Comparing ResNet and Vision Transformers: Supervised vs Semi-Supervised Learning for Pet Breed Classification

Francesco Olivieri
KTH Royal Institute of Technology
olivieri@kth.se

Inês Mesquita
KTH Royal Institute of Technology
inesm@kth.se

Leandro Duarte
KTH Royal Institute of Technology
ldr0@kth.se

Abstract

This project undertakes a comparative study of ResNet50 and Vision Transformer (ViT) architectures for fine-grained pet breed classification on the Oxford-IIIT Pet Dataset. We apply and evaluate a range of techniques, including varied fine-tuning strategies (classifier-only, partial backbone, full backbone, and gradual unfreezing), data augmentation, and semi-supervised learning (SSL) via pseudo-labeling, to assess their impact on both model architectures. The investigation explores performance across binary (cat vs. dog) and multi-class (37 breeds) tasks, particularly examining robustness as the proportion of labeled data is reduced. Our findings indicate that while both ResNet50 and ViT achieve high performance on this dataset, ViT often achieves higher peak performance, though ResNet50 can be more robust under certain challenging conditions. This research emphasizes the practical implementation nuances and provides insights into the relative efficacy of these models and techniques in adapting to specialized visual recognition under varying data availability. The complete codebase for reproducing these experiments is available at `https://github.com/Leandr0Duar7e/kth-DD2424-project`.

1 Introduction

Image classification, assigning labels to images, is crucial in computer vision. However, training accurate models from scratch requires vast labeled data, which is costly to acquire. Transfer learning mitigates this by adapting pre-trained models to new tasks with less data. We explore this for pet breed classification using the Oxford-IIT Pet Dataset, a benchmark for fine-tuning.

Our report compares ResNet50 (a CNN) against Vision Transformer (ViT, Hugging Face implementation). Addressing data scarcity, we extend our analysis to semi-supervised learning (SSL). SSL leverages abundant unlabeled data alongside limited labeled examples. We implement pseudo-labeling and evaluate ResNet50 and ViT with progressively reduced labeled training data. This assesses their robustness and SSL efficacy in data-constrained regimes. Our findings aim to clarify how these architectures and strategies perform on specialized visual recognition tasks.

2 Related Work

Recent studies have explored the transferability and effectiveness of visual representations from CNNs and Transformers. Raghu et al.[1] compared ConvNets and Vision Transformers, showing that Transformers can outperform CNNs in transfer learning tasks, although they typically require more data. Similarly, research comparing CNN, ResNet, and Vision Transformers for chest disease classification[2] found that Transformers often achieved higher accuracy, underscoring their potential in complex image recognition tasks. However, these studies primarily focus on fully supervised learning and are often restricted to specific domains such as medical imaging. This leaves a gap in evaluating these architectures under semi-supervised conditions, particularly in more general-purpose tasks. Additionally, comparisons are frequently limited to multi-class classification, overlooking binary classification scenarios. To address these gaps, our work systematically compares ResNet and Vision Transformer architectures in both supervised and semi-supervised settings, across binary and multi-class classification tasks using the Oxford-IIIT Pet dataset. The semi-supervised learning component is developed using the pseudo-labeling approach proposed by Lee et al.[3], enabling us to evaluate how well these models perform when labeled data is limited.

3 Data

The project utilizes the Oxford-IIIT Pet Dataset [4], a benchmark for fine-grained visual classification. It contains 7,349 images of 37 pet breeds, with around 200 images per class. The dataset features significant variations in scale, pose, and lighting. Annotations include breed, head Region of Interest (ROI), and pixel-level trimap segmentations. For all experiments, images are resized to 224x224 pixels and normalized. Data augmentation techniques such as random horizontal flips and rotations are applied in specific experiments. Vision Transformer (ViT) models utilize specific preprocessing steps via the Hugging Face AutoImageProcessor. The dataset is consistently split into training, validation, and test sets. For the semi-supervised learning (SSL) experiments, the proportion of labeled data in the training set was systematically reduced, with the remainder treated as unlabeled data to assess model performance under data scarcity. Given that the Oxford-IIIT Pet Dataset is a standard benchmark, various methods have been evaluated on it. State-of-the-art results are often achieved by Transformer-based models; for instance, fine-tuned Vision Transformers have reported accuracies around 94% [5]. Other competitive approaches include specialized transformer architectures like OmniVec2 [6]. Zero-shot learning with models like CLIP has also demonstrated strong performance, achieving up to 88% accuracy without dataset-specific fine-tuning [7].

4 Methods

This section details the methodologies employed to compare ResNet50 and Vision Transformer (ViT) architectures. Our approach is rooted in transfer learning, leveraging pre-trained models to adapt to the specific classification tasks. We investigate two primary learning paradigms: fully supervised training utilizing all available labels, and semi-supervised learning (SSL) through pseudo-labeling to assess model performance under conditions of reduced label availability.

4.0.1 ResNet50

ResNet50 [8] (Fig. 1) processes 224x224 RGB images. It starts with a convolutional layer, batch normalization, ReLU, and max pooling. The core comprises four stages of residual blocks (convolutional layers with batch norm, ReLU, and skip connections to mitigate vanishing gradients). It concludes with global average pooling and a final dense layer. We used a ResNet50 pre-trained on ImageNet, leveraging its learned features for our smaller dataset.

