Optimising Data Collection Process in Autonomous Driving Industry Using Machine Learning

MSc Research Project
Data Analytics

Maria Migrova
Student ID: x21146021

School of Computing
National College of Ireland

Supervisor: Jorge Basilio
Industry Supervisor: Mihai Ilie

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Maria Migrova
Student ID:	x21146021
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Jorge Basilio
Submission Due Date:	14/08/2023
Project Title:	Optimising Data Collection Process in Autonomous Driving Industry Using Machine Learning
Word Count:	8151
Page Count:	30

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Maria Migrova
Date:	18th September 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Optimising Data Collection Process in Autonomous Driving Industry Using Machine Learning

Maria Migrova
x21146021

Abstract

This study presents an innovative solution for optimising data collection processes within the realm of autonomous driving. It introduces a high-performance multi-label Convolutional Neural Network (CNN) classification model designed to classify attributes like weather, lighting, and surface conditions. Notably, this model achieves an impressive 99.46% testing accuracy, accompanied by a minimal test loss of 0.0162 and a speedy training time of only 3.25 minutes per epoch. This exceptional performance renders it suitable for deployment within vehicles, enabling real-time driver alerts to prevent over-capturing various categories. This approach effectively mitigates potential human errors that can arise from pre-annotated data collected by drivers in the vehicle.

Furthermore, a unique and robust multi-class dataset comprising over 20,000 images, capturing diverse weather, lighting, and surface conditions, has been meticulously curated from various autonomous driving sources.

The findings of this research not only contribute a novel methodology but also pave the way for extensive future exploration in this field. The optimization of data collection in autonomous driving remains a fertile ground for further investigation, offering opportunities to enhance methodologies, refine datasets, train for additional classes, and unlock new avenues of innovation.

1 Introduction

1.1 Background and Motivation

In the dynamic landscape of transportation, the rise of autonomous driving is shaping a new era of possibilities. This evolution holds the potential to reshape mobility, ensuring safer roads, reduced congestion, and enhanced accessibility. Amidst this exciting transformation, the significance of data collection takes center stage. Data serves as the lifeblood driving the development of cutting-edge technologies in autonomous driving. (Nava; 2023)

However, this vital data collection process is not without challenges. The diverse scenarios and environments that autonomous vehicles must master demand a comprehensive and intricate collection effort. This involves skilled drivers, specialized equipment, sensor technology, annotation teams, and data processing pipelines. Despite these efforts, human errors and inefficiencies can creep in, impacting the accuracy and effectiveness of the collected data.

To address this, we introduce an innovative solution: an image classification machine learning model. By harnessing the power of image classification, this model serves a dual purpose. Firstly, it can be seamlessly integrated into vehicle systems to alert drivers of potential inaccuracies during data capture, preventing unnecessary overcapturing. Secondly, it empowers data collection leads to identify errors and ensure adherence to planned categories, optimizing capture efforts.

This advancement not only corrects errors but transforms data collection itself. By enhancing accuracy and preventing overcapturing, it promises both precision and cost-effectiveness. Our cutting-edge machine learning model not only confronts challenges head-on but also propels the autonomous driving industry toward a future defined by exceptional accuracy and efficiency.

1.2 Research Question

Main Research Question: How can an image classification Machine Learning model be strategically designed and proficiently implemented to optimise the effectiveness and the efficiency of the daily data collection process in the autonomous driving industry?

Sub-research Question: How can the model's footprint be optimised while maintaining high accuracy levels?

Sub-research Question: Can the implementation of Transfer Learning enhance the model's performance and address the challenges posed by limited data availability and computational resources?

1.3 Challenges and Difficulties

Engaging with the data collection process in the real-world setting presents its own set of challenges. Licensing restrictions and stringent General Data Protection Regulation (GDPR) guidelines restrict direct access to Valeo's proprietary data. As an alternative, this research explores the utilization of open-source datasets from other autonomous driving companies, with their distinct characteristics – such as the use of pin-point images compared to Valeo's fish-eye perspective. Notably, due to the lack of sufficient data for certain categories, data augmentation was employed, resulting in a final dataset with a higher count of augmented images compared to the original ones. It's worth noting that the dataset's balance is imperfect due to the absence of diverse weather conditions, which can impact the model's generalisation.

1.4 Structure of the Report

This research project is structured to comprehensively address the various facets of optimising the data collection process for autonomous driving. The document encompasses:

Abstract: A succinct overview of the research's objectives, methodology, and key findings. Introduction: The present section, providing a glimpse into the context, research questions, and challenges. Related Work: A survey of existing literature and research relevant to data collection, image classification, and autonomous driving. Research Methodology: A detailed outline of the research's approach, including data acquisition, model design, and evaluation strategies. Design Specification: Articulation of the model's architecture, technical considerations, and rationale behind design choices. Implementation:

A meticulous account of how the model was built and integrated into the data collection process. Evaluation: A critical assessment of the model's performance, considering accuracy, efficiency, and embedded footprint. Conclusion and Future Work: A synthesis of the findings, implications, and avenues for further research and development. Bibliography: A compilation of all referenced sources providing a solid foundation for the research.

2 Related Work

This section delves into image classification's role in optimizing data collection for autonomous driving. It traces the evolution of image classification techniques and compares models like Convolutional Neaural Network (CNN), Support Vector Machine (SVM), and Random Forest (RF), with CNN as the chosen approach. The study explores weather image classification, multi-label classification, and transfer learning's applicability. Additionally, the inquiry evaluates architecture models, focusing on EfficientNetB1's suitability. This literature review provides a solid foundation for enhancing data collection efficiency in the autonomous driving sector.

2.1 Evolution of Image Classification

In the dynamic evolution of Computer Vision, (Huang; 1996) provides a retrospective on its journey from emulating human visual systems to engineering autonomous systems. It traces the shift from "low-level" tasks to "Purposive Vision," acknowledging challenges in real-world applications.

Deep Learning's surge from 2010 to 2020 is highlighted by (Li et al.; 2021), emphasizing its significance in image analysis and its potential applications. Notably, the evaluation of image classification networks shows AlexNet, GoogleNet, VGGNet, and DenseNet's favorable performance.

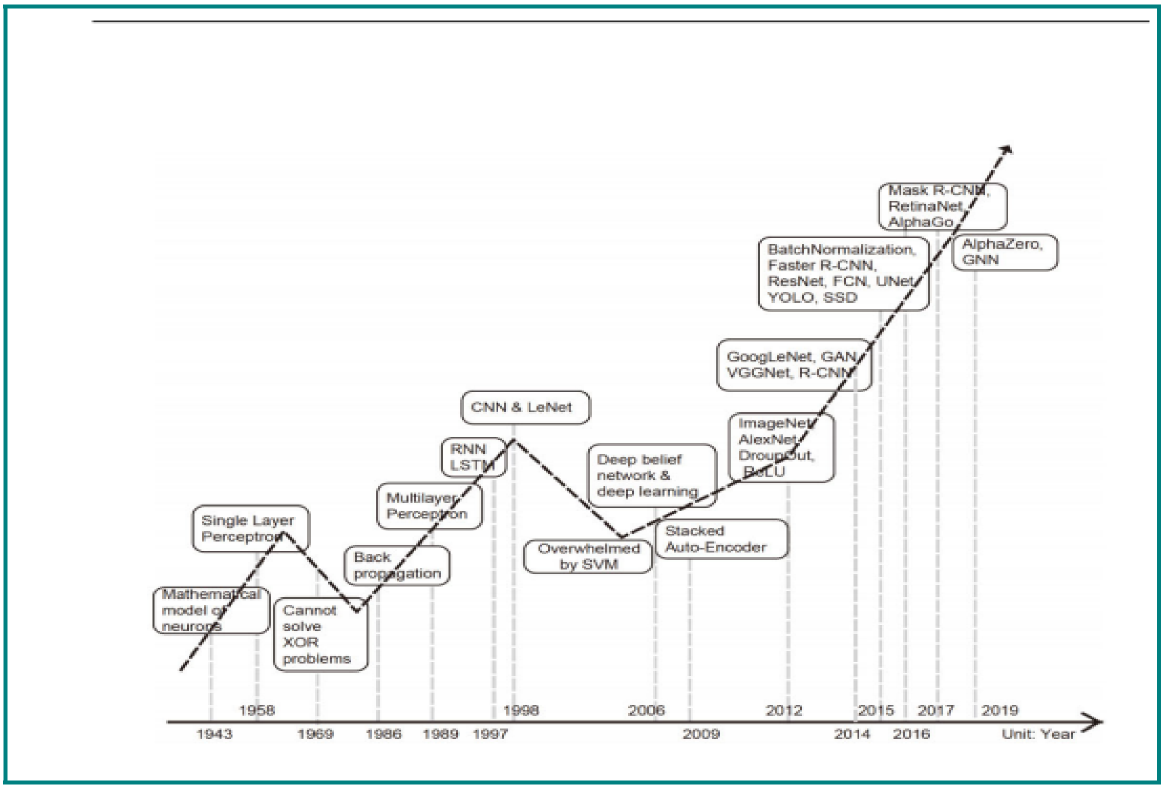


Figure 1: Deep Learning History (Qiang Li et al., 2021)

Introducing automation, (Real et al.; 2017) leverages evolutionary algorithms for autonomous image classification model discovery. The approach yields competitive accuracies on CIFAR-10 and CIFAR-100 datasets, showcasing the potential of advanced techniques in machine learning research.

In the realm of AI algorithm comparison, (Yogitha et al.; 2023) evaluates CNN, SVM, and Random Forests. CNN excels in accuracy, and Random Forest shines in precision. The research emphasizes model and metric considerations, aiding classifier selection for specific tasks.

	CNN	SVM	Random Forest
Accuracy	0.927	0.8809	0.9136
Specificity	0.8717	0.8482	0.9347
Sensitivity	0.9649	0.9057	0.8989
Precision	0.9166	0.8871	0.9518
F1-score	0.9066	0.8601	0.8989

Table 1: Evaluation Metrics comparison (Yogitha et al., 2023)

Addressing complex data analysis, (Sothea et al.; 2020) demonstrates CNN's ability to extract features from hyperspectral and 3D data, achieving an Overall Accuracy of 84.4% and 74.95% in different forest fragments. This superiority highlights CNN's potential for robust and accurate tree species classification.

