Informatics Institute of Technology

In Collaboration With

University of Westminster, UK

University of Westminster, Coat of Arms

DAugtify: Revolutionizing Computer Vision Performance with Automated Data Augmentation

A dissertation by

Mr. Idirimuni Pubudu Mihiranga De Silva

W1761268 / 2019285

Supervised by

Mr. Guhanathan Poravi

May 2023

Submitted in partial fulfillment of the requirements for the

BEng (Hons) Software Engineering degree at the University of Westminster.

ABSTRACT

The performance of modern Convolutional Neural Networks (CNN) heavily depends on the quality and the volume of the dataset. However, collecting proper training datasets in many real-world domains is well-known to be labor-intensive and expensive. Data augmentation (DA) is a widely accepted solution to improve the low diversity datasets. But selecting optimal DA technique combinations (DA policies) based on the given dataset using traditional trial-and-error approach is time-consuming and requires domain expertise.

In this study, the author attempts to automate the process of selecting optimal DA policies based on the given dataset by further enhancing Automated Data Augmentation (AutoDA) technologies. To achieve this goal, the author redefined the traditional AutoDA search space by reducing the number of hyper-parameters and proposing novel search space exploration strategy that utilizes neural network and gradient decent technologies. Through these modifications, the author was able to bridge the existing research gap of achieving efficiency and effectiveness of AutoDA solutions at the same time.

According to the conducted tests, the Wide ResNet 28x10 CNN model achieves 78% accuracy in 9 hours and 47 minutes on CIFAR10 classification when trained without DA. However, when trained with the proposed AutoDA system, the same network achieves 86% accuracy in 9 hours and 50 minutes on CIFAR10 classification. This suggests that by utilizing the proposed AutoDA system, the CNN model is able to achieve significant performance improvements while maintaining a competitive training time.

Subject Descriptions:

- Machine Learning → Low Diversity Datasets → Low Accuracy & Generalizability
- Improve Low Accuracy & Generalizability → Data Augmentation → Difficulties in Traditional Data Augmentation → Automation in Data Augmentation

Keywords: Data Augmentation, Automated Data Augmentation, Differentiable Programming, Convolutional Neural Network

DECLARATION

I confirm that this dissertation is the outcome of my own research and that it has not been submitted or is currently being submitted for any other degree or academic qualification provided by a university or institution. Any information obtained from prior works is appropriately acknowledged and credited to the original authors through proper citation.

Student Name: Idirimuni Pubudu Mihiranga De Silva

Registration Number: w1761268/2019285

Signature: ………………… **Date**: 2nd May 2023

ACKNOWLEDGEMENT

Completing this research project was a challenging yet rewarding experience, requiring extensive hours of researching, documenting, developing, and testing over the eight months. I would like to extend my gratitude to everyone who supported me throughout this journey.

Firstly, I would like to express my sincere thanks to my supervisor and module coordinator, Mr. Guhanathan Poravi, for his unwavering support, guidance, and encouragement. His expertise and dedication have been invaluable in ensuring that I did the research competently. I would also like to thank all the lecturers who provided me with their extensive knowledge throughout the years, which ultimately helped me complete this project.

Furthermore, I am grateful to the domain, technical experts, and evaluators of this research, who provided feedbacks, enabling me to refine my work and make it better.

Finally, I would like to express my heartfelt thanks to my friends and family, who have been my pillars of support throughout the final year project and the entirety of my university life. Their constant encouragement and belief in me have been instrumental in helping me achieve this milestone.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AI	Acritical Intelligence.
DA	Data Augmentation.
AutoDA	Automated Data Augmentation.
DA Policy	Data Augmentation Policy.
ML	Machine Learning.
AutoML	Automated Machine Learning.
DL	Deep Learning.
DP	Differentiable Programming.
MVP	Minimum Viable Product.
UI	User Interface.
SE	Software Engineering.
CV	Computer Vision.
CNN	Convolutional Neural Networks.
DFD	Data Flow Diagram.
LR	Literature Review.
GUI	Graphical User Interface.
GPU	Graphics Processing Unit.

CHAPTER 1: INTRODUCTION

1.1 Chapter Overview

The objective of the research is to present a novel efficient and effective approach to automate the data augmentation (DA) process in computer vision (CV) tasks by solving the limitations of current methodologies. This chapter outlines the research domain, goals, and objectives, as well as the evidence required to establish the research gap and the novelty of the study. Finally, the author identifies the challenges associated with this research and the strategies employed to overcome them.

1.2 Problem Domain

In recent years, the CV domain has focused on developing various Convolutional Neural Network (CNN) architectures that can learn information from digital images/videos to perform artificial intelligence (AI) tasks like image classification, segmentation, and object detection. The performance of these CNNs significantly rely on the quality and volume of the training dataset (Shorten and Khoshgoftaar, 2019). This is because CNNs can learn more information from high-quality and large training datasets (Khalifa et al., 2022; Shorten and Khoshgoftaar, 2019).

Overfitting and poor generalization performance are common issues in CNNs that occur due to poor training datasets (Shorten and Khoshgoftaar, 2019). To mitigate such issues, it is essential to utilize a diverse and comprehensive dataset during the CNN training process. However, collecting proper training datasets in many real-world domains is well-known to be labor-intensive and expensive. For example, medical image analysis. Hence, improving the performance of CNNs without proper training dataset is one of the key challenges (Khalifa et al., 2022; Shorten and Khoshgoftaar, 2019).

1.2.1 Data Augmentation

DA is a widely accepted solution to improve low-diversity datasets (Khalifa et al., 2022; Shorten and Khoshgoftaar, 2019). In general, the aim of DA is to artificially increase the size of the original dataset by extracting more information from the existing training dataset. The augmented data points will provide a wider diversity of possible information to the training dataset. Therefore, DA helps overcome the overfitting problem and performance improvements of CNNs. DA can be used to improve many types of data. For example, images, text, and audio. However, in this research work, the author only considers image DA techniques. According to Shorten and Khoshgoftaar, 2019, image DA can divide into two categories. Which are:

1. Basic/classical image manipulations DA techniques
    - Geometrics transformations
    - Color space transformations
    - Kernel filters
    - Mixing images
    - Random erasing
2. Deep learning DA techniques
    - Neural style transfer
    - Generative Adversarial Network-based DA, also known as GAN
    - Adversarial training

Though above mentioned DA techniques improve the performance of CNNs, not all DA operations are suitable for every dataset (Khalifa et al., 2022). This is because each dataset consists of unique characteristics. For instance, on the MNIST dataset, elastic distortions, translation, and scale DA techniques are commonly used and on natural datasets, such as ImageNet and CIFAR 10/100, image mirroring, color shifting, rotation, and random cropping are more conventional (Yang et al., 2022a).

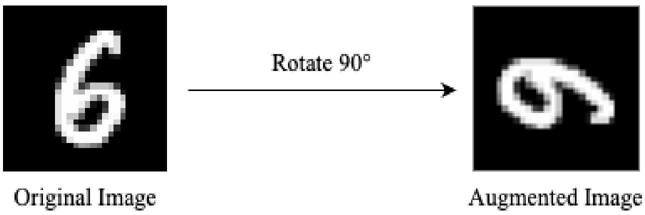


Figure 1: Selection of wrong data augmentation technique

Above figure illustrates that choosing inappropriate DA techniques can lead to the generation of unnecessary data points. Thus, further minimizing the performance of CNNs (Khalifa et al., 2022). The traditional technique for selecting the ideal set of DA operations (combination of multiple DA techniques named as a DA policy) for a given dataset involves a trial-and-error methodology. This process is both time-consuming and requires domain knowledge. Additionally, it may lead to the unwanted data collection, thereby wasting computational resources and efforts.

1.2.2 Automated Data Augmentation

As mentioned above, the traditional way of selecting optimal DA policies for a given dataset has series of difficulties. To mitigate such difficulties, a novel approach is to use automated data augmentation (AutoDA) techniques to automatically learn and design the optimal DA policies based on the given dataset (Cubuk et al., 2019a). In the CV domain, the AutoDA models aim to design the optimal DA policies that boost the performance of a target CNN. Moreover, recent research works on CV domain have shown that DA policies that are generated by AutoDA models perform better than manually designed DA policies (Cubuk et al., 2019a).

1.3 Problem Definition

DA is a widely accepted fruitful method to avoid overfitting issues and enhancing the performance of CNNs. However, currently, DA policies have been designed manually, and the optimal DA policies are dataset specific (Cubuk et al., 2019a; Khalifa et al., 2022). As a result, to design optimal DA policies manually for a given dataset requires a considerable amount of expertise in the DA domain, powerful computational resources, and a lot of time. So far, a significant focus of the researchers has been on refining the architectures of the CNNs to avoid data related issues. Less attention has been put into improving DA technologies, solving difficulties in traditional DA and automatically identifying optimal DA policies based on the dataset and task type (Cubuk et al., 2019a). The advancements in automated machine learning (AutoML) have shown promise in developing AutoDA systems to enhance CNN performance by automatically learning optimal DA policies (Yang et al., 2022a). However, limitations in existing AutoDA works highlight the need for further research to address the important problem of automating the DA process (Yang et al., 2022a).

1.3.1 Problem Statement

DA is a widely accepted fruitful technique to overcome poor dataset-related problems like overfitting, but it is not easy to design optimal DA policies based on the given dataset and task type.

1.4 Research Motivation

DA is a crucial task in the field of CV that has been extensively employed by researchers, students, and industries in various projects. However, standard DA techniques still contain a lot of manual work and need a decent amount of DA domain expertise to achieve the best results. Automating the process of designing DA techniques is a solution for all these shortcomings and will be a huge turning point in the CV domain.

1.5 Research Novelty

1.5.1 Problem Novelty

AutoDA techniques aims to handle the limitations of traditional DA that were first introduced by AutoAugment (Cubuk et al., 2019a) in 2019. However, the practical implementation of this techniques required extensive search time, which made them challenging to apply in day-to-day DA tasks. Later research in the AutoDA field attempted to reduce search time using various technologies. But most of these solutions still rely on high-end GPU resources (Li et al., 2020), which again limits their widespread use. As a result, the AutoDA problem remains unresolved until a new approach is introduced that addresses both search time and computational resource usage challenges (Yang et al., 2022a).

1.5.2 Solution Novelty

Currently, the most practical AutoDA solutions in the literature are search-free AutoDA architectures (Cubuk et al., 2019b; Yang et al., 2022a). However, these methods rely on static hyper-parameter values, which limit the generalization performance of the target CNN model (refer to section 2.5 for more details). Motivated by the findings of search-free AutoDA methodologies, the author of this study attempts to improve the generalization performance of target CNN by redefining the search space hyper-parameters and introducing a lightweight single hyper-parameter neural network for search space exploration, which has never been attempted before. Test results provided in section 8.4 show that the proposed AutoDA architecture can address both effectiveness and efficiency issues, making it widely applicable in the day-to-day DA tasks.

1.6 Research Gap

Upon conducting a comprehensive analysis of the available DA and AutoDA literature, the limitations of existing research works are outlined below.

High resource requirement and implementation complexity: The major drawback of existing AutoDA techniques is that they are highly resource intensive. For instance, Auto Augment (Cubuk et al., 2019a) requires 15000 NVIDIA Tesla P100 GPU hours on the ImageNet dataset to perform AutoDA tasks. Recent AutoDA works were able to reduce this time by leveraging powerful GPU resources and complex AutoDA architectures (Li et al., 2020). However, due to this high resource requirements and implementation complexity constraints, the widespread usability of AutoDA technologies is limited.

Limited generalization performance of target CNNs due to transfer learning: Existing AutoDA systems enhance the performance of state-of-the-art CNNs models on CIFAR-10/100, SVHN and ImageNet datasets (Yang et al., 2022a). In the real-world settings, these AutoDA models function by applying pre-learned DA policies from the aforementioned datasets. This approach is primarily adopted to avoid high resource consumption and implementation complexity. However, these policies may not be as effective on user-specific datasets, as each dataset consists of unique characteristics (Nanni et al., 2021).

1.7 Contribution to the Body of Knowledge

1.7.1 Contribution to the Research Domain (AutoDA)

Redefinition of traditional AutoDA search space and prove effectiveness: It has been demonstrated that the number of hyper-parameters in a traditional AutoDA search space has a direct impact on the time required to search for an optimal DA policy (Cubuk et al., 2019b). Therefore, the redefinition of a traditional AutoDA search space with a single hyper-parameter, and the subsequent demonstration of its effectiveness, represents a contribution.

Optimal DA policy search and evaluation strategy: Since the proposed search space comprises only a single hyper-parameter, it is crucial to develop a new strategy to explore this space. Thus, introducing a lightweight single hyper-parameter neural network as a replacement for the traditional AutoDA search space exploration strategy is another contribution.

Theoretical & technical contribution to standard AutoDA architecture: The combination of above theoretical and technical changes will enable the AutoDA task to be performed with fewer computational resources. Therefore, proving the efficiency and effectiveness of the proposed AutoDA architecture is also a contribution.

1.7.2 Contribution to the Problem Domain (Computer Vision)

DA plays a crucial role in the training pipeline of CNN models, especially when the training dataset is limited. Automating this process utilizing fewer computational resources and time have significant potential for AutoDA architectures to become a standard layer of CNN training pipelines in the future.

1.8 Research Challenges

The primary objective of this research project is to overcome the limitations highlighted in the literature and promote the widespread adoption of AutoDA techniques in practical applications. The following is a list of research challenges based on the proposed methodology.

Improve AutoDA architecture: Automating DA for CV tasks is a relatively novel concept and has not been thoroughly explored for real-world applications (Yang et al., 2022a). Therefore, it is essential to explore theoretical aspects and other ways of defining AutoDA architectures more efficiently and effectively within the given time and limited knowledge is a challenge.

Complexity due to the involvement of advancements in standard CNN training process: Since the AutoDA system aims to incorporate advancements in the standard CNN training process, it is essential to gain knowledge of CNN architecture design and workflow. However, since this involves the two domains of DA and CV, the learning curve is steep, and it is crucial to balance it within the given timeframe.

Measure effectiveness and efficiency: When evaluating the effectiveness and efficiency of the AutoDA system, it is necessary to train various CNN architectures on different datasets. However, the duration of this training varies based on the complexity of the CNN architecture. This makes it challenging to balance rapid development, improvement, and testing while efficiently managing computational resources.

1.9 Research Questions

RQ1: What are the newest improvements in the Software Engineering (SE) domain that can be utilized to perform AutoDA in a practical manner?

RQ2: How to design an AutoDA system that can outperform human-designed data augmentation heuristics and existing AutoDA systems?

RQ3: What potential challenges may arise during the design of an AutoDA system, and how can they be prevented or resolved?

1.10 Research Aim

The aim of this research is to design, develop and evaluate a system that automates the manual process of designing and fine-tuning optimal image data augmentation schemes for computer vision tasks with reduced human intervention.

To further elaborate on the research aim, this study aims to create a new AutoDA architecture that can perform AutoDA tasks while utilizing fewer computational resources and with reduced human intervention. The system will select the most effective DA policies based on the given low-diversity dataset and fine-tune them to enhance the diversity of given dataset further. Additionally, the proposed system and its components will undergo evaluation and testing to assess their impact on output quality and prove the hypothesis.

1.11 Research Objectives

Objective	Description	Learning Outcomes	Research Questions
Problem Domain	In-depth analysis of various research domains was performed. **RO1:** To identify the competitive and complex research domain for undergraduate research project.	LO1, LO4, LO8	RQ1
Literature Review	To fulfill the below requirements, a comprehensive review of the existing literature in the chosen problem domain is carried out. **RO2:** To identify and analyze the existing systems in the AutoDA domain. **RO3:** To find out the limitations, improvements, and research gaps in the AutoDA domain. **RO4:** To identify the ways of reducing computational power and complexity of existing AutoDA works without losing accuracy levels. **RO5:** To identify the technologies, algorithms, frameworks, and other required tools that are necessary for the development phase.	LO1, LO3, LO4, LO5, LO8	RQ1, RQ2, RQ3

	RO6: To identify the evaluation and benchmarking criteria and necessary metrics.		
Requirement Analysis	In-depth requirement analysis was performed to, **RO7:** To determine the most used DA techniques in CV domain. **RO8:** To gather requirements of an AutoDA system and understand end-user expectations. **RO9:** To gather insights and feedback of experts to improve the proposed system and to validate the hypothesis.	LO1, LO2, LO3, LO4, LO6, LO8	RQ1, RQ2, RQ3
Design	To design the proposed AutoDA system architecture, **RO10:** To design search space of proposed AutoDA solution. **RO11:** To design effective and efficient DA policy search and evaluation strategy. **RO12:** To design a method to perform DA using identified DA policies.	LO5, LO7, LO8	RQ1, RQ2, RQ3
Development	To develop the proposed AutoDA system in accordance with the specified design elements, as well as software and hardware prerequisites, **RO13:** To develop search space, search and evaluation strategies of proposed AutoDA system using appropriate software and hardware requirements. **RO14:** To develop other core functionalities of the proposed system using appropriate software and hardware requirements. **RO15:** To develop the GUI of the proposed AutoDA system.	LO5, LO7, LO8	RQ1, RQ2, RQ3
Testing and Evaluation	Testing and evaluating the proposed AutoDA system,	LO4, LO5, LO6,	RQ1, RQ2

	RO16: To create a suitable test plan for functional and non-functional testing. **RO17:** To gather insights of experts on proposed solution, test results and further optimizations. **RO18:** To benchmark the prototype against existing AutoDA works.	LO8	

Table 1: Research Objectives

1.12 Chapter Summary

This chapter discussed about the problem that is going to solve by this research work. Moreover, the research aim, the identified research novelty and gap, and the research challenges are also discussed by providing necessary evidence and a description of the DA problem domain. Additionally, the learning outcomes from this research module were mapped with the research objectives of the project according to the BEng (Hons) Software Engineering degree guidelines provided by the University of Westminster.

CHAPTER 2: LITERATURE REVIEW

2.1 Chapter Overview

This chapter is dedicated to exploring the current AutoDA literature, the various methodologies employed by researchers, and the results of previous studies. Furthermore, this chapter undertakes a critical evaluation of the limitations and advancements of current technologies and studies, as well as evaluation and benchmarking metrics that can be applied in the AutoDA domain.

2.2 Concept Graph

The concept map provides a brief overview of the topics covered during the literature review including existing solutions, potential technologies, and suitable evaluation metrics for the proposed system. The concept map can be viewed in **Appendix A**.

2.3 Problem Domain

2.3.1 Data-Related Challenges in Modern CNN Models

One of the key challenges in modern CNNs is dealing with insufficient training data. Limited training data can lead to several issues in CNNs like overfitting and poor generalization performance (Khalifa et al., 2022).

- **Overfitting** occurs when a CNN is trained on limited training data, causing it to become too specialized in learning the unique characteristics and noise of that specific data.
- **Poor generalization**, on the other hand, can be caused by overfitting, where CNNs are unable to perform well on unseen data, leading to biased and inaccurate predictions.

The following graphs show what overfitting looks like when visualizing training and testing performance over time.

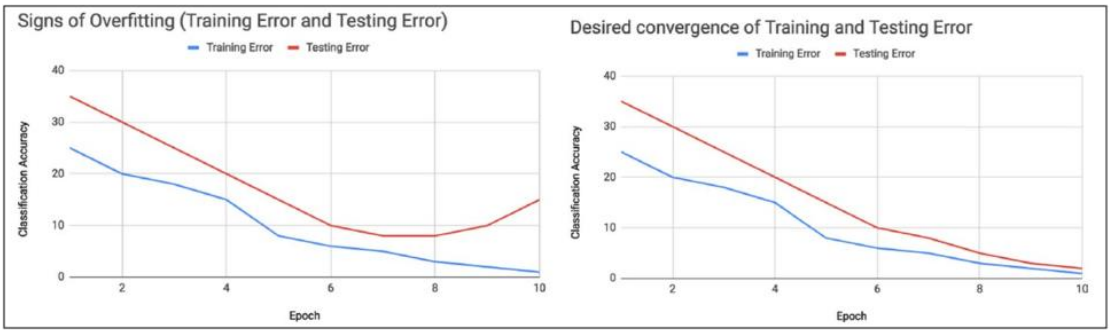


Figure 2: Visualization of model accuracies over training epochs (Shorten and Khoshgoftaar, 2019)

In a better CNN, the training and testing errors must gradually decrease with time (Shorten and Khoshgoftaar, 2019). Hence, a proper dataset is essential when training CNNs, as it ensures that the model has enough diverse data to learn and generalize patterns from.

2.3.2 Data Augmentation

DA is a widely accepted solution to overcome data-related challenges in CNNs. DA aims to increase the quality and quantity of the original training dataset by generating new variations from the existing data (Shorten and Khoshgoftaar, 2019). Data-oversampling and data-warping are two common ways to perform DA task.

- **Data-oversampling** involves generating new data points by replicating or duplicating existing data points.
- **Data-warping** involves modifying existing data points to create new, slightly different versions of original data points.

As a result of DA, CNNs can learn more diverse representations of the data, leading to a more accurate performance on unseen data. Image DA can divide into 2 major categories (Shorten and Khoshgoftaar, 2019):

- Basic/traditional data augmentation
- Deep learning data augmentation approaches

A comprehensive taxonomy of DA domain is placed in the **APPENDIX B**.

2.3.2.1 Basic/Traditional Data Augmentation

Due to the implementation simplicity basic DA is a widely used technique that involves applying geometric and color space transformation to original images to create new data points (Khalifa et al., 2022). The most common geometric transformations include rotation, scaling, and flipping and the most common color space transformation include adjustments like brightness, contrast, and hue. Though, basic DA is easy to use, it requires careful attention to selecting the appropriate magnitude (further elaboration on this can be found in section 2.3.3).

Figure 3: Example basic data augmentation techniques where (A) original Image, (B) horizontal flipping, (C) color space changes, (D) translation (Khalifa et al., 2022)

2.3.2.2 Deep Leaning Data Augmentation Approaches

Deep learning DA approaches use generative models to create synthetic images (Khalifa et al., 2022). Some of the popular deep learning DA approaches are Variational Auto Encoders (VAEs), Generative Adversarial Networks (GANs), and neural style transfer. Deep learning DA approaches are more powerful than the basic DA approaches because the objective of Deep learning DA approaches is to generate more realistic images using the original data (Khalifa et al., 2022). However, one of the noticeable disadvantages of these techniques is the high computational resource requirement.

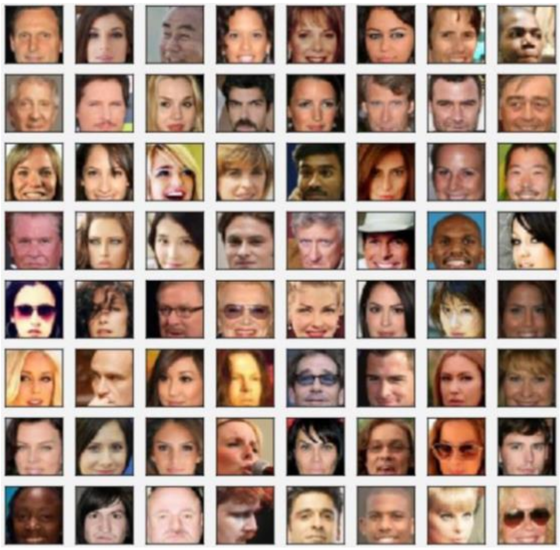


Figure 4: Example GAN generated image (source: https://shorturl.at/ckBU9)

2.3.3 How to Select Optimal Data Augmentation Schemes for Given Dataset?

Each dataset has unique characteristics. As a result, not all DA techniques are appropriate for every dataset. Choosing the wrong DA scheme can lead to a decrease in the performance of the CNN models (Cubuk et al., 2019a). For example, if the dataset contains images of animals, using a rotation data augmentation technique may work well, but if the dataset contains images of written numbers, the rotation data augmentation technique may produce unnecessary data.

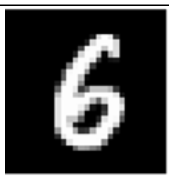
Original Data Point	Unnecessary Augmented Data Point
	
	

Table 2: Selection of wrong data augmentation technique (self-composed)

The naive way to find the optimal DA scheme based on the given dataset is to perform random data augmentation techniques to the given dataset and evaluate the performance of the CNN using augmented data (Cubuk et al., 2019a; Yang et al., 2022a). Therefore, selecting an optimal DA scheme based on the given dataset is time-consuming and requires domain expertise.

2.3.4 Automated Data Augmentation

To address the aforementioned issues, researchers have been exploring ways to automate the process of selecting optimal DA schemes based on the given dataset. This process is named Automated Data Augmentation (AutoDA). AutoDA techniques aim to find the optimal DA scheme that maximizes the performance of the target CNN (Cubuk et al., 2019a).

The recent research effort in the AutoDA domain has proved that AutoDA models can significantly improve model performance (Yang et al., 2022a). Currently, due to the complexity of DL based DA approaches, all the existing AutoDA methods consider automation in basic/traditional DA techniques (Cubuk et al., 2019a; Yang et al., 2022b).

2.3.4.1 Architecture of Standard AutoDA System

The architecture of the standard AutoDA system consists of 3 key components (Yang et al., 2022a):

1. **Search space:** is used to store all possible combinations of basic DA schemes.
2. **Search algorithm:** is responsible for exploring the search space and find optimal DA schemes for a given dataset and CNN model.
3. **Evaluation method:** is used to rank the effectiveness of the selected DA scheme using the search algorithm.

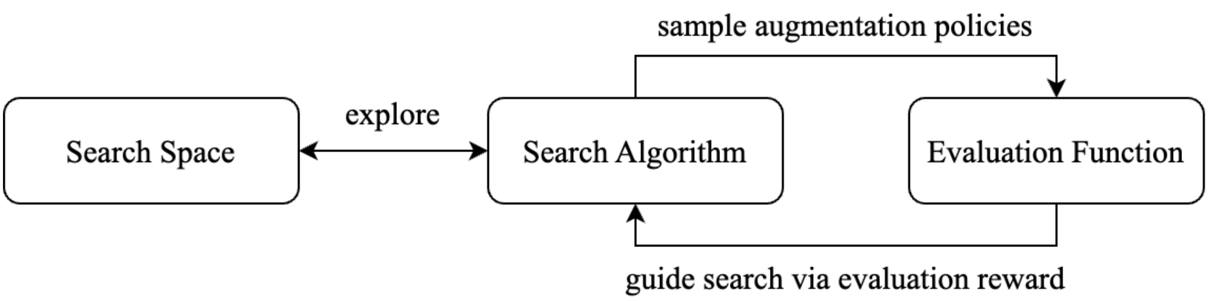


Figure 5: General workflow of AutoDA system (self-composed)

These 3 components work together to explore the search space and find the optimal DA schemes for the given dataset. In addition, the workflow of a standard AutoDA system consists of 2 main stages (Yang et al., 2022a):

1. **Generation stage:** is where the AutoDA model search for optimal DA schemes for the given dataset.
2. **Application stage:** is where optimal DA schemes are applied to the input dataset and generate new data samples.

Based on the aforementioned workflow stages, AutoDA models can divide into 2 major high-level categories (Yang et al., 2022a):

1. **Two-stage AutoDA model:** architecture consists of separate generation and application stages. This architecture is more effective than the one-stage AutoDA applications.

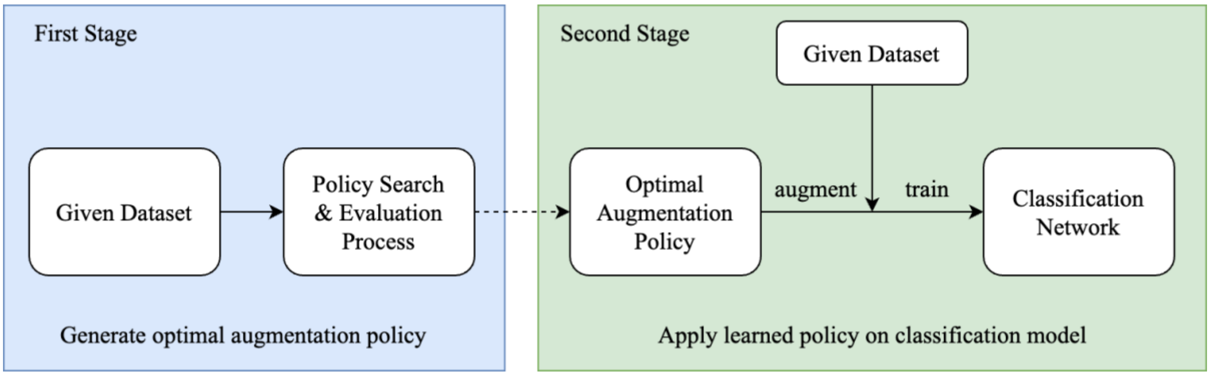


Figure 6: General architecture of two-stage AutoDA model (self-composed)

2. **One-stage AutoDA model:** architecture combines the generation and application stages. This architecture is computationally more efficient than the two-stage AutoDA applications. However, the implementation complexity of this architecture is high.

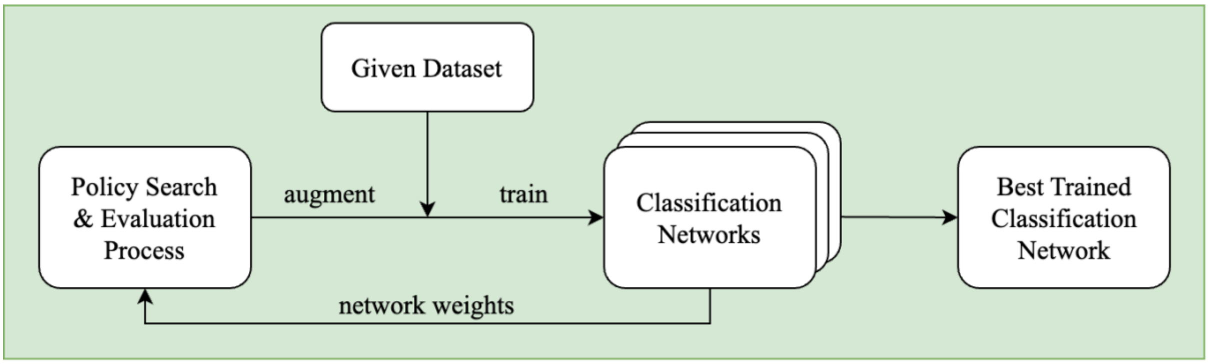


Figure 7: General architecture of one-stage AutoDA model (self-composed)

As mentioned earlier, the objective of AutoDA models is to discover optimal DA policies with the minimal search effort. Existing approaches have proposed various search strategies to enhance search efficiency. Therefore, based on the search strategy, existing AutoDA models can be further categorized as gradient-based, gradient-free, and search-free (Yang et al., 2022a). A detailed explanation for these search strategies is placed in the 2.5 section.

2.4 Existing Systems

2.4.1 General Search Space Definition of AutoDA Systems

As previously stated, existing AutoDA systems focused on automatically searching for optimal DA schemes using basic/traditional DA techniques (Cubuk et al., 2019a). The DA techniques that are mostly utilized in AutoDA systems include:

- Identity
- Auto Contrast
- Contrast
- Rotate
- Equalize
- Shear-X
- Shear-Y
- Translate-X
- Translate-Y
- Solarize
- Brightness
- Sharpness
- Color
- Posterize

The general search space of the AutoDA system consists of a pre-defined set of DA policies, which are a combination of image transformation techniques. Each DA policy consists of multiple sub-policies, each sub-policy consists of multiple specific DA transformations, and each transformation is associated with two hyper-parameters: probability and magnitude. The probability hyper-parameter determines the probability that the transformation will be applied to the image, while the magnitude hyper-parameter controls the magnitude of the transformation (Cubuk et al., 2019a; Yang et al., 2022a).

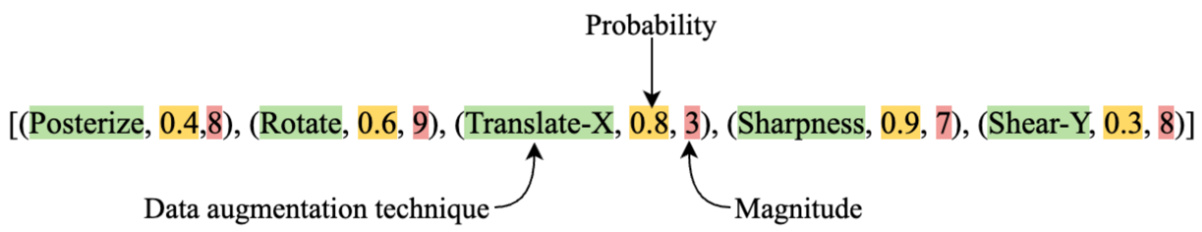


Figure 8: Example DA policy that consists of five sub-policies and each sub-policy consists of one transformation technique.

As aforementioned, the AutoDA models aim to search for the optimal hyperparameters for each DA policy.

2.4.2 Analysis on Existing AutoDA Systems

AutoAugment (Cubuk et al., 2019a) provided a standard problem formulation in the AutoDA domain. However, the key drawback of AA is the high computational resource usage for the search process. As a result, many follow-ups work focused to enhance the efficiency of the optimal DA scheme search process (Yang et al., 2022a). This literature review highlights the major turning points in the AutoDA domain that addressed the efficiency issues of AA.

2.4.2.1 Two-Stage Approaches

AutoAugment (AA) (Cubuk et al., 2019a)

As mentioned above, the key contribution of AA was a standard problem formulation in the domain (Cubuk et al., 2019a), which has been widely adopted by subsequent works.

Search space configuration: Each DA policy in the AA search space consists of five sub-policies and each sub-policy consist of two basic/traditional image transform techniques.

Policy search & evaluation process: AA employed Recurrent Neural Network (RNN) as the search algorithm to discover the optimal DA schemes from the search space. Instead of directly evaluating the selected DA schemes on the target model, AA used a simplified child model for evaluation. This child model was trained with augmented data, and once the training process was complete, the validation accuracy was provided as feedback to the search algorithm (also known as a proxy task). This feedback loop enables the RNN to generate more effective data augmentation schemes for further evaluation (Cubuk et al., 2019a; Yang et al., 2022b).

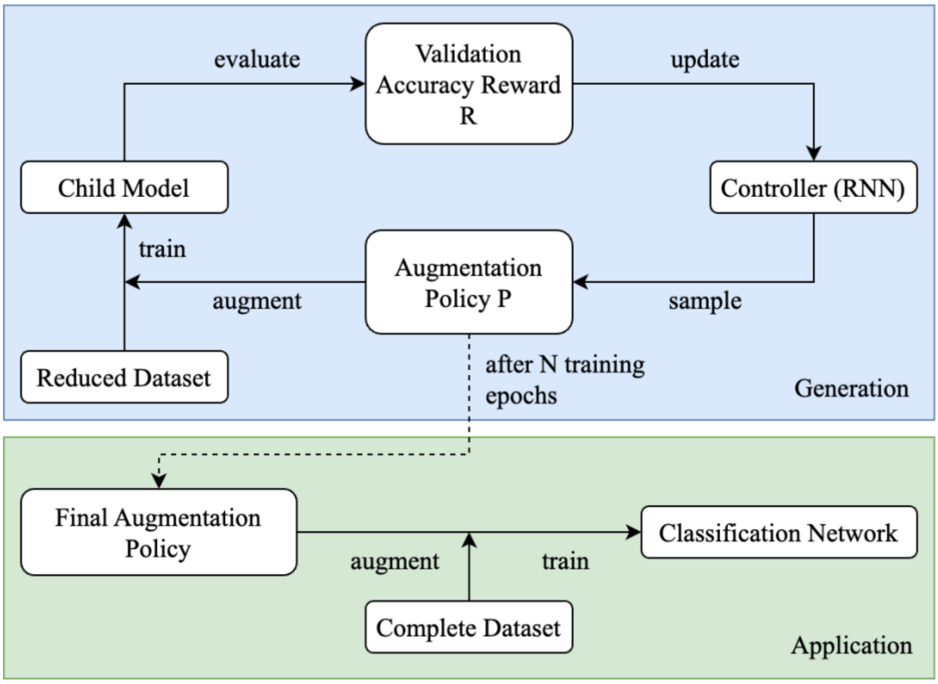


Figure 9: Workflow of AA (self-composed)

Limitations: Even though AA archived state-of-art performance on benchmarking datasets like CIFAR10/100 and ImageNet, AA has serious limitations. One of the key drawbacks is AA requires a large number of computational resources and requires a lot of time to train the policy search model (Cubuk et al., 2019a). Another key limitation of AA is that it relies on the quality of the proxy task, which might not always reflect the true performance of the target model (Cubuk et al., 2019b).

Population-Based Augment (PBA) (Ho et al., 2019)

PBA is one of the primary works that use the Population Based Training (PBT) technique to improve the performance of AA. The search space configuration of PBA is very similar to the AA. To enhance the efficiency, instead of a single child model, PBA trains a group of child models (population) simultaneously with a unique DA policy. After a specific epoch count, PBA evaluates the performance of these child models and ranks them accordingly. The lower-ranked models will then replicate the hyper-parameter configurations of the highest-ranked models and fine-tune them. As a result, PBA eliminates the need to re-initialize child models from scratch, resulting in improved efficiency compared to AA. Despite these benefits, PBA still requires a significant amount of search time, which affects its overall efficiency (Ho et al., 2019; Yang et al., 2022a).

Fast AutoAugment (Fast AA) (Lim et al., 2019)

Fast AA takes a different approach to optimize hyper-parameters of DA policies than the aforementioned AutoDA solutions by focusing on the distribution difference between the original and augmented data. The objective of Fast AA on minimizing the distribution difference between original and augmented data using a density matching algorithm. As a result, Fast AA eliminates the need for child model training, thereby significantly reducing the computational resources required for the optimization process. Although Fast AA has reduced search time and computational resources requirement significantly, further reduction of search time is required to make AutoDA techniques accessible to ordinary users (Lim et al., 2019; Yang et al., 2022a).

RandAugment (RA) (Cubuk et al., 2019b)

Currently, RA is considered the most practical AutoDA solution due to its implementation simplicity and lower computational resource usage. It is the first effort to use a search-free approach for finding optimal DA policies for a given dataset. RA achieves this by reparametrizing the entire traditional AutoDA search space into manually customizable hyper-parameters, including the number of transformations to apply for an image and the static magnitude for all transformations. Furthermore, RA uses a simple grid search to find optimal DA policies from the search space, which drastically reduces computational resource requirements by eliminating the searching process. However, the static magnitude of transformations used in RA may limit its ability to explore high performing DA policies, which remains a bottleneck in its performance (Cubuk et al., 2019b; Yang et al., 2022a).

```
transforms = [
'Identity', 'AutoContrast', 'Equalize',
'Rotate', 'Solarize', 'Color', 'Posterize',
'Contrast', 'Brightness', 'Sharpness',
'ShearX', 'ShearY', 'TranslateX', 'TranslateY']

def randaugment(N, M):
"""Generate a set of distortions.

  Args:
    N: Number of augmentation transformations to
        apply sequentially.
    M: Magnitude for all the transformations.
"""

  sampled_ops = np.random.choice(transforms, N)
  return [(op, M) for op in sampled_ops]
```

Figure 10: Official PyTorch implementation of RA (Cubuk et al., 2019b)

2.4.2.2 One-Stage Approaches

Online Hyper-parameter Learning AutoAugment (OHL-AA) (Lin et al., 2019)

OHL-AA is the preliminary one-stage AutoDA architecture that optimizes the DA policy as well as target CNN model weights simultaneously, thereby reducing search time and cost compared to traditional two-stage methods.

Search space configuration: The search space settings of the OHL-AA similar as the standard search space settings proposed by AA.

Policy search & evaluation process: The OHL-AA framework utilizes a bi-level optimization strategy, whereby the DA policy is selected from a parameterized probability distribution. This distribution is optimized using the REINFORCE estimator in conjunction with the target CNN weights. The model updates the policy hyper-parameters in a forward manner, and the optimization problem involves two interconnected loops. The inner loop trains multiple models concurrently, utilizing different DA policies. The outer loop then updates the policy distribution based on the validation accuracies obtained. Despite maintaining competitive accuracy with the target CNN, the OHL-AA architecture achieves significantly faster search times, with a 60-fold reduction on the CIFAR-10 dataset and a 24-fold reduction on the ImageNet dataset. The gradient based methodology proposed in OHL-AA provides an effective method to solve efficiency issues in AutoDA problems, thereby promoting the development of subsequent one-stage approaches. Below figure illustrates the general workflow of one-stage AutoDA model architecture proposed by OHL-AA (Lin et al., 2019; Yang et al., 2022b).