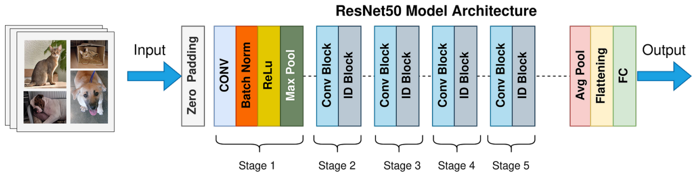

Figure 1: ResNet50 Architecture.

For our experiments, we used a pre-trained ResNet-50 model, originally trained on the ImageNet dataset [9]. To adapt the model to our specific classification tasks, we replaced the original final fully connected layer with a new dense layer tailored to the desired number of output classes. In the binary classification task (e.g., distinguishing between cats and dogs), the final layer was modified to output a single neuron with a sigmoid activation function. For the multi-class classification task (e.g., identifying 37 different breeds of cats and dogs), we used a dense layer with 37 output neurons and a softmax activation function to model the probability distribution over the classes. We fine-tuned the entire model or, in some experiments, froze the earlier layers and only trained the modified classifier head. This allowed us to evaluate the benefit of task-specific tuning versus using fixed pre-trained features. During training, we used categorical cross-entropy for the multi-class case and binary cross-entropy for the binary case. Optimization was performed using the Adam optimizer [10], coupled with a cosine annealing learning rate schedule with warm-up. Additionally, data augmentation techniques were applied to increase robustness and help generalize better to unseen examples.

4.1 ViT

For ViT, we used `google/vit-base-patch16-224` from Hugging Face [11], pre-trained on ImageNet-21k and fine-tuned on ImageNet-1k. It processes 224x224 images into 16×16 patches (Fig. 2).

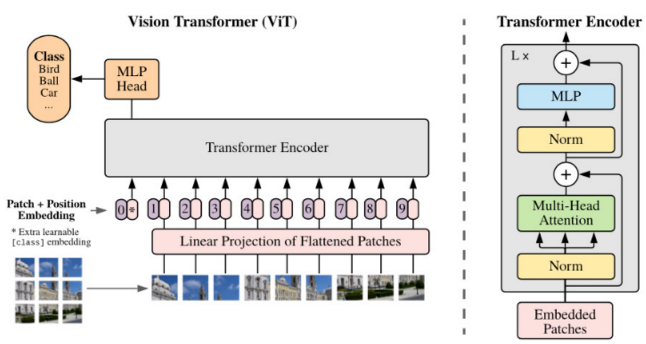

Figure 2: ViT Architecture.

Input images were preprocessed by Hugging Face's `AutoImageProcessor` (resizing to 224x224, model-specific normalization). No further augmentation was used for supervised ViT experiments. The pre-trained ViT was adapted by replacing its classifier head (1 output for binary, 37 for multi-class). For multi-class, we explored two strategies: (1) Unfreezing a fixed number of final encoder layers (0, 1, 3, 6, 12, or entire backbone). (2) Gradual unfreezing, starting with the classifier head and progressively unfreezing deeper layers. ViT models used the Adam optimizer [10]. For supervised learning, LR was 5×10^{-5} (binary), and 5×10^{-5} (multi-class), adjusted to 3×10^{-5} or 1×10^{-5} for more unfrozen layers. Training was typically 2 epochs, batch size 32, using binary or multi-class cross-entropy loss.

4.2 Semi-Supervised Learning

For our semi-supervised experiments, we adopted the pseudo-labeling technique introduced by Lee (2013) [3], which uses a model's own high-confidence predictions on unlabeled data to augment the labeled dataset. First, the model (either ResNet50 or ViT, using the best configuration) was trained on a reduced subset of labeled data (50%, 10%, or 1%) to establish a supervised baseline. Then, it generated pseudo-labels for the remaining unlabeled samples. These pseudo-labeled samples were added to the labeled set, and the model was fine-tuned on this combined dataset. This continued training with pseudo-labels follows the method proposed by Lee and aims to improve generalization. Performance was finally evaluated on the held-out test set.

The process was applied independently for both ResNEt50 and ViT architectures across the different percentages of labeled data explored.

4.3 Imbalanced Class

To investigate the impact of class imbalance on fine-tuning performance, an experiment was conducted where the training dataset was intentionally imbalanced. This was achieved by uniformly reducing the number of training images to 20% of the original set for each cat breed, thereby creating a scenario with limited data per class. The model was initially fine-tuned using a standard cross-entropy loss function, and test performance on classes with this reduced data was specifically evaluated. Subsequently, strategies to mitigate the effects of this imbalance, namely weighted cross-entropy and oversampling of the minority (or underrepresented) classes, were implemented and their impact on final test performance was assessed.