Further advancement in integrating artificial intelligence into practical scenarios is exemplified by an image-based billing system tailored for supermarkets (Shakya; 2020). Through the lens of computer vision, this system identifies fruits and vegetables via a camera, streamlining the billing process. Rigorous evaluation, encompassing classifiers like Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), and Discriminant Analysis (DA), unveils the effectiveness of this innovative approach. Notably, KNN emerges as the frontrunner, achieving an impressive accuracy of 93.103%, promising improved real-time billing procedures. This study underscores the transformative potential of artificial intelligence in revolutionizing mundane tasks and enhancing operational efficiency.

2.2 Image Weather Classification

The research landscape spans diverse aspects of weather classification using deep learning techniques. In the study by (Elhoseiny et al.; 2015), a pioneering approach employs Convolutional Neural Networks (CNNs) to achieve a remarkable normalized classification accuracy of 82.2%. This substantial improvement over previous benchmarks, such as the 53.1% accuracy, underscores CNNs' prowess in addressing intricate weather classification challenges.

Further exploring weather phenomenon recognition, (Xiao et al.; 2021) propose the MeteCNN model, a deep CNN, to attain an impressive 92% accuracy on the Weather Phenomenon Database (WEAPD). This dataset, containing 6,877 images with 11 weather phenomena categories, emphasizes the potential of deep learning methods in advancing accurate weather forecasting.

Taking a distinctive approach, (Roser and Moosmann; 2008) focuses on enhancing machine vision's performance under diverse weather conditions. Their method's remarkable accuracy in discerning weather situations from monocular color images showcases its robustness, making it a promising tool for real-world scenarios.

Similarly, (Dhananjaya et al.; 2021) delves into autonomous driving perception, meticulously curating a dataset encompassing weather, light level, and street type classifications. Their study emphasizes the challenges in weather and light perception, highlighting the need for ongoing research to enhance model performance and alleviate the cost of supervised labeling.

The innovative framework introduced by (Al-Haija et al.; 2022) for weather classification exhibits remarkable accuracy rates and efficient inference times. While a standardized weather dataset might be missing, the emphasis on accuracy underscores its relevance and potential impact.

	SqueezeNet	ResNet-50	EfficientNet-B0
Accuracy	96.05%	98.48%	97.78%
Sensitivity	95.96%	98.41%	97.96%
Precision	95.51%	98.51%	97.74%
F1-score	95.68%	98.44%	97.84%

Table 2: Evaluation Metrics (Abu Al-Haija et al., 2022)

Expanding the horizon, (Chen et al.; 2021) presents a comprehensive review of Convolutional Neural Networks (CNNs) application in image classification. Analyzing the evolution of CNNs and their success in remote sensing image scene classification, the study becomes a pivotal reference for designing effective models in the field.

Table 3. *Cont.*

Model	Year	Depth/Version	Data Set	BFLOPs	Params	Accuracy(%) ImageNet, Top-1	Accuracy(%) CIFAR 10	Accuracy(%) CIFAR 100	Characteristics	Approach
ResNeXt (32×4d) + BAM	2018	101	ImageNet	8.05	44.6M	80.85	-	-	Channel attention and spatial attention Implantable lightweight module	CNN +Attention'
ResNeXt (32×4d) + CBAM	2018	101	ImageNet	7.519	49.2M	80.58	-	-	Channel attention and spatial attention Implantable lightweight module	CNN +Attention'
AmoebaNet-A	2018	-	ImageNet	23.1 104	86.7M 469M	82.8 83.9	-	-	NASNet search space and genetic algorithm	CNN +NAS
GPipe	2018	-	ImageNet CIFAR 10/100	-	557M	84.3	99	91.3	Using Pipeline Parallelism to train effectively	CNN
ResNet + GE	2019	101	ImageNet CIFAR 10/100	7.59	58.4M -	79.26 -	- 95.07	- 79.15	Feature context exploitation Implantable lightweight module	CNN +Attention'
ECA-Net	2019	101	ImageNet	7.35	42.45M	78.65	-	-	No channel dimensionality reduction Portable lightweight modul	CNN +Attention'
SKNet	2019	101 29	ImageNet CIFAR 10/100	8.46	48.9M 27.7M	79.81 -	- 96.53	- 82.67	Adaptively adjust receptive field Implantable lightweight module	CNN +Attention'
GSoP-Net	2019	50	ImageNet	6.56	58.65M	78.81	-	-	Global Second-order Pooling Implantable lightweight module	CNN +Attention'
EfficientNetV1	2019	B0 - B7, L2	ImageNet	0.39, 0.70, 1.0, 1.8, 4.2, 9.9, 19, 37, -	5.3M, 7.8M, 9.2M 12M, 19M, 30M 43M, 66M, 480M	77.1, 79.1, 80.1, 81.6, 82.9, 83.6, 84.0, 84.3, 85.5	98.9 (B7)	91.7 (B7)	Utilizes effective compound coefficient to scale up CNNs	CNN +NAS
BiT	2019	-L	ImageNet CIFAR 10/100	-	-	87.54*	99.37	93.51	Transfer learning Better training and fine-tuning strategies	CNN Transfer-learning
NoisyStudent	2019	-L2	ImageNet	-	480M	88.4*	-	-	Self traning / Semi-Supervised add noise to the student	CNN Semi-Supervised
FixEfficientNet	2020	-B7 -L2	ImageNet	-	66.4M 480M	85.3, 87.1* 85.7, 88.5*	-	-	A training strategy that employs different train and test resolutions	CNN
ViT-H/14	2020	-L, -H	ImageNet CIFAR 10/100	-	307M, 632M	87.76*, 88.55*	99.42, 99.50	90.54, 90.72	Pure transformer model for Computer vision	Vision Transformer
Meta Pseudo Labels	2020	-L2	ImageNet	-	480M	90.2*	-	-	Self traning / Semi-Supervised Update teacher based on student performance	CNN Semi-Supervised
EfficientNetV2	2021	-S, -M, -L	ImageNet CIFAR 10/100	8.8, 24, 53	22M, 54M, 120M 24M, 55M, 121M	83.9, 85.1, 85.7	98.7, 99.0, 99.1	91.5, 92.2, 92.3	FixRes, Fused-MBConv, NAS	CNN +NAS
NFNet	2021	F6+SAM, F4+	ImageNet	337.28, -	438.4M, 527M	86.5, 89.2*	-	-	Adaptive gradient clipping technique, Without Normalization	CNN
ViT-G/14	2021	-G/14	ImageNet	-	1843M	90.45*	-	-	Pure transformer model for Computer vision	Vision Transformer
CoAtNet	2021	-7	ImageNet	2586	2440M	90.88*	-	-	Mixed methods, Convolutional + Transformer	CNN + Transformer

Figure 2: Test results for different CNN models (Chen et al., 2021)

2.3 Transfer Learning for Image Classification Models

Transfer learning is popular in image classification due to its ability to leverage knowledge gained from pre-trained models on large datasets. This approach significantly reduces the need for extensive training data and computational resources, making it feasible to achieve high accuracy even with limited labeled examples. By fine-tuning pre-trained models on specific tasks, transfer learning accelerates model convergence, enhances generalization, and enables the application of deep learning to a wider range of image classification problems with improved efficiency and effectiveness (Figure 5).

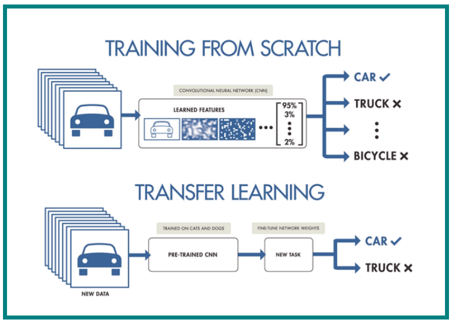


Figure 3: Training from scratch vs Transfer Learning (Medium, 2020)

In a thorough investigation, (Krishna and Kalluri; 2019) conducted an extensive survey on deep learning architectures and transfer learning for image classification. Models like AlexNet, VGG, ResNet, and Inception were assessed across various datasets, revealing the paramount importance of data scale and GPU resources. Transfer learning with models like AlexNet, GoogleNet, and ResNet50 on the CIFAR10 dataset showcased accuracy rates of 13%, 68.95%, and 52.55%, respectively. The study underscored the potential of transfer learning to significantly enhance image classification performance, offering insights into effective strategies.

Another study honed in on CNN architecture refinement, specifically VGG19, through transfer learning. Robust feature extraction by CNNs like AlexNet, VGG16, and VGG19 was highlighted. Through empirical evidence from GHIM10K and CalTech256 databases, VGG19's superiority was validated, solidifying its efficacy in real-world image classification tasks. The study's hybrid approach involving CNN feature extraction followed by SVM classification further enriched the findings, encouraging potential applications in object detection and human action recognition (Shaha and Pawar; 2018).

In pursuit of optimized Convolutional Neural Network (ConvNet) scaling, (Tan and Le; 2020) introduced an innovative approach to balancing depth, width, and resolution. Their compound coefficient-based scaling method showcased remarkable accuracy and efficiency enhancements, particularly in EfficientNets. EfficientNet-B7 achieved 84.3% top-1 accuracy on ImageNet while remaining significantly smaller and faster than alternatives. This systematic exploration of ConvNet scaling yielded state-of-the-art models with enhanced efficiency and accuracy across diverse tasks, solidifying its contribution to the field.

