# SCC.361 Alternative Assessment Coursework

Matthew Kenely (**38834561**)

m.kenely@lancaster.ac.uk

**Abstract**

*This project provides a brief introduction to the fields of artificial intelligence, machine learning and image classification. The performance of 4 traditional machine learning models (K-nearest neighbour, Decision Tree, Discriminant Analysis, Support Vector Machine) and 2 neural network architectures (AlexNet, VGG–19) in the task of image classification on the CIFAR-10 dataset is compared so as to gauge their relative performance and applicability to the real world. A mix of experimentation using MATLAB and theory from past publications is presented. Neural networks are found to outperform the traditional machine learning models in terms of accuracy, but are outperformed by traditional machine learning models in terms of computationally cost and training time. AlexNet is found to be the best model for the task of image classification on the CIFAR-10 dataset, with an accuracy of 70.7.%.*

## 1 Introduction

The aim of this project is to carry out research in the area of image classification (a task concerning the assignment of labels to images based on their content) through experimentation. Image classification is a subfield of computer vision, which is a field of artificial intelligence concerning the development of algorithms capable of interpreting the world through visual data such as images and videos.

**Artificial intelligence** is a scientific field that studies the creation of machines capable of performing tasks which normally require human intelligence. More specifically, it studies the synthesis and analysis of intelligently acting computational agents (agents which produce appropriate actions given specific circumstances and/or goals, are flexible to changes in their environment, learn from experience and make appropriate choices when subject to limitations) [1].

**Machine learning** is a subfield of AI that focuses on the creation and refinement of algorithms that can learn from existing data. The goal of machine learning is to create models capable of finding generalised patterns and relationships within data in order to make a correct prediction about a subject, or correctly classify said subject. [2]

A relevant subfield of machine learning used in the task of image classification is **deep learning**, which involves the use of neural networks, a class of machine learning models modelled after the structure of the biological neural networks found in the human brain [3]. Neural networks are capable of learning complex relationships within data, making them well suited to the task of image classification which involves the detection and extraction of complex features.

In this project, the performance of different machine learning models and neural network architectures will be compared when tested on a particular dataset of images. This approach to experimentation is beneficial as it provides insight into

the relative performance of these machine learning models/neural network architectures and gauges their accuracies on a level playing field. The results produced by such experimentation, however, are limited to the dataset being used, making the generalisability of the findings contingent on the variance in the dataset, and potentially reducing their applicability to the real world.

## 2 Background

Image classification is a rapidly evolving field of research. Despite their application being dormant until the mid-2000s, convolutional neural networks (a type of neural network designed to process data in a grid-like format) have seen rapid progression in recent years given the large amounts of labelled data available, and have set the current state-of-the-art in image classification [4]. The following are two recent research papers in the area of image classification which utilise CNNs:

1. **ResNet** (2015) [5]

   ResNet is a convolutional neural network architecture developed by Microsoft Research which introduced the concept of "residual connections", an architecture feature made to address the problem of vanishing gradients (a problem encountered when training artificial neural networks which use gradient-based optimisation due to updates being proportional to the partial derivative of the error function [6]).

   - Pros:
     - The residual connections used in ResNet counteract the problem of vanishing gradients.
     - ResNet architecture networks can contain many layers without risking an increase in training error.

2. **InceptionNet** (2014) [7]

   InceptionNet is a convolutional neural network architecture developed by Google Researchers which introduced the concept of "inception modules", an architecture feature which allows neural networks to learn image features at multiple levels of abstraction by using multiple filters of different sizes in a single layer.

   - Pros:
     - Inception modules allow more efficient feature extraction.
     - $1 \times 1$ convolutions allow for dimensionality reduction (reducing the number of channels in layer $n$ on which convolution must be performed to achieve the feature maps in layer $n+1$, significantly reducing the network's number of parameters and, hence, computational complexity).

The downsides to the use of ResNet and InceptionNet stem from the downsides of the use of CNNs in general, namely that they take long to train and require a large quantity of data to train effectively [8].

## 3 Methodology

All experimentation in this project is done in MATLAB. The Deep Learning and Statistics and Machine Learning toolboxes are used.

The dataset used in this project is the **CIFAR-10** dataset [9], a dataset consisting of 60000 labelled images of size 32x32, divided into 10 classes: *airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck*.