4.4 Codebase

The project was implemented in Python, leveraging libraries such as PyTorch, Hugging Face Transformers, and Scikit-learn. The codebase is structured modularly, with key components including 'src/main.py' for experiment orchestration via a command-line interface, 'src/dataset.py' for data loading and preprocessing (including specific handling for ResNet50 and ViT, and semi-supervised splits), model definitions within 'src/models/', 'src/trainer.py' for managing the training loops (including pseudo-labeling logic and gradual unfreezing for ResNet), and 'src/evaluation.py' for performance assessment and results visualization. The complete source code is publicly available on GitHub [12].

5 Experiments

In this section we will compare the performances achieved by our fine-tuned ResNet50 and ViT models in two different settings: fully-supervised and semi-supervised learning. Metrics taken into account are: Test Accuracy, Training Accuracy, Validation Accuracy and AUC.

5.1 Fully-Supervised Learning

In the fully-supervised learning setting, models were trained using the entire available labeled portion of the Oxford-IIIT Pet Dataset. This approach serves as a baseline to evaluate the maximum performance achievable with complete label information for both ResNet50 and ViT architectures across the defined tasks.

5.1.1 Binary Classification

For this task, both networks were trained to distinguish between cats and dogs. Initial experiments with single-epoch training identified optimal learning rates: 0.01 for ResNet50 and 5×10^{-5} for ViT. Both models were then trained for 2 epochs using these optimal rates, achieving excellent performance as shown in Table 1.

Table 1: Models performances for fully supervised binary classification

Model	Test Acc. (%)	Train Acc. (%)	Validation Acc. (%)	AUC	Weighted f1
ResNet50	99.59	99.98	99.59	0.9999	0.9959
ViT	99.73	99.27	99.73	0.9998	0.9973

Both architectures surpassed the 99% accuracy target with comparable test performance, though ViT exhibited higher training and validation accuracies than ResNet50, suggesting potentially better generalization capabilities.

5.1.2 Multi-Class Classification

For this task, the aim was to classify the breeds of cats and dogs. After fine-tuning, the performances achieved are reported in Table 4. The Vision Transformer (ViT) generally outperformed ResNet50 in key metrics like test accuracy and AUC, indicating better generalization and class-wise performance. Additionally, ViT showed less overfitting, as reflected by its closer train-test accuracy gap. Both models reached high validation accuracy, but ViT exhibited stronger robustness and balanced performance across metrics.

Table 2: Models performances for fully supervised multi-class classification

Model	Test Acc. (%)	Train Acc. (%)	Validation Acc. (%)	OvR AUC	Weighted f1
ResNet50	94.70	99.23	94.27	0.947	0.9991
ViT	95.11	98.81	94.41	0.9995	0.9512

5.1.3 Imbalanced Classes on the Multi-Class Classification

The following tables report the performance of ResNet50 and ViT under class imbalance, where only 20% of each cat breed was retained. As shown, both models benefited from class rebalancing strategies. Over-sampling yielded strong results for both, with ResNet50 achieving higher accuracy (90.90%) with this method. Weighted loss also proved effective for both models, especially compared to using standard cross-entropy.

Table 3: Performance of the ResNet50 on imbalanced data

Method	Test Acc. (%)	Train Acc. (%)	Validation Acc. (%)	OvR AUC	Weighted f1
Normal Cross-Entropy	77.85	92.50	77.50	0.9964	0.7790
Weighted Cross-Entropy	80.40	97.80	81.00	0.9970	0.7850
Over-sampling	90.90	98.73	92.50	0.9810	0.8700

Table 4: Performance of the ViT on imbalanced data

Method	Test Acc. (%)	Train Acc. (%)	Validation Acc. (%)	OvR AUC	Weighted f1
Normal Cross-Entropy	87.77	99.63	88.96	0.9979	0.8740
Weighted Cross-Entropy	88.72	99.50	90.87	0.9981	0.8840
Over-sampling	88.32	99.11	89.78	0.9983	0.8791

5.2 Semi Supervised Learning

In the semi-supervised learning setting, only a fraction of the labeled data from the Oxford-IIIT Pet Dataset was used for training, while the remaining unlabeled data was incorporated through pseudo-labeling. This approach enables evaluation of how well ResNet50 and ViT architectures can generalize with limited annotated data, providing insights into their robustness and effectiveness in data-scarce scenarios.

5.2.1 Binary Classification

For this task, we tested the best configuration obtained during supervised training in a semi-supervised setting. As shown in Table 5, ResNet50 demonstrated superior performance when only a small percentage of labeled data was available (1% and 10%). However, as the amount of labeled data increased to 50%, ViT outperformed ResNet across all evaluation metrics. Notably, ViT achieved the highest AUC and weighted F1-score, confirming its strong learning capacity when provided with sufficient supervision.

Table 5: Models performances for semi supervised binary classification

Model	Test Acc.(%)	Train Acc.(%)	Validation Acc.(%)	AUC	Weighted f1	Lab. Data(%)
ResNet50	98.64	99.63	97.82	0.9988	0.9863	1
ResNet50	99.59	99.96	99.05	0.9996	0.9959	10
ResNet50	99.59	99.93	99.59	0.9997	0.9959	50
ViT	41.71	87.33	42.59	0.2997	0.4304	1
ViT	85.32	81.68	83.81	0.9302	0.8534	10
ViT	99.86	99.48	99.86	0.9999	0.9986	50

5.2.2 Multi-Class Classification

Table 6 reports the semi-supervised multi-class classification results using ResNet50 and ViT across varying percentages of labeled data. While both models improve as more labeled data becomes available, ViT performs poorly under extreme low-label conditions (1%), whereas it surpasses ResNet50 at higher label percentages, following the same trend showed in binary classification.

