jakubec, E. Lieskovska, B. Bucko, and K. Zabovska, "Pothole detection in adverse weather: leveraging synthetic images and attention-based object detection methods," *Multimedia Tools and Applications*, pp. 1–28, 2024.

[30] H. Xiao, F. Zhang, Z. Shen, K. Wu, and J. Zhang, "Classification of weather phenomenon from images by using deep convolutional neural network," *Earth and Space Science*, vol. 8, 5 2021.

[31] K. Ozcan, A. Sharma, S. Knickerbocker, J. Merickel, N. Hawkins, and M. Rizzo, "Road weather condition estimation using fixed and mobile based cameras," pp. 192–204, 2020.

[32] J. Xia, D. Xuan, L. Tan, and L. Xing, "Resnet15: Weather recognition on traffic road with deep convolutional neural network," *Advances in Meteorology*, vol. 2020, 2020.

[33] F. Garcea, G. Blanco, A. Croci, F. Lamberti, R. Mamone, R. Ricupero, L. Morra, and P. Allamano, "Self-supervised and semi-supervised learning for road condition estimation from distributed road-side cameras," *Scientific reports*, vol. 12, no. 1, p. 22341, 2022.

[34] K. A. Vinodhini and K. R. A. Sidhaarth, "Pothole detection in bituminous road using cnn with transfer learning," *Measurement: Sensors*, vol. 31, p. 100940, 2024.

[35] M. F. Naufal and S. F. Kusuma, "Weather image classification using convolutional neural network with transfer learning." AIP, 2022, p. 050004.

[36] N. An, "Xception network for weather image recognition based on transfer learning." IEEE, 8 2022. ISBN 978-1-6654-9246-1 pp. 330–333.

[37] R. Pillai, N. Sharma, and R. Gupta, "Fine-tuned efficientnetb4 transfer learning model for weather classification." IEEE, 8 2023. ISBN 979-8-3503-0228-8 pp. 1–6.

[38] Q. A. Al-Haija, M. A. Smadi, and S. Zein-Sabatto, "Multi-class weather classification using resnet-18 cnn for autonomous iot and cps applications." IEEE, 12 2020. ISBN 978-1-7281-7624-6 pp. 1586–1591.

[39] P. Mlodzianowski, "Weather classification with transfer learning - inceptionv3, mobilenetv2 and resnet50," 2021, pp. 3–11.

[40] P. Herruzo, A. Gruca, L. Lliso, X. Calbet, P. Ripodas, S. Hochreiter, M. Kopp, and D. P. Kreil, "High-resolution multi-channel weather forecasting – first insights on transfer learning from the weather4cast competitions 2021." IEEE, 12 2021. ISBN 978-1-6654-3902-2 pp. 5750–5757.

[41] S. L. T. Authority, "Datamall." [Online]. Available: https://datamall.lta.gov.sg/content/datamall/en.html (Accessed 2023-09-06).

[42] O. Weather, "Weather api." [Online]. Available: https://openweathermap.org/ (Accessed 2023-09-06).

BIOGRAPHY SECTION

Eleni Aloupogianni received her BS and MS diploma from the National Technical University of Athens, Greece, in 2016. She earned a ME degree in Information and Communications Engineering in 2019 and a PhD in Human-Centered Science and Biomedical Engineering in 2022, from the Tokyo Institute of Technology, Japan. Her research interests include medical image processing, hyperspectral imaging and multi-modal machine learning.

Faiyaz Doctor is a Senior Lecturer in Computer Science at the School of Computer Science and Electronic Engineering, University of Essex, specializing in the design and application of biological inspired machine learning techniques and intelligent agents approaches in real world systems.

Charalampos Karyotis is a senior data scientist at Interactive Coventry Ltd. He has a Ph.D. degree in computer science from Coventry University and over 8 years of experience in data-driven technologies and industrial projects. His research interests include machine learning, smart city systems and big data analytics.

Raymond CL Tang is the founder and CEO of a group of companies that augments human productivity by integrating AI with city scale Digital Twin for remote monitoring, forecast and control of energy grid, flooding and traffic in Asean countries. Originally from Malaysia with Master in reconstructive surgery in UK and Ireland, and Master in electrical engineering in Malaysia.

Rahat Iqbal is a Senior Academic, Director, Inventor and Consultant with 20 years' experience in industrial and academic roles. His main research interests lie in Big Data Analytics, Internet of Things (IoT), Connected and Autonomous Vehicle, Machine Learning and Artificial Intelligence. Prior to joining the University of Dubai, he was Leading the Connected Mobility Research Group at Coventry University, UK.