The dataset is loaded into MATLAB, reshaped into images of size 32 x 32 with 3 channels and exported into subdirectories with names corresponding to class labels to

be used in the training of the neural networks.

Two experiments are carried out to compare model performance based on the quantity of available data, as well as due to hardware limitations. Experiment 1 is carried out on 100% of the dataset, Experiment 2 is carried out on random subset of the dataset consisting of 6000 (10%) elements.

The following machine learning models are tested on the dataset:

1. K-nearest neighbour
    - New input data is labelled based on its proximity to the labelled data in the training set [10].
2. Decision Tree
    - Training data is separated into different classes by a series of lines (decision boundaries), forming a tree wherein each node corresponds to a decision and has two children nodes, traversed based on whether the input data is smaller or larger than the decision boundary.
3. Discriminant Analysis
    - A linear or quadratic function is found such that when training data is projected onto it, the distance between class means is maximised, and the scatter within classes is minimised. Input data is classified based on which side of the function it is on [11].
4. Support Vector Machine
    - A hyperplane (a plane of dimensions $n-1$ in $n$-dimensional space) is found such that the class separation of training data projected onto it is maximised. Input data is classified based on which side of the hyperplane it is on [12].
5. AlexNet
    - A convolutional neural network architecture consisting of 5 convolutional layers, 3 fully-connected layers and an output layer. It introduced the use of Rectified Linear Unit (RLU) as an activation function [13].
6. VGG–19
    - A convolutional neural network architecture consisting of 16 convolutional layers, 3 fully-connected layers and an output layer [14].

A 5:1 training/testing split is used, with 50000 images being used for training and 10000 images being used for testing. With regards to the neural networks (5, 6), the training set is further divided into a training and validation set, with 40000 images being used for training and 10000 images being used for validation.

Taking $TP$ as the number of true positives, $FP$ as the number of false positives, $TN$ as the number of true negatives and $FN$ as the number of false negatives, the following metrics are used to evaluate the performance of the models:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
$$\text{Precision} = \frac{TP}{TP + FP}$$
$$\text{Recall} = \frac{TP}{TP + FN}$$
$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Given that this experiment concerns multi-class classification, metrics must be calculated on a one v.s. all basis and then averaged to produce a single value for each metric. A macro average of this nature is produced using Eugenio Bertolini's `statsOfMeasure` function [15]. Macro average is being used given that there are an equal number of samples of each class in the dataset [9]. 6-fold cross-validation is used to evaluate the performance of models (1

– 4), i.e. the models are trained on 6 random subsets of the dataset (with overlap), with the final results being averages of the metrics calculated during each of the 6 iterations.

Confusion matrices will also be presented for each model.

An untrained AlexNet neural network architecture is used, with the final fully-connected layer, containing 1000 output nodes, being replaced with a fully-connected layer with 10 output nodes, corresponding to the 10 classes in the CIFAR-10 dataset.

Due to hardware and time limitations:

- The Support Vector Machine learning model is only trained in Experiment 2.

- The use of a pretrained VGG–19 model is attempted, with the final fully-connected layer being replaced with a new fully-connected layer with 10 output nodes, corresponding to the 10 classes in the dataset. The model is attempted to be trained using transfer learning.

# 4 Results

## 4.1 Experiment 1 - 100% of the dataset

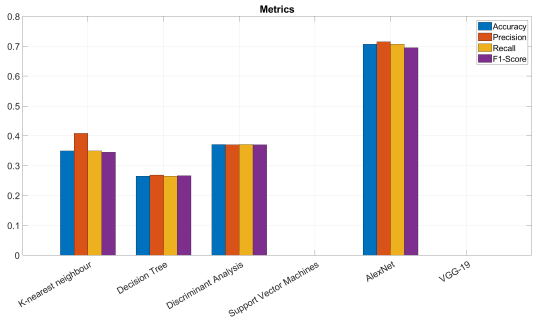


Figure 1: Metrics of K-nearest neighbour, Decision Tree, Discriminant Analysis and AlexNet models.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| K-nearest neighbour | 0.350 | 0.408 | 0.350 | 0.345 |
| Decision Tree | 0.265 | 0.269 | 0.265 | 0.266 |
| Discriminant Analysis | 0.371 | 0.370 | 0.371 | 0.370 |
| Support Vector Machines | | | | |
| AlexNet | **0.707** | **0.715** | **0.707** | **0.695** |
| VGG-19 | | | | |

Figure 2: Metrics of K-nearest neighbour, Decision Tree, Discriminant Analysis and AlexNet models (tabulated).

