FINE-TUNING DEEP LEARNING MODELS: LEARNING RATE AND BATCH SIZE ANALYSIS IN BRAIN TUMOUR CLASSIFICATION

Lau Hong Yi, BEng(DS&E), Year 2
Department of Computer Science, Faculty of Engineering, HKU
Host unit of research internship: School of Computer Science, University of St Andrews, UK
HKU supervisor: Dr. Chim Tat Wing
Internship supervisor: Dr David Harris-Birtill

INTRODUCTION

- Brain tumours are a **significant cause of cancer-related deaths** globally.
- They result from abnormal cell multiplication within the brain.
- The emergence of ChatGPT in late 2022 brought AI to the forefront, but **AI in medicine dates back to the early 2000s** and even earlier, with its roots in the 1950s
- Leveraging AI for early brain tumour detection offers the potential to save lives and streamline diagnosis.
- **Convolutional Neural Networks (CNNs)** are essential tools in medical imaging, excelling at **complex pattern recognition**.
- This study explores different CNN architectures for brain tumour detection using a dataset of 7,023 MRI images, categorized into glioma, meningioma, no tumour, and pituitary classes.

OBJECTIVE

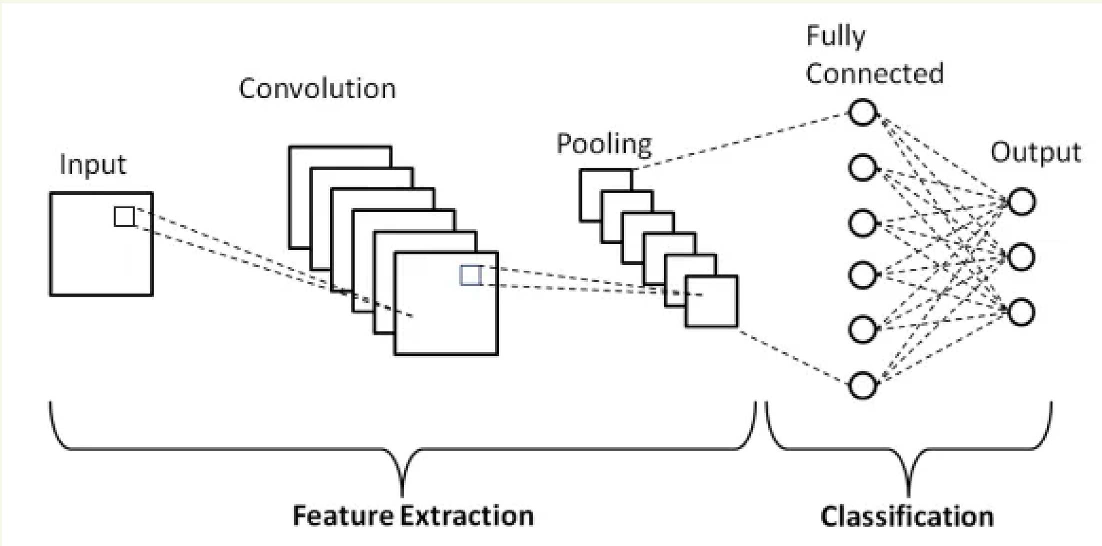
Compare the performance of two CNN architectures: EfficientNet and ResNet.