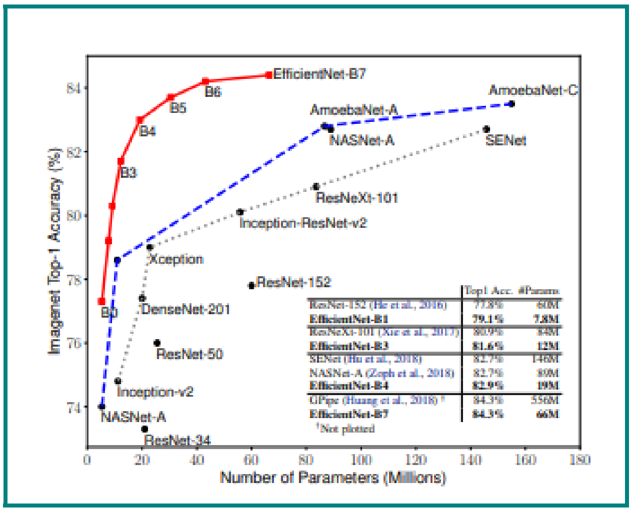


Figure 4: Model Size Comparison (Tan and Le, 2020)

2.4 Multi-label Weather Classification

In their pursuit of multi-label weather recognition, (Zhao et al.; 2018) devised a novel approach to utilizing a CNN-RNN architecture. Curating a multi-label weather classification dataset of 10,000 images from various weather conditions, they integrated a channel-wise attention model and convolutional LSTM into their model. This pioneering strategy showcased superior accuracy, precision, and recall values on both datasets compared to single-label methods. Challenges stemming from ambiguous annotations were acknowledged, spurring discussions on the significance of multiple labels and prospects for integrating additional modalities like humidity. This study serves as a foundation for multi-label weather recognition, opening avenues for improved annotation clarity and multi-modal integration in future research.

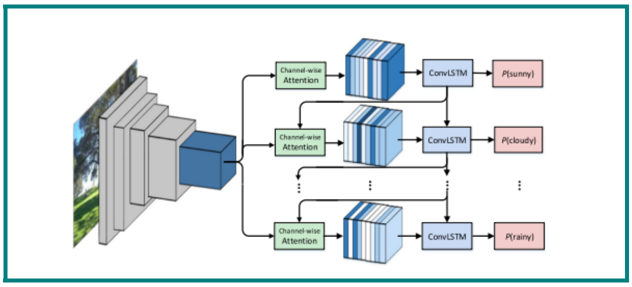


Figure 5: Proposed CNN-RNN Architecture (Zhao et al., 2018)

Pioneering an innovative method for classifying outdoor images into sunny or cloudy conditions using weather cues, (Li et al.; 2017) introduced a collaborative learning framework. They identified five key cues - sky, shadow, reflection, contrast, and haze - and fused them into a 621-dimensional feature vector. Their dataset comprised 10,000 outdoor images, encompassing diverse weather conditions. The study revealed the significance of various cues for accurate classification and employed a collaborative learning strategy to optimize voter contributions, achieving a remarkable normalized accuracy of 53.1%. By segmenting the dataset into clusters and employing a unified optimization framework, the study showcased the potential of computer vision in weather analysis. This comprehensive approach not only offers insights for automated weather classification but also makes their dataset publicly available for future advancements.

2.5 Conclusion

In conclusion, delving into the existing body of work on weather condition classification yields valuable insights. Convolutional Neural Networks (CNNs) emerge as powerful tools in image-based classification, showcasing their ability to capture intricate patterns and details. In the realm of CNN architectures, EfficientNet has truly stood out, offering a well-balanced mix of accuracy, model size, and computational efficiency.

A notable point that surfaces is the limited attention given to multi-label weather condition classification in the research landscape. This scarcity highlights the unique path taken in navigating the less-explored territory of weather classification spanning across multiple labels. This innovative approach prompts the need for custom strategies capable of accurately predicting various weather conditions using multiple visual cues within a single image.

Furthermore, the effectiveness of the model hinges on its compact size and computational speed. This requirement springs from the envisioned application of the model in vehicles capturing real-world scenes, where swift processing and minimal computational load are pivotal. While the allure of transfer learning using pre-trained models like those from ImageNet is strong, potential challenges arise. Adapting models trained on general object recognition tasks to the specialized context of weather classification might lead to concerns about overfitting, given the disparities in relevant features.

In essence, the arena of multi-label weather condition classification remains relatively uncharted, presenting a captivating avenue for pioneering exploration. The aim is to leverage the advancements in CNN architectures, particularly the capabilities of EfficientNet while addressing the nuanced intricacies of weather classification. By adapting and extending existing methodologies to this distinctive context, this research strives to contribute to the evolution of the niche realm of computer vision and weather comprehension.

3 Methodology

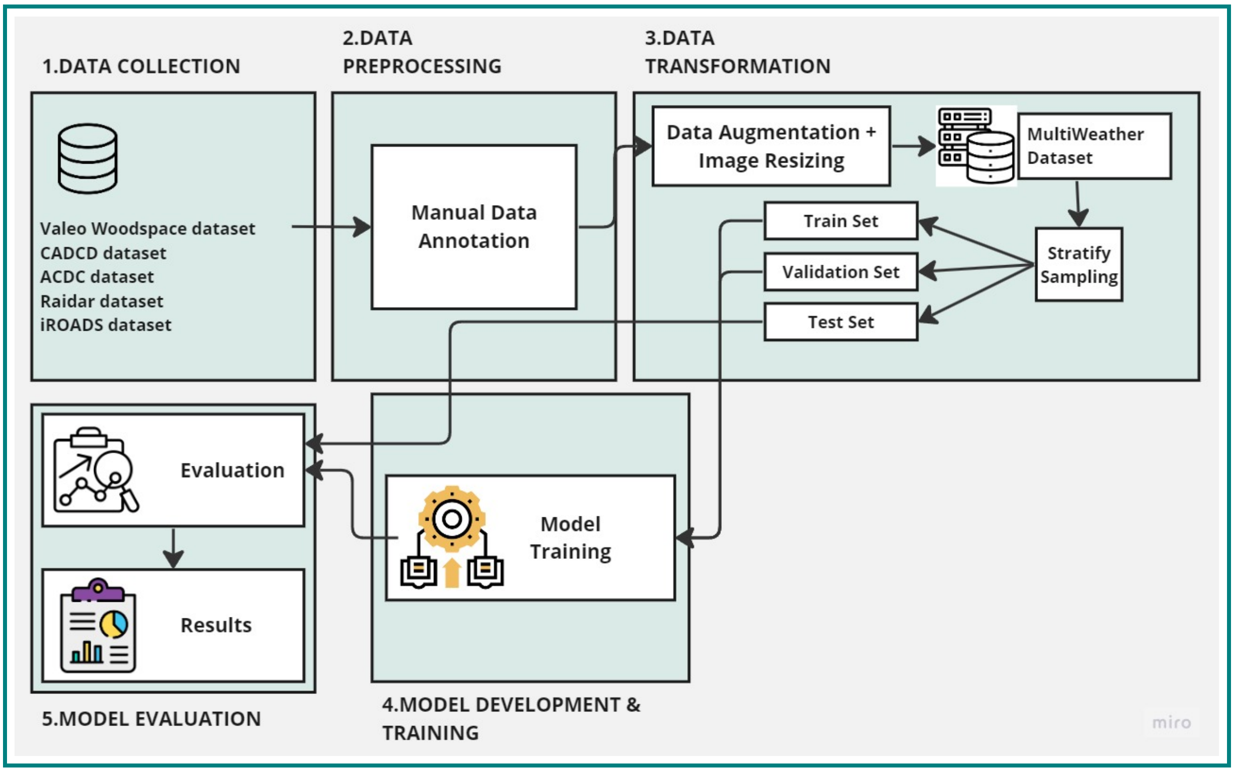


Figure 6: Project Methodology

The methodology employed for this study encompasses distinct phases: Data Collection, Data Preprocessing, Data Transformation, Model Development and Training, as well as Model Evaluation. Each phase plays a crucial role in the process of constructing and training the multi-label weather classification model.

3.1 Data Collection

Initially, the intention was to exclusively utilize the private data provided by Valeo, primarily due to the limited variety present in their public WoodSpace dataset. However, navigating through the intricacies of licensing and GDPR regulations posed significant challenges in obtaining the complete Valeo dataset. Consequently, the decision was made to solely rely on the public dataset offered by Valeo, complemented by the preannotation JSON files derived from drivers' tablets.

Subsequently, a script was executed to extract essential variables from the preannotation JSON files, facilitating the creation of an Excel file housing this comprehensive dataset. Despite these efforts, it became apparent that the dataset's range of categories remained limited, with notable gaps in representation. To address this limitation, an extensive exploration of other publicly available datasets specifically geared towards autonomous driving was undertaken.

The creation of the ultimate training and validation dataset involved a careful curation process aimed at representing a wide spectrum of weather and driving conditions. Various publicly available datasets were combined to achieve a diverse and robust collection.

3.1.1 Valeo WoodScape Dataset

(a) WoodScape Sample 1 (b) WoodScape Sample 2 (c) WoodScape Sample 3

Figure 7: Vale WoodScape Sample Images

Valeo's WoodScape dataset enriches the landscape of autonomous driving technology, providing a pivotal foundation for research and development. In the realm of computer vision algorithms, there's a dearth of adequate public datasets for evaluating fisheye camera applications, despite their value in offering expansive field views. To bridge this gap, we introduce WoodScape: an expansive fisheye automotive dataset named in honor of Robert Wood, the inventor of the fisheye camera. With four surround-view cameras and nine tasks, WoodScape features instance-level semantic annotations for 10,000+ images, along with annotations for over 100,000 images across other tasks. WoodScape's design promotes the development of fisheye-adapted algorithms, moving beyond simplistic rectification methods and propelling advancements in automotive vision systems. (Yogamani et al.; 2019)

Within this framework, we harnessed 2,706 images from the WoodScape dataset, representing diverse weather conditions including rain, dry, cloudy, and clear scenarios. These images provide a comprehensive range of real-world driving situations, further bolstering the dataset's value and applicability in refining autonomous driving technologies.

3.1.2 ACDC Dataset

(a) ACDC Sample 1 (b) ACDC Sample 2 (c) ACDC Sample 3

Figure 8: ACDC Sample Images

The ACDC dataset, or Adverse Conditions Dataset with Correspondences for Semantic Driving Scene Understanding, provides 4,006 images which evenly covers four adverse conditions: fog, nighttime, rain, and snow. This dataset offers high-quality fine pixel-level semantic annotations, along with corresponding normal-condition images and binary masks differentiating clear and uncertain semantic areas. These annotations support both standard and newly introduced uncertainty-aware semantic segmentation tasks.