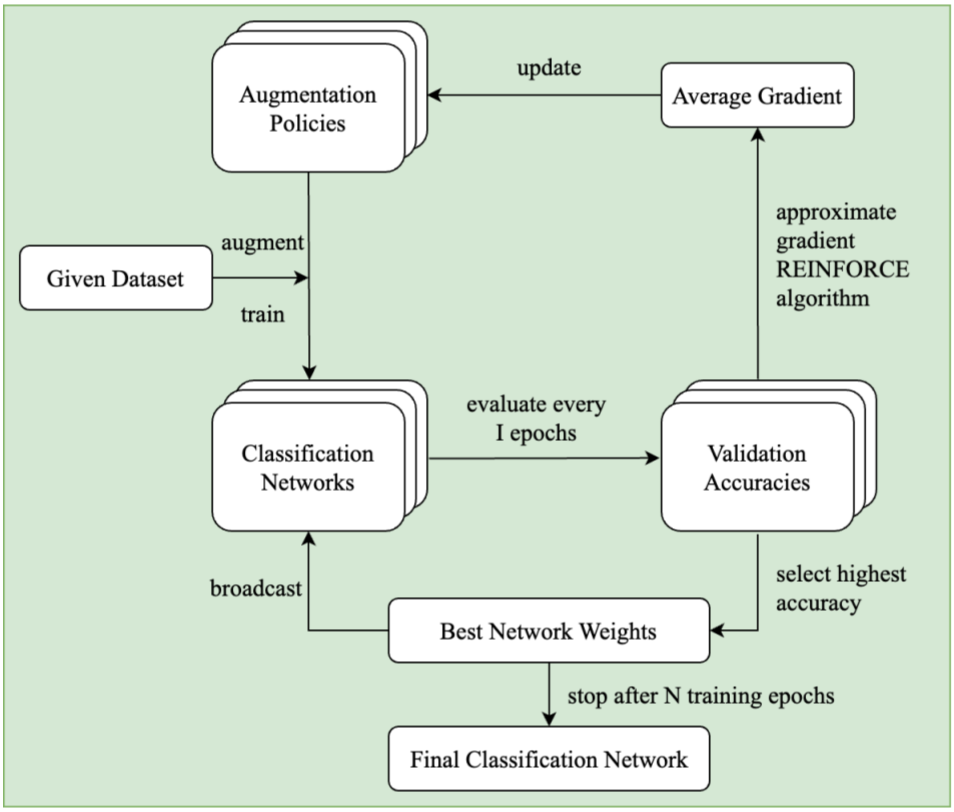


Figure 11: Workflow of OHL-AA (self-composed)

Limitations: Although the proposed one-stage architecture was able to achieve improvements in optimal DA policy search time, it requires high-end GPU resources to perform the AutoDA task as it depends on concurrent process, which limits the wide applicability of the proposed method among ordinary users (Lin et al., 2019; Yang et al., 2022a).

Adversarial AutoAugment (Adv AA) (Zhang et al., 2019)

To improve the efficiency of existing AutoDA techniques, Adv AA was developed as an one-stage version of the original AA. It maintains a similar search space configuration as AA but incorporates Generative Adversarial Network (GAN) concepts into the optimal DA policy search process. Adv AA consists of two networks: Generator and Discriminator. The Generator generates DA policies, while the Discriminator acts as the target classification model. Based on the performance of the Discriminator network, the Generator adjusts the hyper-parameters of the generated DA policies. Although Adv AA reduces the optimal policy search time and computational cost compared to AA, it still requires a significant amount of time to perform AutoDA tasks, and the proposed GAN architecture also requires high computational resources, limiting its applicability for ordinary users. Therefore, further optimization is necessary to make Adv AA more accessible and user-friendly (Yang et al., 2022a; Zhang et al., 2019).

Differentiable Automatic Data Augmentation (DADA) (Li et al., 2020)

DADA is another very efficient one-stage AutoDA solution motivated by the advancements of differentiable NAS technologies. DADA utilized Gumbel-SoftMax Relaxation method for the optimal DA policy selection process and unbiased gradient estimator named RELAX for the hyper-parameter optimization process. The usage of RELAX enables an effective and efficient DA policy optimization process in DADA, resulting in the discovery of accurate DA policies for a given classification model in a shorter amount of time. The primary constraints of DADA can be summarized as the elevated complexity of implementation and substantial computational resources necessary for the search process, which limits the widespread use of DADA among ordinary users (Li et al., 2020; Yang et al., 2022a).

2.4.3 Taxonomy of Existing AutoDA Systems

Existing Work	**Taxonomy**				
	Architecture		**Hyperparameter Optimization Approach**		
	One Stage	**Two Stage**	**Gradient Free**	**Gradient Based**	**Search Free**
(Cubuk et al., 2019a)		X	X		
(Lim et al., 2019)		X	X		
(Ho et al., 2019)		X	X		
(Wei et al., 2020)		X	X		
(Lin et al., 2019)	X			X	
(Zhang et al., 2019)	X			X	
(Li et al., 2020)	X			X	
(Cubuk et al., 2019b)		X			X
(LingChen et al., 2020)		X			X
Proposed	X			X	X (Partially)

Table 3: Taxonomy of existing AutoDA works

2.4.4 Summary of Existing AutoDA Systems

To summarize, standard problem formulation for the AutoDA domain is proposed by AA (Cubuk et al., 2019a), which has key drawbacks such as high computational resource usage . Subsequent research in the AutoDA domain has made significant contributions to reduce computational resource usage by discovering optimal DA policies in less time. Though these subsequent works have reduced the computational resource usage time, they still require powerful GPU computational resources, limiting their applicability to ordinary users (Yang et al., 2022a). Furthermore, all these methods depend on a predefined search space, LR has demonstrated that the predefined search space might be effective on benchmarked datasets such as CIFAR10, CIFAR100, and ImageNet, but it might be negatively impacted on real-world datasets due to unique characteristics (Cubuk et al., 2019b; Yang et al., 2022a). The tables below describe the advantages and disadvantages of existing AutoDA architectures and systems.

AutoDA System Architecture	**Advantages**	**Disadvantages**
Two-Stage	• Better performance in the target model. • Easy to implement.	• Higher computational cost.
One-Stage	• More efficient than the two-stage architecture.	• Requires high computational resources. • Increased complexity.

Table 4: Comparison of AutoDA system architectures

AutoDA System	**Contribution(s)**	**Limitation(s)**
(Cubuk et al., 2019a)	• First Implementation • Provides fundamental theories to implement standard AutoDA model.	• Higher computational cost due to longer search time.
(Ho et al., 2019)	• Reduced the optimal DA policy search time of (Cubuk et al., 2019a).	• Higher computational cost due to longer search time.
(Lim et al., 2019)	• Reduced the optimal DA policy search time of (Cubuk et al., 2019a).	• Higher computational cost due to longer search time.
(Cubuk et al., 2019b)	• Provides fundamental theories to implement search-free AutoDA architecture.	• Performance improvement of target CNN is limited.

	• Reduced the optimal DA policy search time of (Cubuk et al., 2019a).	
(Lin et al., 2019)	• Provides fundamental theories to implement gradient-based one-stage AutoDA architecture. • Reduced the optimal DA policy search time of (Cubuk et al., 2019a)	• Higher computational cost due to high-end GPU resource requirement. • Complex implementation limiting wider usability.
(Zhang et al., 2019)	• Further improves of one-stage AutoDA model architecture by introducing GAN concepts. • Reduced the optimal DA policy search time of (Cubuk et al., 2019a).	• Higher computational cost due to high-end GPU resource requirement. • Complex implementation limiting wider usability.
(Li et al., 2020)	• Further improve gradient-based one-stage AutoDA model efficiency. • Reduced the optimal DA policy search time of (Cubuk et al., 2019a).	• Higher computational cost due to high-end GPU resource requirement. • Complex implementation limiting wider usability.

Table 5: Comparison of existing AutoDA works

2.5 Technology Review

The main objective of AutoDA systems is to efficiently search through a given search space to identify the optimal data augmentation schemes (Cubuk et al., 2019a). To achieve this objective, researchers have developed various hyper-parameter optimization approaches that utilize efficient search algorithms and evaluation methodologies. Based on the perspective of hyper-parameter optimization in the search space, these approaches can be broadly categorized into three main groups: gradient-free, gradient-based, and search-free techniques (Yang et al., 2022a).

2.5.1 Gradient-Free Technologies

Gradient-free technologies are used to find the optimal hyper-parameters without calculating gradients. This can be archived by applying random values for each hyper-parameter, training the model with each set of values, and lastly update the values of hyper-parameters based on the model performance. Gradient-free methodologies used in the AutoDA domain include:

Reinforcement Learning (RL)

RL is a type of ML technique in which an agent learns to interact with the given environment by performing actions and gathering feedback in the way of rewards or penalties (Cubuk et al., 2019a). The aim of the agent is to learn the optimal configuration setup that maximizes the reward value. One of the main advantages of RL is its ability to learn from experience without explicit supervision. However, RL can be computationally intensive and require a large volume of data to achieve good performance.

Population Based Training (PBT)

PBT is a distributed ML technique that trains multiple models simultaneously and periodically ranks the best-performing models. The objective of PBA is to create a population of models, each with unique hyperparameters. The best-performing models are identified based on predefined metrics, and their hyperparameter configurations are recombined with other models in the population to create new configurations (Ho et al., 2019). This process continues until a desired performance is achieved. One of the key advantages of PBT includes efficient hyperparameter optimization. However, PBT requires significant computational resources and can be difficult to implement and tune effectively.

Greedy Breadth-First Search (GBFS)

GBFS is a graph traversal algorithm that follows a specific rule for selecting the next node to visit. This specific rule is a condition of estimating the distance or costs to the goal node and GBFS always chooses the node with the lowest estimated cost as the next node to visit. The main advantage of GBFS is its ability to find solutions quickly (Naghizadeh et al., 2020). However, GBFS can get stuck in local minima or search in the wrong direction if the node traversal rule is not properly designed.

Grid Search (GS)

GS is a simple way of searching the optimal hyperparameters. It works by initializing the grid of all possible hyperparameter values, and then it will train and evaluate each combination in the grid. Finally, the optimal hyper-parameter configuration is selected based on some predefined evaluation metric. For example, in the AutoDA domain, this predefined evaluation metric will be model accuracy (Cubuk et al., 2019b). GS can be computationally expensive for large search spaces.

Generative Adversarial Networks (GAN)

A standard GAN architecture consists generator and discriminator neural networks (Goodfellow et al., 2014). In the image domain, these two neural networks are trained in a competitive manner. The aim of the generator is to produce realistic data that can deceive the discriminator, and the aim of the discriminator is to correctly classify the data as real or generated. The objective of GAN models is to train a generator that can generate realistic data (Goodfellow et al., 2014). Though GANs are powerful models for generating realistic and high-quality images, the GAN training process can be unstable and require significant computational resources.

Density Matching (DM)

DM is a technique used to generate augmented data similar to the original data. The technique operates on the principle that the distribution of the augmented data should match the distribution of the original data, and it works by adjusting the density of the augmented data to match that of the original data (Hataya et al., 2020; Lim et al., 2019). The key disadvantages of DM are, DM can lead to overfitting if the generated data is too similar to the original data, and it may not be effective in capturing rare characteristics of the original data.

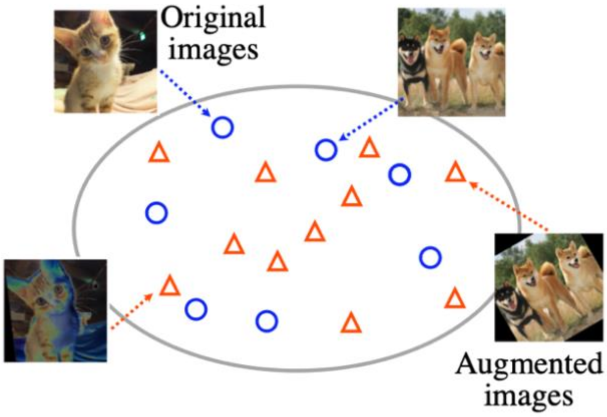


Figure 12: Example distribution of dataset with original & augmented data (Hataya et al., 2020)

2.5.2 Gradient-Based Technologies

Gradient-based hyper-parameter optimization is a method that leverages the gradient of the loss value of the target model with respect to the hyper-parameters to optimize the model. This involves computing the gradients of the loss value of the target model with respect to the hyper-parameters and using these gradients to adjust the hyper-parameters in a way that minimizes the loss value of the target model. By minimizing the loss value through gradient-based hyper-parameter optimization, the target model can achieve better performance and accuracy (Lin et al., 2019; Yang et al., 2022a). Gradient-based methodologies used in the AutoDA domain include:

Stochastic Gradient Descent (SGD)

SGD is a popular optimization algorithm used in ML to train models with large datasets. The technique operates by randomly selecting a mini-batch of data and updating the model hyper-parameters based on the gradients calculated from that mini-batch. One key advantage of SGD is its ability to handle large datasets efficiently (Lin et al., 2019; Summa et al., 2011). However, SGD can be sensitive to the choice of the learning rate, because too high a learning rate can lead to unstable convergence or overshooting the optimal solution.

Relaxed Bernoulli Distribution (RBD)

The RBD is a statistical tool used in ML to represent uncertain binary data. It is similar to the traditional Bernoulli distribution but instead of producing a binary outcome, it generates a real number between 0 and 1. This number is then compared to a threshold, which is a learned hyperparameter, to determine if the output is 1 or 0. This allows the model to control the amount of uncertainty in the result (Hataya et al., 2020; Wang and Yin, n.d.). The RBD is commonly used in VAE and RL to model random decisions or actions. Although the RBD is conceptually simple, it can add computational complexity to ML models.

Reinforce Gradient Estimator (RGE)

The RGE is a method used in RL to find the best policy by optimizing for long-term rewards. It works by taking actions in an environment and receiving rewards for those actions. The estimator then updates the policy function to increase the probability of selecting actions that lead to higher rewards. This process is done by computing the gradient of the policy function and weighting it by the ratio of the probabilities of actions under the current policy and a target policy (Lin et al., 2019). One advantage of the RGE is its ability to handle continuous and discrete action spaces. However, it can suffer from high variance and instability.

Differentiable Architecture Search (DARTS) Estimator

DARTS is a method used in NAS to find the optimal neural network architecture by optimizing a continuous relaxation of the search space. The search process is performed through gradient descent using a differentiable proxy for the validation accuracy as the objective function. The DARTS Estimator has several advantages, including its efficiency in exploring a large search space and its ability to handle complex architectures (Li et al., 2020; Liu et al., 2018). On the other hand, limitations of the DARTS Estimator include sensitivity to the initialization of the search variables and the high computational cost of evaluating the gradients of the objective function.

2.5.3 Search-Free Technologies

The majority of existing AutoDA works consider AutoDA as a hyper-parameter optimization problem (Cubuk et al., 2019b). However, search-free methods reparametrize the search space, allowing for manual customization of hyperparameters and eliminating the need for hyperparameter optimization. These manually customizable hyperparameters include:

- **N:** Number of image transformations operations to apply for each image
- **M:** Static magnitude for all image transformations operations
- **K:** Static probability for all image transformations operations to pick from search space

If N=2, M=10, and K=1 then the example DA policy will be,

[(Rotate, 1, 10), (Sharpness, 1, 10)]

Figure 13: Example DA policy of search-free AutoDA models

As a result, search-free AutoDA technologies offer a promising solution for reducing the computational resource required for AutoDA tasks, while archiving competitive performance (Cubuk et al., 2019b). However, since these technologies rely on image transformation operations that have static magnitude, they may not explore high-optimal DA policies that could further improve the performance of the target model. Hence, the performance of search-free technologies remains in the bottleneck (Yang et al., 2022a).

2.5.4 Analysis on Technologies

As aforementioned, hyper-parameter optimization technologies of the AutoDA domain can be categorized as gradient-free, gradient-based, and search-free techniques.

Technology Category	Advantages	Disadvantages
Gradient-free	• Simple to implement. • Can escape local optima.	• Requires careful tuning of hyper-parameters. • Highly inefficient for high-dimensional search spaces. • The convergence rate is usually slower than gradient-based algorithms.
Gradient-based	• Fast convergence rate. • Efficient for high-dimensional search spaces.	• Can get stuck in local optima. • Computationally expensive for large search spaces.
Search-free	• Simple to implement. • Faster than both gradient-free and gradient-based approaches.	• May not explore all possible optimal hyper-parameter values. • The performance of the target model might be limited.

Table 6: Comparison of hyper-parameter optimization technologies

2.6 Evaluation & Benchmarking

2.6.1 Evaluation Matrices

The primary objective of AutoDA models is to improve the **accuracy** of the classification model provided. Therefore, accuracy is the key evaluation metric for AutoDA models (Cubuk et al., 2019a; Yang et al., 2022a). The accuracy of an ML model is determined by calculating the proportion of correct predictions made by the model out of the total number of predictions. ML models with higher accuracies generally consider as a good models.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made}$$

It is crucial to incorporate other evaluation metrics to obtain a more comprehensive assessment of performance of the classification model. Precision, recall, F1-score, ROC-AUC, and confusion matrix are some examples of alternative evaluation metrics that can be utilized to evaluate the performance of classification model.

- **Confusion matrix:** is a summary table that displays the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) generated by a classification model.
- **Precision:** is a metric used to evaluate the proportion of positive predictions made by the model that are genuinely correct.

$$Precision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

- **Recall:** is a metric that gauges the proportion of all actual positive instances that the model correctly identified as positive.

$$Recall = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

- **F1-Score:** is a metric that assesses a performance of model by combining precision and recall into a unified score. It is computed as the harmonic mean of precision and recall.

$$F1 - score = \frac{2 * (Precision * Recall)}{(Precision + Recall)}$$

- **AUC-ROC curve**: AUC-ROC is a graphical representation of the performance of a classification model at various threshold settings.

2.6.2 Benchmarking

Benchmarking of AutoDA systems typically involves evaluating the performance of standard image classification models on standard benchmarking datasets. In the AutoDA domain, variants of Wide-ResNet, variants of ResNet, Shake-Shake, and PyramidNet are commonly used as standard image classification models for benchmarking, while benchmarking datasets may include CIFAR10, CIFAR100, SVHN, and ImageNet (Cubuk et al., 2019a; Yang et al., 2022a). The below figure described the characteristics of each dataset.

Dataset	Classes	Images per class	Image size	Year
CIFAR-10	10	6,000	32×32	2009
CIFAR-100	100	600	32×32	2009
SVHN	10	630,420	32×32	2011
ImageNet	varying[1]	varying[1]	varying[1]	2009

Figure 14: Characteristics common benchmarking datasets (Yang et al., 2022a)

The table below displays the standard classification tasks proposed by Cubuk et al. 2019 for benchmarking purposes.

Dataset	**Utilized Image Classification Model(s)**
CIFAR10	• Wide-ResNet-28-10 • Shake-Shake (26 2x32d) • Shake-Shake (26 2x96d) • Shake-Shake (26 2x112d) • AmoebaNet-B (6,128) • PyramidNet+ShakeDrop
CIFAR100	• Wide-ResNet-28-10 • Shake-Shake (26 2x96d) • PyramidNet+ShakeDrop
SVHN	• Wide-ResNet-28-10 • Shake-Shake (26 2x96d)
ImageNet	• Wide-ResNet-28-10 • Shake-Shake (26 2x96d)

Table 7: Classification tasks used for benchmarking (Cubuk et al., 2019a)

2.7 Chapter Summary

The primary objective of this chapter is to present a solution to the AutoDA problem domain by identifying the most convenient methodologies and techniques. To systematically collect essential data, a concept map was initially created. The review was classified into four major categories: problem domain, prior research in the domain, technologies and possible evaluation and benchmarking criteria. The critical analysis of each category was conducted with an emphasis on identifying limitations and potential avenues for future improvement.

CHAPTER 3: METHODOLOGY

3.1 Chapter Overview

The primary objective of this chapter is to examine various methodologies for research, project management, and development. Furthermore, it includes the presentation of a project plan and the identification of potential risks, along with their corresponding mitigations.

3.2 Research Methodology

The quality of a research project depends on three crucial factors: scope, cost, and time, all of which must be effectively managed throughout the project. As a result, specific methodologies are necessary, and for this project, the research methodologies were selected from the Saunders Research Onion Model.

Research Philosophy	The **pragmatism** approach was selected from the available options of pragmatism, positivism, interpretivism, and realism as it allows the author to assess and experiment with a range of methods in combination to determine the best approach for achieving the research objective.
Research Approach	From the available research approaches of inductive and deductive, the **deductive** approach was selected for the research project due to the objective of extensively applying existing theories.
Research Strategies	The research strategy pertains to the methodology used to address the research questions. Various methods such as **experiments** and **interviews**, which are assessed through evaluation metrics, will be considered for selection.
Research Choice	Out of the available research options, namely mono, multi methods and mixed, the **mixed** method was selected for this research project. The reason for this selection was to complement qualitative strategies such as interviews with quantitative strategies like surveys, thereby enhancing the overall research outcome.
Time Horizon	Data collection for this research project will be carried out at a single point in time, during the evaluation phase. As a result, a **cross-sectional** method was chosen from the available options of longitudinal and cross-sectional methods.

Techniques and Procedures	To gather and analyze data, this research will employ **interviews**, trial and error **prototyping**, **literature review**, as well as analysis of similar solutions.

Table 8: Research Methodology

3.3 Development Methodology

Among the many available software lifecycle methodologies, the **prototype** method was chosen because the project will be designed, built, and evaluated until a successful outcome is achieved.

3.4 Project Management Methodology

From the available project management methodologies, the **Agile Prince-2** hybrid model was selected for this research project. This is because the author, who is the sole developer, expects multiple interim deliverables, making the Agile Prince-2 method an ideal choice.

3.4.1 Project Plan

Please refer **Appendix C** for Gantt Chart.

3.4.2 Deliverables

Deliverable	Date
Project Proposal Document - The preliminary project proposal for the research study.	3rd November 2022
Literature Review Document - A thorough examination and evaluation of the current AutoDA works and solutions.	11th December 2022
Software Requirement Specification - The requirements that must be fulfilled in the final research prototype.	15th December 2022
System Design Document - A document outlining the architecture of the proposed AutoDA framework.	1st December 2022
Prototype - The working software solution of the research project, comprising the proposed AutoDA framework.	1st February 2023
Thesis - The final research project documentation discusses the research problem, solution, and findings.	2nd May 2023
Research Paper - A research paper that presents the AutoDA framework developed upon the completion of this research project.	15th May 2023

Table 9: Deliverables and Dates

3.5 Resources

Based on the functionalities and objectives of the project, the identified necessary hardware, software, data, and skills to complete this project are as follows.

3.5.1 Hardware Resources

- **Apple M1 Pro (8-core CPU, 14-core GPU, 16-core Neural Engine) processor or above -** To run DA policy search and evaluation engine and perform data augmentation on various CNNs architectures and datasets.
- **16GB RAM or above -** To manage large volumes of data and complex CNN architectures.
- **Disk space of 50GB or more -** To store necessary data and application code.

Note: If hardware resources are not sufficient, and since Apple Silicon processors are not compatible with some of the ML/DL frameworks yet, propose using Google Colab free tire version.

3.5.2 Software Resources

Requirement Description	Primary Software	Secondary Software
To run essential programs such as IDEs and other required tools.	MacOS	Windows
The primary programing language used to develop backend of the proposed system.	Python	N/A
The primary programing language used to develop frontend of the proposed system.	React JS	N/A
The primary software development IDE for speed up and improve the effectiveness of the development process.	Visual Studio Code	Backend: PyCharm Frontend: WebStorm
To version control and backup the application code.	GitHub	Google Drive
To create the documents which are related to the project.	MS Office	Google Docs
To organize, manage, and back up the relevant research papers.	Zotero	Mendeley

Table 10: Software Requirements

3.5.3 Data Requirements

This project is regarding improving diversity of poor image datasets. Therefore, any publicly available image dataset is ideal. However, to do the fair comparison against to similar work, the author will use standard classification datasets, which are:

- Canadian Institute for Advanced Research, 10 classes (CIFAR10)
- Street View House Numbers (SVHN)
- Modified National Institute of Standards and Technology (MNIST)

3.5.4 Technical Skills

The technical skills required for successful project completion are:

- Knowledge of CNN architecture defining training and tuning process.
- Knowledge of image DA techniques.
- Knowledge of efficient search and evaluation techniques.
- UI/UX skills.

3.6 Risk & Mitigation

The following are the risks that were identified before commencing the project, along with potential mitigation measures.

Risk Item	Severity	Frequency	Mitigation Plan
Lack of domain knowledge	5	3	Consult with academic and domain experts, utilize resources such as Stack Overflow, and refer to online tutorials and courses.
Insufficient hardware resources	5	4	Utilize cloud-based solutions like the free tier version of Google Colab.
Misplacing or losing the code under development.	5	2	Use GitHub and external backup of all code.
Misplacing or losing the necessary documents.	5	2	Follow a cloud-first documentation approach.

Any unpredictable risk (such as Covid 19, power interruptions, and natural disasters)	3	5	Maintain a schedule to ensure completion of daily or weekly objectives.

Table 11: Associated Risks And Mitigation

3.7 Chapter Summary

In this chapter, the selected research methodologies based on the Saunders Research Onion Model, required resources to develop the project, deliverables, and potential risks have been thoroughly evaluated with corresponding mitigations provided.

CHAPTER 4: SOFTWARE REQUIREMENT SPECIFICATION

4.1 Chapter Overview

This chapter centers around the identification of potential requirements and stakeholders of the system. Firstly, a rich picture diagram and stakeholder onion model were utilized to define the stakeholders and their interactions of the proposed system. Next, the selected requirement elicitation methods and their corresponding findings were discussed. Use case diagrams and descriptions were then provided. Lastly, the chapter concludes with a discussion of the functional and non-functional requirements of the proposed system.

4.2 Rich Picture Diagram

The rich picture is a way of representing the structure, process, and concerns of a system from a bird's-eye view. The below rich picture diagram visualizes the identified structure, process, and concerns of the proposed AutoDA system.

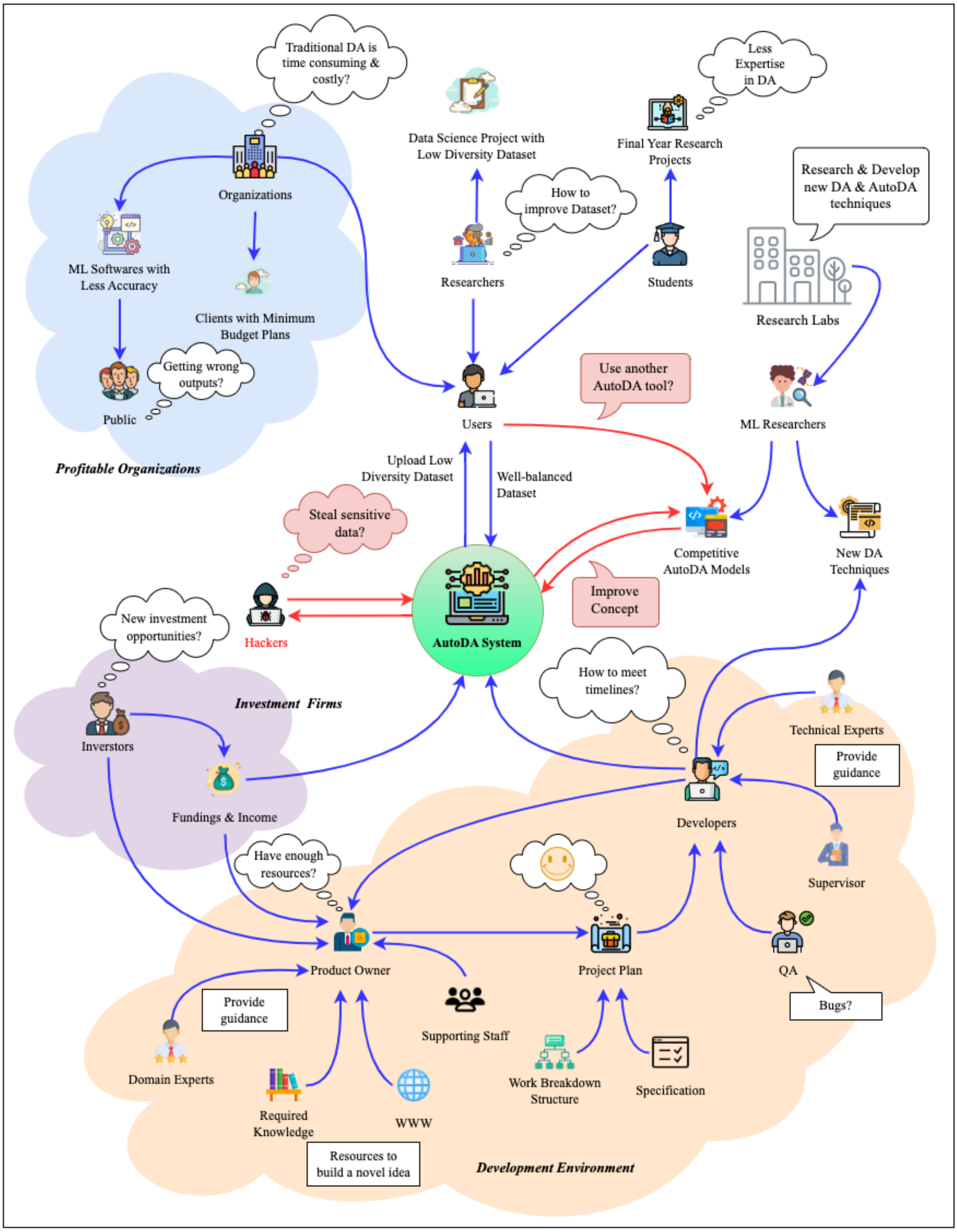


Figure 15: Rich Picture Diagram (self-composed)

4.3 Stakeholder Analysis

The below stakeholder onion model visualizes the identified stakeholders of the proposed AutoDA system. After that, a detailed elaboration of stakeholders of the proposed AutoDA system is also presented.

4.3.1 Stakeholder Onion Model Diagram

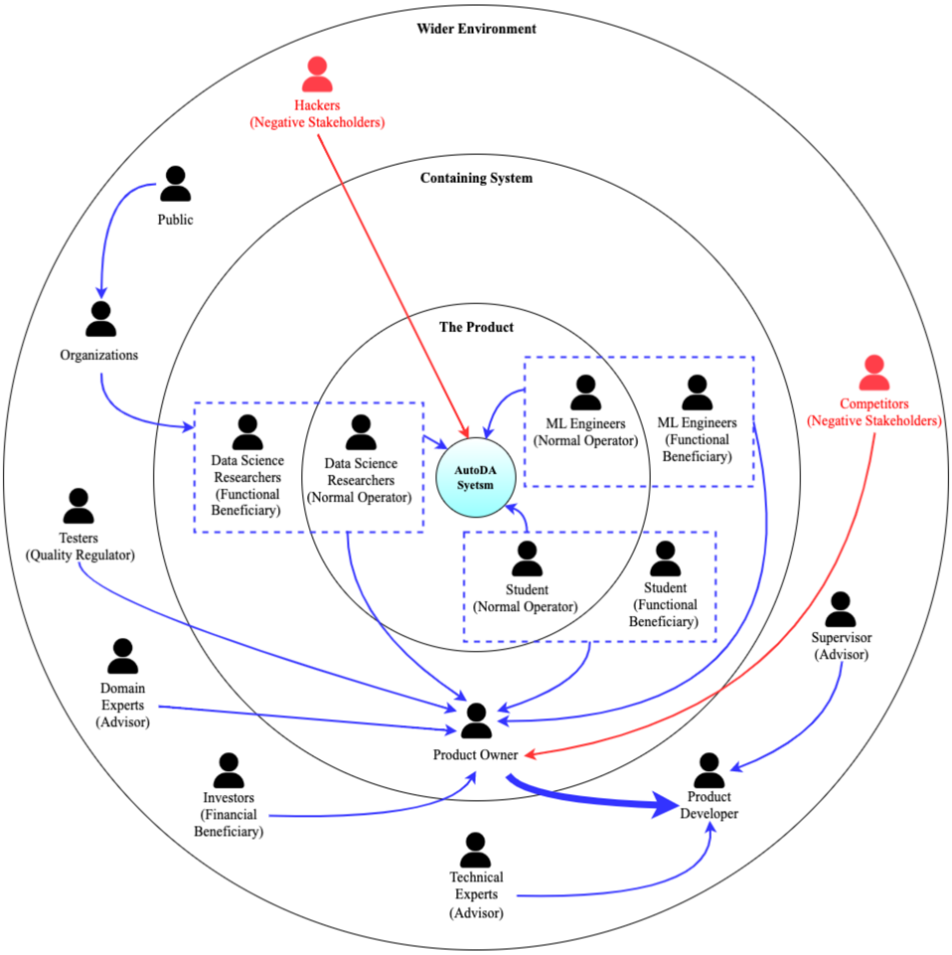


Figure 16: Stakeholder Onion Model (self-composed)

4.3.2 Analysis of the Stakeholders

Stakeholder	Role	Description
The Project Stakeholders		
ML Engineers	Normal operator	Perform AutoDA using the system for their own datasets.
Students		
Data science researchers		
Containing System Stakeholders		
Product owner	Operational beneficiary	Manage and supervise the entire business workflow and product engineers.

Students	Functional beneficiary	Further investigate the research theory presented in the proposed AutoDA system and improve it. Learn about automated data augmentation techniques.
ML Engineers	Functional beneficiary	Get the research hypothesis presented in the proposed system and improve it further. Develop new data augmentation techniques.
Organizations	Functional beneficiary	Use the system to improve their datasets and products.
Wider Environment Stakeholders		
Application developer	Development staff and operational maintainer	Develop and improve the proposed system and fix errors when occurred.
Supervisor	Quality regulator / Advisor	Evaluate the product and provide feedback to improve.
Domain Experts		
Technical Experts		
Testers		
Investors	Financial beneficiary	Invest in the proposed system and gain profit from the audience.
Hackers	Negative stakeholder	Attempt to bring down the system and disable system services.
Competitive AutoDA models	Negative stakeholder	Develop an improved version of AutoDA systems.

Table 12: Analysis of the Stakeholders

4.4 Selection of Requirement Elicitation Methodologies

Requirement elicitation is a way to find out the requirements of a SE project. To gather software requirements for this project, among the available requirement elicitation methodologies, LR, interviews, and prototyping methods were chosen. The justifications for selecting the above requirement elicitation methodologies are outlined below.

Method 1: Literature review
A good LR helps to figure out the research gaps and problems in the existing systems. Hence, at the initial stage of this project, the author did a thorough analysis of the existing AutoDA literature domain, existing AutoDA systems, possible novel approaches, and technologies to solve the current limitation of AutoDA systems.
Method 2: Interviews
Domain and technical experts' insights are very important to validate the identified research idea and research gap. Moreover, with the help of their knowledge, it is easy to identify the best possible way and technologies to solve the identified problem. Additionally, since the AutoDA research domain is still a relatively new research domain. Hence interviews are one of the best approaches to gathering requirements.
Method 3: Prototyping
Since the final goal of this research work is to explore and develop novel architecture to work around the limitations of existing AutoDA systems, prototyping methodology was selected because it allows the author to experiment with various implementation approaches while testing and assessing the prototype to identify potential areas for improvement.

Table 13: : Requirement Elicitation Methodologies

4.5 Discussion of Findings through Different Elicitation Methodologies

4.5.1 Literature Review

Finding	Citation
Poor datasets are led to model overfitting and poor generalizability of neural networks.	(Shorten and Khoshgoftaar, 2019)
Existing AutoDA techniques are highly resource intensive. Hence, it is essential to solving this high resource utilization in existing AutoDA techniques to use AutoDA techniques in real-world commercial applications.	(Lim et al., 2019)
It is challenging to reimplement many published AutoDA methods, limiting the broad applicability of AutoDA techniques. Hence, it is essential to develop plug-and-play AutoDA frameworks which are easy to set up.	(Müller and Hutter, 2021)

Search-free AutoDA methods are more efficient. However, their accuracy is comparatively less than the search-based AutoDA methods.	(Yang et al., 2022a)
The work by Cubuk et al., 2019b consider most practical AutoDA system, but their generalization performance is limited and can be further enhanced.	(Müller and Hutter, 2021; Yang et al., 2022a)

Table 14: Findings through Literature Review

4.5.2 Interviews

To gather opinions of domain and technical experts in ML, DL, and data science were chosen to be interviewed. A Google expert in ML and AI, an AI & data research graduate, two Ph.D. students in data science, a software engineer in AI & data, two data engineers, and two senior students were interviewed. Based on the following themes, **thematic analysis** was processed on the outcome of interviews. Only six participants granted permission to record their insights, and the relevant evidence has been included in **Appendix D**.

Codes	Theme	Analysis
DA Difficulties, Awareness of AutoDA, AutoDA Difficulties, Importance of AutoDA	Research gap and depth of scope	All the participants declared that most DL algorithms do not perform well if they are trained with low-diversity datasets, and data augmentation is a key technique to improve those low-diversity datasets. Also, they stated that selecting specific data augmentation techniques based on the dataset and model is time-consuming. Hence, they thought that resolving the limitations of existing AutoDA methods and making them available for public usage is essential and innovative.
Understating Common DA	Understanding the most used data augmentation techniques	Participants stated that based on their experience working in computer vision and data augmentation, geometric transformations, color space transformations, and kernel filters are the most used data augmentation techniques.
Availability, Tunings, Improvements	Features and suggestions for prototype	Apart from the ability to select the best data augment policies for a given dataset and model, there were a few

		suggestions for the prototype that were stated by the participants, which are listed below. • Rank the data augmentation techniques for the given dataset and results in comparison with the top 3 suggested techniques. • Automatically identify how to balance out the imbalanced dataset. • Availability as a python package.
The necessity of the AutoDA, Applicable Domains	The necessity of plug-and-play AutoDA system and contributions	All the participants clearly stated that AutoDA is the future of DA because AutoDA techniques will save the time of the developer as it limits the number of experiments that need to be performed to identify the best-performing DA technique for a given dataset. As a result, it will reduce project costs. Moreover, it addresses the knowledge limitations of the developer as implementing DA techniques needs technical and domain knowledge. Ultimately, all the participants thought that, since the AutoDA domain is still a relatively new concept and addressing the limitations of it would be very helpful to the domain.

Table 15: Findings through Interviews

4.5.3 Prototyping

The author utilized prototyping to do the following.

Criteria	Findings
To validate number of DA techniques to apply for single images.	Applying too many DA operations to images can create unnecessary augmented data points due to their unique characteristics. Based on experimental findings, it is possible to produce appropriate augmented images by applying only 2 DA operations. Check the *Table 15: Findings for number of DA operations to apply for single image* for evidence.
To monitor the computational resource consumption and verify the	Since the author has narrowed down the DA policy magnitude search range according to the literature findings and prototype findings, the optimal DA policy search time has been reduced.

proposed solution is resource friendly.	Moreover, all the experiments conducted on Apple M1 Pro MacBook with 16GB ram smoothly; therefore, it is proven that the proposed solution can also be used by ordinary users.
To evaluate differentiable DA technique approach will give the expected results.	Cubuk et al., 2019b by the Google Brain team is the widely accepted and most practical approach in the literature; the author has identified that it can be further enhanced by using differentiable DA operations while consuming less computational resources by conducting a series of porotypes. Additionally, differentiable DA operations help to improve the image classification model accuracy since it doesn't use fixed magnitude as Cubuk et al. 2019b.

Table 16: Findings through Prototyping

As previously stated, series of experiments were carried out to verify the number of DA techniques that could be utilized on individual images. The table below illustrates that applying over two DA operations on an image can damage its distinctive features.

Original Image	**Augmented Image**	**Number of DA Techniques Used**	**SSIM Value** (Higher the better)
		2	0.85
		3	0.62
		4	0.58

Table 17: Findings for number of DA operations to apply for single image

4.6 Summary of Findings

Id	Finding	Literature Review	Interviews	Prototyping
1	Identified research gaps in the AutoDA domain need to be filled to use AutoDA techniques in practice.	X	X	
2	Automate the entire data augmentation process, including image classification model training and turning.	X	X	X
3	Should automate the basic image manipulation data augmentation techniques for the initial prototype phase.	X	X	X
4	Provide the ability to add or remove data augmentation techniques from the search space.	X	X	
5	Should consider the class imbalance problem before the data augmentation.	X	X	X
6	Rank the data augmentation techniques for the given dataset and results comparison with the top suggested techniques.		X	
7	Users who don't know much about data augmentation and model training should be able to understand how the system works.	X	X	
8	Should use a standard programing language that utilizes the majority of machine learning models and datasets.	X	X	
9	Availability as a python package.	X	X	
10	Must be developed as a plug-and-play Python Library.	X	X	

Table 18: Summary of Findings

4.7 Context Diagram

The below context diagram (also known as level 0 data flow diagram (DFD)) visualizes the high-level interactions of the DAugtelligent with its high-level environment, including inputs and outputs of the proposed AutoDA system.