METHODOLOGY

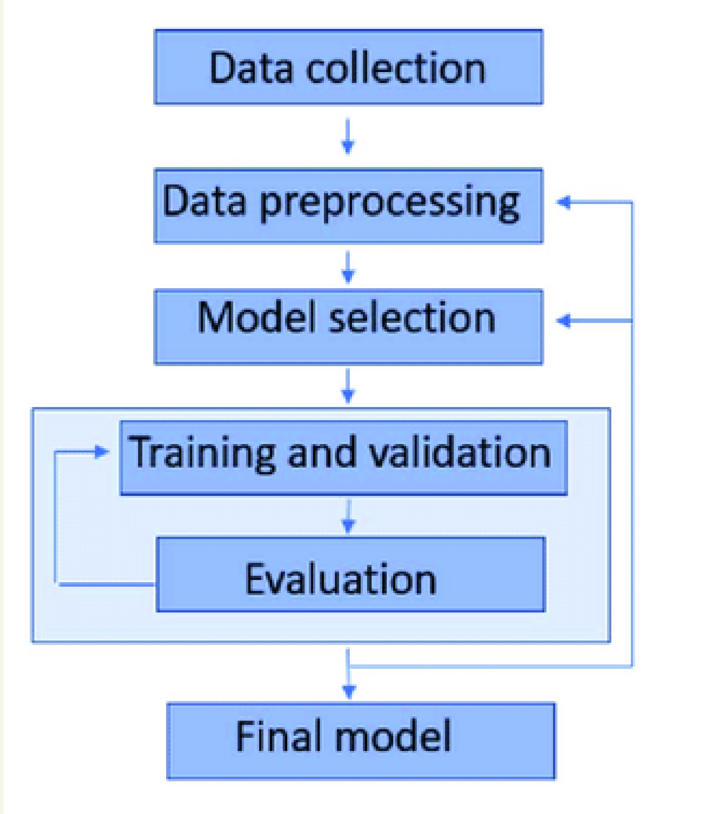
- **Dataset:** Utilized a brain tumour MRI dataset from Kaggle, consisting of approximately 7,000 labelled images -> supervised learning
- **Transfer Learning:** Employed transfer learning with pre-trained models **EfficientNet B0** and **ResNet 50**.
- **Data Split:** Divided the dataset into 5,000 images for training and validation and the remaining for testing.
- **Coding Approach:** Implemented the models using **PyTorch**.
- **Loss Function:** Applied CrossEntropyLoss.
- **Optimizer:** Used AdamW, a variant of the Adam optimizer known to promote better generalization and training loss compared to standard Adam.

Hyperparameter Exploration:

- **Epoch:** Fixed at **100** for all models
- ***Batch size:*** Explored batch sizes of **32** and **64**.
- ***Learning rate:*** Ranged **from 0.00001 to 0.01** to identify optimal values.
- ***Model Architecture:*** **All base layers** of the pretrained models were kept **frozen**. Only the output layer was modified to suit the problem