Table 6: Models performances for semi supervised multi-class classification

Model	Test Acc.(%)	Train Acc.(%)	Validation Acc.(%)	AUC	Weighted f1	Lab. Data(%)
ResNet50	27.03	100.00	29.38	0.8127	0.2265	1
ResNet50	76.63	96.01	75.23	0.9911	0.7654	10
ResNet50	92.25	97.26	90.01	0.9984	0.9233	50
ViT	13.04	94.13	13.61	0.7100	0.1070	1
ViT	81.50	93.50	82.14	0.9240	0.8474	10
ViT	94.20	99.20	93.80	0.9989	0.9410	50

5.3 Ablation Studies

In this section, we explore the effects of fine-tuning strategies, learning rate configurations, data augmentation, and regularization techniques. Our goal is to highlight how each factor contributed to the overall performance and to explain the choices that led to our best-performing models.

5.3.1 ResNet50

We performed extensive ablation studies on ResNet50 to understand the impact of various fine-tuning strategies, learning rates, data augmentation, and L2 regularization on classification performance. Our initial approach involved unfreezing a fixed number of layers beyond the final fully connected (fc) head. Results showed that unfreezing only a few top layers while using higher learning rates (e.g., 5×10^{-4}) led to modest performance gains. However, as deeper layers were unfrozen, higher learning rates became detrimental, leading to unstable training and overfitting. For example, training the last 8 layer with a learning rate of 1×10^{-3} retrieved a test accuracy of 74.86%, while training with 9 layers with a lower learning rate (5×10^{-5}) yielded a test accuracy of 93.88%, but also showed increased variance between training and validation accuracy. Subsequently, we experimented with gradual unfreezing, starting from the fc head and progressively unfreezing deeper layers during training. This method proved more effective, improving test accuracy to 94.42% while also reducing overfitting, as evidenced by a more stable training-validation accuracy gap. To further enhance generalization, we introduced data augmentation and L2 regularization. Applying data augmentation in combination with gradual unfreezing and a layer-specific differential learning rate strategy produced our best result:

a test accuracy of 94.70%, with training and validation accuracy remaining closely aligned (99.23% and 94.28%, respectively). In contrast, using L2 regularization in isolation (e.g., $\lambda = 10^{-3}$) did not consistently yield improvements and sometimes negatively impacted performance, particularly when combined with high learning rates.

We also tested a differential learning rate schedule across all layers, using smaller learning rates for earlier layers and larger ones for later ones. While this method performed well (e.g., 94.02% test accuracy without augmentation), it still fell short of the combined benefit offered by gradual unfreezing with augmentation. Overall, the results indicate that careful management of the fine-tuning depth, combined with selective regularization and data augmentation, can substantially improve performance. Notably, a well-balanced strategy involving gradual unfreezing, moderate learning rates, and augmentation provided the optimal trade-off between adaptation and overfitting mitigation.

5.3.2 ViT

For the Vision Transformer, ablation studies focused on the impact of unfreezing different numbers of encoder layers, fine-tuning strategies, data augmentation, and regularization on the multi-class (37 breeds) classification task. Our investigation began by training only the randomly initialized classifier head, keeping the entire pre-trained ViT backbone frozen, which yielded a baseline test accuracy of 85.87% with a learning rate of 5×10^{-5}. Progressively unfreezing more encoder layers resulted in significant performance gains: the last 1 layer (92.53%), 3 layers (94.29%), and peaking at 6 layers (95.11%). Further unfreezing proved counterproductive, with 12 layers (93.34%) and the entire backbone (91.44%) showing diminishing returns despite reduced learning rates. We compared two fine-tuning strategies: unfreezing a fixed number of layers from the start (Strategy 1) versus gradual unfreezing during training (Strategy 2). Strategy 1 with 6 unfrozen layers consistently outperformed gradual unfreezing (95.11% vs. 93.48%) under similar conditions, suggesting that for this dataset, a carefully selected fixed depth of fine-tuning is more effective than progressive adaptation.

Data augmentation experiments showed mixed results. While augmentation improved performance for certain configurations (e.g., from 93.21% to 94.29% for the 6-layer model with 3 epochs), it couldn't surpass our overall best performance of 95.11% achieved without augmentation. Interestingly, L2 regularization ($\lambda = 10^{-4}$ and $\lambda = 10^{-3}$) consistently reduced performance across configurations, suggesting that the ViT architecture with its inherent self-attention mechanisms may already provide sufficient regularization for this dataset. The model's performance was more sensitive to the depth of fine-tuning than to traditional regularization techniques, emphasizing the importance of careful architecture-specific transfer learning strategies.