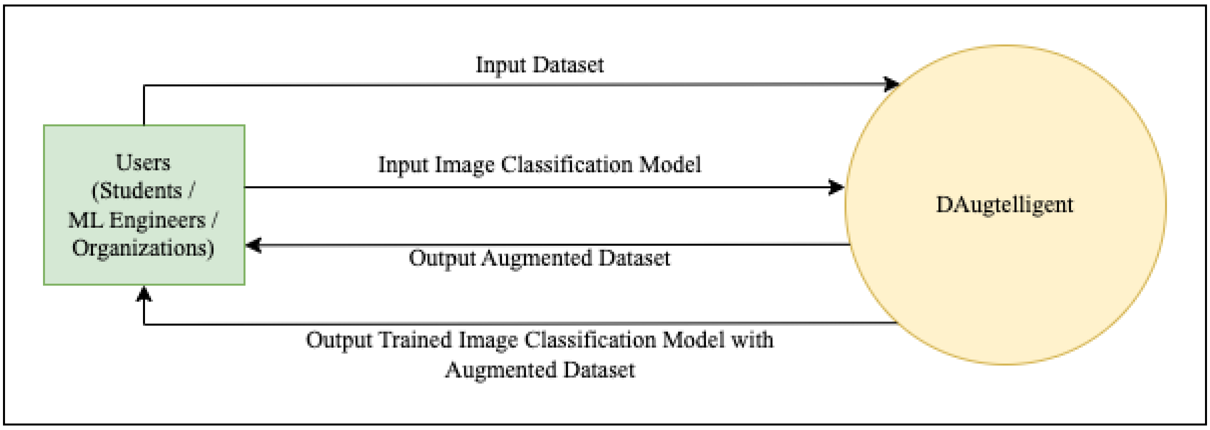


Table 19: Context Diagram (self-composed)

4.8 Use Case Diagram

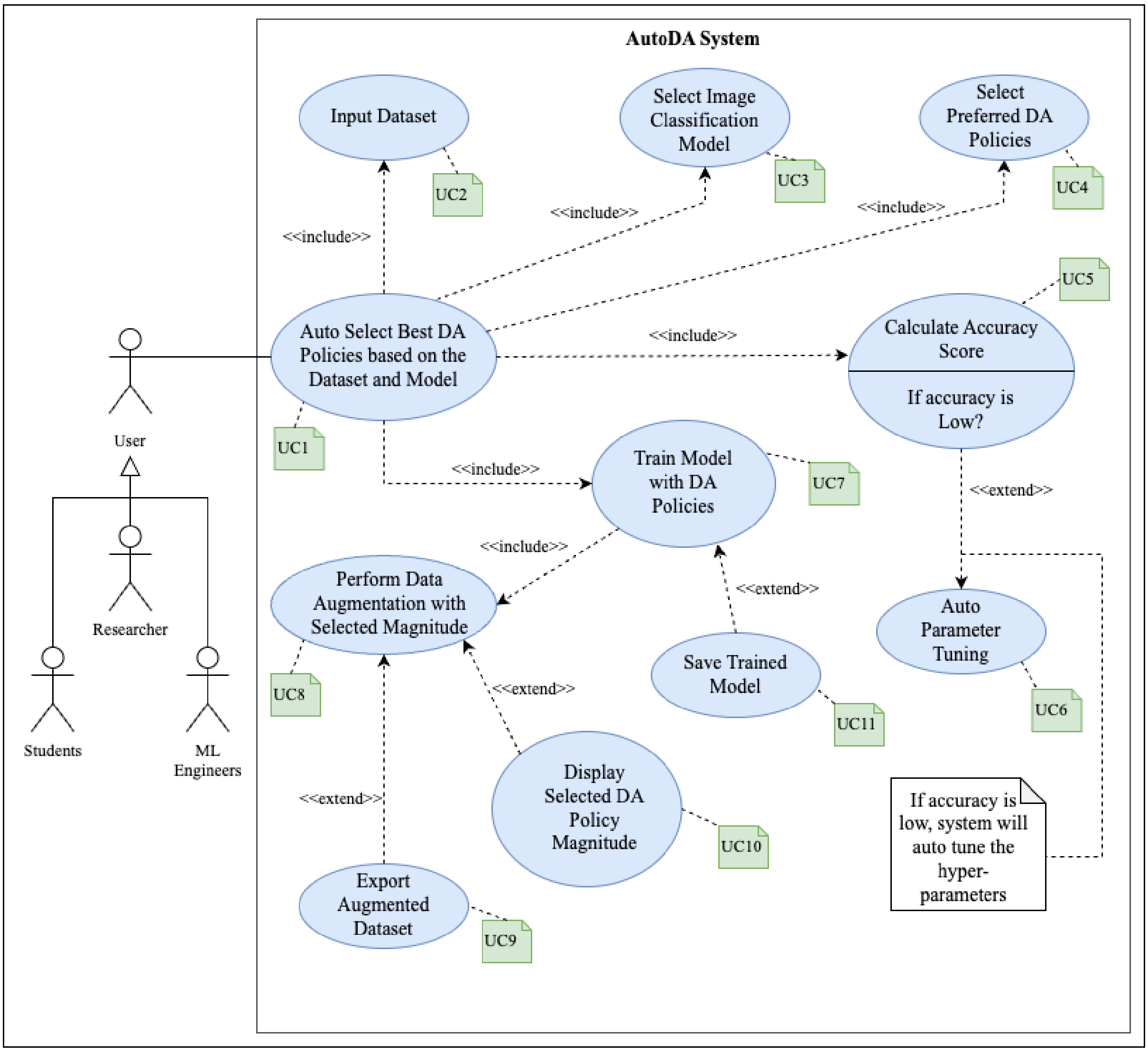


Figure 17: Use Case Diagram (self-composed)

4.9 Use Case Description

For additional explanations of alternative use case scenarios, please refer to **APPENDIX E**.

<table>
<tr><td>Use case name</td><td colspan="2">Auto Select Best DA Policies based on the Dataset and Model</td></tr>
<tr><td>ID</td><td colspan="2">UC1</td></tr>
<tr><td>Description</td><td colspan="2">User is able to find out best-performing DA policies for their dataset that improve the model performance.</td></tr>
<tr><td>Participating actors</td><td colspan="2">User</td></tr>
<tr><td>Preconditions</td><td colspan="2">User must select their dataset, image classification model, and preferred DA policies.</td></tr>
<tr><td>Extended use cases</td><td colspan="2">None</td></tr>
<tr><td>Included use cases</td><td colspan="2">Input Dataset, Select Image Classification Model, Select Preferred DA Policies, Calculate Accuracy Score, Display Selected DA Policy Magnitude</td></tr>
<tr><td rowspan="2">Main Flow</td><td>Actor</td><td>System</td></tr>
<tr><td>1. Select the dataset, image classification model, and preferred DA policies.
2. User executes AutoDA task on their dataset and image classification model.
5. Preview results.</td><td>3. Find common DA policy magnitude for user-preferred DA policies based on dataset and image classification model.
4. Display classification model performance statics and optimal common DA policy magnitude.</td></tr>
<tr><td>Alternative flows</td><td colspan="2">None</td></tr>
<tr><td>Exceptional flows</td><td colspan="2">E1: User did not input the dataset.
2.1 Automated data augmentation will not be executed.
E2: User did not input the image classification model.
2.2 Automated data augmentation will not be executed.
E3: User did not input both the dataset and image classification model.
2.3 Automated data augmentation will not be executed.
E4: User did not select the preferred DA policy set.
2.3 Automated data augmentation will be executed with the default DA policy set.</td></tr>
<tr><td>Post conditions</td><td colspan="2">The system will display the image classification model performance statics and optimal DA policy magnitude.</td></tr>
</table>

Table 20: Use case description for Auto Select Best DA Policies based on the Dataset and Model

4.10 Requirements Specification

The MoSCoW technique is used to categorize the most important requirements of the proposed AutoDA system.

Priority Level	Description
Must have (M)	The requirements that are important and essential to develop a successful Minimum Viable Product (MVP).
Should have (S)	The requirements that are important but not essential to develop a successful MVP. But the system will have some limitations.
Could have (C)	The requirements that are desirable but not essential to develop a successful MVP.
Will not have (W)	The requirements that are not developed during the initial stage of MVP.

Table 21: Summarization of "MoSCoW" prioritization levels

4.10.1 Functional Requirements

FR ID	Requirement	Priority Level	Use Case
FR1	Users must be able to select their preferred dataset.	M	UC2
FR2	Users must be able to select the preferred image classification model.	M	UC3
FR3	Users must be able to select preferred DA techniques.	M	UC4
FR4	Based on the selected dataset, image classification model and the DA techniques, the system must be able to find the optimal combinations of DA techniques with their magnitudes.	M	UC1
FR5	System must be able to train the given image classification model with auto selected DA policies.	M	UC10
FR6	System must be able to display performance statics of trained image classification model.	M	UC7
FR7	Users must be able to save the trained image classification model with augmented data.	M	UC11
FR8	Users must be able to save the augmented data points.	M	UC9
FR9	Users should have the ability to tune the important hyper-parameters of given image classification model through GUI.	S	UC10

FR10	System should be able to identify and solve the class imbalance issue of the selected dataset if present.	S	UC8
FR11	System should be able to show the auto selected DA techniques with their magnitude.	S	UC7
FR12	System could have ability to display previously trained image classification model history and the statics through GUI.	C	UC10
FR13	Users could have ability to select and retrain a previously trained image classification model via GUI.	C	UC10

Table 22: Functional Requirements

4.10.2 Non-Functional Requirements

NFR ID	Requirement	Description	Priority Level
NFR1	Efficiency	This project aims to decrease the computational resources needed for AutoDA tasks and enhance the broad applicability of AutoDA techniques, including for users who do not have access to high-performance GPU resources.	M
NFR2	Effectiveness	while improving the efficiency of AutoDA tasks, it is important to maintain a high level of accuracy. Because unnecessary selection of DA operations can negatively impact the performance of the target CNN model.	M
NFR3	Usability	The ultimate goal of the AutoDA system is to tackle the difficulties in traditional DA. Hence users must be able to perform DA tasks even without expertise in the DA domain.	M
NFR4	Scalability	The system should be able to perform AutoDA techniques in larger datasets and handle the workload smoothly.	S
NFR5	Maintainability	The software development best practices should be followed while building the product.	C
NFR6	Scalability	To manage the significant resource requirements of training a sophisticated image classification	C

		model, the application should be deployed on a cloud server.	

Table 23: Non-Functional Requirements

4.11 Chapter Summary

The main objective of this chapter was to figure out the functional and non-functional requirements of the proposed project. The rich picture diagram and Saunder's Onion model were used to identify and represent possible stakeholders and their interactions with the proposed system. Then, the author discussed about the requirement-gathering methodologies which were utilized to gather required data and opinions from the identified stakeholders. Then, the use case diagram of the proposed AutoDA system and functional as well as non-functional requirements were presented based on the gathered data using the requirement elicitation methodologies. Lastly, the identified requirements were prioritized using the "MoSCoW" requirements prioritization technique.

CHAPTER 5: SOCIAL, LEGAL, ETHICAL, & PROFESSIONAL ISSUES

5.1 Chapter Overview

The aim of this chapter is to identify the social, legal, ethical, and professional challenges that could potentially emerge during the project, and corresponding actions to mitigate such issues.

5.2 Breakdown of Social, Legal, Ethical, and Professional Issues

Social	Legal
• The project is devoid of any religious, political, ethical, or emotional bias. • To prevent communication conflicts and misunderstandings during the interviews necessary evidence and documents were provided.	• User data will not be stored. • The interviewers were informed that their responses will be recorded in the thesis, and their permission was obtained to include their names and designation in the thesis. • The languages, frameworks, tools and datasets utilized to develop the project were governed by an open-source license.
Ethical	**Professional**
• The interview participants and evaluators were informed about the project and their role in contributing to it. • The dissertation does not contain any plagiarism. All information and data presented are genuine, and the knowledge and facts obtained are properly cited.	• The entire source code of the project will be licensed with the MIT license. • The research adhered to both industry and academic standards. • All limitations of the project were identified and emphasized to the evaluators during the feedback collection process.

5.3 Chapter Summary

In this chapter, social, legal, ethical, and professional concerns that may arise in each section were identified, and corresponding mitigation strategies were presented.

CHAPTER 6: SYSTEM ARCHITECTURE & DESIGN

6.1 Chapter Overview

In this chapter, the design decisions of the proposed system are discussed utilizing appropriate design diagrams such as architecture diagrams, components diagram, data flow diagrams, and UI wireframes.

6.2 Design Goals

Design Goal	Description
Performance	Most of the existing AutoDA systems are not able to use by the ordinary audience due to the high computational resource consumption. Hence, it is essential that the system will perform even with limited computational resources.
Output Quality	The correction of selected optimal DA policies and their magnitude should be as good as possible. Describing why a user is getting the suggested DA policies using the given image classification model performance metrics.
Usability	The end objectives of this project is to overcome the optimal DA policy selection phase difficulties and reduce the complexity of the implementation. Therefore, it is important to develop the system straightforwardly to the dwindling number of clicks required to complete AutoDA tasks. Additionally, from the developer's point of view, the proposed AutoDA concept should easy-to-learn even without prior knowledge in DA.
Scalability	The system should be able to work with large datasets properly while optimizing the available computational resources.
Extendibility	Initially proposed system supports 14 DA technologies. However, there are a lot of DA technologies out there. Hence, the system should build with the best software practices (SOLID principles) in mind to make sure that while extending the system with more DA technologies does not cause any problems.

Table 24: Design Goals of the proposed system

6.3 System Architecture

6.3.1 System Architecture Diagram

The below illustration visualizes the architecture of the proposed AutoDA system.

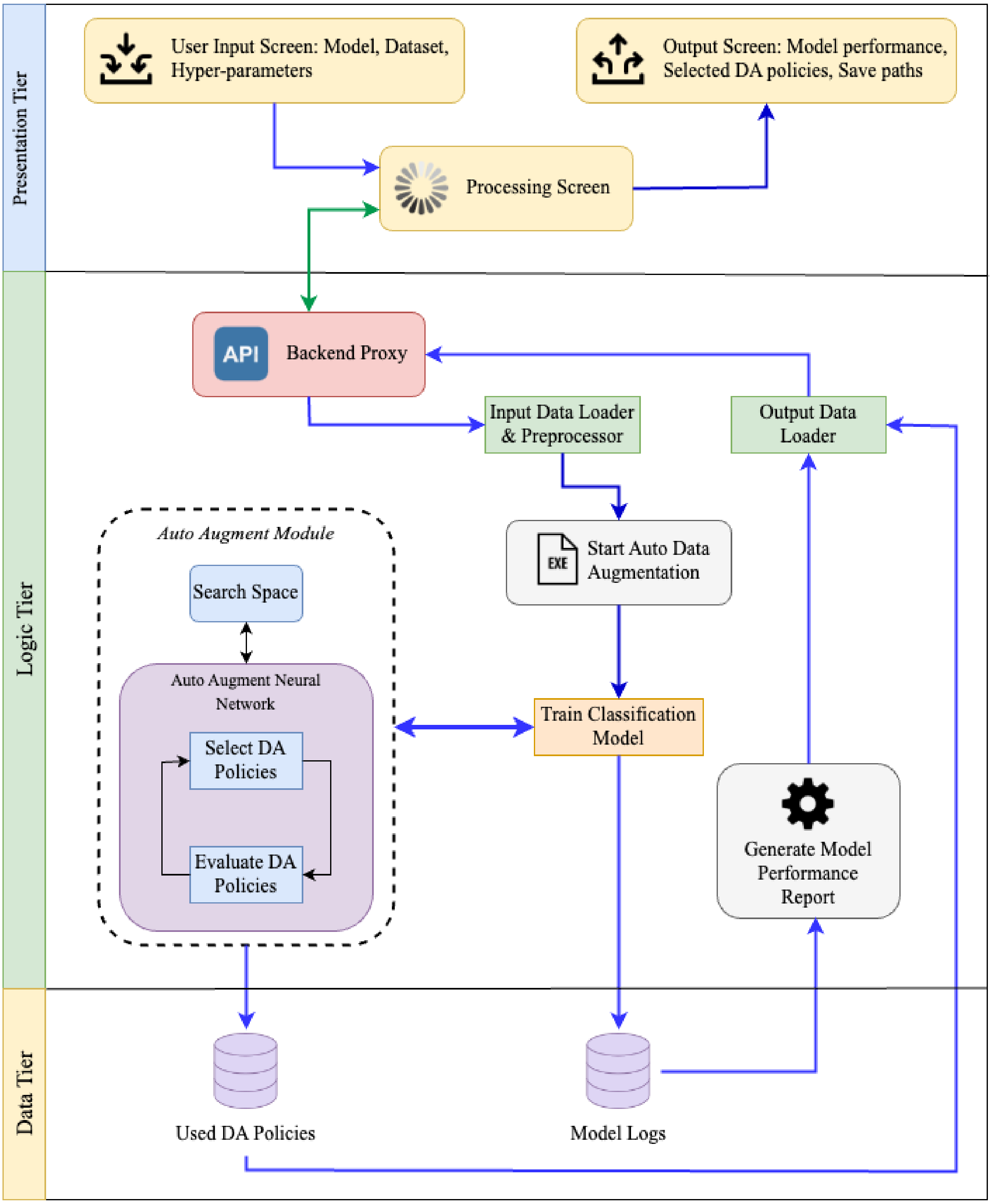


Figure 18: Three Tiered Architecture (self-composed)

6.3.2 Discussion of System Architecture Tiers

As illustrated in the above diagram, the high-level architecture of the proposed AutoDA system was designed based on the three-tier architecture concept (client tier (presentation tier), logic tier, and data tier) to ensure the system's reliability. The client tier is responsible for capturing all the user interactions, while the logic tier is responsible for executing all the logical operations based on the user interactions. And the data tier stores all necessary data for the proposed AutoDA system.

Data Tier

1. Used DA policies – This data storage allows storing already processed DA policies by AutoDA module with their magnitude and the given image classification model accuracy.
2. Model logs – This data storage allows storing classification model performance statics after every epoch.

Logic Tier

1. Backend proxy – Handle all the request from client tier and trigger necessary backend actions.
2. Input loader module – This module is to fetch and pre-process the user-selected dataset and image classification model into the system. Additionally, it will validate the given inputs.
3. Start auto data augmentation – After validating user inputs, this module responsible for executing AutoDA task on given dataset and image classification model.
4. Train image classification – Since the proposed AutoDA architecture is one-stage method both classification model training and optimal DA policy selection will execute simultaneously (refer to section 2.4.2 for additional information). This module is responsible for starting given classification model training.
5. AutoDA module – The core of the system which selects the best-performing DA policies based on the given dataset and image classification model. This component consists of three subcomponents.
    a. Search space – All the available DA policies. At the initial stage, this module consists of 14 DA policies which are Identity, Shear, Translate, Rotate, Cutout, Contrast, Auto Contrast, Equalize, Solarize, Solarize Add, Posterize, Color, Brightness, and Sharpness.
    b. Select DA policies – This is module responsible for explore optimal DA policies based on the evaluation function reward.
    c. Evaluate DA policies – This module is responsible for validating the performance of selected DA policies from the search space.

6. Generate model performance report – This module is to generate final performance report of the trained classification model with AutoDA techniques. This performance report includes accuracy, classification report, confusion metrics and ROC curves.
7. Output data loader – This module prepares the trained model performance report and auto selected DA policies.

Client Tier

1. User input screen – This page will allow users to select their dataset and image classification model and preferred DA techniques. Additionally, this will allow users to configure hyper-parameters of the classification model.
2. Processing Screen Wizard – Since it takes time for the AutoDA module to process and find out the best-performing DA policies, this page will display the progress of that task.
3. Output screen - This page will show the performance report of trained classification model, auto selected DA policies and trained classification model augmented data points saved path.

6.4 System Design

6.4.1 Selection of Design Paradigm

There are two different design paradigms that are commonly used for the software development process that is:

1. Object Oriented Analysis and Design (OOAD)
2. Structured Systems Analysis and Design Method (SSADM)

The goal of this project is to develop a novel approach to solve difficulties in existing AutoDA implementations. To archive this goal, it is essential to conduct a series of experiments and rapid changes to the code. Among the OOAD and SSADM, SSADM provides a more precise and easy process to improve existing software systems. Hence, the **Structured Systems Analysis and Design Method (SSADM)** method is ideal for this project.

6.4.2 Data Flow Diagram

The below illustration visualizes the level 1 DFD of the proposed AutoDA system, which provides additional details of the level 0 DFD (also known as context diagram). The level 0 DFD of the proposed AutoDA system was described during the SRS chapter.

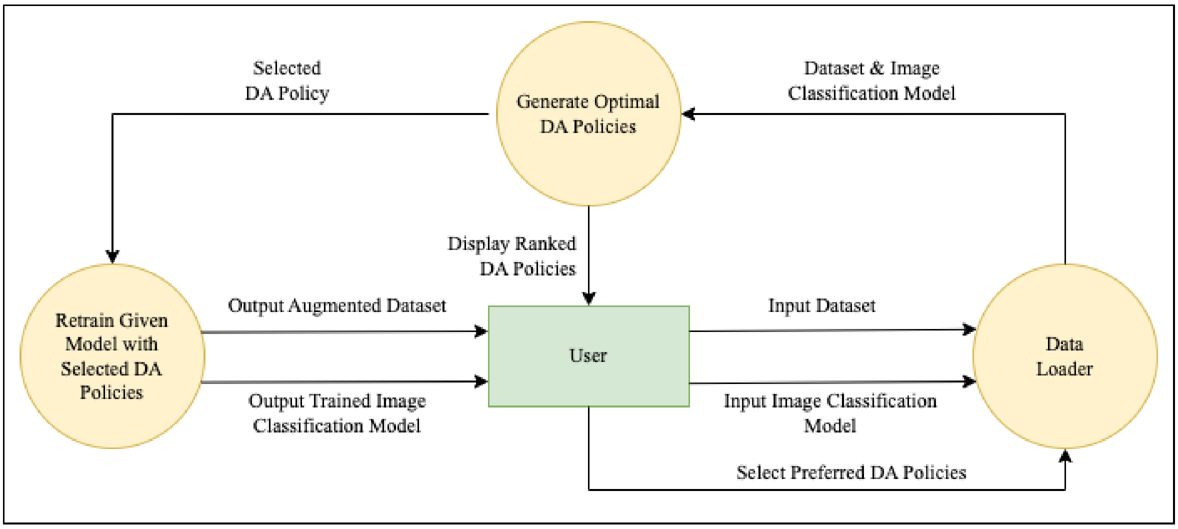


Figure 19: Data Flow Diagram - Level 1 (self-composed)

The level 1 DFD consists of 3 major data processes. Which are

1. Data loader
2. Generate optimal DA policies
3. Train given image classification model with user-preferred DA policies among the generated DA policies

The below illustrations depict level 2 data flow diagrams of the above-mentioned major processes, which provides a more detailed breakdown of the level 1 DFD major processes.

The below illustration visualizes level 2 DFD of the data loader module.

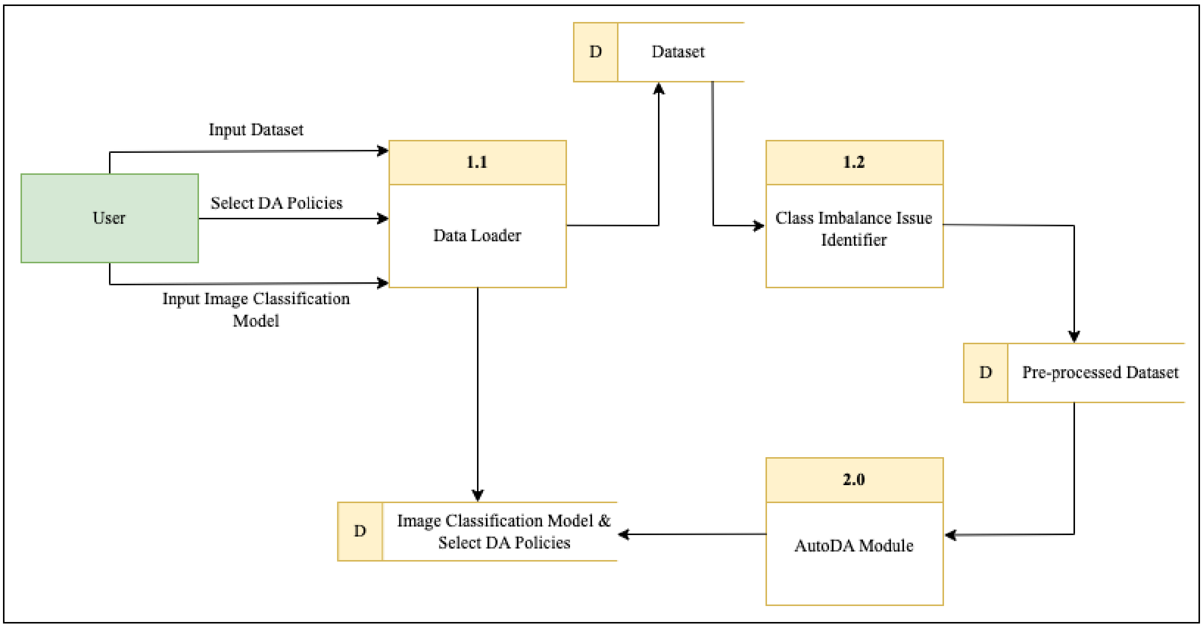


Figure 20: Data Loader - Data Flow Diagram - Level 2 (self-composed)

The below illustration depicts level 2 DFD of the AutoDA (core) module.

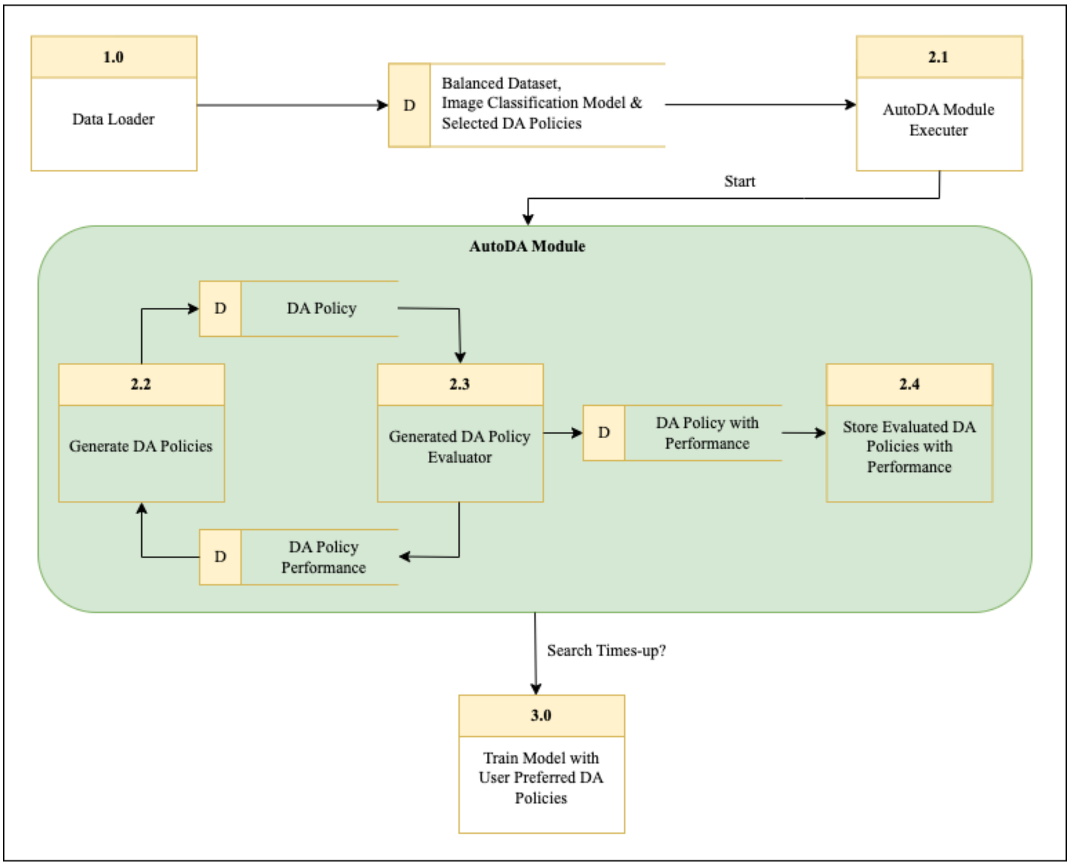


Figure 21: AutoDA Module - Data Flow Diagram - Level 2 (self-composed)

The below illustration depicts level 2 DFD of the given image classification model train module.

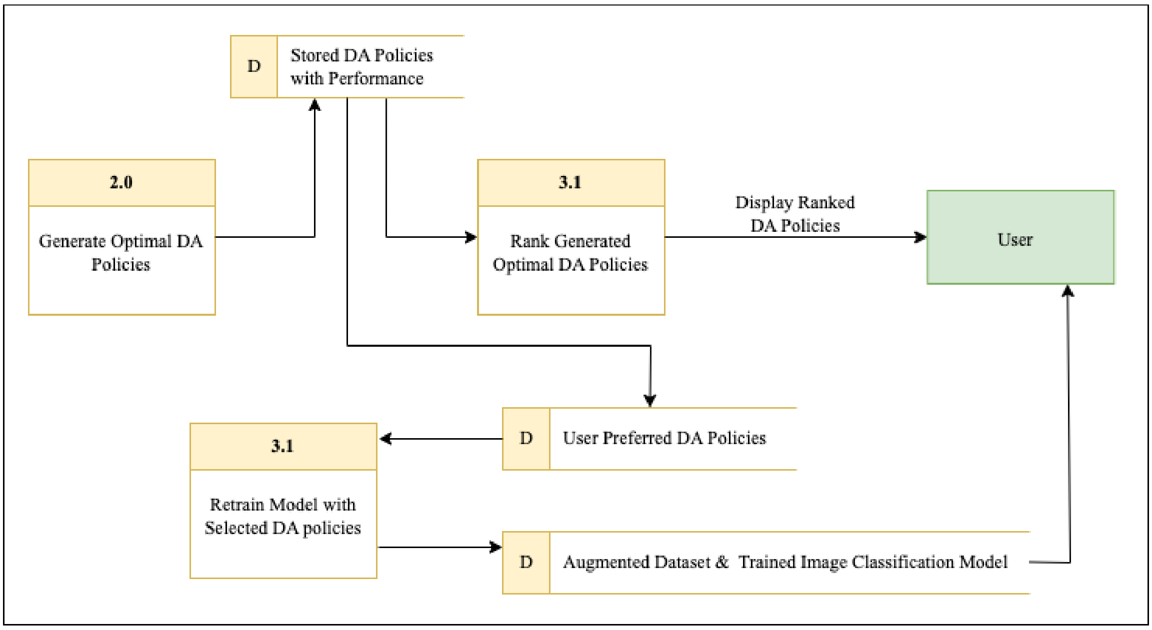


Figure 22: Train Image Classification Module - Data Flow Diagram - Level 2 (self-composed)

6.4.3 Design of the Proposed AutoDA System

As mentioned in Chapter 2, standard AutoDA system consists of three key components: search space, search strategy and evaluation strategy. In this section, the author discusses the designing of proposed AutoDA system search space, search strategy and evaluation strategy.

6.4.3.1 Search Space Design

6.4.3.1.1 DA Techniques Utilized in Search Space

After doing a thorough LR and conducting a series of interviews with domain and technical experts, the author concluded that the majority of users use 14 DA techniques. Which are:

- Identity
- Shear-X
- Shear-Y
- Auto Contrast
- Equalize
- Solarize
- Translate-X
- Translate-Y
- Rotate
- Cutout
- Posterize
- Color
- Brightness
- Sharpness

So, the proposed system will only consider automation in above DA operations during the initial stage.

6.4.3.1.2 Parameterization of Proposed Search Space

As mentioned in Chapter 2, the standard AutoDA system comprises three hyper-parameters: probability (P), magnitude (M), and the number of operations to apply to a single image (N). Below is a detailed explanation of the reparameterization of the traditional AutoDA search space proposed by the author.

P: Equal probability for all DA operations.

To elaborate more, the search space consists of 14 image DA operations, therefore the probability for selecting each operation is 1/14.

N: The value N will always be 2.

Experiments that have been conducted throughout the development have proved that applying more than 2 DA operations sequentially can damage the characteristics of the image. Check Section 4.5 for evidence.

M: Differentiable shared magnitude for all image transformations.

The effectiveness of DA policies is primarily governed by the magnitude hyper-parameter, which makes it a crucial aspect of the process. However, searching for a specific magnitude value for each operation can be time-consuming, and setting a static magnitude value for all operations can lead to poor performance of the target CNN. To address this challenge, the author has suggested a novel approach that involves using a differentiable shared magnitude value for all operations. This approach results in reduced search time while simultaneously enhancing the generalization performance of the target CNN.

6.4.3.2 Search Space Exploration & Evaluation Strategy Design

In Chapter 2, it was discussed that the standard AutoDA workflow comprises two stages, namely generation and application. When these two stages operate concurrently, the resulting architecture is known as a one-stage AutoDA system. The proposed AutoDA architecture is a one-stage system that performs both generation and application stages simultaneously while tuning them based on the performance of the target CNN.

The main aim of the search and evaluation process in the proposed DAutify system is to search for most suitable common magnitude value (M) for all DA operations in the search space, the author has developed a neural network that consists of a single hyper-parameter. This hyper-parameter represents the magnitude of all DA operations, and it is tuned based on the performance of the target CNN.

6.4.3.2.1 Search Space Exploration & Evaluation Stopping Strategy Design

Most of the previous studies conclude the optimal DA policy search process once the desired accuracy of the target model is achieved. However, since the proposed system employs a distinct search and evolution strategy from traditional AutoDA systems, an early stop callback is designed. The purpose of this callback function is to monitor the hyper-parameter M, and if it remains unchanged for five consecutive epochs, the optimal magnitude search process will terminate. The target CNN will then train with the magnitude value at that point until the epoch count limit is reached.

6.4.3.3 Workflow of Proposed AutoDA Module

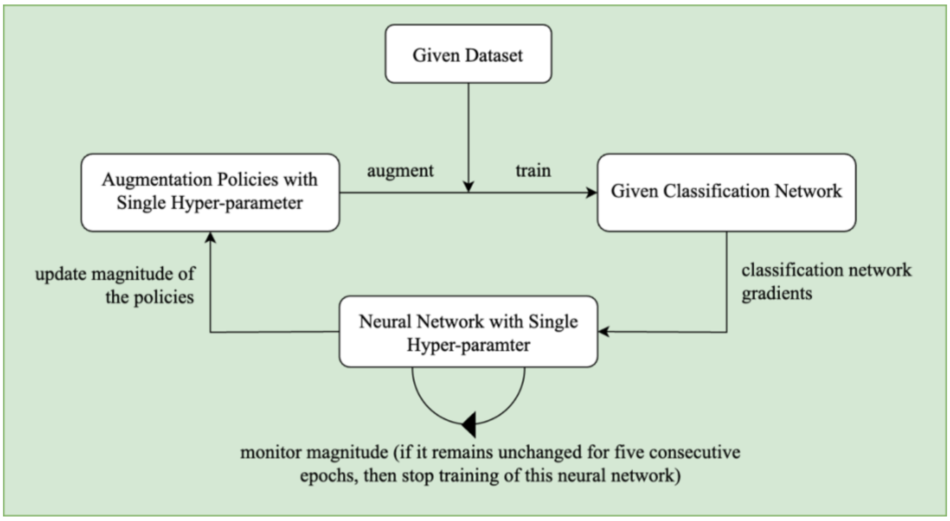


Figure 23: Workflow of DAugtify AutoDA system

6.4.4 System Activity Diagram

The below illustration depicts the activity diagram of the proof-of-concept system. It stimulates the general workflow of the prototype system, including user inputs, processes, and outputs based on the developer's point of view.

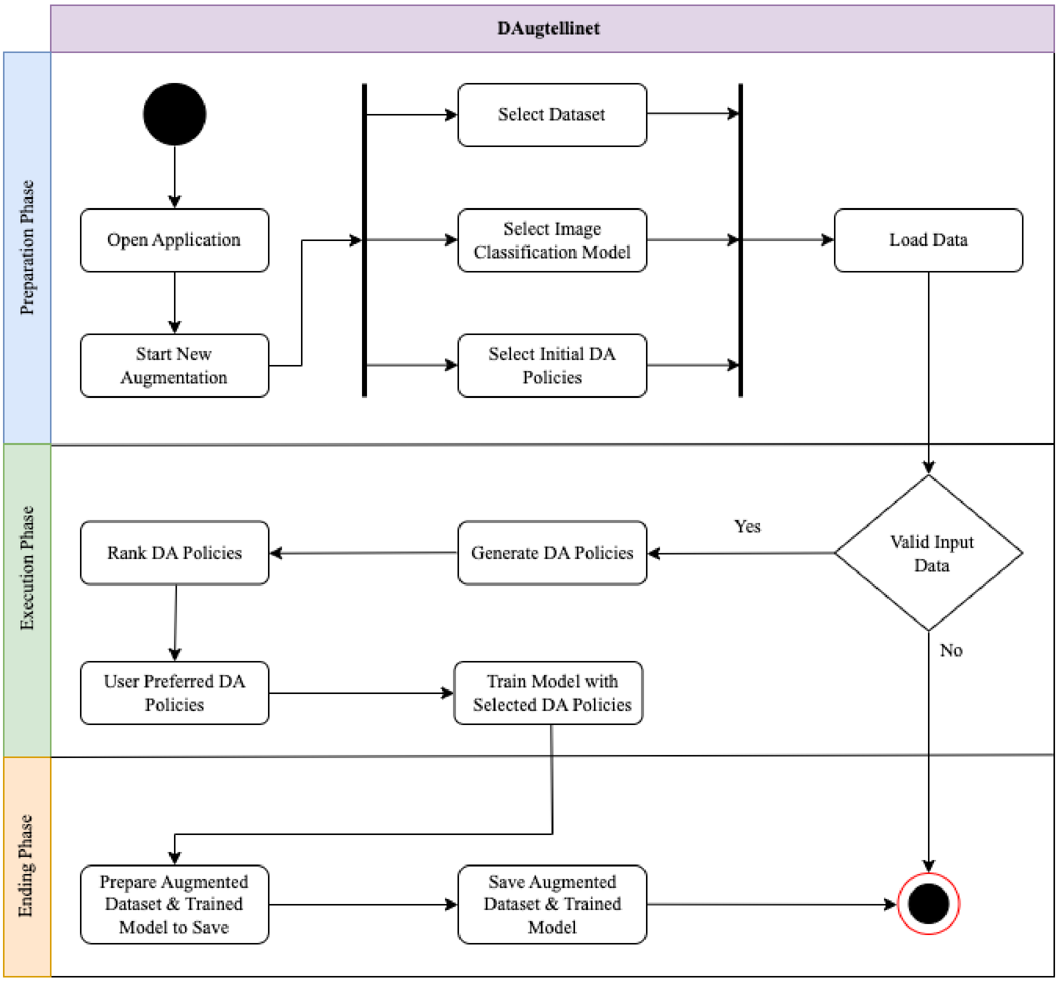


Figure 24: System Activity Diagram (self-composed)

6.4.5 UI Design

The user interface wireframes of the proposed DAugtify system are located in the **Appendix F** section.

6.5 Chapter Summary

The main objective of this chapter was to discuss about design goals and system architecture aspects of the proposed AutoDA system. To begin with, the author has presented the design goals of the proposed system with the proper justification. Then three-tier system architecture diagram and other required design diagrams, along with the data flow diagrams, were presented. Later, the core contribution, which is the novel search-free AutoDA concept, was discussed. Lastly, initial UI design wireframes for MVP have been presented.

CHAPTER 7: IMPLEMENTATION

7.1 Chapter Overview

This chapter discusses the implementation of core functionalities of the proposed AutoDA system, proving the research hypothesis. Moreover, the author discusses the selection of the technology stack, programming languages, and additional tools that are utilized to develop the prototype, along with the respective reasons for each selection. Lastly, this chapter discusses how the design decisions were translated into a working prototype by demonstrating the required code snippets of the initial implementation.

7.2 Technology Selection

7.2.1 Technology Stack

The technology stack that was used to develop the three-tier architecture of the proposed AutoDA system is described in the below illustration.