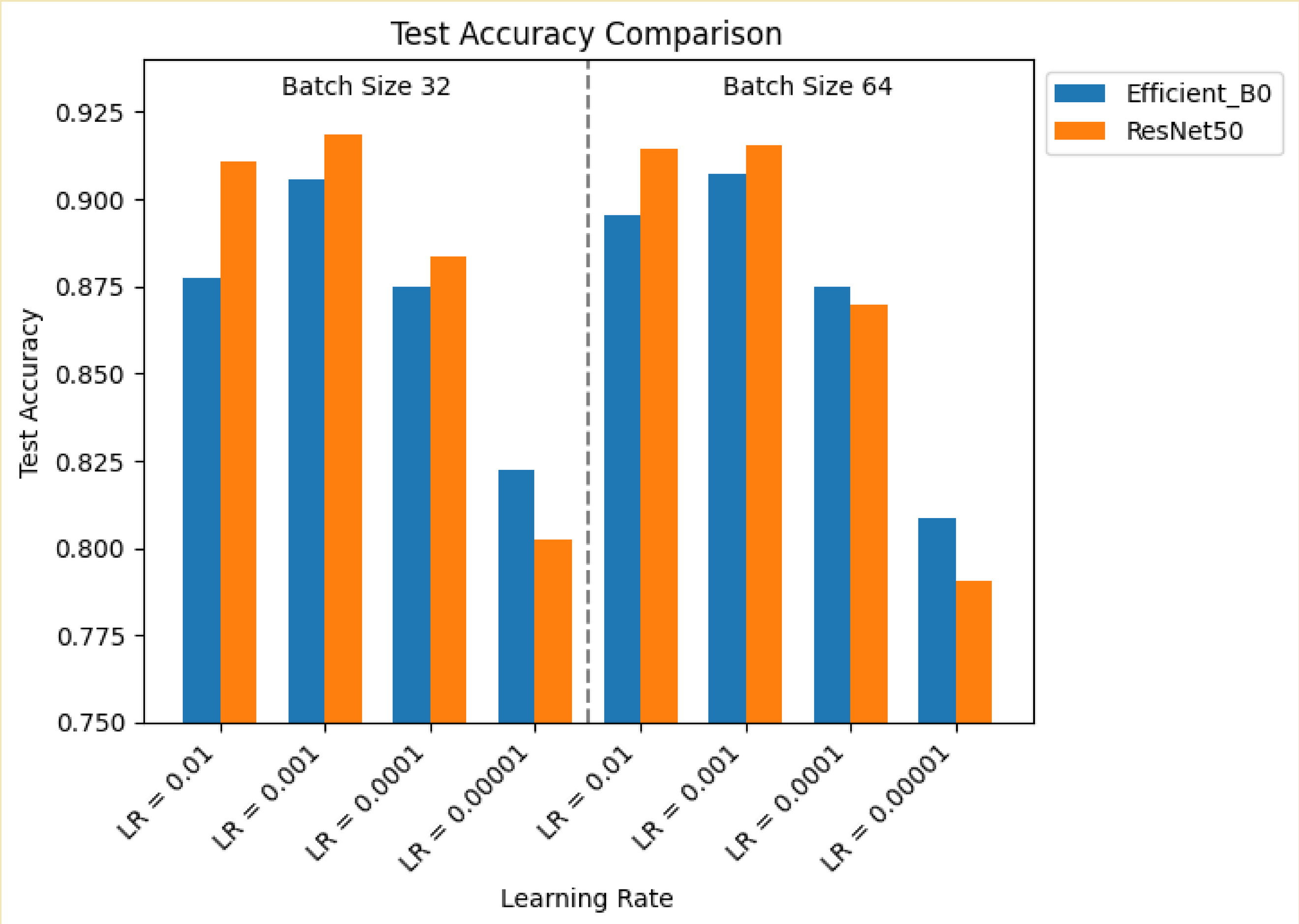


Typical CNN architecture



Typical machine learning workflow

RESULTS



1. **Optimal Learning Rate:**
   - The learning rate of **0.001** consistently produced the best-performing models across experiments.
2. **Learning Rate of 0.01:**
   - Despite **noticeable fluctuations (squiggling)** in the learning curves for models trained with a learning rate of 0.01, they consistently **ranked second in performance** across different batch sizes.
   - Potential Reason: A learning rate of 0.01 may **facilitate faster initial convergence** due to more aggressive updates, even though it leads to fluctuations in later stages.
3. **Overfitting in ResNet Models:**
   - Surprisingly, clear evidence of overfitting was observed in all the ResNet models except those trained with the smallest learning rate of 0.00001.
   - Unexpectedly, **ResNet models outperformed EfficientNet overall, despite the overfitting.**
4. **Batch Size:**
   - **No significant distinction in results** was observed between different batch sizes. Both 32 and 64 yielded similar outcomes.
   - Suggests that distinguishing between these batch sizes may not be necessary in this context.

LIMITATIONS

1. **Model Performance and Complexity:**
   - Achieving **better performance** in certain models may **not necessarily imply** that it is more **proficient at identifying complex medical images**; it could simply excel at recognizing simpler ones.
   - **Further investigation** is required to understand the model's behaviour in distinguishing between **different levels of image complexity**.
   - The study does not provide insights into the nuances of model performance in handling varying degrees of image intricacy.
2. **Model Architectures and Optimization:**
   - Model architectures were used **without substantial modifications** beyond adjusting the output layer, this approach may **not fully optimize the models** for brain tumour detection.
   - Future research opportunities should explore fine-tuning or architectural adjustments to potentially enhance model performance and adaptability to the specific problem domain.
3. **Dataset Size Consideration:**
   - The dataset used in this study, while substantial, **may not be large enough** to capture the full diversity of brain tumour cases.
   - This limitation is **common in medical data research**, where acquiring a massive dataset can be challenging.
   - Future research endeavours should aim to address this limitation by obtaining larger and more diverse datasets for improved model generalization and robustness.

CONCLUSION

- **Optimal learning rate: 0.001** consistently yielded top model performance.
- **Surprising resilience:** Models with a learning rate of **0.01**, despite fluctuations, ranked second.
- Unexpected overfitting: **ResNet models outperformed despite overfitting**, challenging complexity expectations.
- Batch size: **No significant differences between 32 and 64 batches,** suggesting efficiency with smaller batches.
- **Limitations:** Further exploration needed for **model complexity, architecture, and dataset size**.
- Future prospects: Continue refining CNNs for improved brain tumour detection, showcasing AI's potential in healthcare.

ACKNOWLEDGEMENT

I would like to thank Dr. David Harris-Birtill and his team for their patient guidance and great support, and the Laidlaw Foundation for the valuable research opportunity.