6 Conclusion

Both ResNet50 and ViT surpassed the 99% accuracy target for binary classification and achieved strong performance (95%) on multi-class tasks. ViT consistently outperformed ResNet50 in fully supervised settings. Under class imbalance, both models suffered performance drops with standard cross-entropy loss, but these were mitigated by weighted loss or over-sampling, with varied model-specific success. In semi-supervised learning, performance declined with reduced labeled data, with ResNet50 outperforming ViT in low-data regimes, where ViT required more careful tuning. Ablation studies identified optimal strategies: gradual unfreezing with augmentation for ResNet50, and unfreezing six encoder layers for ViT. Overall, ViT proved highly effective for fine-grained visual recognition in supervised scenarios but showed limitations in semi-supervised settings.

Computational constraints prevented accurate training time analysis. Future work should explore more complex datasets, extensive hyperparameter optimization, code efficiency improvements, and more techniques such as early stopping and AdamW optimizer.

References

[1] Hong-Yu Zhou, Chixiang Lu, Sibei Yang, and Yizhou Yu. Convnets vs. transformers: Whose visual representations are more transferable? In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 2230–2238, 2021. doi: 10.1109/ICCVW54120.2021.00252.

[2] Ananya Jain, Aviral Bhardwaj, Kaushik Murali, and Isha Surani. A comparative study of cnn, resnet, and vision transformers for multi-classification of chest diseases. 05 2024. doi: 10.48550/arXiv.2406.00237.

[3] Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICML Workshop on Challenges in Representation Learning (WREPL)*, volume 3, page 896. PMLR, 2013.

[4] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, and C.V. Jawahar. Cats and dogs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3498–3505, 2012.

[5] norburay. Model card for norburay/vit-base-oxford-iiit-pets. `https://huggingface.co/norburay/vit-base-oxford-iiit-pets`.

[6] Sparsh Srivastava, Hao Fan, Lisha Devi, Cihang Xu, Abhinav Shrivastava, Danail Stoyanov, Saining Liu, and Christoph Feichtenhofer. OmniVec2: A Novel Transformer based Network for Large Scale Multimodal Learning and Perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[7] muellje3. Model card for muellje3/vit-base-oxford-iiit-pets. `https://huggingface.co/muellje3/vit-base-oxford-iiit-pets`.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

[10] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. URL `https://arxiv.org/abs/1412.6980`.

[11] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October 2020. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/2020.emnlp-demos.6`.

[12] Francesco Olivieri, Inês Mesquita, and Leandro Duarte. Github repo for kth-dd2424-project. `https://github.com/Leandr0Duar7e/kth-DD2424-project`, 2025.

Appendix provides data from experimental runs conducted with both ViT and ResNet50 architectures across supervised and semi-supervised learning scenarios.