Data collection involved capturing real-world scenes in Switzerland using a GoPro Hero 5 camera. GPS readings facilitated image-level correspondences between adverse and normal conditions. The dataset is divided into training, validation, and test sets for each adverse condition, with a focus on creating a challenging benchmark for semantic segmentation models. (Sakaridis et al.; 2021)

Utilising 3,002 images captured in adverse conditions such as rain, low light, overcast, and snow, these data were integral in training and evaluating our models.

3.1.3 Raidar Dataset

(a) Raidar Sample 1 (b) Raidar Sample 2 (c) Raidar Sample 3

Figure 9: Raidar Sample Images

The dataset "RaidaR: A Rich Annotated Image Dataset of Rainy Street Scenes" introduces a collection of rainy street scene images for advancing autonomous driving research. Comprising an extensive 58,542 rainy images, including 5,000 annotated with semantic segmentations and 3,658 with object instance segmentations, RaidaR encompasses a wide spectrum of authentic rain-induced effects such as fog, droplets, and road reflections. Acquired through a roof-mounted camera platform in Metro Vancouver, Canada, the dataset aids in enhancing data-driven machine perception during adverse weather. To streamline annotation, a semi-automatic approach combining manual segmentation and cross-validation-based automated processing was developed, significantly accelerating the process. This dataset's significance is demonstrated by its ability to boost segmentation algorithm accuracy through data augmentation, alongside the introduction of an innovative image-to-image translation algorithm for manipulating rain artifacts, leveraging the annotated RaidaR dataset. (Jin et al.; n.d.)

The Rich Annotated Image Dataset of Rainy Street Scenes (RAIDAR) also made a valuable contribution, contributing 948 images that predominantly featured rainy and low-light conditions. This dataset enhanced the collection by supplementing instances of rainy weather and enriching the diversity of the dataset.

3.1.4 CADCD Dataset

Furthermore, the Canadian Adverse Driving Conditions Dataset (cadcd) chipped in with 100 images, primarily focusing on snow conditions. Despite its smaller size, it injec-

(a) CADCD Sample 1 (b) CADCD Sample 2 (c) CADCD Sample 3

Figure 10: CADCD Sample Images

ted a crucial element into the dataset by introducing scenarios specific to snow-covered environments, an integral component of comprehensive weather classification.

The Canadian Adverse Driving Conditions Dataset (CADC) dataset is curated for testing autonomous driving algorithms in winter conditions. It features lidar point clouds and images from eight cameras, synchronized with GPS/IMU data. Annotated data includes object positions, dimensions, orientations, and attributes like vehicle types, pedestrian ages, and snowfall levels. The dataset emphasizes diverse scenarios with varying snowfall intensity. With thousands of instances, CADC aids in 3D object detection and localization research. With over 56,000 labeled images, this dataset proves invaluable for advancing autonomous driving algorithms, especially in challenging winter conditions. (Pitropov et al.; 2020)

100 snow scenes from the CADCD dataset were used in this project.

3.1.5 iROADS Dataset

(a) iROADS Sample 1 (b) iROADS Sample 2 (c) iROADS Sample 3

Figure 11: iROADS Sample Images

The iROADS Dataset, a comprehensive collection of 4656 image frames, offers a valuable external resource for evaluating machine learning models. The dataset comprises seven distinct categories, capturing images from moving vehicles in diverse weather and lighting conditions. These categories include Daylight, Night, Rainy day, Rainy night, Snowy, Sun strokes, and Tunnel scenarios. (Rezaei and Terauchi; 2014)

In light of the substantial use of augmented images in our primary dataset due to the scarcity of certain vital scenarios, the prevalence of augmented data surpassed original instances. To mitigate potential concerns of overfitting and ascertain the model's generalization capabilities, we adopted a cautious approach. Consequently, we selected 30 images from the iROADS Dataset, a previously unexplored and distinct dataset, to assess our model's performance. By evaluating our model in an unfamiliar context with diverse conditions, we sought to validate its robustness and confirm its effectiveness beyond its exposure to augmented data. This step offered insights into our model's adaptability to new and challenging scenarios.

3.1.6 Dataset Conclusion

Through the strategic fusion of these datasets, the final MultiWeather image dataset was meticulously designed to encompass a wide array of weather conditions, lighting scenarios, and surface conditions. This compilation mirrors the intricate and multifaceted environments encountered in real-world autonomous driving scenarios.

3.2 Data Preprocessing

During the data preprocessing phase, a crucial and time-intensive aspect was the manual annotation process. While the initial dataset included preannotations from the WoodScape dataset, sourced from the tablets within the vehicle operated by the driver, a dedicated Python script was employed to extract these preannotations and consolidate them into a single Excel file. However, this preliminary step was just the starting point.

The subsequent challenge emerged from the necessity to verify and correct these preannotations. Human errors, whether due to misinterpretation or other factors, had to be diligently identified and rectified to ensure the highest annotation accuracy. This validation process demanded meticulous attention and consumed a substantial amount of time due to the sheer volume of data and the intricacies involved in scrutinizing each annotation for discrepancies.

Furthermore, a significant portion of the dataset lacked any preannotations or labels altogether. This gap mandated an extensive manual annotation effort. Each unlabeled data instance required manual assessment and labeling, a task that involved an inherent understanding of the objects within the data frames, their characteristics, and their spatial relationships.

To accomplish this, annotations were progressively added into two Excel files (one for the main dataset and one for a completely new testing dataset), creating a comprehensive record of labeled data. The Excel files were subsequently transformed into a CSV format for standardized storage and ease of integration with machine learning pipelines.

In summary, manual annotation played a pivotal role in the data preprocessing stage, ensuring the accuracy and reliability of the dataset. This process, while time-consuming, was indispensable in providing a well-annotated dataset that forms the foundation for subsequent model training and evaluation efforts.

3.3 Data Transformation

In the data transformation phase, several essential steps were undertaken to enhance the quality and suitability of the dataset for subsequent machine learning tasks. The process began with image resizing, where all images were uniformly resized to dimensions of (240 x 240) pixels. This standardization not only facilitated computationally efficient processing but also ensured consistent input dimensions for the models.

Recognizing the need to mitigate data scarcity within specific categories, we strategically adopted image augmentation techniques. The rationale behind employing image augmentation was to counterbalance the limited availability of data in certain categories, which could otherwise lead to biased model training. By employing brightness and contrast adjustments with a controlled range of 0.8 to 1.2 with an augmented factor set to 3, a diverse array of image variations was generated. This augmentation process not only expanded the dataset but also bolstered the model's resilience and adaptability to different scenarios. Its profound impact was particularly evident in categories initially constrained by a shortage of original images, enabling the model to gain better insights and generalize effectively.

Ultimately, the final MultiWeather dataset was composed of a total of 6,659 original images, supplemented by an additional 17,439 images generated through augmentation techniques. This comprehensive dataset, containing both original and augmented images, provides a rich and diverse set of data points for training and evaluating machine learning models across various attributes and scenarios.

Table 3: Final Dataset Label count

Class	**Label**	**Count**
`scene_sky_cover`	Cloudy	12003
`scene_sky_cover`	Clear	6845
`scene_sky_cover`	Invisible	5250
`scene_climatic_conditions`	Dry	13161
`scene_climatic_conditions`	Snow	5490
`scene_climatic_conditions`	Rain	5447
`scene_surface_conditions`	Dry	13161
`scene_surface_conditions`	Wet	10937
`lighting_conditions`	Daylight	14458
`lighting_conditions`	Lowlight	9640

Table 4: Final Dataset for Multi-label model label count

Label	**Count**
Clear_Dry_Dry._Daylight	6717
Clear_Dry_Dry._Lowlight	128
Cloudy_Dry_Dry._Daylight	644
Cloudy_Dry_Dry._Lowlight	422
Cloudy_Rain_Wet_Daylight	1604
Cloudy_Rain_Wet_Lowlight	3840
Cloudy_Snow_Wet_Daylight	5490
Invisible_Dry_Dry._Lowlight	5250

In the context of an ever-expanding data landscape, the prominence of sampling has surged. As data volumes continue their exponential growth, the landscape of algorithmic design faces persistent challenges. Despite the strides made in devising scalable algorithms capable of directly accommodating extensive datasets, exemplified by works such as (Boyd et al.; 2011) and (Owen et al.; 2011), certain traditional algorithms mandate the reduction of data scale to manageable proportions. Within this milieu, sampling emerges as a strategic and efficacious approach to diminishing data dimensions while preserving intrinsic data attributes. Noteworthy achievements have arisen from the fusion of conventional algorithms with sampling methodologies. For instance, (Dasgupta et al.; 2009) illuminate the potential of meticulous sampling techniques, whereby solutions to subsampled instances of linear regression problems yield robust approximations to the original challenges, buttressed by sound theoretical underpinnings.

Upon completing image preprocessing, the dataset was split into distinct subsets for training, validation, and testing using the 80/10/10 technique. To achieve a balanced representation of the dataset's attributes, a stratified sampling strategy was employed. The dataset was stratified based on all four attributes: scene sky cover, scene climatic conditions, scene surface conditions, and lighting conditions. This approach ensured that each subset accurately represented the distribution of attributes found in the original dataset, preventing bias and enhancing the model's generalization capabilities.