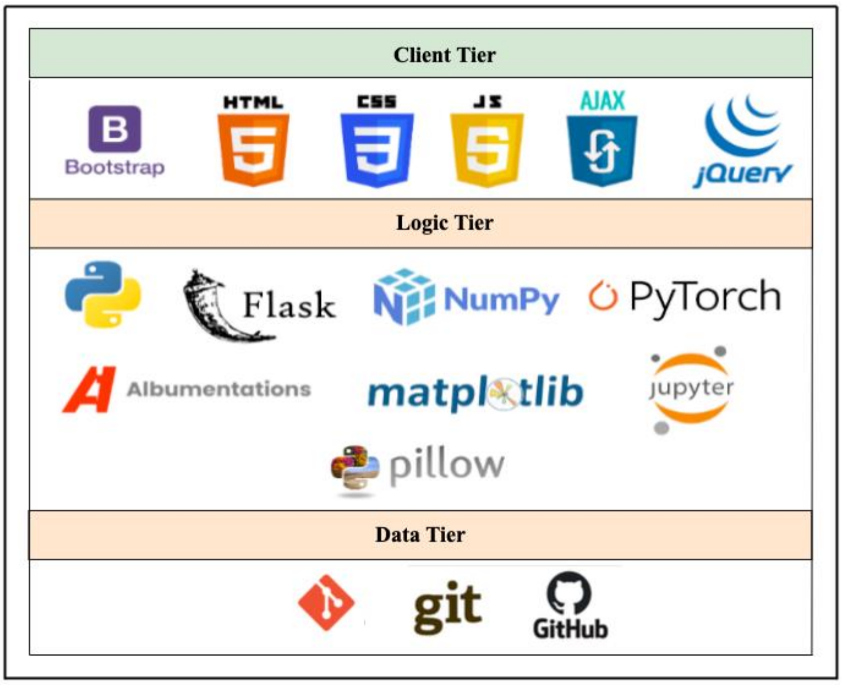


Figure 25: Technology Stack

7.2.2 Programming Language Selection

Programming Language	Justification for Selection
Python	Among the available programming languages, Python has been utilized in most of the ML projects due to the vast collection of community and supporting libraries.

JavaScript	For the front end, it is easy to make a highly interactive and inviting user experience using JavaScript.

Table 25: Justifications for Programming Language Selection

7.2.3 Development Framework Selection

Development Framework	Justification for Selection
PyTorch	Most of the existing AutoDA works used PyTorch for their implementation. Hence to do a fair comparison, PyTorch was selected. Moreover, the benchmark proves that the performance of PyTorch is better than the TensorFlow.
Flask	Easy to build Application Programming Interfaces (APIs) for Python.

Table 26: Justifications for Development Framework Selection

7.2.4 Libraries

Libraries	Justification for Selection
Numpy	Supports an extensive variety of mathematical and algebraic development.
Matplotlib	Supports an extensive variety of visualization methods for use in data analysis.
Albumentations	Supports an extensive variety of DA augmentation methods.
Pillow	Supports an extensive variety of image-related operations like opening and saving images.

Table 27: Justifications for Development Libraries Selection

7.2.5 IDE Selection

IDE	Justification for Selection
Google Colab	Since the author doesn’t have access to GPU resources and Google Colab is free to use, it was utilized to run the experiments on GPUs.
VSCode	Powerful IDE to build both front-end and back-end while being very simple and flexible to use.

Table 28: Justifications for IDE Selection

7.2.6 Summary of Technologies & Tools Selection

Component	Tools
Programming Languages	JavaScript, Python
Development Frameworks	PyTorch, Flask

Libraries	Albumentations, Numpy, Pillow, Matplotlib
IDEs	VSCode
Version Control	Git, GitHub

Table 29: Summary of Technologies and Tools Selection

7.3 Implementation of Core Functionalities

As mentioned in Chapter 2, standard AutoDA system consists of three key components: search space, search strategy and evaluation strategy. In this section, the author discusses the implementation of proposed AutoDA system search space, search strategy and evaluation strategy.

7.3.1 Implementation of search space

As mentioned in Section 6.4.4, the initial implementation of proposed system will consist of 14 DA operations. The implementation of the search space is shown in the code snippets below.

```
DefaultOpSet = [
    Identity(), Brightness(0.1, 1.9),
    Sharpness(0.1, 1.9), Rotate(0, 30),
    Cutout(0, 0.2), Equalize(),
    Posterize(0, 4), TranslateX(0, 0.45),
    TranslateY(0, 0.45), AutoContrast(),
    Contrast(0.1, 1.9), Color(0.1, 1.9),
    Solarize(0, 0xFF), SolarizeAdd(0, 0x6E),
    ShearX(0, 0.3), ShearY(0, 0.3),
]
```

Figure 26: DA techniques in the Search Space

Below code snippets depict the definition of hyper-parameter to calculate common magnitude (M) for all the DA policies.

```
# Hyper-parameter to calculate optimal M value
self.magnitude_logits = nn.Parameter(torch.empty(()).normal_(0, 1))
```

Figure 27: Implementation of search-space - M hyper-parameter

7.3.2 Implementation of search & evaluation strategy of DA policies

As stated in Section 6.4.4, the exploration of the search space and evaluation of selected DA policies are performed through a neural network. The implementation of the forward mechanism for the proposed neural network is as follows:

```python
def forward(self, x: torch.Tensor) -> torch.Tensor:
    # Normalize the image to [0, 1].
    if self.normalized:
        x = (x + 1) / 2

    # Definition of M hyper-parameter
    magnitude = self.magnitude_logits.sigmoid()

    # Generating randomly selected 2 DA operation with M
    temp_policy_set = []
    for _ in range(self.num_ops):
        policies = random.choice(self.opset)
        policyName = policies.get_operation_name()
        temp_policy_set.append(policyName)
        x = policies(x, magnitude)

    # Storing the previously utilized DA policies.
    if SelectedPolicies.full():
        SelectedPolicies.get()
    else:
        SelectedPolicies.put(temp_policy_set)
        temp_policy_set = []

    # Normalize the image to [-1, 1].
    if self.normalized:
        x = 2 * x - 1
    return x
```

Figure 28: Implementation of search & evaluation strategy

7.3.3 Implementation of search & evaluation strategy stop callback

As aforementioned, the exploration of the search space and evaluation of selected DA policies are performed through a neural network. Therefore, an early stop callback is designed for the AutoDA neural network (Check section 6.4.4 additional information).

```python
def on_validation_epoch_end(self, trainer, pl_module):
    logs = trainer.callback_metrics
    current_val = logs.get(self.monitor)
    if self.best_val is None:
        self.best_val = current_val
    elif current_val == self.best_val:
        self.wait += 1
        if self.wait >= self.patience:
            print(f"\nStopping training due to no change in {self.monitor} for {self.patience} epochs.")
            trainer.should_stop = True
    else:
        self.best_val = current_val
        self.wait = 0
```

Figure 29: Implementation of search & evaluation function stop callback

7.3.4 Implementation of executing the AutoDA module and the target CNN training

The PyTorch Lightning Module was used to train both the AutoDA module and the target CNN. To leverage the functionalities provided by PyTorch Lightning Module to train the models, several key functions were overridden based on the requirements of the proposed system. These include the **__init__**, **configure_optimizers**, **forward**, **validation_step**, and **test_step** functions. The implementation of these functions is shown below.

__init__: This function initializes the learning rate, total number of epochs, and the target CNN.

```python
def __init__(self, leaning_rate: float, total_epochs: int, model: nn.Module):
    super().__init__()
    # Initialize learning rate for target CNN
    self.leaning_rate = leaning_rate

    # Initialize total epoch count
    self.total_epochs = total_epochs

    # Initialize DAugtify Module
    self.data_augmentor = DAugtifyModule()

    # self.model = wide_resnet_28x10(10)
    self.model = model
```

Figure 30: Implementation of PyTorch Lightning init function

configure_optimizers: This function initializes the optimizer, and scheduler for both models.

```python
# configure hyper-parameters
def configure_optimizers(self):
    param_groups = [
        {"params": self.augmentor.parameters(), "lr": 0.01},
        {"params": self.model.parameters(), "lr": self.leaning_rate},
    ]

    optimizer = optim.Adam(param_groups, lr=self.leaning_rate)

    #  reduces the learning rate of the optimizer during training
    scheduler = lr_scheduler.CosineAnnealingLR(optimizer, self.total_epochs)

    return [optimizer], [scheduler]
```

Figure 31: Implementation of PyTorch Lightning configure_optimizers function

training_step: This function executes the training for both models, where the DAugtify module performs data augmentation and requests the target CNN to compute its loss value.

```python
# This method takes a batch of input images and their corresponding labels,
# applies data augmentation using DAugtify module,
# and feeds them through the model. Lastly computes the loss and logs it.
def training_step(self, batch, batch_idx: int) -> torch.Tensor:
    x, labels = batch

    # apply data augmentation to input images
    x_augmented = self.augmentor(x)

    #  save augmented images
    img_to_save = TF.to_pil_image(x_augmented[0].cpu())
    augmented_label = labels[0].item()
    img_path = os.path.join(
        f"./datasets/augmented/{augmented_label}", f"augmented_img_{batch_idx}_0.png")
    os.makedirs(os.path.dirname(img_path), exist_ok=True)
    img_to_save.save(img_path)

    # aaugmented images passes through the target CNN to generate the predicted scores for each class
    logits = self.model(x_augmented)

    # computes the loss between the predicted scores and the ground truth labels.
    loss = F.cross_entropy(logits, labels)


    self.log("train_loss", loss, prog_bar=False)
    self.log("train_acc", self._accuracy(logits, labels, topk=1))
    self.log("common_magnitude", self.augmentor.get_magnitude(), prog_bar=True)

    self.common_magnitude = self.augmentor.get_magnitude()
    return loss
```

Figure 32: Implementation of PyTorch Lightning forward pass

validation_step: This function executes validation process on the validation dataset and calculate the target CNN loss value.

```python
# used for validating the model on the validation set.
def validation_step(self, batch, batch_idx: int):
    x, labels = batch

    # calculating loss
    logits = self.model(x)
    loss = F.cross_entropy(logits, labels)

    self.log("val/loss", loss)
    self.log("val/acc", self._accuracy(logits, labels, topk=1), prog_bar=True)

    # return loss for update both traget CNN and augmentor hyper-parameters
    self.predicted_loss = loss
```

Figure 33: Implementation of PyTorch Lightning validation step

7.3.5 Implementation of target CNN accuracy calculation

```python
# This is a helper method that computes the accuracy of the model by
# comparing the predicted labels with the ground truth labels.
def _accuracy(self, logits: torch.Tensor, labels: torch.Tensor, topk: int = 1) -> torch.Tensor:
    _, indices = torch.topk(logits, k=topk)
    correct = indices == labels.unsqueeze(-1)
    return correct.any(-1).float().mean()
```

Figure 34: Implementation of target CNN accuracy calculation

7.4 User Interface

The below figure demonstrates the main user input configuration screen and other user interfaces of the proposed DAugtify system are located in the **Appendix G** section.

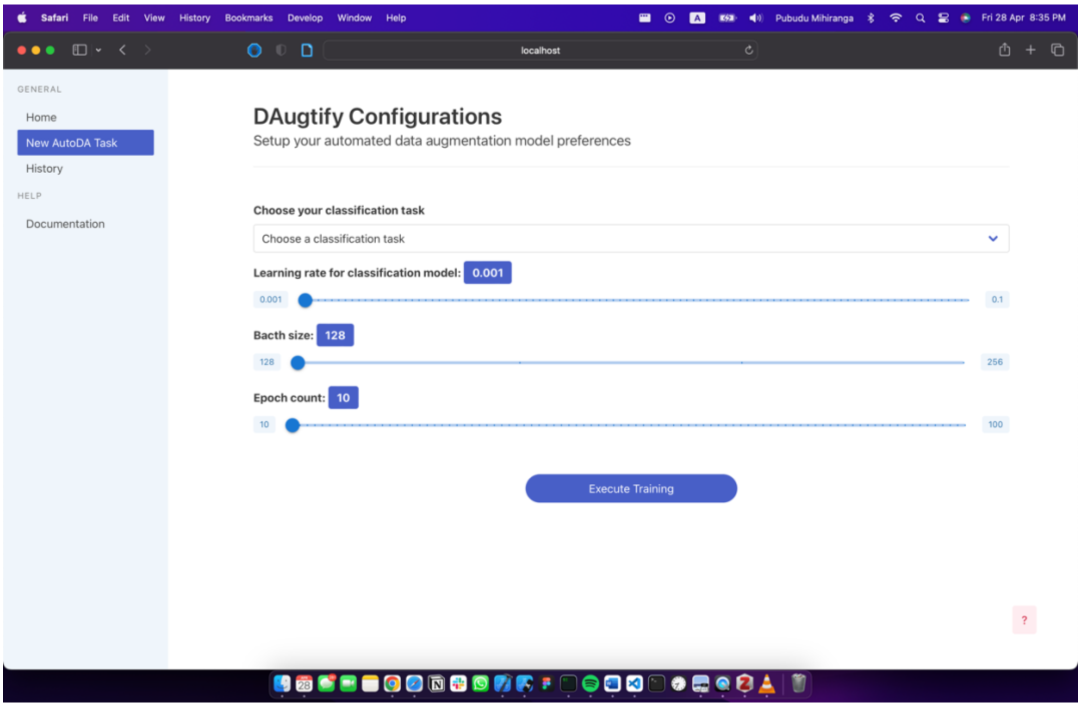


Figure 35: Main user input configuration screen

7.5 Chapter Summary

This chapter discussed the technology stack, programming languages, and additional tools that are utilized to develop the proposed research prototype. Moreover, the implementation of the core functionalities of the proposed research project was discussed with relevant code snippets which prove the research hypothesis and contributions.

CHAPTER 8: TESTING

8.1 Chapter Overview

The primary objective of this chapter is to present an extensive analysis of the testing methodologies implemented to assess the effectiveness of the novel AutoDA approach.

8.2 Objectives & Goals of Testing

The main objective of conducting software testing is to verify that the system is functioning in accordance with the obtained requirements. Below are listed the primary objectives of the testing process for DAugitfy,

- To verify that all models within the DAugitfy system are functioning as intended and have undergone thorough testing in order to achieve optimal results.
- To determine benchmarking methodology for the system and to perform fair benchmarking among existing works.
- To ensure whether the system fulfills the "Must have" and "Should have" functional requirements, as identified through the MoSCoW technique.
- To ensure whether the system fulfills the essential non-functional requirements.
- To recognize potential areas for enhancements & bug fixes that could be implemented in the system.

8.3 Testing Criteria

The testing criterion has been established to evaluate the system in two distinct ways, with the aim of minimizing the gap between expected and implemented systems.

1. **Functional Quality**: entails testing the features and critical technical components of the system to determine how well it adheres to the specified design, as per the functional requirements.
2. **Structural Quality**: involves testing the non-functional requirements of the system while also ensuring that it meets the performance criteria of the functional requirements.

8.4 AutoDA Model Testing & Evaluation

As outlined throughout this document, the primary objective of AutoDA systems is to enhance the performance of a given image classification model by utilizing AutoDA techniques. Hence, the testing of the proposed DAugtify system was conducted using the accuracy of the given classification model as the primary metric to test and evaluate. However, to archive more comprehensive testing and evaluation criteria, additional image classification model evaluation metrics such as F1-score, precision, and recall were also utilized.

The proposed DAugtify system was tested on standard datasets and effective classification models tailored for each dataset. The tested datasets, along with their corresponding tailored CNN architectures, can be summarized as follows:

Dataset	CNN Architecture
CIFAR10	WideResNet-28x10
	ResNet18
SVHN	VGG16
	ResNet18
MNIST	LeNet-5

Table 30: Datasets and CNNs used for testing

8.4.1 Testing Results

Criteria				Accuracy	Time Taken for Training
Description	**Epochs**	**Learning Rate**	**Batch Size**		
CIFAR10 & WideResNet-28x10					
WideResNet-28x10 CNN architecture was trained on the CIFAR10 dataset and without any data augmentation.	10	0.05	128	78%	9H 47Min
WideResNet-28x10 CNN architecture was trained on the CIFAR10 dataset and with proposed DAugtify AutoDA system.	10	0.05	128	86%	9H 50Min
CIFAR10 & ResNet18					
ResNet18 CNN architecture was trained on the CIFAR10 dataset and without any data augmentation.	10	0.05	128	64%	1H 15Min

WideResNet-28x10 CNN architecture was trained on the CIFAR10 dataset and with proposed DAugtify AutoDA system.	10	0.05	128	75%	1H 15Min
SVHN & VGG16					
VGG16 CNN architecture was trained on the SVHN dataset and without any data augmentation.	10	0.05	128	93%	2H 24Min
VGG16 CNN architecture was trained on the SVHN dataset and with proposed DAugtify AutoDA system.	10	0.05	128	96%	2H 34Min
SVHN & ResNet18					
ResNet18 CNN architecture was trained on the SVHN dataset and without any data augmentation.	10	0.05	128	89%	2H 17Min
ResNet18 CNN architecture was trained on the SVHN dataset and with proposed DAugtify AutoDA system.	10	0.05	128	95%	2H 20Min
MNIST & LeNet-5					
LeNet-5 CNN architecture was trained on the MNIST dataset and without any data augmentation.	10	0.05	128	98%	1Min 27S
LeNet-5 CNN architecture was trained on the MNIST dataset and with proposed DAugtify AutoDA system.	10	0.05	128	98%	2Min 7S

Table 31: Testing results

Refer to the **Appendix H** for evidence supporting the testing outcomes mentioned above and a more detailed performance report of each CNN architecture.

8.4.2 Evaluation of Testing Results

While keeping the same epoch count and other hyperparameter values for all scenarios, the DAugtify system significantly improved the performance of the following test cases within a competitive training time:

- The WideResNet-28x10 CNN architecture achieved an accuracy increase from 78% to 86% on the CIFAR10 dataset.
- The ResNet18 CNN architecture achieved an accuracy increase from 64% to 75% on the CIFAR10 dataset.
- The VGG16 CNN architecture achieved an accuracy increase from 93% to 96% on the SVHN dataset.
- The ResNet18 CNN architecture achieved an accuracy increase from 89% to 95% on the SVHN dataset.

Moreover, the LeNet5 CNN architecture achieved the same accuracy of 98% on the MNIST dataset on both scenarios, indicating that the proposed system was able to prevent model overfitting.

The results suggest that the proposed AutoDA system is an effective technique for enhancing CNN performance within a competitive training time.

8.5 Benchmarking

Similar to the testing criteria, the AutoDA systems are typically benchmarked by measuring the accuracy of image classification models on a standard dataset using the AutoDA solutions. However, many existing AutoDA methods can not utilized for benchmarking due to their high computational resource requirements. Therefore, the author proposes the following criteria for benchmarking:

Baseline	Measure the performance of various CNN architectures on the specified dataset without any data augmentation.
AutoAugment (AA) (Cubuk et al., 2019a)	Measure the performance of various CNN architectures on the specified dataset with the Pytorch implementation of AA.
RandAugment (RA) (Cubuk et al., 2019b)	Measure the performance of various CNN architectures on the specified dataset with the Pytorch implementation RA.

Ours	Measure the performance of various CNN architectures on the specified dataset with proposed AutoDA solution.

Table 32: Benchmarking Criteria

By adopting these criteria, the author can assess the effectiveness of proposed AutoDA solution in a more practical and feasible manner. The benchmarking results as follows:

CNN Architecture	**Configurations**			**Accuracy Rates** (higher the better)			
	Epochs	**Learning Rate**	**Batch Size**	**Baseline**	**AA** (Cubuk et al., 2019a)	**RA** (Cubuk et al., 2019b)	**Ours**
Benchmarking for CIFAR10							
WideResNet-28x10	10	0.05	128	78%	78%	82%	**86%**
ResNet18	10	0.05	128	64%	66%	69%	**75%**
Benchmarking for SVHN							
VGG16	10	0.05	128	93%	92%	94%	**96%**
ResNet18	10	0.05	128	89%	87%	87%	**95%**
Benchmarking for MNIST							
LeNet5	10	0.05	128	98%	95%	94%	**98%**

Table 33: Benchmarking Results

The results indicate that the proposed AutoDA system achieved state-of-the-art performance in all scenarios. **Appendix H** provides evidence supporting these findings.

8.6 Functional Requirement Testing

The Black-Box testing methodology was employed to assess the functionalities of DAugtify based on the functional requirements outlined in Chapter 4.

Test Case	**Action**	**Expected Outcome**	**Actual Outcome**	**Status**
1	Select dataset.	Dataset gets stored.	Dataset gets stored.	Pass
2	Select image classification model.	Image classification model gets stored.	Image classification model gets stored.	Pass
3	Adjust hyper-parameters of	New hyper-parameter values get stored.	New hyper-parameter values get stored.	Pass

	selected image classification model.			
4	Select user preferred DA techniques from available ones.	Selected DA techniques gets stored.	Selected DA techniques gets stored.	Pass
5	Click on AutoDA task execute button.	Start search for optimal DA policies.	Start search for optimal DA policies.	Pass
		Concurrently train the selected image classification model.	Concurrently train the selected image classification model.	Pass
6	The system guides the user to the output display page.	Display performance statics of trained image classification model.	Display performance statics of trained image classification model.	Pass
		Display auto selected DA techniques and shared magnitude.	Display auto selected DA techniques and shared magnitude.	Pass
7	System attempts to save trained image classification model.	Save trained image classification model and display local machine path.	Save trained image classification model and display local machine path.	Pass
8	System attempts to save augmented data points.	Save augmented data points and display local machine path.	Save augmented data points and display local machine path.	Pass
9	Navigate to history page.	Navigate to history page and display AutoDA task history if available.	Navigate to history page and display AutoDA task history if available.	Pass
10	Star new AutoDA task.	Clean previous user inputs and configurations.	Clean previous user inputs and configurations.	Pass

Table 34: Functional Requirement Testing

8.7 Module Integration & Testing

Module	Input	Expected Output	Actual Output	Status
Input Loader	User selected dataset.	Dataset get attached for system.	Dataset get attached for system.	Passed
	User selected image classification model.	Image classification model get attached for system.	Image classification model get attached for system.	Passed
	User selected hyper-parameters values for learning rate and epochs count.	hyper-parameters values for learning rate and epochs count get attached for system.	hyper-parameters values for learning rate and epochs count get attached for system.	Passed
AutoDA Module	Dataset & image classification model.	Search for optimal data augmentation techniques.	Search for optimal data augmentation techniques.	Passed
Classification Report Generation Module	Training logs of classification model.	Generate classification report.	Generate classification report.	Passed
Output Loader Module	Classification report.	Display classification report.	Display classification report.	Passed
	Local path for augmented.	Display path for augmented data.	Display path for augmented data.	Passed
	Local path for classification model	Display path for trained classification model.	Display path for trained classification model.	Passed

Table 35: Module Integration & Testing

8.8 Non-Functional Requirement Testing

8.8.1 Efficiency and Effectiveness Testing

The aim of this project is to minimize the computational resource consumption required for the AutoDA task while holding competitive outcome and make it accessible to a broader range of regular users. Therefore, efficiency and effectiveness are the primary non-functional requirement identified for this research. The author has evaluated the overall computational resource consumption and overall CNN architecture training time as part of the non-functional requirement testing. The below table depicts the non-functional testing results.

Similar to aforementioned testing criteria, the term 'baseline' refers to the CNN training without DA.

Training Method	Configurations			Accuracy (%)	Avg RAM Usage (GB)	Avg Training Time
	Epochs	Learning Rate	Batch Size			
CIFAR10 & WideResNet-28x10						
Baseline	10	0.05	128	78	1.07 GB	9H 47Min
With DAugtify	10	0.05	128	86	1.61 GB	9H 50Min
CIFAR10 & ResNet18						
Baseline	10	0.05	128	64	584 MB	1H 15Min
With DAugtify	10	0.05	128	75	690 MB	1H 15Min
SVHN & VGG16						
Baseline	10	0.05	128	93	771 MB	2H 24Min
With DAugtify	10	0.05	128	96	854 MB	2H 34Min
SVHN & ResNet18						
Baseline	10	0.05	128	89	738 MB	2H 17Min
With DAugtify	10	0.05	128	95	835 MB	2H 20Min
MNIST & LeNet5						
Baseline	10	0.05	128	98	220 MB	1Min 27S
With DAugtify	10	0.05	128	98	220 MB	2Min 7S

Table 36: Non-functional test results

Though the computational resource usage and training time with the DAugtify system were little higher than the baseline training process, the improvement in the accuracy is significant enough to justify the additional training time. Therefore, it can be concluded that proposed system was able to archive the goal of reducing computational resource consumption while holding the CNN model accuracies. Necessary evidence for performance testing and other non-functional testing results are placed in **Appendix H**.

8.8.2 Usability Testing

A group of evaluators with no prior knowledge in the DA domain were contacted to assess the usability of the application. Refer to Chapter 9 for additional details.

8.8.3 Maintainability Testing

The codebase of the proposed system has received an A grade after being assessed for code quality using the CodeFactor.io (https://www.codefactor.io/) tool.

Figure 36: Code Quality Analysis Results

8.9 Limitations of Testing Process

Limited resources: The end goal of the proposed system is to reduce the high computational resource consumption required for the AutoDA task. Although the proposed system achieves state-of-the-art performance on the aforementioned test cases, all experiments were limited to the author's MacBook Pro device due to the unavailability of computational resources.

Time constraint: As stated in Chapter 2, the majority of existing works used the ImageNet dataset along with various WideResNet CNN architectures to test and benchmark their AutoDA system. However, this process requires a lot of time for training without high-end GPU resources. Due to time constraints, the author was unable to test and benchmark the proposed AutoDA system on the aforementioned ImageNet dataset.

8.10 Chapter Summary

In this chapter, the testing process and results of the proposed AutoDA system are thoroughly discussed. The author evaluated the proposed AutoDA system using both qualitative and quantitative methods based on the pre-defined testing goals and criteria. Furthermore, this chapter presented a comparative analysis of the proposed AutoDA system with the other solutions available. To conduct functional testing, the black-box testing approach was employed, while different codes and analysis tools were used to perform non-functional testing. Finally, the author provides a summary of the limitations of the testing process.

CHAPTER 9: EVALUATION

9.1 Chapter Overview

After successfully implementing and optimizing the prototype for optimal performance through a vast number of testing combinations, the focus shifted to assessing the research outcomes. This chapter centers on the evaluation aspect of the project, which will encompass self-evaluation and evaluations from industry and domain experts.

9.2 Evaluation Methodology & Approach

To gather feedback and potential enhancements from domain and technical experts regarding the project, a concise document containing the problem, proposed solution, project significance, initial test outcomes, and a brief demonstration video were distributed. The document that was shared can be found via:
https://drive.google.com/file/d/1cWRwVaobErq_1z5PifgHQBlHZbVmqDg3/view?usp=sharing

Due to the limited number of experts available in the AutoDA domain, a **qualitative approach** was primarily adopted to evaluate the project. Nevertheless, it is important to emphasize that domain and technical experts are part of target audience of the project. Therefore, interviews were the main approach for evaluations, using open-ended questions. In addition, the UI/UX of the project was assessed using a **quantitative approach**.

9.3 Evaluation Criteria

The thematic analysis of the qualitative evaluation was based on the following criteria:

Criteria	Objective
Problem, background and problem novelty	To evaluate whether the problem is significant enough for a solution to add value.
The scope, depth, and complexity	To confirm that the project has an adequate scope, depth, and complexity for a bachelor's research.
Development decisions of the proposed solution and solution novelty	To evaluate the theorical and technical decisions made during the development of the proposed solution architecture and to validate the uniqueness of the proposed solution.

Quantitative analysis of results	To verify whether the system performs well enough to make a significant academic contribution and have commercial value.
Limitations and possible improvements	To reveal potential areas for improvement that could be addressed in future work.

Table 37: Evaluation Criteria

9.4 Self-Evaluation

Criteria	Evaluation
Problem, background and problem novelty	Computer vision has become a widely discussed topic in modern world. However, one major obstacle in this field is the lack of sufficient data, which is typically addressed using DA techniques. Currently, the selection of optimal DA techniques is performed manually, but automating this process has the potential to save valuable time and computational resources, benefiting the entire computer vision community. As a result, AutoDA models have a high potential to become an important component of standard computer vision model training pipelines in the future.
The scope, depth, and complexity	As stated in the problem and background evaluation, this research is focused on enhancing the standard CNN model training pipeline, which has the potential to benefit the entire computer vision community. As a result, the scope, depth, and intricacy of the research are substantial enough to qualify as an undergraduate research project.
Development decisions of the proposed solution and solution novelty	The development process of proposed system involved iteration, whereby the design, development, and improvement phases were repeated multiple times to arrive at the best-suited design for the given scenario. This ultimately led to the discovery of innovative methods to perform AutoDA task.
Quantitative analysis of results	The test results indicate that the proposed system achieved state-of-the-art performance. Nevertheless, to publish the findings in a top-tier journal or conference, additional experiments must be conducted using various datasets, CNNs, and hardware configurations.

Limitations and possible improvements	Addressing the class imbalance issue is a critical aspect that can be improved through DA techniques. However, due to time constraints during the initial phase, the proposed system will not address this issue. Therefore, future improvements will involve automatic resolution of the class imbalance issue and conducting additional test cases.

Table 38: Self-evaluation (Self-composed)

9.5 Selection of Evaluators

Image DA is a sub-topic within the computer vision field, making it challenging to find experts who have mastered both domain and technical knowledge of image DA. Hence, the primary evaluators chosen were researchers/engineers with prior experience in computer vision and image DA. The selected evaluators for the core research component are shown in the below table and the selected evaluators for the UI/UX of the product were listed in **Appendix I**. It is important to mention that the names and affiliations were listed only after obtaining explicit consent from the evaluator.

ID	Name	Position	Affiliation
EV1	Confidential	Professor of Software Engineering	Confidential
EV2	Mr. Manik Fernando	Senior Machine Learning Engineer	Surge Global
EV3	Mr. Varatharajah Vaseekaran	Machine Learning Engineer	Iron One Technologies
EV4	Mr. Nisal Perera	Machine Learning Engineer	Veracity AI
EV5	Ms. Areefa Thassim	Analyst - Process Innovation	MillenniumIT ESP
EV6	Mr. Dilum De Silva	iOS Engineer	Hectre Group New Zealand
EV7	Ms. Hansamali Wijayatilake	Senior Quality Engineer	Circles Life Sri Lanka
EV8	Mr. Wishwa Prabodha	Senior Software Engineer	Circles Life Sri Lanka / University of Kelaniya

EV9	Mr. W.D Damitha Dayananda	Senior Software Engineer	Circles Life Sri Lanka / University of Moratuwa
EV10	Mr. Deepak Ranjan	Senior Software Engineer	Circles.Life Singapore
EV11	Mr. Hasal Fernando	Data Engineer	Circles Life Sri Lanka
EV12	Mr. Visal Rajapakse	Software Engineer / Part Time Lecturer	Circles Life Sri Lanka / Informatics Institute of Technology
EV13	Mr. K.T. Yasas Mahima	PhD Candidate	University of New South Wales (UNSW)
EV14	Mr. Dinuka Piyadigama	Software Engineering Masters Student	University of Melbourne
EV15	Mr. Shian Fernando	Software Engineer	Limark Technologies
EV16	Mr. Hamdaan Mohideen	Data Engineer Intern / NBQSA 2022 Gold Award - Tertiary Category, Technology / APICTA 2022 Merit Award - Student Category, Tertiary Applications, SEA NM Regional Semifinals - Education Category	Circles Life Sri Lanka / Informatics Institute of Technology
EV17	Mr. Malitha Randeniya	Data Engineer Intern	Circles Life Sri Lanka

Table 39: Selected evaluators for the core research component

9.6 Evaluation Results from Experts

9.6.1 Qualitative Evaluation Result Analysis

After collecting the expert opinions, a thematic analysis was performed on their feedback to identify codes that satisfied the established evaluation criteria. The supporting evidence for each criterion has been included in the **Appendix J**.

Themes	Codes
Research problem and gap	Addressed problem, timely and essential, existing issues, impact

Novelty	Unique, great, light-weight system
Scope, depth & complexity	Adequate challenges, undergraduate level, complex, overcome
Architecture	Neural networks, efficient solution, approach, resources
Results	Computational resources, performance, accuracy
Suggestions	More test results, review & research paper

Table 40: Themes identified by conducting thematic analysis on expert feedback (Self-composed)

Criteria	**Theme**	**Summary of Feedback**
Problem, background and problem novelty	Research problem and gap, Novelty, Scope, depth & complexity	The initial impression of the project has been outstanding and addresses a timely and essential research problem in the computer vision domain. Moreover, the problem novelty is also evident, mainly in enabling automatic image data augmentation with limited computing resources and high accuracy levels. One evaluator stated that while most undergraduate tends to go for just predictions on algorithms for random dataset but here this student tries to contribute to a research area that can use to improve any ML model. He highly appreciated exploring this kind of rare research area for undergraduate project.
The scope, depth, and complexity	Research problem and gap, Novelty, Scope, depth & complexity	Despite selecting a challenging scope in the wide domain area of data augmentation, the student managed to deliver a pretty decent and well-thought-out project that brings value to the table. Moreover, a crucial aspect of the project was making the complex system resource-friendly, which was a paramount task. One evaluator stated that when considering theoretical aspects of the redefinition of traditional search space and developing an end-to-end tool like this which will benefit for entire computer vision community, making it worthy of being called '**great undergraduate research**'.

Development decisions of the proposed solution and solution novelty.	Novelty, Scope, depth & complexity, Architecture	The proposed solution approach is well-designed from the start to finish, and the author have provided a detailed description of the approach and have justified design decisions through experimental evaluations. Therefore, the novelty of the solution is impressive, and the results also are promising. Further, the author has good understanding about the solution, hence trying to find a balance between the pros like simplicity, efficiency, robustness VS challenges like limitations, overfitting issues and lack of flexibility. Moreover, one evaluator stated that the proposed solution novelty has contributed to the Body of Knowledge (BoK) up to the expected standards.
Quantitative analysis of results	Architecture, Results	The system demonstrates the potential importance of this research in real-world data augmentation scenarios by achieving improved performance compared to other AutoDA techniques, while using significantly fewer computational resources and being simpler to implement. Therefore, the provided results are solid evidence that prove proposed solution can make significant contribution to the domain. One evaluator stated that when thinking from a customer perspective of a DA system, he would like to manage everything with little knowledge and the provided system achieves this. The evaluator also stated that the accuracy of the proposed system is higher than any other system that exists, hence further enhancing the merit of the proposed solution.
Limitations and possible improvements	Suggestions	Suggest presenting more experimental results and benchmark with more solutions as it will support the novel solution and justify it further. Moreover, product

		can be enhanced to measure the effectiveness of other computer vision tasks such as image segmentation and object detection.

Table 41: Summary of experts feedback of the project

9.6.2 Quantitative Evaluation Result Analysis

As mentioned, quantitative assessment was carried out to evaluate the UI/UX of the project and the following presents a qualitative analysis of the responses provided by UI/UX experts.

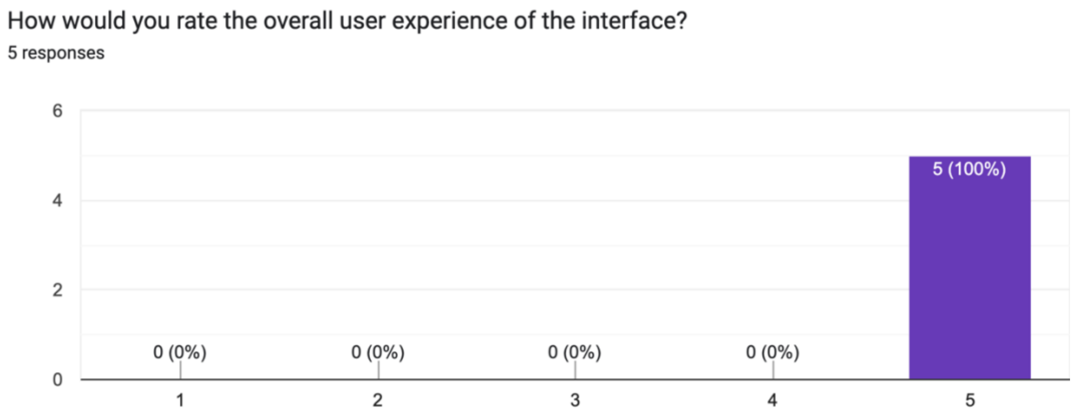


Figure 37: Quantitative analysis of UI/UX of the project

9.7 Evaluation of Functional Requirements

IM: Implemented | **NI**: Not Implemented

FR ID	Requirement	Priority Level	Status
FR1	Users must be able to select their preferred dataset.	M	IM
FR2	Users must be able to select the preferred image classification model.	M	IM
FR3	Users must be able to select preferred DA techniques.	M	IM
FR4	Based on the selected dataset, image classification model and the DA techniques, the system must be able to find the optimal combinations of DA techniques with their magnitudes.	M	IM
FR5	System must be able to train the given image classification model with auto selected DA policies.	M	IM
FR6	System must be able to display performance statics of trained image classification model.	M	IM

FR7	Users must be able to save the trained image classification model with augmented data.	M	IM
FR8	Users must be able to save the augmented data points.	M	IM
FR9	Users should have the ability to tune the important hyper-parameters of given image classification model through GUI.	S	IM
FR10	System should be able to identify and solve the class imbalance issue of the selected dataset if present.	S	NI
FR11	System should be able to show the auto selected DA techniques with their magnitude.	S	IM
FR12	System could have ability to display previously trained image classification model history and the statics through GUI.	C	IM
FR13	Users could have ability to select and retrain a previously trained image classification model via GUI.	C	NI
Functional requirement completion rate = (12/13) * 100% = 92**.3%**			

Table 42: Evaluation of functional requirements

9.8 Evaluation of Non-Functional Requirements

IM: Implemented | **NI**: Not Implemented

NFR ID	Description	Priority Level	Status
NFR1	This project aims to decrease the computational resources needed for AutoDA tasks and enhance the broad applicability of AutoDA techniques, including for users who do not have access to high-performance GPU resources.	M	IM
NFR2	while improving the efficiency of AutoDA tasks, it is important to maintain a high level of accuracy. Because unnecessary selection of DA operations can negatively impact the performance of the target CNN model.	M	IM
NFR3	The ultimate goal of the AutoDA system is to tackle the difficulties in traditional DA. Hence users must be able to perform DA tasks even without expertise in the DA domain.	M	IM
NFR4	The system should be able to perform AutoDA techniques in larger datasets and handle the workload smoothly.	S	IM

NFR5	The software development best practices should be followed while building the product.	C	IM
NFR6	To manage the significant resource requirements of training a sophisticated image classification models, the application should be deployed on a cloud server.	C	NI
Non-Functional requirement completion rate = (5/6) * 100% = **83.3%**			

Table 43: Evaluation of non-functional requirements

9.9 Limitations of Evaluation

The main obstacle in the evaluation process is to showcase the complete training process of the CNN using AutoDA techniques. This is due to the fact that the training duration depends on the complexity of the CNN architecture. However, to overcome this challenge, prior test outcomes and a video recording of the system have been provided. The shared testing outcomes document can be found via this link:
https://docs.google.com/spreadsheets/d/1X7nXCMpr8WLckpUvf7LeYf6Qenh8Trdd6xx7vI0ojxU/edit?usp=sharing

9.10 Chapter Summary

This chapter focused on evaluating the research conducted, discussing the chosen approaches and the rationale behind them. The evaluation criteria were established before conducting the evaluations. The opinions of evaluators were analyzed thematically and presented based on the predefined criteria. Additionally, both functional and non-functional requirements were assessed.

CHAPTER 10: CONCLUSION

10.1 Chapter Overview

This chapter covers the final remarks on the research conducted. It includes a discussion on the aims, objectives, and learning outcomes of the project, and the obstacles that were encountered during its implementation. The chapter also covers the utilization of the skills and knowledge obtained over the four-year period and the new skills acquired through the research. Additionally, it explores any deviations from the initial plan, limitations, potential areas for future work, and the contribution to the existing body of knowledge.

10.2 Achievement of Research Aim & Objectives

10.2.1 Achievement of Aim

The aim of this research is to design, develop and evaluate a system that automates the manual process of designing and fine-tuning optimal image data augmentation schemes for computer vision tasks with reduced human intervention.

The research successfully accomplished aim of creating a new AutoDA architecture that can perform AutoDA tasks while utilizing fewer computational resources and with reduced human intervention. The Testing and Evaluation chapters provide evidence to support the effectiveness of the proposed solution.