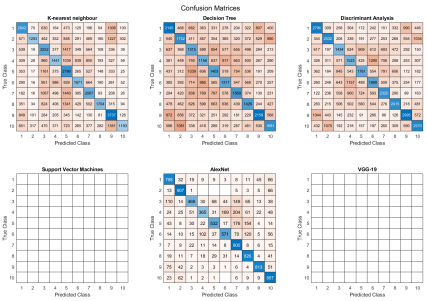


Figure 3: Confusion matrices of K-nearest neighbour, Decision Tree, Discriminant Analysis and AlexNet models.

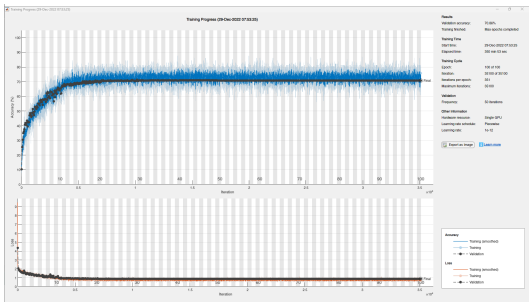

Figure 4: AlexNet learning curve (100 epochs).

## 4.2 Experiment 2 - 10% of the dataset

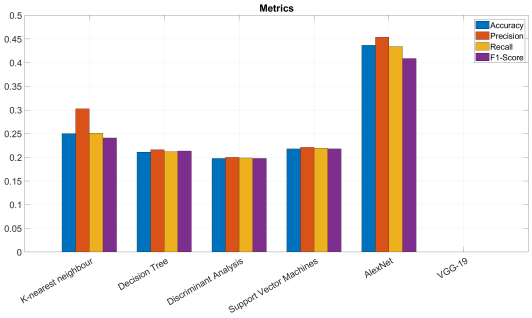


Figure 5: Metrics of K-nearest neighbour, Decision Tree, Discriminant Analysis, Support Vector Machine and AlexNet models trained on 10% of the dataset.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| K-nearest neighbour | 0.250 | 0.302 | 0.251 | 0.241 |
| Decision Tree | 0.211 | 0.216 | 0.212 | 0.214 |
| Discriminant Analysis | 0.198 | 0.200 | 0.199 | 0.198 |
| Support Vector Machines | 0.218 | 0.221 | 0.219 | 0.218 |
| AlexNet | **0.437** | **0.454** | **0.434** | **0.409** |
| VGG-19 | | | | |

Figure 6: Metrics of K-nearest neighbour, Decision Tree, Discriminant Analysis, Support Vector Machine and AlexNet models trained on 10% of the dataset (tabulated).

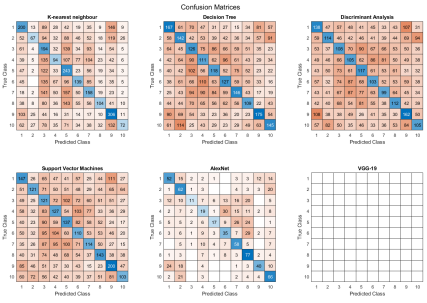


Figure 7: Confusion matrices of K-nearest neighbour, Decision Tree, Discriminant Analysis, Support Vector Machine and AlexNet models trained on 10% of the dataset.

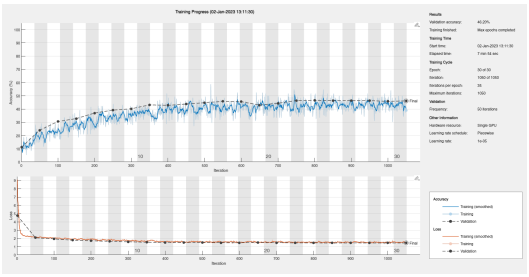

Figure 8: AlexNet learning curve when trained on 10% of the dataset (30 epochs).