Table 5: Final dataset Training/Validation/Test split

Label	**Training Count**	**Validation Count**	**Test Count**
Clear_Dry_Dry._Daylight	6717	6717	6717
Clear_Dry_Dry._Lowlight	128	128	128
Cloudy_Dry_Dry._Daylight	644	644	644
Cloudy_Dry_Dry._Lowlight	422	422	422
Cloudy_Rain_Wet_Daylight	5490	5490	5490
Cloudy_Rain_Wet_Lowlight	3840	3840	3840
Cloudy_Snow_Wet_Daylight	5490	5490	5490
Invisible_Dry_Dry._Lowlight	5250	5250	5250

Table 6: Extra testing dataset combined Label count

Label	**Count**
Clear_Dry_Dry._Daylight	0
Clear_Dry_Dry._Lowlight	0
Cloudy_Dry_Dry._Daylight	10
Cloudy_Dry_Dry._Lowlight	0
Cloudy_Rain_Wet_Daylight	10
Cloudy_Rain_Wet_Lowlight	0
Cloudy_Snow_Wet_Daylight	0
Invisible_Dry_Dry._Lowlight	10

3.4 Model Development and Training

In the domain of model development and training, our approach encompassed the creation of six distinct models tailored to our objectives:

The first model was crafted utilizing EfficientNetB1 and leveraged transfer learning, making use of pretraining on the expansive ImageNet Dataset. This strategy enabled the model to harness the knowledge gained from a vast array of images.

In addition, we devised four separate simplified Convolutional Neural Network (CNN) models. Each of these models was trained individually to classify one of the fundamental classes: sky cover, climatic conditions, surface conditions, and lighting conditions. This granularity allowed us to focus on precise attributes for classification.

Notably, we introduced a sophisticated multilabel model capable of collectively classifying all aforementioned classes. This comprehensive model amalgamated the diverse attributes, enabling simultaneous analysis and identification.

3.4.1 EfficientNetB1 Model

In the preliminary stages of model development, the focus was directed towards harnessing the power of the EfficientNetB1 architecture. Its attractiveness lay in its pre-trained status on the ImageNet dataset, offering a compelling starting point for the classification task. Tailoring this architecture to our specific requirements, a seamless integration was ensured with input data - images characterized by dimensions of 240x240 pixels and featuring 3 RGB color channels.

The journey commenced by streamlining the EfficientNetB1 base model through the removal of its upper layers, channeling its capabilities towards feature extraction. Mitigating overfitting and fostering adaptability, the inclusion of a Global Average Pooling 2D layer was followed by a dropout layer with a 0.5 dropout rate. Subsequently, a dense layer boasting 512 units and ReLU activation was strategically introduced to uncover intricate feature relationships.

A subsequent layer of dropout, characterized by a higher dropout rate of 0.9, was strategically added, acting as a robust buffer against overfitting. Culminating in significance was the output layer, housing a dense layer equipped with a softmax activation function and 3 units tailored to reflect the 'scene_sky_cover' class categories. The utilization of softmax activation facilitated the generation of class scores founded on probabilities.

The intricate details were meticulously orchestrated during the model's assembly. Compilation entailed the Adam optimizer and the 'sparse_categorical_crossentropy' loss function, a tailored combination well-suited for our integer label format. Rigor extended to the training phase, marked by a training duration of 5 epochs due to time constraints. Simultaneously, the ModelCheckpoint callback dutifully safeguarded the best-performing model throughout this abbreviated training endeavor.

Marking its debut as the first experimental model, this architecture emerged as a crucial litmus test to evaluate the viability of transfer learning within our specific context. The primary objective encompassed the establishment of a strong foundational framework, paving the path for the realization of our project's overarching aspirations.

3.4.2 Individual Attribute Classification Models

For attribute-specific classification, we developed 4 individual Convolutional Neural Network (CNN) models, each targeting a distinct attribute prediction. The architecture remains consistent among these models, differing solely in the output class they predict.

The architecture commences with data preprocessing, involving ImageDataGenerator for input transformation. Images are resized to 240x240 pixels and pixel values are normalized. The model architecture features Convolutional layers with 32 and 64 filters, enhanced by ReLU activation. MaxPooling layers downsample the image representation.

A flattened image vector transitions into a Dense layer with 128 units, followed by a Dropout layer to combat overfitting. The model culminates with a Dense layer equal to the class count (3 for scene sky cover) and utilizes softmax activation for class score generation.

Optimization employs the Adam optimizer and 'sparse_categorical_crossentropy' loss.

Training oversight integrates an EarlyStopping callback to curtail overfitting, and a ModelCheckpoint callback preserves the best model checkpoint during training.

This architecture's utilization of convolutional layers and dropout underscores its potential for efficient and generalized attribute-specific classification within our multi-attribute framework.

3.4.3 Multi-Label Classification Model for Simultaneous Attribute Prediction

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 238, 238, 32)      896

 max_pooling2d (MaxPooling2  (None, 119, 119, 32)      0
 D)

 max_pooling2d_1 (MaxPoolin  (None, 59, 59, 32)        0
 g2D)

 conv2d_1 (Conv2D)           (None, 57, 57, 64)        18496

 max_pooling2d_2 (MaxPoolin  (None, 28, 28, 64)        0
 g2D)

 flatten (Flatten)           (None, 50176)             0

 dense (Dense)               (None, 128)               6422656

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 8)                 1032

=================================================================
Total params: 6443080 (24.58 MB)
Trainable params: 6443080 (24.58 MB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 12: Multi-Label model architecture

In response to the limited prior research in this field, the project delved into unexplored territory, adopting a unique approach. A strategic decision was made to implement a multi-label classification framework as a means to achieve our objectives. This approach was chosen upon recognizing the potential benefits of amalgamating attributes into a single label, which could yield novel insights and more comprehensive predictions. To operationalize this strategy, the attributes were combined, leading to the emergence of eight distinct labels spanning the entire dataset.

The architectural foundation of this model (Figure 14) is built upon a Sequential structure—a prevalent choice for constructing neural networks. The model commences with a Convolutional layer, where 32 filters with a 3x3 size are applied to input images. These filters serve to extract specific features, enhanced by the Rectified Linear Unit (ReLU) activation function that introduces non-linearity, empowering the network to capture intricate data relationships. The input shape of (240, 240, 3) signifies the image dimensions of 240x240 pixels with three RGB color channels.

Max-Pooling layers are deployed to downsample image representations, diminishing dimensions while preserving key information. These layers operate on 2x2 regions and identify maximum values within each region.

Subsequently, an additional Convolutional layer, incorporating 64 filters, enriches the network's feature extraction capabilities. Consecutive Max-Pooling layers continue the downsampling process.

In preparation for the final classification phase, the 2D image representation undergoes flattening into a 1D vector. This vector is then linked to a Dense (fully connected) layer comprising 128 units, empowering the network to discern intricate patterns and relationships within the extracted features.

A Dropout layer is introduced, implementing a 0.5 dropout rate. This layer selectively deactivates 50% of neurons during training, serving as a regularization mechanism to combat overfitting.

The ultimate Dense layer encompasses units equal to the number of classes in the classification problem, totaling eight in this instance. Employing the softmax activation function, the model's outputs transform into probabilities for each class, enabling simultaneous multi-label predictions.

Model compilation involves the Adam optimizer, adjusting weights during training, and the sparse categorical cross-entropy loss function—ideal for multi-label classification. Accuracy is adopted as the metric to monitor training progress (Figure 14).

The training phase incorporates the EarlyStopping callback, terminating training when validation accuracy plateaus, and the ModelCheckpoint callback, preserving the optimal model during training.

This architectural blueprint mirrors the design of the four distinct CNN models, ensuring uniformity and leveraging the established effectiveness of this configuration in simultaneous multi-label prediction. Through diligent training, the model endeavors to capture intricate attribute-image relationships, culminating in accurate multi-label predictions.

3.5 Model Evaluation

In the context of our evaluation methodology, a comprehensive assessment of the models was undertaken. This entailed dual evaluations using distinct datasets, executed through a consistent framework. Initially, a subset of 10% from the original dataset was employed to gauge model performance on familiar grounds. This process involved encoding target labels, configuring relevant data generators, and analyzing essential metrics such as test loss and accuracy. These evaluations yielded insights into the models' behaviors within known data.

Subsequently, the evaluation ventured into uncharted domains by subjecting each model to the iROADS dataset—an unexplored collection comprising 30 images. Employing a parallel evaluation protocol, we scrutinized models' abilities to generalize across diverse data sources. This comprehensive approach was consistently applied to all six models, encompassing EfficientNet, four single-label CNN models, and the multi-label model. The outcomes provided valuable perspectives on their adaptability, robustness, and generalization prowess.

This meticulous evaluation strategy facilitated a detailed assessment of each model's performance and enabled a comparative study of their responses in varying data contexts. These insights underpin the analysis of the forthcoming result, establishing a solid foundation of empirically-driven insights.

4 Design Specification

The design specifications of this project encompass a comprehensive delineation of the techniques, architecture, and framework that underpin the implementation. Central to the project's approach is the utilization of diverse Machine Learning models for scene classification. The proposed architecture entails a multi-pronged strategy, including a foundational EfficientNetB1-based model for holistic scene classification, four individual CNN models tailored for distinct attribute classifications, and a pioneering multi-label classification model capable of jointly predicting various labels.

The architecture of the EfficientNetB1-based model entails leveraging transfer learning from the ImageNet dataset. The model embraces Convolutional and Max-Pooling layers for feature extraction, followed by Dense layers for intricate pattern recognition. A balance between model complexity and overfitting is achieved through the integration of Dropout layers. This model serves as a cornerstone to benchmark the effectiveness of the project's approach.

Additionally, four standalone CNN models are developed, each targeting a specific attribute classification. These models emulate the EfficientNetB1-based architecture while focusing solely on one attribute, enhancing attribute-specific accuracy and interpretability.

The innovative multi-label classification model transcends traditional categorical boundaries, consolidating all attributes into a unified label. This approach facilitates nuanced interactions between attributes, enabling comprehensive scene characterization. The model encompasses Convolutional and Max-Pooling layers, followed by Dense and Dropout layers for effective feature learning and generalization.

The comprehensive design specifications encapsulate the project's ambition to unravel the potential of machine learning models in scene classification. By embracing both established and innovative architectures, the project seeks to achieve a holistic and accurate classification framework, illuminating the nuanced interplay between attributes for robust scene understanding.

5 Implementation

In the culmination of the implementation phase, a multifaceted set of outputs was generated, embodying the results of rigorous experimentation and model development. The primary outcome consisted of meticulously trained machine learning models, each tailored to distinct attributes for scene classification. These models were designed, executed, and fine-tuned within a Python environment, specifically leveraging TensorFlow and Keras libraries for seamless model construction.