10.2.2 Achievement of Objectives

Research Objective	Learning Outcomes	Status
Problem Domain: Perform an in-depth review of the necessary areas for the research.	LO1, LO4, LO5	Completed
Literature Review: Comprehensive survey and critical analysis of the existing literature and research on a problem domain.	LO1, LO4, LO5	Completed
Requirement Analysis: Collect user requirements and analyze them in a critical manner.	LO3, LO6, LO7	Completed
Design: Design a suitable architecture for the proposed AutoDA system.	LO2, LO5, LO7	Completed

Development: Development of proposed AutoDA architecture based to the designed architecture.	LO1, LO5, LO6, LO7	Completed
Testing & Evaluation: Conduct testing and evaluation of the proposed AutoDA system to validate the outcomes.	LO5, LO6, LO7, LO8	Completed

Table 44: Achievement of Research Objectives

10.3 Utilization of Knowledge from the Degree Program

Module(s)	Utilized Knowledge
Programming Principle I, II & Object-Oriented Programming	Through these modules, essential programming skills were gained along with the software design and testing
Web Design & Development	The knowledge gained from this module was utilized during the client tier development of the proposed system.
Algorithms: Theory Design and Implementation	The discussion of data structures and search algorithms in this module particularly useful while designing the proposed system.
Software Development Group Project.	This module provided the initial inspiration to work on research by guiding the process from research problem recognition to design, development, and testing of a proposed solution.

Table 45: Utilization of Knowledge of Degree Program

10.4 Use of Existing Skills

- **Industry experience as Software Engineer** - The experience that the author gained during the internship at Circles.Life proved useful in completing the project within the deadlines.
- **ML & DL** - Before the start of the final year, the author learned basic concepts of ML and DL through self-study by watching tutorial videos on platforms such as YouTube and LinkedIn Learning.

10.5 Use of New Skills

- **Data augmentation** - Initially, the author lacked any prior experience in the field of data augmentation. Hence the author was able to expand his knowledge of data augmentation which is a crucial component of the training process for computer vision models.

- **Neural network architecture design & workflow** - The proposed system was implemented as layer of convolutional neural networks (CNNs). Hence the author gained in-depth understanding of neural network architecture design and the workflow.

10.6 Achievements of Learning Outcomes

What has been learned	LOs
The author has gained knowledge about the research project process, which involves adhering to research methodologies and enhancing decision-making skills by analyzing the gathered data.	LO1, LO2, LO4
The author conducted a comprehensive literature review on the DA domain and acquired knowledge on various DA technologies.	LO4, LO5
The author acquired an understanding of the formal data gathering methods while keeping in mind the importance of social, legal, and ethical considerations.	LO3, LO6 LO4
Since the author had to document the research at every stage, it helped to enhance the academic writing abilities.	LO8
During the prototyping phase of the research, the process involved trial and error based experimentations. The author developed problem-solving skills through identifying issues, seeking solutions and eventually obtaining the desired outcomes.	LO5, LO7

Table 46: Achievement of Learning Outcomes

10.7 Problems and Challenges Faced

Problem/Challenge	Solution
Vast project scope.	The computer vision field encompasses image classification, segmentation, and object detection. Additionally, data augmentation techniques in computer vision involve both basic image manipulation and deep learning-based methods. However, due to the time limitations, automating all data augmentation techniques for all computer vision tasks is not feasible. Therefore, the project scope has been narrowed down to AutoDA for image classification tasks using basic image transformations.

Vast learning curve.	Studying all search and evaluation strategies to perform AutoDA tasks, as well as the effectiveness of different DA techniques on various CNNs, would require a significant amount of time. To mitigate that, the proposed system architecture was built by utilizing the literature findings from the last five years.
Long hours of testing time.	To evaluate the effectiveness of the proposed AutoDA system, it is essential to train various CNN architectures on different datasets. However, the training time may vary depending on the complexity of the CNN architecture. To address this issue, selected CNN architectures and datasets were used for testing purposes.
Resource constraints.	Most of the existing works for AutoDA tasks require high GPU resources, making it infeasible to implement them from scratch with limited resources. Therefore, pre-defined AutoDA works in PyTorch were used for benchmarking purposes.

Table 47: Problems and challenges faced.

10.8 Deviations

There were no significant deviations in either the project design or its implementation. However, during the requirement gathering phase, domain/experts suggested that ideal way to demonstrate the research would be through Python library, because training of given CNN model can take a significant amount of time depending on their architecture complexity. However, to fulfil the University of Westminster requirement for the final-year project a web application was implemented with limited features.

10.9 Limitations of the Research

- Although the computer vision domain encompasses several sub-domains, including image classification, image segmentation, and object detection, the research focused only on automated data augmentation within the image classification sub-domain due to time constraints.
- Despite several experiments demonstrating that the proposed system outperformed the state-of-the-art, it was only tested and evaluated on a limited number of datasets and CNN architectures due to time constraints.

- Similar to the above, all experiments were conducted and evaluated on an Apple M1 Pro 2021 MacBook. However further testing with various hardware configurations is necessary to ensure the effectiveness of proposed system.
- During the initial stage, the author has only considered automation in 14 DA techniques.
- The proposed system does not automatically handle the class imbalance problem.

10.10 Future Enhancements

- A significant future improvement would be to evaluate the effectiveness of the proposed system in other sub-domains of computer vision, such as image segmentation and object detection.
- Another future enhancement would be to assess the effectiveness of the proposed system across a wide range of datasets, CNN architectures, and different hardware configurations.
- Currently, including proposed AutoDA system all other existing works in the AutoDA domain have utilized basic image data augmentation techniques within their search space. Another significant future enhancement would be to integration of DL-based image data augmentation into the search space.
- Another potential direction for future research could be to address the issue of imbalanced datasets, which are common in real-world scenarios and can lead to model overfitting. This could involve exploring ways to automatically address class imbalance while performing AutoDA task.
- Frontend of the proposed tool needs to be updated with more options and scalable training process. This can be achieved by deploying the proposed system on high-performing cloud-based servers.

10.11 Achievement of the Contribution to Body of Knowledge

10.11.1Contribution to Computer Vision Domain

Modern CNN architectures can be easily overfit to training data due to insufficient dataset. DA is an effective technique to overcome this issue, and automating the process can save time and resources. However, existing AutoDA solutions have limitations, such as high computational resource usage and implementation complexity, which make them impractical for day-to-day DA tasks. The proposed solution bridges this gap, and its findings suggest that AutoDA models could become an essential part of standard CNN model training pipelines in the future.

10.11.2 Contribution AutoDA Domain

Theoretical and technical contribution can be summarized as follows:

- **Redefinition of traditional AutoDA search space** – Traditional AutoDA systems require a lot of time to perform AutoDA task due to the vast number of hyperparameters involved in the search space. However, in this study, the author redefines the AutoDA search space by limiting it to single hyperparameter, which significantly reduces the time required to perform AutoDA tasks.
- **Development of lightweight neural network to perform AutoDA task** – As mentioned, the search space of proposed system comprises of a single hyperparameter, which the author aims to optimize. The study demonstrates that instead of employing separate search and evaluation strategies, a straightforward neural network consisting of the hyperparameter is adequate for the optimal data augmentation policy search process.

By redefining the search space theoretically and utilizing a neural network with a single hyper-parameter to explore it, a new technical architecture for AutoDA has been created. This architecture leads to a more efficient and effective AutoDA process while using fewer computational resources.

10.12 Concluding Remarks

The author proposed a novel approach to perform AutoDA task named DAugtify that can be used to amplify the performance of image classification models. Thus, making important contributions to the fields of data augmentation and computer vision. The target audience for the project includes general ML researchers, data scientists, and students, and the author was fortunate to receive feedback from all these groups during the evaluation phase. The project received overwhelmingly positive feedback during the evaluation phase, with praise for the novel findings that are likely to benefit the entire computer vision community. The project was challenging, as it involved with advancements in standard CNN model training pipeline, but completing it has been a rewarding learning experience, covering all aspects of the research process and helping to overcome the challenges and limitations encountered to achieve the final goals.

REFERENCES

Cubuk, E.D., Zoph, B., Mané, D., Vasudevan, V., Le, Q.V., 2019a. AutoAugment: Learning Augmentation Strategies From Data, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 113–123. https://doi.org/10.1109/CVPR.2019.00020

Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V., 2019b. RandAugment: Practical automated data augmentation with a reduced search space.

Fawzi, A., Samulowitz, H., Turaga, D., Frossard, P., 2016. Adaptive data augmentation for image classification, in: 2016 IEEE International Conference on Image Processing (ICIP). Presented at the 2016 IEEE International Conference on Image Processing (ICIP), pp. 3688–3692. https://doi.org/10.1109/ICIP.2016.7533048

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Nets, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Hataya, R., Zdenek, J., Yoshizoe, K., Nakayama, H., 2020a. Faster AutoAugment: Learning Augmentation Strategies Using Backpropagation. Lecture Notes in Computer Science 12370, 1–16. https://doi.org/10.1007/978-3-030-58595-2_1

Hataya, R., Zdenek, J., Yoshizoe, K., Nakayama, H., 2020b. Meta Approach to Data Augmentation Optimization.

Ho, D., Liang, E., Stoica, I., Abbeel, P., Chen, X., 2019. Population Based Augmentation: Efficient Learning of Augmentation Policy Schedules.

Khalifa, N.E., Loey, M., Mirjalili, S., 2022. A comprehensive survey of recent trends in deep learning for digital images augmentation. Artif. Intell. Rev. 55, 2351–2377. https://doi.org/10.1007/s10462-021-10066-4

Lee, D., Park, H., Pham, T., Yoo, C.D., 2020. Learning Augmentation Network via Influence Functions, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10958–10967. https://doi.org/10.1109/CVPR42600.2020.01097

Li, C., Zhang, J., Hu, L., Zhao, H., Zhu, H., Shan, M., 2022. In-and-Out: a data augmentation technique for computer vision tasks. J. Electron. Imaging 31, 013023. https://doi.org/10.1117/1.JEI.31.1.013023

Li, P., Liu, X., Xie, X., 2021. Learning Sample-Specific Policies for Sequential Image Augmentation, in: Proceedings of the 29th ACM International Conference on Multimedia. Presented at the MM ’21: ACM Multimedia Conference, ACM, Virtual Event China, pp. 4491–4500. https://doi.org/10.1145/3474085.3475602

Li, Y., Hu, G., Wang, Y., Hospedales, T., Robertson, N.M., Yang, Y., 2020. DADA: Differentiable Automatic Data Augmentation.

Lim, S., Kim, I., Kim, T., Kim, C., Kim, S., 2019. Fast AutoAugment.

Lin, C., Guo, M., Li, C., Xin, Y., Wu, W., Lin, D., Ouyang, W., Yan, J., 2019. Online Hyper-parameter Learning for Auto-Augmentation Strategy.

LingChen, T.C., Khonsari, A., Lashkari, A., Nazari, M., Sambee, J.S., Nascimento, M., 2020. UniformAugment: A Search-free Probabilistic Data Augmentation Approach. ArXiv.

Liu, H., Simonyan, K., Yang, Y., 2018. DARTS: Differentiable Architecture Search. ArXiv.

Müller, S.G., Hutter, F., 2021. TrivialAugment: Tuning-free Yet State-of-the-Art Data Augmentation.

Naghizadeh, A., Abavisani, M., Metaxas, D.N., 2020. Greedy AutoAugment. Pattern Recognit. Lett. 138, 624–630. https://doi.org/10.1016/j.patrec.2020.08.024

Nanni, L., Paci, M., Brahnam, S., Lumini, A., 2021. Comparison of Different Image Data Augmentation Approaches. J. Imaging 7, 254. https://doi.org/10.3390/jimaging7120254

Pham, H., Guan, M., Zoph, B., Le, Q.V., Dean, J., 2018. Efficient Neural Architecture Search via Parameter Sharing. undefined.

Saunders, M., Lewis, P., Thornhill, A., n.d. Research Methods for Business Students 10.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on Image Data Augmentation for Deep Learning. J. Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0

Summa, M.G., Bottou, L., Goldfarb, B., Murtagh, F., Pardoux, C., Touati, M. (Eds.), 2011. Large-Scale Machine Learning with Stochastic Gradient Descent Léon Bottou, in: Statistical Learning and Data Science. Chapman and Hall/CRC, pp. 33–42. https://doi.org/10.1201/b11429-6

Takase, T., Karakida, R., Asoh, H., 2020. Self-paced Data Augmentation for Training Neural Networks.

Terauchi, A., Mori, N., 2021. Evolutionary Approach for AutoAugment Using the Thermodynamical Genetic Algorithm. Proc. AAAI Conf. Artif. Intell. 35, 9851–9858. https://doi.org/10.1609/aaai.v35i11.17184

Wang, X., Yin, J., n.d. Relaxed Multivariate Bernoulli Distribution and Its Applications to Deep Generative Models.

Wei, L., Xiao, A., Xie, L., Zhang, X., Chen, X., Tian, Q., 2020. Circumventing Outliers of AutoAugment with Knowledge Distillation, in: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (Eds.), Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 608–625. https://doi.org/10.1007/978-3-030-58580-8_36

Yang, Z., Sinnott, R.O., Bailey, J., Ke, Q., 2022a. A Survey of Automated Data Augmentation Algorithms for Deep Learning-based Image Classication Tasks.

Yang, Z., Sinnott, R.O., Bailey, J., Ke, Q., 2022b. A Survey of Automated Data Augmentation Algorithms for Deep Learning-based Image Classification Tasks.

Zhang, R., Liang, Y., Somayajula, S.A., Xie, P., 2021. Improving Differentiable Architecture Search with a Generative Model. https://doi.org/10.48550/arXiv.2112.00171

Zhang, X., Wang, Q., Zhang, J., Zhong, Z., 2019. Adversarial AutoAugment.

Zoph, B., Le, Q.V., 2017. Neural Architecture Search with Reinforcement Learning. https://doi.org/10.48550/arXiv.1611.01578

Appendix A: Concept Map

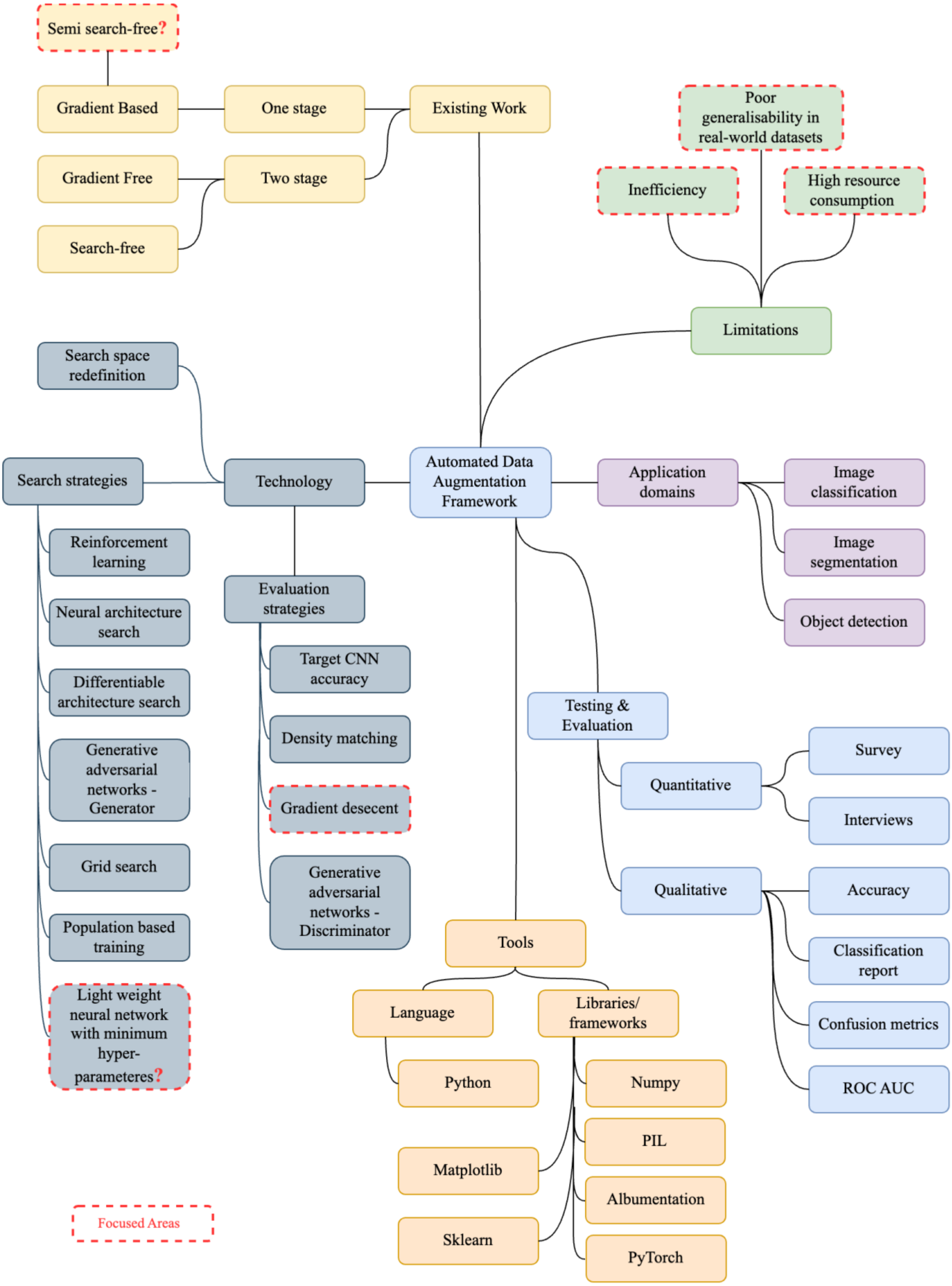


Figure 38: Concept map

Appendix B: Taxonomy of Data Augmentation

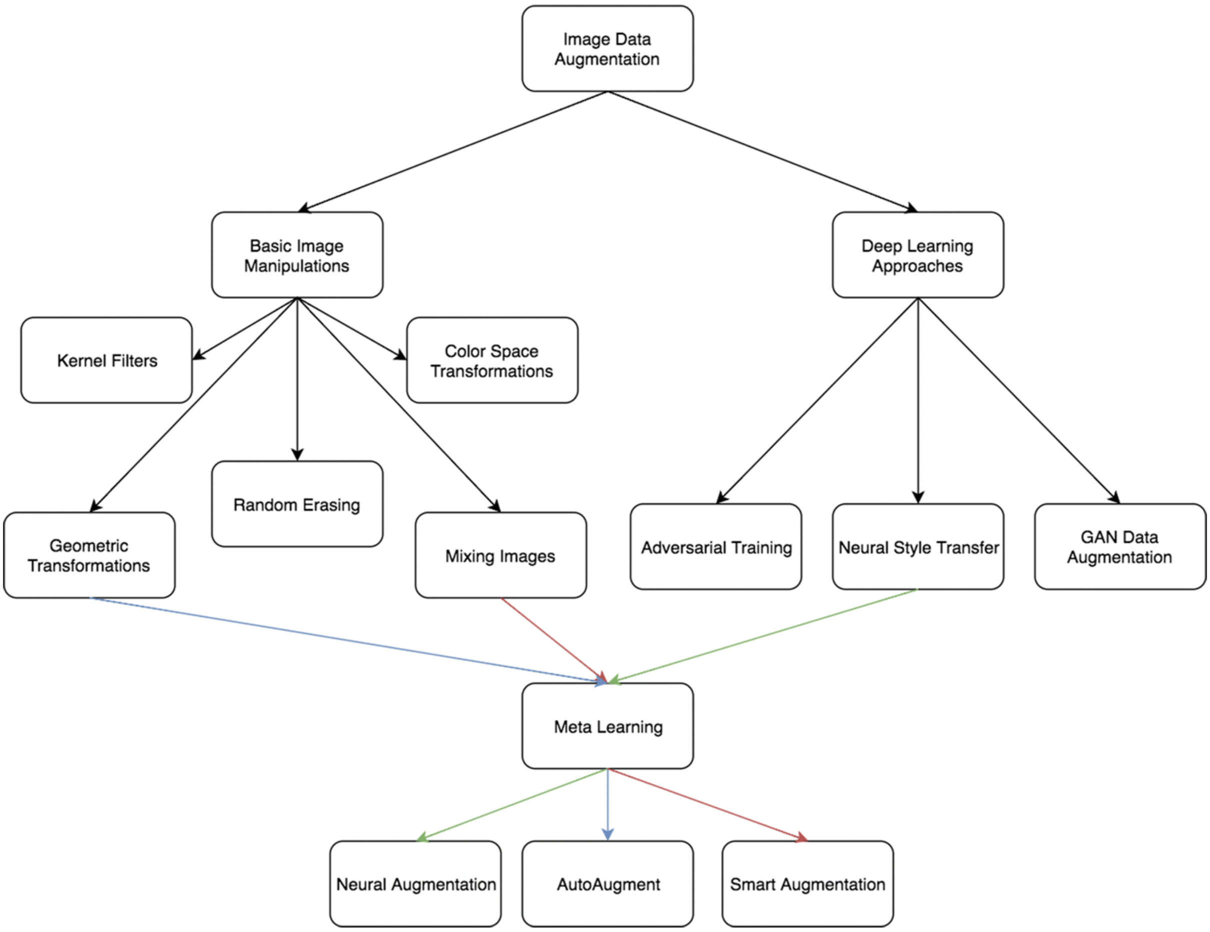


Figure 39: Taxonomy of data augmentation techniques (Shorten and Khoshgoftaar, 2019)

Appendix C: Gantt Chart

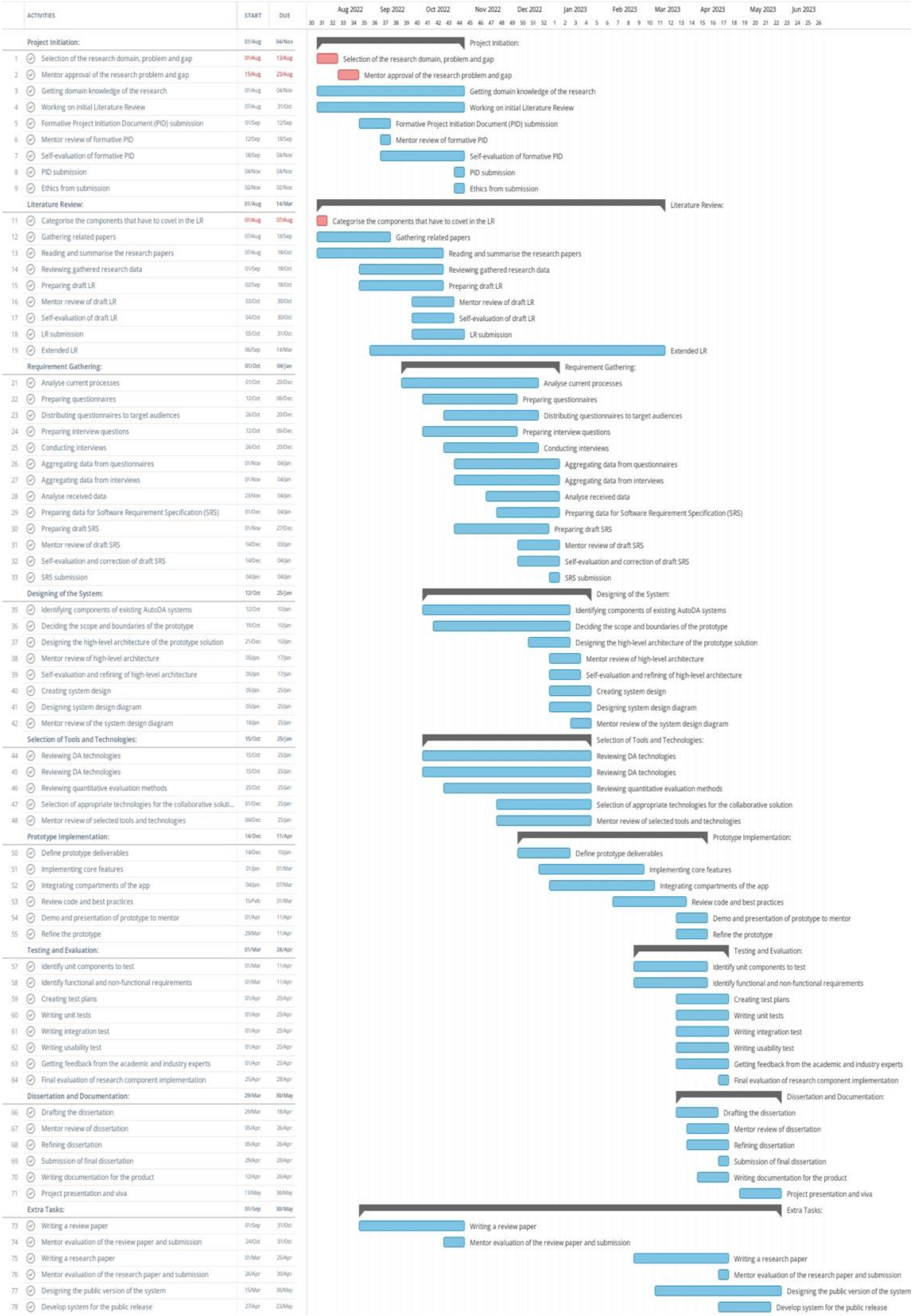


Figure 40: Gantt chart

Appendix D: Interview Findings

Question: Have you pushed back from implementing artificial intelligence solutions due to the dataset limitations?

Participant	Answer
1	Yes
2	Yes
3	Yes
4	Yes
5	Yes
6	Yes

Table 48: Requirements gathering, interview findings for question 1

Question: What are the most used image data augmentation techniques in your projects?

Participant	Answer
1	Geometric transformations (i.e., Flipping, Rotation, Translation), Color space transformations (i.e., Color filters), Kernel filters (i.e., Noise), Random erasing
2	Geometric transformations (i.e., Flipping, Rotation, Translation), Color space transformations (i.e., Color filters)
3	Geometric transformations (i.e., Flipping, Rotation, Translation), Color space transformations (i.e., Color filters), Kernel filters (i.e., Noise), Mixing images, Random erasing
4	Data duplication, Geometric transformations (i.e., Flipping, Rotation, Translation), Adversarial training, Generative adversarial network (GAN) data augmentation
5	Geometric transformations (i.e., Flipping, Rotation, Translation), Color space transformations (i.e., Color filters), Kernel filters (i.e., Noise), Neural style transfer, Adversarial training, Generative adversarial network (GAN) data augmentation
6	Geometric transformations (i.e., Flipping, Rotation, Translation), Color space transformations (i.e., Color filters), Kernel filters (i.e., Noise), Generative adversarial network (GAN) data augmentation

Table 49: Requirements gathering, interview findings for question 2

Question: What challenges have you faced while selecting the image data augmentation methods for your dataset and task?

Participant	Answer
1	Obstructing relevant information that could be necessary for generating desired output. Rotation may result in unrealistic information.
2	What specific augmentations would work the best for my model to perform better, Lots of time being spent in applying different augmentations, Performance issues when model being combined with augmentations.
3	Mostly deciding on which technique could eventually give me the best results in the quickest possible time. This also has been the reason to not explore Neural Network based approaches to Data Augmentation.
4	Time consuming to try multiple techniques
5	Selecting data augmentation techniques is an trial and error process which needs lot of experiments. Hence, project cost can be high since it requires lot of dev hours. When doing data augmentation need to consider the balance between model overfitting and underfitting. (No proper balance between that can occur model underfitting). Have to validate for data bias after DA, if necessary will have to implement some algorithms to correct data bias.
6	Time consuming to try multiple techniques. What specific augmentations would work the best for my model to perform better.

Table 50: Requirements gathering, interview findings for question 3

Question: What potential do you see in automated data augmentation techniques?

Participant	Answer
1	They are useful if applied with a subtle hand. Going too heavy handed could result in worse learning
2	It is an interesting concept and certainly it would help researchers to focus more on the domain they are working and worry less on experimenting on augmenting data for their model's use case, to improve its performance.
3	This should allow me to focus more on the final training aspect rather than spending too much time on carrying out augmentation manually. I am assuming the tool would perform multiple runs for different possible combinations and provide me with the best combination that I can use.

4	It is a good concept, if it can provide a ranking of the concepts used and a score, it would give insight on how to develop models.
5	Automated DA techniques will save time of the developer as it limits the number of experiments need to be performed to identify best performing DA technique for a given dataset. It will reduce project cost. It address knowledge limitations of the developer as implementing DA techniques needs technical and domain knowledge.
6	It is an interesting concept and certainly it would help researchers to focus more on the domain they are working and worry less on experimenting on augmenting data for their model's use case, to improve its performance.

Table 51: Requirements gathering, interview findings for question 4

Question: Have you used automated data augmentation techniques in your projects? If yes, what challenges have you faced while using automated data augmentation techniques?

Participant	Answer
1	Yes. Erasing or obscuring relevant information in the image.
2	No. With the projects I've worked with and gone on to complete, there was no specific need to use augmented data a lot and the domains were not relying much on augmented data as well. But, another factor is the less exposure to data augmentation and knowing it would be lot more of a bigger scope to get to know about all the specific automated data augmentations and to use them. Wherever I have used data augmentation, I have picked a few which would work and implemented those so the training data will be augmented as per those augmentations I needed.
3	Yes. One of the main challenges is deciding on how to augment and balance the dataset. This might seem very simple but at the initial stages, this was very confusing. If the tool can also identify the imbalance in the dataset and help decide on how to balance the dataset out using data augmentation, I think that can be very useful. Another challenge with data augmentation I have faced is when there is too little data where even regular augmentation techniques such as geometric transformations and image manipulation aren't very useful. In this case

	techniques using GANs can be used but they tend to be rather more complicated and time consuming.
4	No, I don't have much confidence in existing techniques to provide the best for my dataset
5	I didn't worked on a project where I need to use image data augmentation. But If I get a need I will definitely use them to save time and utilize the given time to do more experiments with the dataset to improve accuracy. Because DA is a critical data preprocessing task which will directly effect to the model prediction results. Easy integration, performance of the system (Time taking to give the best DA technique) will consider when choosing automated DA system.
6	No, since its not an easy to implement.

Table 52: Requirements gathering, interview findings for question 5

Question: What would you expect from such a system apart from the ability to select best-performing data augmentation techniques?

Participant	Answer
1	Ability to tune parameters
2	Based on the use case, based on the data, automatically suggest augmentations which the system thinks would work and another list of augmentations which can be further applied if user wants to try them out as well. Then, user gets to add or remove any augmentations they want.
3	The best combination of data augmentation techniques and to be able to understand how the different combinations affect the final output. I believe that the same combination would act different given a different kind of dataset. Also, the tool should be able to identify how to balance out the imbalanced dataset automatically and provide suggestions accordingly. Although, data augmentation is a type of technique used for oversampling I think it'd be quite useful if the tool could also suggest if under sampling could work better for the use case.
4	A Ranking of the techniques tried with a score
5	1. Rank the DA techniques for the given dataset and results comparison with the top 3 suggested techniques.

	2. Availability as a python package 3. Try with different data augmentation technique combinations
6	The best combination of data augmentation techniques and to be able to understand how the different combinations affect the final output.

Table 53: Requirements gathering, interview findings for question 6

Appendix E: Use Case Descriptions

Use case name	Input Dataset, Classification Model and Preferred Data Augmentation Techniques	
ID	UC2, UC3, UC4	
Description	User has to input their dataset, classification model and preferred DA techniques to perform AutoDA task.	
Participating actors	User	
Preconditions	None	
Extended use cases	None	
Included use cases	None	
Main flow	Actor	System
	1. Input dataset, classification model and preferred DA techniques.	2. Validate inputs.
Alternative flows	None	
Exceptional flows	Display error message.	
Post conditions	None	

Table 54: Use case description for Input Dataset, Classification Model and Preferred Data Augmentation Techniques

Use case name	Perform Data Augmentation with Selected Magnitude	
ID	UC8	
Description	Once the AutoDA module finds the optimal DA policy magnitude for the user's dataset and image classification model, the user is able to perform DA applying that magnitude.	
Participating actors	User	
Preconditions	AutoDA module must complete the execution and out the optimal DA policy magnitude user selected DA policies.	
Extended use cases	Export Augmented Dataset, Train Model with DA Policies.	
Included use cases	None	
Main flow	Actor	System
	1. Click on the 'Perform DA' button.	2. Start DA using the optimal magnitude found by the AutoDA module.

		3. Display the 'Download Augmented Dataset' and 'Train Image Classification Model with Augmented Dataset' buttons.
Alternative flows	None	
Exceptional flows	None	
Post conditions	Display the 'Download Augmented Dataset' and 'Train Image Classification Model with Augmented Dataset' buttons.	

Table 55: Use case description for Perform Data Augmentation with Selected Magnitude