model_filename	classification_typ	epochs	learning_rate	l2_lambda	training_time_se	test_accuracy	test_loss	final_train_loss	final_val_loss	final_train_acc	final_val_acc	weighted_f1_sco	roc_auc	roc_auc_ovr_wei
vit_binary_1ep_lr	binary	1	5,00E-05	0.0		0.985054347826	0.337755170853	0.341816652890	0.339678694372	98.89436979078	98.77384196185	0.985062105812	0.9973288591724472	
vit_binary_1ep_lr	binary	1	3,00E-05	0.0		0.794836956521	0.531240822180	0.531198755230	0.536734268717	79.77547201905	79.97275204359	0.800934331691	0.9318692142672309	
vit_binary_1ep_lr	binary	1	1,00E-05	0.0		0.762228260869	0.566874339528	0.574617722760	0.580633471841	74.58751488348	74.93188010899	0.754909295372	0.8331121294835516	
vit_binary_2ep_lr	binary	2	5,00E-05	0.0		0.997282608695	0.195441560252	0.201934062025	0.195148691534	99.26858309236	99.72752043596	0.997279744646	0.9997996644379336	
vit_multiclass_2e	multiclass	2	5,00E-05	0.0	456.6657419204	0.858695652173	1.257372964983	1.229083021049	1.244051456451	85.03146793672	83.65122615803	0.8549446999383195		0.995027075774
vit_multiclass_2e	multiclass	2	5,00E-05	0.0	485.0932288169	0.925271739130	0.339024852151	0.302794213852	0.348448704766	94.64194590916	92.09809264305	0.925021789215326		0.999182201533
vit_multiclass_2e	multiclass	2	5,00E-05	0.0	548.8975982666	0.942934782608	0.213808165944	0.135356764693	0.210966463322	97.56761353971	95.50408719346	0.9432880869721307		0.999412307284
vit_multiclass_2e	multiclass	2	5,00E-05	0.0	629.1820299625	0.951086956521	0.199718455257	0.083000096290	0.209956621994	98.80932131314	94.41416893732	0.9512108058115862		0.999491531898
vit_multiclass_2e	multiclass	2	3,00E-05	0.0	767.9399578571	0.933423913043	0.271430553301	0.124696554312	0.277815774083	98.79231161762	94.27792915531	0.9331342519066709		0.999286193187
vit_multiclass_2e	multiclass	2	1,00E-05	0.0	771.5634467601	0.914402173913	0.711682076039	0.590160798119	0.724566135717	95.62850824970	91.96185286103	0.9135854786445201		0.998494106678
vit_multiclass_2e	multiclass	2	5,00E-05	0.0	600.4495110511	0.934782608695	0.232510134901	0.104583576726	0.235198547334	98.41809831604	93.73297002724	0.9343483101765008		0.999257125374
vit_multiclass_3e	multiclass	3	5,00E-05	0.0	870.1967208385	0.925271739130	0.216625232735	0.067409843965	0.222550492895	98.94539887736	93.86920980926	0.9245295503008136		0.999369087330
vit_multiclass_2e	multiclass	2	5,00E-05	0.0	594.5294334888	0.940217391304	0.187884376262	0.048207169813	0.228574917044	98.92838918183	93.05177111716	0.9400651444291899		0.999347908965
vit_multiclass_3e	multiclass	3	5,00E-05	0.0	845.8184728622	0.932065217391	0.207720029046	0.015953374347	0.197376653066	99.79588365368	93.86920980926	0.9318303038407272		0.999263488163
vit_multiclass_2e	multiclass	2	5,00E-05	0.0	631.9706645011	0.932065217391	0.217272568818	0.056780688672	0.236915611702	98.57118557577	92.64305177111	0.93212657770518		0.999203936079
vit_multiclass_3e	multiclass	3	5,00E-05	0.0	1240.409925460	0.936141304347	0.190809348841	0.023068495370	0.207449088601	99.67681578499	93.18801089918	0.9358374703247874		0.999318157274
vit_multiclass_3e	multiclass	3	5,00E-05	0.0	1078.821066856	0.942934782608	0.171426876675	0.053306161107	0.238593173091	98.52015648919	93.46049046321	0.9429321845200317		0.999354671464
vit_multiclass_3e	multiclass	3	5,00E-05	0.0	1116.566211462	0.923913043478	0.237099382540	0.068305969040	0.206035030762	98.28202075182	93.46049046321	0.9243349263927362		0.999097578139
vit_multiclass_5e	multiclass	5	5,00E-05	0.0	1747.357209205	0.942934782608	0.176500462643	0.017482481775	0.213323906711	99.71083517605	93.05177111716	0.9429061686359708		0.999415503811
vit_multiclass_3e	multiclass	3	5,00E-05	0.0	1206.507048845	0.929347826086	0.221247707534	0.046388283898	0.229759561140	98.94539887736	93.18801089918	0.9294605261391307		0.999138368756
vit_multiclass_5e	multiclass	5	5,00E-05	0.0	1788.352621078	0.938858695652	0.206278132193	0.026522843232	0.220635841722	99.45568974315	93.59673024523	0.9390355418762522		0.999150704757
vit_multiclass_5e	multiclass	5	5,00E-05	0.0001	1879.807774305	0.933423913043	0.208554516424	0.017765661211	0.223881568192	99.64279639394	93.59673024523	0.933479696642185		0.999287529126
vit_multiclass_2e	multiclass	2	5,00E-05	0.0001	639.9130232334	0.936141304347	0.213598298313	0.086412639984	0.209383868004	98.70726313998	94.68664850136	0.9361627603674825		0.999390325516
vit_multiclass_2e	multiclass	2	5,00E-05	0.001	885.8352787494	0.940217391304	0.222819740681	0.102090594378	0.225744704837	98.52015648919	94.41416893732	0.9402015599588881		0.999274654836
vit_binary_2l+2c_	binary	2	5,00E-05	0.0		0.417119565217	0.899567451166	0.446652680473	0.906845445218	87.32777683279	42.58503401360	0.430400910209	0.29971034816651226	
vit_binary_2l+2c_	binary	2	0.001	0.0		0.930706521739	0.146913427049	0.112240759696	0.146189496892	95.76458581391	93.33333333333	0.931806808248	0.9934640522875817	
vit_binary_2l+2c_	binary	2	5,00E-05	0.0		0.853260869565	0.419299599917	0.433793655718	0.433823554412	81.68055791801	83.80952380952	0.853412439372	0.9302164458801826	
vit_binary_2l+2c_	binary	2	5,00E-05	0.0		0.998641304347	0.113502435710	0.118973129712	0.121078314988	99.48970913420	99.59183673469	0.998640591926	0.999991652684914	
vit_multiclass_2l-	multiclass	2	5,00E-05	0.0		0.130434782608	4.125751039256	0.535290936574	3.852245361908	94.13165504337	13.60544217687	0.1069933336580311		0.709913218335
vit_multiclass_2e	multiclass	2	5,00E-05	0.0		0.877717391304	0.430122174646	0.070100732161	0.399474432934	99.63344788087	88.96457765667	0.8739562959546915		0.997946531547
vit_multiclass_2e	multiclass	2	5,00E-05	0.0		0.887228260869	0.403482050999	0.068965702623	0.380898695277	99.49599083619	90.87193460490	0.8840138047176765		0.998107860564