To bolster the dataset and facilitate improved model generalization, image augmentation was performed utilizing the OpenCV (cv2) library. This entailed modifying brightness and contrast levels, resulting in a more diverse and robust dataset for training. The augmentation process was orchestrated within Jupyter Notebooks, harnessing the interactive and iterative capabilities of the environment.

The transformation of raw data into processed and augmented datasets, the creation of comprehensive models, and the intricate fine-tuning process culminated in the project's primary deliverables. Notably, the NVIDIA GeForce RTX 2070 GPU was instrumental in expediting the model training process, accelerating computations and enabling efficient exploration of hyperparameters.

Overall, the implementation phase encompassed an amalgamation of Python-based coding, integration of TensorFlow and Keras for model development, and utilization of OpenCV for data augmentation. These tools synergistically facilitated the creation of robust machine learning models poised for rigorous evaluation and subsequent analysis

6 Evaluation

The Evaluation section encompasses a series of rigorous experiments designed to thoroughly assess the performance and efficacy of various Convolutional Neural Network (CNN) classification models. These evaluations were conducted using two key techniques:

Accuracy: To measure the precision of the models, we employed the accuracy metric, defined as:

$$Accuracy = \frac{CorrectPredictions}{TotalPredictions} \times 100 \tag{1}$$

Loss: Additionally, we utilized the Mean Squared Error (MSE) to evaluate the models. MSE is calculated as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2}$$

Our investigation unfolds along four distinct paths:

6.1 Experiment 1: EfficientNetB1 Model with Transfer Learning

The initial experiment involved constructing an EfficientNetB1 model that utilized Transfer Learning, leveraging pre-trained weights from the ImageNet dataset. This exploration sought to determine the viability of Transfer Learning for addressing the unique classification challenges.

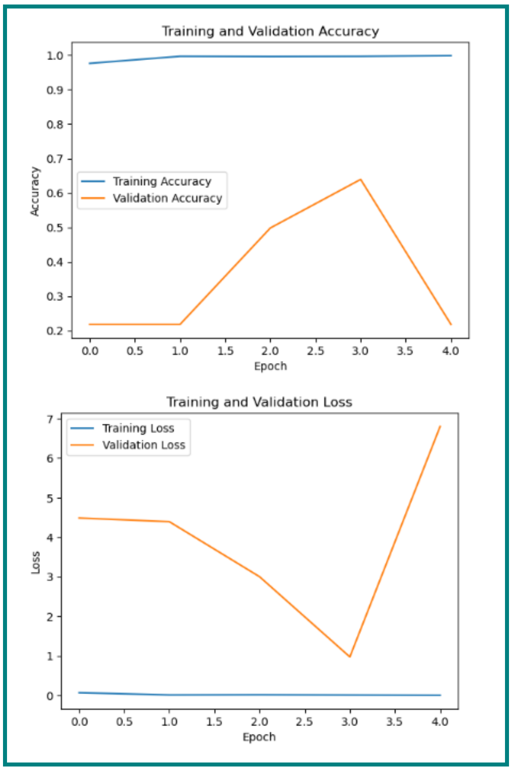


Figure 13: EfficientNetB1 Training and Validation Process

The training process of the EfficientNetB1 model unfolded across five epochs. The initial epoch exhibited a training loss of 0.0674 and an accuracy of 97.66%, while the validation loss and accuracy were recorded at 4.4855 and 21.79%, respectively. As training progressed, substantial improvements emerged, with the model achieving a training accuracy of 99.89% and a remarkably low training loss of 0.0034 in the final epoch. However, the validation accuracy fluctuated across epochs, peaking at 63.93% in the fourth epoch and declining to 21.79% in the fifth epoch. Each epoch took around 40 minutes, and despite the model's impressive training accuracy, achieving a consistent validation performance remains a notable challenge. (Figure 15)

These findings indicate a potential overfitting phenomenon. While the model demonstrated exceptional proficiency in learning from the training data, its validation accuracy did not mirror this performance. This divergence between training and validation accuracy suggests that the model might have become too specialized in the training data, leading to reduced generalization capability on new, unseen data. The overfitting concern prompts the need for further optimization strategies and parameter adjustments to enhance the model's ability to generalize effectively.

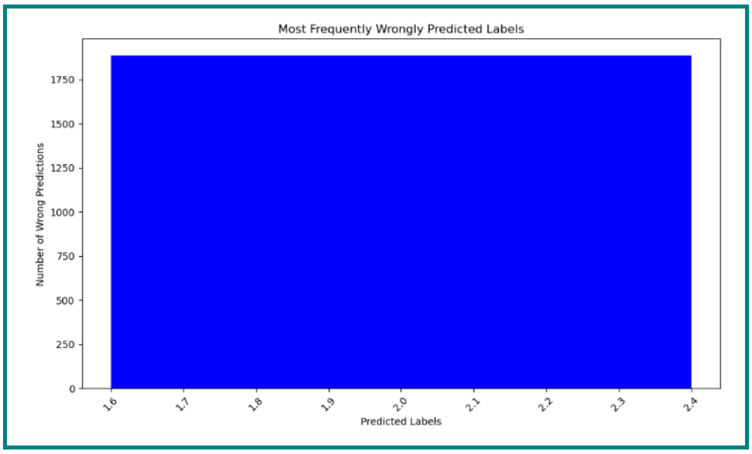


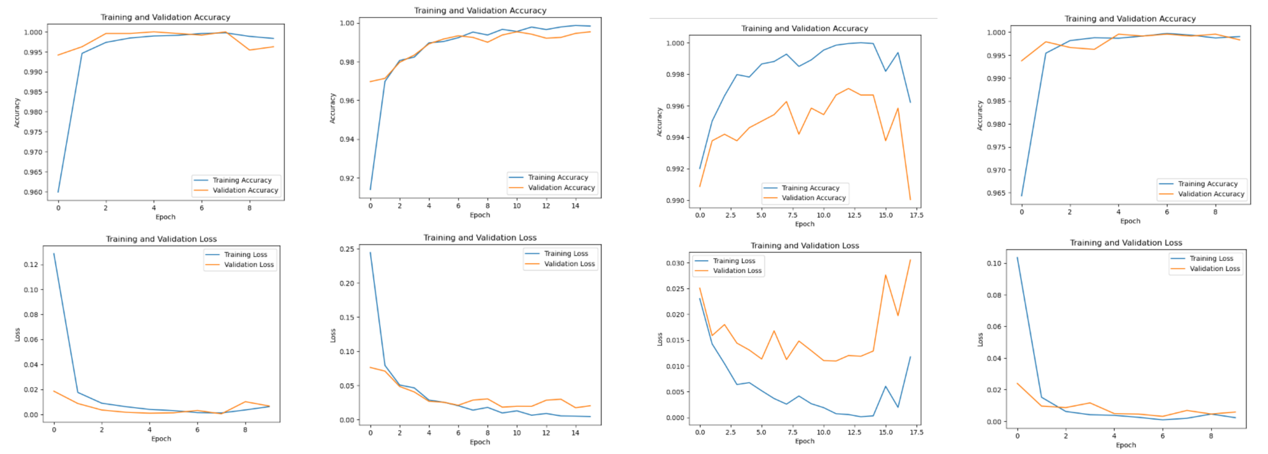
Figure 14: EfficientNetB1 Testing results

The evaluation focused on a test dataset, and the results indicated a test loss of approximately 6.798 and an accuracy of around 21.78%. This performance disparity between the training and testing phases signaled a potential issue of overfitting, where the model excelled in capturing training data patterns but struggled to generalize effectively to new, unseen data.

Among the wrongly predicted labels, a notable challenge emerged in predicting the "Invisible sky cover" label. This label signifies a dark sky with no visible clouds. The frequency of mispredictions for this label, exceeding 1750 instances (Figure 16), underscored the intricacies of recognizing these nuanced conditions. The model's struggles in predicting such complex attributes suggested a need for further enhancement to improve its accuracy and robustness in capturing these specific characteristics.

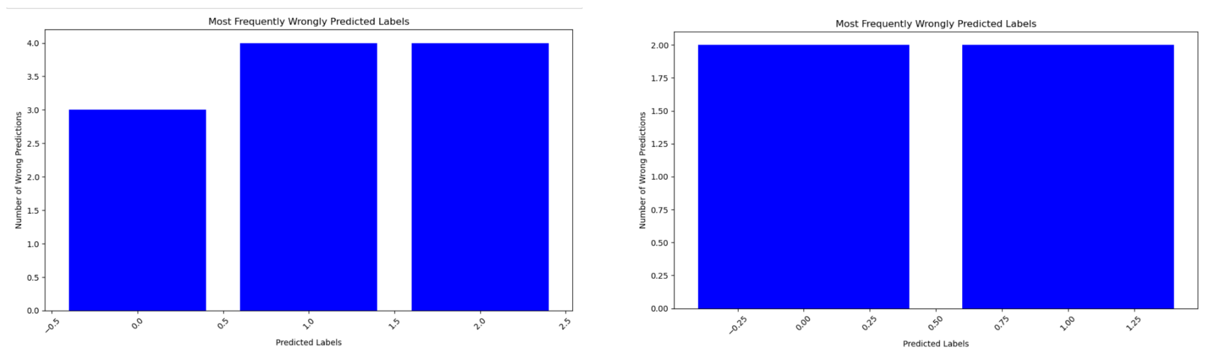
6.2 Experiment 2: Individual CNN Models for Classifying Attributes

Subsequently, four separate Convolutional Neural Network (CNN) models were developed, each tailored to classify a specific attribute. This approach facilitated an isolated analysis of each attribute's classification process.



(a) Sky cover (b) Climatic c. (c) Surface (d) Lighting

Figure 15: CNN Single models accuracy and loss



(a) Scene Climatic Conditions (b) Scene Surface Conditions

Figure 16: CNN Single models wrongly predicted labels

6.2.1 Sky Cover Classification Model

The Scene Sky Cover prediction model underwent 10 epochs of training, showing initial promise with 95.99% training accuracy and 99.42% validation accuracy. Subsequent epochs led to substantial improvement, reaching 99.90% training and 100% validation accuracy. Yet, validation accuracy exceeding 100% from the 5th epoch raised concerns about overfitting, warranting further investigation (Figure 17).