Appendix F: Low Fidelity UI Design

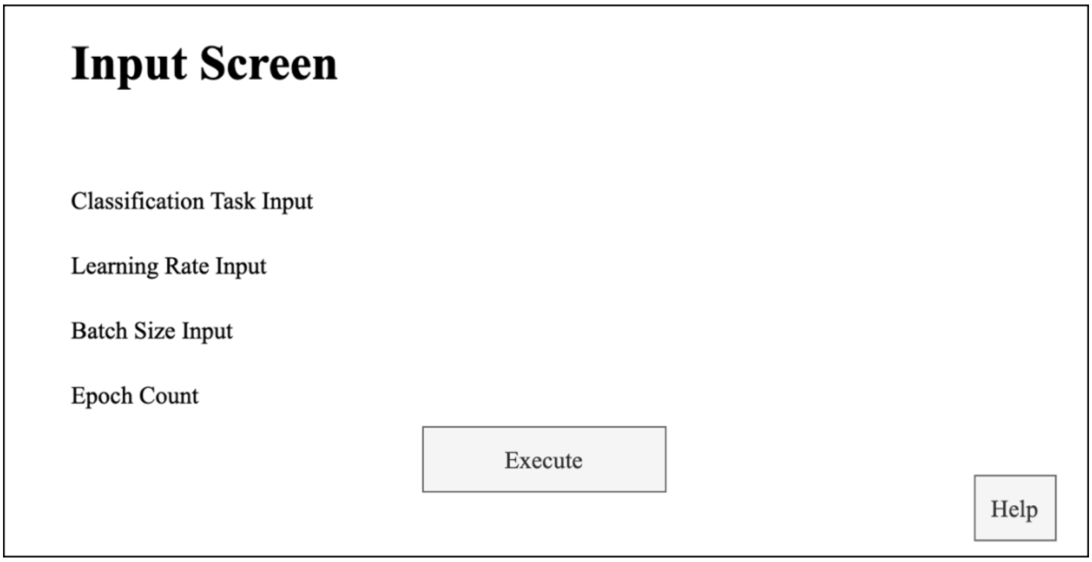


Figure 41: Input screen wireframe

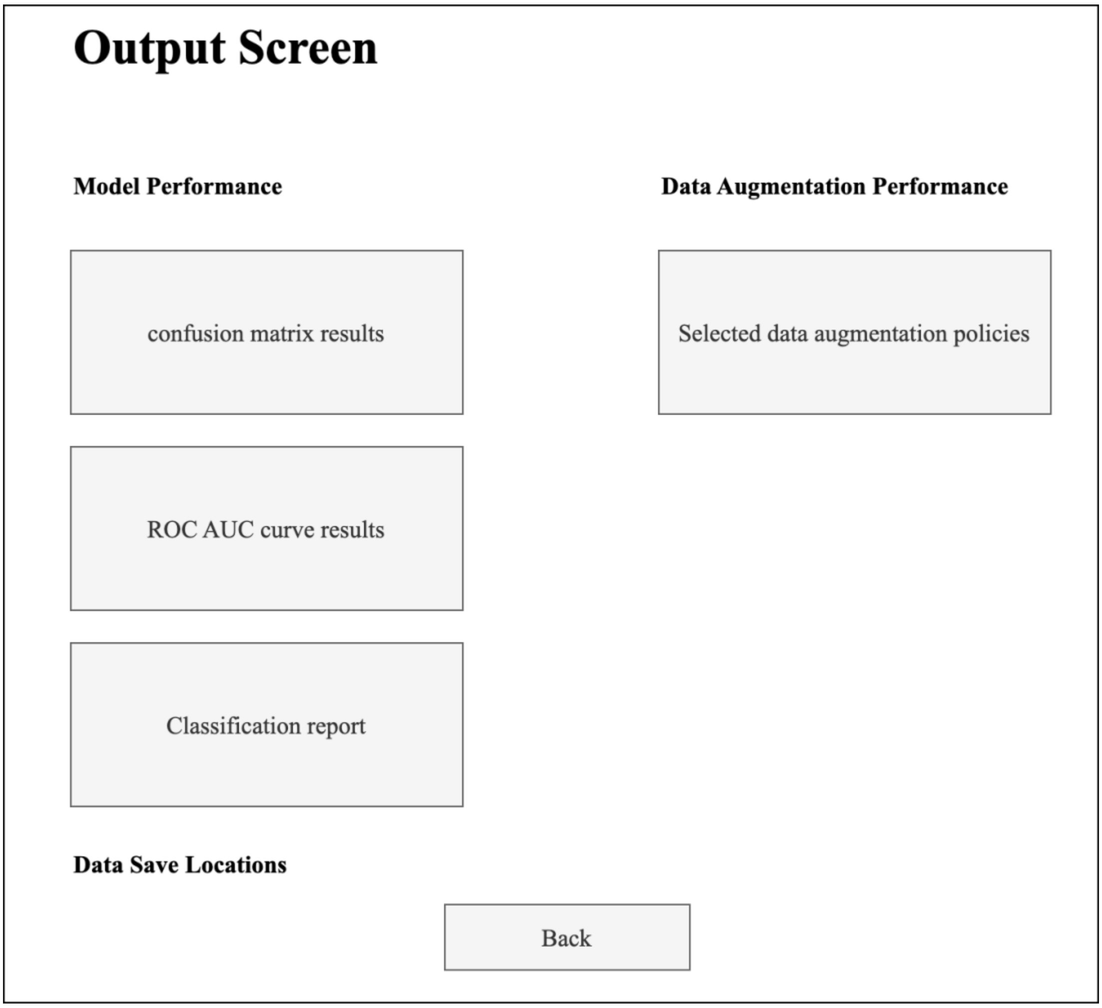


Figure 42: Output screen wireframe

Appendix G: High Fidelity UI Design

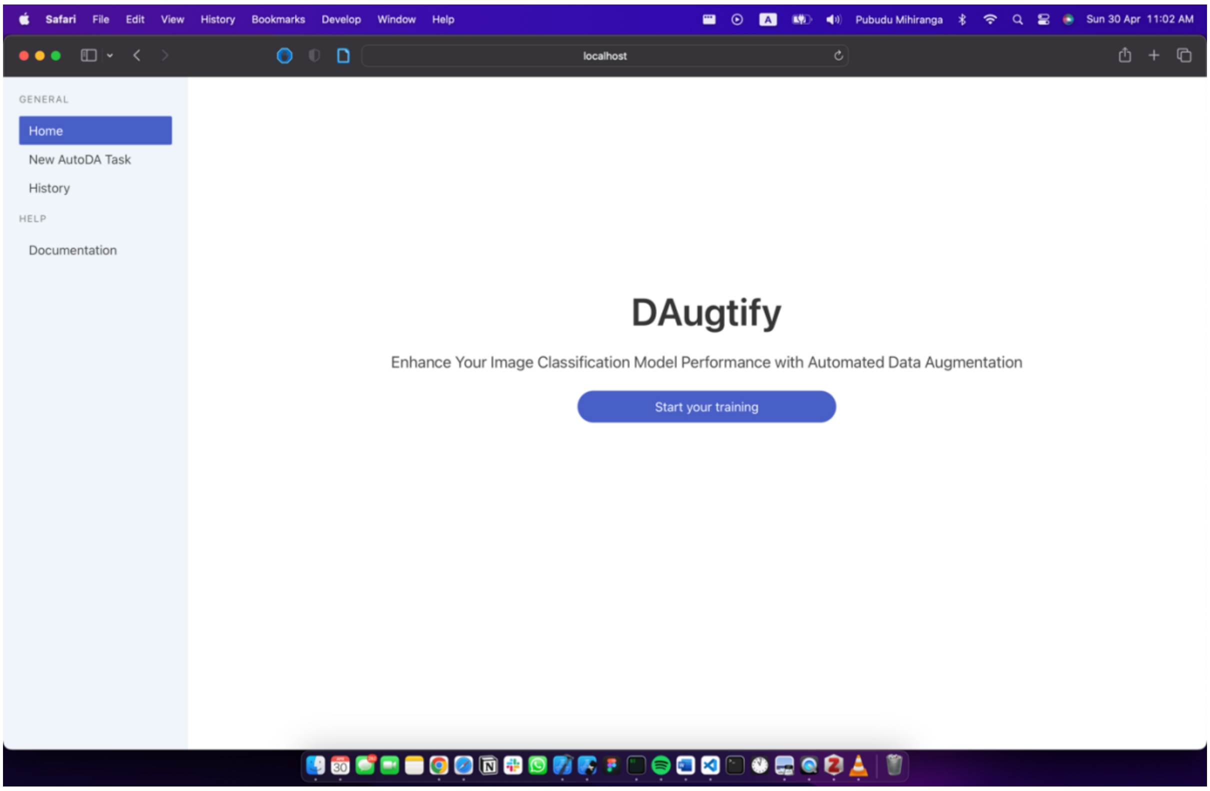


Figure 43: Home screen

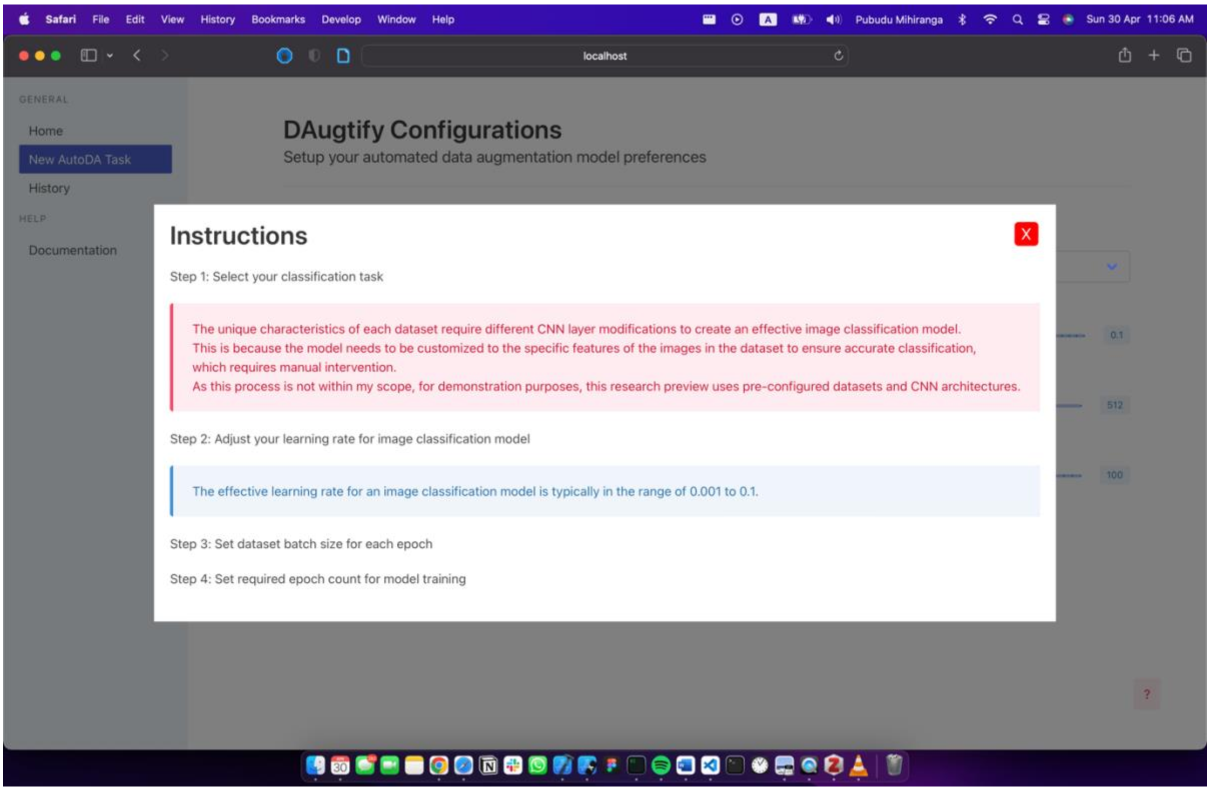


Figure 44: Input instructions

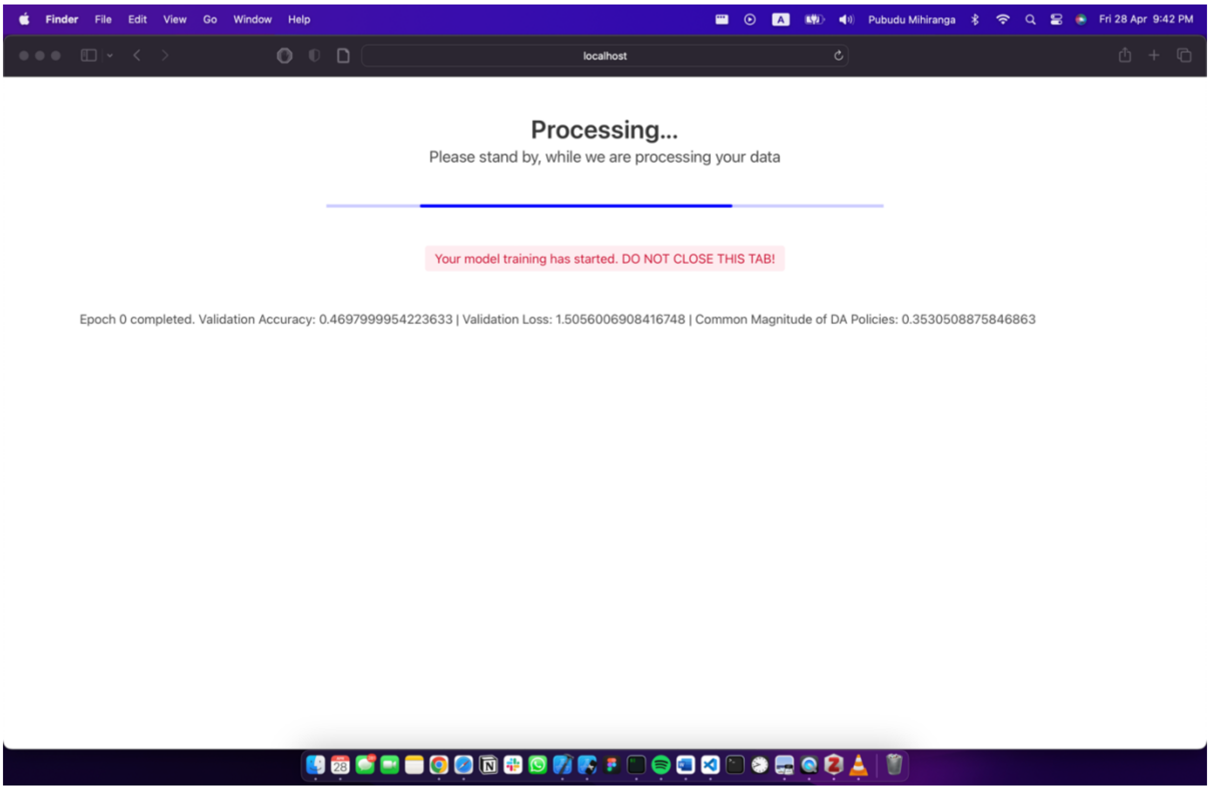


Figure 45: Processing screen

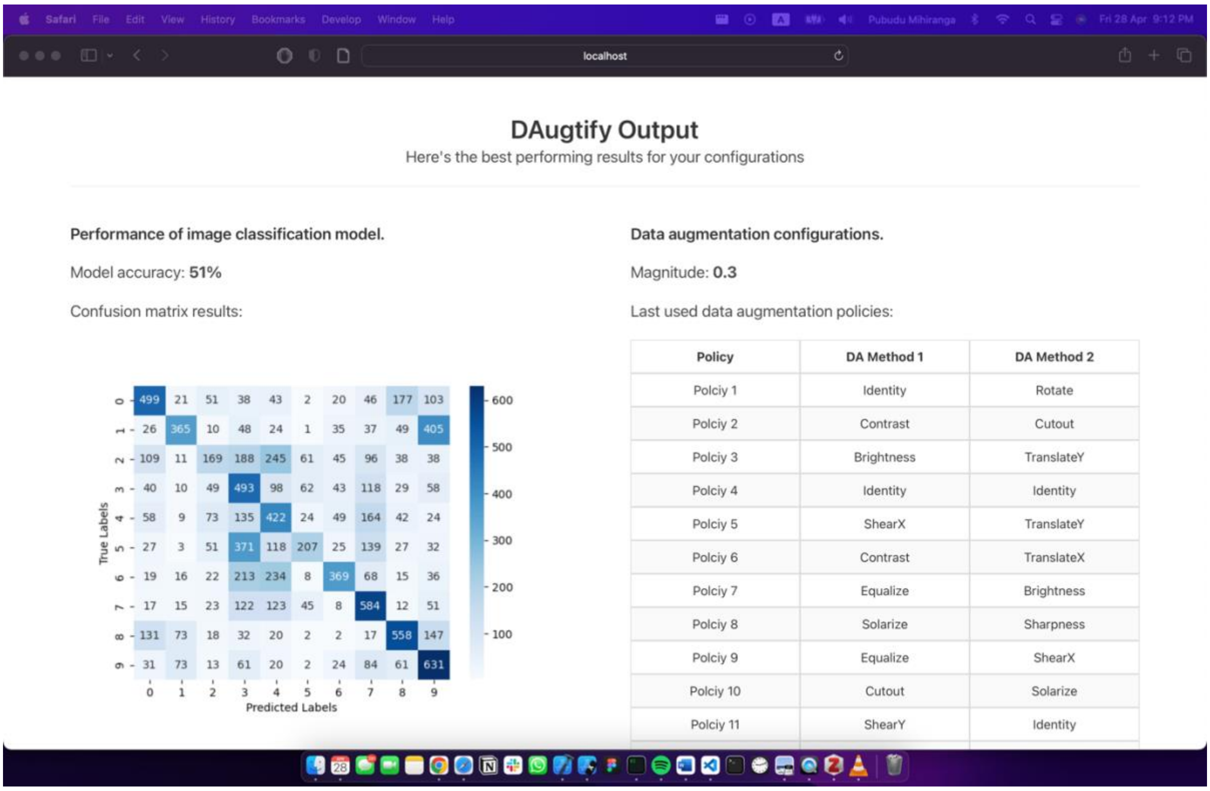


Figure 46: Output screen 1

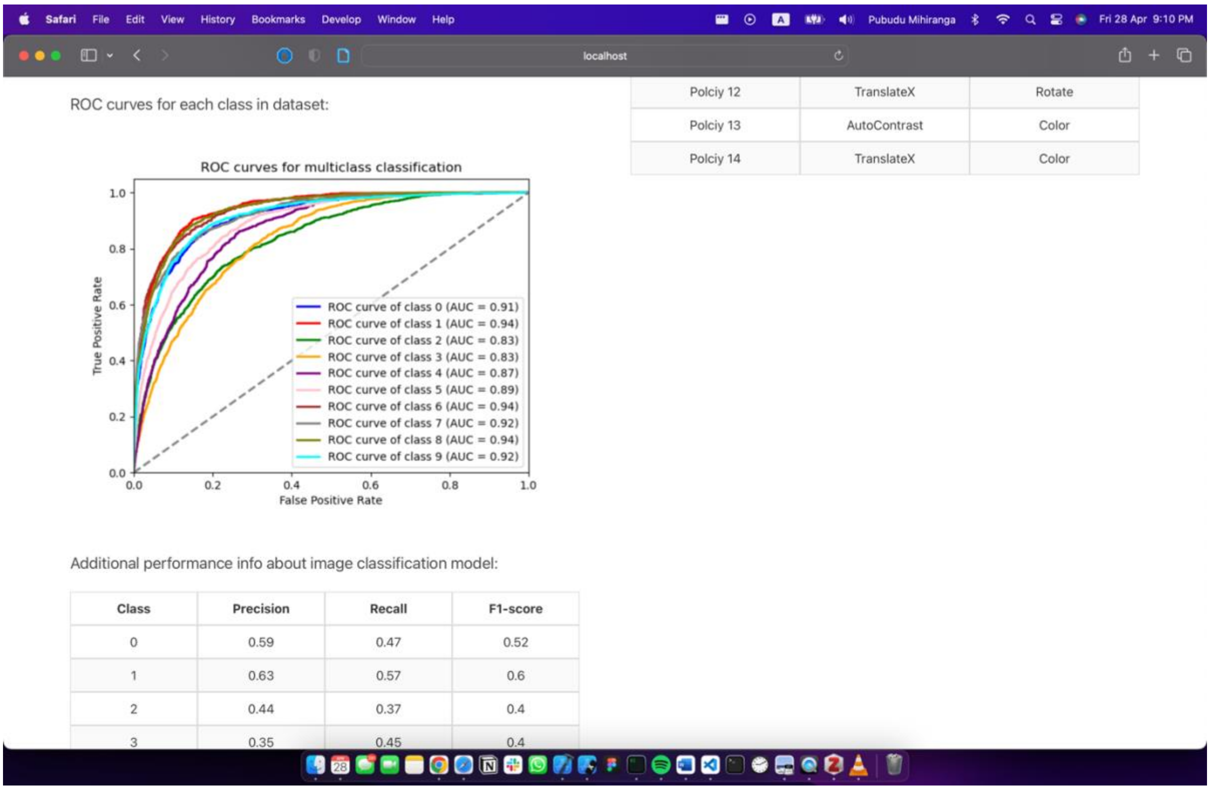


Figure 47: Output screen 2

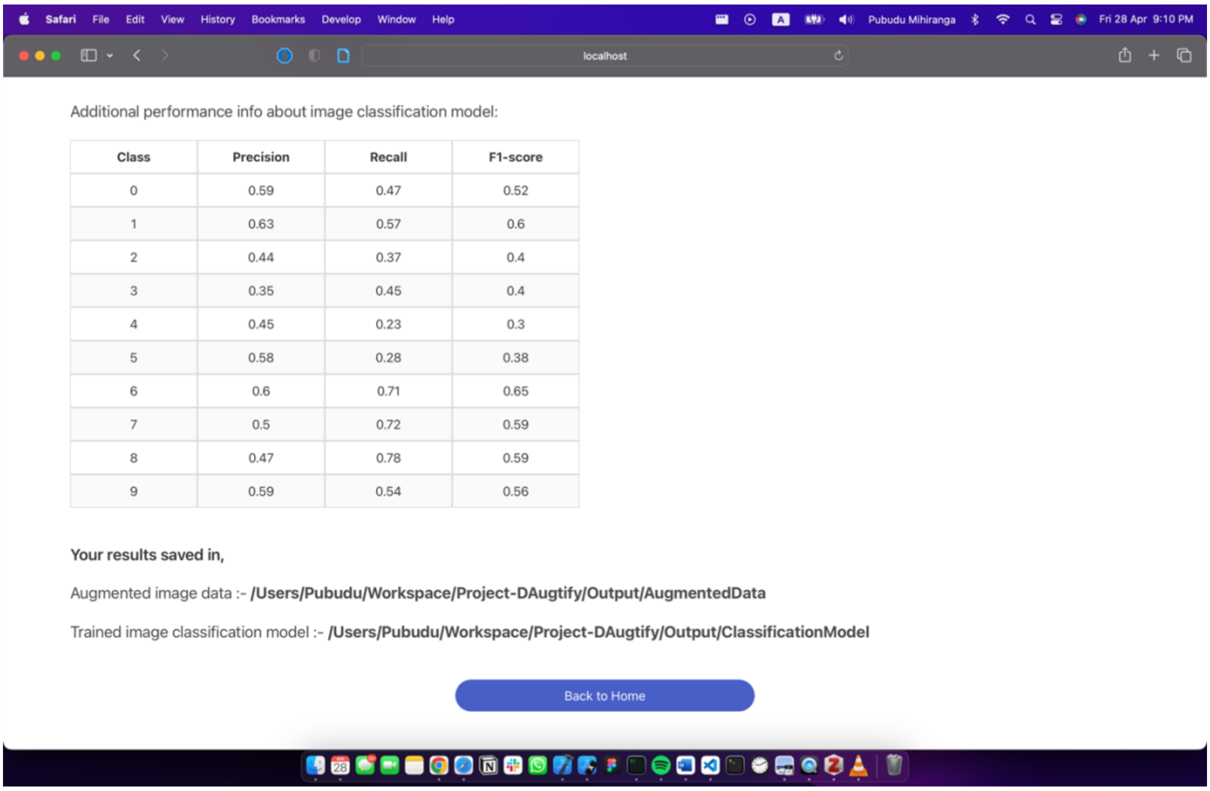


Figure 48: Output screen 3

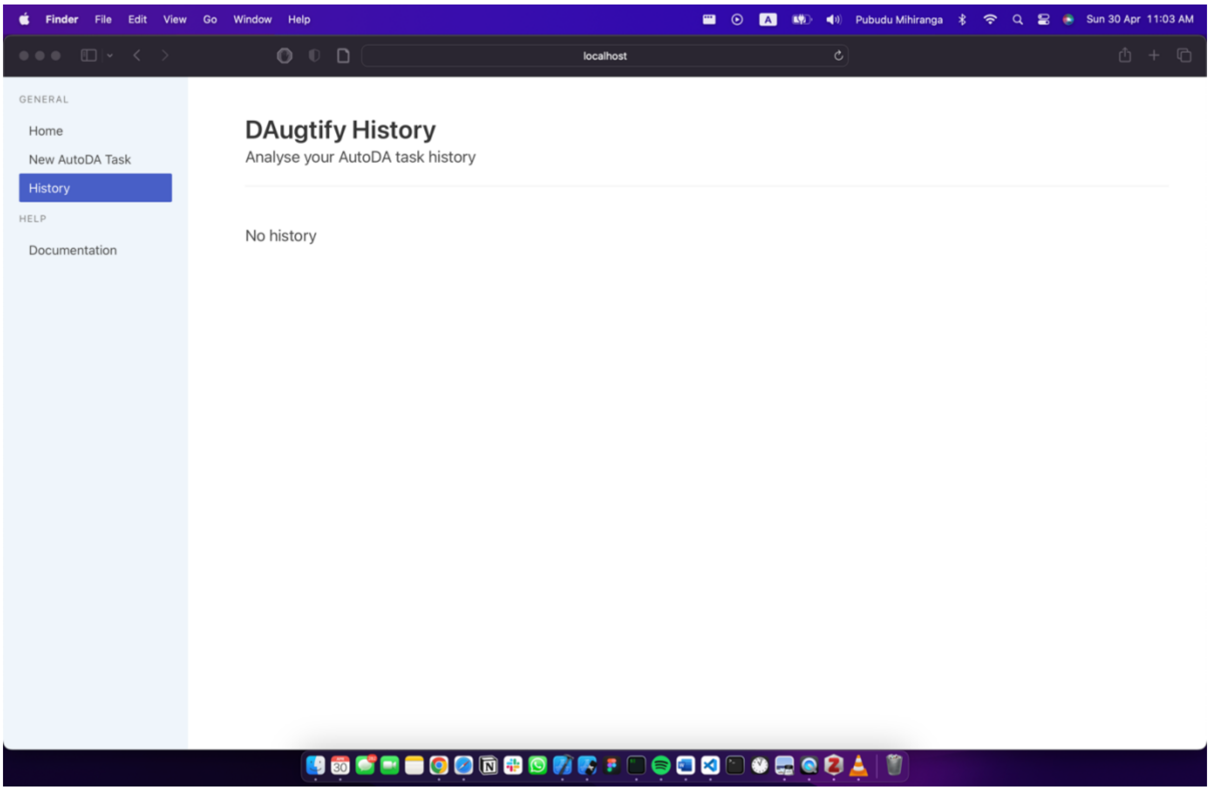


Figure 49: History screen

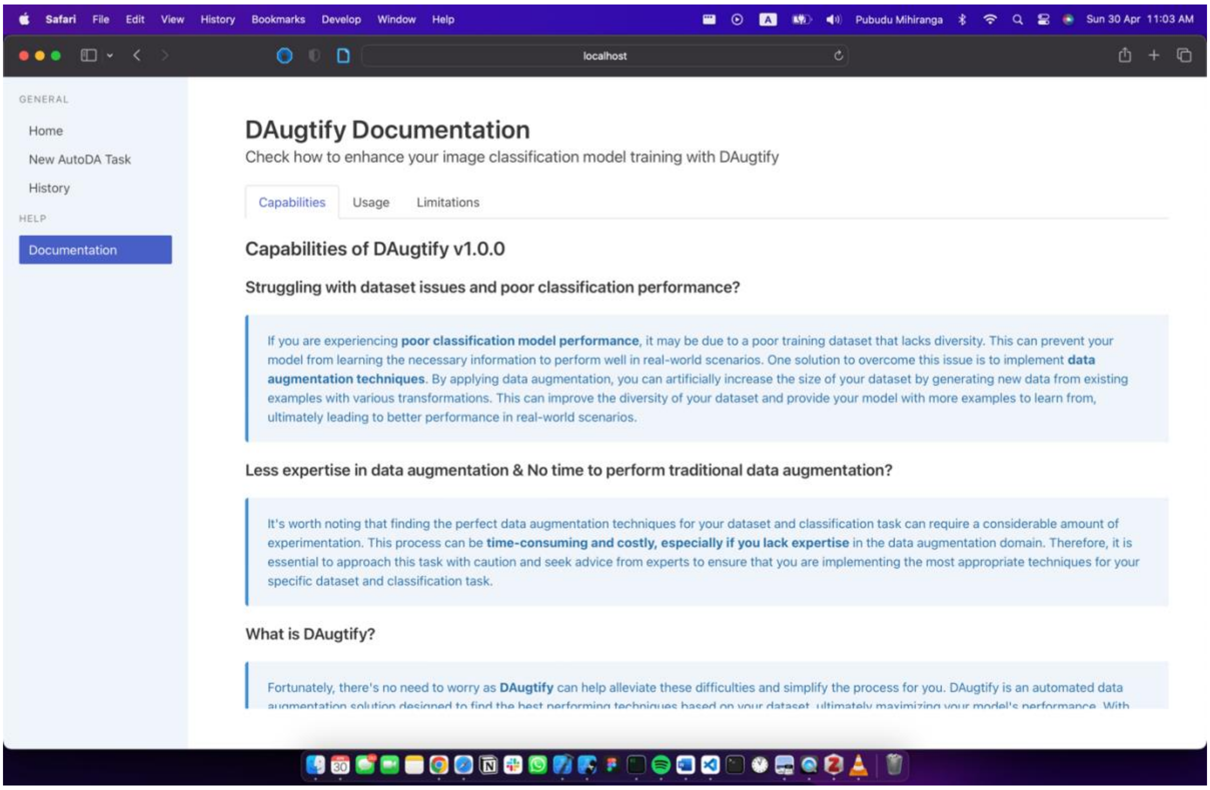


Figure 50: Documentation screen (capabilities)

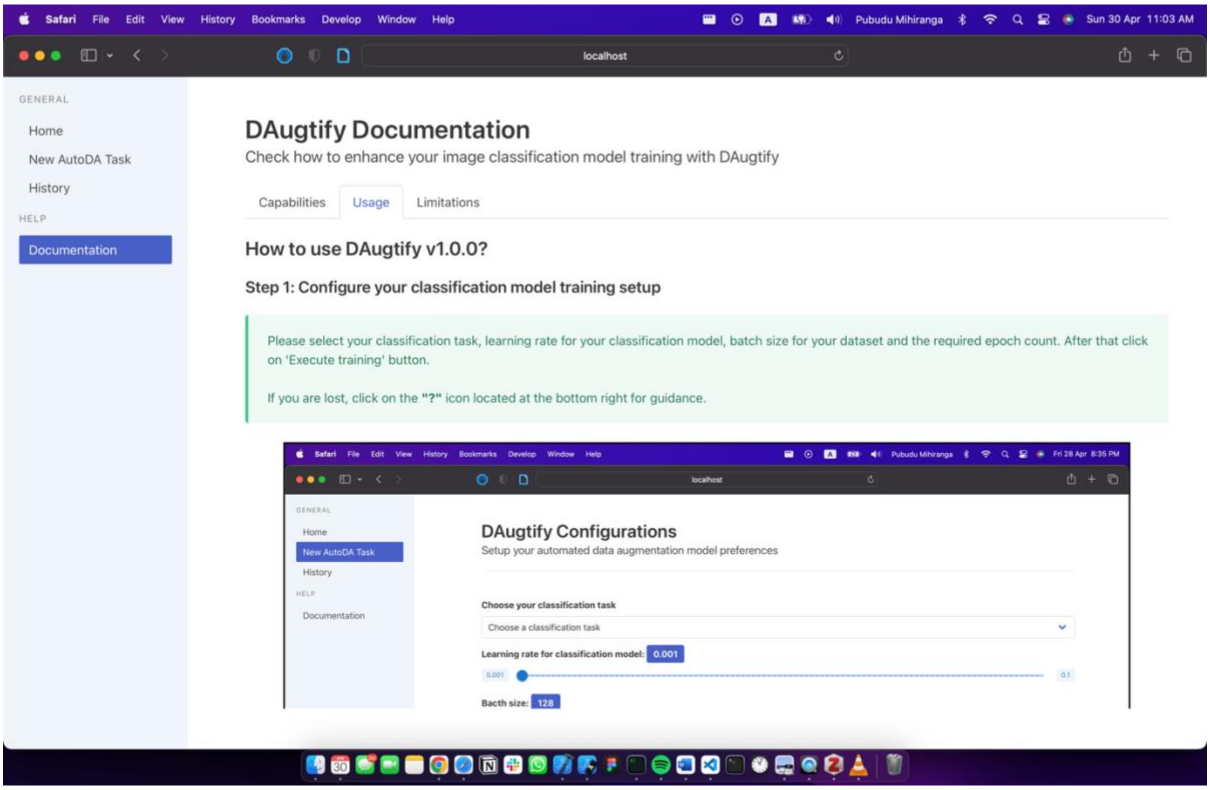


Figure 51: Documentation screen (usage)

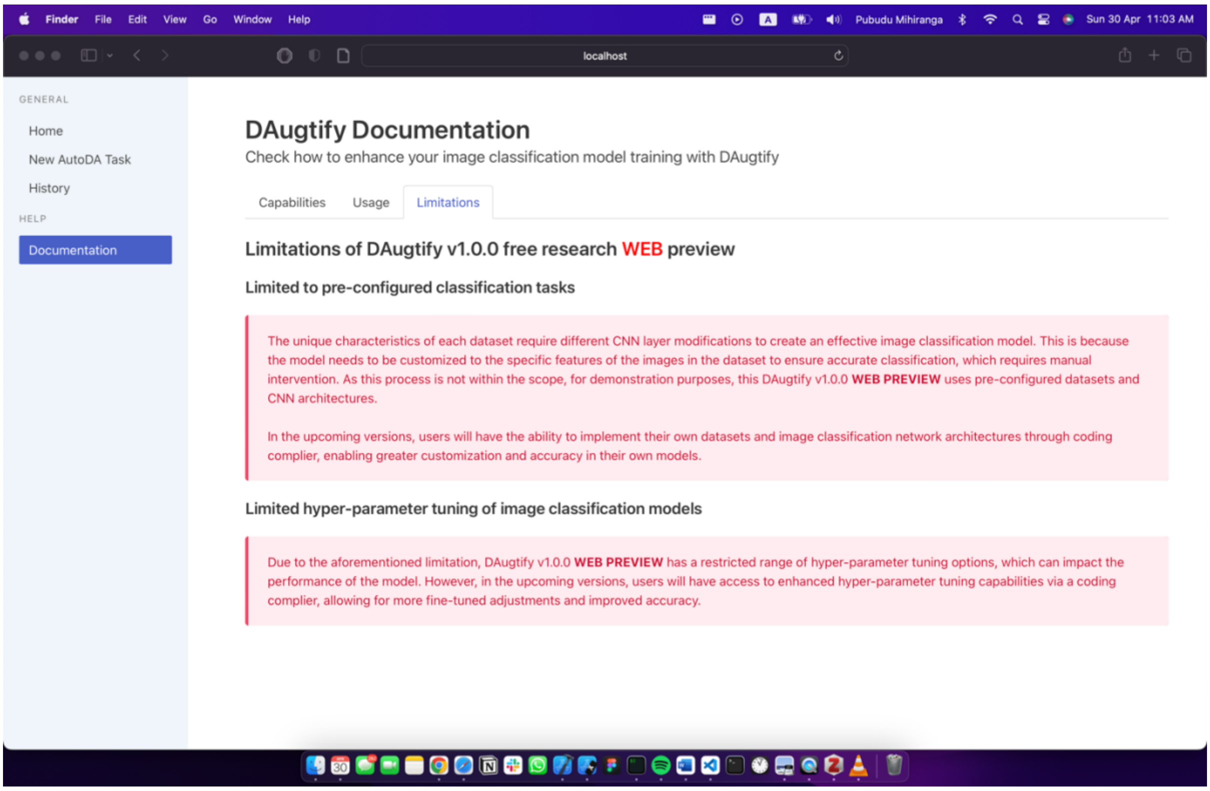


Figure 52: Documentation screen (limitations)

Appendix H: Testing & Benchmarking Results

1. CIFAR10 Classification with WideResNet 28-10

    1.1. Without Data Augmentation

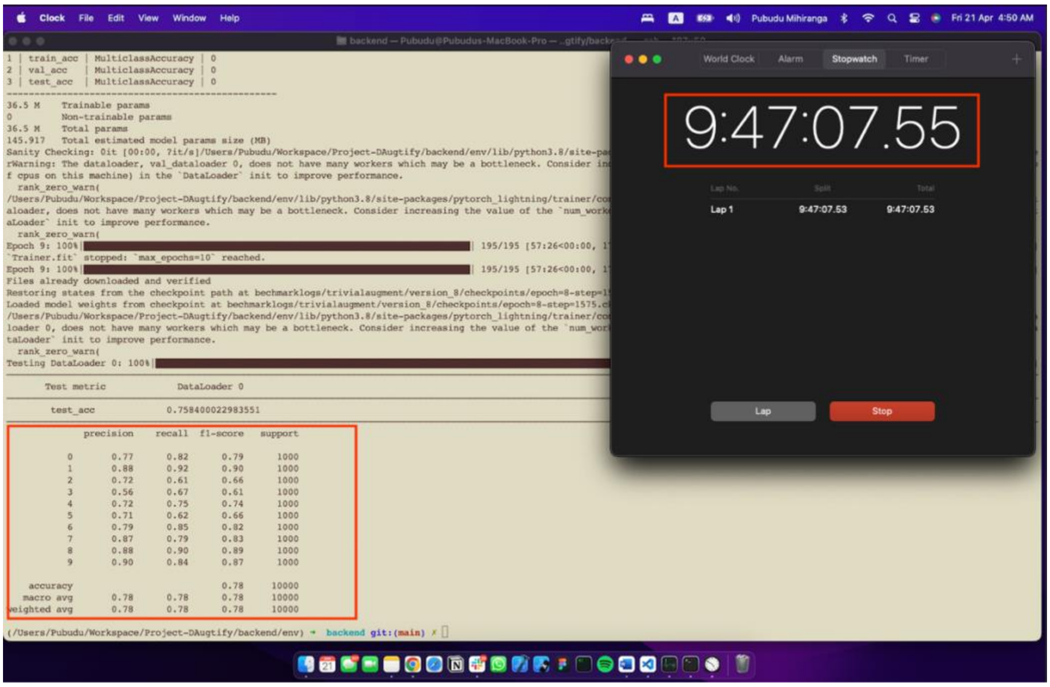


Figure 53: CIFAR10 Classification with WideResNet 28-10 (without data augmentation)

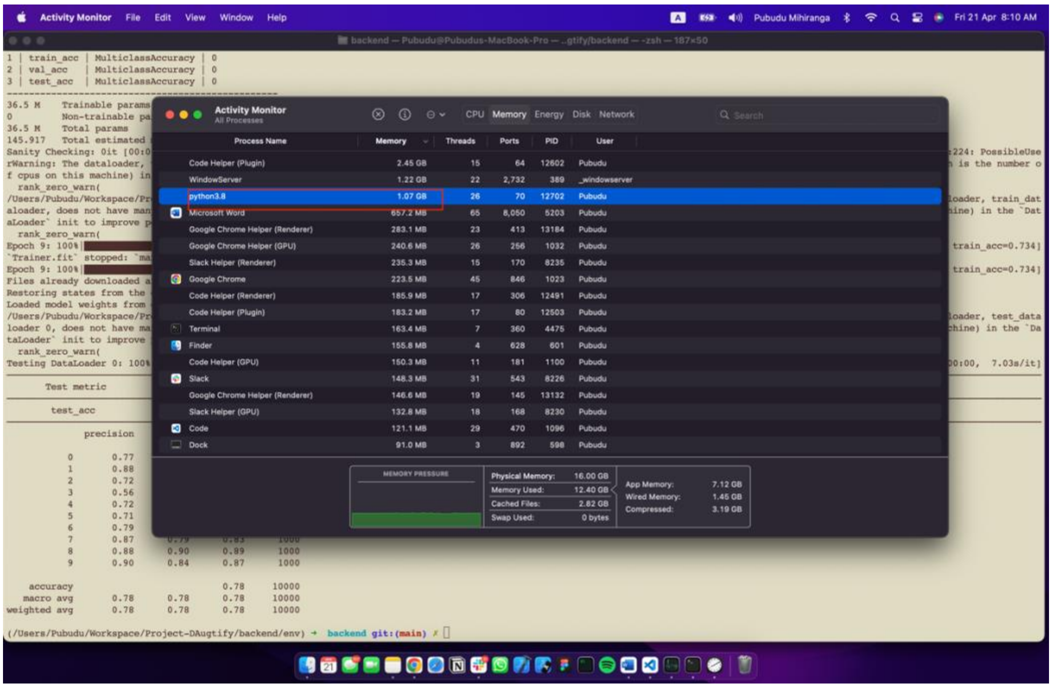


Figure 54: CIFAR10 Classification with WideResNet 28-10 (without data augmentation)

1.2. With DAugtify

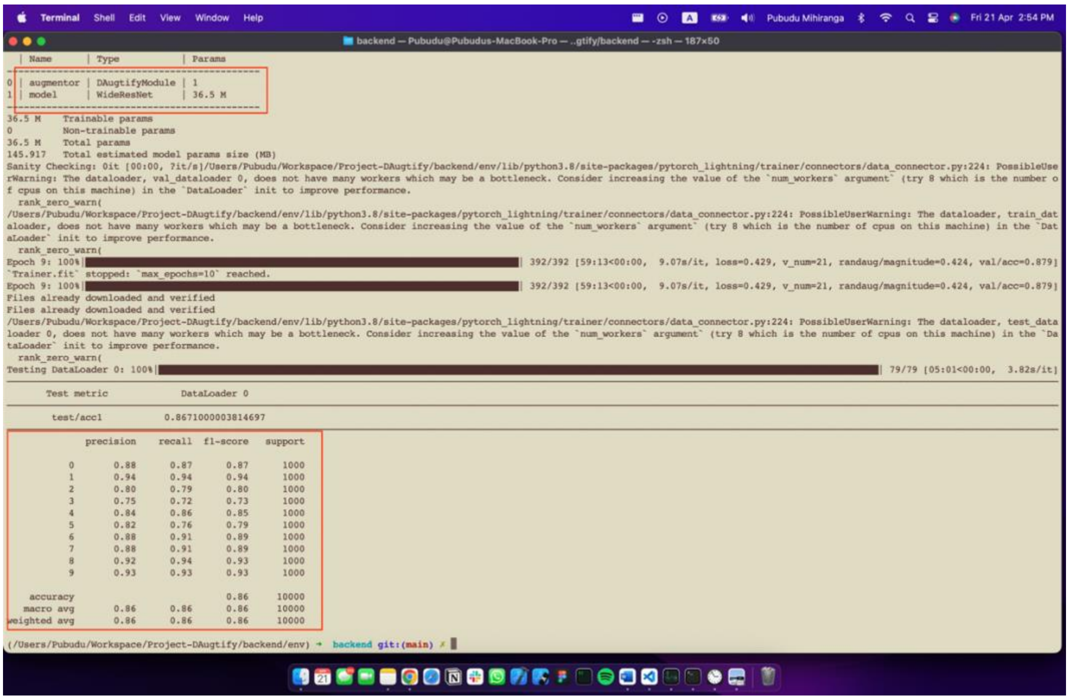


Figure 55: CIFAR10 Classification with WideResNet 28-10 (with DAugtify)

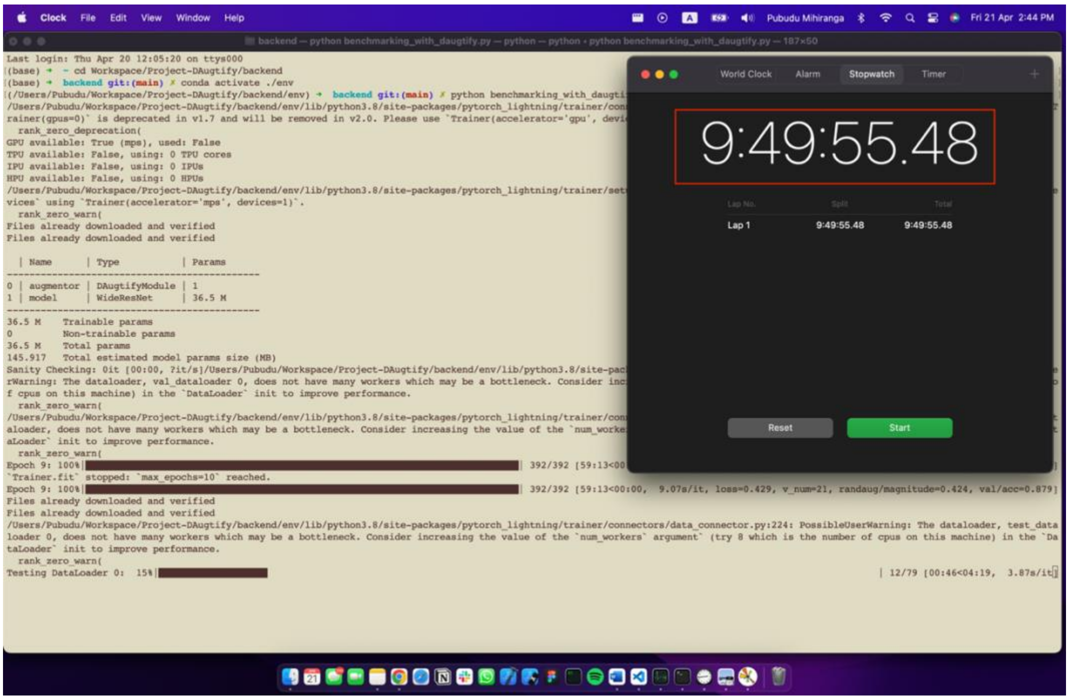


Figure 56: CIFAR10 Classification with WideResNet 28-10 (with DAugtify)

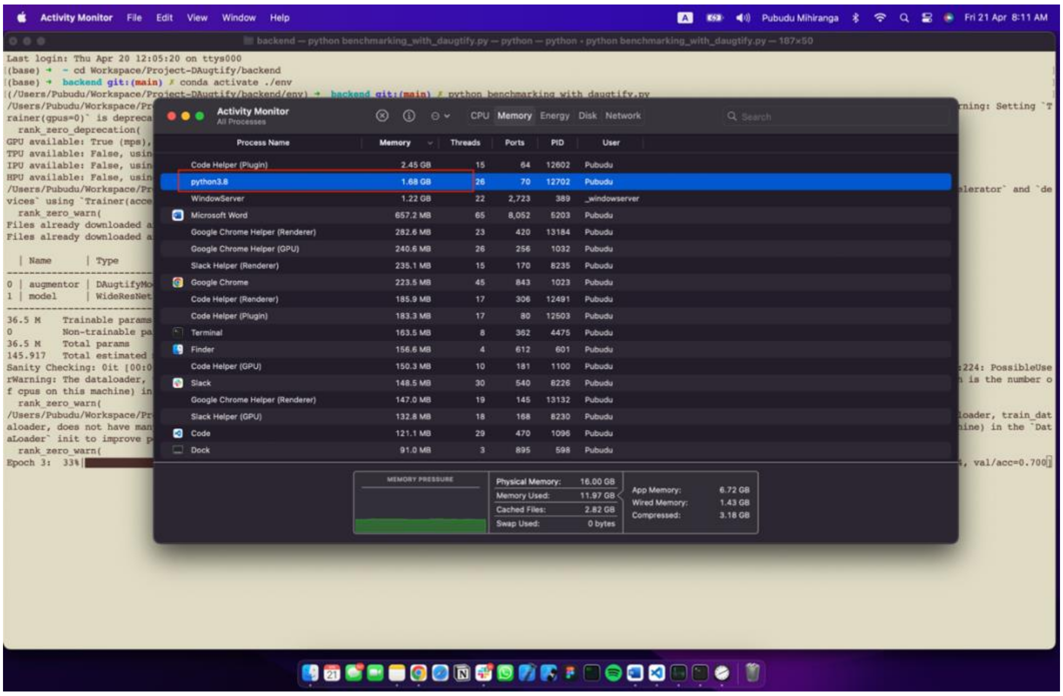


Figure 57: CIFAR10 Classification with WideResNet 28-10 (with DAugtify)