model_filename	classification_typ	epochs	learning_rate	l2_lambda	training_time_se	test_accuracy	test_loss	final_train_loss	final_val_loss	final_train_acc	final_val_acc	weighted_f1_sco	roc_auc	roc_auc_ovr_wei
resnet_multiclass	multiclass	6	[0.01, 0.0001, 4e	0.001	1008.275383472	0.9375	0.206106164695	0.015476078483	0.165872933505	99.69382548052	95.50408719346	0.9376297746312321		0.999126668579
resnet_multiclass	multiclass	6	[0.01, 0.0001, 4e	0.001	1069.263246536	0.936141304347	0.205417289196	0.01727023711	0.186440296714	99.60877700289	94.68664850136	0.9361525670329555		0.999249142141
resnet_multiclass	multiclass	6	[0.01, 0.0001, 4e	0.001	1034.328715562	0.932065217391	0.212941778580	0.017480777814	0.194991984769	99.57475761183	94.00544959128	0.9322678668464883		0.999281303718
resnet_multiclass	multiclass	6	[0.01, 0.0001, 4e	0.003	0.947074298777	0.947010869565	0.195915657622	0.999100280495	0.184090585407	99.23456370130	94.27792915531	0.9470742987777875		0.999100280495
resnet_multiclass	multiclass	2	0.001	0.0	285.7998514175	0.915760869565	0.367858919112	0.253729201124	0.359651950390	96.27487667970	92.09809264305	0.9148169853611283		0.998750758403
resnet_multiclass	multiclass	2	1,00E-05	0.0	277.8032157421	0.135869565217	3.453086448752	3.444427788257	3.449614255324	16.60146283381	14.44141689373	0.10508690150170684		0.796019985645
resnet_multiclass	multiclass	2	0.001	0.0	309.5739302635	0.889945652173	0.356692593706	0.095935085036	0.352584689207	96.83619663208	87.46594005449	0.888487200052503		0.998031064581
resnet_multiclass	multiclass	2	1,00E-05	0.0	434.4331190586	0.831521739130	1.816355083299	1.753346624581	1.860804682192	87.48086409253	83.24250681198	0.8265823374248435		0.991680427901
resnet_multiclass	multiclass	2	0.001	0.0	434.6632308959	0.81114130434	0.578985983262	0.229000403829	0.589535749476	93.06004422520	80.38147138964	0.8112417896686309		0.995576769532
resnet_multiclass	multiclass	2	1,00E-05	0.0	405.0447447299	0.84375	1.897790478623	1.760792889024	1.894743779431	88.29732947780	86.37602179836	0.8391536280877914		0.992871853105
resnet_multiclass	multiclass	2	0.001	0.0	465.9927513599	0.638586956521	1.359628068364	0.842033613149	1.263302844503	76.20343595849	65.53133514986	0.6362734937807454		0.979092792823
resnet_multiclass	multiclass	2	1,00E-05	0.0	903.7435579299	0.846467391304	1.703909454138	1.580998715499	1.722765518271	88.16125191359	84.19618528610	0.8405159328197932		0.993764671767
resnet_multiclass	multiclass	2	0.001	0.0	558.2714419364	0.748641304347	0.851469148760	0.408191649445	0.755262779152	87.10665079095	76.43051771117	0.7417001598071016		0.991694855213
resnet_multiclass	multiclass	2	1,00E-05	0.0	502.9951326847	0.84375	1.725607555845	1.639240476748	1.783440760944	88.36536825990	81.88010899182	0.8372949009530712		0.993668255908
resnet_multiclass	multiclass	2	5,00E-05	0.0	500.8167610168	0.938858695652	0.223480982948	0.076969822190	0.227794830241	98.80932131314	92.37057220708	0.9390645046984759		0.999285594643
resnet_multiclass	multiclass	2	5,00E-05	0.003	407.4568822383	0.940217391304	0.234308369781	0.108653052527	0.245567613969	98.26501105630	93.46049046321	0.9399933791198505		0.999331386588
resnet_multiclass	multiclass	2	5,00E-05	0.0	431.3884816169	0.944293478260	0.235102309480	0.106372101758	0.249017526274	98.43510801156	93.18801089918	0.9444016209358664		0.999238711405
resnet_multiclass	multiclass	2	[0.005, 0.0001, 4	0.0	404.4519326686	0.933423913043	0.205888778457	0.020462991163	0.204720200727	99.65980608947	93.46049046321	0.9330050167230814		0.999209456616
resnet_multiclass	multiclass	2	[0.001, 0.0005, 0	0.0	382.3491680622	0.907608695652	0.286624157882	0.057279217608	0.235515913237	98.75829222656	91.55313351498	0.9071844763777056		0.998756855970
resnet_multiclass	multiclass	2	[0.001, 0.0006, 0	0.0	367.1065328121	0.892663043478	0.333445282733	0.107091667424	0.334027662225	97.60163293077	88.96457765667	0.8898095157387281		0.997860974726
resnet_multiclass	multiclass	2	[0.005, 0.0005, 0	0.0	373.6591541767	0.898097826086	0.292890527974	0.036856711495	0.317367035085	99.01343765946	90.46321525885	0.898280774812606		0.998696522975
resnet_multiclass	multiclass	2	[5e-05, 5e-05, 5e	0.0	371.9603357315	0.9375	0.230065731898	0.105068815851	0.242494217403	98.06089470998	93.18801089918	0.9375641961692358		0.999408587956