In testing, the model maintained exceptional performance. Across 19 batches and 19 steps, the model achieved a test loss of 0.0007065326208248734, coupled with a remarkable test accuracy of 100%. This underscores the model's proficiency in predicting scene sky cover categories. While showcasing impressive results, the sustained validation accuracy beyond 100% prompts the need for addressing potential overfitting for enhanced real-world reliability.

6.2.2 Climatic Conditions Classification Model

Throughout the training and validation process for the Climatic Conditions Classification Model, spanning 16 epochs, the model's accuracy steadily improved. Initial accuracy of 91.39% and validation accuracy of 96.97% ascended to 99.67% training accuracy and 99.54% validation accuracy. However, validation accuracy stabilizing after the 5th epoch may suggest overfitting concerns (Figure 17).

During testing, the model demonstrated exceptional performance, yielding a high test accuracy of 99.54% with a test loss of 0.0166. Despite its robustness, a few mislabellings were observed in predictions: "Dry" was misclassified three times, while "Rain" and "Snow" each incurred four incorrect predictions. These occurrences shed light on areas where further investigation could enhance the model's already impressive climatic conditions classification capabilities (Figure 18).

6.2.3 Surface Conditions Classification Model

The Surface Conditions Classification Model underwent rigorous training over 18 epochs, resulting in consistent improvement. The model's initial accuracy of 99.20% and validation accuracy of 99.09% was followed by an impressive ascent to 99.98% training accuracy and 99.71% validation accuracy. Notably, validation accuracy remained high and stable post the 7th epoch, which may warrant further scrutiny for overfitting. The loss during training also exhibited a consistent decline, indicating the model's capacity to learn and adapt effectively (Figure 17).

During the testing phase, the model exhibited remarkable performance, achieving a test accuracy of 99.83% with a test loss of 0.0083. The model misclassified "Dry" twice and "Wet" twice. (Figure 18).

6.2.4 Lighting Conditions Classification Model

The final model in this set, the Lighting Conditions Classification Model, showcased remarkable performance. Achieving a training accuracy of 99.91% and validation accuracy of 99.96% after just 10 epochs, this model demonstrated rapid convergence and effective learning (Figure 17). During testing, the model achieved a flawless accuracy of 100%, complemented by an impressively low test loss of 0.0008876. This highlights its ability to consistently predict lighting conditions accurately with minimal error, positioning the Lighting Conditions Classification Model as a robust and reliable solution for the task of lighting conditions classification.

6.3 Experiment 3: Multi-Label Classification Model

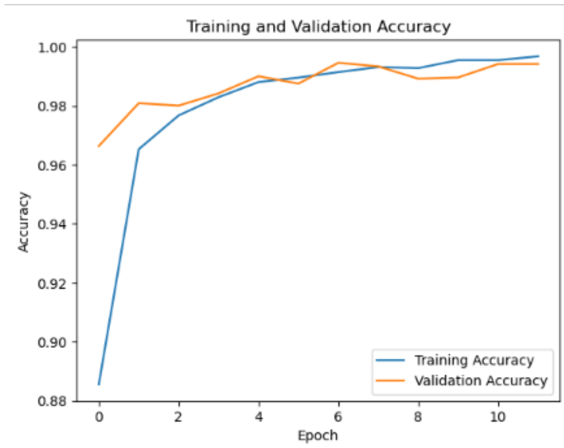


Figure 17: Accuracy

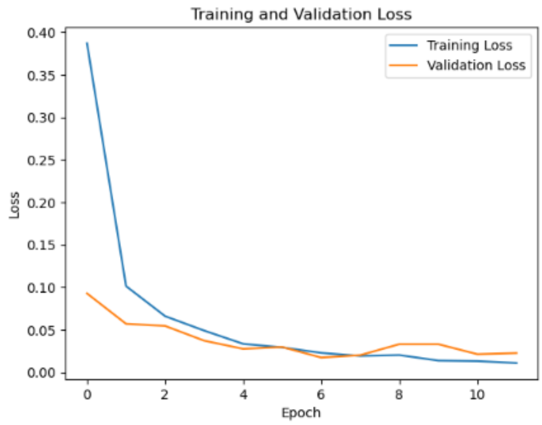


Figure 18: Loss

Figure 19: Multi-label Model Training and Validation Accuracy and Loss

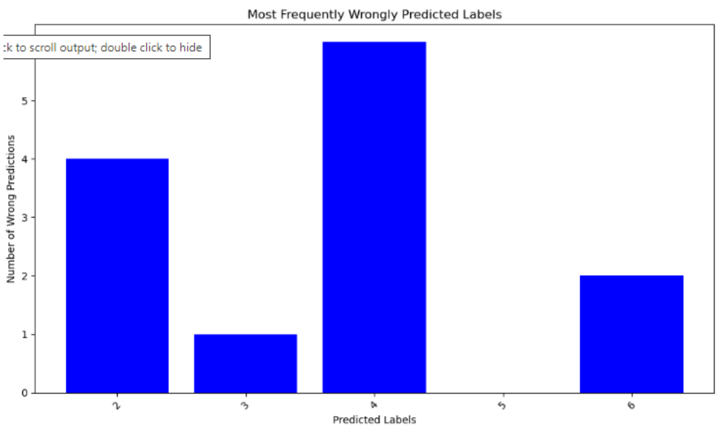


Figure 20: Multi-label Testing Wrongly Predicted Labels

A multi-label image classification model was developed to predict diverse attributes simultaneously, providing a comprehensive view of scene classification. The model showcased rapid improvement in validation accuracy during early epochs, culminating in an impressive test accuracy of 99.46%. The model achieved a low test loss of approximately 0.0161, highlighting its accurate predictions. With training conducted over 12 epochs, each lasting around 200 seconds, the model demonstrated efficient learning. (Figure 21) While the model displayed high validation and test accuracy alongside low loss, it's noteworthy that mispredictions were distributed across different labels (Figure 22), mitigating concerns of a singular label-related issue. This multi-label model not only offers insightful attribute relationships but also indicates promising potential for various applications.

6.4 Experiment 4: Evaluation on an Unknown Dataset

To ensure the robustness of the trained image classification model and address concerns regarding the potential impact of a high number of augmented images in the original

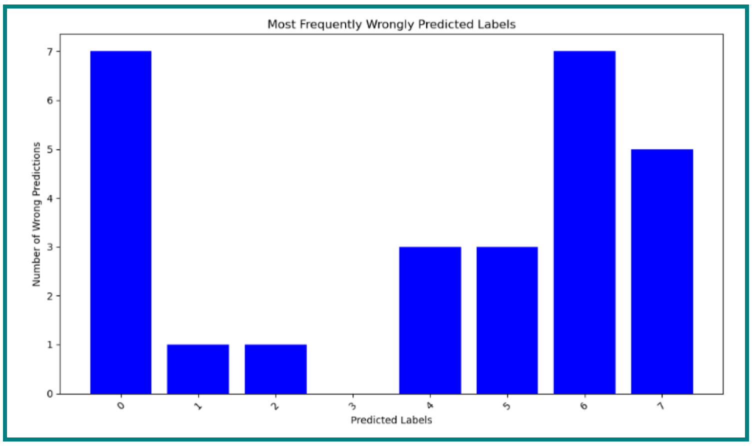


Figure 21: Unknown dataset testing Wrongly Predicted Labels

dataset, a dedicated evaluation was performed on an entirely new and distinct dataset. This dataset comprised a total of 30 images, distributed across three specific categories. The objective was to ascertain the model's adaptability to unseen data without the potential influence of excessive data augmentation.

The selected categories and their respective encoded values in the dataset were as follows:

- Cloudy_Dry_Dry._Daylight (Encoded Value: 2), Count: 10
- Invisible_Dry_Dry._Lowlight (Encoded Value: 7), Count: 10
- Cloudy_Rain_Wet_Daylight (Encoded Value: 4), Count: 10

In this assessment, the model achieved a test accuracy of 53.33% and a corresponding test loss of 5.84. The notable decrease in accuracy can be attributed to the fact that this dataset introduced entirely new and unseen images, encompassing only three categories and consisting of a mere 30 images in total. Despite these challenges, the model demonstrated its capability to generalize its learnings from the original training dataset to diverse and unencountered images. The misclassifications were distributed across different labels, further highlighting the model's resilience to unfamiliar scenarios and its potential for broader real-world applications.

6.5 Discussion

The evaluation of the trained models encompassed a range of experiments that aimed to assess their classification accuracy, generalization capabilities, and adaptability across diverse scenarios. The summarized results provided in Table 5 offer a comprehensive overview of the models' performance in various contexts.

Model	Test Acc.	Test Loss	Training t (per epoch)
EfficientNetB1 with TL	21.78%	6.798	45 min
Single-Label: Sky Cover	100%	0.000706	3.25 minutes
Single-Label: Climatic Conditions	99.54%	0.0166	3.25 minutes
Single-Label: Surface Conditions	99.83%	0.0083	3.25 minutes
Single-Label: Lighting Conditions	100%	0.0009	3.25 minutes
Multi-Label	99.46%	0.0162	3.25 minutes
Evaluation on Unknown Dataset	53.33%	5.84	-

Table 7: Model Performance and Evaluation

The single-label classification models exhibited exceptional accuracy, achieving a high test accuracy of more than 99% for each model. This outcome underscores the effectiveness of these models in accurately predicting individual labels pertaining to weather, road, surface, and lighting conditions. The success of these models can be attributed to the well-structured and balanced dataset, which enabled them to capture intricate patterns inherent in the images. Moreover, the comparatively shorter training time at 3.3 min per epoch on average, underlines the efficiency of these models, making them suitable for real-time applications.