1.3. With AutoAugment

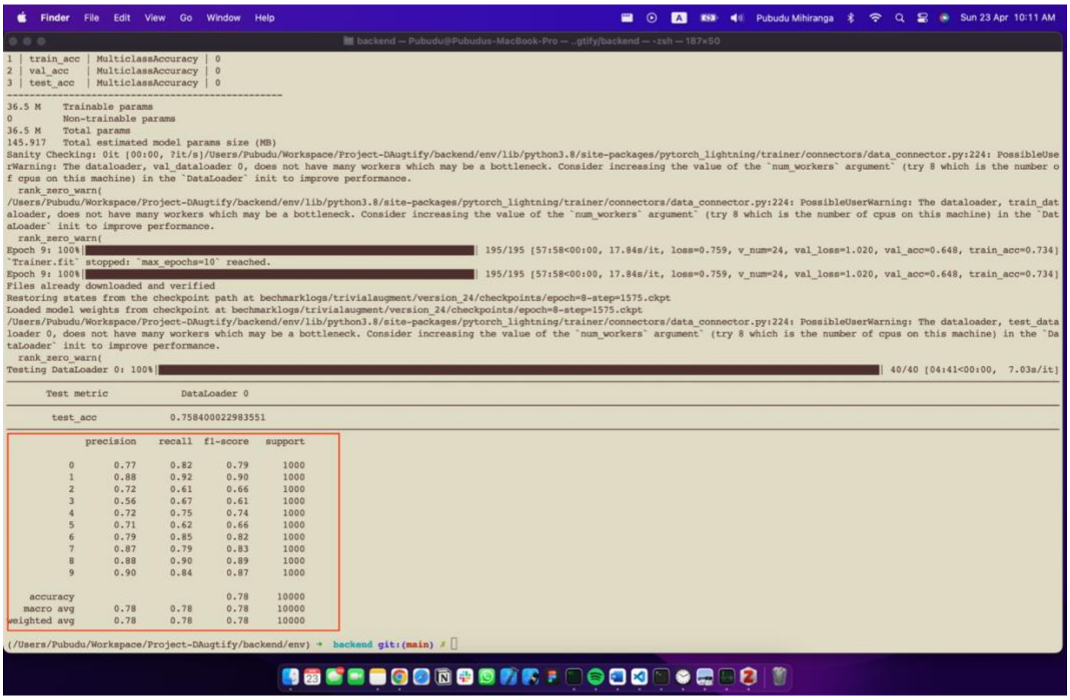


Figure 58: CIFAR10 Classification with WideResNet 28-10 (with AutoAugment)

1.4. With RandAugment

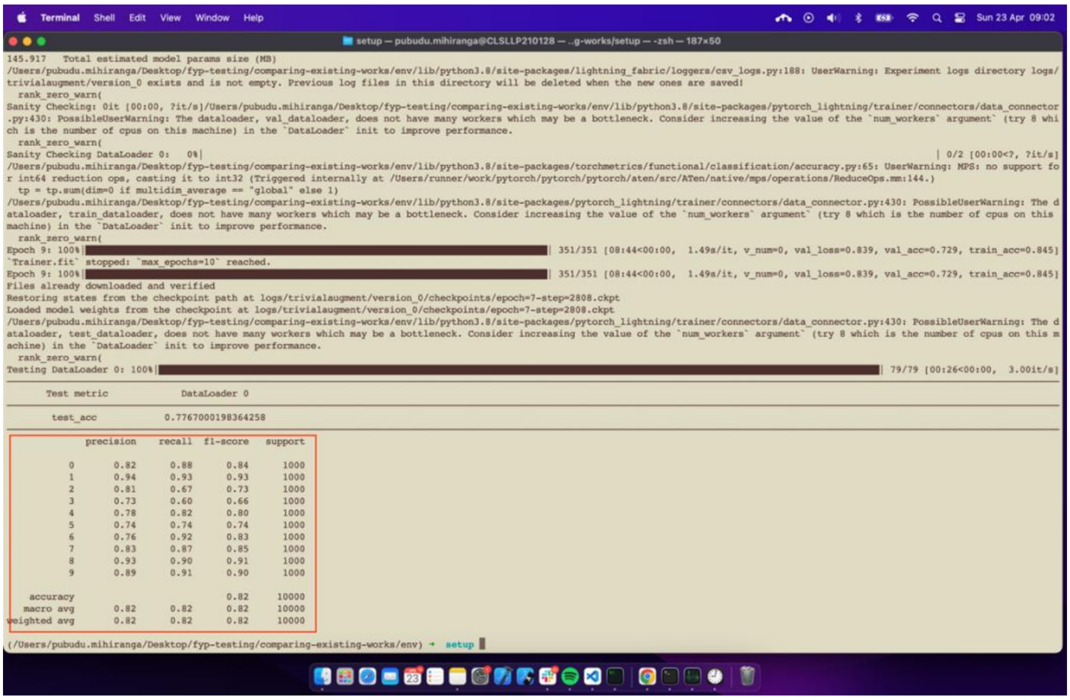


Figure 59: CIFAR10 Classification with WideResNet 28-10 (with RandAugment)

2. CIFAR10 Classification with ResNet18

2.1. Without Data Augmentation

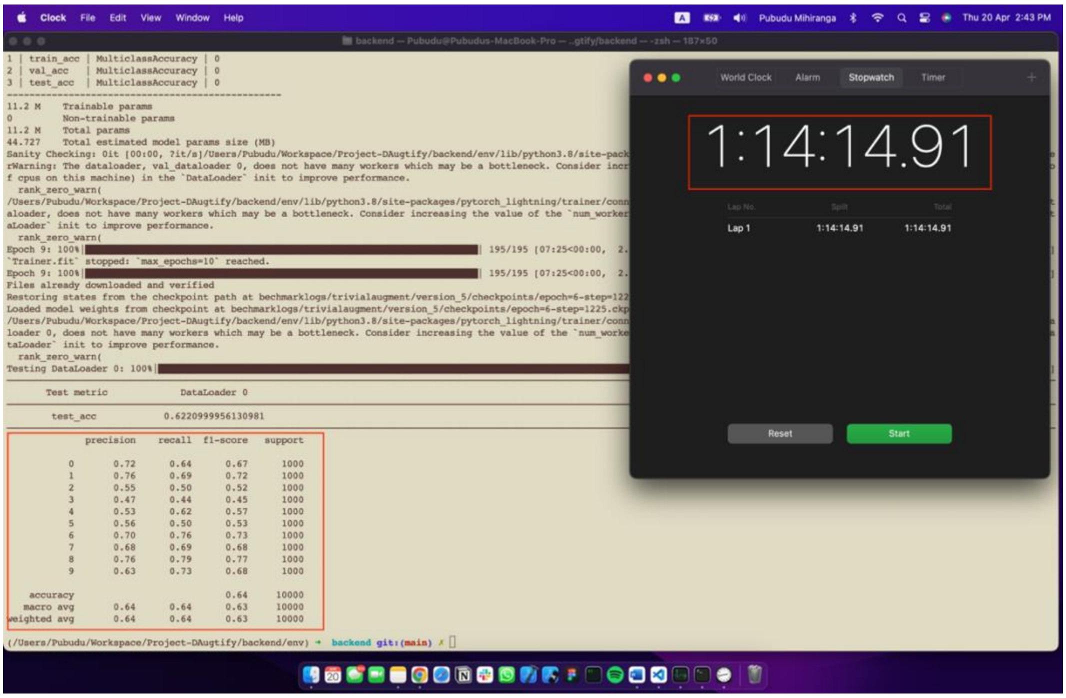


Figure 60: CIFAR10 Classification with ResNet18 (without data augmentation)

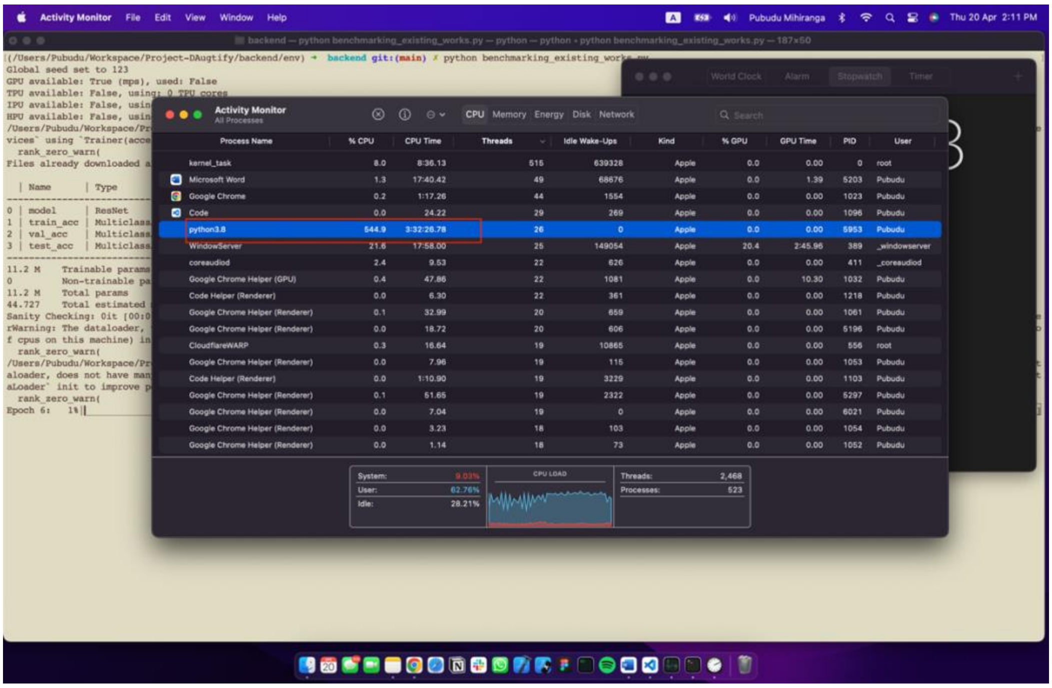


Figure 61: CIFAR10 Classification with ResNet18 (without data augmentation)

2.2. With DAugtify

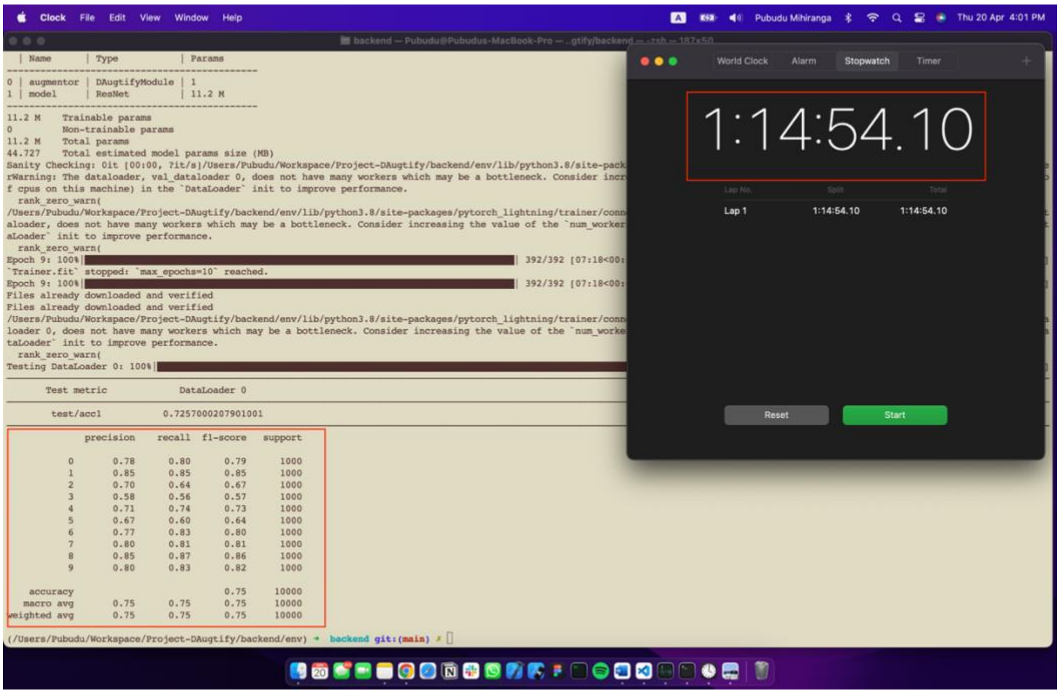


Figure 62: CIFAR10 Classification with ResNet18 (with DAugtify)

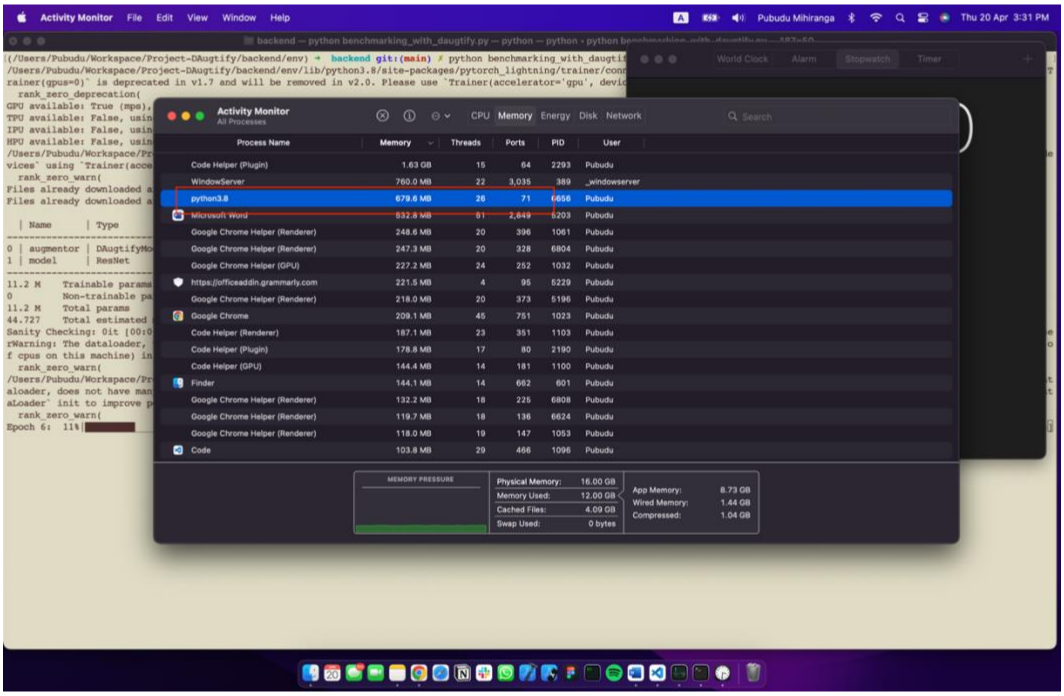


Figure 63: CIFAR10 Classification with ResNet18 (with DAugtify)

2.3. With AutoAugment

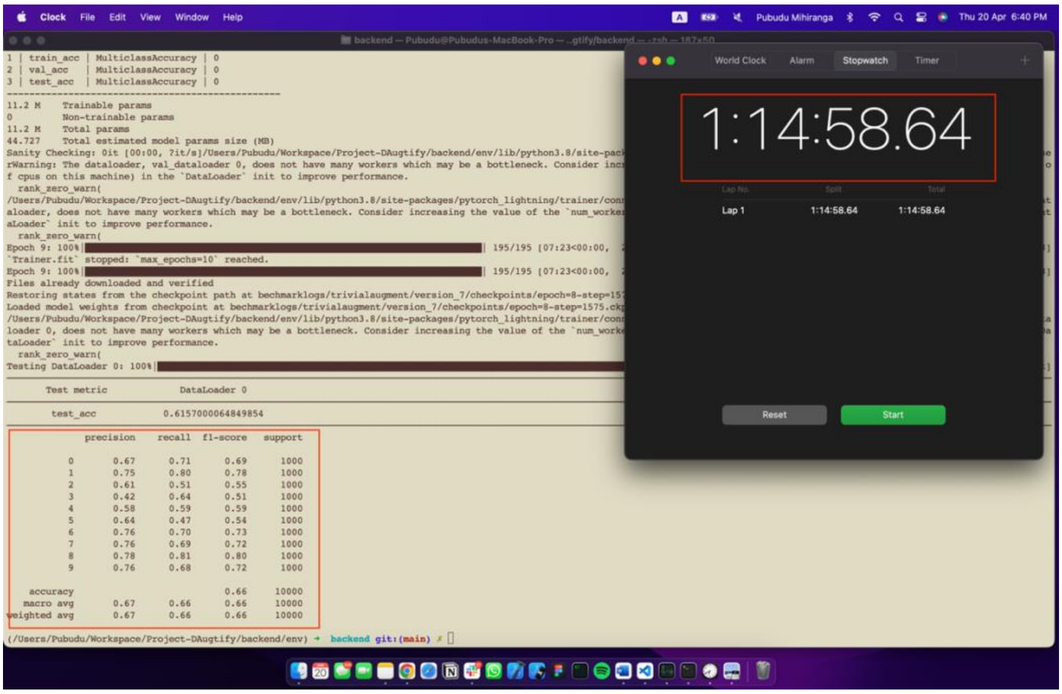


Figure 64: CIFAR10 Classification with ResNet18 (with AutoAugment)

2.4. With RandAugment

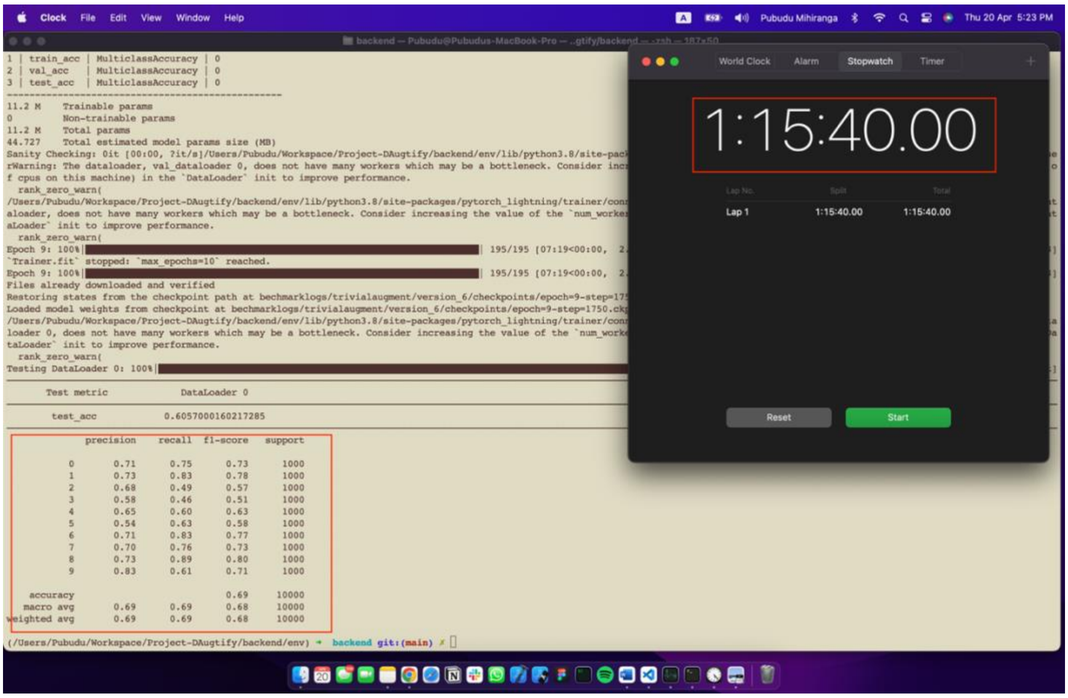


Figure 65: CIFAR10 Classification with ResNet18 (with RandAugment)

3. SVHN Classification with VGG16

3.1. Without Data Augmentation

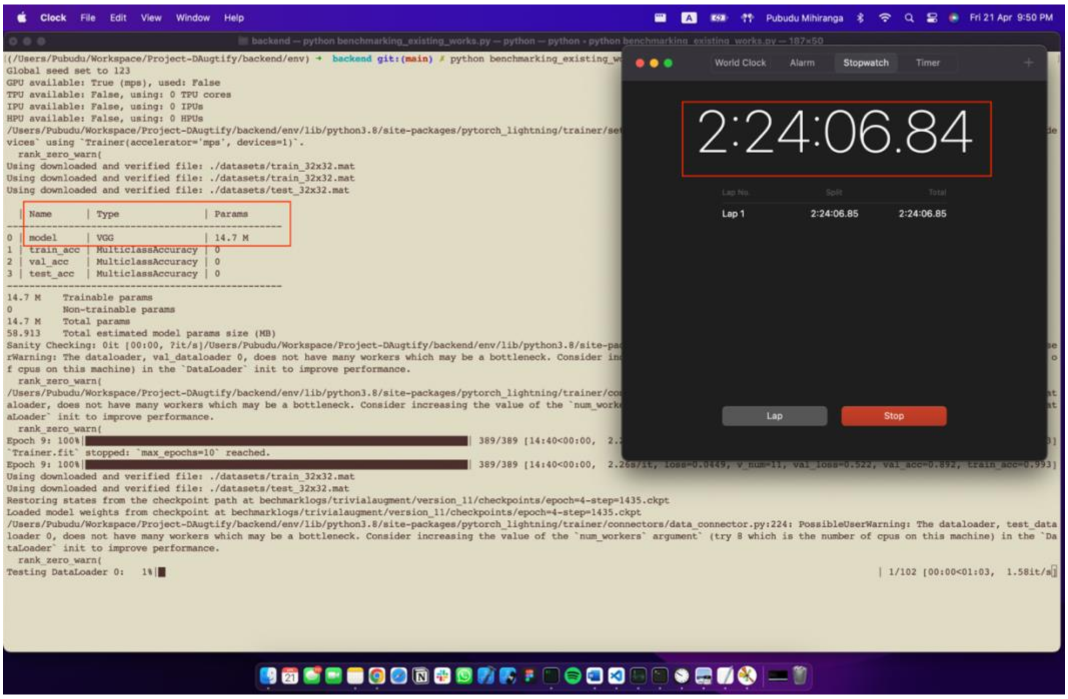


Figure 66: SVHN Classification with VGG16 (without data augmentation)

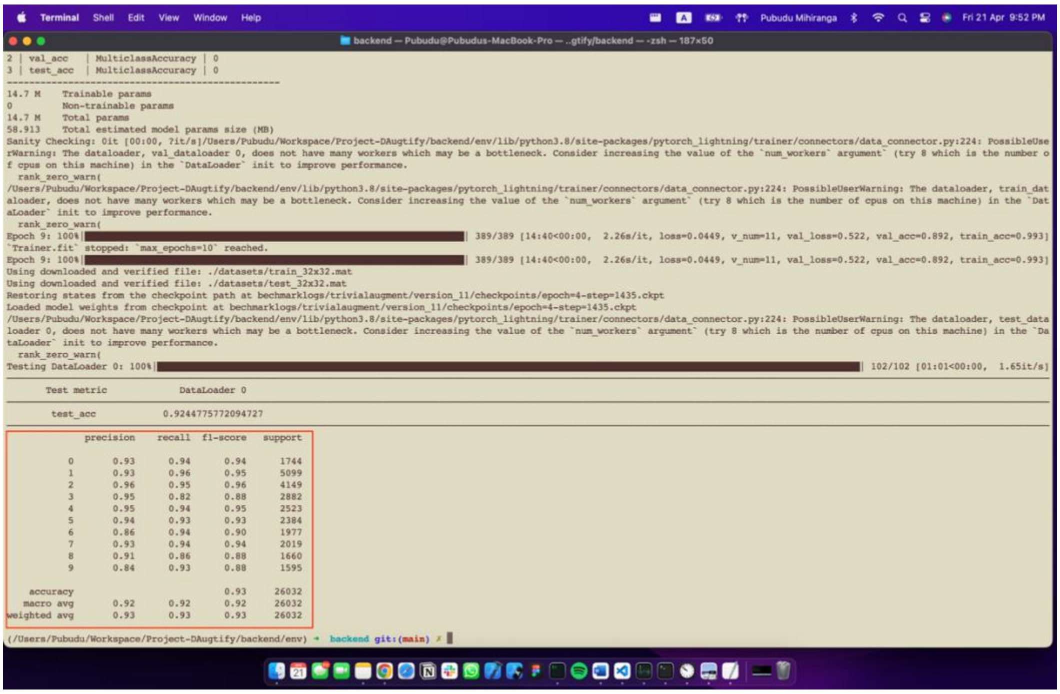


Figure 67: SVHN Classification with VGG16 (with data augmentation)

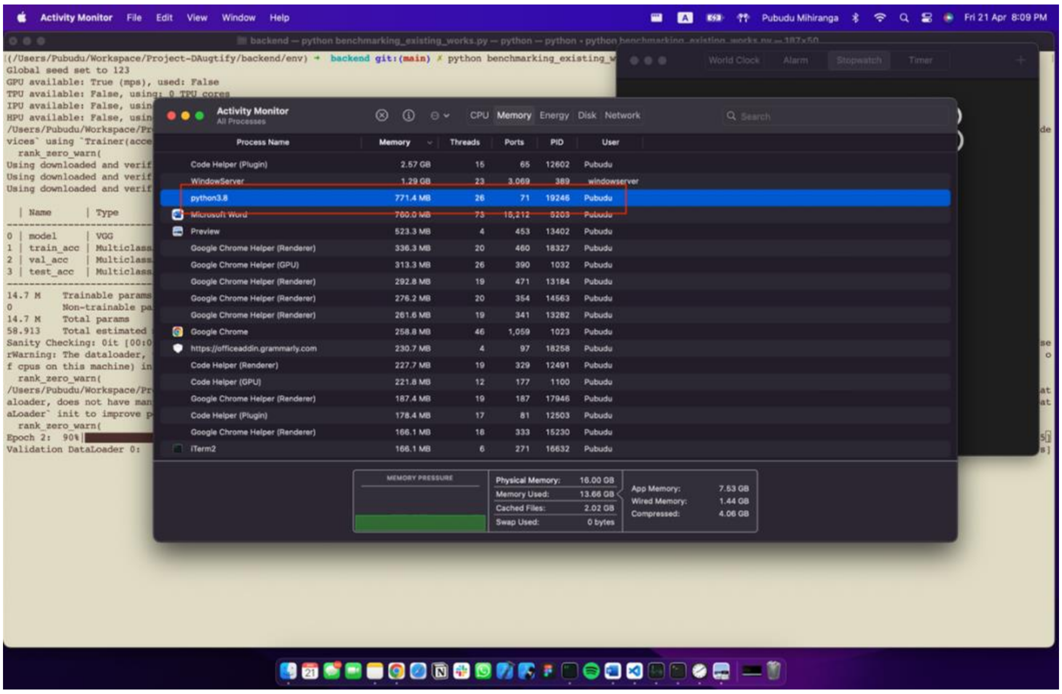


Figure 68: SVHN Classification with VGG16 (without data augmentation)

3.2. With DAugtify

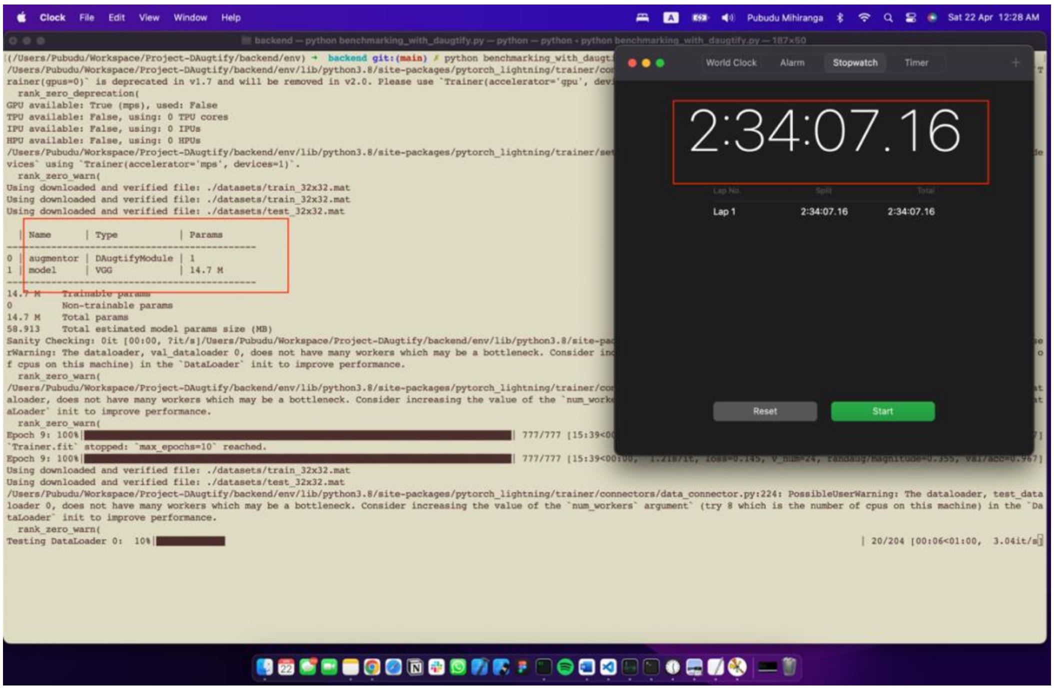


Figure 69: SVHN Classification with VGG16 (with DAugtify)

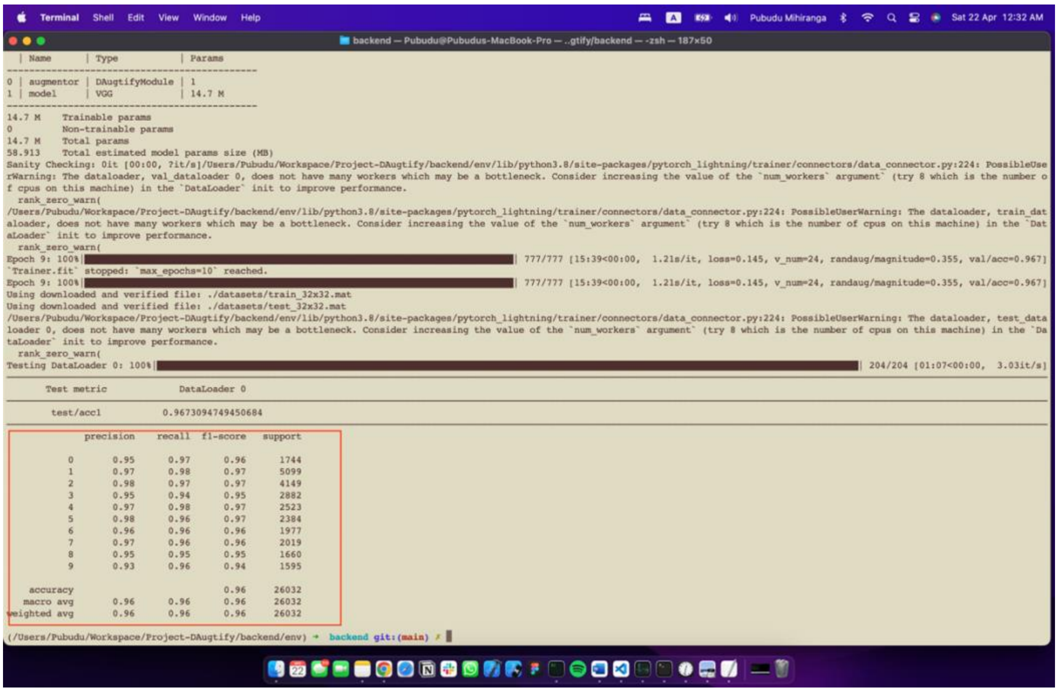


Figure 70: SVHN Classification with VGG16 (with DAugtify)

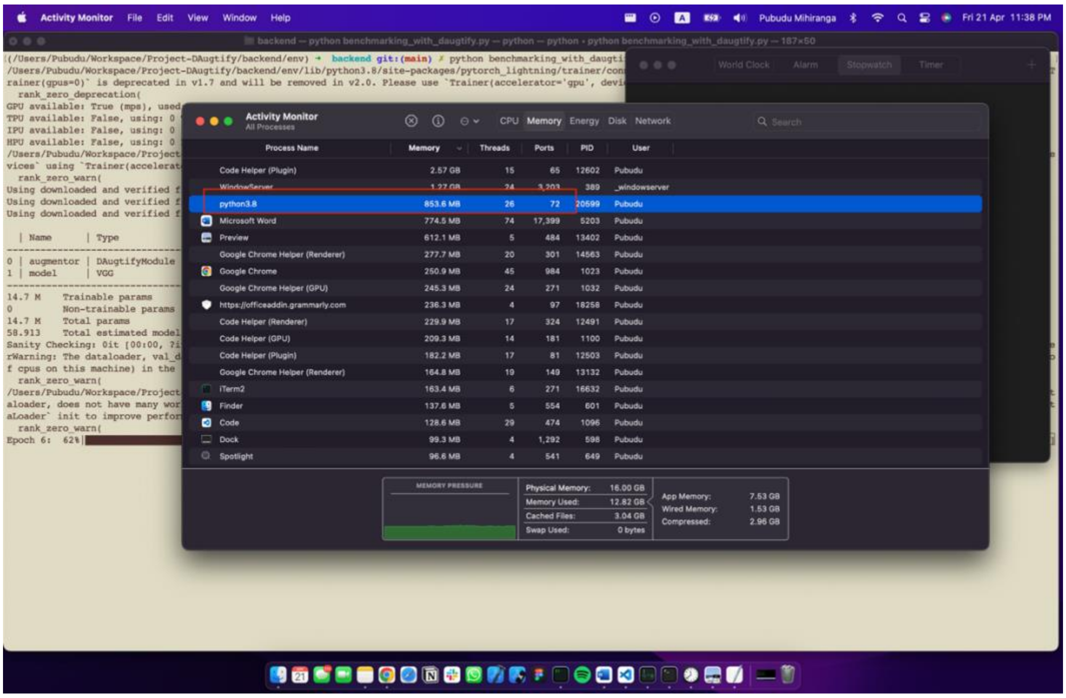


Figure 71: SVHN Classification with VGG16 (with DAugtify)

3.3. With AutoAugment

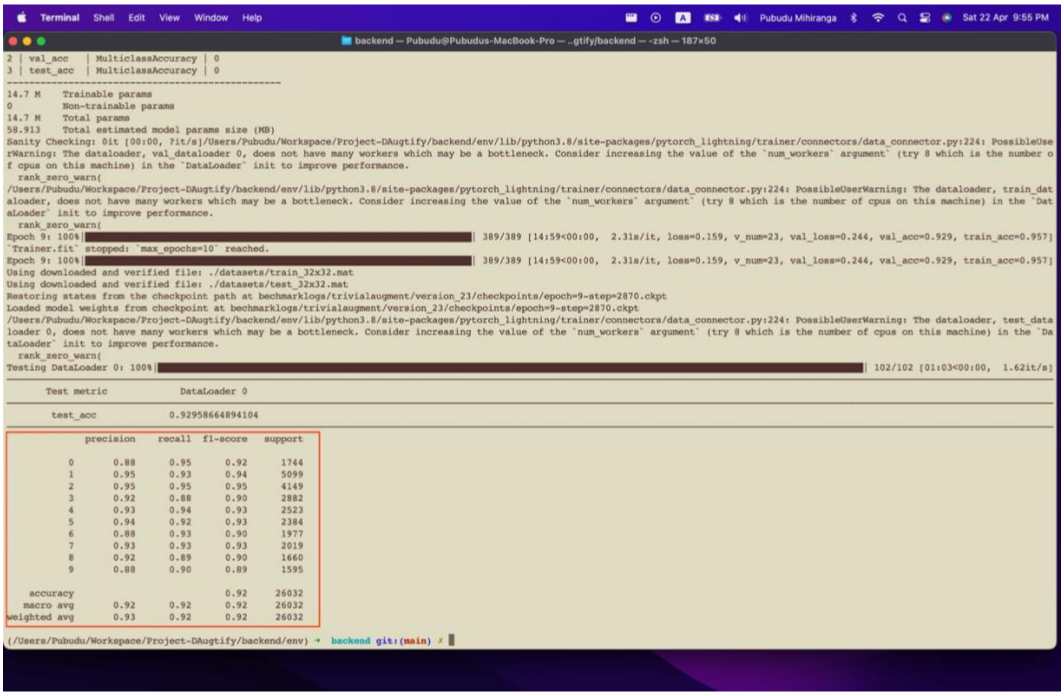


Figure 72: SVHN Classification with VGG16 (with AutoAugment)

3.4. With RandAugment

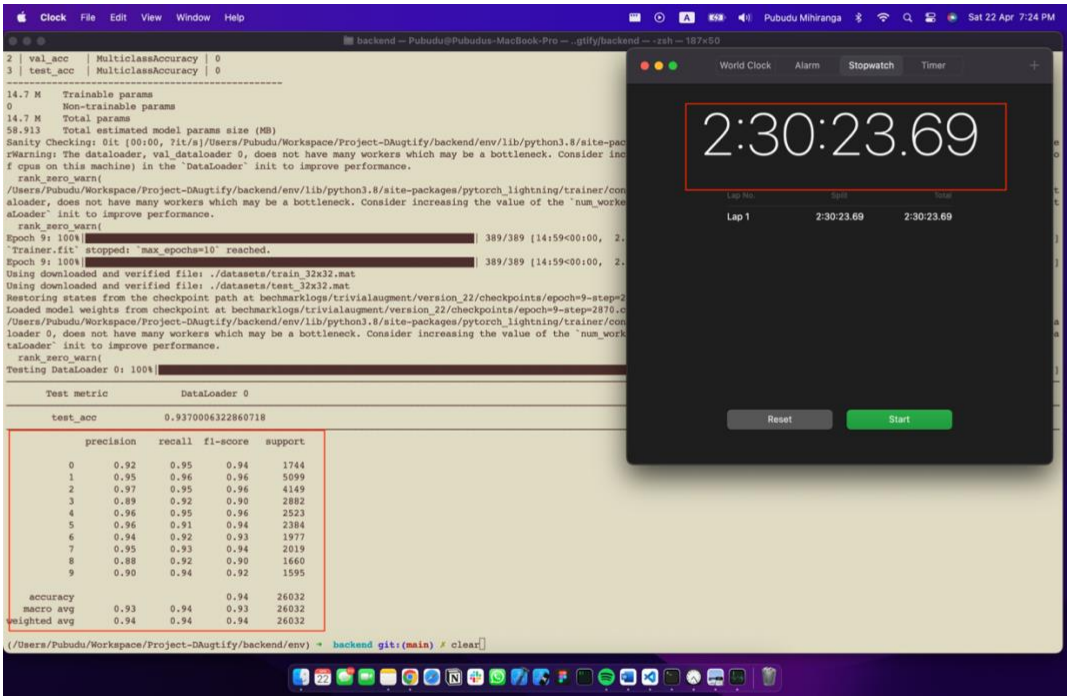


Figure 73: SVHN Classification with VGG16 (with RandAugment)

4. SVHN Classification with ResNet18

4.1. Without Data Augmentation

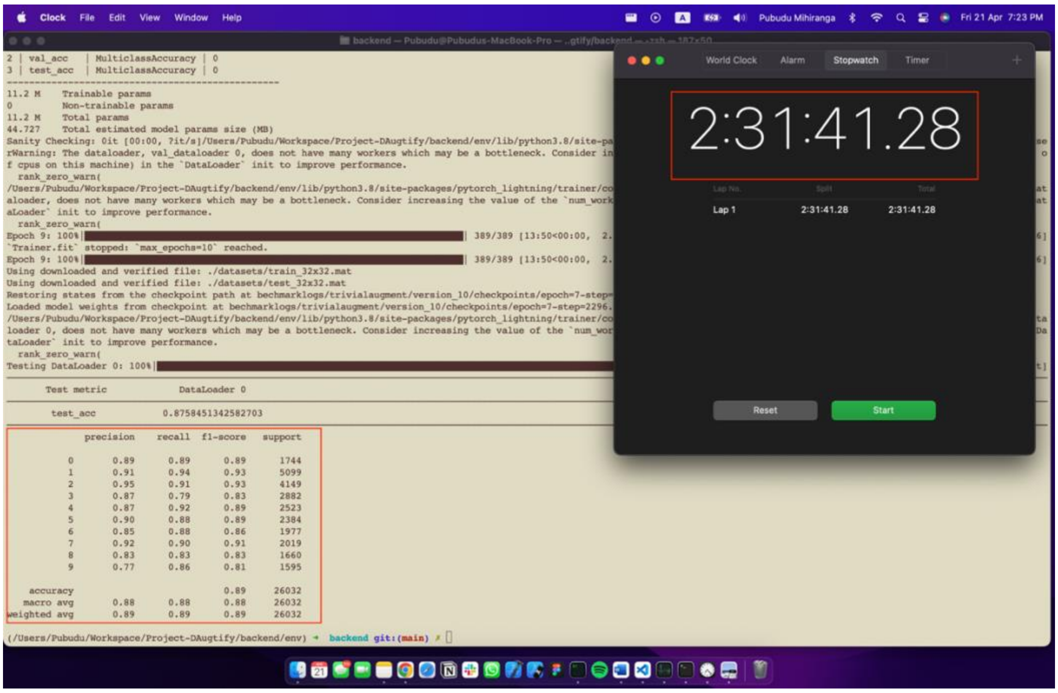


Figure 74: SVHN Classification with ResNet18 (without data augmentation)