## 5 Discussion

*Note: It is was intended for the performance of VGG-19 to be gauged against the other models on the CIFAR-10 dataset in this experiment, however this was rendered infeasible due to time and hardware limitations given the computational complexity of the VGG-19 architecture. Given its outperformance of AlexNet on the ImageNet dataset* [16], *it is expected that VGG-19 would also outperform AlexNet on the CIFAR-10 dataset in terms of accuracy.*

Despite the large quantity of data, it is immediately noticeable that the performance of traditional machine learning models (1 – 4) is poor. Models 1 – 3 achieve average accuracy of 32.9% (1 d.p.) when trained on the entire dataset, and the traditional machine learning models achieve an average accuracy of 21.9% when trained on 10% of the dataset. AlexNet performs significantly better than the traditional learning models, achieving an accuracy of 70.7% when trained on the entire dataset and an accuracy of 43.7% when trained on 10% of the dataset.

These results are to be expected given neural networks' ability to extract complex features given a large enough dataset [17] and the large amount of data in CIFAR-10. Neural network training effectiveness is drastically reduced when the amount of data available is small. Comparing performance between the traditional machine learning models and the neural networks in Experiment 1 and Experiment 2, the traditional machine machine learning models see an average accuracy reduction of 11.8% when trained on 10% of the dataset, whereas the neural networks see an average accuracy reduction of 26.0%.

The time taken to train the models is also of consideration. Traditional machine learning models took significantly less time to train due to the reduced computational complexity of their algorithms (namely, a significantly lower amount of parameters to be optimised). Taking one training iteration as an example, when trained on an Intel(R) Core(TM) i5-9400F, the K-nearest neighbour algorithm took 7.8 minutes to complete, whereas training AlexNet for 25 epochs (around the time the neural network started to converge) on an NVIDIA GeForce RTX 2060 took 1 hour and 27 minutes. It should be noted the use of pre-trained neural networks in conjunction with transfer learning (freezing all but the last learnable layer of the neural network), as well as parallel computing (the use of multiple GPUs in parallel to train a network) can reduce training time significantly.

Transfer learning, as well as neural networks' ability to learn complex features [17] allow them to be more generally applicable to new data.

While VGG-19 can be expected to outperform AlexNet in terms of accuracy, it must, at the minimum, be retrained using s pre-trained model and transfer learning to be able to classify into 10 categories (as is required in the CIFAR-10 dataset), and the feasibility of this training with respect to real-world applications must be considered, specifically, time constraints and access to appropriate hardware. The VGG-19 architecture is significantly more computationally complex than that of AlexNet (VGG-19 contains 19 layers and 138 million parameters, whereas AlexNet contains 8 layers and 60 million parameters), requiring a GPU with enough memory to store these parameters, as well as the use of multiple GPUs in parallel to reduce training time.

## 6 Conclusion

In this project, it has been found that the most appropriate model for classification of the CIFAR-10 dataset is the AlexNet neural network, given that it achieved the highest scores in all metrics when compared to the other models and trained in a feasible amount of time. Assuming the requirement of balance between classification accuracy and feasible real-world applicability, as well as considering the training time reduction capabilities and generalisability mentioned in Section 5, AlexNet can be expected to perform well in real-world scenarios. As stated previously, the results of this project specifically are limited by the nature of the CIFAR-10 dataset, specifically the images being of size $32 \times 32$ while AlexNet takes images of size $227 \times 227$ as input, as well as the limited number of classes within the dataset.

## References


[1] D. L. Poole and A. K. Mackworth, *Artificial Intelligence: foundations of computational agents*. Cambridge University Press, 2010.

[2] T. M. Mitchell and T. M. Mitchell, *Machine learning*. McGraw-hill New York, 1997, vol. 1, no. 9.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[6] S. Hochreiter, "Untersuchungen zu dynamischen neuronalen netzen," *Diploma, Technische Universität München*, vol. 91, no. 1, 1991.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[8] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," in *2017 international conference on communication and signal processing (ICCSP)*. IEEE, 2017, pp. 0588–0592.

[9] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[10] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," *International Statistical Review/Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989.

[11] G. J. McLachlan, *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, 2005.

[12] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[15] E. Bertolini, "Precision, specificity, sensitivity, accuracy f1-score," https://uk.mathworks.com/matlabcentral/fileexchange/86158-precision-specificity-sensitivity-accuracy-f1-score, 2022, retrieved December 30, 2022.

[16] "Imagenet benchmark (image classification) — papers with code," https://paperswithcode.com/sota/image-classification-on-imagenet, retrieved January 01, 2023.

[17] P. Wang, E. Fan, and P. Wang, "Comparative analysis of image classification algorithms based on traditional machine learning and deep learning," *Pattern Recognition Letters*, vol. 141, pp. 61–67, 2021.