The multi-label classification model further demonstrated its competence by achieving a test accuracy of 99.46%. This model's capability to predict multiple labels concurrently underscores its potential to capture complex attribute relationships within the images. Despite the added complexity of predicting multiple labels, the model delivered notable performance, a testament to the strength of the training dataset and the architecture of the model itself.

However, the evaluation on an unknown dataset yielded a lower accuracy of 53.33%. This result can be attributed to the unique characteristics of the unknown dataset, comprising only 30 images distributed across three categories. The limited number of images and categories in this dataset hindered the model's ability to generalize effectively, resulting in a reduced accuracy score. Additionally, the inclusion of categories that were not heavily represented during the dataset construction phase, such as Clear_Dry_Dry._Lowlight and Cloudy_Dry_Dry._Lowlight, may have contributed to the observed decrease in accuracy for those categories.

One notable aspect of the experiment was the combined utilization of fish-eye images and pin-hole images for training. Constructing a dataset that accommodates both imaging techniques required careful curation, yielding a dataset that effectively captures diverse driving scenarios. However, the inclusion of augmented images introduced complexity, which could have influenced model performance to some extent.

It's important to acknowledge the limitations arising from computational constraints. For instance, the training of the EfficientNetB1 model was restricted to a lower number of epochs due to time and GPU limitations. Extended training durations could potentially yield more refined results.

To further enhance the experimental framework, specific measures can be undertaken. Initially, focusing on a unified training dataset from a singular image type (either fish-eye or pin-hole) would amplify model consistency and overall effectiveness. Moreover, expanding the training dataset's size, accompanied by preannotation files, could streamline the labor-intensive manual annotation process. This augmentation should encompass a wider array of categories to ensure comprehensive coverage. This enlarged dataset would not only benefit the single-label models but would also provide essential support for training the multi-label model effectively.

In addition, potential avenues for improving model performance include fine-tuning the architecture, optimizing hyperparameters, and exploring advanced techniques suited to the model type. These refinements, coupled with the use of a robust GPU, would contribute to honing the models' predictive accuracy and their ability to adapt to diverse real-world driving scenarios.

In summary, a concerted effort toward consolidating the training dataset, obtaining preannotation files, exploring advanced techniques, and leveraging powerful GPU resources can collectively propel the models to a higher level of accuracy and applicability. These endeavors align with the evolving demands of accurate scene classification across a spectrum of driving contexts.

7 Conclusion and Future Work

In this research endeavor, we embarked on a journey to address the main research question: "How can an image classification Machine Learning model be strategically designed and proficiently implemented to optimize the effectiveness and efficiency of the daily data collection process in the autonomous driving industry?" Additionally, we delved into two sub-research questions: The first one aimed to refine the model's precision while reducing its computational demands, particularly in resource-constrained environments. The second sub-research question explored the potential of Transfer Learning to enhance the model's performance despite challenges related to limited data availability and computational resources.

7.1 Conclusion

Main Research Question: Our investigation into the primary research question has yielded valuable insights. Through meticulous model architecture design and comprehensive dataset augmentation, we have effectively showcased the feasibility of constructing robust image classification models customized to the dynamic requirements of the autonomous driving industry. Notably, we have developed four distinct single-label machine learning models, each catering to specific attributes within the driving environment. The amalgamation of both fish-eye and pin-hole images has proven pivotal in crafting a versatile model capable of accommodating an array of real-world scenarios. The attained accuracies of these single-label models further affirm the prowess of our approach in elevating the efficacy of the daily data collection process.

Sub-research Question 1: Addressing the first sub-research question, our findings have uncovered a pathway to enhance model precision while simultaneously minimizing its embedded footprint, all without compromising its practical utility. Through the strategic development of a multi-label classification model, we not only streamlined the classification process but also significantly reduced computational demands. This reduction in an embedded footprint not only aligns with our objective of optimising efficiency but also paves the way for more effective real-world applications, including integration within vehicles. The innovative nature of our approach resonates with the industry's need for quick and efficient decision-making capabilities.

Sub-research Question 2: While the implementation of Transfer Learning is a well-regarded approach, our exploration in the context of the second sub-research question unveiled challenges. The model's performance was hindered by the incongruity between the pre-existing image categories in the Transfer Learning dataset and the nuanced attributes of autonomous driving images. This mismatch underscored the importance of dataset relevance, urging us to forge ahead with alternative strategies that better suit the intricacies of our domain.

7.2 Future Work

Our discoveries hold profound implications for revolutionizing the autonomous driving industry. The model's remarkable capacity to precisely classify images captured from diverse perspectives presents a promising avenue for optimising data collection processes. However, similar to any research undertaking, certain limitations have come to the forefront. To fortify our contributions, prospective endeavors must center on the construction of more expansive and contextually relevant datasets, thereby enabling more profound model training. Moreover, venturing beyond the realm of Transfer Learning, possibilities could propel our model's predictive prowess to unprecedented heights.

In conclusion, this research has shed light on the complexities of image classification in the realm of autonomous driving, while also mapping out a trajectory for its enhancement and expansion. As we progress, these findings stand poised to serve as a cornerstone for future investigations, reshaping the landscape of image classification within the autonomous driving industry. With promising avenues for future work on the horizon, the augmentation of datasets to encompass diverse weather conditions, the broadening of the model's predictive scope to encompass attributes like surface material, environmental type (urban, rural), and time of day, along with the exploration of semantic segmentation, holds the promise of further refining and extending the practicality of the model.

It's important to highlight that a comprehensive multi-class weather dataset was meticulously constructed from scratch, consisting of images and corresponding CSV files, significantly enhancing the authenticity and robustness of this research endeavor.

References

Al-Haija, Q., Gharaibeh, M. and Odeh, A. (2022). Detection in adverse weather conditions for autonomous vehicles via deep learning, *AI* **3**: 303–317.

Boyd, S. et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers, *Foundations and Trends R in Machine Learning* **3**(1): 1–122.

Chen, L. et al. (2021). Review of image classification algorithms based on convolutional neural networks, *Remote Sens.* **13**: 4712.

Dasgupta, A. et al. (2009). Sampling algorithms and coresets for p regression, *SIAM Journal on Computing* **38**(5): 2060–2078.

Dhananjaya, M. M., Kumar, V. R. and Yogamani, S. (2021). Weather and light level classification for autonomous driving: Dataset, baseline and active learning, *arXiv preprint arXiv:2104.14042* .

Elhoseiny, M., Huang, S. and Elgammal, A. (2015). Weather classification with deep convolutional neural networks, pp. 3349–3353.

Huang, T. S. (1996). Computer vision: Evolution and promise, *1996 CERN SCHOOL OF COMPUTING* **8**(1): 21–24.

Jin, J., Fatemi, A., Lira, W. M. P., Yu, F., Leng, B., Ma, R., Mahdavi-Amiri, A. and Zhang, H. (n.d.). Raidar: A rich annotated image dataset of rainy street scenes. * indicates equal contribution. † indicates corresponding author.

Krishna, S. T. and Kalluri, H. K. (2019). Deep learning and transfer learning approaches for image classification, *International Journal of Recent Technology and Engineering (IJRTE)* **7**.

Li, Q. et al. (2021). Performance evaluation of deep learning classification network for image features, *IEEE Access* **9**: 9318–9333.

Li, Y., Song, Y. and Luo, J. (2017). Improving pairwise ranking for multi-label image classification, *IEEE Transactions on Image Processing* **26**(8): 4038–4048.

Nava, J. (2023). The evolution of mobility: Self-driving vehicles on the horizon in china, *chinadaily.com.cn* .
URL: $URL_OF_THE_ARTICLE$

Owen, S. et al. (2011). *Mahout in Action*, Manning Publications Co.

Pitropov, M., Garcia, D., Rebello, J., Smart, M., Wang, C., Czarnecki, K. and Waslande, S. (2020). Canadian adverse driving conditions dataset.

Real, E. et al. (2017). Large-scale evolution of image classifiers, pp. 2902–2911.
URL: *https://proceedings.mlr.press/v70/real17a.html*

Rezaei, M. and Terauchi, M. (2014). Vehicle detection based on multi-feature clues and dempster-shafer fusion theory, *Image and Video Technology* **8333**: 60–72.

Roser, M. and Moosmann, F. (2008). Classification of weather situations on single color images, *2008 IEEE Intelligent Vehicles Symposium*, pp. 798–803.

Sakaridis, C., Dai, D. and Gool, L. V. (2021). Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding, *ETH Zurich* . arXiv:2104.13395v3 [cs.CV] 1 Sep 2021.
URL: *https://acdc.vision.ee.ethz.ch*

Shaha, M. and Pawar, M. (2018). Transfer learning for image classification, *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, pp. 656–660.

Shakya, S. (2020). Analysis of artificial intelligence based image classification techniques, *Journal of Innovative Image Processing (JIIP)* **2**: 44–54.

Sothea, C. et al. (2020). Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data, *GIScience & Remote Sensing* **57**(3): 369–394.

Tan, M. and Le, Q. V. (2020). Efficientnet: Rethinking model scaling for convolutional neural networks, *arXiv preprint arXiv:1905.11946v5 [cs.LG]* .
URL: *https://arxiv.org/abs/1905.11946v5*

Xiao, H. et al. (2021). Classification of weather phenomenon from images by using deep convolutional neural network, *Earth and Space Science* **8**(5): e2020EA001604.

Yogamani, S. et al. (2019). Woodscape: A multi-task, multi-camera fisheye dataset for autonomous driving, *Computer Vision Foundation* .
URL: $https://openaccess.thecvf.com/content_ICCV_2019/papers/Yogamani_WoodScape_AMulti-Task_Multi-Camera_Fisheye_Dataset_for_Autonomous_Driving_ICCV_2019_paper.pdf$

Yogitha, R. P. et al. (2023). Breast histopathology image classification: A comparative evaluation of svm, random forest, and cnn, Hubli, India, pp. 1–6.

Zhao, B., Li, X., Lu, X. and Wang, Z. (2018). A cnn–rnn architecture for multi-label weather recognition, *Neurocomputing* **322**: 47–57.
URL: *https://www.sciencedirect.com/science/article/pii/S0925231218304261*