resnet_multiclass	multiclass	2	[0.0001, 9e-05, 8	0.0	448.0752220153	0.940217391304	0.171678098161	0.040576496196	0.184261887617	99.25157339683	94.41416893732	0.9397885234233923		0.999513823045
resnet_multiclass	multiclass	2	[0.0001, 9e-05, 8	0.0	383.6641261577	0.929347826086	0.235081381771	0.062406535938	0.192964492608	98.82633100867	95.23160762942	0.9293326601793875		0.998907944858
resnet_multiclass	multiclass	2	[0.0001, 9e-05, 8	0.001	482.6528890132	0.932065217391	0.220858677573	0.099882232541	0.195786427544	97.44854567103	93.46049046321	0.9318670786482058		0.999046181988
resnet_multiclass	multiclass	2	[0.0001, 9e-05, 3	0.0008	467.9984636306	0.932065217391	0.205862572497	0.108178451191	0.207603690416	97.17639054260	93.05177111716	0.9313980987502323		0.999390686225
resnet_multiclass	multiclass	2	[0.0001, 9e-05, 3	0.0008	468.0644845962	0.927989130434	0.225141825883	0.096903173928	0.205035537481	97.07433236945	92.91553133514	0.9281564501044446		0.998877826517
resnet_multiclass	multiclass	2	[0.0001, 9e-05, 3	0.0008	394.1562347412	0.927989130434	0.215750034412	0.092803699521	0.191165366898	97.29545841129	94.27792915531	0.9279188861138856		0.999022369390
resnet_multiclass	multiclass	2	[0.0001, 9e-05, 3	0.0008	444.7331378459	0.910326086956	0.249941927583	0.163497582204	0.242434916936	94.70998469127	91.28065395095	0.9106598296298916		0.998711221280
resnet_multiclass	multiclass	5	[0.0001, 9e-05, 3	0.0008	805.9901645183	0.930706521739	0.189837125656	0.046616660935	0.179857096594	98.52015648919	94.41416893732	0.9306761025575447		0.999374271155
resnet_binary_2	binary	2	0.01		265.1771619319	0.995923913043	0.012873513851	0.003450954485	0.018622033176	99.98299030447	99.59128065395	0.995930196509	0.9999666107396556	
resnet_binary_2l	binary		0.01			0.986413043478	0.053631805202	0.016098870956	0.076884584501	99.62578669841	97.82312925170	0.986338530959	0.9988898070935484	
resnet_binary_2l	binary		0.01			0.995923913043	0.018623011708	0.004047381791	0.022236357431	99.96598060894	99.04761904761	0.995926028858	0.9996994966569004	
resnet_binary_2l	binary		0.01			0.995923913043	0.015862262873	0.002757639321	0.011030470285	99.93196121789	99.59183673469	0.995926028858	0.9997495805474169	
resnet_multiclass	multiclass		[0.001, 0.0005, 0	0.0		0.270380434782	10.37149199195	0.002633005690	7.807003591371	100.0	29.38775510204	0.22653170486981475		0.812751815540
resnet_multiclass	multiclass		[0.001, 0.0005, 0	0.003		0.766304347826	0.783383101224	0.207114684597	0.810335375692	96.01973124681	75.23809523809	0.7654498951229916		0.991155114626
resnet_multiclass	multiclass		[0.001, 0.0005, 0	0.003		0.922554347826	0.247175956549	0.127785615972	0.378829691721	97.26143902024	88.29931972789	0.9233557881670912		0.998455510959
resnet_multiclass	multiclass	1	0.0001	0.0	127.2504732608	0.516304347826	3.145501976427	2.839227547610	3.146930850070	72.80641466208	51.63487738419	0.4375816171909594		0.930968749552
vit_binary_2l+2c	binary		5,00E-05	0.0		0.417119565217	0.899567451166	0.446652680473	0.906845445218	87.32777683279	42.58503401360	0.430400910209	0.29971034816651226	
vit_binary_2l+2c	binary		0.001	0.0		0.930706521739	0.146913427049	0.112240759696	0.146189496892	95.76458581391	93.33333333333	0.931806808248	0.9934640522875817	
vit_binary_2l+2c	binary		5,00E-05	0.0		0.853260869565	0.419299599917	0.433793655718	0.433823554412	81.68055791801	83.80952380952	0.853412439372	0.9302164458801826	
vit_binary_2l+2c	binary		5,00E-05	0.0		0.998641304347	0.113502435710	0.118973129712	0.121078314988	99.48970913420	99.59183673469	0.998640591926	0.999991652684914	
resnet_multiclass	multiclass	2	5,00E-05	0.0	335.3633058071	0.789402173913	0.649453356214	0.159000716777	0.607312514730	97.29667812142	81.88010899182	0.7708321188373884		0.996453320179
resnet_multiclass	multiclass	2	5,00E-05	0.0	341.6307778358	0.804347826086	0.626551349525	0.164928576698	0.601381562326	97.18213058419	81.88010899182	0.7808500435565557		0.997573260837
resnet_multiclass	multiclass	2	5,00E-05	0.0	349.2771792411	0.917119565217	0.302011661555	0.087518057303	0.308796525001	98.83161512027	91.82561307901	0.9173341103424173		0.999021150933