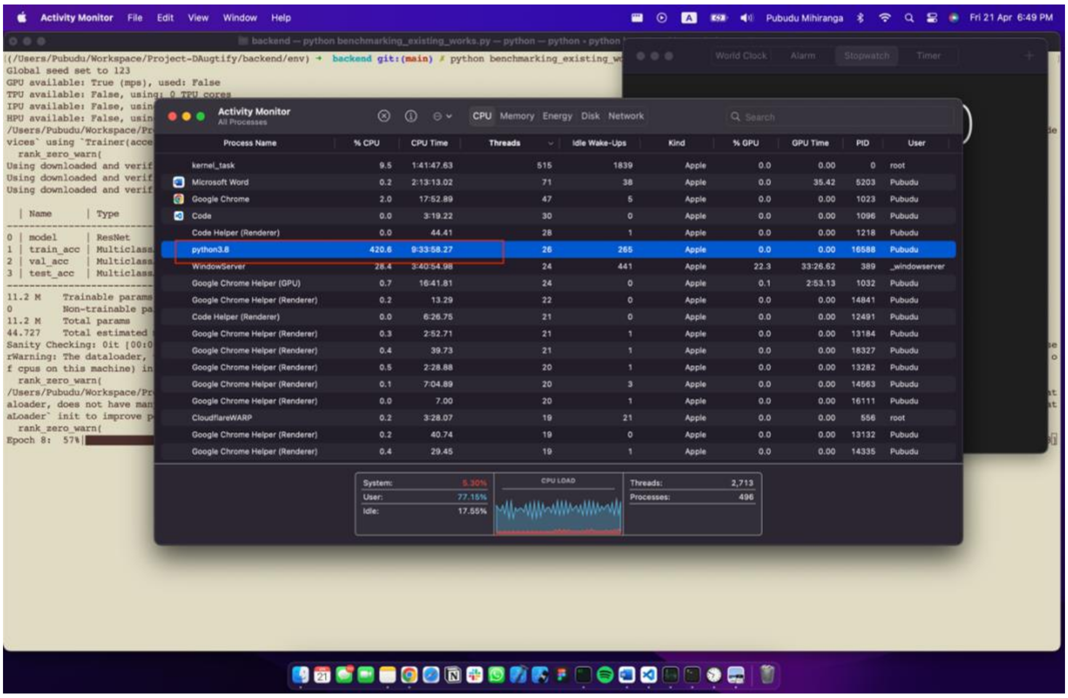


Figure 75: SVHN Classification with ResNet18 (without data augmentation)

4.2. With DAugtify

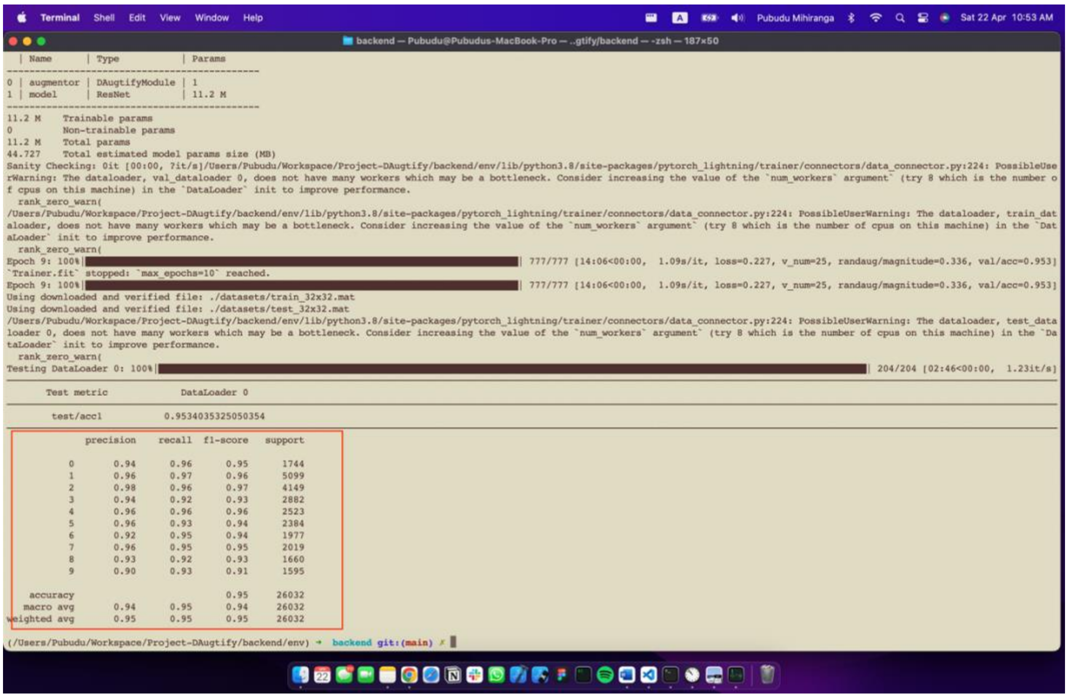


Figure 76: SVHN Classification with ResNet18 (with DAugtify)

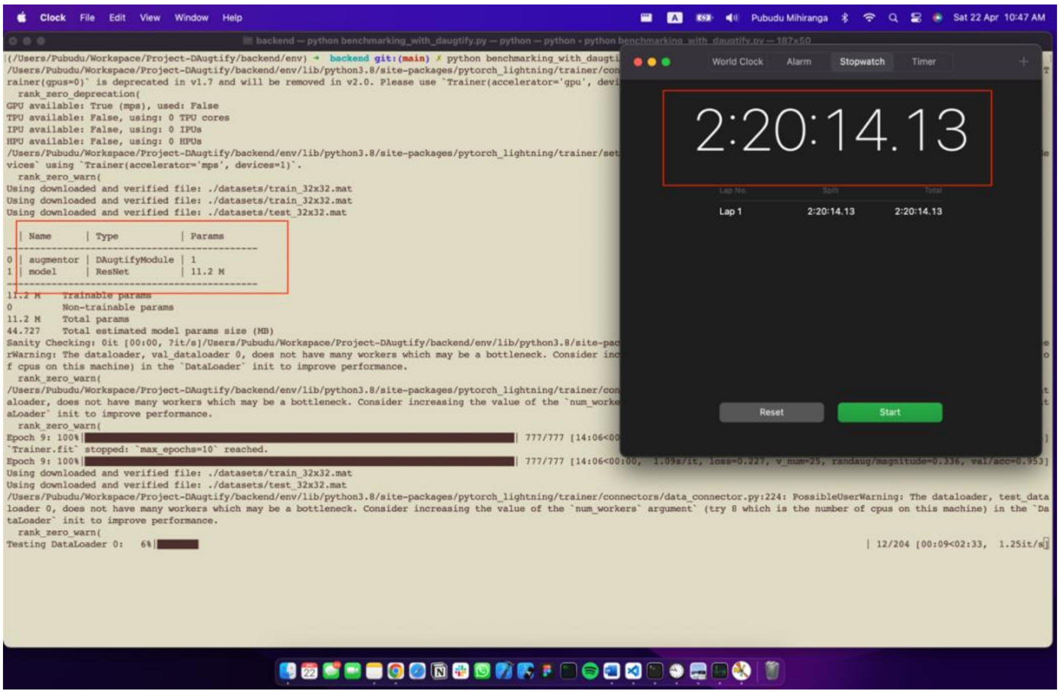


Figure 77: SVHN Classification with ResNet18 (with DAugtify)

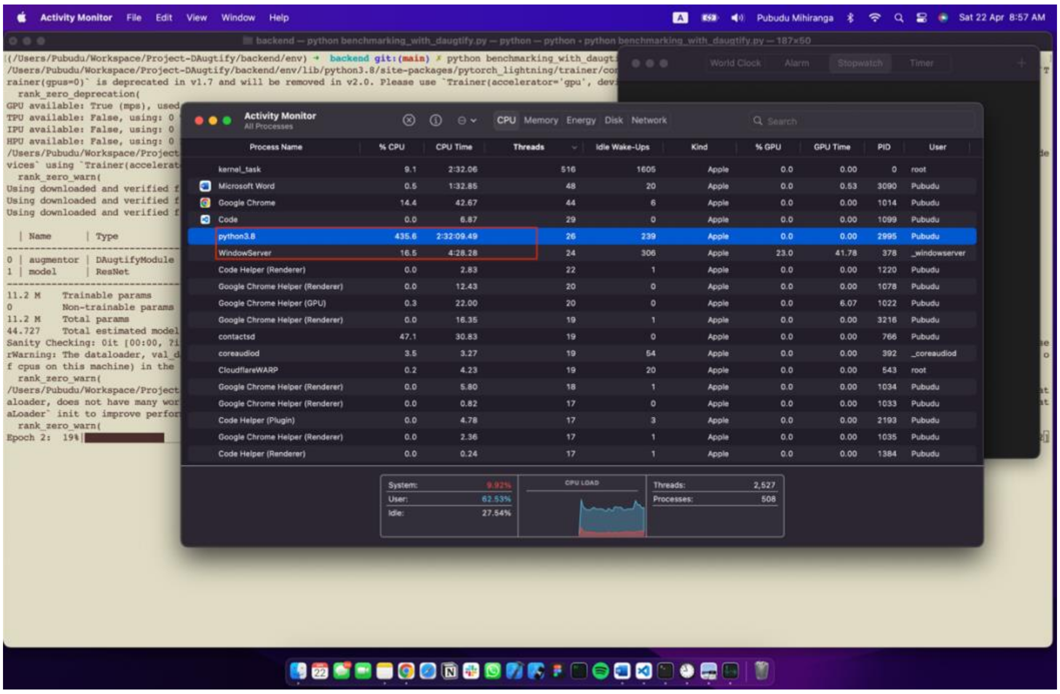


Figure 78: SVHN Classification with ResNet18 (with DAugtify)

4.3. With AutoAugment

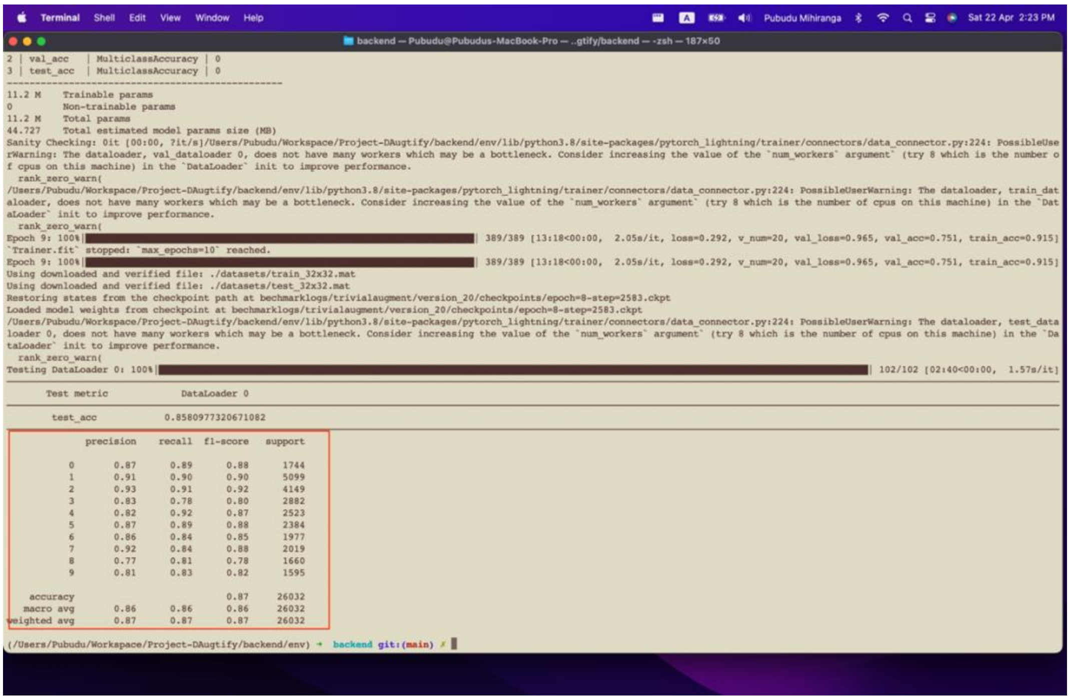


Figure 79: SVHN Classification with ResNet18 (with AutoAugment)

4.4. With RandAugment

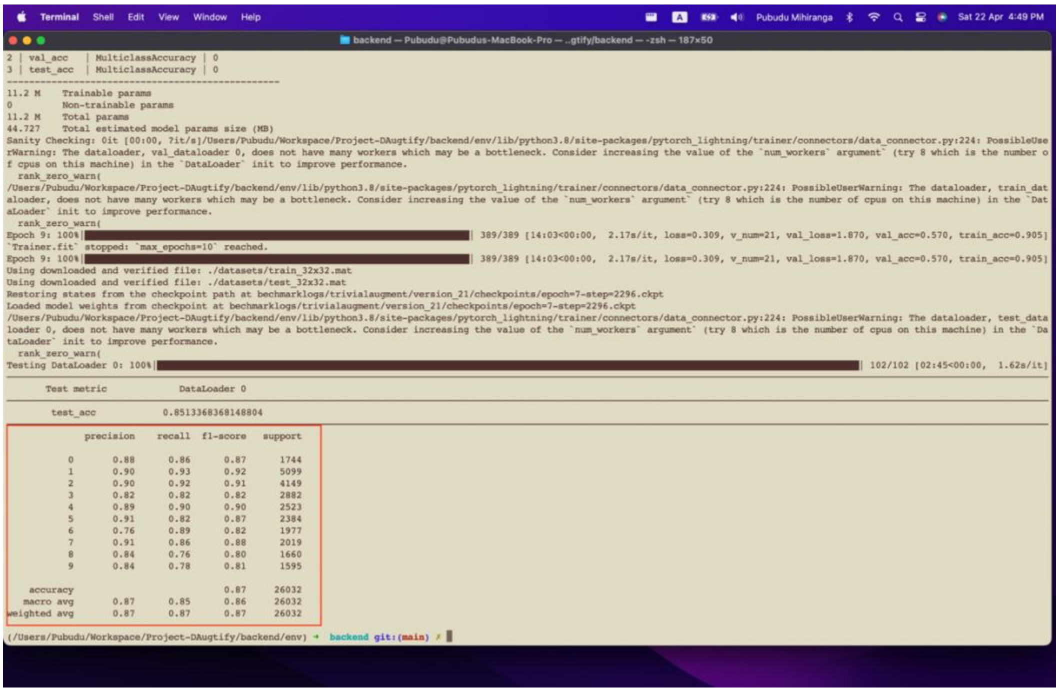


Figure 80: SVHN Classification with ResNet18 (with RandAugment)

5. MNIST Classification with LeNet5

5.1. Without Data Augmentation

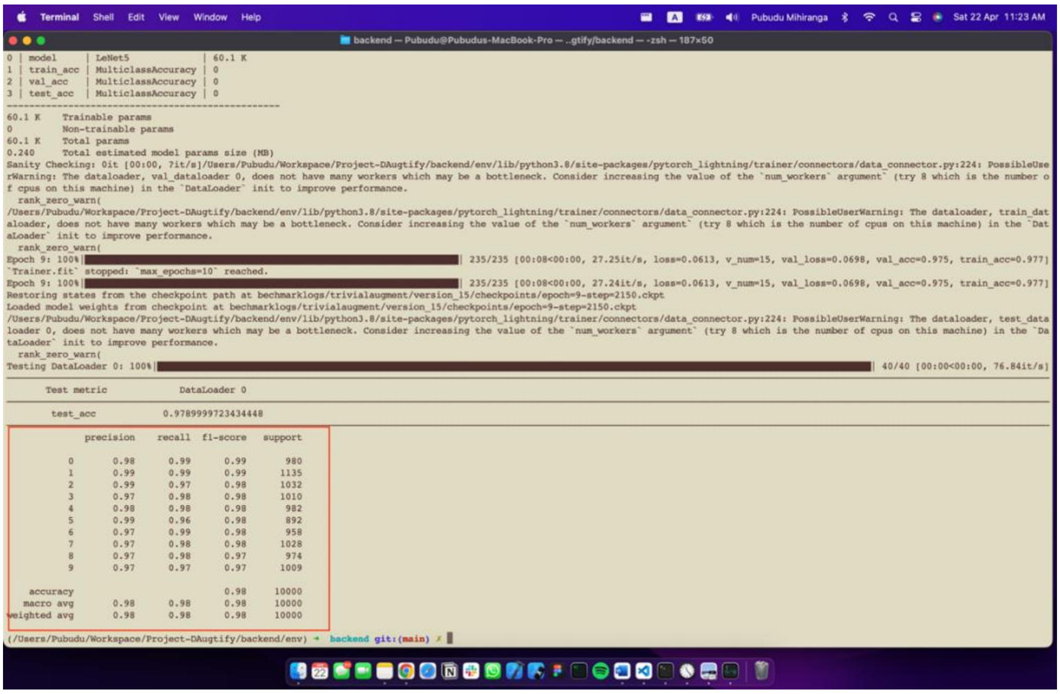


Figure 81: MNIST Classification with LeNet5 (without data augmentation)

5.2. With DAugtify

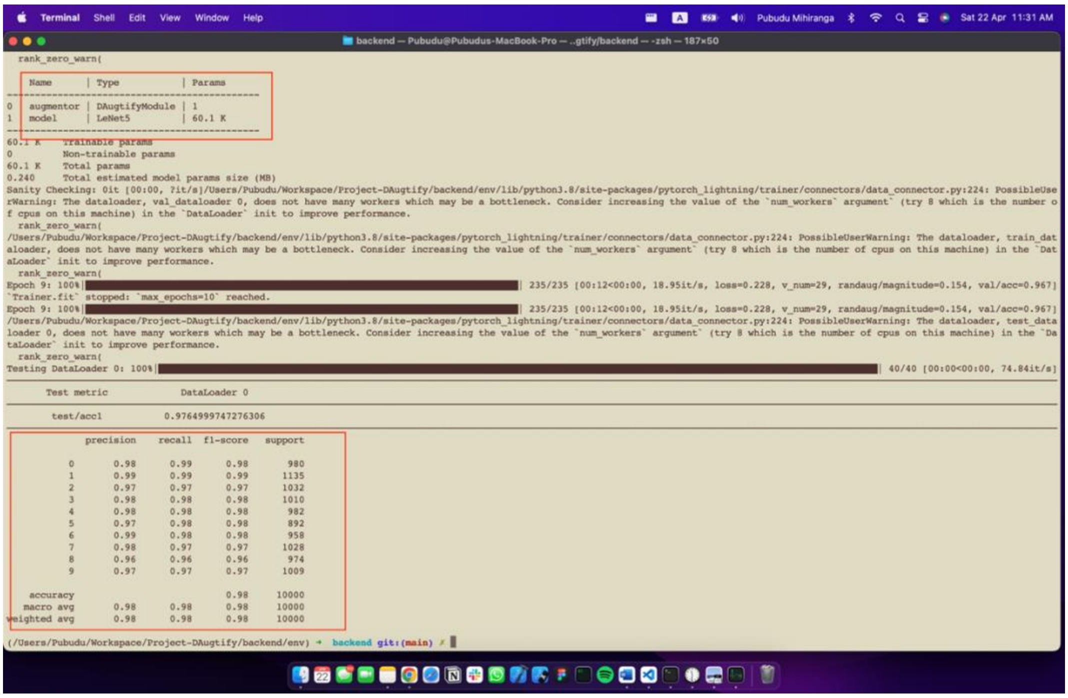


Figure 82: MNIST Classification with LeNet5 (with DAugtify)

5.3. With AutoAugment

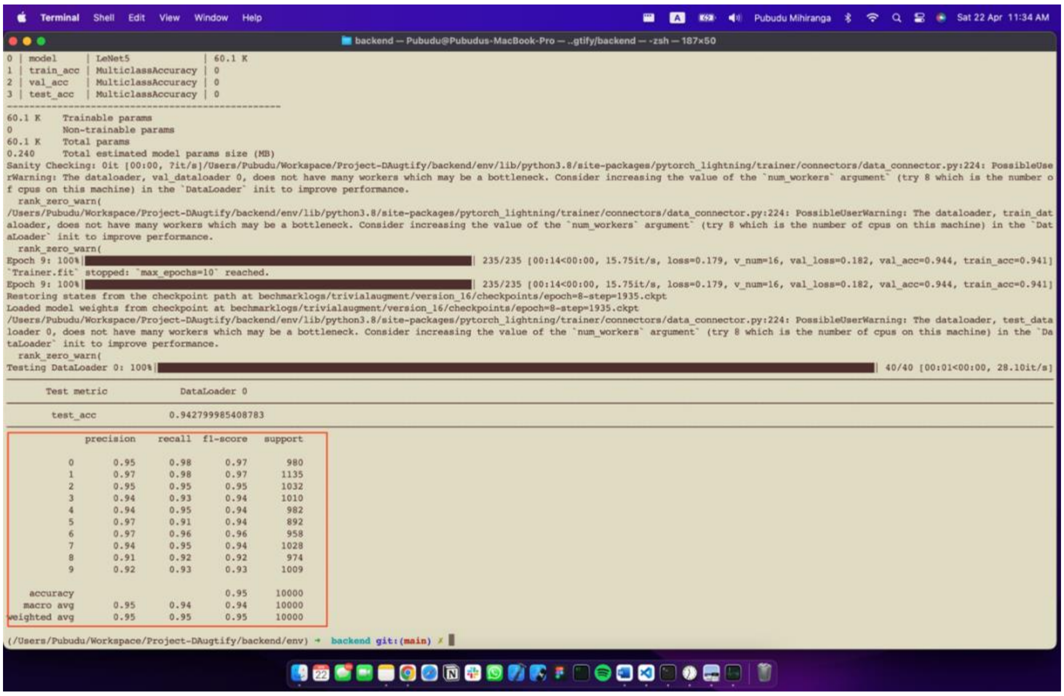


Figure 83: MNIST Classification with LeNet5 (with AutoAugment)

5.4. With RandAugment

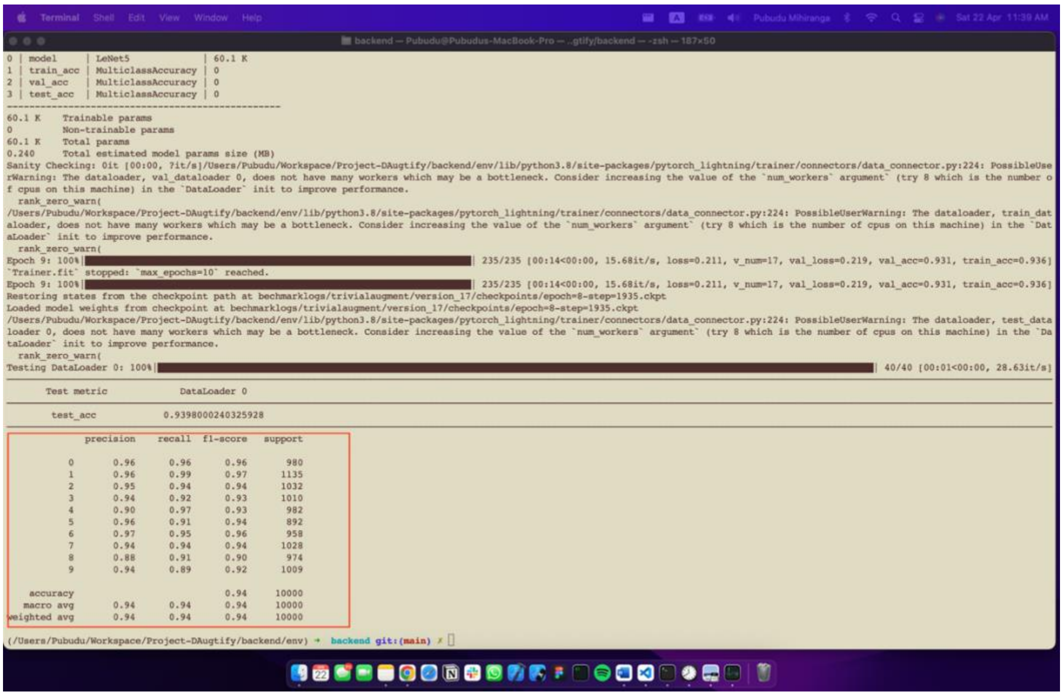


Figure 84: MNIST Classification with LeNet5 (with RandAugment)

Appendix I: Selection of Evaluators for UI/UX of the Project

ID	Name	Position	Affiliation
EV18	Mr. Gayan Kumarage	Staff Engineer in Product Engineering (Web-front)	Circles Life Sri Lanka
EV19	Confidential	Senior UI/UX Engineer	Confidential
EV20	Mr. Nimendra Kariyawasam	UI/UX Designer	Visual Theatre
EV21	Mr. Manoj Lakshan	Software Engineer	N-Ach
EV22	Mr. Malindu Mewan	Software Engineer	Avtra-Soft Technologies

Table 56: Selection of evaluators for UI/UX of the project

Appendix J: Evaluations of Domain & Technical experts

Question: What are your initial impressions on the project, the background and problem?

Evaluator ID	**Feedback**
EV1	The project addresses an important problem and is well informed by relevant background literature in the field.
EV2	I think the author has tried to tackle an interesting area of research.
EV3	The project is intriguing, as the problem chosen is well-defined. The research would be highly beneficial for many image-related tasks, especially when using non-GPU machines.
EV4	This is an important problem Computer Vision field is having including myself. I have worked on many CV projects and had the same issue.
EV5	An interesting approach to a problem faced in the machine learning space. The background and problem are relevant issues faced by researchers.
EV6	Impressive project that addresses a good problem scope.
EV7	The problem addressed in this project will benefit other researcher in many domains, overall, the product seems to be easy to use and provides a quality output.
EV8	Nice work by addressing a rarely touch, yet interesting topic.
EV9	It's great to see applications of machine learning based solutions to replace domain experts and reduce amount of time
EV10	Image augmentation and training is a heavy and costly exercise. The project here attempts to ease the process by cutting down few of the most rigorous exercises like finding the best hyperparameters to train the neural engine and making it affordable to achieve a decent initial result.
EV11	The project has been done based on a domain which has a great importance for machine learning applications. Such could help a lot in machine learning applications with limitations in training data. It can help a lot in improving a lot of model performances by enabling the utilization of varying data for training. The student has done a good research on the background of the domain to find a pain point as a problem to address. Providing a solution to this problem would

	help a lot in reducing performance costs a lot and it can improve the time limitations of domain related research processes as well.
EV12	The initial impression of the project has been outstanding. Practical DA solutions for a general audience are hindered by the sheer resources these systems need. Making a resource-friendly DA solution that can be used by anyone, anywhere, on any system can help propel ML applications
EV13	This is a timely essential research topic/ problem that most computer vision researchers faced while introducing an accurate model. Moreover, this enables lots of novel research directions on the robustness of the intelligent system research domain as model robustness will depend on the distribution of the training data.
EV14	I think the project, background and problem are quite interesting. I would've liked to know what sort of an impact this would bring for real-world applications (would be nicer if the researcher could quantify this by any means of measure, in addition to the test results).
EV15	This a valid and a promising approach towards the solutions for heavy time consumption and lack of domain expertise required for choosing the optimal data augmentation techniques for model training.
EV16	The project seems to be well-informed and grounded in existing research. The proposed solution has the potential to save time and resources while producing better results, and could become an important part of standard machine learning model training pipelines in the future.
EV17	I think its a good initiative and haven't seen much work in this field prior.

Table 57: Expert feedbacks for question 1

Question: What are your impressions on the novelty and relevance of the project in Automated Data Augmentation?

Evaluator ID	**Feedback**
EV1	Some novelty is evident, mainly in enabling image augmentation to be done to a high accuracy level with limited computing resources
EV2	I think the aproach utilized is a novel one
EV3	The novelty is unique, and as it is based on an existing research, the way the project was formulated is really good.

EV4	It's definitely new and relevant. Having a light-weight system is very important.
EV5	DAugtify research covers a widely used scope and presents a viable solution.
EV6	Data augmentation has been in the industry for a while but Automated approach that this project describe towards data augmentation novel and can be useful to many other researches which faces the limitations because of data
EV7	When compared to other products this doesn't require any high end resources and can be easily configured
EV8	Relevance is good. Since most undergraduate tends to go for just predictions on algorithms for random dataset. Here Pubudu tries to contribute to a research area that we can use to improve any model
EV9	applying neural network based solution to finding optimal magnitude in AutoDA is novel approach, And it'll definitely cut the required human effort on that stage
EV10	The project is like a stepping stone for the ones who wants to get into the field but don't know from where to start. It creates a base on top of which further improvement can be planned.
EV11	I think the novelty is great as this research is defining and addressing the problem based on latest researches. The relevance too is very valid as this performance limitation will apply across relevant Data Augmentation models.
EV12	The project's relevance is on par with other AI/ML topics that are trending. AI is meant to make lives more manageable, and such a tool developed by Pubudu can be helpful and relevant. Novelty-wise, he has taken huge strides to make DA to make it as accessible as possible, a leap that is yet to be taken by anyone else, as per my knowledge.
EV13	The introduced algorithm (Selecting the appropriate augmentations from the search space) is a novel approach. Personally, I highly believe that this is excellent work.
EV14	The project seems to be able to address the issue of resource constraint when it comes to Automated Data Augmentation.
EV15	Fusing a single hyper- parameter neural network to achieve the reduction of the search time, complexity and the computational resources is a well researched decision to make.

EV16	Based on my understanding of the project, it appears to be a novel and relevant contribution to the field of AutoDA. The approach was able to outperform other state-of-the-art data augmentation methods, while also being computationally efficient, suggests that it has practical relevance for researchers and practitioners working with machine learning models. I believe that this project represents an important step forward in the development of more effective and efficient automated data augmentation techniques.
EV17	I think both these factors are on par with the main motivation of the project

Table 58: Expert feedbacks for question 2

Question: Is the depth of the research sufficient for undergraduate research?

Evaluator ID	Feedback
EV1	This is not my primary area, but it looks to be a well conceived and executed research project
EV2	Yes, the depth of the research is sufficient for an undergraduate project
EV3	Yes
EV4	I would say so.
EV5	The researcher has used the learnings from RandAugment and his own research to create a solution. This requires significant research and execution.
EV6	Though this is a wide domain area student has selected a challenging scope and managed well to achieve that
EV7	yes, the solution provides an accurate output and the complexity is fair enough for an undergrad research
EV8	Indeed
EV9	yes, author has involved greatly on data augmentation and neural network technologies
EV10	Yes, it's a pretty decent and well thought through project and brings value on the table.
EV11	I am satisfied with the depth of the research.
EV12	Yes, the research has great depth. Making such a complex system resource-friendly is a paramount task. However, as the test results prove, the system has achieved higher results in a day-to-day machine like a MacBook within an acceptable time frame.

EV13	Yes
EV14	Yes
EV15	Yes the approach you have taken forward to overcome the limitations of the state of the art is sufficient for an under graduate research.
EV16	Based on the information provided, it seems that this project has a good level of depth for an undergraduate research project in the field of machine learning and data augmentation.
EV17	I strongly believe yes

Table 59: Expert feedbacks for question 3

Question: Is the proposed solution approach well thought through in terms of design to accomplish its task?

Evaluator ID	Feedback
EV1	The tool design looks good, and the results also are promising.
EV2	When condensing the search space parameters to a single parameter, more details should be given in terms of experimental results to support the decisions taken on the fixed hyperparameters. More experimentation can be done to support the use of a neural network to determine the single parameter. For instance, what other estimators can be used in this setting instead of an NN and how does the results compare with NN and non-NN experiments on
EV3	yes
EV4	yes
EV5	Yes, the researcher has conducted testing and benchmarking to credit his claims.
EV6	Novelty of the solution is impressive as the solution does a significant improvement. Moreover, the solution has been evaluated and benchmarked to further prove the solidity of the research
EV7	yes, its well designed and has addressed its objectives
EV8	yes
EV9	yes, it's going to cut the requirement of domain expertise, And as data show it improved the accuracy too

EV10	Definitely, the author is trying to find a balance between the Pros like Simplicity, Efficiency, Robustness VS challenges like Limitations, Overfitting issues and lack of flexibility
EV11	Yes. I think this has been designed well from the start to finish .
EV12	yes
EV13	Yes. Seems the proposed solution is able to achieve state-of-the-art performance with low resource constraints.
EV14	The proposed solution seems to work and I don't see any literature that suggests that the proposed solution has been attempted prior to this. Therefore, I would agree that the approach to the solution has been well thought out.
EV15	As I went through the demonstration of the Web UI I'm currently in a state to confirm that the thought process has contributed to the Body of Knowledge(BoK) up to the expected standards.
EV16	Yes, The authors have provided a detailed description of the approach and have justified their design decisions through experimental evaluations.
EV17	i think yes but cannot provide a certain answer with regards to the design as a relatively limited amount of information is available in the document

Table 60: Expert feedbacks for question 4

Question: Based on the test results provided, does the system justify the importance of this research in real-world data augmentation scenarios?

Evaluator ID	Feedback
EV1	Not my primary area, but the evidence provided looks plausible. Would have been good to see more details on the experiments performed to evaluate the solution.
EV2	Yes
EV3	yes
EV4	I think the system can improve. But as the 1st version this justifies the impportance.
EV5	Yes, although more benchmarking could be done. The initial test results provide sufficient information to continue further research as well.
EV6	Yes, evaluation and benchmarked results are solid evidence that proposed solution can make significant contribution to the domain

EV7	Yes
EV8	Yes it does
EV9	yes, proposed system shows the highest accuracy in image classification
EV10	The project is able to achieve a pretty decent accuracy of 76% which is good as a starter and can be further enhanced and tested on a variety of datasets, it brings some value for the user to spend quality time on bigger issues of overfitting and flexibility to achieve better results.
EV11	Yes.Based on the evidence provided of the performance, i is evident that this can improve the applications of image data augmentations.
EV12	Yes. Once again, thinking from a customer perspective of a DA system, I would like to manage everything with little knowledge. The provided system achieves this. Further, the accuracies stated are higher than any other system that exists, hence giving it more merit in my opinion
EV13	Yes
EV14	It's difficult for me to think about how this can be mapped to real-world data augmentation scenraios. Maybe it's better to present this in the document/ presentation?
EV15	yes the test results and the respective elaborations on the document seems promising and could be attached to the BoK by means of publication to make it a progressive research area.
EV16	Based on the test results provided, it seems that the proposed system is effective in improving the performance of the classifier on the augmented data. The system achieved higher accuracy on the augmented datasets compared to the original dataset. This indicates that the proposed system has the potential to be useful in real-world data augmentation scenarios where improving the performance of classifiers on limited datasets is a common problem.
EV17	yes

Table 61: Expert feedbacks for question 5

Question: From your perspective, if you think the system can be improved in any way, how would it be?

Evaluator ID	**Feedback**
EV1	Present additional experimental results.

EV2	As mentioned in the improvements by the author, I believe if this solution was applied to other CV tasks such as segmentation, pose estimation it would have been helpful to strengthen the central thesis
EV3	The experiment was done on an M1 Pro CPU (which is a bit over-powered compared to Intel and AMD chips), so conducting tests and optimizing the HP tuning for other CPUs is also suggested.
EV4	I would recommend to improve the hyper-parameter neural network to get more accuracy and take advantage of cloud computing features and GPUs
EV5	Benchmark with more solutions. It will support your research and justify it further.
EV6	If the author of this project can support other computer vision tasks and more data types through this architecture that would be a really useful to many other researches
EV7	product can be enhanced to measure the effectiveness of other computer vision tasks including image segmentation
EV8	Maybe find a approach to find the optimal parameters in less time by optimizing parameter search using methods like information gain and look into algorithms to efficiently search for best hyperparameter combinations
EV9	It may help to improve performance of the system by using augmentation techniques like CutMix, MixUp and GAN(Generative Adversarial Network) may help to improve accuracy furthermore
EV10	A single hyper-parameter may not be sufficient to accurately represent the variations in different image transformations. This can result in the network being less effective at handling certain types of transformations. But the problem this project is solving will loose it's value if we move towards determining more hyperparameters, rather i would focus more on fine tuning the hyperparamters on variety of datasets for better accuracy and improve on training time.
EV11	As a future step I feel, this can be further tested with different hardware configurations as well. That will provide more clarity on the versatility of the model across hardware configurations. Other than that all other improvements for future have been identified by the student.
EV12	Reduction in training time could yield better commercial value

EV13	Try to address the following research question - How to select the most appropriate augmentation policies based on the domain? (Some of these augmentation policies won't work for some safety-critical domains).
EV14	I would like to know what percentage of difference would be made for certain augmentation actions. This would probably be another research.
EV15	For the moment I don't have any suggestions but I could guarantee that this could be a part of a progressive research area.
EV16	It would be useful to test the performance of the proposed method on different types of datasets to determine its scalability and versatility.
EV17	An open-source library as mentioned would be something that can definitely be benefited from

Table 62: Expert feedbacks for question 6

Question: Complexity wise, does the project provide adequate challenges to overcome considering it to be an undergraduate level research? (Please leave your thoughts on the technical difficulties as well)

Evaluator ID	Feedback
EV1	Yes, it looks to be of appropriate complexity.
EV2	I think this project provides an adequate level of depth and complexity for an undergraduate
EV3	Yes. The literature researched to reach the gap is commendable.
EV4	Yes. I think doing a machine learning based research project provided that this from a Software Engineering specialized student, it is complex enough.
EV5	The researcher has presented a thorough solution to a relevant problem. The research complexity is significant for a undergraduate level research.
EV6	Project domain itself complex and proposed solution is well thought and designed to achieve the best results with minimum resources. As I mentioned earlier this is an impressive piece of work for an undergraduate level research with a good level of contribution to the domain
EV7	Complex enough for an undergrad project, mainly it does not require high end specs
EV8	Yes

EV9	yes, it challenging to successfully applying a neural network base solution to resolve a actual problem, author has successfully achived it
EV10	The project aims to reduce the complexity of the image augmentation and training which is indeed a big problem and costly to achieve. With limited resources, it's a great outcome. There is definitely a decent scope of improvement but considering an undergrad level research, the approach and outcomes are very encouraging.
EV11	Yes. I believe that coming up with the solution based on research done on the theoretical aspects of the redefinition of search space had its own challenges and all those had been overcome to come up with the solution. Also, the contribution and the solution has been properly built as an end-to-end application as well. So I believe this research and the project has had adequate challenges to call it a "great undergraduate research".
EV12	Plenty of challenges to overcome. A project as such requires thorough planning. Given the time frame, I'm sure that Pubudu has faced and overcome many obstacles to output the work he has referred me to.
EV13	Yes.
EV14	It does provide adequate challenges as the machine learning pipeline for the automated augmentation process itself has been attempted to be changed and enhanced.
EV15	Drilling down till a single hyperparameter based neural network to find a solution for the persisted data augmentation related problem seems a suffice workaround on the technicality.
EV16	Yes, The project involves designing and implementing a novel technique for automated data augmentation, which requires a good understanding of machine learning concepts and algorithms, as well as programming skills to implement the technique in code. The technical difficulties involved in this project include selecting appropriate data augmentation techniques, designing and implementing the augmentation pipeline, and tuning the hyperparameters of the model. These tasks require a good understanding of the machine learning models being used, as well as experience with designing and implementing machine learning pipelines.

EV17	Yes i agree. One major challenge can be the hardware as everyone may not have sufficient computational power to carry out the procedure.

Table 63: Expert feedbacks for question 7

Question: Additional remarks or thoughts

Evaluator ID	**Feedback**
EV1	N/A
EV2	N/A
EV3	N/A
EV4	A good project.
EV5	N/A
EV6	This project has more potential to improve and investigate more academic avenues. I believe the author (student) of this project will take this research further to contribute more to the subject domain
EV7	Create an open source library so that community can contribute and use this
EV8	Good to see students touching these rare research areas
EV9	Interesting approach to addressed the issue in automated data augmentation field
EV10	N/A
EV11	N/A
EV12	Following along the project for months, he has been exceptional. Good job!
EV13	Try to improve the benchmarking section. Try to publish a review paper.
EV14	N/A
EV15	Great contribution to the BoK of the Computer Vision domain.
EV16	N/A
EV17	N/A

Table 64: Expert feedbacks for question 8